

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

**Université de Batna -2-**  
**Faculté de Technologie**  
**Département d'Électronique**



## **THÈSE**

Présentée pour l'obtention du diplôme de  
**DOCTORAT en SCIENCES**  
Spécialité: **Électronique**  
Option: **Traitement du Signal**

Par  
**GUERMOUI Mawloud**

### **Thème**

---

**Classification de personnes par utilisation des techniques de  
l'intelligence artificielle**

---

Soutenue le .../...../2017

Devant le jury:

SAIDI	Lamir	Prof.	Université Batna -2-	Président
MELAAB	Djamel	Prof.	Université Batna -2-	Rapporteur
BENAKCHA	Abdelhamid	Prof.	Université Biskra	Examineur
DEBILOU	Abderrazzak	Prof.	Université Biskra	Examineur
AOUGHLANET	Said	M.C.A.	Université Batna -2-	Invité

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

**Université de Batna -2-**  
**Faculté de Technologie**  
**Département d'Électronique**



## THÈSE

Présentée pour l'obtention du diplôme de  
**DOCTORAT en SCIENCES**  
Spécialité: **Électronique**  
Option: **Traitement du Signal**

Par  
**GUERMOUI Mawloud**

### Thème

---

**Classification de personnes par utilisation des techniques de  
l'intelligence artificielle**

---

Soutenue le .../...../2017

Devant le jury:

SAIDI	Lamir	Prof.	Université Batna -2-	Président
MELAAB	Djamel	Prof.	Université Batna -2-	Rapporteur
BENAKCHA	Abdelhamid	Prof.	Université Biskra	Examineur
DEBILOU	Abderrazzak	Prof.	Université Biskra	Examineur
AOUGHLANET	Said	M.C.A.	Université Batna -2-	Invité



---

*A mes parents*

*A mes frères et mes sœurs*

*A ma famille*

*A mes amis*

# Remerciements

*Je souhaite remercier en premier lieu mon directeur de thèse, Monsieur MELAAB Djamel, Professeur à l'Université de Batna 2 pour m'avoir accueilli au sein de son équipe. Je lui suis également reconnaissant pour le temps conséquent qu'il m'a accordé, ses qualités pédagogiques et scientifiques, sa franchise et sa sympathie. J'ai beaucoup appris à ses côtés et je lui adresse ma gratitude pour tout cela.*

*Je tiens à remercier Monsieur SAIDI Lamir, Professeur à l'Université de Batna 2 pour m'avoir fait l'honneur de participer à mon jury de thèse en tant que Président. Je le suis sincèrement reconnaissant pour toutes les remarques pertinentes et pour l'intérêt qu'il a porté à ce travail.*

*Je remercie vivement Monsieur BENAKCHA Abdelhamid, Professeur à l'Université de BISKRA, qui m'a fait un grand honneur en acceptant de juger le travail de cette thèse.*

*Mes vifs remerciements vont également à Monsieur DEBILOU Abderrazak, Professeur à l'Université de BISKRA, d'avoir accepté d'être membre du jury de cette thèse.*

*Je tiens à remercier Monsieur AOUGHLANET Said, Maître de conférences (A) à l'Université de Batna 2, d'avoir accepté d'être membre invité du jury.*



---

**Table des matières**


---

<b>Remerciements</b>	
<b>Résumé</b>	
<b>Introduction Générale</b> .....	01
<b>Chapitre I : La biométrie</b> .....	05
<b>1.1 Introduction</b> .....	06
<b>1.2 Les Systèmes biométriques et leurs modes de fonctionnements</b>	
1.2.1 Le module de capture .....	07
1.2.2 Le module d'extraction des caractéristiques .....	07
1.2.3 Le module de correspondance .....	07
1.2.4 Le module de décision .....	07
<b>1.3 Les différentes techniques de la biométrie</b> .....	08
1.3.1 Techniques intrusives .....	08
1.3.2 Techniques non intrusives .....	08
<b>1.4 Analyses biométriques</b> .....	09
1.4.1 Analyses biologiques .....	09
1.4.1.1 L'A.D.N .....	09
1.4.1.2 La reconnaissance de la thermographie faciale.....	10
1.4.2 Analyses morphologiques .....	10
1.4.2.1 Les empreintes digitales .....	10
1.4.2.2 La Géométrie de la main .....	10
1.4.2.3 La reconnaissance de la rétine .....	11
1.4.2.4 La reconnaissance de l'iris .....	12
1.4.2.5 La reconnaissance du visage .....	13
1.4.2.6 La reconnaissance vocale .....	14
1.4.2.7 La reconnaissance de la dynamique de la frappe au clavier .....	14
1.4.2.8 La reconnaissance de l'oreille .....	15
1.4.2.9 Le rythme cardiaque .....	15
1.4.3 Analyse comportementale .....	16
1.4.3.1 La reconnaissance par la signature .....	16
1.4.3.2 La reconnaissance de la démarche .....	16
<b>1.5 Architecture d'un système biométrique</b> .....	17
1.5.1 L'enrôlement .....	17
1.5.2 Le processus de vérification (contrôle) .....	18
<b>1.6 Les applications de la biométrie</b> .....	18
1.6.1 Applications commerciales .....	19
1.6.2 Applications gouvernementales .....	19
1.6.3 Applications juridiques .....	19
<b>1.7 Les limites de la biométrie</b> .....	19
1.7.1 Les limites fonctionnelles .....	19
1.7.2 Les limites techniques .....	20
<b>1.8 Conclusion</b> .....	27
<b>Chapitre II : La Reconnaissance Faciale : L'état de l'art</b> .....	27

<b>2.1 Introduction</b> .....	28
<b>2.2 Pourquoi le visage</b> .....	28
<b>2.3 Systèmes biométriques</b> .....	29
2.3.1 Les phases d'un système biométriques .....	29
2.3.2 Modules des systèmes biométriques .....	29
<b>2.4 Impact des différents types de variabilités</b> .....	29
<b>2.5 Difficultés inhérentes à la reconnaissance automatique</b> .....	30
2.5.1 Condition d'illumination .....	31
2.5.2 Les variations de la pose .....	31
2.5.3 Les expressions faciales .....	32
2.5.4 Les occlusions partielles .....	32
2.5.5 La température du corps .....	33
2.5.6 Autres difficultés .....	33
<b>2.6 Principales bases de données des visages</b> .....	34
<b>2.7 Techniques 2-D de la reconnaissance des visages</b> .....	36
2.7.1 Techniques globales .....	36
2.7.1.A Techniques linéaire .....	36
2.7.1.B Techniques non linéaire .....	38
2.7.2 Techniques locales .....	39
2.7.2.A Méthodes locales basées sur les caractéristiques d'intérêt .....	39
2.7.2.B Méthodes locales basées sur l'apparence du visage .....	42
2.7.3 Techniques hybrides .....	43
<b>2.8. Conclusion</b> .....	45
<b>Chapitre III : La reconnaissance faciale : Mise en œuvre et résultats résultats</b> .....	51
<b>3.1. Introduction</b> .....	52
<b>3.2 Extraction de caractéristiques</b> .....	52
3.2.1 Scale Invariant Feature Transform .....	52
3.2.2 Ondelettes de Gabor .....	55
3.2.3 Local Binary Pattern .....	57
3.2.4 L'analyse en Composantes Principales .....	60
3.2.5 Local Gabor Binary Patterns Histogramms .....	63
<b>3.3 Méthode Proposée</b> .....	64
3.3.1 Apprentissage du classificateur .....	66
3.3.2 Prétraitement .....	70
3.3.2 LBP Modifiée (MLBP) .....	72
3.3.4 Classification .....	74
3.3.5 Approche Sparce Representation .....	75
<b>3.4 Bases de Données utilisées</b> .....	77
3.4.1 La Base ORL .....	77
3.4.2 Base de Données Georgia Tech Visage .....	78
3.4.3 Base de Données YALE .....	78
3.4.4 Base de données YALE-B .....	79
<b>3.5 Résultats Expérimentaux</b> .....	79
3.5.1 Détection de visage .....	80
3.5.2 Prétraitement .....	80

3.5.3 Extraction des caractéristiques .....	81
3.5.4 Performance .....	83
3.5.5 L'opérateur CLBP .....	86
3.5.6 Evaluation .....	88
<b>3.6 Conclusion .....</b>	<b>90</b>
<b>Chapitre IV : Reconnaissance de l'oreille Humaine : Le savoir faire .....</b>	<b>93</b>
<b>4.1 Introduction .....</b>	<b>94</b>
<b>4.2 De qui s'agit-il ? .....</b>	<b>96</b>
<b>4.3 Travaux Précédents et Problématique de la Reconnaissance</b>	
<b>Automatique de l'Oreille Humaine .....</b>	<b>96</b>
4.3.1 Détection de l'oreille .....	97
4.3.2 Reconnaissance de l'oreille .....	103
4.3.2.1 Descripteurs Holistiques .....	103
4.3.2.2 Descripteurs Locaux .....	108
4.3.2.3 Approches Hybrides .....	109
4.3.2.4 Classifications et approches statistiques .....	110
<b>4.4 Reconnaissances d'oreilles 3D .....</b>	<b>112</b>
<b>4.5 Bases de Données Disponibles pour la Détection et la</b>	
<b>Reconnaissance des Oreilles .....</b>	<b>115</b>
4.5.1 USTB Data-Bases .....	115
4.5.2 UND Data-Bases .....	116
4.5.3 WPUT-DB Data-Bases .....	116
4.5.4 IIT Delhi Data-Bases .....	117
4.5.5 IIT Kanpur .....	117
4.5.6 La ScFace .....	118
4.5.7 Sheffield Face Data-Base .....	118
4.5.8 YSU Data-Base .....	119
4.5.7 NCKU Data-Base .....	119
4.5.8 UBEAR Data-Base .....	119
<b>4.6 Défis ouverts et Applications Futures .....</b>	<b>120</b>
4.6.1 Localisation Automatique de l'Oreille .....	120
4.6.2 Occlusion et Variations de Pose .....	121
4.6.3 Évolutivité (Scalability) .....	121
4.6.4 Comprendre la Symétrie et le Vieillessement .....	122
<b>4.7 Conclusion .....</b>	<b>122</b>
<b>Chapitre V : Reconnaissance de l'oreille humaine : Techniques</b>	
<b>Proposées et Résultats .....</b>	<b>131</b>
<b>5.1 Introduction .....</b>	<b>132</b>
<b>5.2 Les différentes étapes de notre approche d'identification .....</b>	<b>133</b>
5.2.1 Description de l'Oreille .....	133
5.2.2 Comparaison des Oreilles .....	134
<b>5.3 Bases de Données Utilisées .....</b>	<b>135</b>
5.3.1 Database1 (DB1) .....	135
5.3.2 Database2 (DB2) .....	135

---

<b>5.4 Méthodes Proposées</b> .....	135
5.4.1 Extraction des Caractéristiques .....	136
5.4.1.A Patterns of Oriented Edge Magnitudes (POEM) .....	137
5.4.1.B Co-Occurrence Local Binary Patterns (Co-LBP) .....	139
5.4.2 Théorie du Sparse Coding .....	143
5.4.3 Stratégies de Reconnaissance par la Méthode Sparse Coding .....	145
<b>5.5 Méthodes Proposées Pour la Reconnaissance de l'Oreille Humaine</b> .....	147
5.5.1 Description de la Première Méthode .....	147
5.5.2 Application et Résultats Obtenus .....	153
5.5.3 Description de la Deuxième Méthode .....	158
5.5.4 Mécanisme de reconnaissance de l'Oreille par la Méthode Proposée .....	158
5.5.5 Application et Résultats Obtenus .....	160
<b>5.6 Conclusion</b> .....	165
<b>Conclusion Générale et perspective</b>	

## Listes des Figures

**Chapitre I**

<b>Figure 1.1</b> Différentes modalités biométriques .....	08
<b>Figure 1.2</b> L'ADN .....	10
<b>Figure 1.3</b> La reconnaissance de l'empreinte digitale .....	11
<b>Figure 1.4</b> La reconnaissance de la main .....	11
<b>Figure 1.5</b> La reconnaissance de la rétine .....	12
<b>Figure 1.6</b> La reconnaissance de l'iris .....	13
<b>Figure 1.7</b> La méthode Eigen face .....	14
<b>Figure 1.8</b> La reconnaissance de la parole .....	14
<b>Figure 1.9</b> La reconnaissance de l'oreille .....	15
<b>Figure 1.10</b> La reconnaissance par rythme cardiaque .....	16
<b>Figure 1.11</b> Une signature scannée .....	16
<b>Figure 1.12</b> La reconnaissance de la démarche .....	17
<b>Figure 1.13</b> Architecture d'un système biométrique .....	18
<b>Figure 1.14</b> Distribution de l'utilisation des systèmes biométriques sur le marché mondial .....	24

**Chapitre II**

<b>Figure 2.1</b> Exemple d'un visage d'une même personne subissant un changement de luminosité .....	31
<b>Figure 2.2</b> Exemple d'un visage d'une même personne subissant des variations de pose .....	32
<b>Figure 2.3</b> Variabilité intra classe due à la présence d'expressions faciales .....	32
<b>Figure 2.4</b> Variabilité intra classe due à la présence d'occlusions partielles .....	33
<b>Figure 2.5</b> Variabilité intra classe due à des variations de la température du corps .....	33
<b>Figure 2.6</b> Localisation des caractéristiques géométriques .....	40
<b>Figure 2.7</b> Cartes de contours utilisés .....	41
<b>Figure 2.8</b> Approche de Price et Gee .....	41
<b>Figure 2.9</b> Modèle actif d'apparence .....	41
<b>Figure 2.10</b> Caractéristiques (MB) LBP pour un visage .....	43
<b>Figure 2.11</b> Graphe appliqué aux visages pour l'approche EBGm .....	44
<b>Figure 2.12</b> Classification des principaux algorithmes utilisés en reconnaissance faciale 2D .....	45

**Chapitre III**

<b>Figure 3.1</b> Construction d'un descripteur SIFT .....	54
<b>Figure 3.2</b> Exemples de détection de points d'intérêt .....	55
<b>Figure 3.3</b> Image et sa transformé de Gabor .....	56
<b>Figure 3.4</b> Opérateur LBP .....	57
<b>Figure 3.5</b> Trois voisinages pour des R et P différents .....	57
<b>Figure 3.6</b> Une image de visage et ses LBP .....	59
<b>Figure 3.7</b> Calcul du descripteur LBP .....	59
<b>Figure 3.8</b> Exemple d'une représentation faciale basée sur le LBP .....	60
<b>Figure 3.9</b> Exemple de décomposition en Eigenfaces .....	61
<b>Figure 3.10</b> Illustration Gabor-LBP .....	63
<b>Figure 3.11</b> Opérateur LTP .....	64
<b>Figure 3.12</b> Architecture générale d'un système de reconnaissance de visage .....	65
<b>Figure 3.13</b> Caractéristiques pseudo-haar à seulement deux caractéristiques .....	67
<b>Figure 3.14</b> Exemple de détection de visage .....	69
<b>Figure 3.15</b> L'apparence du visage change de manière importante cas d'illumination Variable .....	70

<b>Figure 3.16</b> Illustrations pour la détection de visage .....	71
<b>Figure 3.17</b> Les différentes étapes de l'algorithme de tan & triggs .....	72
<b>Figure 3.18</b> Deux structures différentes avec le même Code LBP .....	73
<b>Figure 3.19</b> Fonctionnement de l'approche SRC .....	76
<b>Figure 3.20</b> Vue schématique de l'approche SRC.....	76
<b>Figure 3.21</b> Échantillons de la base AT&T .....	77
<b>Figure 3.22</b> Échantillons de la base Georgia Tech .....	78
<b>Figure 3.23</b> Images d'une personne de la base Yale .....	78
<b>Figure 3.24</b> Echantillon YALE-B data base .....	79
<b>Figure 3.25</b> Viola-Jones .....	80
<b>Figure 3.26</b> La méthode de tan & triggs .....	80
<b>Figure 3.27</b> Exemple d'image faciale divisée en blocs de 8x8 .....	81
<b>Figure 3.28</b> SIFT key point's extraction and matching .....	82
<b>Figure 3.29</b> Filtre de Gabor .....	82
<b>Figure 3.30</b> La méthode proposée .....	86
<b>Figure 3.31</b> Présentation d'une image faciale par l'opérateur CLBP .....	86
<b>Figure 3.32</b> Taux de reconnaissance sur la base de données ORL .....	88
<b>Figure 3.33</b> Taux de reconnaissance sur la base de données YAL .....	89
<b>Chapitre IV</b>	
<b>Figure 4.1</b> Douze mesures dans le «système Iannarelli» .....	95
<b>Figure 4.2</b> Les différentes étapes dans l'enrôlement .....	101
<b>Figure 4.3</b> Les caractéristiques utilisées pour la formation de l'AdaBoost .....	101
<b>Figure 4.4</b> La méthode de HT .....	102
<b>Figure 4.5</b> La méthode de FF de toutes les immunoglobulines.....	104
<b>Figure 4.6</b> Les points extraits de SIFT.....	108
<b>Figure 4.7</b> Exemples de caractéristiques de surface dans des images auriculaires 3D .....	114
<b>Figure 4.8</b> Exemple de caractéristiques de patch de surface local (LSP) .....	115
<b>Figure 4.9</b> Exemples d'images de la base de données d'oreille WPUT .....	117
<b>Figure 4.10</b> Exemples d'images de la base de données de l'oreille de l'IIT de Delhi .....	117
<b>Figure 4.11</b> Exemples d'images de SCface .....	118
<b>Figure 4.12</b> Quelques exemples d'images de la base de données NICU .....	119
<b>Chapitre V</b>	
<b>Figure 5.1</b> Les différentes étapes d'un processus de classification auriculaire .....	135
<b>Figure 5.2</b> Exemple des images de quatre personnes de la base de données IIT Delhi-2 .....	137
<b>Figure 5.3</b> Processus d'extraction du POEM .....	139
<b>Figure 5.4</b> Principales étapes de l'extraction du POEM .....	140
<b>Figure 5.5</b> Différences entre l'histogramme LBP et Co-LBP .....	141
<b>Figure 5.6</b> LBP Original .....	141
<b>Figure 5.7</b> Les combinaisons LBP utilisées .....	142
<b>Figure 5.8</b> Les configurations utilisées.....	143
<b>Figure 5.9</b> Exemple d'extraction de la caractéristique proposée .....	144
<b>Figure 5.10</b> Exemple d'extraction de la caractéristique proposée .....	144
<b>Figure 5.11</b> Illustration de la stratégie de décision résiduelle adoptée .....	147
<b>Figure 5.12</b> Illustration de la stratégie des coefficients adoptée .....	148
<b>Figure 5.13</b> Modèle de travail de l'approche proposée .....	150
<b>Figure 5.14</b> Exemple de différents niveaux d'occlusion .....	156
<b>Figure 5.15</b> Exemple de différents niveaux d'oreilles occluses .....	157
<b>Figure 5.16</b> Techniques de reconnaissance d'oreille proposée .....	160



---

**Liste des Tableaux**


---

**Chapitre I**

<b>Tableau 1.1</b> : Avantages et inconvénients des différentes technologies biométriques .....	23
---	----

**Chapitre II**

<b>Tableau 2.1</b> Base de données 2D .....	35
---	----

**Chapitre III**

<b>Tableau 3.1</b> Le taux de classification de la méthode proposé .....	84
<b>Tableau 3.2</b> Le taux de reconnaissance de la méthode proposée .....	85
<b>Tableau 3.3</b> Taux de reconnaissance de la méthode proposée et algorithmes de comparaison .....	90

**Chapitre IV**

<b>Tableau 4.1</b> Exemples de taux de reconnaissance pour différents systèmes biométriques .....	94
<b>Tableau 4.2</b> Taux de détection pour différents systèmes de biométrie auriculaire .....	98
<b>Tableau 4.3</b> Taux de reconnaissance pour différents systèmes de biométrie auriculaire 2D .....	104
<b>Tableau 4.4</b> Biométrie de l'oreille 3D: étude de détection, extrait de caractéristiques et méthodes de reconnaissance .....	112

**Chapitre V**

<b>Tableau 5.1</b> Comparaison de la performance moyenne des différentes techniques de décision .....	153
<b>Tableau 5.2</b> Comparaison avec les règles de l'art dans des conditions Normales .....	154
<b>Tableau 5.3</b> Comparaison des performances de reconnaissance .....	157
<b>Tableau 5.4</b> Taux de reconnaissance de l'oreille .....	157
<b>Tableau 5.5</b> Comparaison de la performance moyenne des différentes règles de décision .....	161
<b>Tableau 5.6</b> Comparaison de la performance moyenne avec différentes Méthodes .....	162
<b>Tableau 5.7</b> Comparaison des temps d'exécution en secondes .....	162
<b>Tableau 5.8</b> Comparaison avec l'état de l'art dans des conditions normales .....	163
<b>Tableau 5.9</b> Taux de reconnaissance de l'occlusion de l'oreille .....	164

# **Introduction Générale**

## **Introduction Générale :**

La biométrie est la science qui traite l'identification d'une personne à partir d'un ou plusieurs de ses identifiants corporels. Elle commence à être déployée de façon massive dans les nouveaux passeports biométriques qui contiennent les empreintes de deux doigts et une photo numérisée du visage, codés dans une carte à puce intégrée au passeport. D'autres applications non étatiques, comme le commerce électronique, le contrôle d'accès à des zones sensibles (aéroport, centrales nucléaires), la personnalisation d'environnements, etc., sont également concernées.

Un point particulièrement crucial de ces systèmes concerne leur sécurisation. A ce propos, les chercheurs ont été conduits à réaliser des capteurs de plus en plus complexes en y intégrant des contre-mesures (mesure de la capacité superficielle de la peau, test du caractère vivant, oxymétrie, ...) et à diversifier les modalités. Pour autant, le problème de sécurisation des capteurs n'est pas encore résolu.

Pour répondre à ces besoins, la biométrie semble être une solution pratique, efficace et dont le coût en effort et en argent est en constante diminution. En effet, cette technique connaît un développement fulgurant. Cet engouement entraîne le développement de méthodes biométriques très variées: des plus classiques, comme l'étude des empreintes digitales ou de l'iris, aux plus exotiques comme la reconnaissance de la démarche, la reconnaissance de la forme de l'oreille. Les industriels proposent de plus en plus, pour les problèmes exigeant énormément de sécurité, de ne plus utiliser une seule caractéristique mais de mettre en place un système basé sur des combinaisons de différents moyens biométriques afin d'accroître encore la sécurité.

Grace à la puissance de calcul grandissante des ordinateurs, les applications biométriques sont devenues de plus en plus nombreuses et efficaces. Elles permettent d'apporter un niveau de sécurité supérieur en ce qui concerne des accès logiques (ordinateurs, comptes bancaires, données sensibles, etc.), ou des accès physiques (bâtiments sécurisés, aéroports, laboratoires, etc.). Cependant, la biométrie comporte des points d'imperfections. En effet, actuellement, il y a encore bien souvent trop peu de réflexions avant l'implémentation d'une solution biométrique, que ce soit au niveau de la méthode choisie, des contraintes imposées aux usagers ou du niveau de sécurité choisi.

L'autre point critique des systèmes biométriques concerne leur fiabilité et les mécanismes de reconnaissance ou d'authentification à mettre en œuvre.

Au départ de ce travail, nous nous sommes posés beaucoup de questions concernant les différentes modalités biométriques et la différence d'efficacité entre elles; quelle est la modalité la plus efficace et la plus facile en terme d'acquisition et d'acceptation par les utilisateurs pour l'identification ? Cependant, d'importants problèmes ont été rencontrés pour répondre à cette question. Le premier problème a été celui des bases de données. En effet, comme dans toute étude, la disponibilité et la quantité de données est souvent un problème majeur.

Dans ce travail on s'est intéressé à l'analyse de deux modalités biométrique: (I) le visage et (II) l'oreille humains pour la reconnaissance des personnes. Les raisons de ce choix sont :

- Ces deux modalités retiennent des indices biologiques très forts, contenant de nombreuses indications sur l'identité d'une personne ;
- Leurs images peuvent être acquises de manière non invasive, contrairement à l'empreinte digitale et à l'iris, par conséquent elles sont les plus acceptables parce qu'on peut les utiliser à distance sans contact avec l'objet ;
- Ces deux modalités sont complémentaires: peuvent être utilisées en mode multi biométries ;
- Utiliser une seule caméra permet d'acquérir les images de ces deux modalités. Selon le système utilisé, l'individu doit être positionné devant la caméra où en mouvement à une certaine distance de celle-ci.

Les données biométriques extraites, à partir de ces images, sont comparées au fichier référence. L'algorithme d'identification doit être en mesure d'identifier un individu malgré différents artifices physiques. Le visage est une biométrie relativement peu sûre. En effet, le signal acquis est sujet à des variations beaucoup plus élevées que d'autres caractéristiques. Celles-ci peuvent être causées, entre autres, par le maquillage, la présence ou l'absence de lunettes, le vieillissement et l'expression d'une émotion. La méthode de la reconnaissance du visage est sensible à la variation de l'éclairage et le changement de la position du visage lors de l'acquisition de l'image.

La biométrie auriculaire (l'oreille) n'est pas aussi bien connue que d'autres biométries tels que le visage, l'iris et les empreintes digitales. L'avantage de l'oreille sur les autres empreintes est qu'elle a une structure riche et stable qui ne change pas avec le temps et l'âge. Elle ne change pas non plus avec les expressions faciales d'une personne.

Des travaux récents sur la biométrie de l'oreille ont démontré le potentiel prometteur de l'oreille comme biométrie viable. En effet, les oreilles peuvent être beaucoup plus fiables qu'un visage, parce que la recherche a montré que ce dernier peut être le siège d'identification

erronée en raison de la capacité d'un sujet à changer ses expressions faciales ou manipuler son visage.

Dans ce travail de thèse, nous développerons des algorithmes d'identification biométrique à partir des images 2D du visage et de l'oreille humains en explorant des nouvelles techniques. Notre étude se base sur des techniques récentes pour l'extraction des caractéristiques et la classification.

Le manuscrit de la thèse est organisé en cinq chapitres :

Le chapitre 1 exposera la problématique générale de la reconnaissance biométrique, les différentes techniques biométriques qui existent dans la littérature, l'évaluation de ces techniques et les limitations rencontrées.

L'état de l'art de la reconnaissance faciale, les avantages et les inconvénients, le processus de la reconnaissance et les techniques utilisées, pour cette biométrie, seront présentés dans le deuxième chapitre.

Le troisième chapitre est dédié à notre approche de classification faciale 2D. Nous présenterons tout d'abord notre modèle de classification, puis, nous détaillerons les principales étapes de notre approche. Enfin, l'étude expérimentale sur les différentes bases de données et la comparaison avec les méthodes récentes de la reconnaissance faciale seront données.

Le chapitre 4 traitera l'état de l'art de la reconnaissance auriculaire. L'extraction des caractéristiques locales ou globales y seront exposées ainsi que les difficultés inhérentes à de tels systèmes.

Le cinquième chapitre est consacré à la présentation de notre système développé pour la reconnaissance auriculaire. Nous présenterons en détail les différents algorithmes proposés ainsi que les résultats obtenus sur les différentes bases de données. Nous terminerons par une conclusion générale et les perspectives.

# **Chapitre I**

## **La Biométrie**

## 1.1 Introduction

Devant la croissance exponentielle des communications tant physiques que virtuelles et les risques que cela peut représenter, il est apparu nécessaire de contrôler l'identité des acteurs de ces échanges. La biométrie permet de vérifier que l'utilisateur est bien la personne qu'il prétend être. C'est une technologie qui utilise les caractéristiques physiques propres à chaque individu pour établir de façon aussi fiable que possible son identité. Jouissant actuellement d'un certain engouement dû, sans doute, aux différents gadgets d'identification que l'on a pu voir dans certaines productions cinématographiques, la biométrie tend à envahir notre quotidien. Devant cette déferlante, il est nécessaire de faire le point sur ce qu'est exactement la biométrie, quelles techniques existent vraiment et leur degré de fiabilité pour ensuite détailler les plus utilisées.

De plus en plus, notre société éprouve le besoin de se contrôler. Que ce soit pour garantir la sécurité des gens dans les lieux publics ou pour éviter le détournement ou le vol d'informations sensibles. Il n'est pas rare qu'une personne ait à retenir plus d'une dizaine de codes d'accès tant dans son quotidien que dans son milieu professionnel et c'est souvent considéré comme frustrant par l'utilisateur. Il existe plusieurs façons de prouver son identité, mais deux grands modes se dégagent. L'identification par laquelle on cherchera à savoir si la personne est bien monsieur X en vérifiant, par rapport à un code saisi, que les données biométriques analysées sont bien en rapport. L'autre mode est l'authentification par laquelle on cherche simplement à savoir qui est la personne qui cherche à se faire reconnaître du système.

La biométrie est une mesure des caractéristiques biologiques pour l'identification ou l'authentification d'un individu à partir de certaines de ses caractéristiques: comportementales (exemple de la dynamique de frappe au clavier), physiques ou physiologiques (exemple de l'ADN).

En effet, pour qu'un système biométrique puisse fonctionner en environnement réel, les caractéristiques physiologiques doivent satisfaire les conditions suivantes [1]:

- **Etre universelles:** la caractéristique doit être possédée par chaque individu,
- **Uniques:** la caractéristique doit permettre la différenciation d'un individu par rapport à un autre.
- **Permanentes:** la caractéristique doit être invariante dans le temps.
- **Enregistrables:** la caractéristique doit pouvoir être acquise.

- **Performante:** la caractéristique doit permettre au système de reconnaître efficacement un individu, en minimisant autant que possible les fausses alarmes.
- **Acceptable:** la caractéristique doit être acceptée par les utilisateurs, ce point dépend de la façon dont est perçue la caractéristique par la population.

Les différentes caractéristiques par lesquelles il est possible d'identifier un individu sont appelées modalités biométriques.

Les plus couramment utilisées (ou étudiées) sont:

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

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

## 1.2 Les Systèmes biométriques et leurs modes de fonctionnements

En général un système biométrique est un système automatique de mesure basé sur la reconnaissance de caractéristiques propres à un individu (physique ou comportementale). Il est basé sur l'analyse de données liées à l'individu qui peuvent être classées en trois grandes catégories : analyse basée sur la morphologie, analyse de traces biologiques, l'analyse comportementale.

Il peut être représenté par quatre modules principaux :

**1.2.1 Le module de capture** est responsable de l'acquisition des données biométriques d'un individu (cela peut être un appareil photo, un lecteur d'empreintes digitales, une caméra de sécurité, etc.).

**1.2.2 Le module d'extraction de caractéristiques** prend en entrée les données biométriques acquises par le module de capture et extrait seulement l'information pertinente afin de former une nouvelle représentation des données. Généralement, cette nouvelle représentation est censée être unique pour chaque personne et relativement invariante aux variations intra-classes.

**1.2.3 Le module de correspondance** compare l'ensemble des caractéristiques extraites

avec le modèle enregistré dans la base de données du système et détermine le degré de similitude (ou de divergence) entre les deux.

**1.2.4 Le module de décision** vérifie l'identité affirmée par un utilisateur ou détermine l'identité d'une personne basée sur le degré de similitude entre les caractéristiques extraites et le(s) modèle(s) stocké(s).

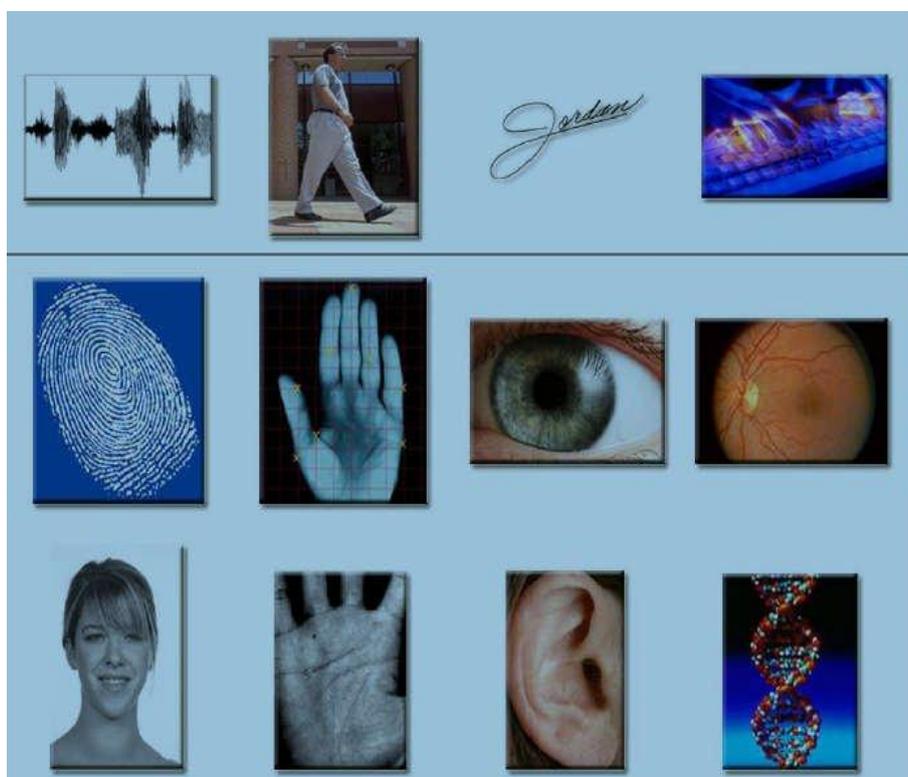
### 1.3 Les différentes techniques de la biométrie

Les techniques biométriques se divisent en deux groupes selon la coopération ou non de l'individu.

**1.3.1 Techniques intrusives:** Ces techniques requièrent un contact physique avec l'individu pour l'identifier, tel que les empreintes digitales, la rétine, l'iris ou la forme de la main. Leur usage est généralement mal accepté.

**1.3.2 Techniques non intrusives:** Ces techniques ne requièrent pas la coopération de l'individu en question. Leur application peut se faire à distance en utilisant des capteurs qui ne nécessitent pas de contact directe avec l'utilisateur (visage, démarche,...).

La biométrie permet l'identification ou l'authentification d'une personne sur la bases de données reconnaissables et vérifiables qui lui sont propres.



**Figure 1.1:** Différentes modalités biométriques.

On peut classer les techniques biométriques en trois catégories :

**A- Celles basées sur l'analyse de traces biologiques:** ce type de biométrie se fait à l'aide de l'ADN d'une personne, de son sang, ou de sa salive...

**B- Celles basées sur l'analyse comportementale:** se base sur l'analyse de certains comportements d'une personne comme le tracé de sa signature, l'empreinte de sa voix, sa démarche et sa façon de taper sur le clavier.

**C- Celles basées sur l'analyse morphologique:** est basée sur l'identification de traits physiques particuliers qui, pour toute personne, sont uniques et permanents. Cette catégorie regroupe la reconnaissance de la forme du visage, de la forme de la main, des empreintes digitales, de la rétine et de l'iris de l'œil.

Toutefois, dans un système biométrique pratique (à savoir, un système qui utilise la biométrie à des fins personnelles de reconnaissance), il y a un certain nombre d'autres questions qui devraient être considérées [2]:

- **La performance**, qui se réfère à la précision de la reconnaissance et de la vitesse possible ainsi que les ressources nécessaires pour obtenir cette précision de la reconnaissance et de la vitesse désirées. La performance se réfère également au fonctionnement et aux facteurs environnementaux qui influent sur la précision et la vitesse;

- **L'Acceptabilité**, qui indique la mesure dans laquelle les gens sont prêts à accepter l'utilisation notamment d'un identifiant biométrique (caractéristique) dans leur vie quotidienne;

- **Le contournement**, ce qui reflète la façon dont le système peut facilement être dupliqué en utilisant des méthodes frauduleuses.

## 1.4 Analyses biométriques

### 1.4.1 Analyses biologiques

#### 1.4.1.1 L'A.D.N

Présent dans les cellules du corps, il est spécifique d'un individu à un autre et permet de l'identifier de manière certaine à partir d'un simple fragment de peau, d'une trace de sang ou d'une goutte de salive.

Actuellement, le temps requis pour une analyse et le coût associé à celle-ci restreignent son utilisation dans des domaines autres que celui de l'identification judiciaire. Cependant, ce procédé biométrique fait l'objet de recherche intensive puisqu'il représente la technologie d'identification par excellence avec une marge d'erreur bien en dessous des autres moyens biométriques.



**Figure 1.2:** L'ADN.

#### **1.4.1.2 La reconnaissance de la thermographie faciale**

Une caméra infrarouge capte la chaleur émise par la peau. Contrairement à la reconnaissance faciale, on peut donc l'utiliser même dans l'obscurité ou de mauvaises conditions de visibilité. Mais les conditions de prise de vue peuvent conduire à des erreurs.

### **1.4.2 Analyses morphologiques**

#### **1.4.2.1 Les empreintes digitales**

Un système biométrique utilisant l'empreinte digitale comme moyen d'identification ou de vérification ne procède pas de la même façon, ce n'est pas l'image de l'empreinte digitale qui sert de point de comparaison, mais l'ensemble des données biométriques qui est tiré à partir des minuties de l'empreinte digitale. Les minuties représentent les fins de crêtes, les bifurcations, les lacs, les Lacs et les points qui composent l'empreinte digitale [3]. La combinaison des minuties est quasi infinie. L'acquisition des données est faite par un capteur électronique de type optique, thermique, capacitif ou à ultrasons. Cette dernière est considérée comme la plus fiable, mais aussi la plus coûteuse (voir figure I.3).

Le recours à l'empreinte digitale compte pour plus du tiers du marché des procédés biométriques. Elle représente nettement la solution préférée des entreprises œuvrant dans ce domaine. La force de ce procédé tient au fait que l'utilisation de l'empreinte digitale est plus facile à accepter par la communauté et qu'elle est une des plus efficaces et des moins coûteuses. La qualité d'image de l'empreinte digitale peut varier selon que la peau du doigt est sale, trop humide ou trop sèche, huileuse ou affligée d'une coupure [4-5].



**Figure 1.3:** La reconnaissance de l’empreinte digitale.

#### 1.4.2.2 La Géométrie de la main

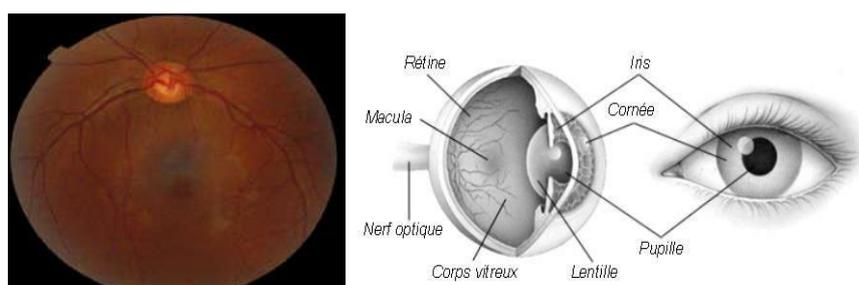
La reconnaissance de la forme de la main est considérée comme l'ancêtre des technologies biométriques. A la fin des années soixante, Robert P. Miller déposa un brevet pour un appareil permettant de mesurer des caractéristiques de la main et de les enregistrer pour une comparaison ultérieure. L'utilisateur place sa main sur un gabarit, le tout est éclairé par une lumière infrarouge et l'image est captée par une caméra digitale. Prés d'une centaine de caractéristiques sont extraites de l'image et converties en données stockées en mémoire, lors de la phase d'enrôlement ou comparées lors de la phase d'identification. Ces données concernent la longueur, la largeur et l'épaisseur de la main, de même que la forme des articulations et la longueur inter articulations [6-7].



**Figure 1.4:** La reconnaissance de la main

### 1.4.2.3 La reconnaissance de la rétine

Cette mesure biométrique est plus ancienne que celle utilisant l'iris, mais elle est très peu utilisée. Elle a été moins bien acceptée par le public et les utilisateurs, sans doute à cause de son caractère trop contraignant : la mesure doit s'effectuer à très faible distance du capteur, à peine quelques centimètres. Le capteur effectue ensuite un balayage de la rétine. Il est physiquement impossible d'effectuer une mesure rétinienne à plus de trente centimètres sur un sujet mobile, comme on peut le voir dans certains films. Cette méthode requiert des sujets coopératifs et entraînés. Pourtant cette technique semble être tout aussi fiable que celle de l'iris. Elle se base sur le fait que le schéma et le dessin formés par les vaisseaux sanguins de la rétine sont uniques pour chaque personne, différents entre jumeaux et assez stables tout au long de la vie. La mesure peut ainsi fournir jusqu'à quatre cents points caractéristiques. C'est une mesure très fiable, très difficile à falsifier. La capture de l'image est très intrusive et mal acceptée par le public [8-9].

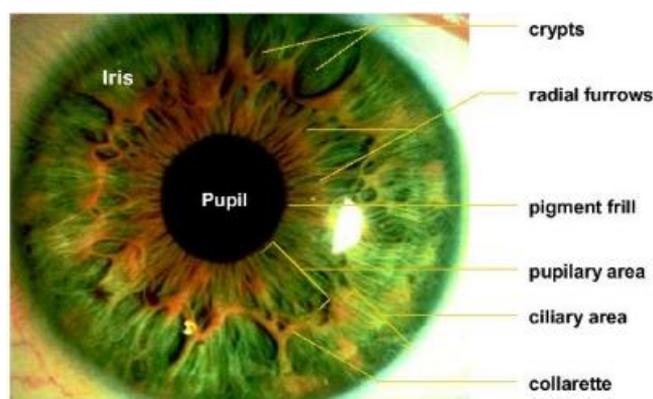


**Figure 1.5:** la reconnaissance de la rétine.

### 1.4.2.4 La reconnaissance de l'iris

La technologie de reconnaissance de l'iris, qui représente six pour cent des applications, est très intéressante pour la biométrie, car elle permet de différencier même des jumeaux ou l'œil gauche de l'œil droit d'une même personne [10]. L'iris est indépendant du code génétique de l'individu, et très difficilement falsifiable. En effet, l'iris présente une quasi-infinité de points caractéristiques qui ne varient pratiquement pas pendant la vie contrairement à sa couleur qui, elle, peut changer. Mais cela n'a aucune influence car les images d'iris obtenues par les capteurs sont en noir et blanc. Le seul problème de cette technique est lié à la mesure en elle-même, qui peut être source d'erreurs ou de problèmes. Ainsi, on peut quasiment dire que le nombre de problèmes rencontrés lors de cette mesure augmente proportionnellement avec la distance entre l'œil et la caméra. D'autres problèmes se posent à cause des reflets. Il faut être très attentif à l'éclairage. Pour palier la détection de faux yeux ou de photographies, on fait appel à des caractéristiques dynamiques de l'œil qui prouveront son authenticité, comme la

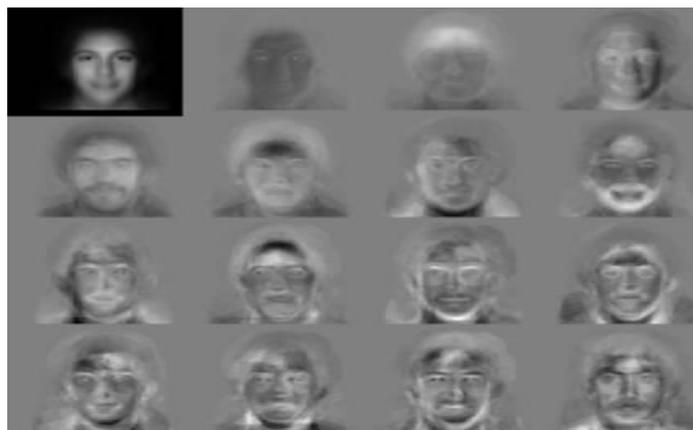
réactivité de la pupille par exemple. Cette technique est assez délicate à mettre en œuvre et très intrusive [11].



**Figure 1.6:** la reconnaissance de l'iris.

#### 1.4.2.5 La reconnaissance de visage

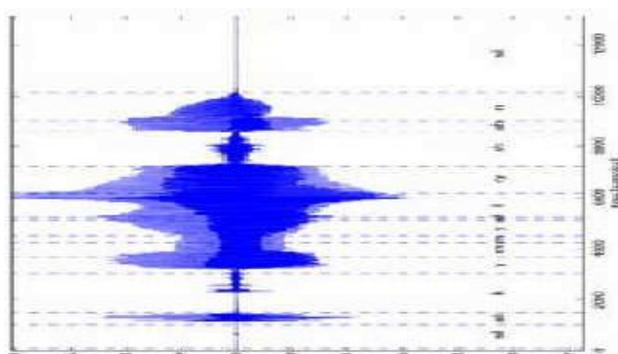
Le développement de systèmes biométriques basés sur la reconnaissance de la forme du visage est des plus récents. En 1982, deux chercheurs Hay et Young affirment que l'humain, pour reconnaître un visage, utilise les caractéristiques globales et locales qui le composent. Des recherches plus avancées furent effectuées afin de voir si cette capacité de reconnaissance pouvait être reproduite informatiquement. C'est à partir des travaux du professeur Teuvo Kohonen (1989), chercheur en réseaux neuronaux de l'Université d'Helsinki, et des travaux de Kirby et Sirovich (1989) de l'Université Brown du Rhode Island, que fut mis au point par le MIT un système de reconnaissance du visage nommé **Eigenface**. L'image du visage est captée par une caméra. Le sujet peut se présenter volontairement devant celle-ci ou encore, son image peut être capturée à son insu pour en dégager certaines particularités. Selon le système utilisé, l'individu doit être positionné devant l'appareil ou peut être en mouvement à une certaine distance. Les données biométriques qui sont obtenues sont par la suite comparées au fichier référence [12-13].



**Figure 1.7:** La méthode Eigen face.

#### 1.4.2.6 La reconnaissance vocale

C'est en 1962 que Lawrence Kersta, un ingénieur du bel Laboratories, établit que la voix de chaque personne est unique et qu'il est possible de la représenter graphiquement (voire figure 1.8). La voix est constituée de composantes physiologiques et comportementales. La mesure biométrique de la voix traite des données qui proviennent à la fois de facteurs physiologiques dépendants de l'âge, du sexe, de la tonalité, de l'accent et de facteurs comportementaux comme la vitesse et le rythme. Ils ne sont en général pas imitables. C'est la seule technique qui permette à l'heure actuelle de reconnaître une personne à distance. Cependant cette technique est très facilement falsifiable, en utilisant un enregistrement. Cette technique nécessite en plus une excellente qualité d'enregistrement ce qui n'est pas possible dans les lieux publics par exemple. Le peu de différences qui existe entre deux voix rend également cette technique peu fiable [14].



**Figure 1.8:** la reconnaissance de la parole

#### 1.4.2.7 La reconnaissance de la dynamique de la frappe au clavier

Le système est basé sur la dynamique de frappe au clavier, il ne nécessite aucun équipement

particulier, chaque ordinateur disposant d'un clavier. Il s'agit d'un dispositif logiciel qui calcule le temps où un doigt effectue une pression sur une touche et le temps où il est levé (entre les frappes).

Cette mesure est capturée environ 1000 fois par seconde. La séquence de frappe est prédéterminée sous la forme d'un mot de passe. Initialement l'utilisateur doit composer son mot de passe à quelques reprises afin que soit constitué un gabarit de référence [15].

#### 1.4.2.8 La reconnaissance de l'oreille

A priori, la technique serait efficace, car il n'existe pas deux formes d'oreilles identiques. Cependant il n'existe encore aucune application commerciale.



Figure 1.9: La reconnaissance de l'oreille

#### 1.4.2.9 Rythme cardiaque

Comme les empreintes digitales, le rythme cardiaque est propre à chaque personne et doit donc permettre de l'identifier.

En effet, des chercheurs ont découvert que le rythme cardiaque, et plus précisément la forme des pics d'un électrocardiogramme, sont propres à chaque individu. Une chercheuse a mis au point un capteur cardiaque, HeartID, identifiant en 1,2 seconde et qui peut être intégré dans tout type d'appareil électronique pour servir de système d'authentification biométrique. On pourrait le voir arriver sur les Smartphones, tablettes et consoles de jeu dans un avenir proche.

Différents laboratoires de recherche situés en Europe et en Amérique du Nord travaillant sur ce sujet confirment que ce profil cardiaque ne varie pas avec l'âge ni même si le rythme du cœur s'emballé suite à un effort ou une émotion. Par ailleurs, les progrès techniques récents sur les électrocardiogrammes ont permis de mettre au point des capteurs miniaturisés et bon marché qui fonctionnent à partir du bout des doigts. L'idée d'utiliser le rythme cardiaque comme outil biométrique a donc rapidement fait son chemin.



**Figure 1.10:** Aperçu de l'interface du système d'identification HeartID. Lorsque l'utilisateur place ses doigts sur les capteurs de rythme cardiaque, un algorithme de reconnaissance traite le signal et le compare avec l'électrocardiogramme de la personne préalablement enregistré.

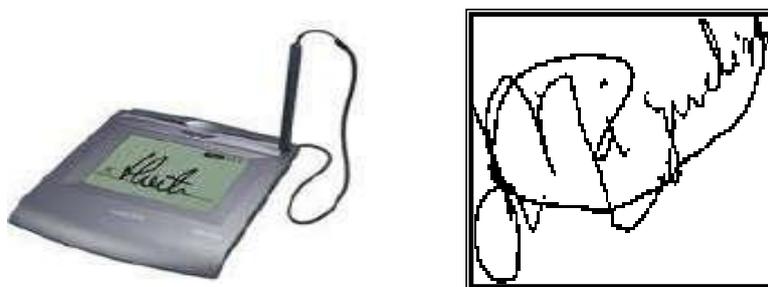
### 1.4.3. Analyse comportementale

#### 1.4.3.1 La reconnaissance de la signature

Des 1929, Osborn établit que l'écriture dépend de plusieurs facteurs caractéristiques. Pour imiter une signature, il faut donc non seulement imiter la forme de l'écriture mais aussi tenir compte de ces facteurs liés notamment à la vitesse, aux conditions environnantes et à la dextérité musculaire. Par la suite, diverses techniques de reconnaissance de la signature furent mises au moins au bénéfice notamment des banques et des corps policiers.

Les systèmes de reconnaissance de l'écriture, analysent les caractéristiques spécifiques d'une signature comme la vitesse, la pression sur le crayon, le mouvement, les points et les intervalles de temps où le crayon est levé [15-16].

L'utilisateur de cette technologie signe généralement avec un stylo électronique sur une tablette graphique. Ces données sont enregistrées pour comparaison ultérieure. Certains systèmes ne font qu'enregistrer l'image statique de la signature pour la comparaison.



**Figure 1.11:** une signature scannée

### I.4.3.2 La reconnaissance de La démarche

L'attraction potentielle de l'identification de démarche se situe dans la capacité d'identifier un individu à distance. Cependant, il y a des défis sérieux à surmonter à cet égard. L'idée qu'un individu marche typiquement avec une démarche unique est intéressante et, sous des conditions de laboratoire, le concept de l'identification de démarche peut être démontré. Cependant, la vie réelle est pleine de désaccords dynamiques qui rendent l'exécution d'un tel système particulièrement difficile.

En plus des complexités de comparaison, il y a des facteurs tels que l'occasion de saisir l'image mobile d'un individu en isolement et dont le détail est suffisant pour pouvoir entreprendre une telle comparaison. La création d'un modèle fiable est également quelque chose qui présente de vrais défis. L'identification de la démarche représente un exemple intéressant de la recherche biométrique conduite par une condition perçue : dans ce cas-ci, pour identifier un individu à une distance au-delà de laquelle la biométrie de contact et à bout-portant ne peuvent fonctionner. C'est peut-être une idée attrayante pour des applications militaires et de très haute sécurité, mais il est douteux que l'identification par la démarche deviendra une technique biométrique courante [17].

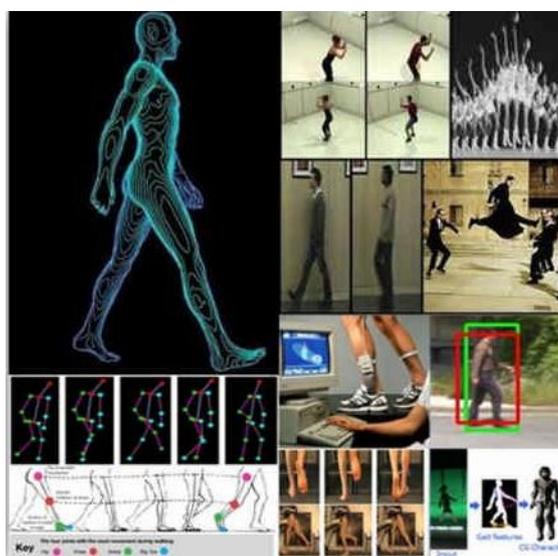


Figure 1.12: la reconnaissance de la démarche.

## 1.5 Architecture d'un système biométrique

Le système biométrique repose sur deux processus : celui dit d'enrôlement biométrique et celui de recherche et de vérification (contrôle).

**1.5.1 L'enrôlement** : cette étape sert à créer la base de données de référence. Ses étapes sont multiples et ordonnées :

capture de l'échantillon biométrique, extraction des données décrivant les caractéristiques de cet échantillon, création d'un gabarit reproduisant l'ensemble des données de l'échantillon original, mise en mémoire de celui-ci sur une base de données centralisée ou dans un dispositif tel que disque dur, carte à puce, code barre.

**1.5.2 Le processus de vérification (contrôle):** cette étape permet de comparer une donnée d'utilisateur à une donnée de référence. Il se décompose de la façon suivante : utilisation du dispositif biométrique pour la capture d'un échantillon, extraction des données numériques, création d'un candidat gabarit reproduisant l'ensemble des données caractéristiques de l'échantillon original qui servira à effectuer la recherche et la vérification, placement du candidat gabarit dans le moteur de vérification biométrique pour qu'il puisse être comparé avec le gabarit biométrique original.

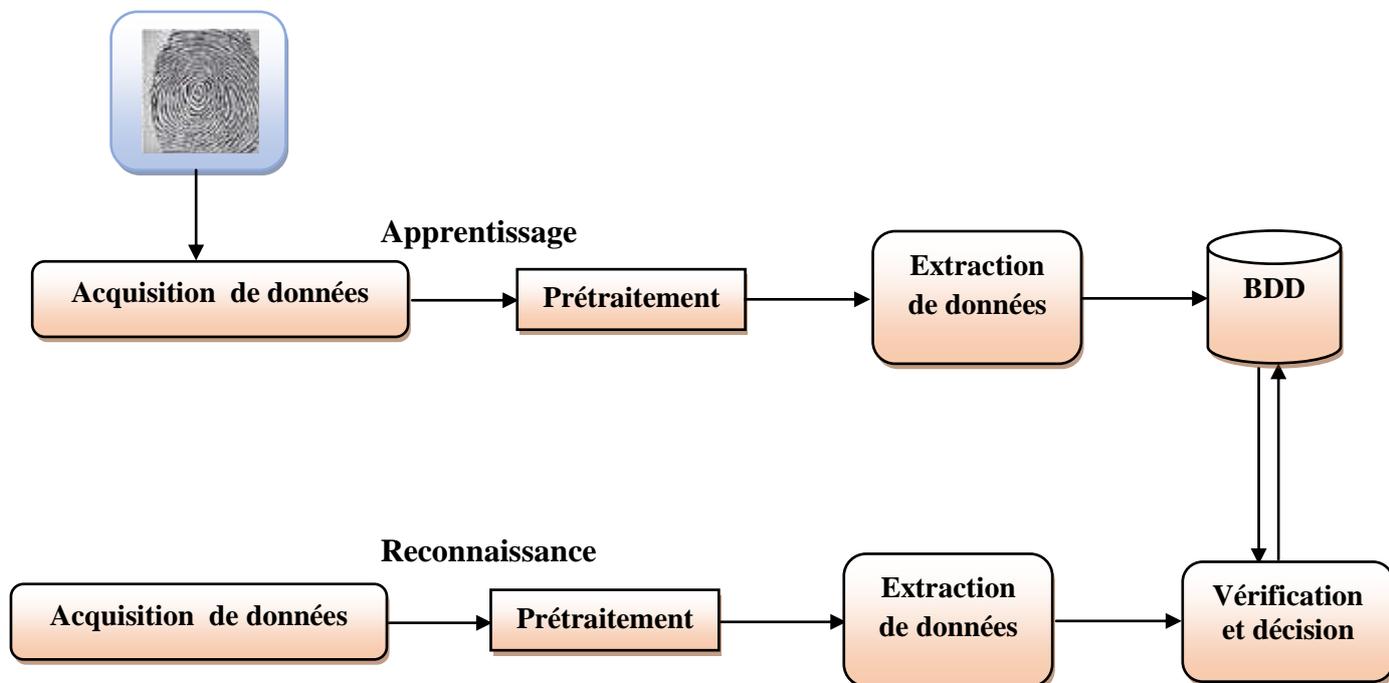


Figure 1.13: Architecture d'un système biométrique.

## 1.6 Les applications de la biométrie

Aujourd'hui, les principales applications sont la production de titres d'identités, le contrôle d'accès à des sites sécurisés, le contrôle des frontières, l'accès aux réseaux, systèmes d'information et stations de travail, le paiement électronique, la signature électronique et même le chiffrement de données. Cette liste n'est pas exhaustive, et de nouvelles applications vont très certainement voir rapidement le jour.

Les techniques biométriques sont appliquées dans plusieurs domaines et leur champ d'application couvre potentiellement tous les domaines de la sécurité où il est nécessaire de

connaître l'identité des personnes. Les applications peuvent être divisées en trois groupes principaux :

**1.6.1 Application commerciales:** telles que l'accès au réseau informatique, la sécurité de données électroniques, le commerce électronique, l'accès d'internet, l'ATM, la carte de crédit, le contrôle d'accès physique, le téléphone portable, le PDA, la gestion des registres médicales, l'étude de distances, etc....

**1.6.2 Applications gouvernementale:** telles que la carte nationale d'identifications, le permis de conduite, la sécurité sociale, le contrôle de passeport, etc....

**1.6.3 Applications juridiques:** telles que l'identification de cadavre, la recherche criminelle, l'identification de terroriste, les enfants disparus, etc.

**I.7 Les limite de la biométrie:** La biométrie présente malheureusement un certain nombre d'inconvénients parmi eux : le problème de la qualité de l'authentification. Ces méthodes ne sont en effet pas toujours fiables à 100%, ce qui empêche des utilisateurs de bonne foi d'accéder à leur système. Car il s'agit bien là d'une des caractéristiques majeures de tout organisme vivant : on s'adapte à l'environnement, on vieillit, on subit des traumatismes plus ou moins importants, bref on évolue et les mesures changent:

Prenons le cas le plus simple, celui des empreintes digitales (mais la même chose s'applique à toute donnée physique). Suivant les cas, nous présentons plus ou moins de transpiration, la température des doigts n'est pas régulière. Il suffit de se couper pour présenter une anomalie dans le dessin de ses empreintes. Dans la majorité des cas, les mesures du capteur et du logiciel associé retourneront un résultat différent de la mesure initiale de référence. Or, il faut pourtant bien réussir à se faire reconnaître. En pratique, cela sera réalisé dans la plupart des cas car le système est amené à autoriser une marge d'erreur entre la mesure et la référence. De manière générale, les faiblesses de ces systèmes ne se situent pas au niveau de la particularité physique sur laquelle ils reposent, mais bien sur la façon avec laquelle ils la mesurent, et la marge d'erreur qu'ils autorisent. Là encore, il convient de ne pas se laisser impressionner par une image illusoire de haute technologie - produit miracle. De plus, les experts techniques mettent au passif cette technologie, d'une part, son coût, d'autre part, la question de sa révocation. En effet, confronté à une personne qui a subtilisé un mot de passe ou une signature manuscrite, le titulaire du mot de passe ou de la signature peut facilement les remplacer ou les révoquer. La chose semble plus complexe pour une empreinte digitale ou rétinienne. Si un tiers s'approprie une identité biométrique du type empreintes digitales ou

identité visuelle, il peut au moyen de ces identités biométriques passer tout type d'actes au nom de la victime. Comment la victime pourrait-elle alors révoquer sa propre empreinte digitale ou identité visuelle ? Les experts en sécurité sont partagés sur la question, même si, en majorité, ils semblent considérer que cette révocation est possible. Tous reconnaissent cependant la difficulté à mettre au passif cette protection technique.

Les données biométriques sont comparables à tout autre système de contrôle d'accès comme des mots de passe, ...etc. Car du point de vue du système informatique, ce ne sont rien d'autres que des séries de bits comme toute donnée. Autrement dit, la difficulté réside dans la contrefaçon de la caractéristique physique et biologique que l'on mesure. Si la biométrie se généralise dans notre environnement, il est dangereux de penser qu'il s'agit de la réponse à tous les problèmes de sécurité. La biométrie, de par ses limites fonctionnelles, techniques et juridiques n'est en aucun cas synonyme de technologie miracle et de sécurité absolue.

### **1.7.1 Les limites fonctionnelles**

Les systèmes d'authentification biométrique représentent une grande partie des limites fonctionnelles. En effet, les systèmes biométriques laissent la place à un certain nombre de faux rejets et de fausses acceptations. Ils ne peuvent à eux seuls garantir à 100% que seules les personnes autorisées pourront passer le contrôle. Ils ne peuvent même pas garantir qu'une personne autorisée ne sera pas rejetée par le système. Il y aura toujours une marge d'erreur à prendre en compte, ce qui n'est pas forcément très rassurant.

### **1.7.2 Les limites techniques**

Bien que cela représente un travail assez conséquent, les données biométriques peuvent être imitées, notamment celles qui laissent des traces sur le passage de l'individu telles que les empreintes digitales. Un individu mal intentionné peut récupérer les empreintes digitales sur un objet tenu par la victime, les imiter et tenter de passer le contrôle biométrique à l'aide de ces empreintes. De plus, les données biométriques sont dans la majeure partie des cas numérisées sur un support, de préférence individuel. Si ce support n'est pas protégé contre les intrusions et le piratage, tout le système biométrique tombe à l'eau.

Nous dressons dans le tableau suivant les avantages et les inconvénients des différentes technologies biométriques :

**Tableau 1.2 :** Avantages et inconvénients des différentes technologies biométriques.

Modalité	Avantages	Inconvénients
<b>L'iris</b>	<ul style="list-style-type: none"> <li>● L'iris n'est pas modifiable même par intervention chirurgicale ;</li> <li>● Les iris sont uniques et différent même pour les vrais jumeaux ;</li> <li>● Grande quantité d'information contenue dans l'iris ;</li> <li>● Iris très difficilement falsifiable.</li> <li>● Dessin de l'iris indépendant du code génétique ;</li> <li>● L'iris ne varie presque pas au cours d'une vie.</li> </ul>	<ul style="list-style-type: none"> <li>● L'iris est aisément visible et peut être photographié. Le problème de sécurité est alors lié aux vérifications effectuées lors de la prise de vue. (Problème identique pour les empreintes, la voix, l'oreille,... Mais moins pour la rétine) ;</li> <li>● Aspect psychologiquement invasif de la méthode ;</li> <li>● Des problèmes peuvent survenir lors de la mesure (reflet, variation de la taille de la pupille, etc.). Une photo ou une lentille de contact reproduisant l'image de l'iris peut affecter la fiabilité.</li> </ul>
<b>La rétine</b>	<ul style="list-style-type: none"> <li>● L'empreinte rétinienne est peu exposée aux blessures (coupure, brûlure) ;</li> <li>● Les taux de faux rejet et de fausse acceptation sont faibles ;</li> <li>● Très difficile, voir impossible, à imiter ;</li> <li>● La rétine est différente chez les vrais jumeaux ;</li> <li>● La rétine est stable durant la vie d'un individu ;</li> <li>● très efficace. Carte vasculaire propre à chaque individu et différente, même entre jumeaux.</li> <li>● Haute sécurité.</li> </ul>	<ul style="list-style-type: none"> <li>● Système intrusif, il faut placer l'œil près du capteur ;</li> <li>● Mauvaise acceptation du public (l'œil est un organe sensible) ;</li> <li>● Coût plus important que d'autres technologies ;</li> <li>● Pas adapté pour un flux de passage important ;</li> <li>● Technique contraignante pour les participants (mesure à courte distance [quelques centimètres] du capteur) ;</li> <li>● Technique invasive et peu acceptée par le public ;</li> <li>● L'aspect des vaisseaux sanguins peut être modifié par la maladie ou l'âge.</li> </ul>
<b>Le visage</b>	<ul style="list-style-type: none"> <li>● Le facial-scan fait par caméra photo est réputé pour être la technique la plus simple et la moins contraignante. Son principal avantage est son côté peu intrusif ;</li> <li>● Comparable à se faire prendre en photo, il est relativement mieux accepté socialement ;</li> <li>● Seule une opération chirurgicale modifiant la forme du visage (ajout de prothèses, transformation du cartilage, etc.) peut affecter la fiabilité. Seule technique utilisable sans le consentement de la personne ;</li> <li>● utilisation facile.</li> </ul>	<ul style="list-style-type: none"> <li>● La reconnaissance des visages ne fonctionnant pas bien incluent l'éclairage pauvre, les lunettes de soleil, les longs cheveux, ou d'autres objets couvrant partiellement le visage du sujet, et des images à basse résolution ;</li> <li>● la distance pour la capture de l'image du visage n'est pas inconfortable ;</li> <li>● Un autre inconvénient sérieux est que beaucoup de systèmes sont moins efficaces si les expressions du visage varient. Même un grand sourire peut rendre le système moins efficace. Par exemple: Le Canada permet maintenant seulement des expressions du visage neutres en photos de passeport ;</li> <li>● Technique qui ne permet pas d'identifier des personnes en</li> </ul>

		<p>mouvement ;</p> <ul style="list-style-type: none"> <li>● Impossibilité de différencier des jumeaux ;</li> <li>● Peu d'efficacité ;</li> <li>● Sensibilité à la variation de l'éclairage et au changement de la position du visage.</li> </ul>
<b>Les empreintes digitales</b>	<ul style="list-style-type: none"> <li>● La technologie la plus éprouvée techniquement et la plus connue du grand public ;</li> <li>● Petite taille du lecteur facilitant son intégration dans la majorité des applications (téléphones portables, PC) ;</li> <li>● Faible coût des lecteurs grâce aux nouveaux capteurs de type "Chip silicium" ;</li> <li>● Traitement rapide</li> <li>● Bon compromis entre le taux de faux rejet et le taux de fausse acceptation.</li> </ul>	<ul style="list-style-type: none"> <li>● Image "policière" des empreintes digitales. Besoin de la coopération de l'utilisateur (pose correcte du doigt sur le lecteur) ;</li> <li>● Certains systèmes peuvent accepter un moulage de doigt ou un doigt coupé (la détection du doigt vivant permet d'éviter ce type d'usurpation) ;</li> <li>● Difficulté de lecture: sensibilité aux altérations pouvant survenir au cours de la vie (égratignure, cicatrice, vieillissement ou autres) et à certaines variations (température, humidité, saleté).</li> </ul>
<b>Forme de la main</b>	<ul style="list-style-type: none"> <li>● Bonne acceptation des usagés ;</li> <li>● Très simple à utiliser ;</li> <li>● Le résultat est indépendant de l'humidité et de l'état de propreté des doigts ;</li> <li>● Fichier "gabarit" de petite taille.</li> <li>● Technique moins coûteuse que la reconnaissance des empreintes digitales: insensibilité à la poussière, aux coupures au doigt, etc ;</li> <li>● D'utilisation très simple.</li> </ul>	<ul style="list-style-type: none"> <li>● Trop encombrant pour un usage sur le bureau, dans une voiture ou un téléphone ;</li> <li>● Risque de fausse acceptation pour des jumeaux ou des membres d'une même famille ;</li> <li>● La forme de la main ou des doigts se modifie avec le vieillissement, ce qui nuit à la mesure à long terme.</li> </ul>
<b>La voix</b>	<ul style="list-style-type: none"> <li>● Technologie biométrique facile à mettre en œuvre ;</li> <li>● Permet de sécuriser une conversation téléphonique ;</li> <li>● Généralement très bien acceptée car la voix est un signal naturel à produire ;</li> <li>● Une des seules techniques permettant de reconnaître quelqu'un à distance et la Seule ;</li> <li>● utilisée pour la reconnaissance par téléphone.</li> </ul>	<ul style="list-style-type: none"> <li>● La voix n'est pas un attribut permanent (elle change bien entendu avec l'âge) ;</li> <li>● Technologie biométrique vulnérable aux attaques ;</li> <li>● Il est très facile d'enregistrer ou de reproduire la voix. - Nécessite une excellente qualité audio ;</li> <li>● Sensible aux bruits ambiants.</li> <li>● La voix change dans le temps et peut être altérée (rhume, fatigue, forte émotion, etc.) ;</li> <li>● Faible niveau de différenciation entre deux voix ;</li> <li>● Taux élevés de faux rejets et de fausses acceptations.</li> </ul>
<b>Frappe du clavier</b>	<ul style="list-style-type: none"> <li>● Non intrusif, geste naturel pour un individu ;</li> <li>● Pas de matériel supplémentaire, un simple logiciel suffit ;</li> <li>● Mise en œuvre rapide pour un grand nombre d'utilisateur ;</li> <li>● Réduit sensiblement la nécessité de changement de mot de passe et la sollicitation des services</li> </ul>	<ul style="list-style-type: none"> <li>● L'état de santé et la fatigue peuvent altérer la façon de frapper les touches ;</li> <li>● Sensibilité à la différence entre les claviers.</li> </ul>

	informatique ; <ul style="list-style-type: none"> <li>● Permet d'identifier une personne à distance, à partir de son ordinateur.</li> </ul>	
<b>signature</b>	<ul style="list-style-type: none"> <li>● La signature écrite sur un document peut être conservée des certains documents ;</li> <li>● Action qui implique (responsabilité) le demandeur ;</li> <li>● Facile à utiliser. Très acceptée par les usagers.</li> </ul>	<ul style="list-style-type: none"> <li>● Besoin d'une tablette graphique ;</li> <li>● Sensible aux émotions de l'individu ;</li> <li>● Pas utilisable pour du contrôle d'accès en extérieur par exemple ;</li> <li>● Technique peu utilisée jusqu'à maintenant ;</li> <li>● La signature étant changeante, une combinaison de données (vitesse d'exécution ou autres) est nécessaire.</li> </ul>
<b>ADN</b>	<ul style="list-style-type: none"> <li>● Une Très grande précision ;</li> <li>● Il impossible que le système a fait des erreurs ;</li> <li>● Il est standardisé.</li> </ul>	<ul style="list-style-type: none"> <li>● Très cher.</li> </ul>

Cette comparaison nous permet de choisir une technologie appropriée en fonction des contraintes liées à l'application demandée. Par exemple, on remarque que l'iris et l'empreinte digitale sont les modalités les plus discriminantes. Cela est utile pour les systèmes d'identification à grande-échelle nécessitant un haut niveau de sécurité.

Une brève comparaison des techniques biométriques les plus utilisées ci-dessus basées sur sept facteurs est fournie dans le tableau. L'applicabilité d'une technique biométrique spécifique dépend fortement des conditions du domaine d'application. Par exemple, il est bien connu que la technique basée sur l'empreinte digitale est plus précise que la technique basée sur la voix. Cependant, dans une application de transaction bancaire à distance, la technique basée sur la voix peut être préférée puisqu'elle peut être intégrée dans le système de téléphone existant.

Une question qui se pose souvent dans ce domaine est la suivante :

**« Quelle est la meilleure technique biométrique ? »**

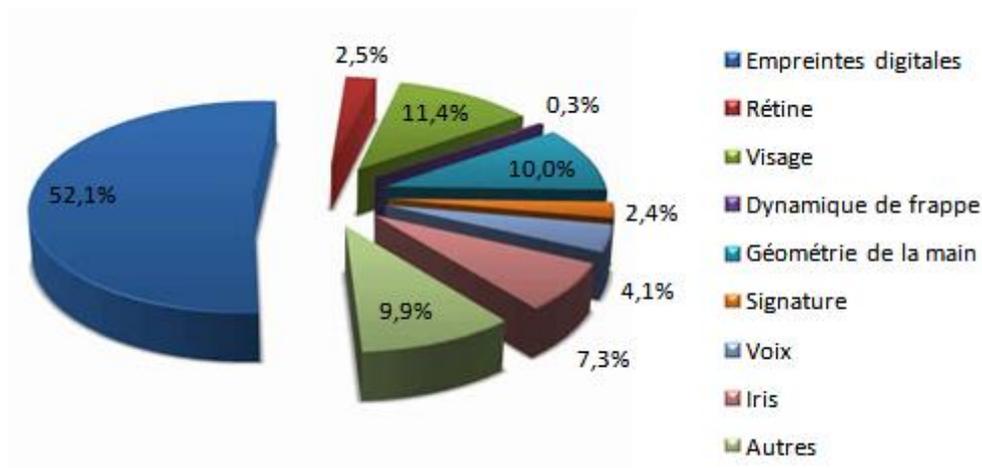
La réponse naturellement est qu'il n'y a aucune meilleure technique biométrique en termes absolus, tout dépend de la nature précise de l'application et des raisons de son exécution. Néanmoins, nous pouvons analyser leur distribution du point de vue utilisation, et les comparer dans leurs propres contextes selon leurs critères de performance.

Le Tableau I.2 montre qu'il n'existe pas de méthode idéale. Les méthodes se divisent en deux grands groupes.

Actuellement, pour la mise en place des grands projets de passeports biométriques, les

systèmes retenus par l'Europe semble être un stockage de la photo d'identité, des empreintes digitales et de l'iris sous forme numérique. A noter que le choix du ou des dispositifs biométriques peut aussi dépendre de la culture locale. Ainsi, en Asie, les méthodes nécessitant un contact physique comme les empreintes digitales sont rejetées pour des raisons d'hygiène alors que les méthodes basées sur l'iris sont très bien acceptées. La biométrie basée sur l'image rétinienne est très précise et très fiable mais c'est la méthode la plus intrusive et nécessite un effort et un coût importants par rapport aux autres méthodes.

La figure I.14 illustre une comparaison entre les systèmes biométriques les plus répandus. Nous pouvons constater que l'utilisation du système biométrique rétinien est rare malgré son excellente efficacité. Le retard de cette technique est dû surtout, comme déjà évoqué, au caractère invasif de l'acquisition de l'image rétinienne.



**Figure 1.14:** Distribution de l'utilisation des systèmes biométriques sur le marché mondial.

## 1.8 Conclusion

En résumé, l'exigence accrue pour des systèmes d'authentification fiables et commodes, la disponibilité des ressources informatiques peu coûteuses, le développement des capteurs biométriques bon marché, et les avancements dans le traitement du signal, ont contribué au déploiement rapide des systèmes biométriques dans les établissements s'étendant des épiceries aux aéroports.

Dans ce chapitre nous avons décrit les technologies utilisées dans les systèmes biométriques pour l'identification des personnes, leurs architectures et leurs différentes applications, ainsi nous avons donné un aperçu sur les différentes modalités biométriques tout en soulignant les avantages et les inconvénients de chacune. Nous avons constaté aussi que les performances

des systèmes biométriques dépendent de plusieurs facteurs et qu'elles varient d'un système à un autre.

En d'autres termes, la reconnaissance dans la plupart des systèmes biométriques n'est pas fiable à 100% comme c'était le cas pour les systèmes de reconnaissance classiques (Badge, carte à puce, mot de passe...), mais présentent bien des avantages qui leur donnent un intérêt d'une grande importance dans la sécurité des infrastructures et des systèmes informatiques.

**Bibliographie :**

- [1] Delac, K., & Grgic, M. (2004, June). A survey of biometric recognition methods. In *Electronics in Marine, 2004. Proceedings Elmar 2004. 46th International Symposium* (pp. 184-193). IEEE.
- [2] Jain, A. K., Ross, A., & Prabhakar, S. (2004). An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1), 4-20.
- [3] Field, A. T., & Field, A. T. (1959). *Fingerprint handbook*. Charles C. Thomas Publisher.
- [4] Karu, K., & Jain, A. K. (1996). Fingerprint classification. *Pattern recognition*, 29(3), 389-404.
- [5] Hong, L., Wan, Y., & Jain, A. (1998). Fingerprint image enhancement: algorithm and performance evaluation. *IEEE transactions on pattern analysis and machine intelligence*, 20(8), 777-789.
- [6] Kumar, A., Wong, D. C., Shen, H. C., & Jain, A. K. (2003, June). Personal verification using palmprint and hand geometry biometric. In *International Conference on Audio-and Video-Based Biometric Person Authentication* (pp. 668-678). Springer Berlin Heidelberg.
- [7] Jain, A., Flynn, P., & Ross, A. A. (Eds.). (2007). *Handbook of biometrics*. Springer Science & Business Media.
- [8] Borgen, H., Bours, P., & Wolthusen, S. D. (2008, August). Visible-spectrum biometric retina recognition. In *Intelligent Information Hiding and Multimedia Signal Processing, 2008. IHHMSP'08 International Conference on* (pp. 1056-1062). IEEE.
- [9] Choras, R. S. (2010). Hybrid Iris and Retina recognition for biometrics. In *2010 3rd International Congress on Image and Signal Processing*.
- [10] Wildes, R. P. (1997). Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, 85(9), 1348-1363.
- [11] Ma, L., Tan, T., Wang, Y., & Zhang, D. (2004). Efficient iris recognition by characterizing key local variations. *IEEE Transactions on image processing*, 13(6), 739-750.
- [12] Zhao, W., Krishnaswamy, A., Chellappa, R., Swets, D. L., & Weng, J. (1998). Discriminant analysis of principal components for face recognition. In *Face Recognition* (pp. 73-85). Springer Berlin Heidelberg.
- [13] Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12), 2037-2041.
- [14] Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303.
- [15] Huang, K., & Yan, H. (1997). Off-line signature verification based on geometric feature extraction and neural network classification. *Pattern Recognition*, 30(1), 9-17.
- [16] Wu, H., Siegel, M., & Khosla, P. (1998, May). Vehicle sound signature recognition by frequency vector principal component analysis. In *Instrumentation and Measurement Technology Conference, 1998. IMTC/98. Conference Proceedings. IEEE* (Vol. 1, pp. 429-434). IEEE.
- [17] Ashbourn, J. (2011). *Guide to Biometrics for Large-Scale Systems*. Springer.

# **Chapitre II**

## **La Reconnaissance Faciale**

### **L'état de l'art**

## 2.1. Introduction

La reconnaissance automatique du visage, en tant qu'une des technologies biométriques de base, a pris une part de plus en plus importante dans le domaine de la recherche, ceci est dû aux avancées rapides dans les technologies de développement des appareils photo numériques, Internet et les dispositifs mobiles, le tout associé à des besoins en sécurité sans cesse en augmentation. Selon IBG (International Biometric Group) [1], la modalité la plus utilisée après l'empreinte est le visage.

Le nombre d'applications possibles a pris beaucoup d'importance depuis plusieurs années. Deux types d'application peuvent être différenciés: celles dites du monde ouvert et celles dites du monde fermé. Alors que ces dernières ne traitent qu'avec un nombre limité de personnes connues, les premières quant à elles peuvent traiter avec des personnes inconnues. Pour reconnaître une personne à partir de son image, il faut passer par certaines étapes. Tout d'abord il faut détecter et/ou localiser le visage dans l'image. Ensuite il faut éventuellement le normaliser pour ramener le visage à une taille standard. Enfin, il s'agit d'aborder la phase de reconnaissance proprement dite.

De nombreuses méthodes de reconnaissance du visage ont été proposées au cours des 30 dernières années. La reconnaissance faciale automatique est un challenge et a suscité de nombreuses recherches dans différentes disciplines: psychologie, neurologie, mathématiques, physique et informatique (reconnaissance des formes, réseaux de neurones, vision par ordinateur). C'est la raison pour laquelle la littérature sur la reconnaissance du visages est vaste et diversifiée. Dans notre travail nous nous sommes intéressés par l'analyse d'images fixes, puisqu'elles représentent la base de tous les travaux.

Les systèmes de reconnaissance du visages sont très souvent classés à partir des conclusions d'études psychologiques sur la façon dont les hommes utilisent les caractéristiques faciales pour se reconnaître. De ce point de vue, on distingue les trois catégories suivantes: (1) les méthodes globales, (2) les méthodes locales et (3) les méthodes hybrides.

## 2.2. Pourquoi le visage

La reconnaissance faciale est la technique la plus commune et la plus utilisée. Elle reste la plus acceptable puisqu'elle correspond à ce que les humains utilisent dans l'interaction visuelle. Par rapport aux autres méthodes, la reconnaissance du visage s'avère plus avantageuse, d'une part c'est une méthode non intrusive, c'est-à-dire qu'elle n'exige pas la coopération du sujet (en observant les individus à distance), et d'autre part les capteurs utilisés

sont peu coûteux (une simple caméra) contrairement à l’empreinte digitale et l’iris où le sujet devra être très proche du capteur et devra coopérer pour l’acquisition de l’image sans oublier le coût de l’équipement nécessaire pour l’acquisition (équipement spécial coûteux). Malgré que certains disent que la reconnaissance du visage est une biométrie relativement peu sûre, sur le fait que le signal acquis est sujet à des variations beaucoup plus élevées en dépendance à d’autres paramètres tel que la variation de l’éclairage, le changement de la position du visage, la présence ou l’absence de lunettes et autres; au cours de ces dernières années plusieurs techniques de traitements d’images sont apparues, pour minimiser l’effet de ces paramètres, telle que la détection du visage, la normalisation de l’éclairage, etc.

### 2.3. System biométriques

#### 2.3.1. Les phases d’un système biométrique

Chaque système biométrique comporte deux phases d’exécution: (i) une phase d’apprentissage et (ii) une phase de test. Chacune des deux phases contient trois modules de base plus un module propre à la phase de test.

#### 2.3.2. Modules des systèmes biométriques

Typiquement tout système biométriques peut être présenté par quatre modules principaux:

- a. Module de capture:** responsable de l’acquisition des données biométriques d’un individu (cela peut être un appareil photo, un lecteur d’empreintes digitales, une caméra de sécurité, etc).
- b. Module d’extraction de caractéristiques:** Qui prend en entrée les données biométriques acquises par le module de capture et extrait seulement l’information pertinente afin de former une nouvelle représentation des données. Idéalement, cette nouvelle représentation est censée être unique pour chaque personne et relativement invariante aux variations intra-classe.
- c. Module de correspondance:** Il compare l’ensemble des caractéristiques extraites avec le modèle enregistré dans la base de données du système et détermine le degré de similitude (ou de divergence) entre les deux.
- d. Module de décision :** vérifie l’identité affirmée par un utilisateur ou détermine l’identité d’une personne basée sur le degré de similitude entre les caractéristiques extraites et le(s) modèle(s) stocké(s).

### 2.4. Impact des différents types de variabilités

Il est également très intéressant d’étudier l’impact des différentes sources de variabilité sur les performances du système visuel humain en termes de reconnaissance de visage. Une personne

ayant vu un visage en une seule occasion peut le reconnaître dans des conditions d'orientation, d'expression faciale ou de luminosité très différentes [1].

La direction d'éclairage influe sur l'aptitude humaine à reconnaître un visage. Les conditions d'éclairage les plus favorables consistent en un éclairage par le haut [2]. Johnston et al. [3] ont montré qu'un visage familier éclairé par une source lumineuse situé sous le visage est plus difficile à identifier. Par contre, l'expression faciale a peu d'influence sur nos capacités de reconnaissance, pour autant qu'elle reste raisonnable. D'autre part, de nombreux scientifiques s'accordent à dire que la pose n'a d'influence significative sur la reconnaissance que si son amplitude est très importante. En effet il est difficile de mettre en correspondance les deux profils d'un même visage, tandis que pour deux vues de trois quart cette tâche est relativement aisée. Cette constatation tend à prouver que notre cerveau est capable d'utiliser, dans une certaine mesure, la symétrie du visage par rapport à son axe central vertical [4].

Ces études illustrent l'importance de l'apprentissage pour la reconnaissance humaine. Nous avons, depuis notre enfance, observé suffisamment de visages pour avoir complètement intégré leurs symétries et être capable d'inférer une vue à partir de son opposée. A l'inverse, on a peu l'occasion de rencontrer dans la nature des sources d'éclairage provenant du sol. Par conséquent, il nous est difficile de reconnaître un visage dans ces conditions. Ces remarques montrent que les dis similarités entre deux visages différents ainsi que les variations possibles entre deux vues d'un même visage sont nécessaire pour construire un outil de reconnaissance automatique performant.

### **2.5. Difficultés inhérentes à la reconnaissance automatique**

De nombreuses propriétés du visage ainsi que les conditions dans lesquelles ils ont été capturés rendent le traitement automatique difficile. Dans le cadre de la reconnaissance, le principal problème sous-jacent est la variance intra classe, c'est à dire la variabilité que peut prendre le visage d'une même personne.

Les systèmes automatiques de reconnaissance des visages doivent rester invariants à tout facteur indépendant de l'identité du visage, même si ce facteur engendre des changements d'apparence du visage. Or, de nombreux facteurs, extérieurs au visage ou en lien avec sa nature intrinsèque, peuvent influencer sur celle-ci. Les conditions de prise de vue, notamment l'angle sous lequel le visage est observé et la puissance des sources de luminosité, influent considérablement sur l'apparence d'un visage [5].

Nous détaillons ici les principales difficultés rencontrées par un système de reconnaissance faciale automatique dans des conditions réelles.

### 2.5.1. Condition d'illumination

L'intensité et la direction d'éclairage lors de la prise de vue influent énormément sur l'apparence du visage dans l'image. Dans la plupart des applications réelles, des changements dans les conditions d'éclairage sont néanmoins inévitables, notamment lorsque les vues sont collectées à des dates différentes, à l'intérieur ou à l'extérieur. Etant donné la forme intrinsèque d'un visage humain, des changements d'éclairage peuvent faire apparaître sur le visage des ombres accentuant ou masquant certaines caractéristiques faciales.

Les variations d'illumination viennent entraîner des variations considérables dans l'apparence d'un visage. Deux types d'éclairage peuvent influencer celle-ci: l'illumination globale (ou ambiante) et l'illumination locale. Alors que l'illumination globale affecte tout le visage de manière uniforme (ou presque), l'illumination locale entraîne la création d'ombre et de zones éclairées et ce de manière non linéaire. La figure 2.1 présente un exemple de visage dont la source lumineuse l'éclairant se déplace.



**Figure 2.1:** Exemple d'un visage d'une même personne subissant un changement de luminosité

### 2.5.2. Les variations de la pose

La pose d'un visage définit la rotation qu'a pu subir un visage lors de la capture. Les variations de pose peuvent être de deux types selon le type de rotation: rotation dans le plan où l'axe de rotation est l'axe de la caméra se situent dans le même plan et rotation hors plan sinon. La figure 2.2 présente un exemple d'un visage subissant une rotation hors plan. Les variations de pose affectent grandement les systèmes de reconnaissance automatique du visage, c'est pourquoi nombre d'entre eux se limitent aux poses frontales, ou à des poses spécifiques requérant cependant une estimation au préalable. Dans le cas d'une rotation dans le plan, l'apparence du visage n'est pas déformée, et une bonne estimation de l'angle de rotation peut suffire à recalculer l'image par simple rotation inverse, et ainsi obtenir une pose frontale (front en haut de l'image, menton en bas). Le cas de la rotation hors plan est souvent bien plus complexe, sauf si les visages utilisés pour l'enrôlement et la reconnaissance présentent la même pose.



**Figure 2.2:** Exemple d'un visage d'une même personne subissant des variations de pose (hors plan).

### 2.5.3. Expressions faciales

Les visages sont des objets non rigides. L'expression faciale de l'émotion (figure 2.3), combinée avec la parole, peut produire des changements d'apparence importants des visages. Le nombre de configurations possibles est incalculable. L'influence de l'expression faciale sur la reconnaissance est donc difficile à évaluer. Puisque l'expression faciale affecte la forme géométrique et les positions des caractéristiques faciales, il semble logique que les techniques globales ou hybrides y soient plus robustes que la plupart des techniques géométriques. Gross et al. [6] soutiennent que les expressions faciales n'ont pas une grande influence sur les algorithmes de reconnaissance, pour autant qu'elles restent raisonnables. Alors que les cas extrêmes qui engendrent d'importantes déformations de la bouche (tel que le cri) et le rétrécissement ou la fermeture complète des yeux dégradent considérablement les performances de la reconnaissance automatique.



**Figure 2.3:** Variabilité intra classe due à la présence d'expressions faciales.

### 2.5.4. Les occlusions partielles

Le visage peut être partiellement masqué par des objets dans la scène (figure 2.4), ou par le port d'accessoires tels que les lunettes de soleil, une écharpe, un chapeau, ... Les

occultations peuvent être intentionnelles ou non. La plupart des systèmes déployés doivent être non intrusifs, i. e. ne comptant pas sur une coopération du sujet. Par conséquent, il est important de savoir reconnaître des visages partiellement occultés. Gross et al. ont évalué dans [6] l'impact du port de lunettes de soleil, et d'un cache nez occultant la partie inférieure du visage, par le biais de l'utilisation de la base AR [7]. Leurs résultats expérimentaux montrent que les performances des algorithmes testés sont relativement faibles dans ces conditions par rapport à l'utilisation de visages entièrement découverts.



Figure 2.4 : Variabilité intra-classe due à la présence d'occlusions partielles.

### 2.5.5. Température du corps

Les variations de la température du corps peuvent altérer grandement le rendu d'un visage capturé dans la modalité infrarouge (grandes longueurs d'onde). Cette modalité reflétant la chaleur émise par les objets : certaines parties du visage (nez ou oreille notamment) peuvent être à des températures différentes, leur rendu peut donc être très différent d'une capture à l'autre. La figure 2.5 présente un exemple d'une même personne dont les captures ont été réalisées à différents moments.



Figure 2.5: Variabilité intra-classe due à des variations de la température du corps.

### 2.5.6. Autres difficultés

D'autres types de difficultés peuvent apparaître pour un système automatique de reconnaissance faciale. Ainsi, des variations peuvent être dues à la présence de maquillage, d'opérations chirurgicales, de différentes coupes de cheveux, ou encore la

présence (absence) de moustaches, de barbes, . . . Un autre point très important est l'âge des captures, c'est à dire le moment à laquelle les captures ont été réalisées. En effet, l'apparence d'un visage peut changer au cours du temps (notamment lors de l'adolescence), et un écart de temps important entre deux captures peut engendrer des difficultés de reconnaissance. Cet effet n'est pas nouveau mais la constitution de bases de données s'étalant sur plusieurs années est difficile.

## 2.6 Principales Bases de Données de Visages

Plusieurs bases de données contenant des informations qui permettent l'évaluation des systèmes de reconnaissance de visages sont disponibles sur le marché. Toutefois, ces bases de données sont généralement adaptées aux besoins de quelques algorithmes spécifiques de reconnaissance. Le tableau (2.1) récapitule les principales caractéristiques de ces bases. Il résume, pour chacune d'elle, les conditions d'acquisition des images de visages (changements d'illumination, de pose, d'expression faciale), la disponibilité ou non des données 3D et aussi le nombre de sessions d'acquisition. Par ailleurs, plusieurs protocoles expérimentaux ont été élaborés afin d'évaluer le nombre important d'algorithmes de reconnaissance proposés ces dernières années. Des efforts particuliers de standardisation ont été aussi déployés pour définir des protocoles d'évaluation standardisés. Ainsi, la série d'évaluations FERET [Phi89b] menées par le « National Institute of Standards and Technology (NIST) » a permis la comparaison entre neuf systèmes de reconnaissance de visages proposés par des institutions et des entreprises dans le contexte de l'identification et de la vérification.

Les différentes bases de données peuvent différer entre elles sur plusieurs points:

- le nombre d'images disponibles est probablement le critère le plus important d'une base de données.
- le nombre d'images disponibles par personne.
- la modalité (2D-visible, 2D-infrarouge, 3D, couleur, niveaux de gris, maillage, carte de profondeur, . . .).
- la taille des images.
- les poses et orientations des visages.
- les variations de l'illumination.
- le sexe des personnes présentes.
- la présence d'artefacts (lunettes, barbes, . . .).
- la présence d'un fond uniforme.

- la période entre les prises de vues.

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

Les bases de données de visages peuvent être classées en trois catégories selon l'objectif recherché: reconnaissance, détection de visages ou analyse des expressions faciales.

**Tableau 2.1:** Base de données 2D. Les variations de l'image sont indiqués par (I) l'éclairage, (P) pose, (E) l'expression, (O) occlusion, (T) retard de temps.

Nom	RVB/Gris	Taille	Nombre de Personnes	Image/ Personne	Variations de l'image
AR	RVB	576*768	126 70 Males 56 Femelles	26	I,E,O,T
CVL	RVB	640*480	114 108 Males 6 Femelle	7	P,E
Yale-B	Gris	640*480	10	576	P, I
Yale	Gris	320*243	15 14 Males 1 Femelle	11	I,E
PIE	RVB	640*486	68	608	P, I, E
UMIST	Gris	220*220	20	19-36	P
Olivetti-ORL	Gris	92*112	40	10	E
JAFFE	Gris	256*256	10	7	I
The Human Scan Database	Gris	384*286	23	66	P
The Univ of Oulu Face Data	Gris	428*569	125		I
XM2VTS	RVB	576*720	295		P
FERET	RVB	256*384	30000		P,I,E,T
LFW	RVB	150*150	13233		P,I,E,O,T

## 2.7. Techniques 2-D de la reconnaissance des visages

Plusieurs méthodes de reconnaissance du visage ont été proposées durant les vingt dernières années. Elles peuvent être placées en deux grands axes: la reconnaissance à partir d'images fixes et la reconnaissance à partir de séquence d'images (vidéo). Les travaux portant sur ce domaine, sous différentes conditions d'éclairage, d'expressions faciales et d'orientations, peuvent être répertoriés en trois grandes catégories:

### 2.7.1 Techniques Globales (holistiques)

Ces techniques d'identification utilisent l'image entière du visage comme entrée du système de reconnaissance. L'idée est généralement de projeter l'image d'entrée du visage, préalablement vectorisée, dans un espace de plus faible dimension, où la reconnaissance est supposée être plus facile. Notant, que la projection est souvent conçue pour ne sélectionner que les caractéristiques discriminative (importantes) pour chaque classe dans la base de données.

La plupart de ces méthodes réalisent une analyse de sous-espaces du visage (ou de la variété définie par les visages). Cette analyse découle de la constatation d'un fait relativement simple: la classe des visages réside dans un sous-espace de l'espace de l'image d'entrée. Prenons par exemple une image de taille  $100 \times 100$  en niveaux de gris. Le nombre de configurations possibles est égal à 25610000. Cependant, parmi toutes ces configurations possibles, seule une petite partie correspond aux visages. L'information contenue dans les images du visage est donc très redondante, la dimension de ces images peut donc être réduite en se concentrant uniquement sur ce qui nous intéresse (les visages). Le sous-espace est souvent appelé espace du visage (ou « face space »).

Les méthodes globales peuvent se décomposer en deux types de techniques: les techniques linéaires et les techniques non linéaires.

#### 2.7.1.a Techniques linéaires

Comme leur nom l'indique, les techniques linéaires réalisent une projection linéaire des visages sur un espace de faible dimension. Parmi les inconvénients de ces techniques linéaires leur sensibilité aux conditions de luminosité, et plus généralement aux variations non convexes.

La méthode la plus célèbre de ces approches et la technique des Eigen Faces qui a été proposé par Turk et Pentland [8]. L'Analyse en Composantes Principales (Principal Component Analysis « PCA ») est réalisée sur un ensemble d'apprentissage d'images du visage. Le nouvel espace est définie par les principaux vecteurs propres résultant de la l'PCA.

Par la suite, plusieurs chercheurs ont basé sur le choix des vecteurs propres à retenir pour déterminer le nouvel espace. Kirby et al [9] proposent un critère basé sur l'énergie des valeurs propres associées aux vecteurs propres. Les vecteurs propres correspondant aux plus grandes valeurs propres sont retenus jusqu'à ce que la somme des valeurs propres dépasse un certain seuil de l'énergie totale (90%).

Une autre méthode proposée par Belhumer et al. [10] réalise une Analyse Discriminante Linéaire (LDA), elle est aussi connue sous le nom de *Fisherfaces*. En effet la LDA consiste à maximiser, sur un ensemble d'apprentissage, le critère de Fisher, à savoir le quotient de la différence inter-classe par la variance intra-classe. Ainsi, contrairement à la technique des Eigenface où la meilleure représentation (celle maximisant la variance) est recherchée, le but est ici une meilleure séparation des classes. Cependant, étant donné que le nombre d'images est souvent inférieur à leur dimension, la matrice de variance intra-classe peut être singulière, et son inversion pose donc problème. Ce problème est connu sous le nom de Small sample size problem. Des méthodes ont été proposées pour contourner ce problème, la plus utilisée étant de réaliser une PCA au préalable pour diminuer la dimension des échantillons.

De nombreuses variantes à ces techniques linéaires ont été proposées dans la littérature [11-15].

D'autres techniques linéaires ont également été utilisées pour le calcul des vecteurs caractéristiques:

- l'Analyse en Composantes Indépendantes (Independent Component Analysis « ICA ») dans [16].
- la Factorisation de Matrices non Négatives (Non-Negative Matrices Factorization « NMF ») dans [17].
- l'Analyse Discriminante Bilinéaire (Bilinear Discriminant Analysis « BDA ») dans [18].
- la technique dite de « Vecteurs Communs Discriminants » (Discriminant Common Vectors « DCV ») dans [19]

Bien que ces méthodes linéaires soient assez efficaces, elles manquent de précision dès lors que les images de visages subissent des transformations non linéaires. Une simple modification de la luminosité transforme celui-ci de façon non linéaire étant donné la complexité de la forme.

### 2.7.1.b Techniques non linéaires

Les méthodes précédentes trouvent un espace de représentation fidèle aux données lorsque la structure de ces données est linéaire ce qui n'est généralement pas le cas. L'idée est d'utiliser une fonction noyau pour construire un espace de grande dimension dans lequel le problème devient linéaire. On peut ainsi appliquer des méthodes linéaires de réduction de dimension lorsque la structure intrinsèque des données n'est pas linéaire. Ces méthodes utilisent généralement « l'astuce du noyau » (kernel trick) qui établit que tout algorithme formulé avec une fonction noyau peut être reformulé avec une autre fonction noyau. Une démarche courante est d'exprimer la méthode avec un produit scalaire, puis de le remplacer par une fonction noyau. Le kernel trick permet alors de travailler dans l'espace transformé sans avoir à calculer explicitement l'image de chaque donnée. L'Analyse en Composantes Principales à Noyaux (Kernel Principal Component Analysis, KPCA) est la reformulation non linéaire de la technique linéaire classique qu'est la PCA en utilisant des fonctions à noyaux [20]. Depuis plusieurs années, la reformulation de techniques classiques à l'aide de « l'astuce du noyau » a permis l'émergence de nombreuses techniques comme les machines à support de vecteurs (Support Vector Machine, SVM) [21]. LA PCA à noyaux calcule les principaux vecteurs propres de la matrice du noyau plutôt que la matrice de covariance. Cette reformulation de la PCA classique peut être vue comme une réalisation de PCA sur l'espace de grande dimension transformée par la fonction noyau associée.

La technique KPCA est une méthode basée sur les noyaux et ses performances dépendent alors grandement du choix de la fonction noyau  $k$ . Les noyaux classiquement utilisés sont le noyau linéaire (cela revient alors à effectuer la PCA classique), le noyau polynomial ou encore le noyau gaussien [21]. L'analyse en composantes principales à noyaux a été appliquée avec succès à plusieurs problèmes comme la reconnaissance de la parole [22], ou la détection de nouveaux éléments d'un ensemble [23]. Un gros défaut de l'analyse en composantes principales à noyau est que la taille de la matrice du noyau est le carré du nombre d'échantillons de l'ensemble d'apprentissage ce qui peut rapidement être prohibitif.

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

- le Multi Dimensional Scaling (MDS) dans [24] ou [25].
- l'Isomap dans [26].
- les diffusion maps dans [27].
- le Local Linear Embedding (LLE) dans [28] ou [29].

- les Laplacian eigenmaps dans [30].
- le Hessian LLE dans [31].
- le Local Tangent Space Analysis (LTSA) dans [32].
- les approches neuronales dans [33] (autoencodeurs), dans [34] (cartes de Kohonen), et dans [35] (réseaux de neurones convolutionnels).

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

### 2.7.2 Techniques Locales

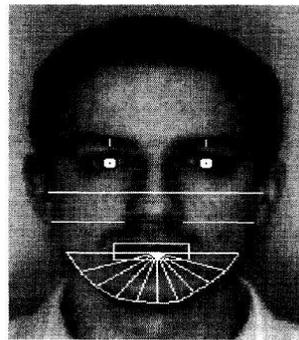
Les méthodes locales peuvent être classées en deux catégories, les méthodes basées sur les points d'intérêt et celles basées sur l'apparence du visage. Dans le premier cas, on détecte tout d'abord les points d'intérêt et ensuite on extrait des caractéristiques localisées sur ces points. Dans le second cas, on divise le visage en petites régions (ou patches) sur lesquelles les caractéristiques locales sont directement extraites. En comparaison avec les approches globales, les méthodes locales présentent certains avantages. Tout d'abord, les approches locales peuvent fournir des informations supplémentaires basées sur les parties locales. De plus, pour chaque type de caractéristiques locales, on peut choisir le classificateur le plus adapté. Malgré ces avantages, l'intégration d'informations de structure plus globale est nécessaire. En général, il y a deux façons de procéder pour atteindre cet objectif. Premièrement, les informations globales sont intégrées dans les algorithmes en utilisant des structures de données, telles qu'un graphe où chaque nœud représente une caractéristique locale alors qu'une arête entre deux nœuds représente la relation spatiale entre eux. La reconnaissance du visage apparaît comme un problème d'appariement de deux graphes. Deuxièmement, les algorithmes peuvent utiliser des techniques de fusion de scores: des classificateurs séparés sont utilisés sur chaque caractéristique locale pour calculer une similarité et ensuite les similarités obtenues sont combinées afin d'obtenir un score global pour la décision finale.

#### 2.7.2.A Méthodes locales basées sur les caractéristiques d'intérêt

Les méthodes les plus anciennes en reconnaissance du visages, appartiennent à cette catégorie [36,37]. Elles s'appuient toutes [38] sur l'extraction de caractéristiques géométriques

spécifiques telles que la largeur de la tête, les distances entre les yeux, etc. Ces données sont ensuite utilisées par des classificateurs afin de reconnaître des individus. Ces méthodes présentent les deux inconvénients suivants: (1) les caractéristiques géométriques sont difficiles à extraire dans certains cas puisque la tâche de détection précise de points caractéristiques n'est pas facile, en particulier dans les cas où des occultations ou des variations (pose, expression) du visage sont présentes et (2) les caractéristiques géométriques seules ne sont pas suffisantes pour représenter entièrement un visage et d'autres informations utiles telles que les valeurs des niveaux de gris de l'image sont complètement écartées.

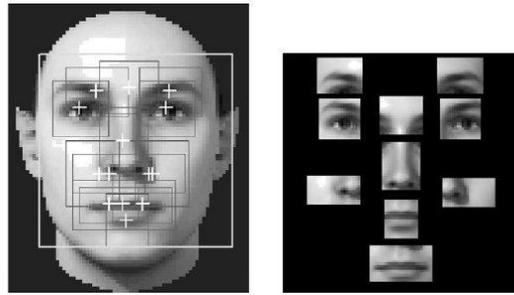
Ces deux limites ont engendré deux directions de recherche. La première se concentre sur les performances des détecteurs de points caractéristiques du visage. Brunelli et Poggio proposent [38], une technique qui extrait automatiquement un ensemble de 35 caractéristiques géométriques d'une image du visage (voir la figure 2.6). Ces ensembles de caractéristiques sont ensuite comparés deux à deux via la distance de Mahalanobis pour réaliser la reconnaissance.



**Figure 2.6:** Localisation des caractéristiques géométriques utilisées dans [38].

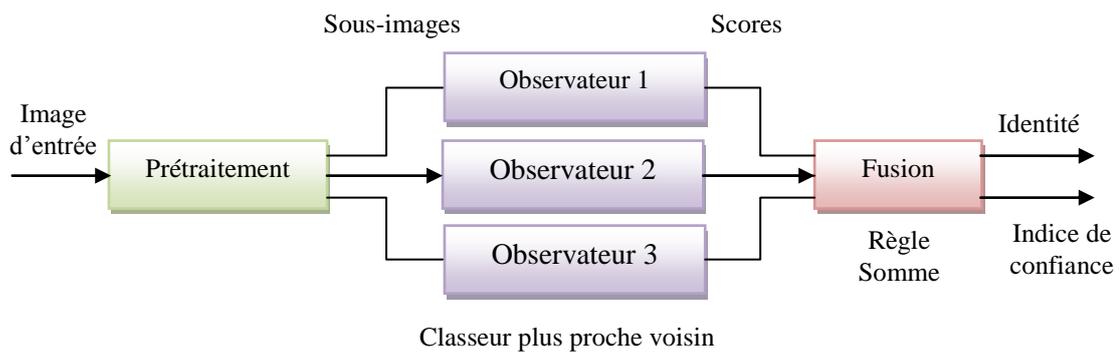
Rowley et al. [39] ont utilisé plusieurs détecteurs de traits spécifiques correspondant à chaque partie du visage, telles que les yeux, le nez, la bouche, etc.

Heisele et al. [40] a proposé une autre technique basée sur la détection de la région contenant le visage dont dix points de caractéristiques sont extraits. Les zones autour de ces points sont ensuite extraites (voir la figure 2.7), et concaténées pour former le vecteur caractéristique du visage. La classification est finalement réalisée grâce à l'utilisation d'une machine à vecteurs de support (SVM).



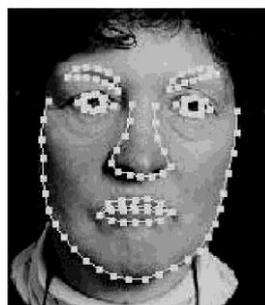
**Figure 2.7:** Cartes de contours utilisées dans [40].

Price et Gree proposent également une méthode [41] se basant sur des zones extraites du visage. Ici, trois régions sont considérées: une bande rectangulaire comprenant les yeux et le nez, une deuxième bande rectangulaire ne contenant que les yeux et enfin une région contenant le visage entier. Une analyse linéaire discriminante est ensuite appliquée à chacune de ces régions (voir la figure 2.8)



**Figure 2.8:** Approche de Price et Gee [41].

Les modèles actifs d'apparence (ou AAM pour « Active Appearance Models ») sont présentés par Cootes et al. Dans [42]. Ils consistent en la création d'un modèle statistique d'un visage, (voir la figure 2.9). Le modèle est ensuite déformé pour « coller » au plus près des traits du visage. La reconnaissance est effectuée sur le résidu calculé correspondant à l'erreur de prédiction du modèle.



**Figure 2.9:** Modèle actif d'apparence.

Plus tard, Wiskott et al. [43] ont étendu l'utilisation de LDA à une méthode très connue appelée Elastic Bunch Graph Matching (EBGM), où les noeuds des graphes sont situés sur un certain nombre de points sélectionnés du visage. De manière similaire à la méthode de [44], Wiskott et al. ont utilisé les ondelettes de Gabor pour extraire les caractéristiques des points détectés car les filtres de Gabor sont robustes aux changements d'illumination, aux distorsions et aux variations d'échelle.

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

### **2.7.2.B Les méthodes locales basées sur l'apparence du visage**

Une fois les régions locales définies, il s'agit de choisir la meilleure manière de représenter les informations de chaque région. Cette étape est critique pour les performances du système de reconnaissance. Les caractéristiques couramment utilisées sont: les coefficients de Gabor [45-47], les ondelettes de Harr [48], les transformées de Fourier, SIFT (Scale Invariant Feature Transform) [49], les caractéristiques basées sur la méthode LBP (Local Binary Pattern) [50], LPQ (Local Phase Quantization) [51], WLD (Weber Law Descriptor) [52], BSIF (Binarized Statistical Image Features) [53]. Dans cette section on discutera en détails la méthode LBP et ses variantes.

#### **2.7.2.B.1 LBP et sa variante récente**

Les LBP (pour « Local Binary Patterns ») ont également été utilisés pour la reconnaissance faciale, notamment dans [50]. Le visage est subdivisé en sous-régions carrées de taille égale sur lesquelles sont calculées les caractéristiques LBP. Les vecteurs obtenus sont ensuite concaténés pour obtenir le vecteur de caractéristiques final.

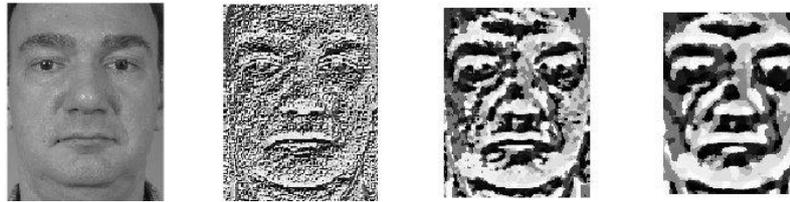
La méthodologie de LBP a été développée récemment avec un grand nombre de variations pour l'amélioration des performances dans différentes applications. Ces variations portent sur différents aspects de l'opérateur LBP original: 1) l'amélioration de sa capacité discriminatoire; 2) l'amélioration de sa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

La méthode LBP définit un certain nombre de modèles pour décrire les structures locales. Pour améliorer leur capacité discriminante, plus de modèles ou d'informations peuvent être codés. Jin et al. [54] ont amélioré l'opérateur de LBP pour décrire plus d'informations de

structure locale dans certaines circonstances. Plus précisément, ils ont proposé un LBP amélioré (Improved LBP, ILBP), qui compare tous les pixels avec l'intensité moyenne de tous les pixels (y compris le pixel central).

Yang et Wang [55] ont proposé Hamming LBP pour améliorer la capacité discriminative de la LBP originale. Ils ont reclassé les motifs non uniformes sur la base de la distance de Hamming, au lieu de les rassembler en un seul bin comme le fait LBP<sup>u2</sup>.

Des extensions des LBP comme les MB-LBP (pour « Multi-Scale Block Binary Pattern ») ont été proposées et appliquées aux visages par Liao et al. [56] (voir la figure 2.10). Le gros avantage des méthodes locales de reconnaissance du visage est qu'elles peuvent modéliser facilement les variations de pose, d'illumination ou encore d'expressions que peut subir un visage. Cependant, elles nécessitent souvent le placement manuel de nombreux points d'intérêts pour une bonne précision, et sont donc lourdes à mettre en œuvre.



**Figure 2.10:** Caractéristiques (MB) LBP pour un visage, respectivement pour un masque de taille  $3 \times 3$ ,  $9 \times 9$  et  $15 \times 15$ .

### 2.7.3 .Techniques Hybrides

Les techniques hybrides sont des techniques qui combinent les caractéristiques holistiques et locales afin d'améliorer les performances de la reconnaissance du visage. En effet les caractéristiques locales et les caractéristiques globales ont des propriétés différentes. On peut espérer pouvoir exploiter leur complémentarité pour améliorer la classification.

Wiskott et al. [57]. Propose une techniques basée sur deux types de caractéristiques Elastic Bunch Graph Matching (EBGM) et l'ondelette de Gabor ou le visage est représenté par des Face Bunch Graph (FBG). Chaque nœud du graphe correspond à une certaine caractéristique faciale (l'œil droit ou gauche...), voir la figure 2.11. A chaque nœud du graphe est associée l'apparence du voisinage de la caractéristique via un jet. Les jets représentent l'ensemble des 40 coefficients issus de la convolution du voisinage de la caractéristique par un filtre de Gabor spécifique. Les arêtes du graphe sont pondérées par la distance relative des caractéristiques adjacentes. Une fois que le graphe est créé pour chaque personne de la base d'apprentissage, un algorithme spécifique de mise en correspondance permet d'identifier une personne inconnue. L'algorithme essaie itérativement de faire correspondre le graphe créé à chaque

graphe de la base d'apprentissage en minimisant une fonction de coût prenant en compte à la fois une mesure de similarité géométrique ainsi qu'une mesure de similarité de l'apparence modélisée via les jets.

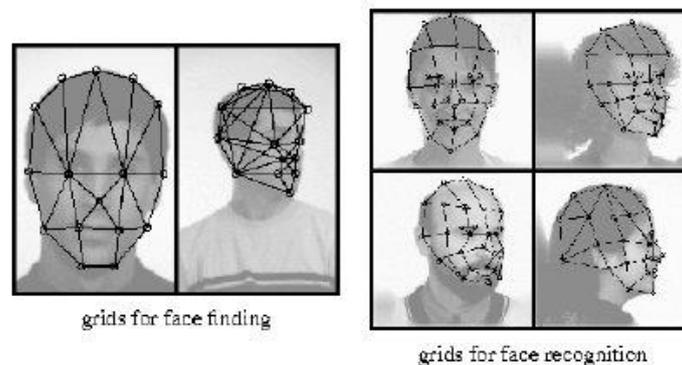


Figure 2.11 : Graphe appliqué aux visages pour l'approche EBM.

Perlibakas présente plus récemment l'algorithme **Log Gabor PCA** dans [58]. Une convolution par des ondelettes de Gabor orientées est réalisée autour de certains points caractéristiques du visage. Les vecteurs ainsi créés contiennent à la fois la localisation ainsi que les amplitudes des énergies locales. Une Analyse en Composantes Principales est ensuite réalisée afin de réduire la dimension de ces vecteurs.

Pentland et al. Présentent dans [59] l'approche dite des espaces propres modulaires (Modular Eigenspaces) Cette technique réalise une Analyse en Composantes Principales et une classification sur des régions distinctes du visage, comme les yeux, le nez, la bouche ou encore le visage entier. La zone de la bouche subit de grosses déformations dues aux expressions faciales, ainsi l'ajout de cette région au processus entier fait décroître les taux de reconnaissance.

Zhang et al ont présenté dans [60] une combinaison entre LBP et l'ondelette de Gabor (Local Gabor Binary Patterns LGBPH) pour améliorer la performance de l'histogramme spatial et exploiter l'analyse multi-résolution et multi-orientation. Le vecteur caractéristique LGBPHS est construit à partir de chaque image résultant de la convolution du visage et le filtre de Gabor. Chaque image est divisée en sous-régions, ensuite LBP est appliquée pour chaque région. Finalement tous les sous-histogrammes sont concaténés pour construire LGBPH globale pour le visage.

La figure 2.12 résume les principaux algorithmes utilisés en reconnaissance du visage 2D.

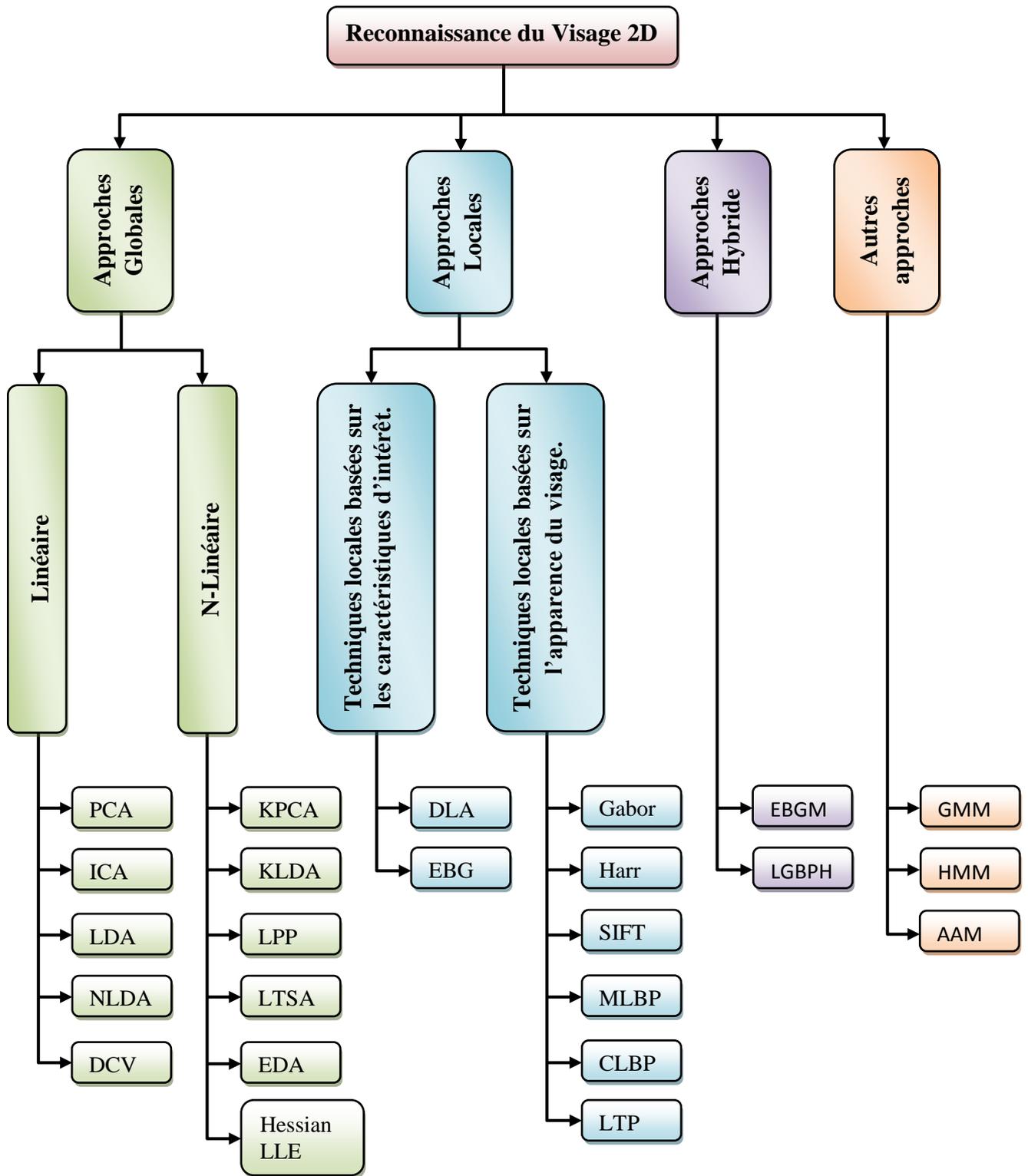


Figure 2.12: Classification des principaux algorithmes utilisés en reconnaissance faciale 2D.

### 2.7.4 Conclusion

Dans ce chapitre nous avons mis en évidence les avantages et les inconvénients de la reconnaissance faciale, le processus de la reconnaissance du visage et les techniques utilisées.

Les méthodes peuvent principalement se classer en deux catégories: les méthodes globales et les méthodes locales (les méthodes hybrides essayant de faire un lien entre ces types d'approches).

Les méthodes globales présentent un certains nombre d'avantages:

- Le problème de la reconnaissance faciale automatique est transformé en un problème d'analyse du sous-espaces de visages, pour lequel de nombreuses méthodes statistiques existent.
- Elles sont souvent applicables à des images basses résolution ou de mauvaises qualités.

Certains inconvénients sont cependant à signaler pour ces méthodes:

- Il est nécessaire de disposer de suffisamment de données représentatives des visages.
- Il n'y a pas d'a priori sur le physique d'un visage.
- Ces méthodes ne sont robustes qu'à des variations limitées (pose, illumination, expression).

De la même manière les méthodes locales présentent certains avantages:

- Le modèle créé possède des relations intrinsèques bien définies avec les visages réels.
- Les modèles créés peuvent prendre en compte explicitement les variations telles que la pose, l'illumination ou les expressions. La reconnaissance est ainsi plus efficace dans le cas de fortes variations.
- La connaissance a priori sur les visages peut être intégrée aux modèles afin d'améliorer leur efficacité.

Ces méthodes présentent également quelques inconvénients :

- La construction du modèle, reposant souvent sur la détection de points caractéristiques faciaux, peut être laborieuse.
- L'extraction des points caractéristiques peut être difficile dans le cas de variations de pose, d'illumination, d'occlusion . . .
- Les images doivent être relativement de bonne qualité, et/ou être de résolution suffisante afin de pouvoir extraire les points caractéristiques.

**Bibliographie :**

- [1] Visani, M. (2005). Vers de nouvelles approches discriminantes pour la reconnaissance automatique de visages (Doctoral dissertation, INSA de Lyon).
- [2] Hill, H., & Bruce, V. (1996). The effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4), 986.
- [3] Lee, K. C., Ho, J., & Kriegman, D. J. (2005). Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on pattern analysis and machine intelligence*, 27(5), 684-698.
- [4] Troje, N. F., & Bühlhoff, H. H. (1998). How is bilateral symmetry of human faces used for recognition of novel views?. *Vision research*, 38(1), 79-89.
- [5] Hatture, S. M., & Patil, S. I. (2015). Recent Advances In Face Recognition.
- [6] Gross, R., Shi, J., & Cohn, J. F. (2001). *Quo vadis face recognition?*. Carnegie Mellon University, The Robotics Institute.
- [7] Martinez, A. M. (1998). The AR face database. *CVC Technical Report*, 24.
- [8] Turk, M. A., & Pentland, A. P. (1991, June). Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on* (pp. 586-591). IEEE.
- [9] Kirby, M., & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern analysis and Machine intelligence*, 12(1), 103-108.
- [10] Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7), 711-720.
- [11] Roweis, S. (1998). EM algorithms for PCA and SPCA. *Advances in neural information processing systems*, 626-632.
- [12] Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3), 611-622.
- [13] Moghaddam, B., & Pentland, A. (1995, June). Probabilistic visual learning for object detection. In *Computer Vision, 1995. Proceedings., Fifth International Conference on* (pp. 786-793). IEEE.
- [14] Cai, D., He, X., & Han, J. (2007, October). Semi-supervised discriminant analysis. In *2007 IEEE 11th International Conference on Computer Vision* (pp. 1-7). IEEE.
- [15] Torralba, A. B., & Oliva, A. (1999). Semantic organization of scenes using discriminant structural templates. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* (Vol. 2, pp. 1253-1258). IEEE.
- [16] Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Transactions on neural networks*, 13(6), 1450-1464.
- [17] Buciu, I., & Pitas, I. (2004, August). Application of non-negative and local non negative matrix factorization to facial expression recognition. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (Vol. 1, pp. 288-291). IEEE.

- [18] Visani, M., Garcia, C., & Jolion, J. M. (2005, September). Normalized radial basis function networks and bilinear discriminant analysis for face recognition. In *IEEE Conference on Advanced Video and Signal Based Surveillance, 2005*. (pp. 342-347). IEEE.
- [19] Cevikalp, H., Neamtu, M., Wilkes, M., & Barkana, A. (2005). Discriminative common vectors for face recognition. *IEEE Transactions on pattern analysis and machine intelligence*, 27(1), 4-13.
- [20] Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5), 1299-1319.
- [21] Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge university press.
- [22] Maurer, T., Guigonis, D., Maslov, I., Pesenti, B., Tsaregorodtsev, A., West, D., & Medioni, G. (2005, June). Performance of Geometrix ActiveID<sup>TM</sup> 3D Face Recognition Engine on the FRGC Data. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops* (pp. 154-154). IEEE.
- [23] Hoffmann, H. (2007). Kernel PCA for novelty detection. *Pattern Recognition*, 40(3), 863-874.
- [24] Kim, H., Park, H., & Zha, H. (2007, January). Distance Preserving Dimension Reduction for Manifold Learning. In *SDM* (pp. 527-532).
- [25] Biswas, S., Bowyer, K. W., & Flynn, P. J. (2010, September). Multidimensional scaling for matching low-resolution facial images. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on* (pp. 1-6). IEEE.
- [26] Yang, M. H. (2002). Face recognition using extended isomap. In *Image Processing. 2002. Proceedings. 2002 International Conference on* (Vol. 2, pp. II-117). IEEE.
- [27] Hagen, G., Smith, T., Banasuk, A., Coifman, R. R., & Mezie, I. (2007, December). Validation of low-dimensional models using diffusion maps and harmonic averaging. In *Decision and Control, 2007 46th IEEE Conference on* (pp. 5353-5357). IEEE.
- [28] Wang, J., Zhang, C., & Kou, Z. (2003). An Analytical Mapping for LLE and Its Application in Multi-Pose Face Synthesis. In *BMVC* (pp. 1-12).
- [29] Socolinsky, D. A., & Selinger, A. (2004, June). Thermal face recognition in an operational scenario. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on* (Vol. 2, pp. II-1012). IEEE.
- [30] He, X., Yan, S., Hu, Y., Niyogi, P., & Zhang, H. J. (2005). Face recognition using Laplacianfaces. *IEEE transactions on pattern analysis and machine intelligence*, 27(3), 328-340.
- [31] Kim, H., Park, H., & Zha, H. (2007, January). Distance Preserving Dimension Reduction for Manifold Learning. In *SDM* (pp. 527-532).
- [32] Wang, Q., & Li, J. (2009). Combining local and global information for nonlinear dimensionality reduction. *Neurocomputing*, 72(10), 2235-2241.
- [33] Teli, M. N. (2007). Dimensionality reduction using neural networks. In *Intelligent Engineering Systems Through Artificial Neural Networks, Volume 17*. ASME Press.
- [34] Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A

- convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1), 98-113.
- [35] Duffner, S., & Garcia, C. (2007, September). Face recognition using non-linear image reconstruction. In *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on* (pp. 459-464). IEEE.
- [36] Manjunath, B. S., Chellappa, R., & von der Malsburg, C. (1992, June). A feature based approach to face recognition. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on* (pp. 373-378). IEEE.
- [37] Rowley, H. A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. *IEEE Transactions on pattern analysis and machine intelligence*, 20(1), 23-38.
- [38] Brunelli, R., & Poggio, T. (1993). Face recognition: Features versus templates. *IEEE transactions on pattern analysis and machine intelligence*, 15(10), 1042-1052.
- [39] Rowley, H. A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. *IEEE Transactions on pattern analysis and machine intelligence*, 20(1), 23-38.
- [40] Heisele, B., Ho, P., Wu, J., & Poggio, T. (2003). Face recognition: component-based versus global approaches. *Computer vision and image understanding*, 91(1), 6-21.
- [41] Price, J. R., & Gee, T. F. (2005). Face recognition using direct, weighted linear discriminant analysis and modular subspaces. *Pattern Recognition*, 38(2), 209-219.
- [42] Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6), 681-685.
- [43] Wiskott, L., Fellous, J. M., Kuiger, N., & Von Der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7), 775-779.
- [44] Manjunath, B. S., Chellappa, R., & von der Malsburg, C. (1992, June). A feature based approach to face recognition. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on* (pp. 373-378). IEEE.
- [45] Wiskott, L., Fellous, J. M., Kuiger, N., & Von Der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7), 775-779.
- [46] Manjunath, B. S., Chellappa, R., & von der Malsburg, C. (1992, June). A feature based approach to face recognition. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on* (pp. 373-378). IEEE.
- [47] Brunelli, R., & Poggio, T. (1993). Face recognition: Features versus templates. *IEEE transactions on pattern analysis and machine intelligence*, 15(10), 1042-1052.
- [48] Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, 57(2), 137-154.
- [49] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.
- [50] Ahonen, T., Hadid, A., & Pietikäinen, M. (2004, May). Face recognition with local binary patterns. In *European conference on computer vision* (pp. 469-481). Springer

Berlin Heidelberg.

- [51] Ojansivu, V., & Heikkilä, J. (2008, July). Blur insensitive texture classification using local phase quantization. In *International conference on image and signal processing* (pp. 236-243). Springer Berlin Heidelberg.
- [52] Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X., & Gao, W. (2010). WLD: A robust local image descriptor. *IEEE transactions on pattern analysis and machine intelligence*, 32(9), 1705-1720.
- [53] Kannala, J., & Rahtu, E. (2012, November). Bsif: Binarized statistical image features. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 1363-1366). IEEE.
- [54] Jin, H., Liu, Q., Lu, H., & Tong, X. (2004, December). Face detection using improved LBP under bayesian framework. In *Image and Graphics (ICIG'04), Third International Conference on* (pp. 306-309). IEEE.
- [55] Yang, H., & Wang, Y. (2007, August). A LBP-based face recognition method with Hamming distance constraint. In *Image and Graphics, 2007. ICIG 2007. Fourth International Conference on* (pp. 645-649). IEEE.
- [56] Liao, S., Zhu, X., Lei, Z., Zhang, L., & Li, S. Z. (2007, August). Learning multi-scale block local binary patterns for face recognition. In *International Conference on Biometrics* (pp. 828-837). Springer Berlin Heidelberg.
- [57] Wiskott, L., Fellous, J. M., Kuiger, N., & Von Der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7), 775-779.
- [58] Perlibakas, V. (2006). Face recognition using principal component analysis and log-gabor filters. *arXiv preprint cs/0605025*.
- [59] Pentland, A., Moghaddam, B., & Starner, T. (1994, June). View-based and modular eigenspaces for face recognition. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on* (pp. 84-91). IEEE.
- [60] Zhang, W., Shan, S., Gao, W., Chen, X., & Zhang, H. (2005, October). Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* (Vol. 1, pp. 786-791). IEEE.

# **Chapitre III**

## **Base de Donn e et Mise en Œuvre et R sultats**

### 3.1 Introduction

Après avoir présenté les différentes techniques de reconnaissance du visage dans le chapitre précédent, il est question dans ce chapitre de mettre en œuvre les différentes techniques récentes dans ce domaine et discuter les avantages et les inconvénients de chaque algorithme en termes de taux de classification et temps de calcul dans le processus d'identification du visage. En effet, la performance de ces algorithmes dépend beaucoup plus des résultats de détection des visages et les conditions de capture. Parmi toutes les méthodes présentées dans la littérature, certaines demeurent plus avantageuses que d'autre et les critères utilisés dans la sélection reposent notamment sur le temps de calcul et la performance de classification. Pour cela on est intéressé aux approches locales dû à leur performance et leur simplicité d'implémentation.

Après une présentation des bases de données utilisées dans notre travail, les différentes caractéristiques locales, qui font l'objet des travaux récents, seront exposées pour présenter ensuite notre technique et sa comparaison avec celles ci.

### 3.2 Extraction de caractéristiques

L'extraction des caractéristiques représente le point le plus délicat dans tous les systèmes de reconnaissance faciale. La performance d'un système biométrique dépend essentiellement du choix de la caractéristique faciale la plus pertinente, pour cela plusieurs méthodes locales et globales sont proposées dans la littérature. L'utilisation des caractéristiques locales présente plusieurs avantages par rapport aux caractéristiques globales. C'est pour cette raison que les systèmes les plus récents s'appuient sur celles-ci pour développer des systèmes de reconnaissance du visage encore plus performants. Ces caractéristiques locales peuvent se décomposer en deux grandes catégories à savoir Sparse et Dense caractéristiques.

#### 3.2.1 Scale Invariant Feature Transform

La Scale Invariant Feature Transform (**SIFT**), que l'on peut traduire par « transformation de caractéristiques invariante à l'échelle » est proposé en 1999 par David Lowe [1]. Cette caractéristique a gagné beaucoup d'attention par les chercheurs dans différents domaines.

L'algorithme SIFT, est un algorithme de traitement d'image qui permet de détecter et de décrire les caractéristiques d'une image. Il permet de transformer une image en un ensemble de caractéristiques, chacune étant invariante aux transformations suivantes : translation, changement d'échelle, rotation et partiellement invariante au changement d'éclairage.

Comme il est proposé par Lowe, les principales étapes pour calculer la SIFT des images à étudier sont :

La première étape de l'algorithme est la représentation espace-échelle définie, par une convolution entre une fonction gaussienne paramétrable et une fonction  $f$  représentant l'image d'entrée.

La détection des points d'intérêt présent dans l'algorithme SIFT se fait dans l'espace des échelles. Les emplacements des points d'intérêt sont définis comme étant l'extremum du résultat de la Différence de Gaussiennes (DoG), appliquée, dans l'échelle spatiale sur, une série d'image lissées et ré-échantillonnées.

L'espace d'échelle d'une image  $L(x, y, \sigma)$  et le produit de convolution entre la fonction gaussienne  $G(x, y, \sigma)$  avec l'image d'entrée  $I(x, y)$  :

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (3.1)$$

Avec \* le produit de convolution, G la fonction gaussienne.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3.2)$$

Pour extraire des points d'intérêt stable dans l'espace échelle, Lowe a proposé d'utiliser l'extremum de la fonction de différences gessiennes, qui peut être calculée à partir de la différence de deux échelles séparées d'une constante multiplicative k. La fonction DoG est donnée par :

$$G(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3.3)$$

Cette première étape permet d'obtenir un grand nombre de points d'intérêt. Selon la méthode SIFT, une phase de filtrage de ces points est effectuée pour extraire les plus intéressants. Pour qu'un point soit sélectionné, il doit satisfaire deux conditions: La première, exige que la valeur  $D(x, y, \sigma)$  soit supérieur a un certain seuil prédéfinie (fixée à 0.03 par Lowe), la seconde, s'intéresse à la matrice **hessienne** de D :

$$M^L = \begin{bmatrix} \frac{\partial D^2}{\partial x} & \frac{\partial D}{\partial x} \frac{\partial D}{\partial y} \\ \frac{\partial D}{\partial x} \frac{\partial D}{\partial y} & \frac{\partial D^2}{\partial y} \end{bmatrix}$$

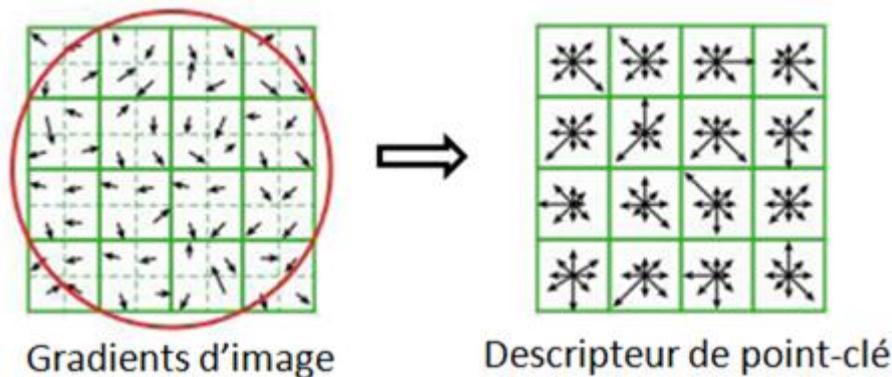
A partir de cette matrice, le critère de sélection de Lowe est le suivant :

$$R^L = \frac{T_r(M^H s)}{\text{Det}(M^L)} \leq \frac{(r+1)^2}{r} \quad (3.4)$$

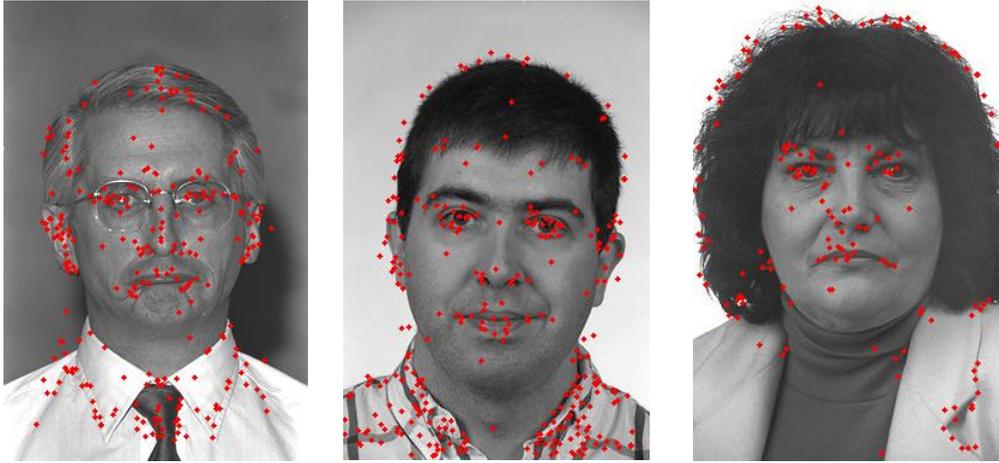
$$\text{O  } M^L = e^{-\frac{x^2+y^2}{2\sigma^2}} * \begin{bmatrix} \frac{\partial I^2}{\partial x} & \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \\ \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} & \frac{\partial I^2}{\partial y} \end{bmatrix}$$

Avec  $\text{Det}(\cdot)$ , le d terminant,  $T_r$  la trace et  $r$  un param tre de crit re. Une valeur  lev e du param tre  $r$  permet de s'assurer que le point consid r  est un point d'int r t.

Un point d'int r t est d fini par 5 param tres  $(x, y, \sigma, \theta, \vartheta)$ , ou les point  $(x, y)$  repr sentent les coordonn es dans l'image originale. Les param tres  $(\sigma, \theta)$  d crivent l' chelle et respectivement l'orientation. Le vecteur  $\vartheta$  est le vecteur descripteur qui est calcul  en utilisant son voisinage qui est divis  par une grille de  $4 \times 4$ . Ensuite, le gradient est calcul  sur chacune des 16 localisations de la grille puis est quantifi  selon un histogramme   8 orientations. La concat nation de ces  l ments nous permet d'obtenir un vecteur de 128  l ments (Figure 3.1). La figure 3.2 pr sente des exemples de d tection des points d'int r t g n r e par SIFT sur des visages.



**Figure 3.1:** Construction d'un descripteur **SIFT** (ref [1])



**Figure 3.2:** Exemples de d tection de points d'int r t.

### 3.2.2 Ondelettes de Gabor

Le filtre de Gabor, d fini par Dennis Gabor [2], a gagn  beaucoup d'attention dans plusieurs domaines de traitement d'image [3.4]. Les ondelettes de Gabor pr sentent deux propri t s int ressantes: la localisation fr quentielle et la s lectivit  en orientation. Les repr sentations en fr quence et en orientation du filtre Gabor s'apparentent   celles du syst me visuel humain. Les ondelettes en dimension deux de Gabor ont  t  introduit dans le domaine de la biom trique par Daugman [5] pour la reconnaissance de l'iris. Un noyau de filtre de Gabor est le produit d'une onde complexe sinuso dale avec une enveloppe gaussienne. L'ondelette de Gabor est d finie comme suit :

$$\psi_{\mu,\nu} = \frac{\|K_{\mu,\nu}\|}{\sigma^2} e^{\frac{\|K_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2}} \left[ e^{iK_{\mu,\nu}z} - e^{\frac{-\sigma^2}{2}} \right] \quad (3.5)$$

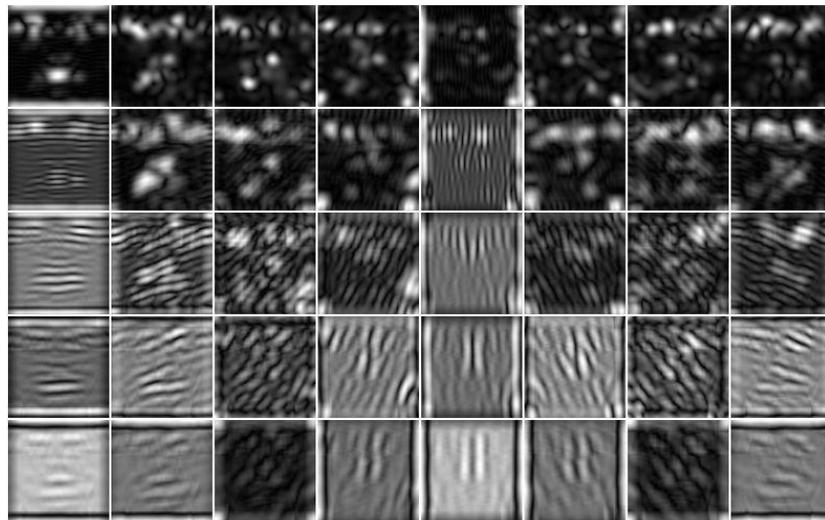
Ou  $z = (x, y)$  est le point de coordonn es  $(x, y)$  dans le plan image. Les param tres  $\mu$  et  $\nu$  d finissent l'orientation et la fr quence des noyaux de Gabor.  $\| \cdot \|$  est l'op rateur norme et  $\sigma$  l' cart-type de l'enveloppe gaussien. La repr sentation en ondelettes de Gabor d'une image r sulte du produit de convolution de l'image avec une famille de noyaux de Gabor de fr quences et d'orientations diff rentes comme d finis par l' quation 3.5. La convolution de l'image  $I$  et d'un noyau de Gabor  $\psi_{\mu,\nu}z$  est d finie par :

$$G_{\mu,\nu}(z) = I(Z) * \psi_{\mu,\nu}(Z) \quad (3.6)$$

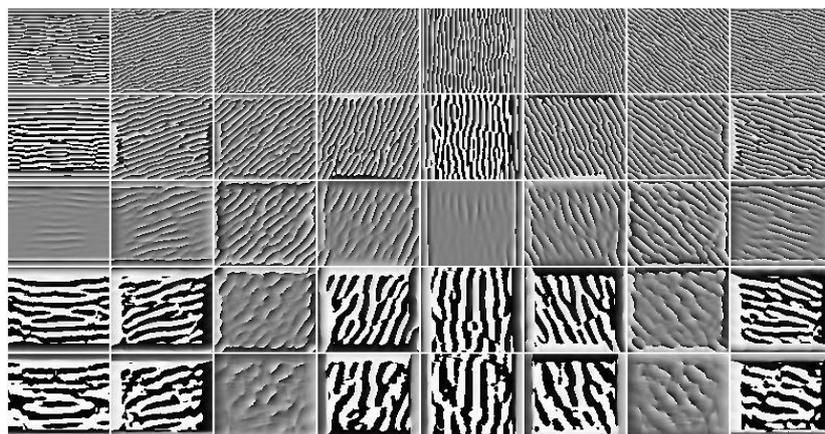
Si nous considérons l'exemple de l'image ci-dessous, l'application des ondelettes de Gabor pour cinq échelles et huit orientations donne:



Input Image



Amplitudes correspondantes



Phase correspondantes

**Figure 3.3:** Image et sa transformé de Gabor (5 scales and 8 orientations)

L'avantage de représenter le visage avec le filtre de Gabor est que ce dernier peut extraire des caractéristiques faciales et capturer les informations du visage dans des orientations et des résolutions différentes. De plus, ces filtres sont robustes aux changements d'illuminations, aux distorsions et aux

variations d'échelles. En effet, la convolution d'une image avec une banque de 40 noyaux de Gabor (5 échelles et 8 orientations) résulte en 40 cartes d'amplitude et 40 autres cartes de phase qui sont de même taille que l'image d'origine (figure 3.3). Par exemple, si on ne considère que les amplitudes, chaque pixel est décrit par un vecteur de dimension 40, ce dernier est également appelé « Jet » et il a été utilisé largement dans les systèmes les plus anciens, tel que EGBM.

### 3.2.3 Local Binary Pattern

#### A. Généralité

L'opérateur LBP a été proposé initialement par Ojala et al ;[6], dans l'objectif de caractériser la texture d'une image. Le calcul de l'opérateur LBP consiste pour chaque pixel à seuiller ses huit voisins directs avec un seuil dont la valeur est le niveau de gris du pixel courant. Un voisin prendra la valeur 1 si sa valeur est supérieure ou égale au pixel courant et 0 si sa valeur est inférieure.

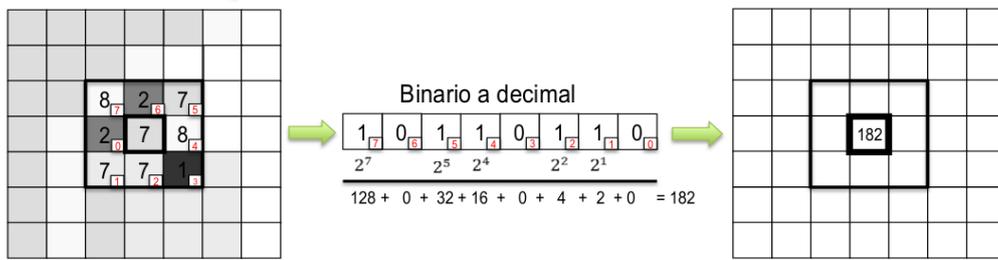


Figure 3.4: L'opérateur LBP.

Par la suite l'opérateur LBP a été étendu en utilisant des voisinages de taille différente. Dans ce cas, un cercle de taille R autour du pixel est considéré. Les valeurs des P voisins sont comparées avec le pixel central. On adopte la notation (P,R) pour définir le voisinage de P points de rayon R d'un pixel. La figure 2.4 illustre trois voisinages pour des valeurs de R et de P.

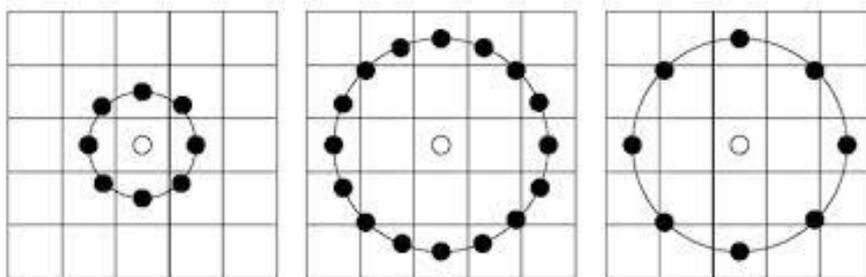


Figure 3.5: Trois voisinages pour des R et P différents.

## B. Principe de LBP

Soient  $g_c$  le niveau de gris du pixel central et  $g_p (p = 1 \dots p)$  les niveaux de gris de ses voisins. L'indice LBP du pixel courant est calcul e comme :

$$LBP_{P,R}(x_c, y_c) = \sum_{p=1}^p S(g_p - g_c) 2^{p-1} \quad (3.7)$$

Ou :

$$S(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (3.8)$$

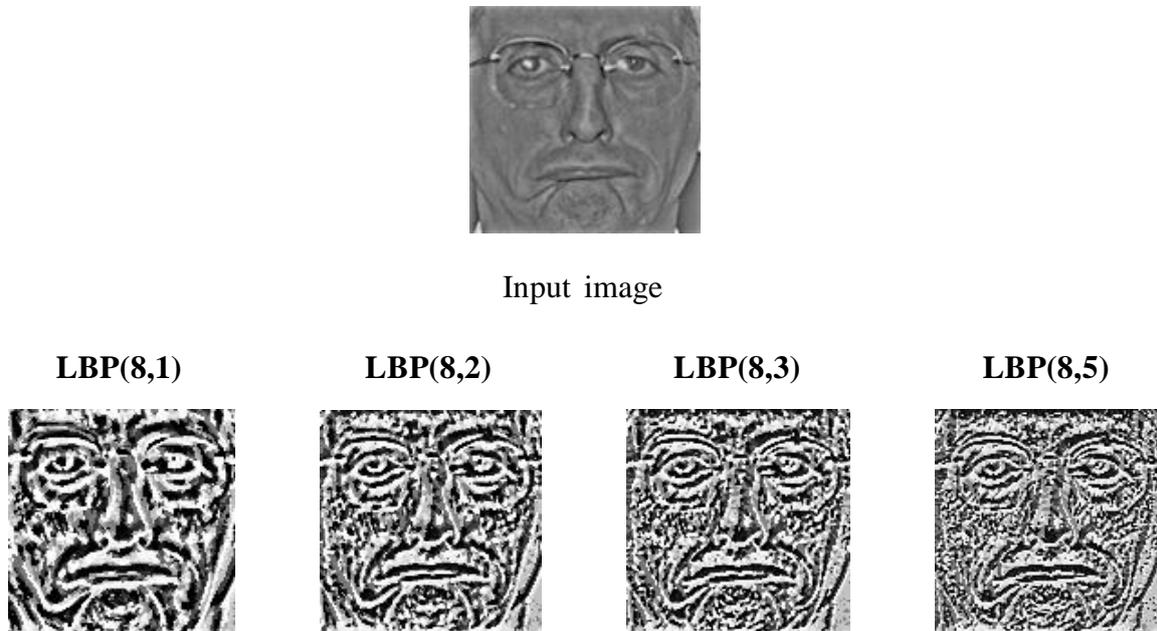
Ou  $x_c, y_c$  sont les coordonn es du pixel courant,  $LBP_{P,R}$  est le code pour le rayon R et le nombre de voisins P. L'op rateur LBP obtenu avec P=8 et R=1 s' crit  $LBP_{8,1}$ . La principale diff rence est que les pixels doivent d'abord  tre interpol es pour obtenir les valeurs des points sur le cercle (voisinage circulaire au rectangulaire).

Une autre extension   l'op rateur LBP d'origine est le LBP uniforme. Un code LBP est appel e uniforme s'il contient au plus deux transitions de bit de 1   0 ou vice versa lorsque la chaine binaire et consid er e circulaire. Par exemple, 00011110 et 10000011 sont des codes uniformes. L'utilisation d'un code LBP uniforme, not e  $LBP^{U2}$  a deux avantages. Le premier c'est de conserver l'espace m moire ainsi que le temps de calcul, le deuxi me est que le LBP<sup>U2</sup> permet de d tecter uniquement les textures locales importantes, comme les spots, les fins lignes, les bords et les coins (pour des exemples de ces textures particuli res). En effet, Ojala et al ont montr e que les LBPs uniformes contiennent plus de 90% de l'information d'une image.

## C. Application de LBP pour la reconnaissance du visage

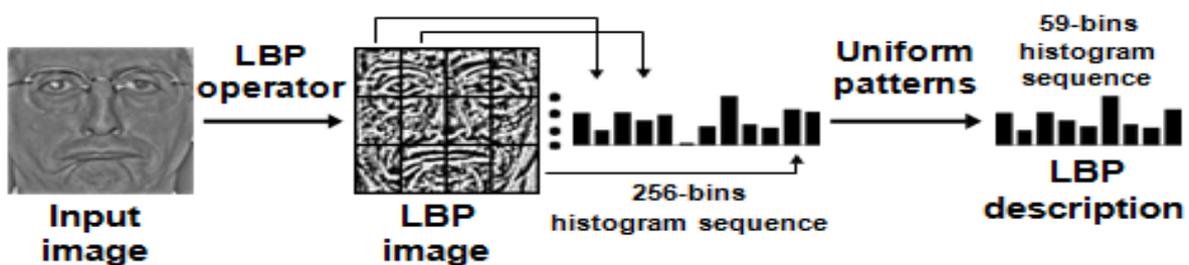
Initialement le LBP a  t e con u comme un descripteur pour r soudre le probl me de classification de texture [6]. Rapidement, il a  t e devenue l'une des caract ristiques les plus populaires pour la reconnaissance faciale. Dans LBP original [7], chaque pixel d'une image d'entr e est attribu e   un nombre d cimal (appel e label LBP) qui est calcul e par seuillage binaire de son niveau de gris avec ses P voisins situ es sur un cercle de rayon r centr e sur le pixel lui-m me. Une interpolation bilin aire est exploit e pour calculer les valeurs des pixels voisins si ils ne sont pas au centre d'un pixel. Ce sch ma de codage est appel e op rateur de LBP et not e LBP (P, R) (pour plus de d tails, voir Figure. 2.3 et Figure 2.4). L'application de l'op rateur LBP sur chaque pixel d'une image du visage produit une image de LBP, qui contient des informations tr s importantes pour la reconnaissance faciale: les textures faciales

micro locales. Si nous considérons le visage de la figure 3.3, l'application de l'opérateur LBP à huit voisins et plusieurs rayons donne les figures LBP ci-dessous.



**Figure 3.6:** Une image de visage et ses LBP

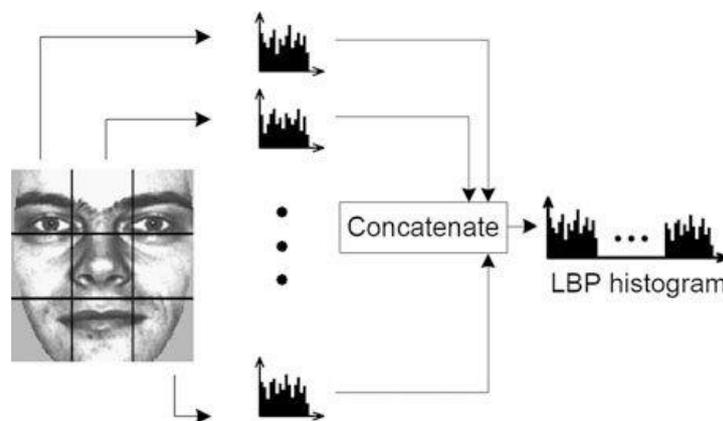
Généralement, le vecteur LBP est représenté par  $2^P$  bins, mais les études statistiques des étiquettes LBP sur les différents types d'images ont révélé que certains bins sont plus fréquents que les autres. Ces principaux bins qui ne contiennent pas plus de 2 transitions de 1 à 0 et vice versa sont appelés LBP uniforme et sont utilisés pour réduire la taille du vecteur caractéristique de LBP. Cette opération de compression résulte en un vecteur de caractéristiques plus court que l'originale et ainsi une classification très rapide, mais au contraire, elle provoque une légère diminution des performances de précision. Toutes ces étapes sont rassemblées dans la figure 3.7.



**Figure 3.7:** Calcul du descripteur LBP

Une fois le code LBP calculé pour tous les pixels de l'image, on calcule l'histogramme de cette image LBP pour former un vecteur de caractéristiques représentant l'image faciale. En

réalité, afin d'incorporer plus d'information spatiales au vecteur représentant le visage, on divise tout d'abord l'image codée par l'opérateur LBP en petites régions et l'histogramme est construit pour chaque région. Finalement, on concatène tous les histogrammes des régions afin de former un histogramme global représentant les caractéristiques facial (voir Figure 3.8). L'efficacité du code LBP comme indice facial s'explique par le fait que le LBP permet de caractériser les détails fins d'un visage. Quand seules les LBPs uniformes sont utilisés, tous les codes LBPs non-uniformes sont étiquetés avec une étiquette unique, alors que chacun des codes uniformes est regroupé dans un histogramme unique. Par exemple, quand  $P = 8$ , nous avons 58 codes uniformes mais l'histogramme est de dimension 59. De la même manière,  $P = 6$  produit un histogramme de dimension 33.



**Figure 3.8.** Exemple d'une représentation faciale basée sur le LBP [8].

Etant donnés deux histogrammes de LBP  $H^1$ ,  $H^2$  de deux visages, l'étape suivante est d'utiliser une métrique pour calculer la similarité entre ces deux histogrammes. En testant les trois métriques  $\chi^2$ , Histogrammes intersection et Log-likelihood statistic, Ahonen et al ; [7] ont observé que la première métrique donne les meilleurs résultats :

$$\chi^2(H^1, H^2) = \sum_i \frac{(H_i^1 - H_i^2)^2}{(H_i^1 + H_i^2)} \quad (3.9)$$

### 3.2.4 L'analyse en Composantes Principales (Eigenface) pour la reconnaissance du visage

La technique Eigen face est l'une des techniques les plus utilisée pour la reconnaissance du visage. Sa popularité est due à sa nature relativement simple, son fondement mathématique fort et ses bons résultats.

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

$$S = \{x_1, x_2, \dots, x_M\} \quad (3.10)$$

La méthode Eigen face est constituée de deux processus ou phases: un processus d'apprentissage et un processus de reconnaissance. Ils peuvent être résumés comme suit :

#### A. Processus d'apprentissage

- Déterminer la moyenne des vecteurs images, tel quel :

$$\psi = \frac{1}{M} \sum_{n=1}^M x_n \quad (3.11)$$

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

$$\Phi_i = x_i - \psi \quad 1 \leq i \leq M \quad (3.12)$$

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

$$\lambda_k = \frac{1}{M} \sum_{n=1}^M (\mu_k^T \Phi_n)^2 \quad (3.13)$$

soit maximale. ou  $\mu_k, \lambda_k$  sont respectivement les vecteurs propres et les valeurs propres de la matrice de covariance  $C$  définie par :

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T \quad (3.14)$$

La figure 2.9 montre les cinq premières Eigenfaces associées aux plus grandes valeurs propres, et calculées sur une base d'apprentissage.



**Figure 3.9:** Exemple de décomposition en Eigenfaces.

## B. Processus de reconnaissance

Plusieurs travaux [9] ont démontrés qu'en pratique, un nombre  $M \ll M$  de Eigenfaces est généralement suffisant pour identifier efficacement les visages. Les  $M$  visages propres considérés correspondent aux plus grandes valeurs propres  $\lambda_k$  car elles encodent les directions dans lesquelles les variations sont les plus marquées.

Le processus de reconnaissance se subdivise en trois étapes: l'image d'entrée  $I(N, N)$  contenant le visage à identifier, est transformée en un vecteur  $X$  de dimension  $(N^2 * 1)$  puis comparé avec le vecteur moyen  $\omega$  défini par l'équation 3.15. La différence entre le vecteur  $X$  et le vecteur moyen  $\omega$  est ensuite multipliée par l'ensemble des vecteurs propres  $\mu_k$  pour générer le vecteur de poids  $\Omega^T$ , tel que:

$$\omega_k = \mu_k^T (X - \omega) \quad k \in \{1, \dots, M\} \quad (3.15)$$

$$\Omega_X^T = [\omega_1, \omega_2, \dots, \omega_M] \quad (3.16)$$

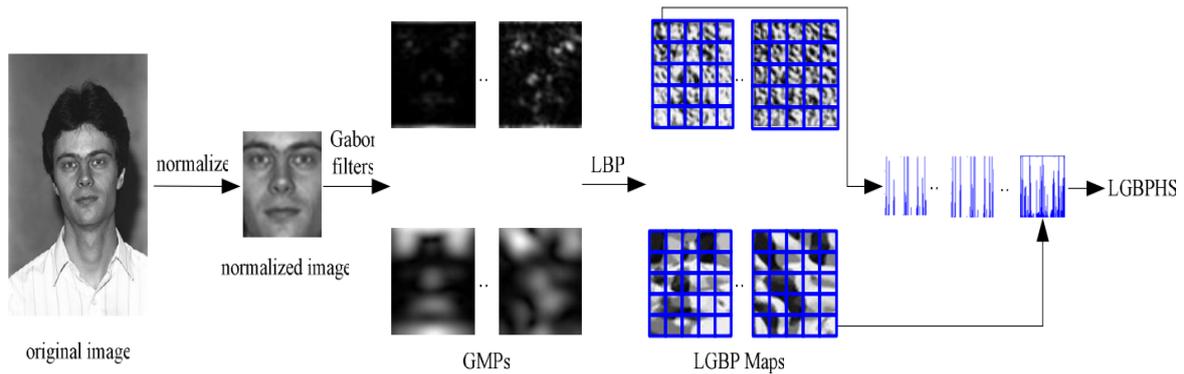
La classe du visage qui fournit la meilleure description de l'image d'entrée est déterminée en minimisant la mesure de similarité entre le poids  $\Omega_X^T$  et les poids de la base d'apprentissage.

$$\Omega_{X_k}^T, \quad k = 1, \dots, M$$

### 3.2.5 Local Gabor Binary Patterns Histogrammes

Local Gabor Binary Patterns Histogrammes Sequence (LGBPHS) [10], qu'on peut traduire par Séquence d'Histogramme Local à Motif Binaire de Gabor est une variante du LBP. Elle représente le visage à partir de deux descripteurs de texture locaux composés d'ondelettes de Gabor et LBP. Les statistiques d'ordre élevé de LBP ont démontré une bonne robustesse à de nombreuses variations du visage, et en particulier à la variation de l'occlusion partielle, comme le montre la figure 2.10. La représentation du visage à base de LGBPHS peut être obtenue par la procédure suivante:

- L'image d'entrée est normalisée et filtrée par un ensemble d'ondelettes de Gabor (famille de noyaux de Gabor).
- Chaque image résultant de la convolution de l'image d'entrée et l'ensemble d'ondelettes de Gabor est divisée en  $K$  blocs non chevauchés et l'histogramme de LBP est calculé à partir de ces blocs locaux.
- En fin, tous les histogrammes LBP de toutes les régions locales sont concaténés pour former la représentation finale du visage (Figure 3.10)



**Figure 3.10:** Illustration Gabor- LBP ([10])

Après la réussite du descripteur LBP pour la reconnaissance du visage [6], de nombreuses variantes ont été proposées pour la reconnaissance du visage au cours des dernières années. Afin de résoudre le problème de sensibilité au bruit dans les régions uniforme Tan et al. [11], ont proposé, suite aux travaux de Ahonen et al., un nouveau descripteur appelé Local Ternary Patterns (LTP) (Modèle Ternaire Locale), qui s'inspire du LBP dans le calcul des variations locales des textures et peut être considéré comme une généralisation de celui-ci au système ternaire. Comme le descripteur LBP, il donne une mesure de texture invariante calculée à partir de l'analyse d'un voisinage local. Tan et al. proposent 3 nouveaux concepts qui permettent d'améliorer significativement les performances (plus de 40% sur la Data Base FRGC-104 selon leur publication). Ces trois concepts sont le LTP, une méthode de prétraitement de l'image et enfin une méthode de mesure de distance pour la comparaison d'échantillons au format LBP ou LTP. L'LTP a été proposée comme solution au problème de sensibilité qu'éprouve le LBP face au bruit aléatoire et celui de quantification. Le principe est le suivant: alors que les LBP appliquent un seuil égal à la valeur du pixel central, la conversion en Local Ternary Patterns (LTP) attribue la valeur 0 aux pixels dont la valeur se trouve dans un voisinage de la valeur du pixel central, 1 à ceux dont la valeur est au-delà de ce voisinage et -1 à ceux dont la valeur est en dessous. La formulation mathématique est la suivante :

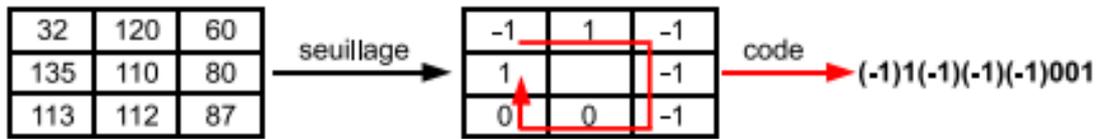
Si  $u$  est le niveau d'un pixel périphérique d'un entourage à convertir,  $i_c$  le niveau du pixel central et  $t$  le seuil :

$$S(u, i_c, t) = \begin{cases} 1 & \text{si } u \geq i_c + t \\ 0 & \text{si } |u - i_c| < t \\ -1 & \text{si } u \leq i_c - t \end{cases} \quad (3.17)$$

Afin de conserver un codage binaire similaire aux LBP, le LTP est séparé en deux codes, un pour les valeurs positives et un pour les valeurs négatives. Ainsi le code (-1)1(-1) (-1)001

donnera 0100001 et 1011001.

La mesure de similarité associée aux LBP et LTP s'inspire de la distance de Hausdorff. Comme fait pour l'opérateur LBP basique, une illustration de l'opérateur LTP basique est donnée à la figure 3.11.

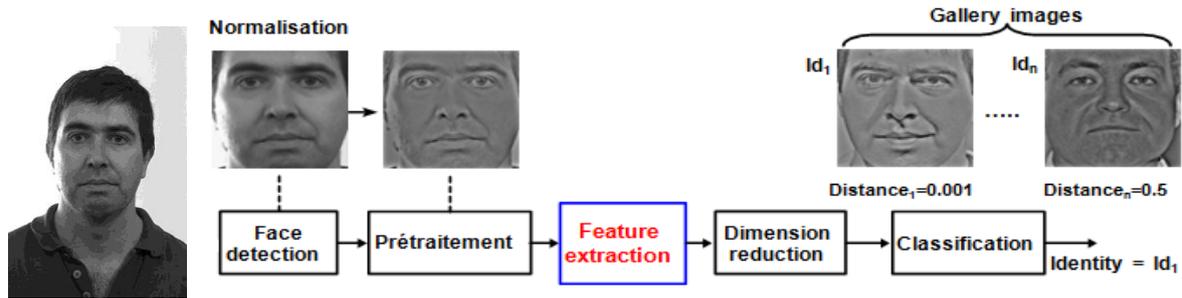


**Figure 3.11:** L'opérateur LTP

Dans [12], un LBP multi-échelle (MB-LBP) est proposé en utilisant des blocs de régions au lieu de pixels. L'LBP original peut être considéré comme un cas particulier du MB-LBP lorsque le bloc de région est un pixel. Une autre version du LBP, appelée Improved LBP (ILBP) est proposée, dans [13], où les auteurs représentent chaque pixel par la moyenne de son voisinage. Comparé à l'LBP, l'ILBP donne de bons résultats quant à la détection du visage. Dans [14], heikkila et al. comparent le centre symétriques des pixels pour former le centre symétriques LBP CS-LBP. Maturana et al. [15], exploitent une technique d'apprentissage supervisé pour chercher les voisins les plus discriminants lors du calcul de LBP appelé Discriminants LBP (DLBP). Cela se fait sur la base de la maximisation d'un critère de séparabilité de classes Fisher-like. Dans [16], un descripteur appelé Patterns of Oriented Edge Magnitude (POEM) a été développées.

### 3.3 Méthode Proposée :

Après une présentation des différentes techniques récentes de reconnaissance du visage, il est temps de présenter notre approche. Comme nous l'avons déjà cité, un système automatique de reconnaissance du visage doit intégrer plusieurs étapes. La première étape consiste à détecter la présence du visage dans l'image et d'en extraire sa position. Ensuite, un ensemble de prétraitement est appliqué à chacun de ces visages extraits. Ils permettent d'instaurer un référentiel commun entre eux. Ainsi, un alignement et une normalisation de l'illumination permettent de faciliter leur comparaison. Des caractéristiques faciales sont ensuite extraites de ces visages prétraités. Ces caractéristiques doivent être suffisamment discriminantes et robustes aux variations intra-identité. L'ensemble des caractéristiques ainsi extraites forment la signature biométrique du visage. Finalement, la comparaison de deux visages est faite à partir de leurs signatures biométriques.



**Figure 3.12:** Architecture générale d'un system de reconnaissance de visage.

Les performances des systèmes de reconnaissance du visage dépendent en grande partie de l'efficacité de ces étapes. La précision de détection du visage et l'extraction des caractéristiques discriminatives ainsi que le choix du classificateur (classifier) adéquat augmente le taux de réussite de la classification.

Dans toutes les applications de reconnaissance du visage (la vidéos surveillance, l'identification et l'authentification de personnes ...), la détection et la localisation du visage est une étape cruciale. En effet, pour reconnaître une personne, il faut d'abord localiser son visage dans l'image, puis extraire les caractéristiques importantes qui seront utilisées pour interroger une base de données afin d'identifier la personne. La détection du visage dans une image est cependant une tâche difficile à cause de la variabilité de la taille, de l'apparence et de l'orientation que peut avoir un visage. De nombreux travaux concernant la détection du visage sont proposés dans la littérature [17]. La méthode la plus utilisée est celle de Viola et Jones [18]. Il s'agit d'une méthode de détection en temps réel d'objet rigides (voiture, visages...). A l'heure actuelle, c'est la méthode la plus performante et c'est elle qu'on a adoptée dans notre travail. Elle consiste à balayer une image à l'aide d'une fenêtre de détection de taille initiale de 24px par 24 px (dans l'algorithme originale) et de déterminer si un visage y est présent. Lorsque l'image a été parcourue entièrement, la taille de la fenêtre est augmentée et le balayage recommence. Ce procédé est étiré jusqu'à ce que la fenêtre fasse la taille de l'image. L'augmentation de la taille de la fenêtre se fait par un facteur multiplicatif de 1.25. Le balayage, quant à lui, consiste simplement à décaler la fenêtre d'un pixel. Ce décalage peut être changé afin d'accélérer le processus, mais un décalage d'un pixel assure une précision maximale.

Cette méthode est une approche basée sur l'apparence, qui consiste à parcourir l'ensemble de l'image en calculant un certain nombre de caractéristiques dans des zones rectangulaires qui se chevauchent. Elle a la particularité d'utiliser des caractéristiques très simples mais très nombreuses.

Il existe d'autres méthodes mais celle de Viola et Jones est la plus performante actuellement. Ce qui la différencie des autres est notamment:

- L'utilisation d'images intégrales qui permettent de calculer plus rapidement les caractéristiques.
- La sélection par boosting des caractéristiques.
- La combinaison en cascade de classifieurs boostés, apportant un net gain de temps d'exécution.

### 3.3.1 Apprentissage du classificateur

Une étape préliminaire et très importante est l'apprentissage du classificateur. Il s'agit d'entraîner le classificateur afin de le sensibiliser à ce que l'on veut détecter, ici des visages. Pour cela, il est mis dans deux situations :

La première où une énorme quantité de cas positifs lui sont présentés et la deuxième où, à l'inverse, une énorme quantité de cas négatifs lui sont présentés. Concrètement, une banque d'images contenant des visages de personnes est passée en revue afin d'entraîner le classificateur. Ensuite, une banque d'images ne contenant pas de visages humains est passée. Dans le cas présent, Viola et Jones ont entraîné leur classificateur à l'aide d'une banque d'images du MIT. Il en résulte un classificateur sensible au visage humain. Il se présente sous la forme d'un fichier XML.

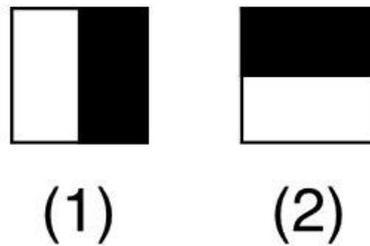
Dans l'absolu, on serait en mesure de détecter n'importe quel signe distinctif à partir d'un classificateur entraîné à cela.

#### A. Les caractéristiques

Une caractéristique est une représentation synthétique et informative, calculée à partir des valeurs des pixels. Les caractéristiques utilisées ici sont les caractéristiques pseudo-Haar. Elles sont calculées par la différence des sommes de pixels de deux ou plusieurs zones rectangulaires adjacentes. Prenons un exemple: Considérons deux zones rectangulaires adjacentes, la première en blanc, la deuxième en noire.

Les caractéristiques seraient calculées en soustrayant la somme des pixels noirs à la somme des pixels blancs. Ces caractéristiques sont calculées à toutes les positions et à toutes les échelles dans une fenêtre de détection de petite taille, typiquement de 24x24 pixels ou de 20x15 pixels. Un très grand nombre de caractéristiques par fenêtre est ainsi généré, Viola et Jones donnent l'exemple d'une fenêtre de taille 24 x 24 qui génère environ 160 000 caractéristiques. L'image précédente présente des caractéristiques pseudo-Haar a seulement deux caractéristiques mais il en existe d'autres, allant de 4 à 14, et avec différentes

orientations. Malheureusement, le calcul de ces caractéristiques de manière classique coûte cher en terme de ressources processeur, c'est là qu'interviennent les images intégrales.



**Figure 3.13:** Caractéristiques pseudo-haar à seulement deux caractéristiques

### B. L'image intégrale

Pour calculer rapidement et efficacement ces caractéristiques sur une image, les auteurs proposaient également une nouvelle méthode, qu'ils appelaient image intégrale. C'est une représentation sous la forme d'une image, de même taille que l'image d'origine, mais qui contient en chacun de ses points la somme des pixels situés au-dessus et à gauche du pixel courant. Plus formellement, l'image intégrale  $ii$  au point  $(x, y)$  est définie à partir de l'image  $i$  par:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (3.18)$$

Le calcul de la somme des valeurs des pixels appartenant à une zone rectangulaire s'effectue donc en accédant seulement à quatre pixels de l'image intégrale: soit un rectangle  $ABCD$  dont les sommets sont nommés dans le sens des aiguilles d'une montre en commençant par le sommet supérieur gauche et soit  $x$  la valeur sous la représentation intégrale d'un sommet  $X$  du rectangle ( $X \in \{A, B, C, D\}$ ). La somme des valeurs des pixels appartenant à  $ABCD$  est, quelle que soit sa taille, donnée par  $c - b - d + a$ . Une caractéristique de Haar étant une combinaison linéaire de tels rectangles  $ABCD$ , son calcul se fait alors en un temps indépendant de sa taille.

### C. Algorithme d'apprentissage basé sur Adaboost

Le deuxième élément clé de la méthode de Viola et Jones est l'utilisation d'une méthode de boosting afin de sélectionner les meilleures caractéristiques. Le boosting est un principe qui consiste à construire un classificateur fort à partir d'une combinaison pondérée de classificateurs faibles, c'est-à-dire, donnant en moyenne une réponse meilleure qu'un tirage aléatoire. Viola et Jones adaptent ce principe en assimilant une caractéristique à un

classificateur faible, en construisant un classificateur faible qui n'utilise qu'une seule caract ristique. L'apprentissage du classificateur faible consiste alors   trouver la valeur seuil de la caract ristique qui permet de mieux s parer les exemples positifs des exemples n gatifs. Le classificateur se r duit alors   un couple (caract ristique, seuil). L'algorithme de boosting utilis  est en pratique une version modifi e d'AdaBoost, qui est utilis e   la fois pour la s lection et pour l'apprentissage d'un classificateur fort. Les classificateurs faibles utilis s sont souvent des arbres de d cision. Un cas remarquable, fr quemment rencontr , est celui de l'arbre de profondeur 1, qui r duit l'op ration de classification   un simple seuillage.

L'algorithme est de type it ratif,   nombre d'it rations d termin .   chaque it ration, l'algorithme s lectionne une caract ristique, qui sera ajout e   la liste des caract ristiques s lectionn es aux it rations pr c dentes, et le tout va contribuer   la construction du classificateur fort final. Cette s lection se fait en entra nant un classificateur faible pour toutes les caract ristiques et en s lectionnant celui avec l'erreur la plus faible sur l'ensemble d'apprentissage. L'algorithme tient  galement   jour une distribution de probabilit  sur l'ensemble d'apprentissage, r valu e   chaque it ration en fonction des r sultats de classification. En particulier, plus de poids est attribu  aux exemples difficiles   classer, c'est- -dire ceux dont l'erreur est  lev e. Le classificateur fort final construit par AdaBoost est compos  de la somme pond r e des classificateurs s lectionn s.

Plus formellement, on consid re un ensemble de  $n$  images  $(x_1, \dots, x_n)$  et leurs  tiquettes associ es  $(y_1, \dots, y_n)$ , qui sont telles que  $y_i = 0$  si l'image  $x_i$  est un exemple n gatif et  $y_i = 1$  si  $x_i$  est un exemple de l'objet   d tecter. L'algorithme de boosting est constitu  d'un nombre  $T$  d'it rations, et pour chaque it ration  $t$  et chaque caract ristique  $j$ , on construit un classificateur faible  $h_j$ . Id alement, le but est d'obtenir un classificateur  $h$  qui pr dise exactement les  tiquettes pour chaque  chantillon, c'est- -dire  $y_i = h(x_i) \forall i \in \{1, \dots, n\}$ .

En pratique, le classificateur n'est pas parfait et l'erreur engendr e par ce classificateur est donn e par :

$$\epsilon_j = \sum_{i=1}^n w_i |h_j(x_i) - y_i| \quad (3.19)$$

Les  $w_i$   tant les poids associ s   chaque exemple et ils sont mis   jour   chaque it ration en fonction de l'erreur obtenue   l'it ration pr c dente. On s lectionne alors   l'it ration  $t$  le classificateur  $h_t$  pr sentant l'erreur la plus faible  $\epsilon_t = \min(\epsilon_j)$ . Le classificateur fort final  $h(x)$  est construit par seuillage de la somme pond r e des classificateurs faibles s lectionn s :

$$h(x) = \begin{cases} 1 & \text{si } \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{si non} \end{cases} \quad (3.20)$$

Les  $\alpha_t$  sont des coefficients calcul s   partir de l'erreur  $\epsilon_t$ :

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \epsilon_t}{\epsilon_t} \quad (3.21)$$

#### D. Cascade de classificateurs

La m thode de Viola et Jones est bas e sur une approche de recherche exhaustive sur l'ensemble de l'image. Elle teste la pr sence de l'objet dans une fen tre   toutes les positions et   plusieurs  chelles. Cette approche est cependant extr mement co teuse en calcul. L'une des id es-cl s de la m thode pour r duire ce co t r side dans l'organisation de l'algorithme de d tection en une cascade de classificateurs. Appliqu s s quentiellement, ces classificateurs prennent une d cision d'acceptation; la fen tre contient l'objet et l'exemple est alors pass  au classificateur suivant, ou de rejet; la fen tre ne contient pas l'objet et dans ce cas l'exemple est d finitivement  cart . L'id e est que l'immense majorit  des fen tres test es  tant n gatives (c.- -d. ne contiennent pas l'objet), il est avantageux de pouvoir les rejeter avec le moins possible de calculs. Ici, les classificateurs les plus simples, donc les plus rapides, sont situ s au d but de la cascade, et rejettent tr s rapidement la grande majorit  des exemples n gatifs. Cette structure en cascade peut  galement s'interpr ter comme un arbre de d cision d g n r , puisque chaque n ud ne comporte qu'une seule branche [18].

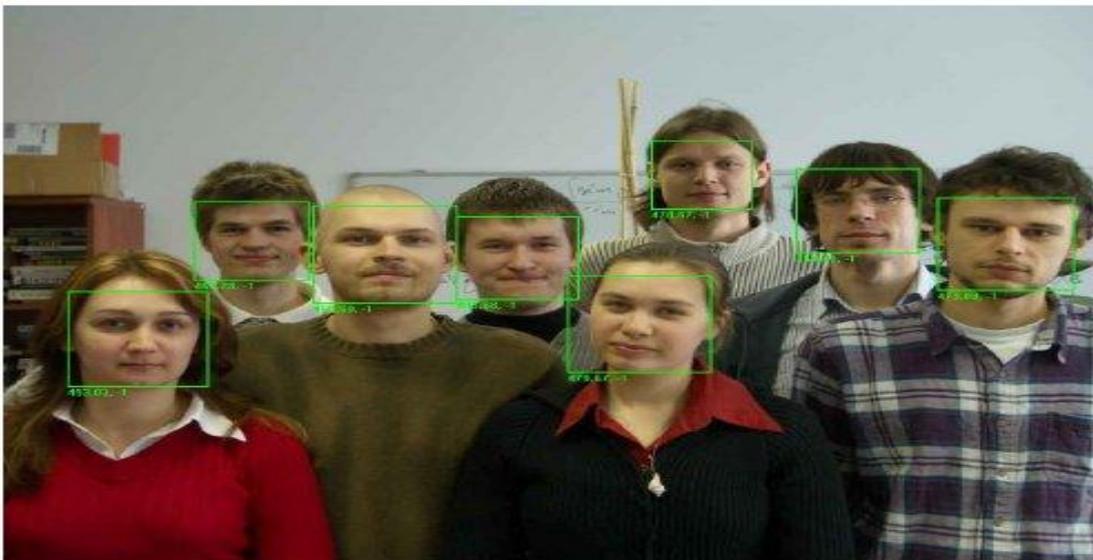
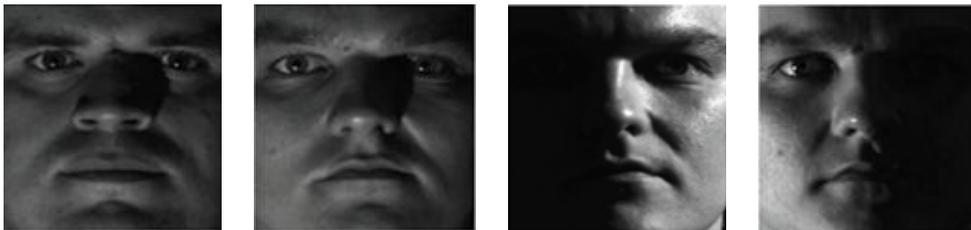


Figure 3.14: Exemple de d tection de visage.

### 3.3.2 Prétraitement

La phase de prétraitement d'une modalité biométrique est une étape très importante. Si on prend le problème des conditions d'éclairage, il est évident que les variations d'illumination affectent les performances des systèmes de biométrie faciale, puisque des conditions différentes d'éclairage peuvent conduire à des images très différentes d'une même personne. Dans [19], Adini et al. ont observés que les variations d'illumination sont souvent plus importantes que les variations d'identité. Visuellement, on peut voir sur la figure que les différences entre deux images de visage de deux individus différents (paire de gauche) peuvent être moins importantes que les différences entre deux images d'un même individu prises avec des conditions d'illumination différentes (paire de droite). Ceci explique la chute des performances des systèmes de reconnaissance faciale en cas de variations d'illumination: les variations interpersonnelles deviennent moins significatives que les variations d'illumination intra personnelles.

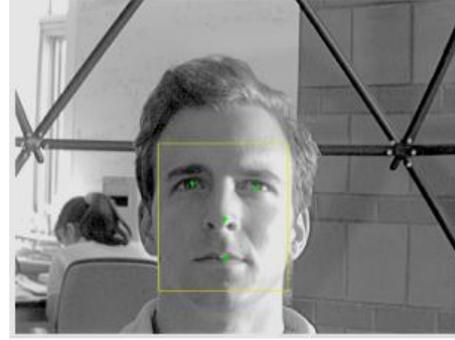


**Figure 3.15:** L'apparence du visage change de manière importante dans le cas d'illumination Variable.

En plus de son effet négatif sur la performance de reconnaissance, un autre problème est posé par les conditions d'illumination : c'est la détection du visage (voire la figure 3.16). Il est clair que les conditions d'illumination affectent également les performances des algorithmes de détection du visage. La littérature propose plusieurs méthodes en réponse à ce problème. Celle dont on a adoptée est proposée par tan et triggs [11]. Il semble que c'est la plus performante, comparée aux autres, en terme de résultats obtenus. Elle est constituée de trois étapes: la correction gamma, le filtrage avec un filtre à Différence de Gaussiennes (DoG) et enfin une normalisation. Etant donnée l'image  $I$ , elle est traitée comme suit:



(a) Sans prétraitement, le détecteur ne fonctionne pas



(b) Avec prétraitement, le détecteur fonctionne

**Figure 3.16** : Illustration pour la détection de visages.

### 1-Correction de gamma

$$I(x, y) \leftarrow I(x, y)^\gamma, \text{ ou } \gamma \in [0, 1] \quad (3.22)$$

### 2-Filtrage DoG

L'image est convoluée à un filtre DoG

$$I = DoG * I \quad (3.23)$$

L'expression du filtre est donnée par:

$$DoG = \frac{1}{2\pi\sigma_{ph}^2} e^{-\frac{-x^2+y^2}{2\sigma_{ph}^2}} - \frac{1}{2\pi\sigma_H^2} e^{-\frac{-x^2+y^2}{2\sigma_H^2}} \quad (3.24)$$

Les termes  $\sigma_{ph}, \sigma_H$  correspondent aux écarts-types des filtres passe-bas gaussiens.

### 3. Normalisation

$$I(x, y) \leftarrow \frac{I(x, y) - \mu_I}{\sigma_I}$$

$$I(x, y) \leftarrow \frac{I(x, y)}{(\text{mean}(|I(x, y)|^a))^{1/a}} \quad (3.25)$$

$$I(x, y) \leftarrow \frac{I(x, y)}{(\text{mean}(\min(\tau, |I(x, y)|^a))^{1/a}}$$

$$I(x, y) \leftarrow \tau \tanh(I(x, y)/\tau)$$

Où  $a$  et  $\tau$  sont choisis de manière empirique.



**Figure 3.17:** les différentes étapes de l'algorithme de tan & triggs (image [11])

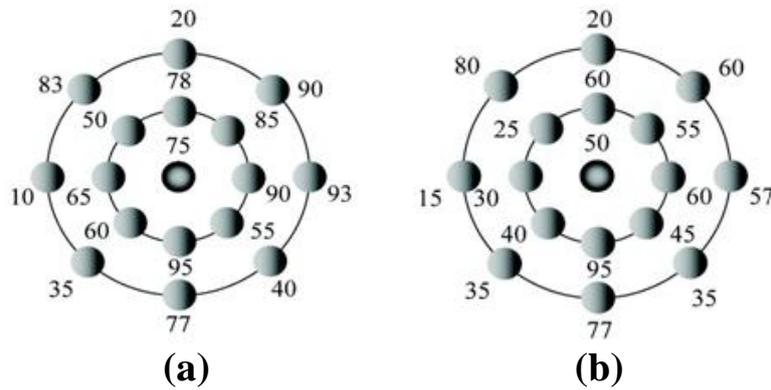
### 3.3.3 LBP Modifiée (MLBP)

Dans la recherche d'une amélioration des performances du descripteur LBP classique, nous proposons dans cette section une nouvelle variante de ce descripteur appelée MLBP, dédié à la reconnaissance faciale. Nous prouvons sa performance par rapport à d'autres techniques populaires.

Dans l'opérateur LBP original, l'information est codé par la différence entre le pixel central  $f_c$  et ces voisins. La notation LBP (P, R) signifie un voisinage de P points de rayon R par rapport au pixel central. Cette notation est généralement écrite sous la forme  $LBP_{P,R}$ . L'opérateur  $LBP_{P,R}$  produits 2P valeurs différentes, il a été démontré que certaines valeurs sont plus importantes que les autres; il est donc possible d'utiliser uniquement un sous-ensemble des 2P LBPs pour représenter une image. Ojala et al .[7], appelaient ces ensembles Uniform LBP et ils les ont noté  $LBP_{P,R}^{U2}$ . Dans notre variante MLBP, l'information locale est codée en prenant la différence entre les pixels situé au rayon R et leurs voisins adjacents situés au rayon R-1.

Le but d'extraction des caractéristiques dans un système de RF est de capter les caractéristiques les plus intrinsèques et discriminantes, pour former une représentation puissante de la forme considérée (le visage dans notre cas). Ces caractéristiques devraient être robustes aux différentes altérations que peut subir la forme. Les dites caractéristiques devraient, en fait, maximiser la variation extra-classe entre les images de différents individus tout en minimisant la variation intra-classe entre ceux d'un même individu.

Dans l'opérateur LBP classique, le code est calculée par seuillage des voisins en faisant référence au pixel central, qui se révèle produire une séquence binaire. L'histogramme des codes est utilisé comme descripteur de texture. Dans notre approche, on extrait le code par seuillage des pixels dans le voisinage du rayon R avec les pixels voisins du plus petit rayon R-1. L'histogramme généré par ce code est construit de la même manière que celle utilisée dans l'opérateur LBP d'origine.



**Figure 3.18** : Deux structures différentes avec le même Code LBP.

En se référant à la figure 3.18, nous pouvons voir que lors de l'application du descripteur LBP classique, les deux structures locales (a) et (b), ont le même code: (01110100) pour LBP<sub>8,1</sub> et (10110100) pour LBP<sub>8,2</sub> quoiqu'elles sont tout à fait différentes. Un autre inconvénient de LBP d'origine est sa sensibilité au bruit [11]. Afin d'améliorer la capacité discriminative du LBP et améliorer son insensibilité au bruit, nous proposons dans ce travail une version simple et efficace appelé MLBP [20].

Dans notre approche, l'information locale des régions est codée en prenant la différence entre les pixels situés à un rayon R et leurs voisins adjacents situés à un rayon R-1 et le résultat est stocké sous forme d'un nombre binaire. Ainsi, l'indice LBP<sub>(P,R1,R2)</sub> désigne le voisinage des P points situés à un rayon R1 par rapport à ceux situés à un rayon R2. L'opérateur LBP<sub>(P,R1,R2)</sub> peut être défini comme:

$$LBP_{(P,R1,R2)} = \sum_{P=0}^{P-1} S(f_{P,R1} - f_{P,R2}) 2^P \quad , S(f_{P,R1} - f_{P,R2})$$

$$= \begin{cases} 1 & \text{if } f_{P,R1} \geq f_{P,R2} \\ 0 & \text{otherwise} \end{cases} \quad (3.26)$$

$$LBP_{P,R1,R2}^{U2} = \begin{cases} z & \text{if } U(LBP_{P,R1,R2}) \leq 2, LBP_{P,R1,R2} = I_z, I_z \in I \\ & \text{where } |I| = (P-1)P + 1 \\ (P-1)P + 2 & \text{otherwise} \end{cases} \quad (3.27)$$

Où:

$$U(LBP_{P,R1,R2}) = \|S(f_{P-1,R1} - f_{P-1,R2}) - S(f_{0,R1} - f_{0,R2})\|$$

$$+ \sum_{P=0}^P \|S(f_{P,R1} - f_{P,R2}) \cdot S(f_{P-1,R1} - f_{P-1,R2})\| \quad (3.28)$$

Il est évident que notre descripteur LBP proposé est insensible au bruit, étant donné que le multi-seuil est utilisé par rapport à l'utilisation de la valeur de gris central, et peut également caractériser différents motifs structuraux plus que le descripteur LBP d'origine. En effet, comme mentionné précédemment, avec le descripteur LBP d'origine, les motifs (a) et (b) ont le même code. Mais si nous appliquons notre descripteur  $LBP_{p,R1,R2}$ , ces deux motifs auront des codes distincts: pour  $LBP_{8,2,1}$  les codes résultants sont 01001001 et 01101001 pour (a) et (b) respectivement. Nous pouvons même étendre notre descripteur aux cas multi-rayons et multi-voisinages en utilisant:  $LBP_{8,2,4}^{U2}$ ,  $LBP_{16,1,2}^{U2}$ ,  $LBP_{12,1,2}^{U2}$ ,  $LBP_{8,1,2}^{RI}$ ,  $LBP_{8,2,3}^{RIU}$ ...

### 3.3.4 Classification

Le problème de la classification consiste à trouver l'identité d'un objet parmi d'autres. Des bases de données sont ainsi spécialement conçues. Elles contiennent tout un ensemble de forme correspondant aux différents objets. Il s'agit dans notre cas de bases de données qui contiennent les visages d'individus. L'objectif est de développer des algorithmes de reconnaissance qui permettent de reconnaître chacun de ces individus parmi tous les autres. La méthode classique consiste à extraire l'empreinte biométrique de la personne, ensuite de la comparer aux empreintes biométriques de toutes les personnes de la base de données.

Soit la matrice  $D \in R^{N \times M}$  dont les colonnes  $d_k \in R^N$  correspondent aux empreintes biométriques des  $M$  personnes enrôlées. Avec un système biométrique idéal, une empreinte biométrique test  $y$  doit satisfaire :

$$y = Dx \quad (3.29)$$

Avec  $x \in R^M$  un vecteur dont toutes les entrées sont nulles sauf l'entrée correspondante à l'identité de  $y$  qui vaut 1.

Dans la réalité, les empreintes biométriques sont obtenues via des algorithmes d'extraction de caractéristiques. Ceux-ci ne sont pas forcément robustes à tous les changements que peut subir une image, ainsi deux empreintes biométriques d'une même personne donnent bien souvent (pour ne pas dire toujours) deux vecteurs de caractéristiques différents (variance intra classe). Pire, deux images de personnes différentes peuvent donner deux vecteurs de caractéristiques proches (variance inter\_classe).

L'identification d'une personne  $P$  à partir d'une image  $I_P$  se résume en trois étapes :

- Calculer l'empreinte biométrique  $y$  de  $I_P$

- Rechercher, parmi les empreintes de la base de données  $D$ , l'empreinte  $d_f$  la plus proche de  $y$
- Retour de l'identité correspondant à  $d_f$ .

La recherche de l'identité peut se formaliser ainsi:

$$d_f = \min_k \|d_k - y\|_2^2, \forall k \text{ avec } \|d_k - y\|_2^2 \quad (30)$$

$d_f$  étant la mesure de la dite similarité de deux empreintes biométriques.

### 3.3.5 Approche « Sparse Representation–based Classification»

Dans cette section, nous résumons l'approche SRC (Sparse Representation-based Classification) initialement présentée par Wright et al. dans [21]. Cette approche repose sur une décomposition parcimonieuse d'une image test (ou son vecteur caractéristique extrait à partir de cette image) de la base de données des personnes déjà enrôlées.

L'idée est qu'en minimisant la norme  $l^1$  lors de la décomposition, de nombreuses classes vont être rejetées et la classification ne sera que plus pertinente. L'approche est schématiquement présentée à la figure 2.19.

Etant donné le vecteur caractéristique  $y \in R^n$  d'une image test et  $D \in R^{n \times m}$  la matrice dont les colonnes sont les vecteurs caractéristiques des personnes enrôlées, l'approche **SRC** peut alors se décomposer en 3 étapes :

- Décomposition parcimonieuse de  $y$  sur  $D$  (la matrice  $D$  peut alors être vue comme un dictionnaire dont les atomes sont les vecteurs caractéristiques de la base de données).

$$\hat{x} = \arg \min_x (\|y - Dx\|_2 + \lambda \|x\|_1) \quad (3.31)$$

Note: Le vecteur  $x$  contient beaucoup de zéros.

- Calcule du vecteur des résidus  $r \in R^M$  (à partir de  $\hat{x}$ ) correspondant à la dissimilarité entre  $y$  et les colonnes de  $D$ .
- Déduction de l'identité correspondante à  $y$  en sélectionnant l'identité correspondante au minimum du vecteur des résidus  $r$ .

Les différentes étapes de l'approche sont résumées dans l'algorithme 1 et schématisées à la figure 3.19.

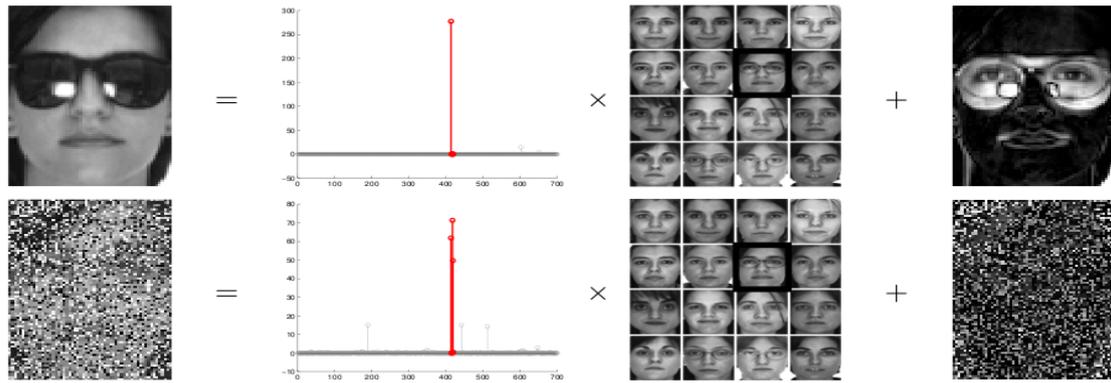


Figure 3.19: Fonctionnement de l'approche SRC (extrait de [20]).

**Algorithme 1**

**Entrées:** La matrice de la galerie  $D \in R^{n \times m}$  et l’empreinte biométrique test  $y \in R^n$ .

**Sorties:** L’identité de  $y$ .

- Normaliser les colonnes de  $D$  de sorte que  $\|d_k\|_2 = 1, \forall k$
- Décomposer  $y$  sur  $D$  selon:  $\hat{x} = \arg \min_x (\|y - Dx\|_2 + \lambda \|x\|_1)$
- Calcul des résidus  $r = \{r_k\}, k = 1, \dots, M; r_k = \|y - D_k \hat{x}\|_2$
- Déduction de l’identité de  $y$ :  $Identité(y) = \operatorname{argmin}_k (r)$

*Retourner Identité(y)*

L’étape critique de l’algorithme est la minimisation  $l^1$  de  $y$  sur  $D$ . Cette étape est traitée dans [21] à l’aide d’un algorithme primal-dual. Dans notre travail, nous avons adopté l’approche fondée sur un seuillage (Orthogonal Matching Pursuit OMP).

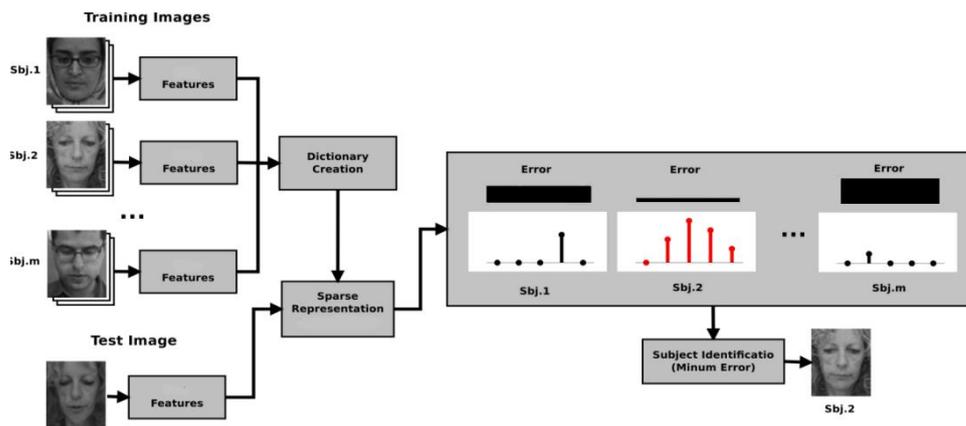


Figure 3.20: Vue schématique de l'approche SRC.

### 3.4 Bases de Données Utilisées:

Les bases de données du visage peuvent être classées en trois catégories selon l'objectif recherché: reconnaissance, détection du visage ou analyse de l'expression faciale. Plusieurs bases de données contenant des informations qui permettent l'évaluation des systèmes de reconnaissance du visage sont disponibles. Toutefois, ces bases de données sont généralement adaptées aux besoins de quelques algorithmes spécifiques de reconnaissance. Chacune d'elles a été construite avec diverses conditions d'acquisition d'images des visages (changements d'illumination, de pose, d'expressions faciales) ainsi que le nombre de sessions pour chaque individu.

#### 3.4.1 La Base ORL

Les données utilisées pour réaliser les tests sur notre système proviennent de la base ORL [22]. Cette base a été collectée entre avril 1992 et avril 1994 par le laboratoire AT&T de l'université de Cambridge. La base contient 40 personnes, chacune étant enregistrée sous 10 vues différentes (figure 3.21). Les images sont de taille  $112 \times 92$  pixels en format JPG et BMP. Pour quelques sujets, les images ont été collectées à des dates différentes, avec des variations dans les conditions d'éclairage, les expressions faciales (expression neutre, sourire et yeux fermés) et des occultations partielles par les lunettes. Toutes les images ont été collectées sur un fond foncé. Les poses de la tête présentent quelques variations en profondeur par rapport à la pose frontale.



Figure 3.21 : Échantillons de la base AT&T.

### 3.4.2 La Base Georgia Tech Visage

Cette base contient 750 images au totale de 50 classe, 15 images pour chaque classe en couleur; prise en deux sessions entre 01/06/99 et 15/11/99 au Centre de Traitement du Signal et de L'image au Georgia Institute de technologie [23]. Les images sont au format JPEG et sont toutes de tailles différentes. Les images de cette base de données montrent des variations des conditions d'éclairage, des expressions faciales et d'échelles (figure 3.22).



**Figure 3.22:** Échantillons de la base Georgia Tech.

### 3.4.3 La Base YALE

Cette base contient 165 images aux niveaux de gris, représentant les visages de 15 individus. Il y a 11 images par individu, chacune correspond à une prise de vue particulière. Les 11 catégories d'images varient selon les critères suivants:

- Variation de l'expression du visage: normal, sad, sleepy, surprised, wink, happy.
- Variation des accessoires: glasses, noglasses.
- Variation de l'éclairage: center-light, left-light, right-light.

Les 15 individus sont déjà numérotés. Afin de nous retrouver dans les 11 catégories, nous associons à chacun un index (voir figure 3.23)



**Figure 3.23:** Images d'une personne de la base Yale

### 3.4.4 La Base YALE-B

Cette base contient 2432 (38 sujets pris sous 9 poses et 64 conditions d'éclairage) [24], chaque classe contient 64 images. Les images sont divisées en cinq sous-groupes en fonction de l'angle entre la direction de la source lumineuse et l'axe central de la caméra ( $12^\circ$ ,  $25^\circ$ ,  $50^\circ$ ,  $77^\circ$ ,  $90^\circ$ ). Des échantillons de cette base sont montrés dans la figure. 3.24.



**Figure 3.24:** Echantillon YALE-B data-base

### 3.5 Résultats Expérimentaux

Les expériences ont été menées sur les quatre bases de données, dont les images sont prises dans des conditions différentes (conditions d'illumination, de pose, expression faciale ....). Le choix principal de ces bases de données est dû au fait qu'elles ont été le siège de multiple sources de variations possibles qui rendent leur classification automatique très difficile pour valider notre approche.

L'approche qu'on a proposée est une approche local basée essentiellement sur deux parties : la première est l'utilisation d'une nouvelle variante de LBP qu'on a appelé MLBP et la deuxième est l'utilisation de la méthode SRC pour la classification des vecteurs MLBP extraits à partir des visages des différents individus. Notre système comprend quatre étapes importantes:

### 3.5.1. D tection du visage

La d tection du visage est la premi re  tape de tout syst me de reconnaissance faciale.  tant donn e une image ou une s quence d'images, l'objectif de cette  tape est de d terminer la pr sence ou non d'un visage dans l'image ainsi que sa localisation. Comme nous l'avons d j  mentionn , la performance d'un syst me de classification automatique du visage d pend de la pr cision de d tection de celui-ci. Le d tecteur de Viola-Jones [18] est utilis  pour extraire la r gion du visage   partir des images des bases ORL et Georgia Tech (figure 3.25), puis toutes les images sont redimensionn es en 130\*150 pixels.



Figure 3.25: D tection du visage par la m thode de Viola-Jones

### 3.5.2. Pr traitement

Apr s la d tection du visage, la phase de pr traitement est ensuite appliqu e. Elle implique g n ralement une normalisation g om trique des visages dans un but d'alignement et une normalisation d' claircissement dans un but de compensation des variations d'illumination. Les zones des visages normalis es sont ensuite utilis es pour l'extraction des caract ristiques faciales. Dans le pr sent travail on a utilis  la m thode de tan and triggs [11], car elle est simple, efficace et   faible complexit  de calcul ce qui permet un traitement des visages en temps r el (voire figure 3.26).

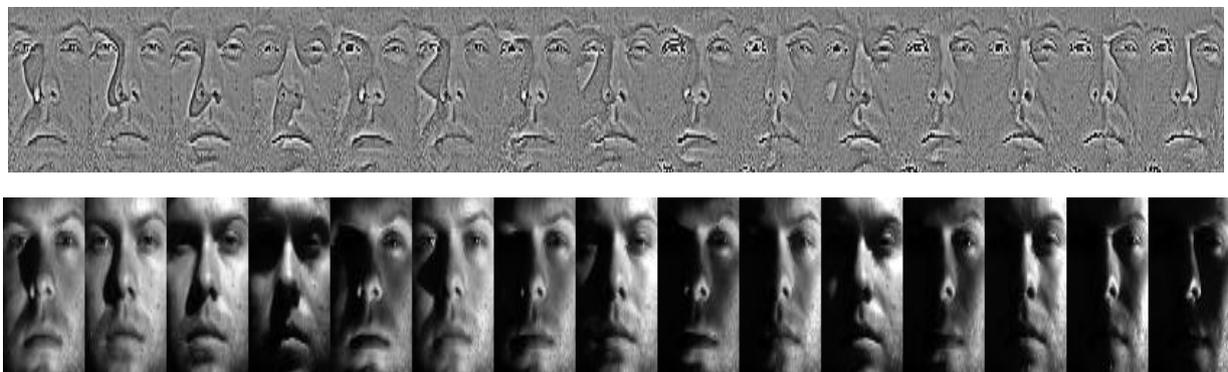


Figure 3.26: La m thode de tan & triggs ( $\alpha=0.1$  and  $\tau=10$ ).

### 3.5.3. Extraction des caractéristiques

Les caractéristiques extraites sont les informations utiles à la phase de reconnaissance. Elles doivent être, dans la mesure du possible, discriminantes et robustes aux changements extérieurs, tels que la pose, l'expression, etc. Les caractéristiques faciales sont modélisées pour fournir la signature biométrique du visage qui est ensuite utilisée dans la phase de classification. L'étude comparative entre notre système (**MLBP+CSC**) et les autres algorithmes, décrits dans le présent document, a d'abord été effectuée sur la base de données des visages **Yal-B**, qui contient 38 sujets (environ 64 images pour chaque classe). Après l'étape de prétraitement des visages, 14 échantillons par sujet sont extraits au hasard pour former l'ensemble d'entraînement et le reste (50 images par classe) sont utilisés dans la phase du test. On opère en suite une sélection des paramètres de notre classificateur dans le but d'optimiser ses performances. Ca concerne les paramètres du descripteur MLBP proposé et le nombre de régions utilisées pour diviser l'image du visage en sous-blocs. Dans la présente étude, nous avons remarqué que l'utilisation de grand rayon pour LBP ne donne pas de bons résultats à cause de la perte d'information, tandis qu'un petit rayon donne de bons résultats et rend l'information spatiale plus locale. Le nombre de bins pour un voisinage de 8 pixels dans LBP uniforme est de 59 bins et 243 bins dans un voisinage de 16 pixels. Le vecteur de l'histogramme global devient  $H = m * 59$  pour un voisinage de 8 pixels et  $H = m * 243$  pour un voisinage de 16 pixels où  $m$  est le nombre de blocs dans l'image. Il est évident que l'utilisation d'un grand nombre de régions résulte en un vecteur de caractéristiques de dimension assez importante provoquant une importante consommation de mémoire et rend la classification plus lente. Un petit nombre de blocs, quant à lui, provoque une perte d'informations spatiales ce qui entraîne une diminution des performances du classificateur. L'art est de trouver un compromis d'équilibre entre les performances de reconnaissance et la longueur du vecteur caractéristique. Nous avons ainsi divisé l'image du visage en 64 sous-blocs (Figure 3.27) et  $LBP_{8,1,2}^{U2}$  est choisi pour extraire le vecteur des caractéristiques.

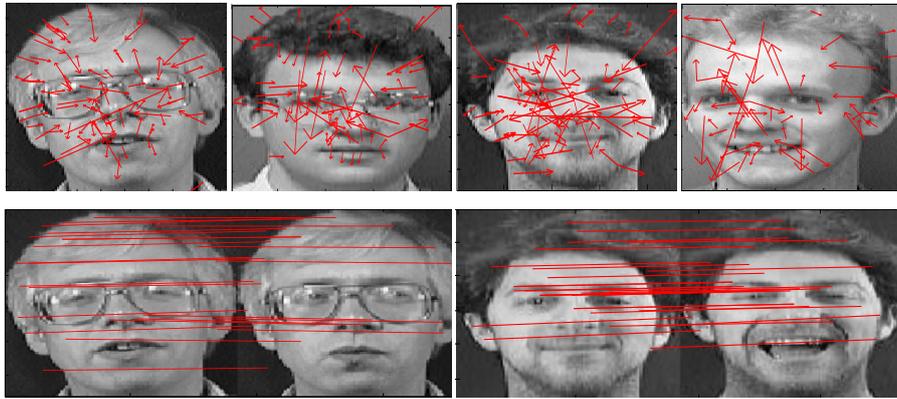


**Figure 3.27:** Exemple d'image faciale divisée en blocs de 8x8

Les résultats de la reconnaissance de la méthode proposée et les algorithmes décrits dans le

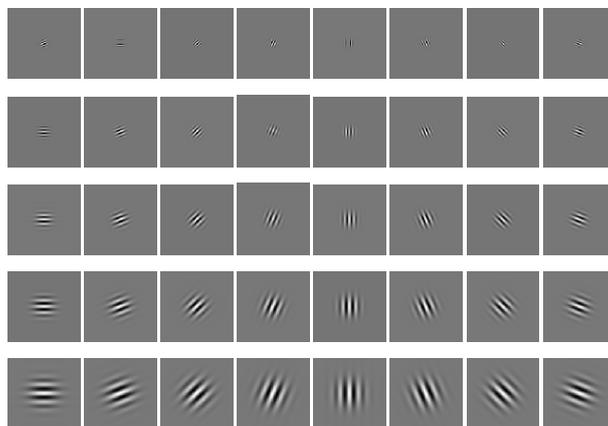
présent document, pour la base Extended Yale-B, sont présentés dans le Tableau.1, avec et sans utilisation de la méthode de Tan &Triggs, pour voir l'influence de l'effet des conditions d'éclairage sur le système proposé.

Avant d'interpréter les résultats, expliquons d'abord l'implémentation des algorithmes de comparaison. Commençons par le SIFT, on a suivi (implémenté) le travail de J. Luo et al. [25]. Après l'extraction des points caractéristiques SIFT, chaque point de l'image de test est comparé indépendamment avec tous les points de toutes les images d'entraînement. La meilleure correspondance pour chaque point se trouve en prenant la distance par rapport à son plus proche voisin. On calcule en suite le rapport de la distance avec le second plus proche voisin. Si ce rapport est inférieure à un certain seuil, le point est retenu sinon le point est rejeté. En fin l'image de test est affectée à la classe disposant de plus de correspondance. L'application du SIFT sur la base ORL est présentée dans la figure 3.28.



**Figure 3.28:** SIFT key point's extraction and matching

Dans une seconde étape, on a combiné le descripteur LBP avec l'ondelette de Gabor [10]. Chaque image est multipliée par une famille de filtres de *Gabor* (voire figure 3.26).



**Figure 3.29 :** Filtre de Gabor

Chaque image, résultante de la convolution de l'image d'entrée et l'ensemble d'ondelettes de Gabor, est divisée en  $K$  blocs non chevauchés (dans notre cas on a utilisé 36 blocs) et l'histogramme LBP est calculé pour chaque bloc. L'histogramme GLBP globale est ensuite construite par une concaténation des histogrammes locaux de chaque bloc. La dimension du descripteur GLBP uniforme dans le voisinage de 8 pixels est :  $HGLBP = m * 59 * 40$  (151040 bins) où  $m$  est le nombre des blocs et 40 est la famille d'ondelette de *Gabor*. Pour la classification des vecteurs caractéristiques  $HGLBP_s$  on a utilisé la distance Chi-Square.

### 3.5.4. Performance

Pour montrer la performance du système proposé, on s'est proposé de le comparer aux méthodes de l'état de l'art : system Ahnen (AH) [6], la célèbre PCA [9], le Local Gabor Binary Pattern [10], Sparse Coding (les éléments du dictionnaire ont été formés par les colonnes des visages eux-mêmes) [21] et le SIFT [24]. La comparaison des résultats expérimentaux a été d'abord menée sur la base de données Extended Yale-B. Après l'étape de prétraitement, on sélectionne au hasard 14 échantillons par sujet pour former l'ensemble d'entraînement. Le reste des échantillons est utilisé comme pour le test.

Nous avons remarqué que l'utilisation d'un petit rayon, ex:  $MLBP_{8,1,2}^{U2}$ ,  $MLBP_{16,1,2}^{U2}$  dans la méthode proposée, fournit une excellente performance, alors que l'utilisation d'un grand rayon dégrade de manière assez significative les performances du descripteur. Ceci peut s'expliquer par le fait qu'un rayon assez important entraîne des pertes d'informations. Maintenant, le nombre de régions, utilisé pour diviser l'image du visage en sous-blocs, doit être défini. Comme nous l'avons déjà signalé, l'utilisation d'un grand nombre de régions génère un grand vecteur de caractéristiques entraînant une consommation élevée de mémoire. Un petit nombre de régions entraîne quant à lui des pertes d'informations spatiales [6]. Afin d'équilibrer les performances de reconnaissance et la longueur du vecteur caractéristique, l'image faciale est divisée en 64 sous-blocs (figure 3.27) et l'élément  $MLBP_{8,1,2}^{U2}$  est retenu pour extraire le vecteur caractéristique. Le Tableau 3.1 résume les résultats de reconnaissance, avec et sans normalisation de l'éclairage par l'algorithme de Tan et Triggs. Comme c'est prévu, LGBP est plus performant sous différentes conditions d'éclairage et présente le meilleur taux de classification. Cependant, l'inconvénient de LGBP est la longueur du vecteur de caractéristique, par exemple l'utilisation d'une famille de Gabor de 40 filtres avec une grille de 64 régions résulte en un vecteur de caractéristiques de dimension 151.040. Cela pourrait causer une forte consommation de mémoire et une classification très lente.

**Tableau 3.1:** Comparaison du taux de classification entre la m thode propos e et les autres M thodes pour la base de donn es Yale-B face database.

Method	Extended Yale-B face database	
	Without Tan and Triggs preprocessing (%)	With Tan and Triggs preprocessing (%)
Original LBP [6]	90.263	97.4211
PCA+Mahdistance [9]	34.008	98.52
SIFT [24]	45.572	67.05
LGBPHS [10]	<b>95.4805</b>	99.3707
SRC [21]	89.89	96.95
Proposed approche	92.4211	<b>99.4158</b>

Sous les m mes conditions, nous pouvons constater que notre approche est plus performante que l’LBP classique, la PCA, l’LGBP et l’algorithme Sparse Coding.

Nous pouvons  galement constater que l’algorithme de pr traitement Tan et Triggs renforce la pr cision des algorithmes et la m thode propos e a les meilleures performances en termes de taux de reconnaissance ainsi que le temps de calcul. Une autre observation int ressante est que l’algorithme de Tan et Triggs renforce les performances de la PCA par une augmentation importantes (64,512%), ce qui signifie que les m thodes globales sont sensibles aux variations de l’ clairage et confirme ainsi les  tudes de Heisele [26].

La m thode LBP suppose que certaines parties de l’image telles que les yeux et le nez correspondent toujours   la m me r gion du visage dans tous les visages de la base de donn es. Ce n’est pas toujours le cas, car ces parties sont souvent r parties dans diff rentes r gions comme c’est le cas des bases de donn es **ORL** et **Georgia-Tech**. Ainsi, les visages de ces bases ne peuvent  tre caract ris s par de petites r gions. En cons quence, nous utilisons des blocks de taille  $32 \times 30$  pixels pour caract riser ces bases de donn es et le Tableau 3.2 montre le taux de classification r sultant. Cela a  t  fait sans l’algorithme de normalisation de lumi re de Tan & Triggs puisque ces bases ne souffrent pas de diff rentes conditions d’ clairage.

**Table.3.2:** Comparaison du taux de reconnaissance entre la méthode proposée et les autres méthodes pour les bases de données ORL et Georgia-Tech data bases.

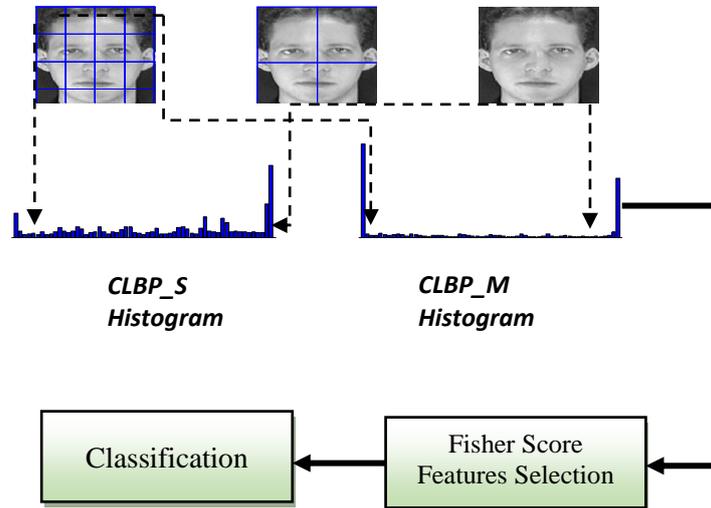
Method	Recognition rate (%)	
	ORL face database	Geor-Tech face database
Original LBP [6]	95	85.4286
PCA+Mahdistance [9]	91.5	62.5
SIFT [24]	90	80.8571
LGBPHS [10]	96.5	79.7143
SRC [21]	94	85.29
<b>Proposed approche</b>	<b>98</b>	<b>88.8571</b>

A partir de ces résultats, nous pouvons constater que le meilleur taux de reconnaissance, et pour les deux bases de données, est obtenu par notre approche. Dans le cas de la base **ORL**, notre approche est plus performante que: l'LGBP, l'LBP original, la PCA, le SIFT est le SC de 1.5%, 3%, 6.5%, 8% et 4% respectivement. Pour la deuxième base **Georgia-Tech**, la meilleure performance est également obtenue par la méthode proposée. La méthode de Ahonen's occupe la deuxième position en terme du temps de calcul tandis que le taux de reconnaissance le plus faible est obtenu par la PCA. Ceci confirme bien le fait que les approches locales sont plus discriminatives que les approches globales.

Concernant le temps de calcul, il est à signaler que notre méthode est plus rapide comparée aux SC [21], LGBP [10] et SIFT [24]. Elle est par contre moins rapide que la PCA et le système de Ahonen's.

Une autre variante de l'LBP appelée Complete Local Binary Patterns **CLBP** a été proposée par Guo et al. [27] pour la classification des textures. La CLBP diffère de l'opérateur LBP par la représentation d'une région d'une image avec l'utilisation du signe et l'amplitude (Magnitude) au lieu de l'utilisation du signe seul comme le fait l'LBP original. La performance de l'opérateur CLBP dans la classification des textures nous a motivés à l'utiliser dans la classification des images faciales. La plupart des descripteurs basés sur l'opérateur LBP reposent sur une grille fixe pour coder une image donnée. Ceci les fragilise et les rend moins robustes pour présenter des images faciales sujettes de variations de poses et d'expression. Pour contourner ce problème, on a utilisé l'histogramme CLBP multi-résolution. Il s'agit de diviser chaque image du visage en des sous-régions de plus en plus fines et pour chaque sous-région on calcule l'histogramme du CLBP correspondant.

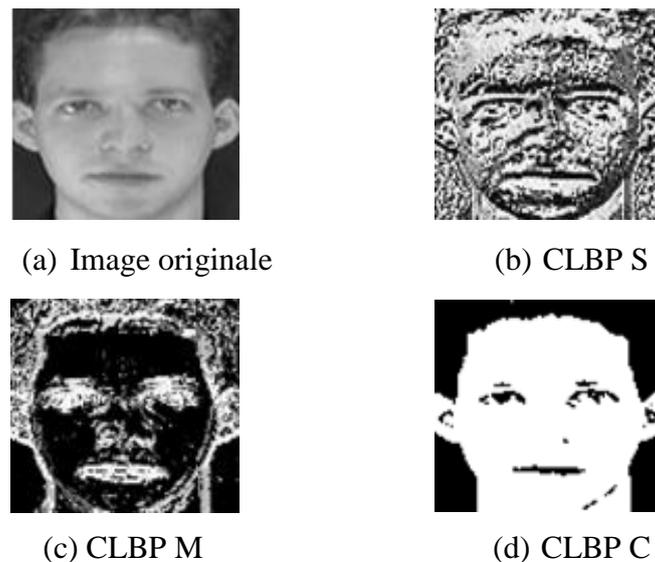
L'histogramme global est la concaténation de tous ces histogrammes. Ensuite, une étape de sélection des caractéristiques est opérée pour sélectionner les bins les plus discriminatives (Figure 3.30). Les résultats expérimentaux, obtenus par l'application de cette nouvelle technique sur les deux bases de données précédentes, ont révélé la supériorité de notre méthode proposée par rapport aux autres techniques.



**Figure 3.30 :** Principe de La méthode CLBP multi résolution proposée.

### 3.5.5 L'opérateur CLBP

Comme nous l'avons déjà mentionné, à la différence de l'opérateur LBP classique qui utilise uniquement le signe, l'opérateur CLBP utilise le signe et l'amplitude (Magnitude) pour caractériser une image donnée (Figure 3.31).



**Figure 3.31:** Présentation d'une image faciale par l'opérateur CLBP:

(a) Image original, (b) Image de signe, (c) Image d'amplitude, (d) image binaire.

Pour un pixel donn e  $fc$  et sont voisinage  $fp$  l'opérateur CLBP calcule la diff rence entre  $fc$  et  $fp$  :

$$LBP_{P,R}^{U2} = \begin{cases} \sum_{p=0}^{P-1} S(g_p - g_c) & \text{if } U(LBP_{P,R}) \ll 2 \\ P + 1 & \text{otherwise} \end{cases} \quad (3.32)$$

$D_p$  peut  tre divis  en deux partie comme :

$$D_p = S_p \times M_p \quad (3.33)$$

O  :

$$M_p = |D_p|, \quad S_p = \begin{cases} 1 & D_p \geq 0 \\ -1 & < 0 \end{cases} \quad (3.34)$$

$S_p$  ici est la composante signe et  $M_p$  la composante amplitude. Le descripteur CLBP est d fini   partir de ces composantes :

$$CLBP_{M_{P,R}} = \sum_{p=0}^{p-1} t(M_p - c) 2^p \quad (3.35)$$

$$t(x - c) = \begin{cases} 1, & x \geq c \\ 0, & x < c \end{cases} \quad (3.36)$$

O   $c$  repr sente la valeur moyenne de  $M_p$  de l'image enti re.

$$CLBP_{S_{P,R}} = \sum_{p=0}^{p-1} t(S_p, 0) 2^p \quad (3.37)$$

$$CLBP_{C_{P,R}} = t(f_c - c_g) \quad (3.38)$$

La dimension de l'histogramme correspondant au  $CLBP_{M_{P,R}}$  est  $P+2$ , la dimension de l'histogramme correspondant au  $CLBP_{S_{P,R}}$  est  galement  $P+2$  alors que la dimension de l'histogramme correspondant au  $CLBP_{C_{P,R}}$  est  gale 2 bins.

Pour le calcul de ces histogrammes, une grille multi-r solution est utilis e. Elle partage chaque image du visage en sous r gions   diff rentes  chelles. C'est une repr sentation en pyramide. Pour chaque sous-r gion de la repr sentation en pyramide, nous calculons les histogrammes  $CLBP_S$  et  $CLBP_M$  s par ment. L'histogramme global est calcul  par concat nation de ces deux histogrammes. Par la suite, une s lection des caract ristiques, bas e sur Fisher-Score [27], est appliqu e afin de r duire la dimension du vecteur de caract ristiques et  liminer les caract ristiques redondantes et bruyantes.

Fisher Score est une méthode de sélection de caractéristiques supervisée. Elle sélectionne les caractéristiques avec la meilleure capacité de discrimination. Si on suppose que  $l_i$  représente le nombre d'échantillons dans chaque classe  $i$ , et  $\mu_r^i$  et  $\sigma_r^i$  représente la moyenne et la variance pour la classe  $i, i = 1, \dots, c$  correspondant à la  $r^{\text{th}}$  caractéristiques. Fisher Score pour la  $r^{\text{th}}$  caractéristiques  $F_r$  et calculé à partir de l'équation suivant:

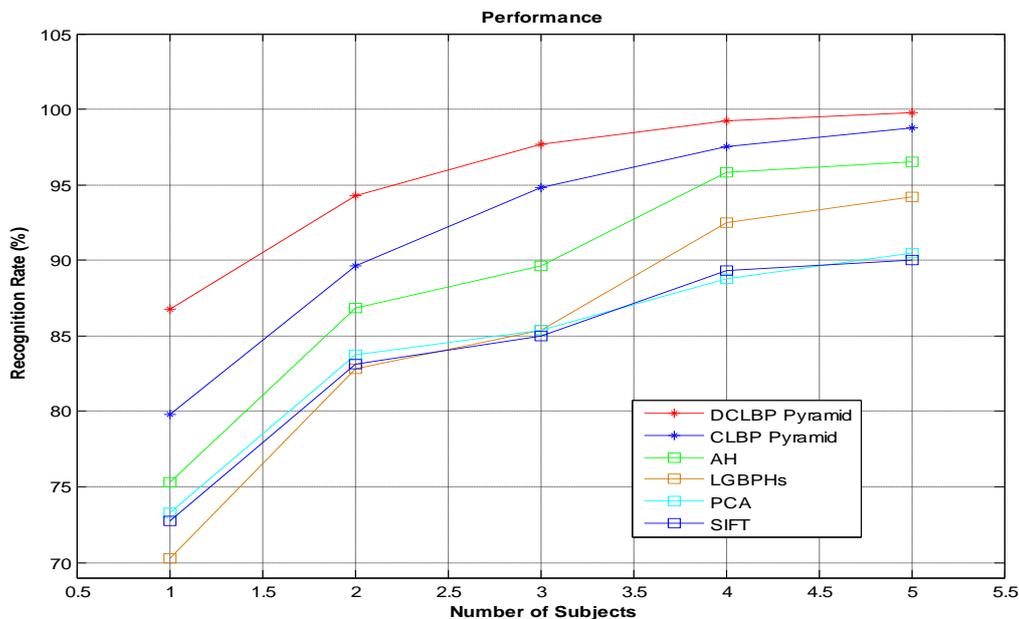
$$F_r = \frac{\sum_{i=1}^c l_i (\mu_r^i - \mu_r)^2}{\sum_{i=1}^c l_i (\sigma_r^i)^2} \quad (3.38)$$

Ensuite, les caractéristiques correspondant au Fisher Score supérieur à un certain seuil fixé a priori sont sélectionnées.

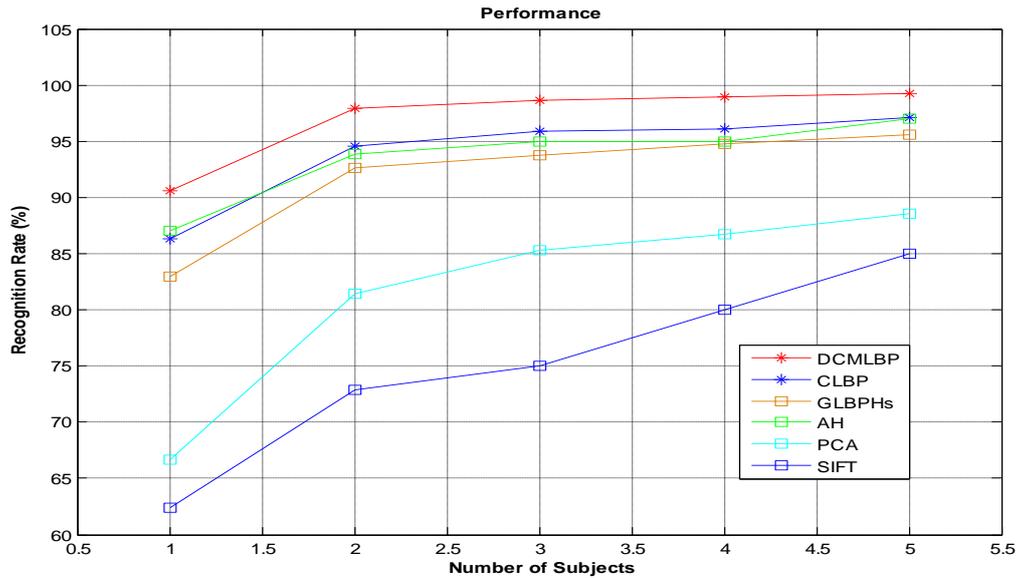
### 3.5.6 Evaluation

Les expériences sont réalisées sur les bases de données ORL et Yale. Ces ensembles de données couvrent diverses conditions telles que l'éclairage, les accessoires, la pose, l'expression et la variation du vieillissement.

Dans la première expérience, nous sélectionnons un nombre  $\mathbf{d}$  d'images de chaque sujet pour l'entraînement. Les images restantes sont utilisées pour le teste. Pour comparer notre méthode avec d'autres méthodes, nous varions  $\mathbf{d}$  d'une image à cinq images par sujet. Les résultats de la classification sont illustrés dans les figures 3.32 et 3.33 pour la base **ORL** et la base **Yale** respectivement.



**Figure 3.32:** Taux de reconnaissance sur la base de données **ORL**.



**Figure 3.33:** Taux de reconnaissance sur la base de données **YAL**.

On peut voir à partir de ces figures que la CLBP pyramide proposée sans sélection de caractéristiques atteint le meilleur taux de reconnaissance par rapport aux méthodes de l'état de l'art.

En considérant la sélection de caractéristiques (c'est-à-dire Fisher Score), la CLBP pyramide Discriminative sélectionnée renforce la performance de la CLBP pyramide sans réduire la capacité de représentation du descripteur. Pour la base de données **ORL**, la **DCLBP** réduit le vecteur des caractéristiques de 2478 à 1169 bins; le seuil utilisé pour sélectionner les caractéristiques discriminantes est fixé à 0,8. Pour la base de données **Yale**, le seuil utilisé est de 1,4 et le vecteur des caractéristiques est ramené de 2478 à 342 bins.

Dans la deuxième expérience, la précision rapportée est calculée en moyenne sur 10 séries, avec 5 images d'apprentissage pour chaque classe et le reste est utilisé comme image de test. Les résultats de classification sont présentés dans le tableau 3.3, où l'on peut voir que la CLBP pyramide a donné la plus grande précision pour les deux bases de données, c'est aussi dire que le taux de reconnaissance est plus stable par rapport aux autres méthodes de l'état de l'art. Une autre remarque intéressante à faire ressortir du tableau 3.3 est que l'utilisation de la pyramide DCLBP augmente la performance de la classification de 0,95% et 2,11% pour les bases de données **ORL** et **YALE**, respectivement. En outre, la classification est plus rapide dans le cas du CLBP pyramide.

**Tableau 3.3:** Taux de reconnaissance de la m thode propos e et algorithmes de comparaison

M�thode	Mean Recognition rate	
	ORL database	Yale database
PCA [9]	94.2±2.39	88.55±4.2
SIFT [24]	93.70±1.23	85.62±3.34
AH [6]	97.85±1.16	95.55±4.71
LGBPHS [10]	95.75±1.36	97.44±2.67
HMLBP [29]	<b>98.8±0.63</b>	<b>97.11±2.23</b>
DHMLBP [29]	<b>99.75±0.35</b>	<b>99.22±0.92</b>

### 3.6 Conclusion :

Nous avons pr sent  dans ce chapitre, un nouveau syst me de reconnaissance faciale bas e sur une nouvelle version de **LBP** classique et la m thode SRC pour la classification. Le syst me propos  est plurimodale (Visage, empreinte digitale ...), et ind pendant du syst me de classification utilis . Nous avons montr  la performance de notre syst me de classification dans diff rentes situations telles que: les variations des expressions faciale, diff rentes conditions d' clairage, la rotation.... Les r sultats obtenus sont tr s satisfaisants. Pour la validation de notre approche, on l'a compar  avec d'autres techniques r centes. Chaque m thode a des avantages et des inconv nients en terme de taux de classification, de rapidit , de simplicit  des calculs et la taille du vecteur des caract ristiques qui est diff rent d'une m thode   l'autre. Le meilleur taux de classification est retenu par notre m thode pour diff rentes bases de donn es.

On   utilis   galement une autre variante de l'LBP appel  CLBP, qui est propos  dans [27], renforc e par une pr sentation pyramidale (multi- chelle) pour obtenir plus d'information sur la r partition des diff rentes structures dans le visage. Le Fisher Score est utilis  pour la s lection des caract ristiques ce qui a permet de r duire consid rablement la taille du vecteur des caract ristiques et d'am liorer le taux de classification.

En r sum  et d'apr s les r sultats obtenus, les caract ristiques locales sont plus informatives que les caract ristiques globales sp cialement dans les conditions incontr lables (variation d'illumination, rotation....).

**Bibliographie :**

- [1] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.
- [2] Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Electrical Engineers-Part III: Radio and Communication Engineering, Journal of the Institution of*, 93(26), 429-441.
- [3] Serrano,  ., de Diego, I. M., Conde, C., & Cabello, E. (2010). Recent advances in face biometrics with Gabor wavelets: A review. *Pattern Recognition Letters*, 31(5), 372-381.
- [4] Huo, Q., Ge, Y., & Feng, Z. D. (2001). High performance Chinese OCR based on Gabor features, discriminative feature extraction and model training. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on* (Vol. 3, pp. 1517-1520). IEEE.
- [5] Daugman, J. G. (1993). High confidence visual recognition of persons by a test of statistical independence. *IEEE transactions on pattern analysis and machine intelligence*, 15(11), 1148-1161.
- [6] Ahonen, T., Hadid, A., & Pietik inen, M. (2004, May). Face recognition with local binary patterns. In *European conference on computer vision* (pp. 469-481). Springer Berlin Heidelberg.
- [7] Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7), 971-987.
- [8] Pietik inen, M., Hadid, A., Zhao, G., & Ahonen, T. (2011). *Computer vision using local binary patterns* (Vol. 40). Springer Science & Business Media.
- [9] Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1), 71-86.
- [10] Zhang, W., Shan, S., Gao, W., Chen, X., & Zhang, H. (2005, October). Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* (Vol. 1, pp. 786-791). IEEE.
- [11] Tan, X., & Triggs, B. (2010). Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE transactions on image processing*, 19(6), 1635-1650.
- [12] Liao, S., Zhu, X., Lei, Z., Zhang, L., & Li, S. Z. (2007, August). Learning multi-scale block local binary patterns for face recognition. In *International Conference on Biometrics* (pp. 828-837). Springer Berlin Heidelberg.
- [13] Jin, H., Liu, Q., Lu, H., & Tong, X. (2004, December). Face detection using improved LBP under bayesian framework. In *Image and Graphics (ICIG'04), Third International Conference on* (pp. 306-309). IEEE.
- [14] Heikkil , M., Pietik inen, M., & Schmid, C. (2006). Description of interest regions with center-symmetric local binary patterns. In *Computer vision, graphics and image processing* (pp. 58-69). Springer Berlin Heidelberg.

- [15] Maturana, D., Mery, D., & Soto, A. (2011, March). Learning discriminative local binary patterns for face recognition. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* (pp. 470-475). IEEE.
- [16] Vu, N. S., & Caplier, A. (2010, September). Face recognition with patterns of oriented edge magnitudes. In *European conference on computer vision* (pp. 313-326). Springer Berlin Heidelberg.
- [17] Yang, M. H., Kriegman, D. J., & Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Transactions on pattern analysis and machine intelligence*, 24(1), 34-58.
- [18] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-511). IEEE.
- [19] Adini, Y., Moses, Y., & Ullman, S. (1997). Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7), 721-732.
- [20] Mawloud, G., & Djame, M. (2014). Modified local binary pattern for human face recognition based on sparse representation. *International Journal of Computers and Applications*, 36(2), 64-71.
- [21] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2), 210-227.
- [22] The ORL Database of Faces, AT & T (Olivetti) Research Laboratories, Cambridge, U.K., [Online]. Available: <http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/ attfaces.577 tar.Z>
- [23] Georgia Tech Face Database". Available at: <http://www.face-rec.org/databases>.
- [24] Nefian, A. V., Khosravi, M., & Hayes III, M. H. (1997, January). Real-Time detection of human faces in uncontrolled environments. In *Electronic Imaging'97* (pp. 211-219). International Society for Optics and Photonics.
- [25] Luo, J., Ma, Y., Takikawa, E., Lao, S., Kawade, M., & Lu, B. L. (2007, April). Person-specific SIFT features for face recognition. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07* (Vol. 2, pp. II-593). IEEE.
- [26] Heisele, B., Ho, P., Wu, J., & Poggio, T. (2003). Face recognition: component-based versus global approaches. *Computer vision and image understanding*, 91(1), 6-21.
- [27] Guo, Z., Zhang, L., & Zhang, D. (2010). A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing*, 19(6), 1657-1663.
- [28] Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.
- [29] Mawloud, G., & Djamel, M. On the Use of Complete Local Binary Patterns for Face Recognition. International Conference on Information Processing and Electrical Engineering ICIPEE14 November 24-25, 2014 University Tebessa, Algeria

# **Chapitre IV**

## **Reconnaissance de l'oreille humaine**

### **Le savoir faire**

#### 4.1 Introduction :

Toute caractéristique physiologique ou comportementale humaine peut être utilisée comme indicateur pour faire une identification personnelle, pourvu qu'elle satisfasse aux exigences suivantes [1]: universalité, unicité, permanence, recevabilité et acceptabilité.

Le tableau 4.1 montre quelques exemples de performances de reconnaissance pour différentes approches biométriques.

Ces Techniques ont été sélectionnées dans la plus récente édition de la Conférence **International Joint Conference on Biometrics (IJCB)**, qui est l'une des conférences les plus importantes dans ce domaine.

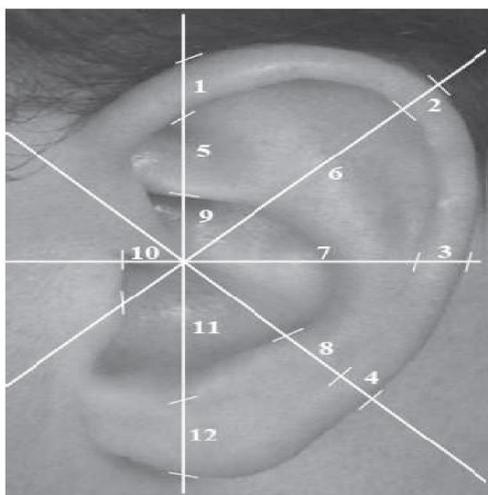
**Tableau 4.1:** Exemples de taux de reconnaissance pour différents systèmes biométriques.

Biometrics	Référence	Taux de Reconnaissance
Ear	Hurley et al. [2]	99.2%
	Cadavid et al. [3]	97.6%
face	Juefei-Xu et al. [4]	100%
	Smeets et al. [5]	98.6%
Iris	Daugman [6]	100%
	Ko et al. [7]	98.2%
Fingerprint	Shen and Kot [8]	97.6%
	Win and Sein [9]	99%

La plupart des recherches actuelles en biométrie ont porté sur le visage, les empreintes digitales, la démarche, l'iris et la signature ou la géométrie de la main. Cependant il ya eu ces dernières années un intérêt croissant pour la biométrie auriculaire. En effet, l'oreille de part sa forme externe, ses lobes et sa structure osseuse représente un candidat intéressant et promis de bons résultats.

En revendiquant l'oreille comme une classe de biométrie, nous devons montrer qu'elle est viable comme une classe de biométrie c'est-à-dire, qu'elle répond aux exigence d'universalité, d'unicité et de permanence cités ci dessus. Après des décennies de recherche de mesures anthropométriques de photographies d'oreille de milliers de personnes, Iannarelli a constaté qu'on ne peut en aucun cas trouver deux oreilles semblables. Cette constatation est même vraie dans le cas de jumeaux identiques et fraternels, triplets et quadruplets [10]. Il a également constaté que la structure de l'oreille ne change pas radicalement avec le temps.

Iannarelli [10] a développé le système célèbre de classification d'oreille basé sur des mesures anthropométriques manuelles des photographies d'oreilles de milliers de personnes. Le «**système Iannarelli**» est basé sur 12 mesures illustrées à la figure 4.1.



**Figure. 4.1:** Les douze mesures du «système Iannarelli (figure extraite de [10, chap. 13].

Les emplacements indiqués sont mesurés à partir de photographies spécialement alignées et normalisées de l'oreille droite. Ce système est utilisé depuis plus de 40 ans, bien que sa fiabilité ait récemment été contestée devant les tribunaux.

L'oreille externe est un trait biométrique émergent qui attire l'attention de la communauté de recherche depuis plus d'une décennie. La structure unique de l'oreillette est connue depuis bien longtemps parmi les médecins légistes et a été utilisée pour l'identification des suspects dans de nombreux cas. Les oreilles humaines ont gagné beaucoup d'attention en biométrie en raison de :

- La robustesse de la forme de l'oreille: La forme de l'oreille ne change pas en raison de l'émotion et reste constante pendant la vie d'une personne [1] ;
- Ont une structure riche qui semble être cohérente tout au long de la vie d'une personne à partir de quelques mois après la naissance ;
- Ne sont pas affectées par des changements de l'expression faciale ;
- L'image des oreilles peut être acquise sans la participation du sujet ;
- La répartition des couleurs est plus uniforme pour l'oreille que pour le visage humain, l'iris ou la rétine.

Ces propriétés font des oreilles un identifiant biométrique très attrayant. Ainsi, ces dernières années, les chercheurs ont commencé à considérer le problème de développement de systèmes automatiques pour la reconnaissance de l'oreille et il va sans dire que les domaines d'applications sont si vastes que variés.

Les images d'oreille peuvent être capturées à distance. Cependant il existe un grand obstacle potentiel qui est l'occlusion par les cheveux et les boucles d'oreilles, ce qui est presque certain de se produire dans les environnements incontrôlés. Bien que les oreilles soient souvent soumises à des occlusions, notre recherche montrera que cela peut être surmonté dans une certaine mesure par les méthodes présentées dans cette étude.

#### **4.2 De qui s'agit-il ?**

Notre but est de préconiser l'utilisation des oreilles comme technique biométrique non invasive pour l'identification humaine. Nous voulons donc développer un système de reconnaissance basé sur l'oreille, de quoi a-t-on besoin ?

D'abord, nous aurons besoin de localiser l'oreille dans une image de profil. Il va de soi que cette localisation devrait être automatique et exécutée avec un algorithme robuste à la présence du bruit. Ensuite, nous aurons à choisir la configuration optimale pour la représentation de l'image. Il s'agit de l'utilisation de caractéristiques locales récentes qui ont prouvé leur capacité de discrimination, par rapport aux caractéristiques globales, dans plusieurs domaines. En fin nous aurons à choisir le classificateur optimal qui garantit de bien placer chaque image de test dans sa classe appropriée et en temps réel.

Notre travail est basé sur une nouvelle approche de segmentation spatiale des images auriculaires 2D: L'image de l'oreille est segmentée en plusieurs segments non chevauchés. L'identification est opérée sur chacun de ces segments individuellement et la décision finale est prise par un système de fusion de classificateurs pondérés. Nous présentons différentes techniques de segmentation visant à améliorer le taux de reconnaissance tout en prenant en considération les occlusions.

#### **4.3 Travaux précédents et problématique de la reconnaissance automatique de l'oreille humaine**

La possibilité d'utiliser l'apparence de l'oreille comme moyen d'identification personnelle a été reconnue et préconisée par le criminologue français Alphonse Bertillon dès 1890, dans le cadre de sa nouvelle méthode scientifique d'identification criminelle [11].

Alfred Iannarelli [10] a effectué deux études en 1989 pour fournir une preuve de soutien pour la reconnaissance par l'oreille. Dans la première étude, Iannarelli a comparé plus de 10 000 oreilles tirées d'un échantillon sélectionné au hasard, et dans la deuxième étude il a examiné des jumeaux fraternels et identiques. Ces deux études étaient basées sur la mesure des distances entre des points spécifiques de l'oreille. Iannarelli a développé une méthode où 12 mesures sont utilisées comme caractéristiques pour distinguer les individus (comme illustré à

la figure 3.1). Les résultats de ces études ont prouvé que l'oreille contient des caractéristiques uniques. Même les jumeaux identiques avaient des oreilles semblables mais pas identiques.

Il existe beaucoup d'autres études sur la biométrie auriculaire 2D et 3D décrivant différentes techniques de détection et de reconnaissance des systèmes biométriques auriculaires [12-15]. Nous examinerons d'abord les techniques de recherche de l'oreille: détection de l'oreille. On distinguera alors les approches qui considèrent l'oreille comme une structure plane, les approches 2D et celles qui considèrent l'information.

### 4.3.1 Détection de l'oreille

La détection de l'oreille est l'étape la plus importante dans un système de reconnaissance de l'oreille et la qualité de la détection affectera directement les performances de l'ensemble du système de reconnaissance.

Cette section résume l'état de l'art dans la détection automatique de l'oreille dans les images 2D et 3D, respectivement. Fondamentalement, toutes les approches de détection auriculaire s'appuient sur les propriétés mutuelles de la morphologie des oreilles, comme l'occurrence de certains contours caractéristiques ou des modèles de fréquence. Le tableau 4.2 donne une petite introduction sur les différentes méthodes de détection de l'oreille décrites ci-dessous. La partie supérieure de la table contient les algorithmes pour la localisation de l'oreille 3D, tandis que la partie inférieure liste les algorithmes conçus pour la détection de l'oreille dans les images 2D.

Chen et Bhanu proposent trois approches différentes pour la détection des oreilles. Dans leur approche [16], on forme un classificateur qui reconnaît une distribution spécifique des indices de forme, caractéristique de la surface de l'oreille. Cependant, cette approche ne fonctionne que sur des images de profil et est sensible à tout type de variation de rotation, d'échelle et de pose. Dans leur travail ultérieur, pour la détection de l'oreille, ils ont détecté des régions d'image avec une technique qu'ils ont appelée **step edge magnitude** [17]. Ensuite, un modèle qui contient la forme typique de l'hélice externe et de l'anti-hélice est adapté et sélectionné. Dans [18], Chen et Bhanu ont rétréci le nombre de candidats possibles d'oreille en détectant la région de peau d'abord avant que l'adaptation du modèle d'hélice soit appliqué. Ils ont fusionné les informations de couleur et de forme. Le taux de détection est porté à 99,3% sur la base de données **UCR** et 87,71% sur la collection **UND F** et un sous-ensemble de la collection **UND G**.

Un autre exemple de détection d'oreille, utilisant des lignes de contour de l'oreille, est décrit par Attrachi et al. [38]. Ils localisent le contour extérieur de l'oreille en recherchant le contour

le plus long connecté dans l'image. En sélectionnant les points haut, bas et gauche de la limite détectée, ils forment un triangle avec les points sélectionnés. Ansari et al. utilisent également un détecteur de contour dans la première étape de leur approche de localisation de l'oreille [39]. Les contours sont séparés en deux catégories, à savoir convexe et concave. Les contours convexes sont choisis comme candidats pour représenter le contour extérieur. Enfin, l'algorithme connecte les segments de contour et sélectionne la figure entourant la zone la plus grande pour être le contour de l'oreille externe.

**Tableau 4.2:** Taux de détection pour différents systèmes de biométrie auriculaire.

Référence	Méthode	Nombre de classes de la base de données	Type	Performance
Chen & Bhanu [16]	Shape model and ICP	700	3D	87.7%
Chen & Bhanu [17]	Helix Shape Model	213	3D	92.3%
Zhou et al. [19]	Histograms of Categorized Shapes	942	3D	100%
Prakash & Gupta [20]	connectivity graph	1604	3D	99.38%
Abaza et al. [21]	Cascaded adaboost	940	2D	88.72%
Ansari and Gupta [22]	Edge detection and curvature estimation	700	2D	93.34%
Alvarez et al. [23]	Ovoid model	NA	2D	NA
Arbab-Zavar & Nixon [24]	Hough Transform	942	2D	91%
Arbab-Zavar & Nixon [25]	Log-Gabor filters and wavelet transform	252	2D	88.4%
Attarchi et al. [26]	Edge detection and line tracing	308	2D	98.05%
Chen & Bhanu [27]	Template Matching with Shape index histograms	60	2D	91.5%
Cummings et al. [28]	Ray transform	252	2D	98.4%
Islam et al. [29]	Adaboost	942	2D	99.89%
Jeges & Mate [30]	Edge orientation pattern	330	2D	100%
Kumar et al. [31]	Edge clustering and active contours	700	2D	94.29%
Liu & Liu [32]	Adaboost and skin color filtering	50	2D	96%
Prakash & Gupta [33]	Skin color and graph matching	1780	2D	96.63%
Shih et al. [34]	Arc-Masking and AdaBoost	376	2D	100%
Yan & Bowyer [35]	Concha Detection and active contours	415	2D	97.6%
Yuan & Mu [36]	CAMSHIFT and a contour fitting	Video	2D	NA

Yuan et Mu ont développé une méthode de suivi d'oreille en temps réel, dans des séquences vidéo, en appliquant l'approche de changement de vitesse adaptatif continu (**Continuously Adaptive Mean Shift : CAMSHIFT**) aux séquences vidéo [40]. L'algorithme CAMSHIFT est fréquemment utilisé dans les applications de suivi du visage. Pour une segmentation précise de l'oreille, la méthode d'ajustement du contour, basée sur des modèles de forme active modifiée, proposés par Alvarez et al. est appliquée [41]. Yuan et Mu rapportent un taux de détection de 100%, cependant la base de données de test ne se composait que de deux sujets. Néanmoins, leur approche semble très prometteuse pour les applications de surveillance mais doit être évaluée plus en détail dans des scénarios de test plus réalistes.

Burge et Burger [42], ont utilisé une approche relativement simple basée sur l'application de la technique du contour déformable sur une représentation pyramidale gaussienne du gradient d'image. Une image en échelle de gris a été obtenue à partir de la tête du sujet en profil. Puis l'emplacement de l'oreille dans l'image a été localisé par des contours déformables. L'algorithme du contour déformable est basé sur la minimisation de l'énergie d'un modèle du contour actif généralisé.

Abdel-Mottaleb and Zhou [43] ont détecté automatiquement l'oreille à partir de l'image de profil du visage. La détection de l'oreille a été réalisée en trois étapes: détection de la région de la peau, détection des arêtes et adaptation des modèles. La détection de la région de la peau a utilisé un filtre de peau qui s'est appuyé sur l'information de couleur et de texture. La région de peau de sortie a été étendue pour englober l'oreille entière. Ceci a été réalisé en appliquant une opération de dilatation morphologique à la sortie du filtre de peau. Le détecteur des arêtes (bord) de Canny a été appliqué. Le résultat a été filtré, pour enlever les bords courts et isolés qui sont apparus en raison du bruit. Enfin, la région de l'oreille a été détectée en utilisant une adaptation de modèle. La distance de Hausdorff a été utilisée pour rechercher un modèle d'oreille dans la carte de bordure de la région de peau. Cette technique faisait partie d'un système de reconnaissance de l'oreille qui reconnaissait l'oreille après sa détection.

Yuan et Mu [44] ont combiné le modèle de couleur de peau et l'information de contour d'intensité pour localiser et suivre l'oreille humaine dans une séquence vidéo. Une méthode de suivi des oreilles en deux étapes a été proposée. Dans la première étape, l'approche CAMSHIFT est utilisée pour suivre le profil du visage dans les séquences de trame vidéo. Dans la deuxième étape, l'algorithme d'ajustement, basé sur le contour, a été utilisé pour la détection précise de l'oreille. Le but de suivi du profil était de repérer rapidement la région de couleur de la peau principale comme région d'intérêt qui contenait l'oreille.

Ansari et Gupta [45] ont présenté une méthode de détection de l'oreille basée sur les bords des hélices de l'oreille externe. Ils ont utilisé le parallélisme entre les courbes d'hélice externe de l'oreille pour détecter l'oreille. Les bords de l'oreille ont été obtenus en utilisant le détecteur de Canny.

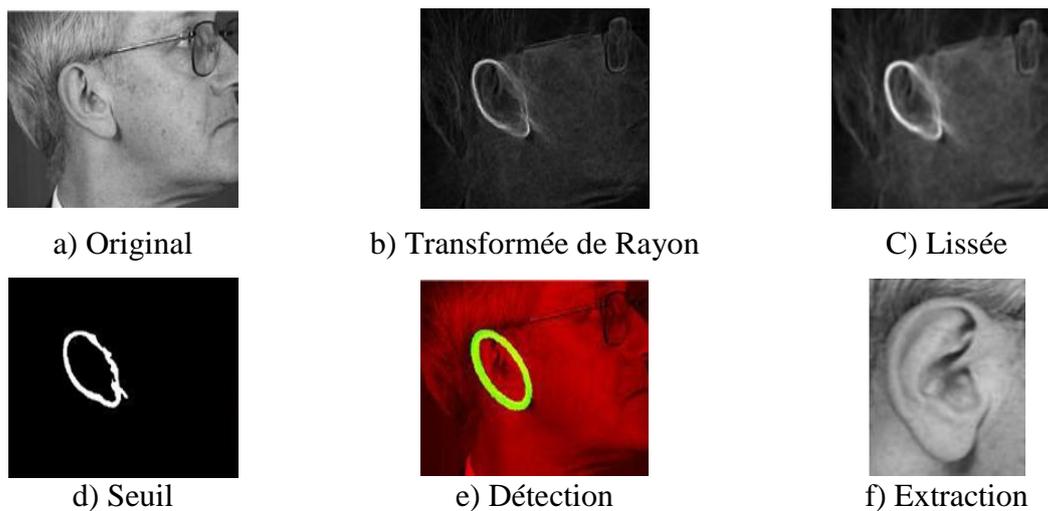
Prakash et al. [45, 46] ont présenté une technique efficace de transformée de distance et de modèle, pour la localisation automatique de l'oreille à partir d'une image latérale de face. La technique a d'abord procédé par segmenter des régions en deux groupes: peau et non-peau de l'image, sur la base d'un modèle de couleur. Ce modèle a transformé une image en couleur en une image en échelle de gris de sorte que la valeur de gris à chaque pixel montrait la probabilité du pixel appartenant à la peau. Ensuite, l'image de l'échelle de gris a été transformée en une image binaire avec un seuil approprié. La segmentation de la région de la peau a été ensuite obtenue à l'aide du détecteur de Canny. Enfin, une similarité entre deux séries de moments de Zernike, une pour le modèle et l'autre pour l'oreille détectée, a été estimée pour déterminer si l'oreille détectée était réellement une oreille ou non. La technique proposée a été testée sur la base de données IIT Kanpur contenant des images de face latérale de 150 individus et a atteint un taux d'exactitude de 95,2%.

Cummings et al. [47] ont présenté une nouvelle technique d'enrôlement de l'oreille utilisant la transformée du rayon d'image en analogie aux rayons lumineux. La transformée est capable de mettre en évidence des structures tubulaires telles que l'hélice de l'oreille et, en exploitant la forme elliptique de l'hélice, peut être utilisée comme une méthode pour la biométrie de l'oreille. L'étape initiale était d'appliquer la transformée du rayon d'image à l'image d'oreille d'origine (Figure 4.2-a). Ceci a produit une image dans laquelle l'hélice de l'oreille a été soulignée (figure 4.2-b) en plus d'autres dispositifs tubulaires tels que des lunettes. Ensuite, un lissage gaussien a été appliqué au résultat (Figure 3.2-c) pour réduire le bruit. L'image a alors été seuillée (Figure 4.2-d) pour produire une image avec une hélice forte.

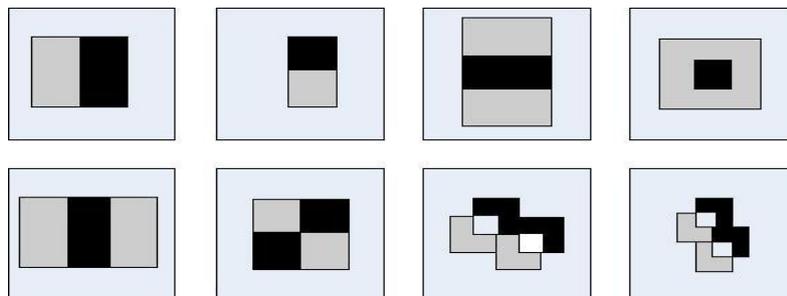
L'égalisation de l'histogramme a été utilisée pour permettre l'utilisation d'un seuil uniforme pour toutes les images. Le modèle simple a ensuite été utilisé avec un modèle elliptique à travers une gamme de rotations et d'échelles (Figure 4.2-e), puis la section correspondante a ensuite été normalisée et extraite (Figure 4.2-f). Ils ont obtenu un taux de 99,6% lors de l'utilisation de 252 images de la base de données XM2VTS.

Islam et al. [48] ont modifié l'approche en cascade AdaBoost pour détecter l'oreille à partir d'images de visage de profil 2D. L'approche a été divisée en trois étapes: prétraitement, entraînement et détection. Dans l'étape de prétraitement, les images d'entrée ont été collectées à partir de différentes sources avec des valeurs de taille et d'intensité variables. Par

conséquent, toutes les images d'entrée ont été normalisées à l'échelle pour une taille d'entrée choisie. Les valeurs d'intensité des images ont également été normalisées pour minimiser l'effet de l'éclairage. Au cours de l'étape d'entraînement, des classificateurs faibles ont été construits en fonction de certaines caractéristiques rectangulaires (figure 4.3). Ces caractéristiques étaient la dérivée de l'ondelette de Haar, qui est un ensemble naturel de fonctions de base calculé à partir de la différence d'intensité dans les régions voisines. Huit types de caractéristiques ont été proposés pour la construction des classificateurs faibles.



**Figure 4.2** : Les différentes étapes dans l'enrôlement [47]

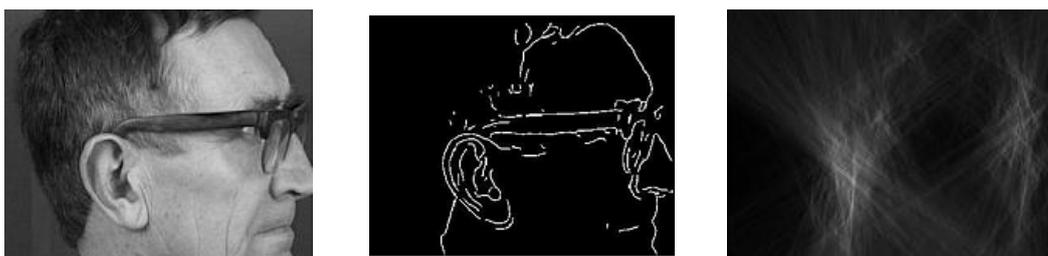


**Figure 4.3**: Les caractéristiques utilisées pour la formation de l'AdaBoost.

L'ensemble d'entraînement positif a été construit avec 5000 images d'oreilles, recueillies à partir du profil des images de visages de diverses bases de données et couvrant un large éventail de races, sexes, apparences, orientations et illuminations. Cet ensemble contenait des images des bases de données **UND**, **MID**, **XM2VTS**, **USTB**, **MIT-CBL** et **UMIST**. Il contenait également environ 3000 images synthétisées par rotation de **-15** à **+15** degrés de certaines images de bases de données **USTB**, **UND** et **XM2VTS**. L'ensemble d'entraînement négatif comprenait 10000 images d'images non-auriculaires. Ces images ont été

principalement recueillies à partir d'images de visage de profil, à l'exclusion de la zone de l'oreille. La formation avec une grande variété d'images d'oreille de diverses bases a rendu le détecteur robuste à la rotation légère, l'occlusion et aussi à la dégradation de la qualité d'image dans une mesure significative. La performance de détection du détecteur en cascade a été testée contre 203 images de la base de données **UND**; Toutes les oreilles ont été extraites avec précision par le détecteur proposé avec un très faible taux de faux positifs. En présence d'occlusions partielles impliquant des cheveux et des boucles d'oreilles, sur 104 images occluses sélectionnées de la base de données **XM2VTS**, 54 images ont été correctement détectées.

Arbab-Zavar et Nixon [50] ont proposé une technique de détection de l'oreille basée sur la localisation de la forme elliptique de celle-ci en se servant de la Transformée de Hough (HT). La détection des têtes était la première étape de l'algorithme (Figure 4.4-a). L'éclairage diffus ou la détection de la peau en couleur a été utilisé pour éliminer la tête de l'arrière-plan en s'appuyant sur la base de données qu'ils utilisaient. Le détecteur de Canny est par la suite appliqué pour détecter le contour des oreilles (Figure 4.4-b). La HT génère un certain nombre de pics correspondant à la position des centres d'ellipses (Figure 4.4-c). L'oreille apparaissant habituellement sous la forme d'une ellipse verticale, avec un axe vertical environ deux fois plus grand que l'axe horizontal, Les pics correspondant aux ellipses non verticales ont été alors éliminés. Ensuite, les informations de direction du bord ont été utilisées pour déterminer si une paire de points peut éventuellement faire partie d'une ellipse.



**Figure 4.4:** Détection de l'oreille en utilisant la transformée de Hough

L'algorithme d'enrôlement proposé a été testé sur deux bases de données. La première était la base de données XM2VTS qui comprenait 252 images de 63 individus. La deuxième était la base de données biométriques de l'oreille UND qui comprenait 942 images 2D de 302 sujets. Le processus d'entraînement a détecté avec succès la région de l'oreille dans toutes les 252 images de profil de la base XM2VTS. 83% seulement des oreilles ont été détecté avec succès

dans le cas d'occlusion de 40% du haut qui a été étudié. Pour la base de données UND, l'algorithme offre un taux de réussite de 91%.

Une approche récente pour la détection d'oreille 2D est décrite dans [50]. Kumar et al. proposent à extraire des oreilles à partir d'images 2D en utilisant le contour actif. Ils évaluent leur approche sur une base de données qui se compose de 100 sujets avec 7 images par sujet. Un dispositif d'imagerie spécial a été utilisé pour collecter les données. Ce dispositif assure que la distance à la caméra est constante et que les conditions d'éclairage sont les mêmes pour toutes les images. Dans ce contexte, un taux de détection de 94,29% est rapporté

### 4.3.2 Reconnaissance de l'oreille

La partie précédente traitait les différentes techniques de détection et d'extraction de l'oreille. Disposant de cette empreinte, il est question dans cette partie de passer en revue les différentes techniques de reconnaissance de celle-ci.

Chaque système de reconnaissance d'oreille se compose d'une étape d'extraction de caractéristiques et d'une étape de classification. Dans cette enquête, nous divisons les approches de reconnaissance de l'oreille en quatre sous-classes différentes, à savoir les approches holistiques (globales), les approches locales, les approches hybrides et les approches statistiques. Dans les tableaux 3.2 et 3.3, toutes les approches de reconnaissance auriculaire 2D mentionnées dans ce chapitre sont résumées dans l'ordre chronologique.

#### 4.3.2.1 Descripteurs Holistiques

Une approche qui a gagné une certaine popularité est la transformée de champ des forces (Force Field) proposé par Hurley [51]. La méthode de transformation de champ des forces suppose que les pixels ont une attraction mutuelle proportionnelle à leurs intensités et inversement au carré de la distance entre eux, comme la loi universelle de gravitation de Newton. Le champ d'énergie associé prend la forme d'une surface lisse avec un certain nombre de pics joints par des crêtes (figure 4.5).

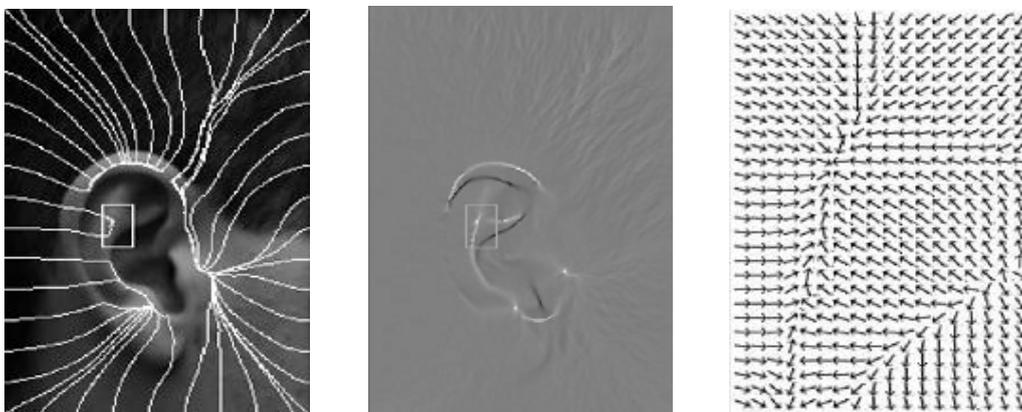


Figure 4.5: La méthode de FF

Utilisant cette méthode, Hurley et al. ont obtenu une performance de rank-1 de plus de 99% sur la base de données XM2VTS (252 images). Sur la base de ces résultats, Abdel-Mottaleb et Zhou utilisent une représentation 3D du champ des forces pour extraire des points situés sur le pic de celui-ci [43]. Comme le champ des forces converge au contour de l'oreille, les pics de la représentation 3D représentent essentiellement le contour de l'oreille. A souligner que la méthode du champ des forces est plus robuste au bruit que les autres détecteurs de contour, tels que Sobel ou Canny. En utilisant cette approche, Abdel-Mottaleb et Zhou ont obtenu une performance de rang-1 (rank-1) de 87,93% sur un ensemble de données comprenant 103 images d'oreille de 29 sujets.

Tableau 4.3: Taux de reconnaissance pour différents systèmes de biométrie auriculaire 2D.

Référence	Méthode	Nombre de classes de la base de données	Image	Performance
Dong and M[52]	Force Field Transform and NKFDA	29	711	75.3%
Guo and Xu [53]	Local Binary Pattern and CNN	77	308	93.3%
Nasseem et al. [54]	Sparse representation	32	192	96.88%
Wang et al. [55]	Haar Wavelets and Local Binary Patterns	79	395	92.41%
Xie and Mu [56]	Locally Linear Embedding	79	1501	80%
Yaqubi et al. [57]	HMAX and SVM	60	180	96.5%
Zhang and Mu [58]	Geometrical Features, ICA and PCA with SVM	77	308	92.21%
Badrinath and Gupta [59]	SIFT landmarks from ear model	106	1060	95.32%
Kisku et al. [60]	SIFT from different Color Segments	400	800	96.93%
Wang and Yuan [61]	Low-Order Moment Invariants	77	308	100%
Alaraj et al. [62]	PCA with MLFFNNs	17	85	96%
Bustard et al. [63]	SIFT Point Matches	63	252	96%
De Marisco et al. [64]	Partitioned Iterated Function System (PIFS)	114	228	61%
Gutierrez et al. [65]	MNN with Sugeno Measures and SCG	77	308	97%
Wang et al. [66]	Moment Invariants and BP Neural Network	NA	60	91.8%
Wang and Yuan [67]	Gabor Wavelets and GDA	77	308	99.1%
Prakash and Gupta [68]	SURF and NN classifier	300	2066	2.25%
Kumar et al. [69]	SIFT	100	700	95%

Wang and Yan [70]	Local Binary Pattern and Wavelet Transform	77	308	100%
Kumar and Wu [71]	Phase encoding with Log Gabor filters	221	753	95.93%
Abate et al. [72]	General Fourier Descriptor	70	210	88%
Sana et al. [73]	Haar Wavelets and Hamming Distance	600	1800	98.4%
Arbab-Zavar et al. [76]	SIFT points from ear model	63	252	91.5%
Liu et al. [80]	Edge-based features from different views	60	600	97.6%
Lu et al. [82]	Active Shape Model and PCA	56	560	93.3%
Yuan and Mu [83]	Full Space LDA with Outer Helix Feature Points	79	1502	86.76%
Victor et al. [84]	PCA	294	808	40%
Chang et al. [85]	PCA	114	464	72.7
Alaraj et al.. [86]	PCA with MLFFNNs	17	85	96%
Zhang and Mu [87]	Geometrical Features, ICA and PCA with SVM	77	308	92.21%
Xie and Mu [88]	Locally Linear Embedding	79	1501	80%
Yaqubi et al. [90]	HMAX and SVM	60	180	96.5%
Gutierrez et al. [91]	MNN with Sugeno Measures and SCG	77	308	97%
Nasseem et al. [92]	Sparse representation	32	192	96.88%

Dong et Mu [52] ajoutent l'invariance de pose aux contours extraits en utilisant la méthode de champ des forces. Ceci est réalisé avec une analyse discriminante qui a la propriété de représenter des relations non linéaires entre deux ensembles de données. Dong et Mu ont mené des expériences sur la base de données USTB IV. Avant l'extraction des caractéristiques, la région auriculaire a été découpée manuellement à partir des images avec une normalisation de pose. Pour des variations de pose de 30 degrés, ils rapportent un taux de classification rang-1 de 72,2% et pour une variation de pose de 45 degrés ils en obtiennent un taux de classification rang-1 égale au 48,1%.

Dans [71], Kumar et Wu présentent une approche de reconnaissance de l'oreille qui utilise l'information de phase des filtres Log-Gabor pour encoder la structure locale de l'oreille. Les informations de phase codées, sont stockées dans des images aux niveaux de gris normalisées. Les résultats obtenus sont meilleurs que celles obtenus par la méthode de champ des forces.

De plus, différentes combinaisons des filtres Log-Gabor ont été comparées entre elles. La performance de rang-1 pour ces approches se situe entre 92,06% et 95,93% sur une base de données qui contient 753 images de 221 sujets.

La structure riche de l'oreille externe donne des informations de texture spécifiques, qui peuvent être mesurées à l'aide de filtres de Gabor. Wang et Yuan [67], extraient des caractéristiques de fréquences locales en utilisant une banque de filtres de Gabor. Ils sélectionnent les caractéristiques les plus discriminatives en utilisant une analyse discriminante générale. Dans leurs expériences sur la base de données USTB II, ils ont comparé l'impact, sur la performance, de différents paramètres des filtres de Gabor. Différentes combinaisons d'orientation et d'échelles, dans les ensembles des filtres, sont comparées les unes avec les autres. Il a été constaté que ni le nombre d'échelles ni le nombre d'orientations n'a d'impact majeur sur la performance de rang-1. La performance totale de rang-1 de cette approche est de 99,1%. Dans une approche similaire, Arbab-Zavar et Nixon [26] ont mesuré la performance des filtres de Gabor sur la base de données XM2VTS où ils rapportent une performance de rang-1 de 91,5%.

Abate et al. [72] utilisent un descripteur de Fourier générique pour la représentation des fonctions invariantes de rotation et d'échelle. L'image est transformée en un système de coordonnées polaires puis ramenée dans l'espace des fréquences. Afin de s'assurer que le centroïde de coordonnées polaires du système est toujours à la même position, les images auriculaires doivent être alignées avant qu'elles puissent être transformées en système de coordonnées polaires. L'approche a été testée sur un ensemble de données qui contient 282 images d'oreilles au total. Les images ont été prises en deux jours différents et dans différentes poses. La performance de rang-1 de cette approche varie en fonction de l'angle de pose. Pour une variation de 0 degré, la performance est de 96%, mais pour des poses de 15 et de 30 degrés, elle tombe à 44% et 19% respectivement.

Sana et al. utilisent des coefficients sélectionnés d'ondelettes de Haar, extraits pendant la compression pour la représentation des caractéristiques [73]. Lors de l'application de la transformée en ondelettes à quatre niveaux à plusieurs reprises sur l'image de l'oreille, chaque itération stocke l'un des coefficients dérivés dans un vecteur caractéristique. La performance obtenue par leur algorithme est de 96% et a été obtenue à partir de la base de données IITK et de la base de données Saugor (350 sujets).

Un système d'extraction de caractéristiques appelé Partitioned Iterated Function System (PIFS) est proposé par De Marisco et al. [64]. PIFS mesure l'auto-similarité dans une image entre sous-régions similaires de celle-ci. Pour rendre leur système robuste à l'occlusion,

Marisco et al. divisent l'image de l'oreille en segments également grands. Si un segment est obstrué, les autres segments contiennent encore un ensemble suffisamment distinctif de caractéristiques. De Marisco et al. ont montré que leur approche est supérieure à d'autres méthodes d'extraction de caractéristiques en présence d'occlusion. Ils mènent plusieurs expériences afin d'évaluer la performance du système dans différents scénarios d'occlusion. Les tests sont effectués sur la base de données UND collection E et les 100 premiers sujets de la base de données FERET. Si l'occlusion se produit sur l'image de référence, une performance de rang-1 de 61% est obtenue. Sans occlusion, la performance est de 93%.

Les moments invariants sont une mesure statistique pour décrire les propriétés spécifiques d'une forme. Wang et al. [66] composent six vecteurs de caractéristiques différentes en utilisant sept moments invariants. Ils montrent également que chacun des moments est robuste aux changements d'échelle et de rotation. Les vecteurs caractéristiques sont utilisés comme entrée pour un réseau neuronal à rétro-propagation, entraîné pour classer les ensembles de caractéristiques extraites. Basé sur une base de données propriétaire de 60 images d'oreilles, ils rapportent une performance de rang-1 de 91.8%.

Dans [61] Wang et Yuan comparent différentes méthodes d'extraction de caractéristiques sur la base de données USTB I. Ils comparent la performance de rang-1 des descripteurs de Fourier, de la transformée de Gabor, des Moment Invariants et les caractéristiques statistiques. Ils arrivent à la conclusion que le taux de reconnaissance le plus élevé peut être atteint en utilisant les moments invariants et la transformée de Gabor. Pour ces deux méthodes, Wang et Yuan rapportent une performance rang-1 de 100%.

#### 4.3.2.2 Descripteurs locaux

Scale Invariant Feature Transform (SIFT), qui peut être traduite par Transformation à Caractéristique Invariante à l'échelle, est connue comme une méthode robuste d'extraction des points discriminants pour de petites variations de pose et de luminosité [74].

La SIFT contient une mesure d'orientation locale; elle peut également être utilisée pour estimer la rotation et la translation entre deux images d'oreille normalisées. Bustard et al. montrent que SIFT peut manipuler des variations de pose jusqu'à 20 degrés [75]. Dans les régions d'image très structurées, la densité et la redondance des repères SIFT sont très élevées. Par conséquent, les points SIFT doivent être filtrés avant que la comparaison réelle puisse commencer. Arbab-Zavar et al. [76] ainsi que Badrinath et Gupta [59] forment donc un modèle de repère de référence, qui ne contient qu'un petit nombre de points non redondants. Ce modèle est utilisé pour filtrer les points SIFT, qui ont été initialement détectés dans l'oreille de référence (ensemble d'entraînement). Ayant les repères filtrés, il est possible

d'assigner chacun des points du repère à son homologue correspondant. La figure 4.6 montre un exemple de repères SIFT extrait d'une image d'oreille. Ils ont été utilisés comme données d'entraînement pour le modèle repère de référence dans le travail d'Arbab-Zavar et al.. Parce que Arbab-Zavar et al. ont utilisé la même base de données XM2VTS que celle utilisée par Bustard et al., leurs résultats peuvent donc être directement comparés. La performance de rang-1 rapportée par Arbab-Zavar et al. est de 91,5%, alors que celle du même rang rapportée par l'approche plus récente de Bustard et Nixon, est de 96%.



**Figure 4.6:** Les points extraits de SIFT

Kisku et al. résolvent le problème de la répartition correcte des points en décomposant l'image de l'oreille en différents segments de couleur [60]. Les repères SIFT sont extraits de chaque segment séparément, ce qui réduit les chances d'attribuer des repères SIFT qui ne représentent pas les mêmes caractéristiques. En utilisant cette approche, Kisku et al. atteignent une performance de rang-1 de 96,93%.

Une approche récente de Prakash et Gupta [77] fusionne les points, extraits par Speeded Up Robust Features (SURF) [78], provenant de différentes images d'un même sujet. Ils proposent d'utiliser plusieurs images d'entrée pour l'enrôlement et de stocker tous les points de fonctionnalités SURF dans le vecteur des caractéristiques fusionnées. Ces ensembles de caractéristiques sont ensuite utilisés pour former un classificateur du plus proche voisin pour assigner deux points de caractéristique corrélés. Si la distance entre deux points caractéristiques de SURF est inférieure à un seuil donné, ils sont considérés comme étant corrélés. L'évaluation de cette approche a été effectuée sur la collection UNDE et sur les deux sous-ensembles de la base de données IIT Kanpur. Prakash et Gupta ont testé l'influence de différents paramètres sur les caractéristiques du SURF et pour le classificateur du plus proche voisin. Selon la composition des paramètres, l'EER varie entre 6,72% et 2,25%.

Local Binary Pattern (LBP) est une technique locale d'extraction des caractéristiques au niveau des pixels. L'LBP encode le voisinage local d'un pixel en stockant la différence entre le pixel examiné et ses voisins. Guo et al. [53] extraient l'LBP des images d'oreilles et créent des histogrammes décrivant la distribution de l'LBP locale. Dans le cas de Wang et Yan [79], la dimension du vecteur caractéristique est réduite avec une analyse discriminante linéaire avant qu'une mesure de distance euclidienne vient quantifier la similarité de deux vecteurs caractéristiques. Wang et Yan ont évalué leur approche sur l'ensemble de données USTB II et ont rapporté une performance de rang-1 de 100%.

#### 4.3.2.3 Approches Hybrides

Liu et al. combinent la vue en avant et en arrière de l'oreille en extrayant les caractéristiques par utilisation de la méthode du rapport triangulaire et la caractéristique des moments de Tchebichef [80]. Les moments de Tchebichef sont un ensemble de fonctions orthogonales basées sur des polynômes discrets de Tchebichef et ont été introduit comme méthode de représentation en 2001 [81]. L'arrière de l'oreille est décrit par un certain nombre de lignes qui sont perpendiculaires à l'axe le plus long dans le contour de l'oreille. Ces lignes mesurent le diamètre local de l'oreillette à des points prédéfinis. La performance de rang-1 rapportée par cette approche combinée est de 97,5%. Si seule la vue de face est utilisée, la performance est de 95% et pour les images de l'arrière, Liu et al. obtiennent un taux de classification 86,3% de rang-1.

Lu et al. [82] ainsi que Yuan et Mu [83] utilisent le Modèle à Forme Active (Active Shape Model « ASM ») pour extraire le contour de l'oreille. Lu et al. utilisent des images d'oreilles segmentées manuellement à partir de 56 sujets dans différentes poses. Le vecteur de caractéristiques stocke des points sélectionnés sur le contour de l'oreille avec leur distance. Avant d'appliquer un classificateur linéaire, la dimension des vecteurs caractéristiques est réduite par application de l'Analyse en Composantes Principales (PCA). Lu et al. Comparent le rendement de rang-1 des pipelines où seulement l'oreille gauche ou l'oreille droite a été utilisé pour l'identification. Ils montrent que l'utilisation des deux oreilles augmente la performance de 93,3% à 95,1%. Dans la base de données USTB III, Yuan et Mu rapportent une performance de rang-1 de 90% si la rotation de la tête est inférieure à 15 degrés. Pour des angles de rotation compris entre 20 et 60 degrés, le rendement chute à 80%.

#### 4.3.2.4 Classifications et Approches Statistiques

Victor et al. ont été les premiers à avoir ramener l'utilisation de la PCA depuis la reconnaissance faciale à la reconnaissance de l'oreille [84]. Ils ont rapporté que la

performance de l'oreille comme une caractéristique est inférieure à celle du visage. Cela peut être dû au fait que dans leurs expériences ils supposent une symétrie de l'oreille gauche et de l'oreille droite. Ils ont utilisé l'une des deux oreilles pour la formation et l'autre pour les tests, ce qui aurait pu diminuer la performance de l'APC dans ce cas. La performance de rang-1 obtenue est de 72,2% dans la collection UND.

Chang et al. [85] rapportent un rendement significativement meilleur que Victor et al. Alaraj et al. [86] ont publié une autre étude, où la PCA est utilisé pour la représentation des caractéristiques dans la reconnaissance de l'oreille. Dans leur approche, un réseau de neurones multicouche a été formé pour la classification des composantes des caractéristiques. La performance de rang-1 obtenue est de 96%. Ils améliorent donc les résultats antérieurs de Victor et al. et Chang et al. Cependant, il convient de noter que ce résultat est uniquement basé sur un sous-ensemble de la collection UND, qui comprend 85 images d'oreilles de 17 sujets.

Zhang et Mu ont mené des études sur l'efficacité des méthodes statistiques en combinaison avec les classificateurs. Dans [87], ils montrent que l'Analyse en Composantes Indépendantes (ICA) est plus efficace sur la base de données USTB I que la PCA. Ils ont d'abord utilisé la PCA et l'ICA pour réduire la dimension des images d'entrée et ensuite formé une Machine à Vecteur de Support (Support Vector Machine « SVM »), pour classer les vecteurs de caractéristiques extraites. De plus, ils ont étudié l'influence de la taille de l'ensemble d'entraînement sur la performance. En fonction de cette taille, la performance de rang-1 varie entre 85% et 94,12% pour la PCA, alors que celle de rang-2 varie entre 91,67% et 100% pour l'ICA.

Xie et Mu [88] proposent un algorithme nommé Improved Locally Linear Embedding (ILLE) pour réduire la dimension des caractéristiques de l'oreille. L'ILLE est une technique pour projeter des points de données de grande dimension dans un système de coordonnées de dimension inférieure tout en préservant la relation entre eux. Cela nécessite que les points de données soient étiquetés d'une certaine façon, de sorte que leur relation soit fixe. La version améliorée de ILLE présentée par Xie et Mu a éliminé le problème en utilisant une fonction de distance différente. En outre, Xie et Mu montrent que ILLE est supérieur à la PCA et la PCA à noyau (Kernel PCA), si les données d'entrée contiennent des variations de pose.

Leurs études, menées sur la base de données USTB III, a montré que la performance au rang-1 de l'ILLE régulière (43%) est améliorée de manière significative par leur méthode et atteint 60,75%. Si la variation de la pose est de seulement 10 degrés, l'approche ILLE améliorée a obtenu une performance de rang-1 de 90%.

Dans leur approche, Nanni et Lumini [89] proposent d'utiliser la Sequential Forward Floating Selection (SFFS), qui est une méthode itérative statistique, pour la sélection des caractéristiques. SFFS essaie de trouver le meilleur ensemble de classificateurs en créant un ensemble de règles qui correspond le mieux à l'ensemble de fonctionnalités actuelle. Les ensembles sont créés en ajoutant un classificateur à la fois et en évaluant sa puissance discriminative avec une fonction de fitness. Si le nouvel ensemble de règles surpasse la version précédente, la nouvelle règle est ajoutée à l'ensemble final de règles. Les expériences ont été effectuées sur la collection UND E et les classificateurs uniques sont fusionnés en utilisant la règle de somme pondérée. SFFS sélectionne les sous-fenêtres les plus discriminatives qui correspondent à l'ensemble de règles le plus adapté. Nanni et Lumini rapportent un taux de reconnaissance de rang\_1 de 80% et un taux de reconnaissance de rang\_5 de 93%. L'EER varie entre 6.07% et 4.05% selon le nombre de sous-fenêtres utilisées pour la reconnaissance.

Yaqubi et al. utilisaient des caractéristiques obtenues par une combinaison de détecteurs de contour de position et d'échelle sur plusieurs positions et orientations de l'image [90]. Cette méthode d'extraction de caractéristiques est appelée modèle HMAX. Les caractéristiques extraites sont classées avec SVM et K-NN. La performance de rank-1, sur un petit ensemble de données de 180 images d'oreilles de 6 sujets, varie entre 62% et 100% selon le type de caractéristiques de base.

Gutierrez et al. [91] divisaient les images d'oreilles coupées en trois parties de tailles égales. La partie supérieure, la partie médiane et la partie inférieure. Chacune de ces sous-images est décomposée par la transformée en ondelettes puis introduite dans un réseau neuronal. Dans chaque module du réseau, différentes architectures et fonctions d'apprentissages ont été utilisées. Les résultats de chacun des modules sont fusionnés dans la dernière étape pour obtenir la décision finale. Les résultats de rang-1 varient entre 88,4% et 97,47% sur la base de données USTB I.

Dans [92] Nasseem et al. Proposent un algorithme de classification général basé sur la théorie de la compression sensorielle. Ils supposent que la plupart des signaux sont compressibles et que toute fonction de compression donne lieu à une représentation éparsée de ces signaux. Dans leurs expériences, sur les bases de données UND et FEUD, ils montrent que leur méthode de représentation éparsée est robuste contre les variations de poses et les conditions d'éclairage variables. La performance de rang-1 variait entre 89,13% et 97,83%, selon la base de données utilisée dans l'expérience.

#### 4.4 Reconnaissance d'oreille 3D

Dans la reconnaissance de l'oreille 2D, les variations de pose de position de la caméra, appelées rotations hors plan, sont encore des défis non résolus. Une solution possible consiste à utiliser des modèles 3D plutôt que des photos comme références, car une représentation 3D du sujet peut être adaptée à n'importe quelle rotation, échelle et translation. De plus, les informations de profondeur contenues dans les modèles 3D peuvent être utilisées pour améliorer la précision d'un système de reconnaissance d'oreille. Cependant, la plupart des systèmes de reconnaissance d'oreille 3D ont tendance à être coûteux en termes de calcul. Dans le tableau 4.4, tous les systèmes de reconnaissance d'oreilles 3D décrits dans cette section sont résumés.

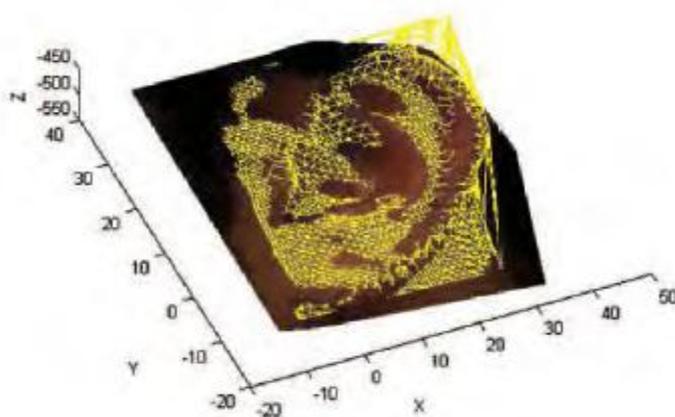
**Table 4.4** : Biométrie de l'oreille 3D: étude de détection, extrait de caractéristiques et méthodes de reconnaissance.

Référence	Méthode	Nombre de classes de la base de données	Image	Performance
Cadavid et al. [17]	ICP and Shape from shading	462	-	95%
Chen and Bannu [49]	Local Surface Patch	302	604	96.36%
Chen and Bhanu [18]	ICP Contour Matching	52	213	93.3%
Liu and Zhang [80]	Slice Curve Matching	50	200	94.5%
Islam et al. [93]	ICP with reduced meshes	415	830	93.98%
Islam et al. [94]	Local Surface Features with ICPMatching	415	830	93.5%
Passalis et al. [95]	Reference ear model with morphing	525	1031	94.4%
Yan and Bowyer [96]	ICP using voxels	369	738	97.3%
Yan and Bowyer [36]	ICP using Model Points	415	1386	97.8%
Zheng et al. [97]	Local Binary Patters	415	830	96.39%
Zhou et. al. [98]	Surface Patch Histogram and voxelization	415	830	98.6%, 1.6% EER

Bien que l'ICP soit initialement conçue pour être une approche d'enregistrement d'images, l'erreur d'enregistrement peut également être utilisée comme une mesure pour la dissimilarité de deux images 3D. Parce que l'ICP est conçue pour être un algorithme d'enregistrement, il est robuste contre tous les types de translation ou de rotations. Chen et Bhanu extraient des nuages de points du contour de l'hélice externe et enregistrent ces points avec le modèle de

référence en utilisant ICP [17]. Dans une approche ultérieure, Chen et Bhanu utilisent des patches de surface locaux (LSP) au lieu de points situés sur l'hélice externe [18]. Comme le LSP se compose de moins de points que l'hélice externe, cela réduit le temps de traitement tout en améliorant la performance de rang-1 de 93,3% avec les points de l'hélice externe à 96,63% avec LSP.

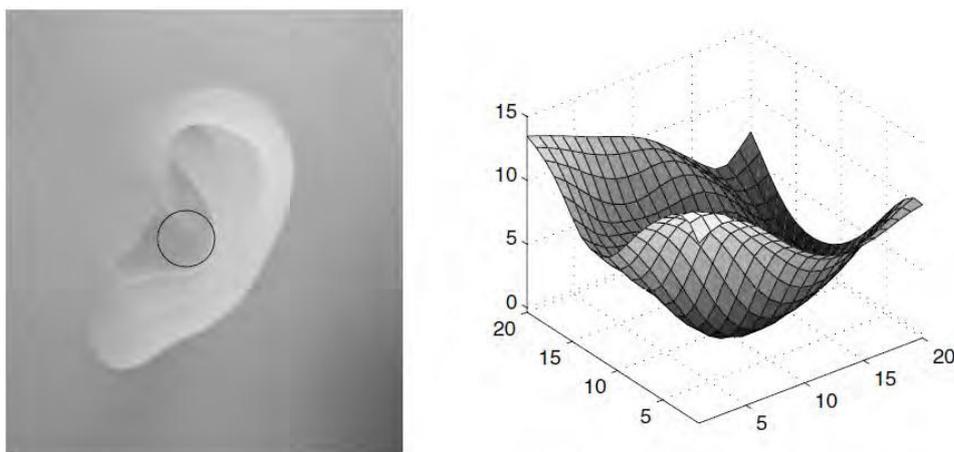
Cadavid et al. Proposent un système de reconnaissance d'oreilles en temps réel. Ils reconstituent les modèles 3D à partir des images 2D CCTV en utilisant la technique d'ombrage [99]. Le modèle 3D est ensuite comparé aux images 3D de référence, qui sont stockées dans la galerie. L'alignement du modèle ainsi que le calcul de la mesure de dissimilarité est effectué par l'ICP. Cadavid et al. rapportent un taux de reconnaissance de 95% sur une base de données de 402 sujets. Il est indiqué dans [99] que l'approche a des difficultés avec les variations de poses. Dans [98], Zhou et. Al. utilisent une combinaison d'histogrammes locaux avec des modèles voxel. Ils rapportent que leur approche est plus rapide avec un EER de 1,6%, et qu'elle est également plus précise que les algorithmes de comparaison basés sur l'ICP proposés par Chen et Bhanu et Yan et Browyer.



**Figure 4.7:** Exemples de caractéristiques de surface dans des images auriculaires 3D. L'image montre un exemple de comparaison basée sur l'ICP tel que proposé dans [17],

Dans une approche ultérieure, Islam et al. extraient des patches de surface locaux comme le montre la figure 4.8 et les utilisent comme caractéristiques [93]. Pour extraire ces LSP, un certain nombre de points est sélectionné aléatoirement à partir du modèle 3D. La PCA est alors appliquée pour trouver les caractéristiques les plus descriptives dans la LSP. La procédure est répétée jusqu' à ce que le nombre désiré de caractéristiques ait été trouvé. L'approche a été évaluées sur la base UND et le taux de reconnaissance trouvé est de 93,5%.

Zheng et al. extraient l'indice de forme à chaque point du modèle 3D et l'utilisent pour projeter le modèle 3D dans l'espace 2D [97]. L'indice de forme 3D à chaque pixel est représenté par une valeur de gris à la position correspondante dans l'image 2D. Les fonctions SIFT sont ensuite extraites à partir de l'indice de forme. Pour chacun des points SIFT, un système de coordonnées local est calculé où l'axe z correspond à la normale du point caractéristique. Par conséquent, les valeurs z de l'image d'entrée sont normalisées en fonction de la normale du point caractéristique SIFT auquel elles ont été affectées. Dès que les valeurs z ont été normalisées, elles sont transformées en une image de niveau de gris. En conséquence, Zheng et al. obtiennent une image aux niveaux de gris local pour chacune des fonctions SIFT sélectionnée. Ensuite l'LBP sont extraites pour la représentation des caractéristiques dans chacune de ces images aux niveaux de gris. La comparaison est d'abord effectuée grossièrement entre les points clés de l'indices de forme, puis ils utilisent la distance Earth mover's pour comparer les histogrammes LBP des images grises normalisées correspondantes. Zheng et al. ont évalué leur approche sur un sous-ensemble de UND-J2 et ont obtenu une performance de rang-1 de 96,39%.



**Figure 4.8:** Exemple de caractéristiques de patch de surface local (LSP) comme proposé dans [93]

#### 4.5 Bases de Données Disponibles pour la Détection et la Reconnaissance des Oreilles

Afin de tester et de comparer les performances de détection ou de reconnaissance d'un système de vision par ordinateur en général et d'un système biométrique en particulier, des bases de données d'images de taille suffisante sont mises à la disposition du public. Ces bases de données peuvent être téléchargées gratuitement. Dans cette section, nous donnons un aperçu de ces bases de données..

##### 4.5.1 USTB Data-Bases

L'Université des sciences et de la technologie de Beijing (USTB) offre quatre collections <sup>1,2</sup> d'images de profil d'oreilles et de visage 2D à la communauté de la recherche. Toutes les bases de données USTB sont disponibles sous licence.

- **Base de données I:** L'ensemble de données contient 180 images au total. Elles ont été prises de 60 sujets en 3 sessions entre juillet et août 2002. La base de données ne contient que des images de l'oreille droite de chaque sujet. Pendant chaque session, les images ont été prises dans des conditions d'éclairage différentes et avec une rotation différente. Les sujets étaient des étudiants et des enseignants de l'USTB.
- **Base de données II:** De la même façon que la base de données I, cette collection contient des images d'oreilles droites provenant d'étudiants et d'enseignants de l'USTB. Cette fois-ci, le nombre de sujets est de 77 et il y a eu 4 séances différentes entre novembre 2003 et janvier 2004. La base de données contient donc 308 images qui ont été prises dans des conditions d'éclairage différentes.
- **Base de données III:** Dans cet ensemble de données 79 sujets sont considérés. Les étudiants et les enseignants de l'USTB ont été photographiés dans différentes poses entre novembre 2004 et décembre 2004. Certaines oreilles sont occluses par les cheveux. Chaque sujet tournait sa tête de 0 à 60 degrés vers la droite et de 0 à 45 degrés vers la gauche. Cela a été répété deux jours différents pour chaque sujet, ce qui a abouti à 1600 images au total.
- **Base de données IV:** Composé de 25500 images de 500 sujets pris entre juin 2007 et décembre 2008, il s'agit du plus grand ensemble de données de l'USTB. Le système de capture se compose de 17 caméras et, est capable de prendre 17 photos du sujet simultanément. Ces caméras sont réparties en cercle autour du sujet, qui est placé au centre. L'intervalle entre les caméras est de 15 degrés. Chaque volontaire a été invité à regarder vers le haut, le bas et au niveau de l'œil, ce qui signifie que cette base de données contient des images à différentes poses. Cette base de données ne contient qu'une seule session pour chaque sujet.

#### 4.5.2 UND Data-Bases

L'Université de Notre Dame (UND) offre une grande variété de bases de données d'images différentes, qui peuvent être utilisées pour l'évaluation de la performance biométrique. Parmi celles-ci, cinq bases de données contiennent des images 2D et des images en profondeur, qui conviennent à l'évaluation des systèmes de reconnaissance auriculaire. Toutes les bases de données d'UND peuvent être mises à disposition sous licence <sup>3</sup>.

- **Collection E:** 464 images de profil à droite de 114 sujets, capturées en 2002. Pour chaque sujet, entre 3 et 9 images ont été prises à des jours différents et sous différentes pose et éclairage.
- **Collection F:** 942 images 3D (images en profondeur) et images de profil 2D correspondantes à 302 sujets, capturés en 2003 et 2004.
- **Collection G:** 738 images 3D et images de profil 2D correspondantes à 235 sujets, capturés entre 2003 et 2005
- **Collection J2:** 1800 images 3D et images de profil 2D correspondantes à 415 sujets, capturés entre 2003 et 2005 [36].
- **Collection NDOff-2007:** 7398 images 3D et images 2D correspondantes à 396 sujets. La base de données contient différentes poses et différentes condition d'éclairage [100].

#### 4.5.3 WPUT-DB

West Pommeranian University of Technology a collecté une base de données sur les oreilles dans le but de fournir des données plus représentatives [101]. La base de données contient 501 sujets de tous âges et 2071 images au total. Pour chaque sujet, la base de données contient entre 4 et 8 images, prises sur des jours différents et dans des conditions d'éclairage différentes. Les sujets portent également des coiffures, des boucles d'oreilles et des aides auditives, et en plus de cela, certaines oreilles sont occluses par les cheveux. Dans la figure 4.9, certaines images d'exemples de cette base sont montrées.



**Figure 4.9:** Exemples d'images de la base de données d'oreille WPUT [101].

La base de données contient des photographies de qualité variable et prises dans des conditions d'éclairage différentes. En outre, la base de données contient des images, où l'oreille est occluse par les cheveux ou par des boucles d'oreilles.

#### 4.5.4 IIT Delhi

La base de données IIT Delhi est fournie par l'Université Polytechnique de Hong Kong [71].

Elle contient des images auriculaires qui ont été recueillies entre octobre 2006 et juin 2007 à

<sup>1</sup>[http://www1.ustb.edu.cn/resb/en/doc/Imagedb\\_123\\_intro\\_en.pdf](http://www1.ustb.edu.cn/resb/en/doc/Imagedb_123_intro_en.pdf)

<sup>2</sup>[http://www1.ustb.edu.cn/resb/en/doc/Imagedb\\_4\\_intro\\_en.pdf](http://www1.ustb.edu.cn/resb/en/doc/Imagedb_4_intro_en.pdf)

<sup>3</sup>[http://cse.nd.edu/cvrl/CVRL/Data\\_Sets.html](http://cse.nd.edu/cvrl/CVRL/Data_Sets.html)

l'Indian Institute of Technology de Delhi à New Delhi (voir figure 4.10). La base de données contient 121 sujets, et au moins 3 images ont été prises par sujet dans un environnement intérieur, ce qui signifie que la base de données se compose de 421 images au total.



**Figure 4.10:** Exemples d'images de la base de données de l'oreille de l'IIT de Delhi [71].

#### 4.5.5 IIT Kanpur

La base de données IITK a été apportée par l'Indian Institute of Technology à Kanpur<sup>6</sup>[77]. Cette base de données se compose de deux sous-ensembles.

**A. Sous-ensemble I:** Cette base de données contient 801 images de face latérale collectées à partir de 190 sujets. Le nombre d'images acquises par individu varie de 2 à 10.

**B. Sous-ensemble II:** Les images de ce sous-ensemble ont été prises à partir de 89 individus. Pour chaque sujet 9 images ont été prises avec trois poses différentes. Chaque pose a été capturée à trois échelles différentes. Toutes les images ont été prises le même jour. Il n'est pas précisé si le sous-ensemble II contient les mêmes sujets que le sous-ensemble I.

#### 4.5.6 La ScFace

La base de données SCface est fournie par l'Université Technique de Zagreb<sup>7</sup> [102] et contient 4160 images de 130 sujets. Le but est de fournir une base de données, qui convient pour tester des algorithmes dans des scénarios de surveillance. Malheureusement, toutes les images de la caméra de surveillance ont été prises à un angle frontal de sorte que les oreilles ne sont pas visibles sur ces images.

Cependant, la base de données contient également un ensemble de photographies haute résolution de chaque sujet, qui montre le sujet à différentes poses. Ces poses comprennent des vues du profil droit et gauche, comme le montre la figure 4.11.

---

<sup>6</sup><http://www.cse.iitk.ac.in/users/biometrics/>

<sup>7</sup><http://www.scface.org/>



**Figure 4.11:** Exemples d'images de SCface [102]. Ces exemples proviennent d'images photographiées et non pas d'images collectées avec la caméra de surveillance.

#### 4.5.7 Sheffield Face

Cette base de données était connue sous le nom UMIST<sup>8</sup> data-base et se compose de 564 images de 20 sujets de race et de genre mixtes. Chaque sujet est photographié dans une gamme de poses différentes, y compris une vue frontale et des vues de profil.

#### 4.5.8 YSU

La Youngston State University a collecté un nouveau type de base de données biométriques pour l'évaluation des forensic identification systems [103]. Pour chacun des 259 sujets, 10 images sont fournies. Les images sont capturées à partir des vidéos et montrent le sujet dans des poses entre 0 et 90 degrés. Cela signifie que la base de données contient des images de profil droit et une image de vue frontale pour chaque sujet.

#### 4.5.7 NCKU

L'Université nationale Cheng Kung de Taiwan a collecté une base de données d'images, composée de 37 images pour chacun des 90 sujets. Elle peut être téléchargée à partir du site Web de l'université. Chaque sujet est photographié dans des angles différents entre -90 degrés (profil de gauche) et 90 degrés (profil de droite), par pas de 5 degrés. Dans la Figure 4.12, certains exemples sont affichés. Une telle série d'images est collectée à deux jours différents pour chacun des sujets. Toutes les images ont été prises dans les mêmes conditions d'éclairage et avec la même distance entre le sujet et la caméra. Comme ces données ont été initialement recueillies pour la reconnaissance faciale, certaines des oreilles sont partiellement ou totalement occluses par les cheveux, ce qui rend ces données difficiles pour les approches de détection de l'oreille. Par conséquent, seul un sous-ensemble de cette base de données est adapté à la reconnaissance de l'oreille.

<sup>8</sup><http://www.sheffield.ac.uk/eee/research/iel/research/face>



**Figure 4.12:** Quelques exemples d'images de la base de données NICU, montrant le même sujet à des angles différent.

#### 4.5.8 UBEAR

L'ensemble de données présentes dans [104] contient des images de l'oreille gauche et de l'oreille droite de 126 sujets. Les images ont été prises dans des conditions d'éclairage variables et les sujets n'ont pas été invités à enlever les cheveux, les bijoux ou les coiffes avant de prendre les photos. Les images sont groupé à partir des Stream vidéo, ce qui montre le sujet dans des poses différentes, comme regarder vers la caméra, vers le haut ou vers le bas.

#### 4.6 Défis ouverts et Applications Futures

Comme le montrent les publications les plus récentes sur la reconnaissance d'oreille 2D et 3D, l'application principale de cette technique est l'identification personnelle dans des environnements non contraints. Cela inclut des applications pour la surveillance intelligente, mais aussi l'identification des criminelles (forensic) sur les images de vidéosurveillance ou pour les systèmes de contrôle. Ces champs d'application font partie des systèmes de reconnaissance faciale, mais comme l'oreille est située à côté du visage, elle peut fournir de précieuses informations supplémentaires pour compléter les images faciales.

Les systèmes multi-modes d'identification de l'oreille et du visage peuvent servir à obtenir une invariance de pose et une plus grande robustesse contre l'occlusion dans des environnements non contraints. Dans la plupart des lieux publics, des caméras de surveillance sont placées au-dessus pour capturer le plus de personnes possible. En outre, la plupart des personnes ne regardent pas directement dans la caméra, donc dans la plupart des cas, aucune image frontale des personnes ne sera disponible. Ce fait pose de sérieux problèmes aux systèmes biométriques, en utilisant des traits faciaux pour l'identification. Si le visage n'est pas visible à partir d'un angle frontal, l'oreille peut servir de caractéristique supplémentaire utile dans ces scénarios. En raison de la proximité physique du visage et de l'oreille, il existe également de nombreuses possibilités pour la fusion biométrique de ces deux modalités. Les images de

visage et d'oreille peuvent être fusionnées. Dans le contexte de cette demande, il ya des défis non résolus, qui devraient être abordés par les recherches futures dans ce domaine.

#### **4.6.1 Localisation Automatique de l'Oreille**

Le fait que de nombreux systèmes présentés dans la littérature utilisent des images d'oreille pré segmentées, montre que la détection automatique des oreilles, en particulier dans les images de la vie réelle, est toujours un problème non résolu. Si des systèmes de reconnaissance auriculaire doivent être mis en œuvre dans des systèmes d'identification automatique, des approches rapides et fiables pour la détection automatique des oreilles sont importantes. Comme première étape vers cet objectif, certains chercheurs ont publié des données qui simulent des variations typiques dans des environnements incontrôlés tels que des conditions d'éclairage variables, des poses et des occlusions. Sur la base de ces ensembles de données, les approches actuelles et futures de la reconnaissance auriculaire devraient être testées dans des conditions réalistes afin d'améliorer leur fiabilité. En outre, les systèmes d'imagerie 3D deviennent de plus en plus bon marché ces dernières années. Par conséquent, la reconnaissance de l'oreille 3D devient importante et avec elle la nécessité de localiser les oreilles dans les images en profondeur ou des modèles 3D. Actuellement, une seule approche pour la détection de l'oreille en profondeur a été publiée, ce qui est une première étape vers la détection de l'oreille dans les images 3D.

#### **4.6.2 Occlusion et Variations de Pose**

Contrairement au visage, l'oreille peut être partiellement ou entièrement couverte par les cheveux ou par d'autres éléments tels que des coiffures, des prothèses auditives, des bijoux ou des écouteurs. En raison de la surface convexe de l'oreille externe, des parties de celle-ci peuvent également être occluses si la pose du sujet change. Dans certaines publications, la robustesse contre l'occlusion est explicitement traitée, mais il n'existe aucune étude sur l'effet de certains types d'occlusion comme les cheveux ou les boucles d'oreilles sur le taux de reconnaissance d'un système de reconnaissance de l'oreille. Une fois de plus, la disponibilité de bases de données publiques contenant des images d'oreilles occluses est susceptible de favoriser le développement de solutions pour proposer des algorithmes invariants et robustes pour la détection d'oreilles et l'extraction de caractéristiques.

De plus, à notre connaissance, il n'existe aucune étude sur la visibilité de l'oreille externe dans différents environnements publics. Afin de développer des algorithmes pour la détection et la reconnaissance de l'oreille, des informations supplémentaires sur les parties généralement occluses de l'oreille est nécessaire. L'occlusion due à des variations de pose est un autre défi

non satisfait dans le système de reconnaissance de l'oreille. De même que pour la reconnaissance faciale, certaines parties de l'oreille peuvent être obstruées si la pose change. Récemment, certaines méthodes d'extraction de caractéristiques ont été proposées, qui sont robustes contre des variations de pose dans une certaine mesure. Cependant, cette question n'est pas encore entièrement résolue. Une autre possibilité de compenser les variations de pose pourrait être l'utilisation de modèles 3D au lieu d'images en profondeur de photographies.

#### **4.6.3 Évolutivité (Scalability)**

Les bases de données actuellement disponibles ne contiennent que moins de 10 000 images d'oreille. La seule exception est la collection USTB IV, qui n'a pas encore été mise à la disposition du public. Dans des environnements réalistes, la taille de la base de données sera considérablement plus grande, ce qui rend la recherche exhaustive dans les scénarios d'identifications. Par conséquent, non seulement la précision mais aussi la vitesse de comparaison des systèmes de reconnaissance auriculaire seront intéressantes pour la future recherche. Afin de rendre la reconnaissance de l'oreille applicable aux systèmes à grande échelle, les recherches approfondies devraient être remplacées par des structures de données appropriées permettant la complexité logarithmique pendant la recherche. Cela pourrait par exemple être réalisé en explorant les possibilités d'organisation des modèles d'oreille dans les arbres de recherche.

#### **4.6.4 Comprendre la Symétrie et le Vieillessement**

Parce que la reconnaissance de l'oreille est l'un des nouveaux domaines de la recherche biométrique, la symétrie de l'oreille gauche et l'oreille droite n'a pas encore été entièrement comprise. Une étude menée par Abaza et Ross [12] indique qu'il existe un certain degré de symétrie entre les oreilles gauche et droite, ce qui pourrait être exploité en comparant les oreilles gauche et droite. Leur résultat encourage une recherche plus poussée sur les contraintes de symétrie entre l'oreille gauche et l'oreille droite.

Les études d'Iannarelli indiquent que certaines caractéristiques de l'oreille externe peuvent être héritées et le vieillissement affecte légèrement l'aspect de l'oreille externe. Les deux hypothèses pourraient être confirmées par des études plus récentes, mais à cause de l'insuffisance des données, l'effet de vieillissement sur l'apparence de l'oreille externe n'est pas encore complètement compris. En outre, il n'existe pas encore d'études à grande échelle sur la relation de symétrie entre l'oreille gauche et l'oreille droite.

## 4.7 Conclusion

Nous avons présenté une étude sur l'état de l'art en biométrie auriculaire 2D et 3D, couvrant les systèmes de détection d'oreille et de reconnaissance auriculaire. Nous avons catégorisé le grand nombre d'approches de reconnaissance d'oreille 2D en holistique, local, hybride et méthodes statistiques, discuté de leurs caractéristiques et leurs performances.

La reconnaissance de l'oreille est encore un nouveau domaine de recherche. Bien qu'il y ait un certain nombre d'approches prometteuses, aucun d'entre eux n'a été évalué dans des scénarios réalistes qui incluent des facteurs perturbateurs comme les variations de pose, l'occlusion et les conditions d'éclairage variables. Dans les approches récentes, ces facteurs sont pris en compte, mais plus de recherche à ce sujet est nécessaire jusqu'à ce que les systèmes de reconnaissance auriculaire puissent être utilisés dans la pratique. La disponibilité de bases de données d'essai appropriées, qui ont été recueillies dans des scénarios réalistes, contribuera davantage à la maturation de l'oreille en tant que caractéristique biométrique.

Nous avons présenter dans ce chapitre les différentes bases de données disponibles, les approches existantes de détection et de reconnaissance de l'oreille et les problèmes non encore résolus pour la reconnaissance de l'oreille dans le contexte du système de surveillance intelligente, que nous considérons comme l'application la plus importante pour la biométrie auriculaire. Nous pensons que cette nouvelle empreinte est une extension valable pour les systèmes de reconnaissance faciale sur la façon de poser une identification automatique invariante.

**Bibliographie :**

- [1] Maltoni, D., Maio, D., Jain, A., & Prabhakar, S. (2009). Handbook of fingerprint recognition. Springer Science & Business Media.
- [2] Hurley, D. J., Nixon, M. S., & Carter, J. N. (2005). Force field feature extraction for ear biometrics. *Computer Vision and Image Understanding*, 98(3), 491-512.
- [3] Cadavid, S., Fathy, S., Zhou, J., & Abdel-Mottaleb, M. (2011, October). An adaptive resolution voxelization framework for 3d ear recognition. In *Biometrics (IJCB), 2011 International Joint Conference on* (pp. 1-6). IEEE.
- [4] Juefei-Xu, F., Luu, K., Savvides, M., Bui, T. D., & Suen, C. Y. (2011, October). Investigating age invariant face recognition based on periocular biometrics. In *Biometrics (IJCB), 2011 International Joint Conference on* (pp. 1-7). IEEE.
- [5] Smeets, D., Keustermans, J., Hermans, J., Claes, P., Vandermeulen, D., & Suetens, P. (2011, October). Symmetric surface-feature based 3D face recognition for partial data. In *Biometrics (IJCB), 2011 International Joint Conference on* (pp. 1-6). IEEE.
- [6] Daugman, J. (2007). New methods in iris recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 37(5), 1167-1175.
- [7] Ko, J. G., Gil, Y. H., Yoo, J. H., & Chung, K. I. (2007). A novel and efficient feature extraction method for iris recognition. *ETRI journal*, 29(3), 399-401.
- [8] Shen, L., & Kot, A. (2009). A New Wavelet Domain Feature for Fingerprint Recognition (< Special Issue> BIOMETRICS AND ITS APPLICATIONS). *Biomedical fuzzy and human sciences: the official journal of the Biomedical Fuzzy Systems Association*, 14(1), 55-59.
- [9] Win, Z. M., & Sein, M. M. (2011, November). Texture feature based fingerprint recognition for low quality images. In *Micro-NanoMechatronics and Human Science (MHS), 2011 International Symposium on* (pp. 333-338). IEEE.
- [10] Iannarelli, A. V. (1989). Ear identification. Paramont Publishing Company.
- [11] Bertillon, A. (1890). *La photographie judiciaire: avec un appendice sur la classification et l'identification anthropométriques*. Gauthier-Villars.
- [12] Abaza, A., Ross, A., Hebert, C., Harrison, M. A. F., & Nixon, M. S. (2013). A survey on ear biometrics. *ACM computing surveys (CSUR)*, 45(2), 22.
- [13] Islam, S. M., Bennamoun, M., Owens, R., & Davies, R. (2008). Biometric approaches of 2D-3D ear and face: A survey. In *Advances in computer and information sciences and engineering* (pp. 509-514). Springer Netherlands.
- [14] Pun, K. H., & Moon, Y. S. (2004, May). Recent Advances in Ear Biometrics. In *FGR (Vol. 4, p. p164)*.
- [15] Choras, M. (2007, June). Image feature extraction methods for ear biometrics--a survey. In *Computer Information Systems and Industrial Management Applications, 2007. CISIM'07. 6th International Conference on* (pp. 261-265). IEEE.

- [16] Chen, H., & Bhanu, B. (2005, June). Shape model-based 3D ear detection from side face range images. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops (pp. 122-122). IEEE.
- [17] Chen, H., & Bhanu, B. (2005, January). Contour matching for 3D ear recognition. In Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on (Vol. 1, pp. 123-128). IEEE.
- [18] Chen, H., & Bhanu, B. (2007). Human ear recognition in 3D. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4), 718-737.
- [19] Zhou, J., Cadavid, S., & Abdel-Mottaleb, M. (2010, September). Histograms of categorized shapes for 3D ear detection. In Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on (pp. 1-6). IEEE.
- [20] Prakash, S., & Gupta, P. (2012, March). An efficient technique for ear detection in 3D: invariant to rotation and scale. In 2012 5th IAPR International Conference on Biometrics (ICB) (pp. 97-102). IEEE.
- [21] Abaza, A., & Harrison, M. A. F. (2013, May). Ear recognition: a complete system. In SPIE Defense, Security, and Sensing (pp. 87120N-87120N). International Society for Optics and Photonics.
- [22] Ansari, S., & Gupta, P. (2007, March). Localization of ear using outer helix curve of the ear. In Computing: Theory and Applications, 2007. ICCTA'07. International Conference on (pp. 688-692). IEEE.
- [23] Ansari, S., & Gupta, P. (2007, March). Localization of ear using outer helix curve of the ear. In Computing: Theory and Applications, 2007. ICCTA'07. International Conference on (pp. 688-692). IEEE.
- [24] Alvarez, L., González, E., & Mazorra, L. (2005, October). Fitting ear contour using an ovoid model. In Proceedings 39th Annual 2005 International Carnahan Conference on Security Technology (pp. 145-148). IEEE.
- [25] Arbab-Zavar, B., & Nixon, M. S. (2007, November). On shape-mediated enrolment in ear biometrics. In International Symposium on Visual Computing (pp. 549-558). Springer Berlin Heidelberg.
- [26] Arbab-Zavar, B., & Nixon, M. S. (2008, December). Robust log-gabor filter for ear biometrics. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on (pp. 1-4). IEEE.
- [27] Attarchi, S., Faez, K., & Rafiei, A. (2008, October). A new segmentation approach for ear recognition. In International Conference on Advanced Concepts for Intelligent Vision Systems (pp. 1030-1037). Springer Berlin Heidelberg.
- [28] Cummings, A. H., Nixon, M. S., & Carter, J. N. (2010, September). A novel ray analogy for enrolment of ear biometrics. In Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on (pp. 1-6). IEEE.
- [29] Cummings, A. H., Nixon, M. S., & Carter, J. N. (2010, September). A novel ray analogy for enrolment of ear biometrics. In Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on (pp. 1-6). IEEE.

- [30] Islam, S. M., Bennamoun, M., & Davies, R. (2008, January). Fast and fully automatic ear detection using cascaded adaboost. In *Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on* (pp. 1-6). IEEE.
- [31] Ernő, J., & László, M. (2007). Model-based human ear localization and feature extraction. *International Journal of Intelligent Computing in Medical Sciences & Image Processing*, 1(2), 101-112.
- [32] Kumar, A., Hanmandlu, M., Kuldeep, M., & Gupta, H. M. (2011, December). Automatic ear detection for online biometric applications. In *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2011 Third National Conference on* (pp. 146-149). IEEE.
- [33] Liu, D. K., & Liu, H. (2010). Combining Skin Color Model and Multi-template Matching for Enhancing Adaboost Ear Detection. *Journal of Southwest University of Science and Technology*, 4, 021.
- [34] Prakash, S., & Gupta, P. (2012). An efficient ear localization technique. *Image and Vision Computing*, 30(1), 38-50.
- [35] Shih, H. C., Ho, C. C., Chang, H. T., & Wu, C. S. (2009, September). Ear detection based on arc-masking extraction and AdaBoost polling verification. In *Intelligent Information Hiding and Multimedia Signal Processing, 2009. IHH-MSP'09. Fifth International Conference on* (pp. 669-672). IEEE.
- [36] Yan, P., & Bowyer, K. W. (2007). Biometric recognition using 3D ear shape. *IEEE Transactions on pattern analysis and machine intelligence*, 29(8), 1297-1308.
- [37] Mu, Z. C. (2007, August). Ear detection based on skin-color and contour information. In *2007 International Conference on Machine Learning and Cybernetics (Vol. 4, pp. 2213-2217)*. IEEE.
- [38] Attarchi, S., Faez, K., & Rafiei, A. (2008, October). A new segmentation approach for ear recognition. In *International Conference on Advanced Concepts for Intelligent Vision Systems* (pp. 1030-1037). Springer Berlin Heidelberg.
- [39] Ansari, S., & Gupta, P. (2007, March). Localization of ear using outer helix curve of the ear. In *Computing: Theory and Applications, 2007. ICCTA'07. International Conference on* (pp. 688-692). IEEE.
- [40] Mu, Z. C. (2007, August). Ear detection based on skin-color and contour information. In *2007 International Conference on Machine Learning and Cybernetics (Vol. 4, pp. 2213-2217)*. IEEE.
- [41] Alvarez, L., González, E., & Mazon, L. (2005, October). Fitting ear contour using an ovoid model. In *Proceedings 39th Annual 2005 International Carnahan Conference on Security Technology* (pp. 145-148). IEEE.
- [42] Burge, M., & Burger, W. (1996). Ear biometrics. In *Biometrics* (pp. 273-285). Springer US.
- [43] Abdel-Mottaleb, M., & Zhou, J. (2006, January). Human ear recognition from face profile images. In *International Conference on Biometrics* (pp. 786-792). Springer Berlin Heidelberg.

- [44] Mu, Z. C. (2007, August). Ear detection based on skin-color and contour information. In 2007 International Conference on Machine Learning and Cybernetics (Vol. 4, pp. 2213-2217). IEEE.
- [45] Prakash, S., Jayaraman, U., & Gupta, P. (2008, November). Ear localization from side face images using distance transform and template matching. In 2008 First Workshops on Image Processing Theory, Tools and Applications (pp. 1-8). IEEE.
- [46] Prakash, S., Jayaraman, U., & Gupta, P. (2009, February). A skin-color and template based technique for automatic ear detection. In Advances in Pattern Recognition, 2009. ICAPR'09. Seventh International Conference on (pp. 213-216). IEEE.
- [47] Cummings, A. H., Nixon, M. S., & Carter, J. N. (2010, September). A novel ray analogy for enrolment of ear biometrics. In Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on (pp. 1-6). IEEE.
- [48] Islam, S. M., Bennamoun, M., & Davies, R. (2008, January). Fast and fully automatic ear detection using cascaded adaboost. In Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on (pp. 1-6). IEEE.
- [49] Arbab-Zavar, B., & Nixon, M. S. (2007, November). On shape-mediated enrolment in ear biometrics. In International Symposium on Visual Computing (pp. 549-558). Springer Berlin Heidelberg.
- [50] Kumar, A., Hanmandlu, M., Kuldeep, M., & Gupta, H. M. (2011, December). Automatic ear detection for online biometric applications. In Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2011 Third National Conference on (pp. 146-149). IEEE.
- [51] Hurley, D. J., Nixon, M. S., & Carter, J. N. (2002). Force field energy functionals for image feature extraction. *Image and Vision computing*, 20(5), 311-317.
- [52] Dong, J., & Mu, Z. (2008, December). Multi-pose ear recognition based on force field transformation. In Intelligent Information Technology Application, 2008. IITA'08. Second International Symposium on (Vol. 3, pp. 771-775). IEEE.
- [53] Guo, Y., & Xu, Z. (2008, October). Ear recognition using a new local matching approach. In 2008 15th IEEE International Conference on Image Processing (pp. 289-292). IEEE.
- [54] Naseem, I., Togneri, R., & Bennamoun, M. (2008, December). Sparse representation for ear biometrics. In International Symposium on Visual Computing (pp. 336-345). Springer Berlin Heidelberg.
- [55] Wang, Y., Mu, Z. C., & Zeng, H. (2008, December). Block-based and multi-resolution methods for ear recognition using wavelet transform and uniform local binary patterns. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on (pp. 1-4). IEEE.
- [56] Xie, Z., & Mu, Z. (2008, December). Ear recognition using lle and idlle algorithm. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on (pp. 1-4). IEEE.
- [57] Yaqubi, M., Faez, K., & Motamed, S. (2008, May). Ear recognition using features inspired by visual cortex and support vector machine technique. In Computer and

- Communication Engineering, 2008. ICCCE 2008. International Conference on (pp. 533-537). IEEE.
- [58] Zhang, H., & Mu, Z. (2008, September). Compound structure classifier system for ear recognition. In 2008 IEEE International Conference on Automation and Logistics (pp. 2306-2309). IEEE.
- [59] Badrinath, G. S., & Gupta, P. (2009, February). Feature level fused ear biometric system. In Advances in Pattern Recognition, 2009. ICAPR'09. Seventh International Conference on (pp. 197-200). IEEE.
- [60] Kisku, D. R., Mehrotra, H., Gupta, P., & Sing, J. K. (2009, July). SIFT-based ear recognition by fusion of detected keypoints from color similarity slice regions. In Advances in Computational Tools for Engineering Applications, 2009. ACTEA'09. International Conference on (pp. 380-385). IEEE.
- [61] Xiaoyun, W., & Weiqi, Y. (2009, October). Human ear recognition based on block segmentation. In Cyber-Enabled Distributed Computing and Knowledge Discovery, 2009. CyberC'09. International Conference on (pp. 262-266). IEEE.
- [62] Alaraj, M., Hou, J., & Fukami, T. (2010, November). A neural network based human identification framework using ear images. In TENCON 2010-2010 IEEE Region 10 Conference (pp. 1595-1600). IEEE.
- [63] Bustard, J. D., & Nixon, M. S. (2010). Toward unconstrained ear recognition from two-dimensional images. *IEEE transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 40(3), 486-494.
- [64] De Marsico, M., Michele, N., & Riccio, D. (2010, June). HERO: human ear recognition against occlusions. In Proc. of Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 178-183).
- [65] Gutierrez, L., Melin, P., & Lopez, M. (2010, July). Modular neural network integrator for human recognition from ear images. In The 2010 International Joint Conference on Neural Networks (IJCNN) (pp. 1-5). IEEE.
- [66] Wang, X. Q., Xia, H. Y., & Wang, Z. L. (2010, June). The Research of Ear Identification Based On Improved Algorithm of Moment Invariant. In 2010 Third International Conference on Information and Computing (Vol. 1, pp. 58-60). IEEE.
- [67] Wang, X., & Yuan, W. (2010, July). Gabor wavelets and general discriminant analysis for ear recognition. In Intelligent Control and Automation (WCICA), 2010 8th World Congress on (pp. 6305-6308). IEEE.
- [68] Prakash, S., & Gupta, P. (2012, March). An efficient technique for ear detection in 3D: invariant to rotation and scale. In 2012 5th IAPR International Conference on Biometrics (ICB) (pp. 97-102). IEEE.
- [69] Kumar, A., Hanmandlu, M., Kuldeep, M., & Gupta, H. M. (2011, December). Automatic ear detection for online biometric applications. In Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2011 Third National Conference on (pp. 146-149). IEEE.
- [70] Wang, Z. Q., & Yan, X. D. (2011, April). Multi-scale feature extraction algorithm of ear image. In Electric Information and Control Engineering (ICEICE), 2011 International Conference on (pp. 528-531). IEEE.

- [71] Kumar, A., & Wu, C. (2012). Automated human identification using ear imaging. *Pattern Recognition*, 45(3), 956-968.
- [72] Fabate, A., Nappi, M., Riccio, D., & Ricciardi, S. (2006, August). Ear recognition by means of a rotation invariant descriptor. In *18th International Conference on Pattern Recognition (ICPR'06) (Vol. 4, pp. 437-440)*. IEEE.
- [73] Sanaa, A., Gupta, P., & Purkait, R. (2007). Ear biometrics: A new approach. *Biometrics*, 1, Introduction.
- [74] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on (Vol. 2, pp. 1150-1157)*. Ieee.
- [75] Bustard, J. D., & Nixon, M. S. (2010). Toward unconstrained ear recognition from two-dimensional images. *IEEE transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 40(3), 486-494.
- [76] Arbab-Zavar, B., Nixon, M. S., & Hurley, D. J. (2007, September). On model-based analysis of ear biometrics. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on (pp. 1-5)*. IEEE.
- [77] Prakash, S., & Gupta, P. (2013). An efficient ear recognition technique invariant to illumination and pose. *Telecommunication Systems*, 52(3), 1435-1448.
- [78] Bay, H., Tuytelaars, T., & Van Gool, L. (2006, May). Surf: Speeded up robust features. In *European conference on computer vision (pp. 404-417)*. Springer Berlin Heidelberg.
- [79] Wang, Z. Q., & Yan, X. D. (2011, April). Multi-scale feature extraction algorithm of ear image. In *Electric Information and Control Engineering (ICEICE), 2011 International Conference on (pp. 528-531)*. IEEE.
- [80] Liu, H., & Yan, J. (2007, December). Multi-view ear shape feature extraction and reconstruction. In *Signal-Image Technologies and Internet-Based System, 2007. SITIS'07. Third International IEEE Conference on (pp. 652-658)*. IEEE.
- [81] Mukundan, R., Ong, S. H., & Lee, P. A. (2001). Image analysis by Tchebichef moments. *IEEE Transactions on image Processing*, 10(9), 1357-1364.
- [82] Lu, L., Zhang, X., Zhao, Y., & Jia, Y. (2006, August). Ear recognition based on statistical shape model. In *First International Conference on Innovative Computing, Information and Control-Volume I (ICICIC'06) (Vol. 3, pp. 353-356)*. IEEE.
- [83] Yuan, L., & Mu, Z. C. (2007, September). Ear recognition based on 2D images. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on (pp. 1-5)*. IEEE.
- [84] Victor, B., Bowyer, K., & Sarkar, S. (2002). An evaluation of face and ear biometrics. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on (Vol. 1, pp. 429-432)*. IEEE.
- [85] Chang, K., Bowyer, K. W., Sarkar, S., & Victor, B. (2003). Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Transactions on pattern analysis and machine intelligence*, 25(9), 1160-1165.

- [86] Alaraj, M., Hou, J., & Fukami, T. (2010, November). A neural network based human identification framework using ear images. In TENCON 2010-2010 IEEE Region 10 Conference (pp. 1595-1600). IEEE.
- [87] Zhang, H., & Mu, Z. (2008, September). Compound structure classifier system for ear recognition. In 2008 IEEE International Conference on Automation and Logistics (pp. 2306-2309). IEEE.
- [88] Xie, Z., & Mu, Z. (2008, December). Ear recognition using lle and idlle algorithm. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on (pp. 1-4). IEEE.
- [89] Nanni, L., & Lumini, A. (2007). A multi-matcher for ear authentication. Pattern Recognition Letters, 28(16), 2219-2226.
- [90] Yaqubi, M., Faez, K., & Motamed, S. (2008, May). Ear recognition using features inspired by visual cortex and support vector machine technique. In Computer and Communication Engineering, 2008. ICCCE 2008. International Conference on (pp. 533-537). IEEE.
- [91] Gutierrez, L., Melin, P., & Lopez, M. (2010, July). Modular neural network integrator for human recognition from ear images. In The 2010 International Joint Conference on Neural Networks (IJCNN) (pp. 1-5). IEEE.
- [92] Naseem, I., Togneri, R., & Bennamoun, M. (2008, December). Sparse representation for ear biometrics. In International Symposium on Visual Computing (pp. 336-345). Springer Berlin Heidelberg.
- [93] Islam, S. M., Davies, R., Bennamoun, M., & Mian, A. S. (2011). Efficient detection and recognition of 3D ears. International Journal of Computer Vision, 95(1), 52-73.
- [94] Islam, S. M., Davies, R., Mian, A. S., & Bennamoun, M. (2008, October). A fast and fully automatic ear recognition approach based on 3D local surface features. In International Conference on Advanced Concepts for Intelligent Vision Systems (pp. 1081-1092). Springer Berlin Heidelberg.
- [95] Passalis, G., Kakadiaris, I. A., Theoharis, T., Toderici, G., & Papaioannou, T. (2007, September). Towards fast 3D ear recognition for real-life biometric applications. In Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on (pp. 39-44). IEEE.
- [96] Yan, P., & Bowyer, K. W. (2007). A fast algorithm for ICP-based 3D shape biometrics. Computer Vision and Image Understanding, 107(3), 195-202.
- [97] Zeng, H., Dong, J. Y., Mu, Z. C., & Guo, Y. (2010, October). Ear recognition based on 3d keypoint matching. In IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS (pp. 1694-1697). IEEE.
- [98] Zhou, J., Cadavid, S., & Abdel-Mottaleb, M. (2011, June). A computationally efficient approach to 3d ear recognition employing local and holistic features. In CVPR 2011 WORKSHOPS (pp. 98-105). IEEE.
- [99] Cadavid, S., Mahoor, M. H., & Abdel-Mottaleb, M. (2009, November). Multi-modal biometric modeling and recognition of the human face and ear. In 2009 IEEE

- International Workshop on Safety, Security & Rescue Robotics (SSRR 2009) (pp. 1-6). IEEE.
- [100] Faltemier, T. C., Bowyer, K. W., & Flynn, P. J. (2008, September). Rotated profile signatures for robust 3d feature detection. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on* (pp. 1-7). IEEE.
- [101] Frejlichowski, D., & Tyszkiewicz, N. (2010, June). The west pomeranian university of technology ear database—a tool for testing biometric algorithms. In *International Conference Image Analysis and Recognition* (pp. 227-234). Springer Berlin Heidelberg.
- [102] Grgic, M., Delac, K., & Grgic, S. (2011). SCface—surveillance cameras face database. *Multimedia tools and applications*, 51(3), 863-879.
- [103] Al Nizami, H. A., Adkins-Hill, J. P., Zhang, Y., Sullins, J. R., McCullough, C., Canavan, S., & Yin, L. (2009, September). A biometric database with rotating head videos and hand-drawn face sketches. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on* (pp. 1-6). IEEE.
- [104] Raposo, R., Hoyle, E., Peixinho, A., & Proença, H. (2011, April). UBEAR: A dataset of ear images captured on-the-move in uncontrolled conditions. In *2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM)* (pp. 84-90). IEEE.

# **Chapitre V**

## **Reconnaissance de l'Oreille Humaine**

### **Techniques Proposées et Résultats**

## 5.1 Introduction

Après une présentation des différentes techniques de reconnaissance de l'oreille 2D, 3D et les différentes techniques de détection de l'oreille, dans le chapitre précédent, dans ce chapitre nous présentons notre approche de reconnaissance de l'oreille 2D. Les avantages et les inconvénients de chaque algorithme seront présentés en termes de taux de réussite et temps de calcul pour le processus de la classification de l'oreille. Le processus classique de classification/identification consiste à comparée la signature de l'identité recherchée avec toutes celles de la galerie. Cette tâche se trouve complexe et très couteuse, en temps de calcul, notamment avec de grandes bases de données biométriques. Notre système proposé a pour objectif de simplifier la complexité de calcul et d'améliorer la performance de classification.

Pour cela, nous nous sommes basé sur l'utilisation des caractéristiques locales, qui ont prouvé leur robustesse discriminative, par apport aux caractéristiques globales, en présence des différentes conditions d'acquisition tell que la rotation, la variation d'échelle, les différentes conditions d'éclairage etc... et ce dans différents domaines de classification. On a utilisé la théorie du Sparse Coding (SC) dans la phase de classification pour ranger chaque image de test dans sa propre classe. La raison d'arrière ce choix est que cette dernière est largement utilisée dans le domaine de la vision par ordinateur, à cause de sa performance en terme de précision et sa simplicité d'utilisation par rapport aux autres techniques de l'intelligence artificielle telles que les réseaux de neurones artificielles ANN et les machines à vecteur de support (SVM). En effet, ces techniques ont besoin de plusieurs échantillons dans la phase d'entraînement et beaucoup de temps afin de choisir le model optimal. A l'inverse, la SC peut travailler avec peut d'échantillon et avec un temps de calcul acceptable qui lui permettra d'être utilisé dans des applications en temps réel.

Les expériences ont été menées sur les oreilles de la base DEL1 et DEL2 dont les images sont prises dans des différentes conditions de rotation, d'éclairage... et dans des différentes sessions. Le choix principal de cette base de données est du à sa grande taille et sa popularité puisqu'elle est devenue une norme dans la communauté biométrique.

Un autre problème qu'on a essayé de résoudre, est le problème de l'occlusion qui est un problème majeur dans la biométrie en général et spécialement dans la reconnaissance de l'oreille. Pour cela on a ajouté différents niveaux d'occlusion aux images de test et évaluer notre approche en conséquence.

Dans les sections suivantes, nous allons d'abord essayer de décrire la configuration de notre système d'expériences. Nous examinerons tous les aspects de la configuration de ce système.

Les résultats expérimentaux montreront qu'une amélioration significative des taux de reconnaissance peut être obtenue en utilisant notre approche spécialement contre l'occlusion.

## **5.2 Les différentes étapes de notre approche d'identification**

L'architecture générale d'un système biométriques auriculaire à objectif de classification d'individus est illustrée dans la figure 5.1. Nous avons développé principalement les étapes de partitionnement et la combinaison des différentes caractéristiques locales, ainsi que l'utilisation de différents classificateurs. Ces combinaisons ont pour objectif de regrouper les données (les identités) qui ont des caractéristiques similaires et à séparer celles qui ont des caractéristiques dissemblables dans des classes différentes.

En générale, notre technique de reconnaissance d'oreille contient deux étapes principales: la première consiste à décrire l'oreille comme un vecteur caractéristique et la deuxième s'occupe de la comparaison de ces vecteurs.

### **5.2.1 Description de l'Oreille**

Avant l'étape d'extraction des caractéristiques locales ou globales, une série d'étapes de prétraitement est effectuée. Premièrement, toutes les oreilles sont détectées par la méthode proposée dans [1], dans laquelle le ROI peut être capturé et utilisé comme entrée pour l'étape d'extraction de caractéristiques. Deuxièmement, un prétraitement est également nécessaire avant d'effectuer l'extraction de caractéristiques. Il s'agit de la transformée en pyramide, qui nous permet d'analyser l'image à différentes échelles (résolutions) pour extraire le maximum d'information discriminative à partir d'une image donnée.

L'étape suivante consiste à construire un descripteur d'entité auriculaire pour représenter localement la surface de l'oreille. Cela devrait être fait avec un choix approprié des caractéristiques, qui sont très distinctives tout en restant invariant à d'autres changements, comme la pose, l'encombrement du fond et le bruit.

Dans la littérature, il existe une variété de caractéristiques qui ont montré des performances différentes dans différents scénarios. Certaines d'entre elles dépendent du contexte dans lequel elles sont appliquées, tandis que d'autres conservent un comportement de stabilité.

Le Local Binary Pattern (LBP), a été initialement introduit par Ojala et al. pour l'analyse de texture, Les principaux avantages de cet opérateur est son invariance à la rotation, sa robustesse face aux conditions d'éclairage et aussi son faible temps de calcul, ce qui est un avantage significatif par rapport aux autres caractéristiques.

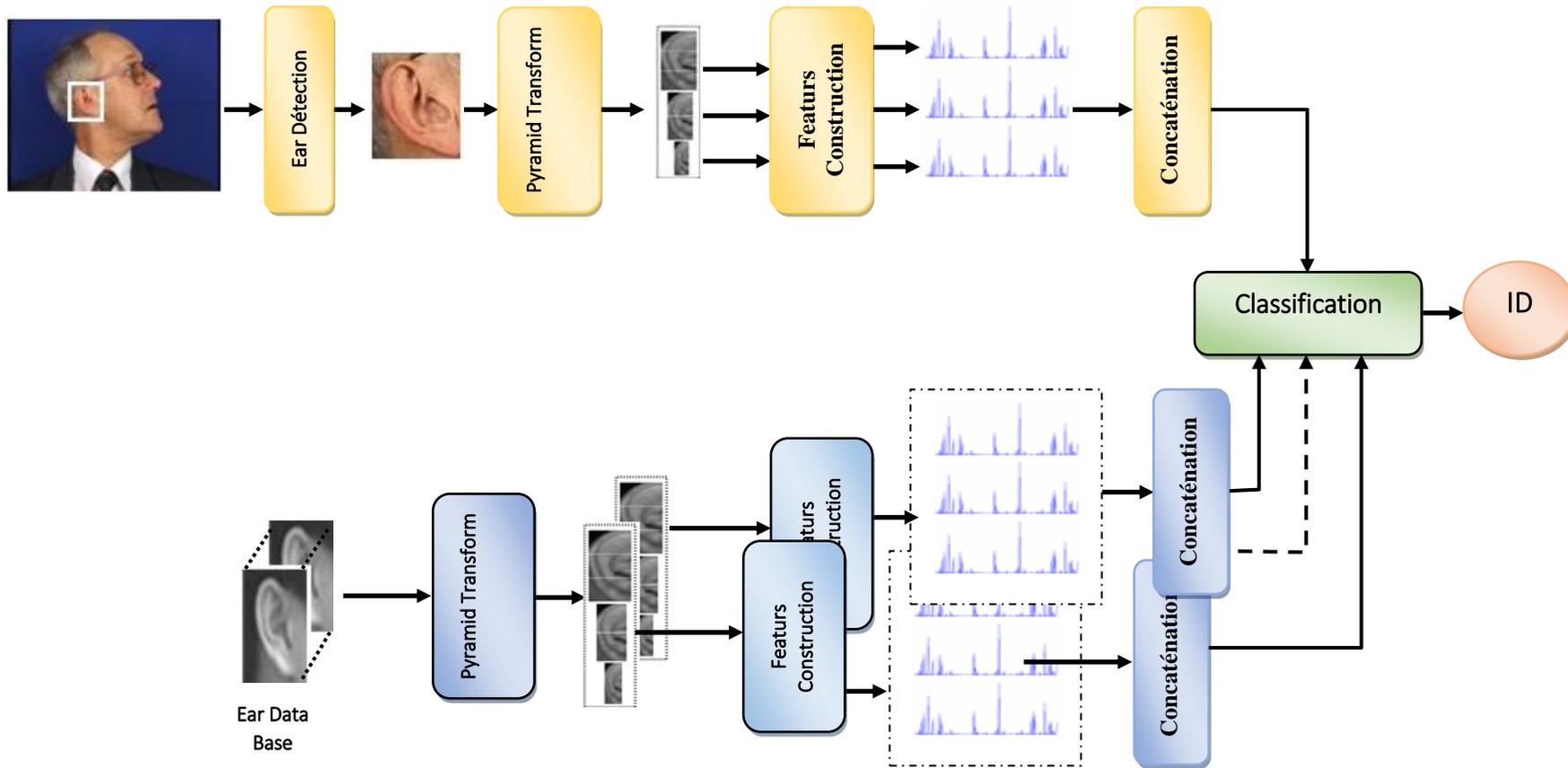


Figure 5.1: Les différentes étapes d'un processus de classification auriculaire.

Autre caractéristiques locales sont proposées telle que : Patterns of Oriented Edge Magnitudes (POEM) [7,8] et Co-Occurrence Local Binary Patterns (Co-LBP) [9], Histogram of Oriented Gradients (HOG) [10], Local Phase Quantization (LPQ) [11], Binarized Statistical Images Features (BSIF) [12].

Parmi ces caractéristiques, nous avons opté pour Patterns of Oriented Edge Magnitudes (POEM) [7,8] et Co-Occurrence Local Binary Patterns (Co-LBP) [9].

Ces deux descripteurs ont prouvé leur efficacité en termes de temps de calcul et taux de reconnaissance, ce qui nous a donc inspiré de les exploiter dans la configuration de la reconnaissance de l'oreille.

### 5.2.2 Comparaison des Oreilles

Une fois les caractéristiques représentant les images d'oreilles sont extraites, on procède à la classification des différentes images de tests. Beaucoup de techniques de classification sont présentes dans la littérature. Parmi ces techniques on s'est intéressé à la théorie Sparse Coding, qui a ouvert une large voie dans la théorie de l'information et traitement du signal. La puissance de cette technique est sa capacité à reconstruire un signal donné à partir d'un dictionnaire de taille limitée. En conséquence, elle a également été adoptée en vision par ordinateur et s'est révélée utile dans différents domaines [13-16].

Le principe de la SC est de récupérer un signal inconnu à partir d'une bibliothèque (dictionnaire) de nombreux échantillons (Atomes). Le Processus de la récupération est habituellement évalué en effectuant les résidus. En d'autres termes, la déduction du signal récupéré, à partir du signal d'entrée, indique dans quelle mesure ce dernier est reconstruit. Par la suite plusieurs travaux ont fait appel à la SC comme outil de classification. Un bon exemple est le travail présenté dans [14], où les auteurs exploitent la SC pour la reconnaissance faciale. Le travail a inspiré de nombreuses tentatives de reconnaissance basée sur cette technique. En partant de [14], on croit que le travail présenté en [13] améliore la performance de reconnaissance faciale en étendant les dictionnaires adoptés avec des échantillons virtuels dits aléatoires, qui sont générés sur la base des images de visage originales.

À cet égard, nous proposons une nouvelle technique de décision basée sur la SC pour la reconnaissance de l'oreille. En général, les systèmes de reconnaissance de l'oreille sont plus précis sur les images qui sont collectées avec une coopération des sujets. C'est donc un problème difficile dans les lieux publics ou dans un environnement non contrôlé (comme dans la vidéo surveillance, qui peut contenir des motifs partiels ou corrompus, Anneaux d'oreille, etc.). Dans ce cas, la performance de reconnaissance se dégrade gravement. Nous tentons dans notre travail de développer une méthode de reconnaissance de l'oreille simple mais efficace

qui peut préserver une meilleure capacité de reconnaissance mêmes dans les cas des occlusions partielles.

### 5.3 Bases de données utilisées

Dans cette section, nous étudions les performances de nos approches pour la reconnaissance de l'oreille dans différentes conditions. Des expériences approfondies sont menées sur deux bases de données, à savoir IIT Delhi-1 et IIT Delhi-2 pour valider la performance de la méthode proposée.

**5.3.1 Database1 (DB1):** Fait référence à la base de données IIT Delhi-1 qui contient 465 images à partir de 125 sujets (au moins 3 images par sujet). Les images ont été acquises sur une période de 9 mois en milieu intérieur. Tous les sujets de la base de données sont dans l'intervalle d'âge de 14-58 ans. Les images de cette base sont caractérisées par un changement d'échelle significative, la translation et des changements de rotation [1].

**5.3.2 Database2 (DB2):** se réfère à la deuxième version IIT Delhi-2, qui contient 793 images de 221 sujets différents et comprend {0,3, -3,6, -6,9, -9} degrés de rotation d'images [1].

Afin d'évaluer la performance du projet WSRC, nous rapportons les résultats expérimentaux du protocole d'essai A [1], (par rapport à la littérature), qui génère une moyenne de trois tests où chacune des trois premières images de chaque sujet sont utilisées comme des images de test. Les images restantes de chaque sujet sont utilisées comme images de formation.

### 5.4 Méthodes Proposées

À partir de l'état de l'art, il est constaté que le problème majeur réside dans le choix du type de caractéristiques ainsi que le classificateur utilisé. Comme les approches locales donnent des informations sur les parties locales de l'image de l'oreille, ils sont considérés comme appropriés pour l'extraction de caractéristiques bien plus que les approches globales. Les caractéristiques locales mettent en évidence des distinctions individuelles mieux que les techniques globales. Le travail proposé repose donc sur des méthodes locales d'extraction de caractéristiques ainsi que la communication entre eux et les différentes méthodes de classification pour atteindre la meilleure performance possible.



**Figure 5 .2:** Exemple d'images de quatre personnes de la base de données **IIT Delhi-2**.

#### 5.4.1 Extraction des caractéristiques

Une fois que l'image de l'oreille est prétraitée, la prochaine étape est l'extraction de caractéristiques (voir la figure 5.1). Cela implique de caractériser la zone de l'oreille par un vecteur qui se réfère à la représentation / descripteur de l'oreille. Une bonne représentation est celle qui minimise les dis similarités intra personnelles tout en élargissant la marge entre les personnes différentes. C'est une question critique, car les variations de poses et d'échelles peuvent être plus grandes que les variations d'identité. Pour les systèmes de reconnaissance réels, tels que les applications de surveillance, une bonne représentation devrait aussi être à la fois rapide et compacte: si on teste une image contre une base de données très grande, l'extraction et le stockage de la représentation de l'oreille doit être suffisamment rapide pour que les résultats puissent être remis à l'utilisateur final en temps utile. Malheureusement, tous les algorithmes de description actuels équilibrent ces critères: les caractéristiques qui peuvent produire des résultats de reconnaissance de haute qualité sont très intensives en termes d'exigences de stockage et de temps de calcul, tandis que les algorithmes de faible complexité ne sont pas suffisamment fiables. Il s'agit à la fois d'un défi et d'une motivation à trouver un descripteur satisfaisant aux trois critères, à savoir le caractère distinctif, la robustesse et le coût de calcul peu coûteux. Parmi les différentes caractéristiques locales, nous proposons

d'utiliser deux caractéristiques récentes, qui ne sont pas encore utilisées dans le domaine de la reconnaissance auriculaire, appelées Patterns of Oriented Edge Magnitudes (**POEM**) et Co-Occurrence Local Binary Patterns (**Co-LBP**). Notre motivation derrière le choix de ces deux caractéristiques est qu'elles sont très informatives et très acceptable en termes de stockage et temps de calcul.

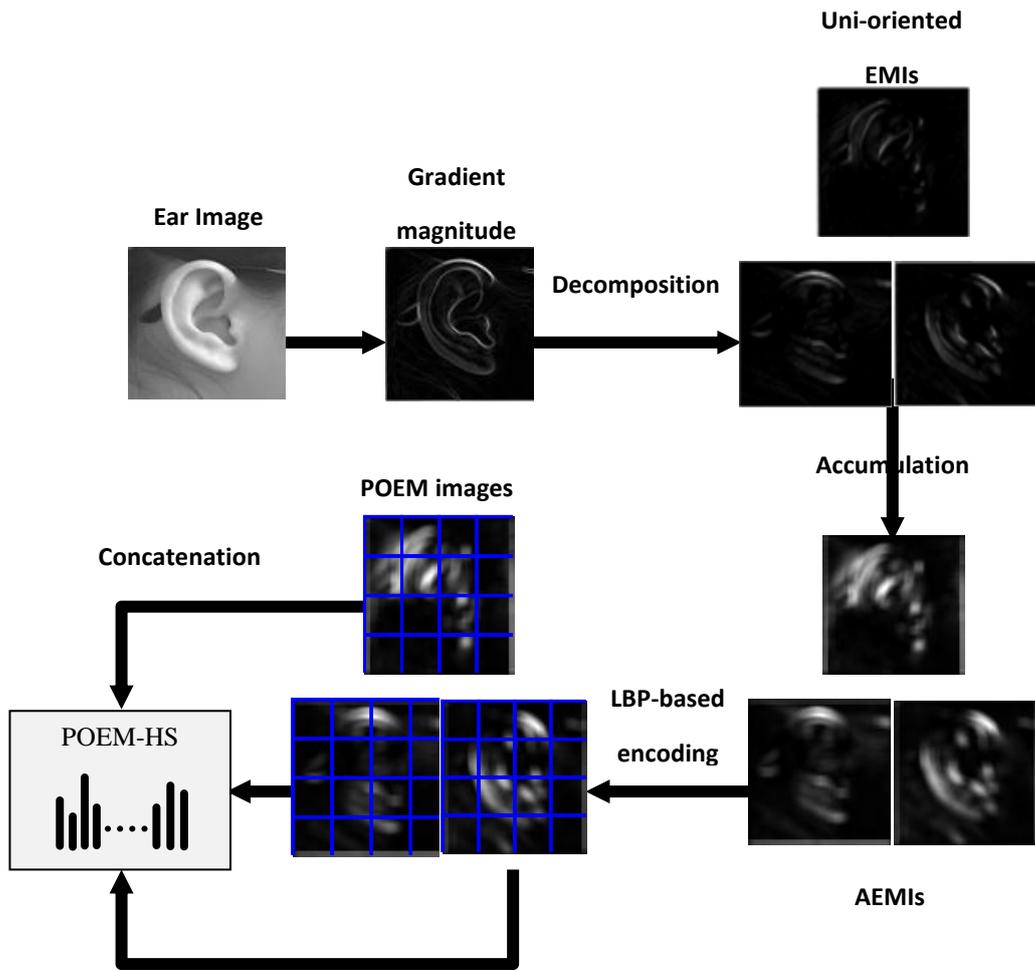
#### 5.4.1.A Patterns of Oriented Edge Magnitudes (POEM)

Dans leur travail [7,8] Ngoc Son et al, ont introduit le descripteur POEM pour résoudre le problème de la reconnaissance faciale. Evalué sur différentes base des données populaires, le descripteur POEM a prouvé son efficacité en termes du taux de reconnaissance et du temps de calcul. En outre, ils ont obtenu des résultats comparables à ceux des anciens travaux. Le succès de POEM nous a donc inspiré de l'adapter à la configuration de reconnaissance d'oreilles (voir figure 5.3). L'idée principale du descripteur POEM est que l'apparence d'un objet peut être bien représentée en combinant des informations de bord et la forme local. Ceci est bien conforme au problème de reconnaissance de l'oreille en capturant des informations utiles, car la structure auriculaire est composée de lignes courbées entourant une zone concave centrale. Ainsi, cette technique tire des avantages conjointement de l'histogramme bien connu des gradients orientés (HOG) de Dalal et al. [10] et la LBP [9]. Nous rappelons ici les étapes clés de l'extraction du descripteur POEM à partir d'une image donnée. Pour une perception plus approfondie du POEM, le lecteur est référé à [8].

Pour construire un descripteur POEM, les valeurs d'intensité dans le calcul de LBP classique sont remplacées par le gradient de l'amplitude, calculé en accumulant un histogramme local des directions du gradient sur tous les pixels d'un patch spatial (cellule).

La génération du descripteur POEM commence par le calcul de l'image d'amplitude du contour orientée (Edge Magnitude Image : EMI), où l'orientation du gradient à chaque pixel est répartie soit sur  $0-\pi$  (discrétisation non signée), ou  $0-2\pi$  (discrétisation signé).

La deuxième étape consiste à incorporer des informations du gradient à partir des pixels voisins en calculant un histogramme local des orientations du gradient sur tous les pixels d'une cellule. Les auteurs ont utilisé des fenêtres pondérées telle que le filtre Gaussien ou des fonctions binomiales. D'après les résultats, la technique de l'image intégrale, utilisée dans l'algorithme de Viola et Jones [17], donne de bons résultats par rapport aux fenêtres pondérées.



**Figure 5.3:** Processus d'extraction du POEM

Finallement, nous encodons les grandeurs accumulées en utilisant l'opérateur LBP dans un bloc (figure (5.4)). Rappelons que l'opérateur LBP d'origine étiquette les pixels d'une image par seuillage au voisinage 3x3 autour du pixel, avec la valeur d'intensité du pixel central et en considérant la séquence de 8 bits résultants comme un nombre. Seuls les modèles uniformes sont généralement considérés pour réduire le nombre de motifs et accélérer la méthode (en utilisant 8 bits binaires pour coder les résultats LBP donne 256 LBP, parmi ceux-ci seulement 58 LBP sont uniformes).

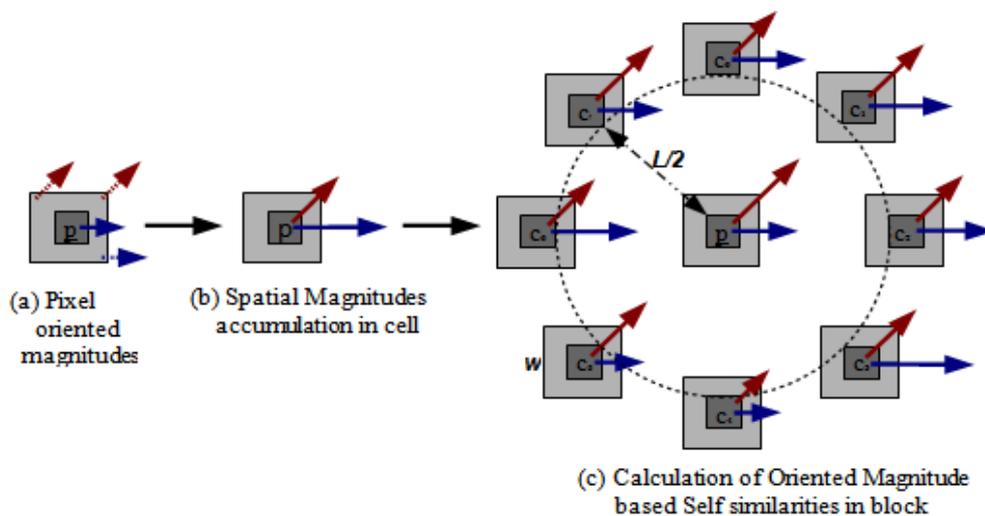
Nous appliquons ce calcul sur les amplitudes du gradient accumulé et sur différentes directions pour construire les caractéristiques du POEM. Tout d'abord, à la position du pixel  $p$ , une caractéristique POEM est calculée pour chaque direction discrétisée  $\theta_i$  comme suit:

$$POEM_{L,w,n}^{\theta_i}(p) = \sum_{j=1}^n f\left(s\left(m_p^{\theta_i}, m_{c_i}^{\theta_i}\right)\right) 2^j \quad (5.1)$$

Où  $m_p$  et  $m_{c_i}$  sont les magnitudes du gradient accumulées des pixels centraux et environnants le pixel  $p$ ,  $c_i$  respectivement;  $f(\cdot)$  est la fonction de similarité (par exemple la différence de deux grandeurs du gradient);  $L$ ,  $w$  se rapportent respectivement à la taille des blocs et des cellules;  $n$  est le nombre de pixels entourant le pixel considéré  $p$ .

Le descripteur POEM final est obtenu en concaténant tous les POEM unidirectionnels à différentes orientations.

$$POEM_{L,w,n}(p) = \{POEM^{\theta_1}, \dots, POEM^{\theta_k}\} \quad (5.2)$$



**Figure 5.4:** Principales étapes de l'extraction du POEM.

#### 5.4.1.B Co-Occurrence Local Binary Patterns (Co-LBP)

Afin de conserver plus d'informations sur la relation spatiale des modèles LBP, Nosaka et al. [9] ont proposé une nouvelle alternative à ce descripteur qu'ils ont appelée Co-occurrence du motif local binaire (Co-LBP). Ce descripteur étudie la Co-Occurrence de LBPs sur une image donnée et considère leur information spatiale. D'après leurs expériences; ils ont montré que la méthode proposée Co-LBP donne des performances plus élevées par rapport à la LBP traditionnelle en particulier pour la reconnaissance des visages et des textures.

Puisque, la relation spatiale entre les différents LBPs est manquante au cours du processus de génération de l'histogramme LBP, car les LBPs sont forcés dans un seul histogramme, il en résulte une perte d'informations dans l'image considéré. Ceci suggère qu'il y a encore une place pour une amélioration de la performance des caractéristiques basées sur LBP.

La figure 5.5 montre la différence entre l'histogramme LBP et l'histogramme de la Co-LBP. Les trois exemples d'image sont composés de trois motifs LBP **A** et trois motifs LBP **B**,

comme représenté sur la figure 5.5 (a) de cette figure. Comme le nombre de motifs LBP A et LBP B dans chaque image est le même, les histogrammes LBP sont générés à partir des trois images qui coïncident les une avec les autres, comme représenté sur la figure 5.5 (b).

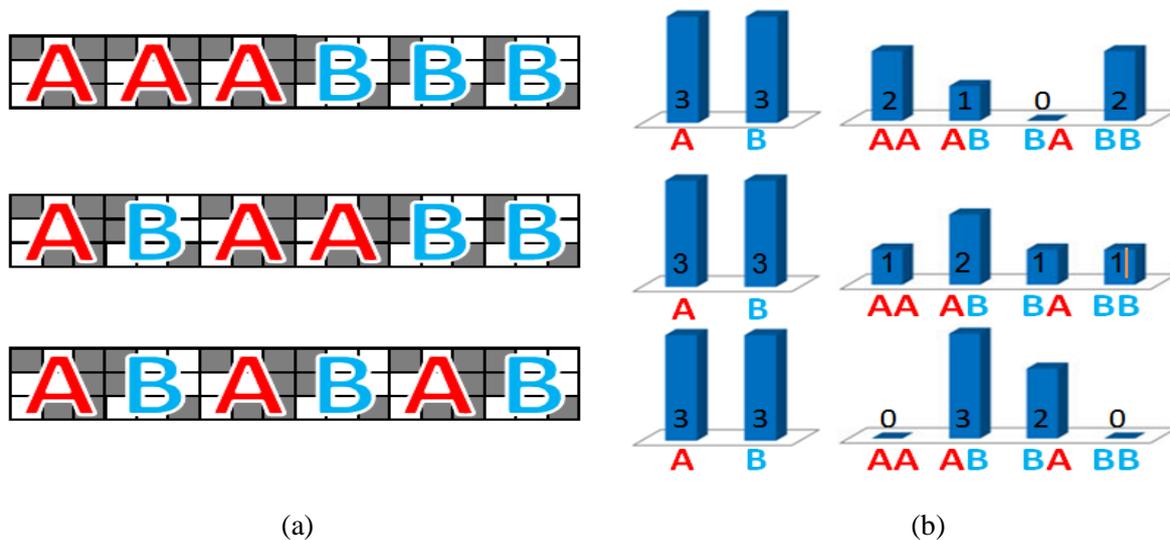


Figure.5.5 : Différences entre l'histogramme LBP et l'histogramme de Co-LBP LBP [9].

En revanche, les histogrammes de la Co-Occurrence spatiale extraits de chaque image sont très différents, comme le montre la figure 5.5.

A partir de cet exemple simple, nous pouvons constater que la capacité discriminative du descripteur LBP original est insuffisante et que la Co-Occurrence spatiale de LBP est une variante valide pour réaliser une capacité de représentation plus élevée.

Le LBP original indique une relation d'amplitude entre un pixel central et ses pixels voisins (figure 5.6). Le code LBP '1000111' est obtenu par seuillage entre le pixel central (5) et son voisinage 3×3. Le motif binaire est alors converti en son équivalent décimal, 135. Des histogrammes LBP sont générés à partir des valeurs décimales de tous les ensembles LBPs.

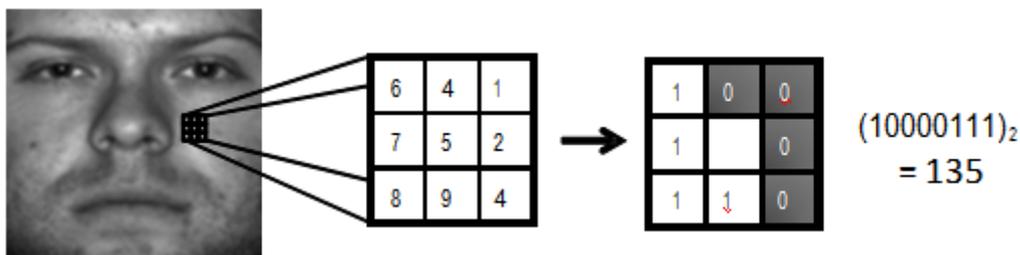


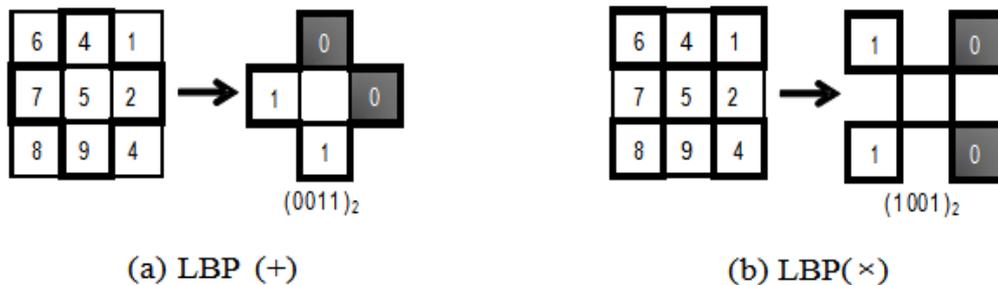
Figure 5.6: LBP Original.

Soit  $I$  une intensité d'image et  $r = (x, y)^T$  un vecteur de position dans  $I$ . Le LBP  $b(r) (\in R^N)$  est défini comme suit:

$$b(r) = \begin{cases} 1, & \text{si } I(r) > I(r + \Delta s_i) \\ 0, & \text{sinon} \end{cases}, \quad (i = 1, \dots, N_n) \quad (5.3)$$

Où:  $N_n$  est le nombre de pixels voisins et  $\Delta s_i$  sont les vecteurs de déplacement de la position du pixel central  $r$  aux pixels voisins. Dans la LBP originale, ces paramètres sont définis comme suit :  $N_n = 8$  ,  $\Delta s_i = \{(\pm\Delta s, \pm\Delta s)^T, (\pm\Delta s, 0)^T, (0, \pm\Delta s)^T\}$  ,  $\Delta s = 1$ . Ensuite, le LBP  $b(r)$  est converti en nombre décimal. Enfin, l'histogramme des LBP est généré en considérant les décimales comme des étiquettes.

Le Co-LBP est défini comme un indice de la fréquence des différentes combinaisons de LBP qui se produit dans l'image entière. Le nombre de combinaisons possibles de LBP est significativement supérieur à celui des LBP originales. Il est donc difficile d'utiliser un programme basé sur des règles pour calculer la Co-Occurrence de toutes les combinaisons quand il existe de nombreux types de LBP. Nosaka et al [9], ont introduit une matrice d'auto corrélation comme méthode efficace de calcul de la co-occurrence des LBP. Tout d'abord, bien que le LBP d'origine utilise huit pixels voisins d'un pixel central donné, ils ont modifié la configuration LBP pour considérer deux configurations, réduisant ainsi le coût du calcul. Une configuration est le LBP (+), qui ne considère que deux pixels horizontaux et deux pixels verticaux, comme le montre la figure 5.7 (a). L'autre configuration est le LBP ( $\times$ ), qui considère les quatre pixels diagonaux représentés sur la figure 5.7 (b). Dans le LBP (+), les paramètres sont définis comme suit:  $N_n = 4$ , si  $\in \{(\pm \Delta s, 0)^T, (0, \pm \Delta s)^T\}$ . Dans le LBP ( $\times$ ), les paramètres sont définis comme suit:  $N_n = 4$ , si  $\in \{(\pm \Delta s, \pm \Delta s)^T, (\pm \Delta s, \mp \Delta s)^T\}$ .



**Figure 5.7:** Les combinaisons LBP utilisées

Ensuite, pour calculer efficacement le Co-LBP, chaque LBP est converti en vecteur  $f \in \mathbb{R}^{Np}$ , défini par :

$$f_i(r) = \delta_{i,I(b(r))} \quad (5.4)$$

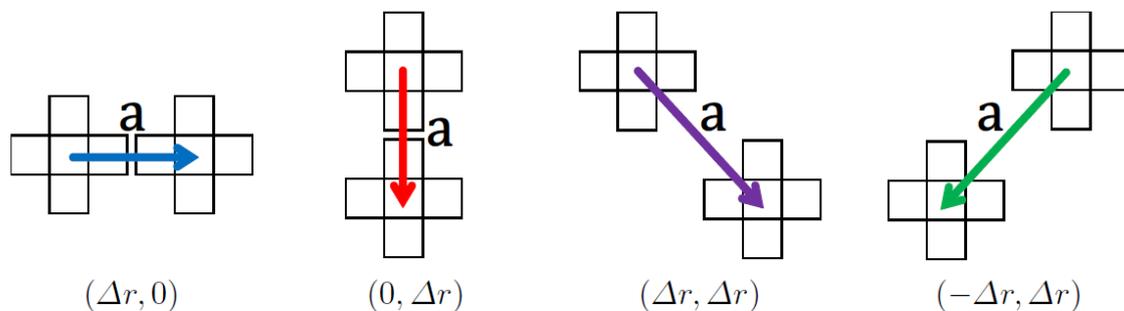
Où  $N_p$  est le nombre de toutes les combinaisons possibles  $N_p = 2N_n$  par exemple. Dans l'LBP original,  $N_p = 2^8 = 256$ ,  $N_n$  est le nombre des pixels voisins,  $\delta_{i,l}$  représente la fonction de Kronecher's delta et  $I(b(r))$  est l'indice de  $b(r)$ .

Pour calculer le CoLBP, Nosaka et al. [9] ont utilisé une matrice d'auto-corrélation de  $N_p \times N_p$  définie par l'équation suivante :

$$H(a) = \sum_{r \in I} f(r) f(r + a)^T \quad (5.5)$$

Où  $a$  est le vecteur de déplacement de la référence LBP à son voisin LBP. L'élément  $H_{i,j}(a)$  de l'équation (5.5) indique le nombre de paires de LBP adjacentes  $LBP_{i,j}$ .

$a$  est défini comme suit:  $\{(\Delta r, 0)^T, (\Delta r, \Delta r)^T, (0, \Delta r)^T, (-\Delta r, \Delta r)^T\}$ . La figure 5.8 montre toutes les configurations de  $r$  et  $r + 1$  dans l'équation (5.5).



**Figure 5.8:** Les configurations des combinaisons LBP utilisées [9].

La figure 5.8 montre le processus d'extraction de la caractéristique proposée dans [9] à partir d'une image. L'image de l'exemple a quatre LBPs (Fig.5.9). Les étiquettes de ces LBP sont respectivement 2, 8, 9 et 14. Dans le cas où le vecteur de déplacement  $a = (\Delta r, 0)^T$ , il y a deux paires LBP ({En haut à gauche, en haut à droite} et {En bas à gauche, en bas à droite} dans l'image. Comme les étiquettes sont (2, 14) et (8, 9), les éléments correspondant à ces étiquettes dans l'équation (5.4) sont mis à 1 et les autres éléments sont mis à 0. Pour les autres vecteurs de déplacement:  $a = (0, \Delta r)^T$ ,  $(\Delta r, \Delta r)^T$  et  $(-\Delta r, \Delta r)^T$ , une matrice d'auto-corrélation  $H(a)$  est générée de la même façon que celle représentée sur la figure 5.10.

La figure 5.10 montre le flux du processus de la caractéristique proposée. Tout d'abord, les LBP sont extraits de l'image d'entrée comme représenté sur la figure 5.10 (a). Ensuite, nous calculons quatre matrices d'auto corrélation  $N_p \times N_p$  des Co-Occurrences spatiales des LBPs adjacentes,  $H(a)$ , comme le montre la Figure 4.10 (b). Enfin, ces matrices sont vectorisées et combinées à un vecteur de caractéristique de dimension  $2N_p^2$  (Figure 5.10 (c)).

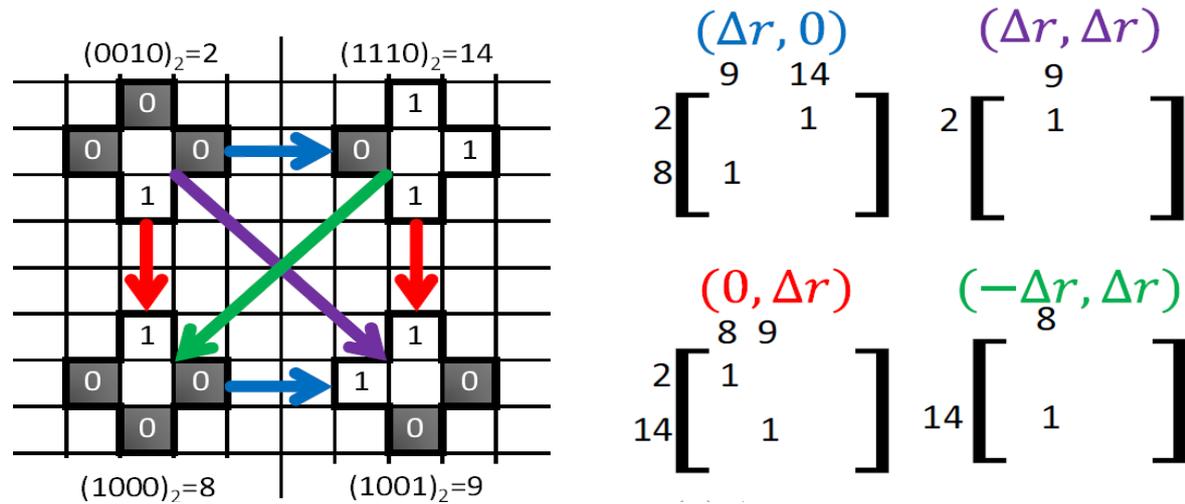


Figure 5.9 : Exemple d'extraction de la caractéristique proposée [9].

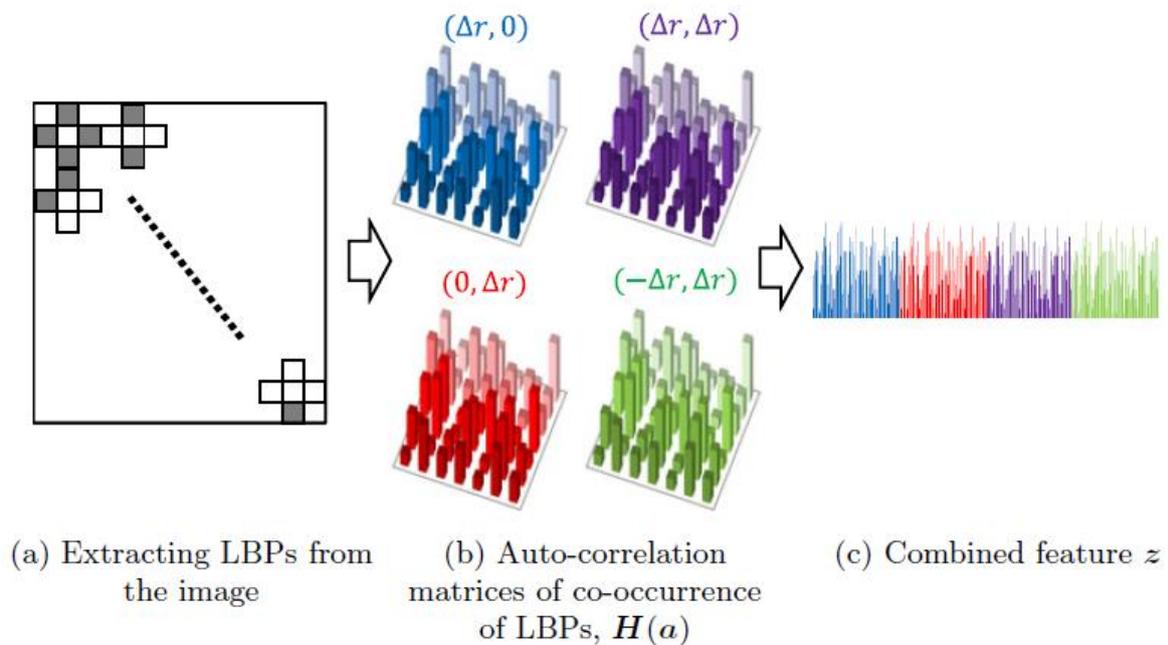


Figure 5.10: Exemple d'extraction de la caractéristique proposée [9].

### 5.4.2 Théorie du Sparse Coding

L'idée générale de la théorie SC est de reconstruire (à travers une combinaison linéaire) un signal inconnu  $H$  donné (dans notre cas, il représente l'histogramme POEM ou Co-LBP), à partir d'un dictionnaire bien établi  $D$  (des descripteurs POEM ou Co-LBP). Ce processus se traduit par la production d'une suite de coefficients  $\alpha$  constituée principalement de valeurs proches ou égales à zéro, calculées par l'utilisation de  $L_0$  minimisation, selon l'équation (5.6):

$$\min \|\alpha\|_0 \text{ subject to } H = D \cdot \alpha \tag{5.6}$$

La littérature propose plusieurs techniques pour résoudre le problème d'optimisation exprimé dans l'équation (5.6), telles que Matching Pursuit (MP), Orthogonal Matching Pursuit (OMP) et Basis Pursuit (BP) [18]. Une autre alternative est le Stagewise Orthogonal Matching Pursuit (St-OMP) [19]. C'est celle-ci qui est utilisée dans notre travail dû à ses performances en terme de temps de calcul et de précision. Les étapes principales de l'algorithme St-OMP sont :

**Etape 1 :**

Considérons une solution initiale  $\alpha_0 = 0$ , un résidu initial  $r_0 = V$ , un compteur d'étage  $s$ , initialisé à 1, et une séquence d'index notée  $T_1, \dots, T_s$  qui contient les emplacements des non-zéros dans  $\alpha_0$ .

**Etape 2 :**

Calculer le produit entre le résiduel actuel et le dictionnaire considéré.

$$\mathbf{C}_s = \mathbf{D}^t \cdot r_{s-1} \quad (5.7)$$

**Etape 3 :**

Effectuer un seuillage afin de trouver les non-zéros significatifs dans  $\mathbf{C}_s$  en recherchant les emplacements correspondant aux grandes coordonnées  $J_s$  :

$$j_s = \{j : \mathbf{C}_s(j) > t_s \sigma_s\} \quad (5.8)$$

Où  $\sigma_s$  représente un niveau de bruit et  $t_s$  est un paramètre de seuil prenant des valeurs dans la gamme  $2 \leq t_s \leq 3$ .

**Etape 4 :**

Fusionner les coordonnées sélectionnées avec le support précédent :

$$T_s = T_{s-1} \cup j_s \quad (5.9)$$

**Etape 5 :**

Projeter le vecteur  $V$  sur les colonnes de  $D$  qui correspondent aux  $T_s$  précédent. Cela donne une nouvelle approximation:

$$(\alpha_s)_{T_s} = (\mathbf{D}_{T_s}^t \mathbf{D}_{T_s})^{-1} \mathbf{D}_{T_s}^t V \quad (5.10)$$

**Etape 6:**

Calculer le résidu selon:

$$r_s = V - \mathbf{D} \cdot \alpha_s \quad (5.10)$$

**Etape 7:**

Vérifier si la condition itérative d'arrêt (par exemple,  $s_{\max} = 10$ ) est satisfaite. Si c'est le cas,  $\alpha_s$  est considéré comme la solution finale et comme le résidu laissé de côté du processus de récupération. Sinon, le compteur est incrémenté et le processus est répété à partir de l'étape 2. En terme général, une prise de décision basée sur CS peut être faite par l'une des deux façons. La première consiste à opter pour la valeur résiduelle (somme des éléments du vecteur résiduel) comme mécanisme. Dans ce cas, plus le résidu d'une certaine classe du dictionnaire est petit, plus il est probable que cette classe soit proche du modèle d'essai V. La deuxième stratégie repose sur les coefficients de reconstruction  $\alpha_s$ . Ici le coefficient le plus élevé d'un certain atome, à partir du dictionnaire, traduit sa ressemblance avec le modèle de test.

### 5.4.3 Stratégies de Reconnaissance par la Méthode Sparse Coding

Jusqu'à présent, nous avons expliqué que la sortie de l'algorithme sont des vecteurs caractéristiques locaux (POEM, Co-LBP) résultant de différentes échelles de l'image. La façon dont les histogrammes sont comparés fait une différence essentielle dans notre travail. La théorie SC, en dépit de sa simplicité conceptuelle, s'est révélée efficace lorsqu'il s'agit de problèmes de reconnaissance. En fait, la SC a montré un grand potentiel dans notre travail, à cet égard, nous expliquons ci-dessous les deux stratégies différentes basées sur la SC pour la classification. La première est une stratégie basée sur les résidus (RDMS) tandis que la deuxième est une stratégie basée sur les coefficients (CDMS).

#### ● Décision résiduelle

Considérons le problème de la reconnaissance de l'oreille dans un scénario de N individus (classes). L'idée sous-jacente du RDMS est de reconstruire une image test d'une d'oreille donnée sur chaque classe parmi N classes et de décider celle qui est la plus proche en fonction de ses résidus restants (non reconstruits). Plus en détail, les caractéristiques de chaque classe sont regroupées (colonne) afin de développer un dictionnaire de la galerie de cette classe individuelle. Ainsi, N dictionnaires sont envisagés. Maintenant, étant donné une image inconnue, le RDMS procède en la reconstruisant sur chaque dictionnaire unique selon l'équation 5.9. Ensuite, en considérant le résidu, selon l'équation 5.10, la classe qui donne le plus faible résidu est considéré comme la plus proche de la l'image de test. La représentation RDMS est donnée à la Fig. 5.11.

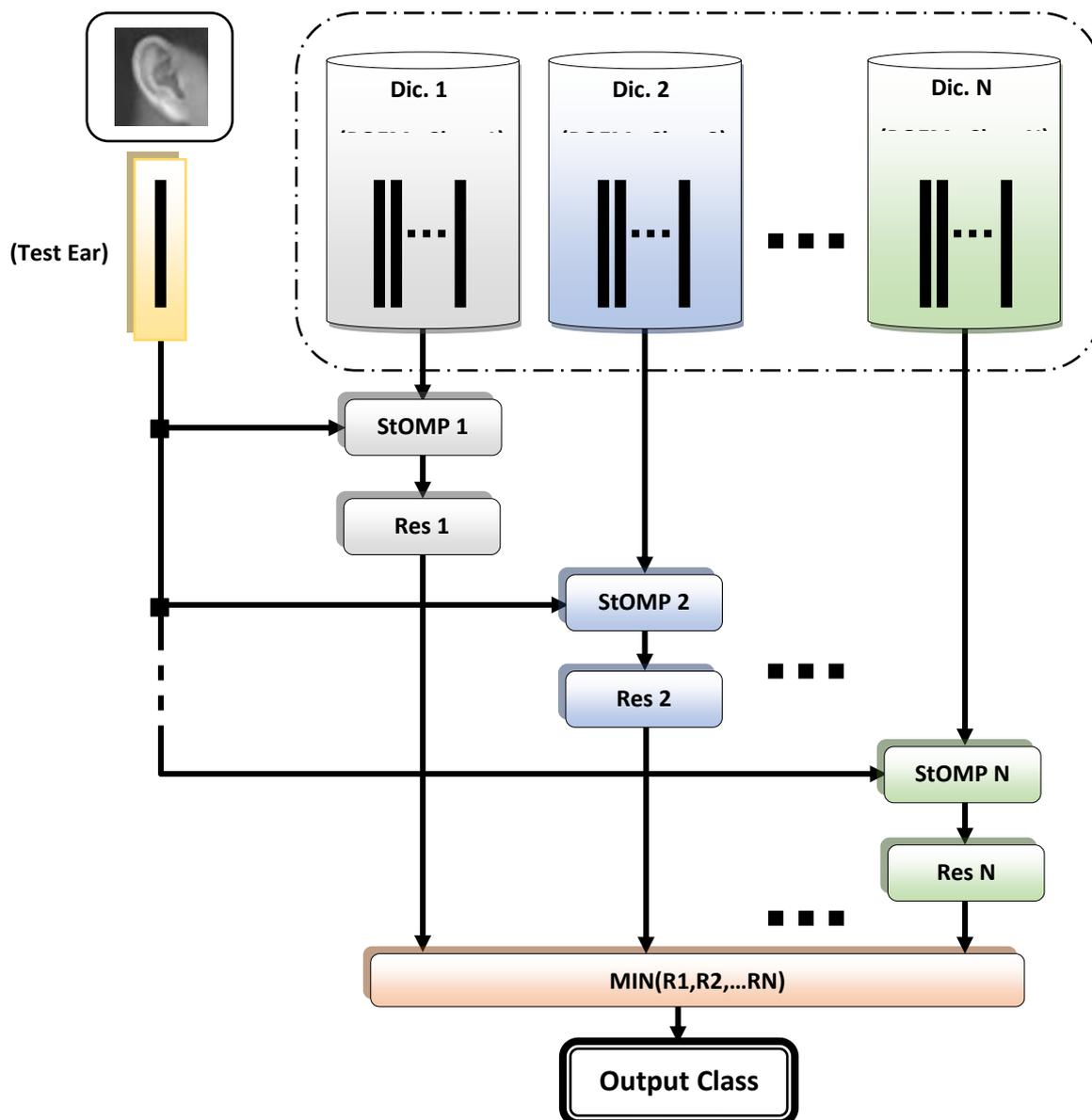


Figure 5.11: Illustration de la stratégie de décision résiduelle adoptée.

- **Décision basée sur les coefficients**

Contrairement au RDMS, qui est basé sur l'analyse des résidus, la décision CDMS est basée sur la récupération SC elle-même. Toujours en considérant le problème des  $N$  classes, les histogrammes de toutes les classes sont maintenant regroupés dans un seul dictionnaire (non séparés comme dans RDMS). Après avoir effectué la reconstruction de l'image de test, on génère autant de coefficients que le nombre de vecteurs colonnes du dictionnaire selon l'équation 5.12. Ensuite, la classe est choisie selon l'idée que plus le coefficient est élevé dans la séquence de coefficients SC obtenue  $\alpha_s$ , plus sa galerie correspond à celle du test. La Figure. 5.12 résume ce procédé.

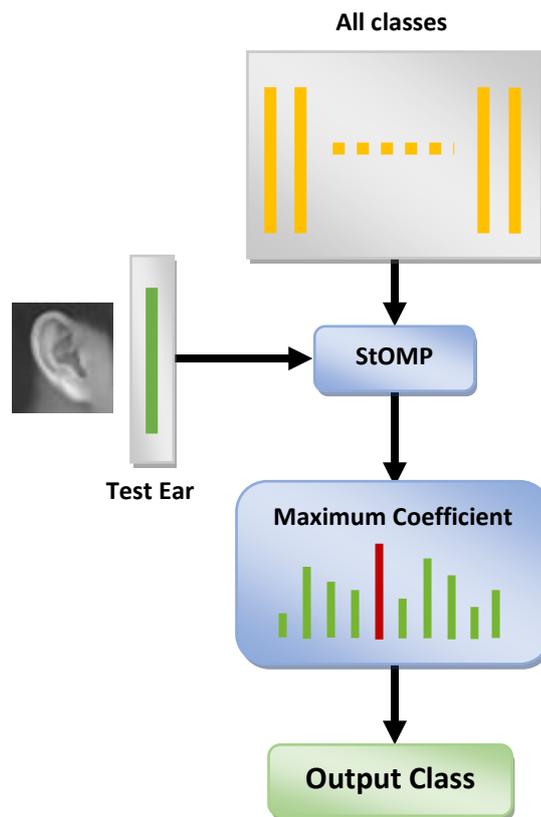


Figure.5.12: Illustration de la stratégie des coefficients adoptée

## 5.5 Méthodes proposée pour la reconnaissance de l'oreille humaine

Dans cette partie on va expliquer en détail les approches proposées, pour la reconnaissance de l'oreille. La mie en œuvre est faite sur des images d'oreilles 2-D.

Deux techniques sont proposée la première est basé sur deux descripteur locaux (POEM , Co-LBP) et deux différents classificateurs: K-NN et la méthode du Sparse Coding (SC) appelé (Weighted Sparse Representation) [20]. La deuxième techniques est basée sur un seul descripteur Co-LBP et une nouvelle technique de classification basée sur SC [21]. Ci-dessous est donné le détail de chaque technique.

### 5.5.1 Description de la première méthode

Cette méthode contient deux étages principaux comme illustré à la figure 5.13.

- **Etape 1 :**

Elle se résume aux points suivants:

- L'image de l'oreille est projetée dans un domaine multi-résolution à l'aide d'une transformation pyramidale. Cela permet une analyse approfondie en recueillant plus

d'informations et de détails sur la répartition des informations locales et à différentes résolutions. En fait, la représentation multi-résolution est basée sur une transformation pyramidale gaussienne. Nous procédons à un échantillonnage descendant de l'image de l'oreille pour obtenir 3 images avec les rapports d'échantillonnage suivants (1, donc la taille de l'original est conservée, 0,5 et 0,25). Ensuite, la première image (de rapport 1) est divisée en trois segments égaux. La seconde image (de rapport 0,5) est divisée en deux, tandis que la troisième image (de rapport 0,25) est retenue telle quelle. Ainsi, six images secondaires (segments) sont générées.

- Par la suite, le descripteur POEM (Pattern of Oriented Edge Magnitude) est appliqué aux segments préalablement préparés. Le descripteur POEM est adopté en raison de ses caractéristiques satisfaisantes, telles que (i) la capture des changements d'image dans n'importe quelle direction, (ii) un temps de calcul très rapide. Les descripteurs POEM ont été extraits de chaque segment (6 vecteurs caractéristiques).
- Les POEM de tous les segments de chaque image sont concaténés de sorte qu'ils constituent un vecteur caractéristique pour cette image (figure 5.13). Il s'agit d'un descripteur POEM à plusieurs échelles. Par la suite, les distances, entre l'image de test et les «N» images de la base de données, sont calculées avec les K-NN à base de la distance de Chi-square. Il en résulte un vecteur de distances:  $d = [d_1, d_2, \dots, d_N]$ .
- Ensuite un sous-ensemble (S), composés des distances les plus faibles, est constitué. Ce sont simplement les «S» premières distances les plus petites qui sont conservés parmi les «N» distances. Ces distances sont par la suite injectées dans la deuxième étape pour une contribution dans la classification finale.

## • Etape 2

Cette étape traite la représentation et la classification des classes des sous-ensembles sélectionnés (S) provenant de la première étape. Ici, le descripteur Co-LBP a été adopté pour sa capacité à fournir plus de détails sur la répartition des microstructures et aussi robuste à la variation d'illumination que le descripteur LBP traditionnel. Les points qui reflètent cette étape peuvent être résumés comme suit:

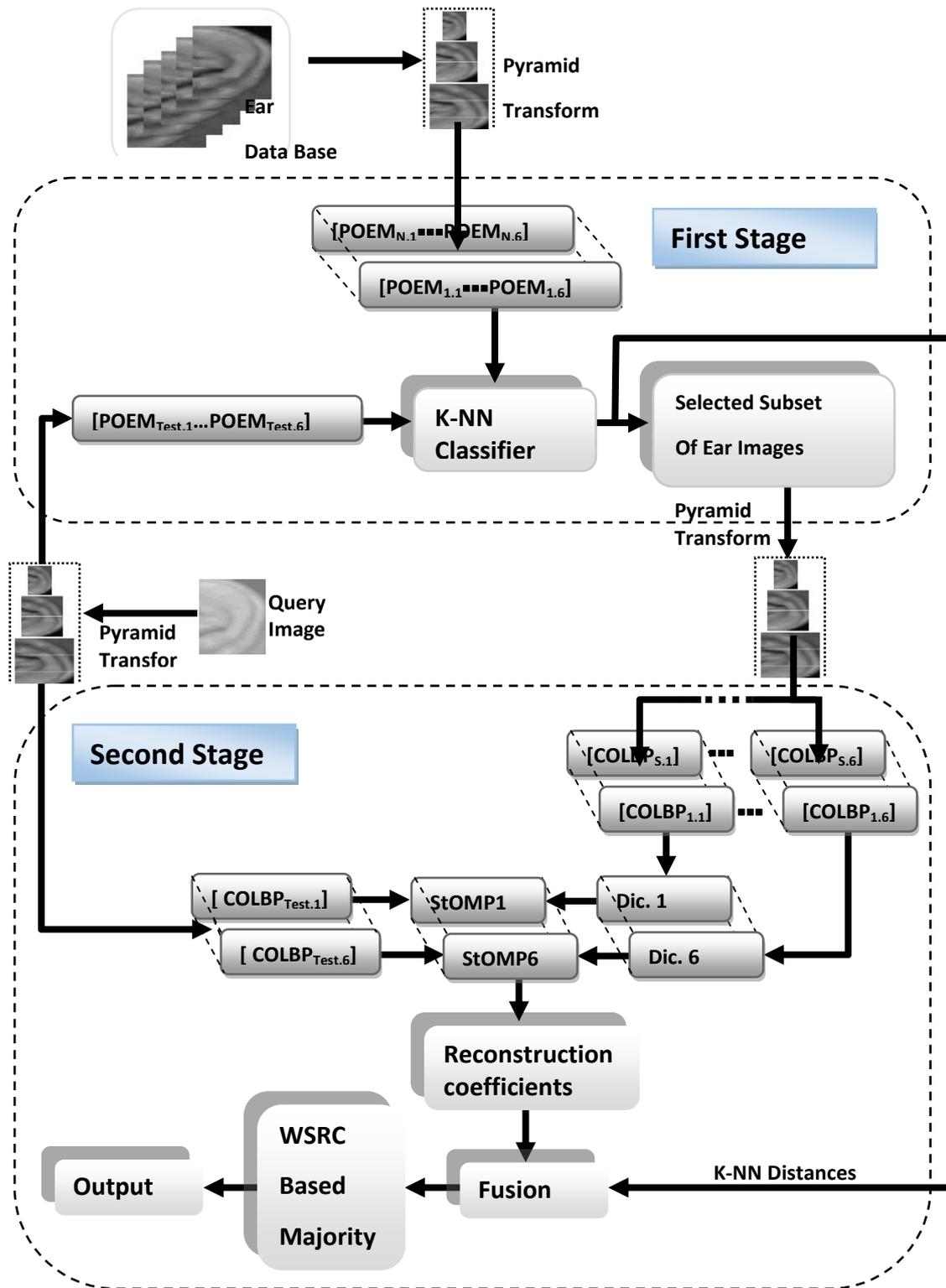


Figure. 5.13: Modèle de travail de l'approche proposée

- Une transformation pyramidale est appliquée au sous-ensemble des images d'apprentissage (S) et à l'image de test également. Le descripteur Co-LBP est alors appliqué à ces images. Pour le sous-ensemble d'images d'apprentissage, l'application du

descripteur Co-LBP donne six dictionnaires dont chacun contient les caractéristiques Co-LBP correspondant à un segment (Fig.5.13).

- La Sparse Representation est alors utilisée pour représenter chaque segment de l'image 'Test' comme une combinaison des segments du dictionnaire correspondant. L'algorithme St-OMP est utilisé pour cette représentation linéaire; il est appliqué pour extraire les coefficients de reconstruction de chaque segment de l'image 'Test'. Il en résulte six vecteurs  $\alpha_1 = [\alpha_{1,1}, \dots, \alpha_{1,S}], \dots, \alpha_6 = [\alpha_{6,1}, \dots, \alpha_{6,S}]$  où les composantes de chacun représentent les coefficients de reconstruction du segment correspondant.
- Pour placer l'image de test dans sa classe correspondante, on applique une nouvelle stratégie de prise de décision basée sur la SRC et la distance du plus proche voisin K-NN, appelée WSRC : On prend les coefficients de reconstruction ' $\alpha_i$ ', de la deuxième étape, et leurs distances correspondantes ' $d_i$ ', de la première étape, pour la fusion selon l'équation:

$$class_i = \operatorname{argmax}[\alpha_i/D_i] \quad (5.12)$$

Où  $class_i$  désigne la classe estimée du segment 'i'.

- Enfin, la règle de vote majoritaire est utilisée pour finaliser la décision et placer l'image de test dans sa classe comme suit:

$$class_{Test} = \operatorname{Majority Voting} \{class_1, \dots, class_6\} \quad (5.13)$$

Dans nos expériences, nous avons observé que le sous-ensemble sélectionné «S» des classes joue un rôle important dans la représentation du test. Une autre observation importante est que la fusion des distances de la première étape avec leurs coefficients correspondants de la deuxième étape (SRC), donne une meilleure performance de classification que le cas où l'un des d'eux est utilisé seul comme nous le montrera la suite de ce chapitre.

On notera que le vecteur POEM est dérivé de la concaténation du vecteur des caractéristiques POEM<sub>i</sub> de chaque segment d'une image d'oreille donnée dans un domaine de transformation pyramidale.

Le pseudo algorithme qui traduit les deux étapes de la figure 1 est donné ci-dessous

## Weighted Sparse Representation Classifier (WSRC) Algorithm

### First stage:

% Here, we coarsely determine the labels of the candidate classes in which the test sample could % belong. Suppose we have L classes and N training samples.

For i=1: #of training data do

- Project the ear image into a pyramid transform domain and split each sub-image.
- $POEM_i = [ \ ]$ ; % initializes the POEM vector of image 'i' with an empty vector

For j=1: 6 do % 6 is the number of segments

- $POEM_{i,j} = \text{Convert segment 'j' of the image 'i' to POEM vector}$
- $POEM_i = [POEM_i, POEM_{i,j}]$ ; % concatenation

End for

End for

$POEM_i = [POEM_{i,1}, \dots, POEM_{i,6}]$

% Represent a test sample 'Test' in the same way.

- $POEM_{Test} = [ \ ]$ ; % Initializes the POEM vector of 'Test' image with an empty vector
- 

For j=1: 6 do

- $POEM_{Test,j} = \text{Convert segment 'j' of the 'Test' image to POEM vector}$
- $POEM_{Test} = [POEM_{Test}, POEM_{Test,j}]$ ; % concatenation

End for

$POEM_{Test} = [POEM_{Test,1}, \dots, POEM_{Test,6}]$ ; % Select a subset classes (S) of training samples.

% Neighbor Classifier

$Distances = [ \ ]$ ;

% initializes the global vector

'Distances'

% with an empty vector

For i=1: #of training data do

- $Distance_i = \text{KNN}(POEM_{Test}, POEM_i)$ ;
- $Distances = [Distances, Distance_i]$ ; % concatenation

End for

- $[Distances, Labels] = \text{sort}(Distances)$ ; % sort the vector distances

- $Distances = Distances(1:S)$ ;

% Only keep the 'S' smallest distances

•  $Labels = Labels(1:S)$  ;

% with their corresponding 'S' Labels

### Second stage:

% In the second stage, we use St-OMP algorithm to represent each segment of the test sample

% 'Test' as a linear combination of the corresponding segments of the selected 'S' classes from

% the first stage.

For i=1: 6 do

%

initializes the dictionaries with empty vectors

$Dic_i = [ ]$ ;

End for

% below, the image "i" is the image whose index in the database is  $Labels(i)$

For i=1: S do

% S is the number of the selected images from the first stage

• Project the ear image 'i' into a pyramid transform domain and split each sub-image.

For j=1: 6 do

•  $CoLBP_{i,j} = \text{Convert segment 'j' of the image 'i' to CoLBP vector}$

•  $Dic_j = [Dic_j, CoLBP_{i,j}]$ ; % Concatenation = Construct the dictionaries

End for

End for

% Represent a test sample 'Test' in the same way.

For j=1: 6 do

•  $CoLBP_{Test,j} = \text{Convert segment 'j' of the 'Test' image to CoLBP vector}$

End for

% Apply sparse representation (St-OMP).

▪  $\alpha_1 = StOMP\{CoLBP_{Test,1}; Dic_1 = [CoLBP_{1,1}, \dots, CoLBP_{S,1}]\}$

▪  $\vdots$

▪  $\alpha_6 = StOMP\{CoLBP_{Test,6}; Dic_6 = [CoLBP_{1,6}, \dots, CoLBP_{S,6}]\}$

% The resulting coefficients from St-OMP are:

%  $\alpha_1 = [\alpha_{1,1}, \dots, \alpha_{1,S}], \dots, \alpha_6 = [\alpha_{6,1}, \dots, \alpha_{6,S}]$

% Fuse the obtained coefficients with their corresponding distances,

% from the first stage, to construct **WSRC** for each segment

% below  $D_j = Distances(j)$ , see algorithm of stage 1 for Distances(j)

For i=1: 6 do

For j=1: S do

•  $WSRC_i = [WSRC_i, \alpha_{1,j}/D_j]$ ; % concatenation

End for

```

End for

% This results in a set of vectors  $WSRC_i$  as:
%  $WSRC_1 = [\alpha_{1,1}/D_1, \dots, \alpha_{1,S}/D_S], \dots, WSRC_6 = [\alpha_{6,1}/D_1, \dots, \alpha_{6,S}/D_S]$ 
% Classify each segment into its appropriate class

▪  $classe\ 1 = \text{argmax}(WSRC_1)$  % Class for segment 1
▪      ⋮
%      ⋮
▪  $classe\ 6 = \text{argmax}(WSRC_6)$  % Class for segment 6

% Finally, majority voting rule is used for final classification of the test sample

▪  $Class\ Test = \text{majority voting } \{classe\ 1, classe\ 2, classe\ 3, classe\ 4, classe\ 5, classe\ 6\}$ 

```

### 5.5.2 Application et résultats obtenus

Dans cette section, nous étudierons les performances de l'approche proposée (WSRC) pour la reconnaissance de l'oreille dans des conditions contrôlées et non-contrôlées (occlusion). Des expériences approfondies sont menées sur deux bases de données, à savoir IIT Delhi-1 et IIT Delhi-2 pour valider la performance de la méthode proposée.

La première expérience est exécutée dans des conditions normales. Les résultats de toutes les stratégies sont présentés dans le Tableau 5.1. Ici, les «Alphas» sont les coefficients de reconstruction du St-OMP.

**Table 5.1:** Comparaison de la performance moyenne des différentes techniques de décision.

Method	POEM-Based K-NN	POEM-Based SRC (Alphas)	POEM-Based SRC (Residual)	Co-LBP-Based K-NN	Co-LBP-Based SRC (Alphas)	Co-LBP-Based SRC (Residual)	Proposed (WSRC)
DB1	95.73±1.66	96.27±0.92	97.33±0.92	97.13±1.2	97.86±0.46	97.86±1.22	<b>99.2±0.00</b>
DB2	95.93±0.45	96.68± 0.94	96.98±1.38	96.27±1.57	97.58±1.45	96.98±1.38	<b>98.34±0.26</b>

Comme on peut le voir dans ce tableau, la classification des caractéristiques extraites a été évaluée en utilisant différentes stratégies. La première est basée sur les résidus, tandis que la deuxième est articulée sur les coefficients de reconstruction du (SRC) et le classificateur K-NN. En outre, toutes les stratégies sont comparées à la proposition WSRC. Il est clair, à partir de ce tableau, que la méthode proposée WSRC atteint les meilleures performances en terme de taux de reconnaissance et de stabilité. Cela s'explique par le fait que la représentation d'image a été faite à l'aide de deux descripteurs locaux: ils encodent et contribuent à la représentation de l'image de l'oreille de différentes façons. D'autre part, la stratégie de décision proposée joue un rôle important dans l'augmentation du taux de classification (fusion des coefficients de reconstruction et des distances du sous-ensemble sélectionné).

Comme le WSRC obtient les meilleurs résultats, nous le considérons pour une comparaison avec les travaux plus récents de la reconnaissance d'oreilles présentés dans [22-24] et [1,25-31]. Les résultats respectifs sont présentés dans le tableau 5.2. Il convient de mentionner que, au meilleur de nos connaissances, les algorithmes présentés dans [22, 23] ont produit les taux de reconnaissance les plus élevés. A partir de la même table, parmi toutes les méthodes, il apparaît clairement que notre stratégie obtient les meilleurs résultats suivis par les travaux proposés dans [22] et [23] respectivement.

**Tableau 5.2:** Comparaison avec les règles de l'art dans des conditions normales

Dataset	Ref.	Method	Rate (%)
DB1	Kumar & Chan (2013) [22]	SR of local texture descriptor	97.06
	Hanmandlu (2013)[23]	Sigmoid feature (SF) and SVM	96.00
		Energy features (EF) and SVM	98.20
		Principal component analysis	79.00
	Benzaoui et al (2014) [24]	BSIF descriptor+KNN	97.26
	Kumar & Wu (2012) [1]	Orthogonal log-Gabor filter	96.27
	Kumar & Wu (2012) [1] Hurley et al (2002) [25]	Force Field Transform	74.93
	Chan & Kumar (2012) [26]	Quaternionic log-Gabor filters	96.53
		Monogenic log-Gabor filters	96.27
		1-D quadrature filters	95.73
	Damer & Benedikt (2012) [28]	Multi-Scale HOG and KNN	92.00
	Tariq et al (2011) [29]	Haar wavelets and fast normalized cross correlation	95.20
	Boodoo et al (2013) [30]	LBP based Ear Recognition	93.00
	Basit et al (2014)[31].	Non linear curvelet features based K-NN	97.77
<b>Proposed (WSRC) Method</b>			<b>99.2±0.00</b>
DB2	Kumar & Chan (2013) [22]	SR of local texture descriptor	97.93
	Hanmandlu (2013) [23]	Sigmoid feature (SF) and SVM	98.20
		Energy features (EF) and SVM	97.20
		Principal component analysis	82.60
	Benzaoui et al (2014) [24]	BSIF descriptor+Knn	97.34
	Kumar & Wu (2012) [1]	Orthogonal log-Gabor filter	95.93



Ces différents niveaux d'occlusion sont évalués pour démontrer l'efficacité de notre méthode, par rapport à d'autres travaux de reconnaissance de l'oreille occluse. En particulier, Yuan et al. [32] ont proposé une factorisation de matrice non négative basée sur la contrainte de SR pour la reconnaissance de l'oreille avec occlusion. Ces mêmes auteurs [33], ont proposé une approche de reconnaissance de l'oreille 2D en utilisant la fusion d'informations locales pour traiter la reconnaissance de l'oreille en occlusion partielle. Dans leur travail, Neighborhood Preserving Embedding a été utilisée pour l'extraction de caractéristiques. Ils ont utilisé les bases de données USTB (24 sujets) et UND (100 sujets). Ils signalent que les régions les plus importantes de l'oreille pourraient être représentées en utilisant seulement quelques sous-fenêtres (la région médiane de l'oreille). Zhang et al. [34] ont proposé une nouvelle technique d'extraction de caractéristiques utilisant l'information d'échelle des ondelettes de Gabor. La fonction Gabor extraite est traitée comme un dictionnaire par le classificateur de représentation non-négative (G-NSR). Dans des conditions normales et avec les bases de données UND J2 et USTB, l'algorithme de classification proposé présente des performances de reconnaissance élevées que le LBP et le filtre de Gabor. Il est également robuste aux occlusions d'oreille à grande échelle.

Pour les occlusions de la Figure. 5.14, la comparaison des performances de reconnaissance entre notre méthode et celle proposée dans [34] est résumée dans le tableau 5.3.

**Tableau 5.3:** Comparaison des performances de reconnaissance entre la méthode proposé WSRC et G-NSRC sous occlusion partielle: T = Haut, M = Milieu, B = Bas

Approaches	Database	Pourcentage de l'occlusion										
		15% (T)	15% (M)	15% (B)	30% (T)	30% (M)	30% (B)	40% (T)	40% (M)	40% (B)	50% (T)	50% (B)
G-NSRC[34]	USTBIII (24 subject)	-	97.7	-	87.6	-	-	-	-	75.2	-	51.9
Proposed WSRC	DB1 (125 subjects)	96.8 ±0.52	96.07 ±0.25	98.13 ±0.46	94.4 ±0.56	88.8 ±0.94	94.4 ±1.2	86.7 ±0.92	56 ±0.92	81.07 ±1.22	60.94 ±1.5	63.5 ±1.66
	DB2 (221 subjects)	95.47 ±0.64	98.13 ±0.5	96.83 ±0.92	90.8 ±1.22	82.8 ±0.92	94.41 ±0.94	82.7 ±0.94	59.6 ±1.2	87.33 ±1.57	60.03 ±2.3	77.5 ±1.38

Pour les occlusions de la figure 5.15, la comparaison des performances de reconnaissance entre notre méthode et celles proposées dans [32-34] sont rapportées dans le tableau 5.4.

**Tableau 5.4:** Taux de reconnaissance de l'oreille

Approaches	Database	Percent Occluded			
		0%	15%	25%	35%
<b>NPE [32]</b>	USTBIII (24 subjects)	94	91	<b>87</b>	<b>80</b>
<b>INMF [33]</b>	USTBIII (24 subjects)	88	83	<b>67</b>	<b>53</b>
<b>G-NSRC [34]</b>	USTBIII (24 subjects)	98	96	<b>91</b>	<b>86</b>
<b>Proposed WSRC</b>	USTBIII (24 subjects)	<b>100</b>	<b>97</b>	<b>93</b>	<b>88</b>

Comme on peut le constater à partir de ce tableau, le WSRC proposé atteint le meilleur taux de classification. Avec l'augmentation du niveau d'occlusion, WSRC réalise les performances les plus élevées. Ceci est dû à la division de l'image de l'oreille en différents segments puis représenter chaque segment avec deux descripteurs locaux différents ce qui permet de coder l'information de différentes façons. Le tableau 5.3 indique également que l'occlusion de la partie moyenne est le pire cas pour la classification. Ceci est dû au fait que les informations les plus distinctives sont comprises dans cette partie.

Dans notre travail précédent [20], nous proposons une technique de reconnaissance d'oreilles de type grossier vers fin (coarse-to-fine). Elle est basée sur deux descripteurs locaux (POEM et Co-ALBP) et deux classificateurs différents (K-NN et SRC). Le système proposé obtient une reconnaissance élevée et présente une robustesse significative contre l'occlusion. Un inconvénient d'une telle méthode est dû principalement à sa complexité et au temps de traitement élevés. Sur ce point, notre objectif principal dans la suite de notre travail est de développer un système de reconnaissance d'oreille simple mais efficace dans des conditions normale et anormale.

La deuxième méthode proposée [21], prend à la fois les coefficients résiduels et les coefficients de reconstruction. Les deux éléments sont liés par un mécanisme de décision unique. La représentation d'image (ainsi que le développement des dictionnaires) est traitée au moyen d'un descripteur local, à savoir la (**Co-ALBP**). Les résultats expérimentaux, réalisés sur deux ensembles de données publiques, ont mis en évidence que (i) des améliorations peuvent être notées par rapport aux méthodes basées sur SC, présentées dans 12,17,18 et, 20 (ii) des précisions supérieures ont été obtenues par rapport aux travaux récents de reconnaissance auriculaire.

### 5.5.3 Description de la deuxième méthode

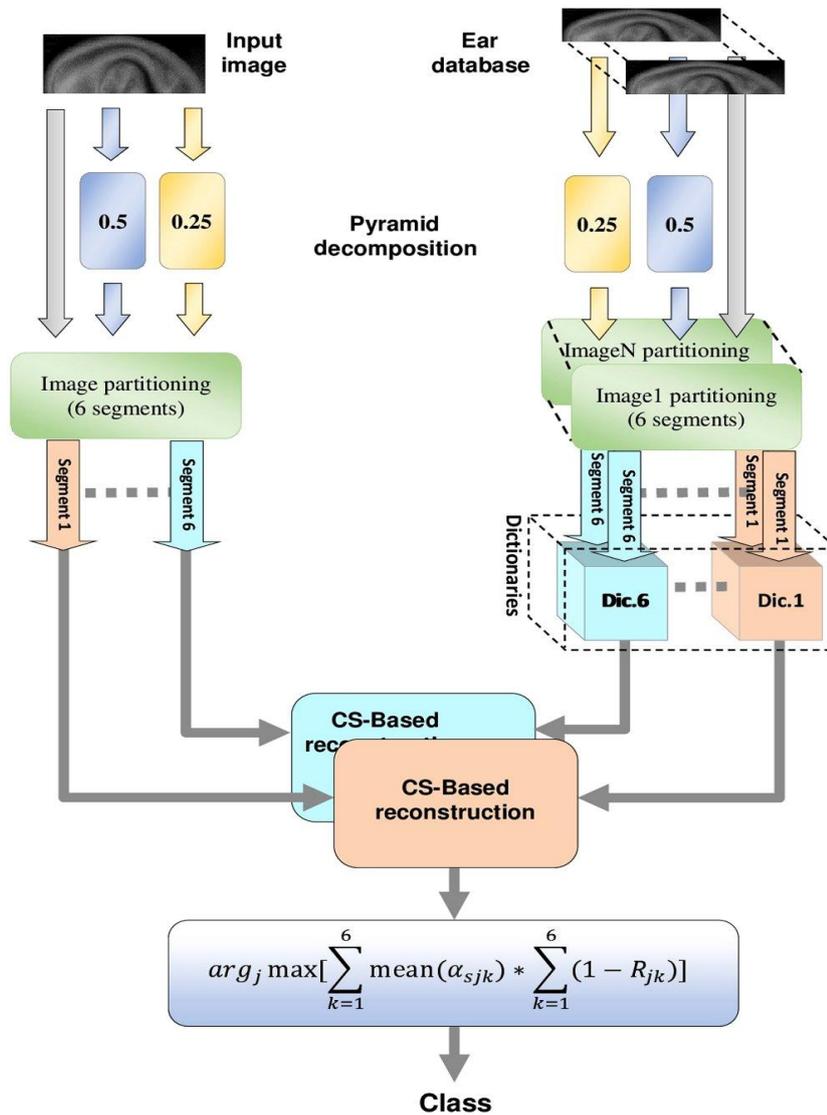
En termes générale, une prise de décision basée sur CS peut être faite par l'une des deux façons. Le premier, basée sur les valeurs résiduelles (somme des éléments du vecteur résiduel) comme un mécanisme ie, plus le résidu d'une certaine classe du dictionnaire est petit, plus il est proche de la classe du test. La deuxième stratégie, qui est relativement moins étudiée, s'appuie sur les coefficients de reconstruction où les coefficients les plus élevés d'une certaine classe du dictionnaire. Dans cette étude, nous proposons une troisième alternative. Elle considère conjointement les techniques mentionnées ci-dessus pour augmenter encore la performance. Si l'on considère le problème de N classes, alors pour un modèle de test, N résidus et N séquences de coefficients (*i. e.*,  $\alpha_s$ ) sont générées. Ensuite, les résidus indiqués dans toutes les classes du dictionnaire sont normalisés par rapport au résidu maximum. La règle de décision est basée sur la maximisation des coefficients de reconstruction (leur valeur moyenne) et est exprimée comme suit:

$$class = arg_j max [mean(\alpha_{sj}) * (1 - R_j)] \quad (5.14)$$

Où j est l'indice de classe du dictionnaire,  $\alpha_s$  est la séquence des coefficients de reconstruction pour la classe j et  $R_j$  est le résidu de la même classe. L'équation (5.14) exprime la forme généralisée de notre règle de décision. Cependant, dans notre travail, nous considérons également une analyse multi-résolution, qui sera détaillée dans la suite.

### 5.5.4 Mécanisme de reconnaissance de l'oreille par la méthode proposée

La méthode de reconnaissance de l'oreille proposée passe par deux étapes. Dans la première étape, les caractéristiques sont extraites à partir des régions d'image (figure 5.16). Dans la deuxième étape, les caractéristiques extraites sont classées selon la règle de décision expliquée ci-dessus. Pour l'étape d'extraction des caractéristiques, les étapes mentionnées ci-dessous sont suivies:



**Figure 5.16:** Techniques de reconnaissance d'oreille proposée [21].

- L'image de l'oreille de test est échantillonnée par décomposition pyramidale (c'est-à-dire une analyse multi-résolution) en convoluant cette image avec un filtre gaussien. Trois images sont ainsi obtenues.
- Ensuite, l'image originale est divisée, en ligne, en trois bandes égales; la deuxième image est divisée en deux moitiés, tandis que la troisième image est retenue telle quelle, il en résulte six segments.
- Les caractéristiques considérées (vecteurs) sont extraites de chaque segment (c'est-à-dire, 6 vecteurs caractéristiques).
- En construisant six dictionnaires de caractéristiques à partir des images d'entraînement, où chaque dictionnaire est dédié à un segment d'images, la reconstruction des vecteurs de caractéristiques de test déjà extraits devient possible (chaque segment d'image est reconstruit sur son dictionnaire d'entraînement correspondant).

- La procédure précédente calcule les valeurs résiduelles ainsi que les séquences de coefficients de chaque segment, respectivement. Ensuite, les quantités  $mean(\alpha_{Sj})$  et  $1-R_j$  de tous les 6 segments sont calculées et la décision de la classe correspondante est faite à partir de l'équation 5.15) comme suit:

$$class = arg_j max \left[ \sum_{k=1}^6 mean(\alpha_{SJK}) * (1 - R_{JK}) \right] \quad (5.15)$$

### 5.5.5 Application et résultats obtenus

Des expériences approfondies sont menées, sur deux bases de référence, à savoir IIT Delhi-1 et IIT Delhi-2, pour valider la performance de la méthode proposée.

Afin d'évaluer la performance de la méthode proposée, nous rapportons les résultats expérimentaux du protocole de test A [1] (par rapport à la littérature). Ce protocole génère le taux de reconnaissance moyen de trois tests, où chacune des trois premières images de chaque sujet est fixée comme image de test tandis que les autres sont retenues comme des images d'apprentissage (galerie).

Comme on l'a déjà motionnée, la caractéristique adoptée dans ce travail est la Co-ALBP, qui est une alternative à la LBP. La Co-ALBP étudie la Co-Occurrence des LBP sur une image donnée et considère en même temps leur information spatiale. Ainsi, il a été démontré qu'elle a une capacité de représentation robuste par rapport aux autres descripteurs locaux.

Pour valider la précision de la méthode proposée, nous devons d'abord démontrer que la règle de décision proposée est plus performante que dans le cas où les résidus ou les coefficients de reconstruction sont considérés uniquement. Ainsi, pour la technique basée sur les résidus, nous comparons notre méthode à celle présentée dans [12], où les valeurs résiduelles sont considérées comme un mécanisme de décision. Alors que pour la technique basée sur les coefficients, nous la comparons à la stratégie proposée dans [17], en prenant la somme des coefficients de reconstruction comme règle de décision. Nous comparons également notre méthode au cas où la moyenne des coefficients ( $\sum_{k=1}^6 mean_{\alpha_{sjk}}$ ) est prise comme règle de décision.

Les résultats obtenus des deux ensembles de données sont présentés dans le tableau 5.5.

**Tableau 5.5:** Comparaison de la performance moyenne des différentes règles de décision

Rule/Dataset	IIT Delhi-1	IIT Delhi-2
Proposed in <sup>13</sup>	98.93	98.49
Proposed in <sup>35</sup>	99.20	98.64

<b>Average</b>	99.2	98.94
<b>Ours</b>	<b>99.47</b>	<b>98.94</b>

A partir de ce tableau, on peut voir que la décision basée sur les résidus donne les taux les plus faibles et que la décision basée sur les coefficients semble avoir une meilleure performance en somme et en moyenne. Cependant, la décision conjointe présentée dans notre travail semble garantir une meilleure efficacité que tous. La raison derrière cela pourrait être le fait que la reconstruction basée sur le SC repose sur un nombre limité d'itérations, afin d'obtenir les résidus et les coefficients, ce qui rend la décision par un seul mécanisme (résidus ou des coefficients) moins précis que l'utilisation des deux qui consolide la règle de décision et assure une performance plus fiable.

Pour mettre en évidence la précision de la méthode proposée, on compare sa performance avec le K-NN et les classificateurs d'intelligence artificielle tel que: Support Vector Machine (SVM) et les Réseaux de Neurones Radial Basic Function (RBF). Le classificateur K-NN tente de trouver l'échantillon d'entraînement le plus proche de l'échantillon d'essai en fonction d'une mesure de distance donnée. Dans l'étude en cours, on a utilisé une mesure de distance Chi-Square avec  $K = 1$  (c'est-à-dire le plus proche voisin).

SVM a été initialement conçu pour les problèmes de classification binaire [36]. Afin de l'étendre au cas de classification multi-classe, une structure hiérarchique comprenant un groupe des SVM binaires est utilisée. Dans notre travail, on a adopté la stratégie un contre tous [37], avec une fonction gaussienne comme fonction d'activation. En utilisant la Lib-SVM [38], les paramètres du modèle ont été ajustés en utilisant une recherche en grille sur l'ensemble d'apprentissage.

Les réseaux de neurones RBF sont utilisés dans de nombreuses applications telles que la classification et la régression [39]. [40] a également été utilisé pour valider notre technique de classification proposée. Un modèle RBF contient trois couches: une couche d'entrée, une couche cachée avec une fonction d'activation non linéaire et une couche de sortie linéaire. La méthode du Crosse validation a été utilisée pour sélectionner les paramètres optimaux du modèle (nombre de neurones dans la couche cachée, la largeur et le centre de la fonction RBF).

Nous notons ici qu'une image d'oreille de chaque classe a été utilisée comme un ensemble de validation et le reste pour l'entraînement, nous remarquons également que toutes les techniques de classification utilisent les mêmes caractéristiques (Co-LBP) ce qui justifie la

comparaison de capacité de classification avec nos techniques proposées. Les résultats sont présentés dans le tableau 5.6.

**Tableau 5.6:** Comparaison de la performance moyenne avec différentes méthodes

Method	SVM	RBF	K-NN	Proposed Method
IIT Delhi-1	96.7±1.15	94.49±1.32	97.13±1.2	<b>99.47±0.46</b>
IIT Delhi-2	96.98±1.82	95.63±0.6918	96.27±1.57	<b>98.94±0.92</b>

Au vue des résultats de ce tableau, la méthode proposée atteint les taux de classification les plus élevés suivis du classificateur K-NN. En plus de la performance de reconnaissance, l'avantage de la méthode proposée est principalement lié à sa simple mise en œuvre (sans entraînement) et à l'exécution rapide, ce qui la rend adéquate à des applications en temps réel, contrairement aux réseaux neuronaux SVM et RBF qui nécessitent plus de temps et plus des échantillons pour l'entraînement et la validation.

En ce qui concerne le temps de traitement et de classification, le tableau 5.7 indique les temps d'exécution des K-NN, RBF, SVM et notre méthode de classification. Notons que les temps d'exécution énumérés dans ce tableau représentent le temps moyen total requis pour extraire les caractéristiques de Co-LBP et celui nécessaire pour construire des classificateurs basés sur l'ensemble de données d'entraînement plus le temps moyen requis pour prédire les classes des échantillons de test.

**Tableau 5.7:** Comparaison des temps d'exécution en secondes

Method	SVM	RBF	K-NN	Proposed Method
<b>Database1 (DB1)</b>	129.20	105.70	0.56	<b>1.74</b>
<b>Database1 (DB2)</b>	193.45	173.85	0.83	<b>3.75</b>

Comme le montre le tableau 5.7, le temps de construction du classificateur RBF et SVM est sensiblement plus élevé que le temps pris dans K-NN et la méthode proposée. Ces résultats impliquent que le temps nécessaire pour sélectionner le meilleur modèle de l'RBF et de l'SVM pourrait être inacceptable pour la reconnaissance de l'oreille, en particulier dans le cas où de nouveaux sujets (classes) sont continuellement ajoutés dans une base de données.

En plus du temps de traitement exigeant de la RBF et de la SVM, leur performance par rapport à notre technique de classification proposée, valide notre méthode de fusion proposée pour la reconnaissance de l'oreille.

Pour la troisième partie de la comparaison, nous considérons des travaux récents concernant la reconnaissance auriculaire [20-22-24] et [1,25-31]. La procédure d'évaluation est la même que ce qui a été indiqué précédemment. Tous les travaux et les résultats sont présentés au tableau 5.8. Il est clair que l'approche proposée surpasse les autres méthodes. L'interprétation des résultats obtenus pourrait être double. D'un côté, les caractéristiques adoptées pour représenter les images de l'oreille jouent un rôle principal dans la mesure dans laquelle la discrimination interclasse serait grande. Dans notre travail, nous avons exploité la puissance discriminative de la Co-LPB. D'autre part, une bonne caractéristique représentative ne suffit pas à produire des résultats satisfaisants, la prise de décision joue également la moitié du rôle de reconnaissance et le SC a démontré son efficacité dans le contexte de la reconnaissance de l'oreille dans notre travail. Cela est dû à sa capacité à récupérer de façon restreinte le modèle de test en maintenant les échantillons potentiellement les plus proches des dictionnaires adoptés. En plus, l'analyse multi-résolution a permis d'analyser l'image sur différentes échelles pour collecter plus d'informations sur la structure de l'oreille humaine.

**Tableau 5.8:** Comparaison avec l'état de l'art dans des conditions normales.

Dataset	Ref.	Method	Recognition Rate(%)
DB1	Kumar & Wu (2012) [1] Hurley et al. (2002) [25]	Force Field Transform	74.93
	Hanmandlu (2013) [23]	Haar wavelets and fast normalized cross correlation	95.20
	Kumar & Wu (2012) [1]	Orthogonal Log-Gabor filter	96.27
	Damer & Benedikt (2012) [28]	Multi-Scale HOG and KNN	92.00
	Chan & Kumar (2012) [26]	Quaternionic log-Gabor filters	96.53
		Monogenic log-Gabor filters	96.27
		1-D quadrature filters	95.73
	Kumar & Chan (2013) [22]	SR of local texture descriptor	97.06
	Boodoo et al (2013) [30]	LBP based Ear Recognition	93.00
	Hanmandlu (2013) [23]	Sigmoid feature (SF) and SVM	96.00
		Energy features (EF) and SVM	98.20
		Principal component analysis	79.00
	Basit et al. (2014)[31].	Non linear-curvelet features based K-NN	97.77
	Benzaoui et al. (2014) [24]	BSIF descriptor+KNN	97.26
<b>Mawloud et al. (2016) [20]</b>	<b>Weighted sparse representation (WSR)</b>	<b>99.2</b>	

	Proposed Method		99.47±0.46
DB2	Kumar & Wu (2012) [1] Hurley et al. (2002) [25]	Force Field Transform	66.67
	Kumar & Wu (2012) [1]	Orthogonal log-Gabor filter	95.93
	Chan & Kumar (2012) [26]	Quaternionic log-Gabor filters	95.32
		Monogenic log-Gabor filters	96.08
		1-D quadrature filters	94.72
	Kumar & Chan (2013) [22]	SR of local texture descriptor	97.93
	Hanmandlu (2013) [23]	Sigmoid feature (SF) and SVM	98.20
		Energy features (EF) and SVM	97.20
		Principal component analysis	82.60
	Jacob L. & Raju (2014)[27]	Texture features	94.12
	Basit et al. (2014)[31].	Non linear curvelet features based K-NN	96.22
	Mawloud et al. (2016)[20]	<b>Weighted sparse representation (WSR)</b>	<b>98.34±0.26</b>
<b>Proposed Method(21)</b>		<b>98.94±0.92</b>	

Dans la quatrième partie de l'expérience, nous appliquons l'approche proposée sur les images de l'oreille partiellement occluse de l'USTB [55]. Cette base de données contient 24 sujets. Trois niveaux d'occlusion différents sont présents pour l'image de test. Pour démontrer la performance de notre méthode, les résultats de reconnaissance obtenus en occlusion sont comparés avec quatre travaux pertinents sur le même ensemble de données. Dans le processus de reconnaissance, les cinq images d'oreilles normales, sans occlusion, sont utilisées pour la l'entraînement tandis que les images occluses sont conservées pour le test.

**Tableau 5.9:** Taux de reconnaissance en fonction de l'occlusion de l'oreille.

Approches	Pourcentage de l'occlusion			
	0%	15%	25%	35%
NPE [ 46]	94	91	87	80
INMF [ 47]	88	83	67	53
G-NSRC [ 48]	98	96	91	86
WSRC [20]	100	97	93	88
<b>Proposed Method [21]</b>	<b>100</b>	<b>98</b>	<b>95</b>	<b>89.5</b>

Comme le montre le tableau 5.9, notre approche de classification dépasse les trois autres méthodes. Avec l'augmentation du niveau d'occlusion, notre approche atteint les

performances les plus élevées. Les résultats expérimentaux du tableau 5.9 montrent que la méthode proposée est plus robuste à l'occlusion de l'oreille, en particulier à grande échelle.

## 5.6 Conclusion

Ce chapitre traitait de nouvelles méthodes de classification auriculaire 2D. Nous avons bâti ces méthodes sur la combinaison entre les descripteurs locaux, POEM et Co-LBP. Nous avons développé également une nouvelle méthode de classification basée sur la méthode de SRC.

Nous avons exploré dans ce chapitre les différentes techniques locales dans le but de partitionner l'espace de l'oreille en des classes homogènes et bien séparées. Si notre approche a bien réussi pour la reconnaissance auriculaire, nous pensons qu'elle reste valable pour toutes les modalités biométriques.

L'objectif de la première approche est de réduire l'espace de recherche par élimination des classes les plus dissemblables de l'identité requête avec une stratégie de combinaison des caractéristiques et de classificateurs. Ces classificateurs diffèrent aussi bien par leurs données d'entrées, à travers les différentes représentations, que par les méthodes de classification mises en place. L'identification finale consiste à apparier l'identité-requête avec un sous-ensemble de la galerie de la phase de réduction. Comme on a montré précédemment cette technique donne des bons résultats, spécialement dans les cas où l'oreille est prise dans des conditions anormales (avec occlusion). Pour améliorer la performance de cette technique on a proposé une autre technique de classification, basée sur la fusion entre les coefficients de reconstructions et les résidus du descripteur Co-LBP. D'après les résultats obtenus la méthode proposée est plus performante que la première en terme du taux de classification et du temps de calcul.

**Bibliographie :**

- [1] Kumar, A., & Wu, C. (2012). Automated human identification using ear imaging. *Pattern Recognition*, 45(3), 956-968.
- [2] Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1), 71-86.
- [3] Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7), 711-720.
- [4] Etemad, K., & Chellappa, R. (1997). Discriminant analysis for recognition of human face images. *JOSA A*, 14(8), 1724-1733.
- [5] Yang, J., Zhang, D., Frangi, A. F., & Yang, J. Y. (2004). Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE transactions on pattern analysis and machine intelligence*, 26(1), 131-137.
- [6] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.
- [7] Vu, N. S., Dee, H. M., & Caplier, A. (2012). Face recognition using the POEM descriptor. *Pattern Recognition*, 45(7), 2478-2488.
- [8] Vu, N. S., & Caplier, A. (2012). Enhanced patterns of oriented edge magnitudes for face recognition and image matching. *IEEE Transactions on Image Processing*, 21(3), 1352-1365.
- [9] Nosaka, R., Ohkawa, Y., & Fukui, K. (2011, November). Feature extraction based on co-occurrence of adjacent local binary patterns. In *Pacific-Rim Symposium on Image and Video Technology* (pp. 82-91). Springer Berlin Heidelberg.
- [10] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE.
- [11] Ahonen, T., Rahtu, E., Ojansivu, V., & Heikkila, J. (2008, December). Recognition of blurred faces using local phase quantization. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* (pp. 1-4). IEEE.
- [12] Kannala, J., & Rahtu, E. (2012, November). Bsim: Binarized statistical image features. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 1363-1366). IEEE.
- [13] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2), 210-227.
- [14] Gemmeke, J. F., Virtanen, T., & Hurmalainen, A. (2011). Exemplar-based sparse representations for noise robust automatic speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7), 2067-2080.
- [15] He, Z., Xie, S., Ding, S., & Cichocki, A. (2007). Convolutional blind source separation in the frequency domain based on sparse representation. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(5), 1551-1563.
- [16] Candes, E. J., & Tao, T. (2005). Decoding by linear programming. *IEEE transactions on information theory*, 51(12), 4203-4215.

- [17] Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, 57(2), 137-154.
- [18] Eldar, Y. C., & Kutyniok, G. (Eds.). (2012). *Compressed sensing: theory and applications*. Cambridge University Press.
- [19] Donoho, D. L., Tsaig, Y., Drori, I., & Starck, J. L. (2012). Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 58(2), 1094-1121.
- [20] Mawloud, G., & Djamel, M. (2016). Weighted sparse representation for human ear recognition based on local descriptor. *Journal of Electronic Imaging*, 25(1), 013036-013036.
- [21] Guermoui, M., Melaab, D., & Mekhalfi, M. L. (2016). Sparse coding joint decision rule for ear print recognition. *Optical Engineering*, 55(9), 093105-093105.
- [22] Kumar, A., & Chan, T. S. T. (2013). Robust ear identification using sparse representation of local texture descriptors. *Pattern recognition*, 46(1), 73-85.
- [23] Kumar, A., & Chan, T. S. T. (2013). Robust ear identification using sparse representation of local texture descriptors. *Pattern recognition*, 46(1), 73-85.
- [24] Hanmandlu, M. (2013). Robust ear based authentication using local principal independent components. *Expert Systems with Applications*, 40(16), 6478-6490.
- [25] Benzaoui, A., Hadid, A., & Boukrouche, A. (2014). Ear biometric recognition using local texture descriptors. *Journal of Electronic Imaging*, 23(5), 053008-053008.
- [26] Hurley, D. J., Nixon, M. S., & Carter, J. N. (2002). Force field energy functionals for image feature extraction. *Image and Vision computing*, 20(5), 311-317.
- [27] Chan, T. S., & Kumar, A. (2012). Reliable ear identification using 2-D quadrature filters. *Pattern Recognition Letters*, 33(14), 1870-1881.
- [28] Jacob, L., & Raju, G. (2014). Ear Recognition Using Texture Features-A Novel Approach. In *Advances in Signal Processing and Intelligent Recognition Systems* (pp. 1-12). Springer International Publishing.
- [29] Damer, N., & Führer, B. (2012, July). Ear recognition using multi-scale histogram of oriented gradients. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on* (pp. 21-24). IEEE.
- [30] Tariq, A., Anjum, M. A., & Akram, M. U. (2011, December). Personal identification using computerized human ear recognition system. In *Computer Science and Network Technology (ICCSNT), 2011 International Conference on* (Vol. 1, pp. 50-54). IEEE.
- [31] Boodoo-Jahangeer, N. B., & Baichoo, S. (2013, November). LBP-based ear recognition. In *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on* (pp. 1-4). IEEE.
- [32] Basit, A., & Shoab, M. (2014). A human ear recognition method using nonlinear curvelet feature subspace. *International Journal of Computer Mathematics*, 91(3), 616-624.
- [33] Yuan, L., Mu, Z. C., Zhang, Y., & Liu, K. (2006, August). Ear recognition using improved non-negative matrix factorization. In *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 4, pp. 501-504). IEEE.
- [34] Yuan, L., & chun Mu, Z. (2012). Ear recognition based on local information fusion. *Pattern Recognition Letters*, 33(2), 182-190.

- [35] Zhang, B., Mu, Z., Li, C., & Zeng, H. (2014). Robust classification for occluded ear via Gabor scale feature-based non-negative sparse representation. *Optical Engineering*, 53(6), 061702-061702.
- [36] Li, J., & Lu, C. Y. (2013). A new decision rule for sparse representation based classification for face recognition. *Neurocomputing*, 116, 265-271.
- [37] Vapnik, V. (2013). *The nature of statistical learning theory*. Springer Science & Business Media.
- [38] Guermoui, M., Mekhalfi, M. L., & Ferroudji, K. (2013, May). Heart sounds analysis using wavelets responses and support vector machines. In *Systems, Signal Processing and their Applications (WoSSPA), 2013 8th International Workshop on* (pp. 233-238). IEEE.
- [39] Lib-SVM toolbox, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/oldfiles/index-1.0.html> (15 March 2016)
- [40] Er, M. J., Wu, S., Lu, J., & Toh, H. L. (2002). Face recognition with radial basis function (RBF) neural networks. *IEEE transactions on neural networks*, 13(3), 697-710.
- [41] RABEHI, A., GUERMOUI, M., DJAFER, D., & ZAIANI, M. (2015). Radial basis function neural networks model to estimate global solar radiation in semi-arid area. *Leonardo Electronic Journal of Practices and Technologies*, (27), 177-184.

# **Conclusion Générale et Perspectives**

## **Conclusion Générale et Perspective :**

La biométrie est un domaine à la fois passionnant et complexe. Elle tente, par des outils mathématiques souvent très évolués, de faire la distinction entre les individus, nous obligeant ainsi à travailler dans un contexte de très grande diversité. Cette diversité se trouve également dans le nombre considérable d'algorithmes qui ont été proposés pour différentes modalités biométrique spécialement dans la reconnaissance faciale.

Le travail présenté dans cette thèse est destiné à la classification des personnes à partir des images bidimensionnelles. On a étudié et analysé deux différentes modalités biométriques: le visage et l'oreille puisqu'elles offrent plusieurs avantages: le système de capture (caméra) est facile à installer est accepté dans les lieux publics et par les utilisateurs aussi elles sont complémentaire (peuvent être utilisées en mode multi biométries).

Le but principale de ce travail est de prouver l'importance de l'utilisation de l'oreille humaine comme une modalité biométrique alternative /principale dans la reconnaissance des personnes.

Une analyse des différentes techniques de reconnaissance développées au cours de ces dernières années a été présentée pour les deux modalités, et cela pour mettre en évidence les particularités ainsi que les avantages et les inconvénients de chacune d'entre elles.

Au cours de cette thèse, une présentation générale de la biométrie, ses caractéristiques, ses limitations, ainsi que les enjeux liés à ses évaluations est effectuée dans le chapitre 1. Une présentation des travaux de la littérature liés à la reconnaissance faciale est fournie dans le chapitre 2. Ces deux premiers chapitres définissent le contexte de la première partie de cette thèse, ainsi que les outils préalables à notre étude. Nous avons ensuite présenté notre approche pour la reconnaissance faciale dans le chapitre 3. Dans le chapitre 4, on a présenté la modalité auriculaire et les différents travaux et problèmes liés à la reconnaissance de l'oreille humaine. Dans le chapitre 5 on a présenté deux différentes techniques pour la classification auriculaire. Les résultats obtenus montrent l'intérêt des techniques proposées spécialement dans le cas de l'occlusion qui a un impact majeur sur la performance globale des systèmes biométriques.

La première contribution, présentée dans le chapitre 3, concerne la proposition d'une nouvelle méthode d'extraction des caractéristiques basée sur le descripteur LBP originale. Au lieu d'utiliser la différence des pixels au rayon  $R$  et le pixel centrale, la méthode proposée utilise l'information des pixels, au rayon  $R$  et  $R-1$ , pour extraire le code caractéristique d'une région donnée de l'image. Cette technique est plurimodal (visage, empreinte digitale, la main...), et

indépendante du système de vérification utilisé. Nous l'avons validé sur quatre bases de données publiques du visage. Les résultats expérimentaux obtenus montrent l'intérêt et la capacité discriminative de notre approche par rapport à la LBP originale et d'autres techniques récentes utilisées dans la biométrie faciale.

On a utilisé une autre variante de LBP appelé CLBP avec une présentation pyramidale est une méthode de sélection des caractéristiques (Fisher Score). Les résultats obtenus sont très satisfaisants et montrent l'avantage des méthodes locales dans l'extraction des caractéristiques informatives pour la représentation d'une image donnée.

Une méthode de classification basée sur deux descripteurs locaux et deux différents classificateurs a été développée pour la reconnaissance auriculaire. Dans la première étape un sous-ensemble des classes est sélectionné en utilisant le descripteur POEM et la méthode de K-NN pour la classification. L'ensemble sélectionné par la première étape est traité par le descripteur Co-LBP. Pour la classification finale une fusion des coefficients de reconstruction (Alpha) avec les distances correspondantes de la première étape est proposée. La technique proposée est évaluée sur deux bases de données: IIT Delhi-1 et IIT Delhi-2 et les résultats obtenus sont très performants avec ou sans occlusion.

La quatrième contribution dans ce travail est la proposition d'une nouvelle technique de classification des oreilles humaines, basée sur la fusion des coefficients de reconstruction est les résidus en utilisant le descripteur Co-LBP comme dictionnaire. Les résultats obtenus sont également très performants par rapport à la première méthode en termes de précision et temps de calcul.

En générale on peut conclure que la biométrie auriculaire présente beaucoup d'avantages par rapport aux autres modalités biométriques et est la modalité la plus adéquate pour la fusion avec le visage dans le cas multimodale.

Une autre conclusion, l'utilisation de techniques locales d'une façon locale ou globale donne de bons résultats en termes de temps de calculs et précision.

### **Perspectives**

Sur la base des contributions de cette thèse, les orientations de recherche suivantes semblent prometteuses. Nous les articulons autour des axes suivants :

- Des caractéristiques supplémentaires telles que la morphologie ou la couleur de l'appariement doivent être considérées.
- Nous souhaitons développer d'autres méthodes de description des images faciales plus discriminatives, spécialement dans le cas d'un écart de temps très important entre les différentes captures comme le cas de la base **FGNET** ou l'écart de temps et 58 ans.

- Bien qu'il semble que les oreilles sont cohérentes avec l'âge, pour une analyse concrète, nous avons besoin d'une base de données avec un grand écart de temps entre les échantillons capturés.
- Puisque les deux modalités (Visage, Oreille) sont complémentaires, il serait intéressant d'étudier la fusion entre eux (en mode multimodal) soit : au niveau des images, au niveau des caractéristiques, au niveau des scores ou encore au niveau des classificateurs.

## ملخص:

أعمال البحث المقدمة في هذه الأطروحة تتعلق بتصنيف الأشخاص بالاعتماد على صور ثنائية الأبعاد. اعتمدنا بصمتين: الوجه والأذن. لأنهما توفران عدة مزايا مقارنة بالبصمات التقليدية، وأنهما متكاملتان بمعنى أنه يمكن دمجهما بشكل متعدد البصمات. على وجه التحديد طريقتان تشكلان جوهر عملنا. من جهة، استخدمنا الخصائص المحلية من أجل وصف أدق لصورة الوجه/الأذن ومن جهة أخرى استثمرنا نجاح تقنية Sparse Coding (SC) لتمثيل صورة معينة (الوجه/الأذن) بدلالة مجموعة من صور التدريب، لاستخدامها في عملية التصنيف. بشكل أدق، في ما يخص تصنيف الوجه لقد عرضنا طريقة جديدة تمثل شكل جديد لعامل التصنيف المتداول Local Binary Pattern (LBP)، والتي أطلقنا عليها اسم Modified Local Binary Pattern (MLBP)، مدموجة مع مصنف من نوع SC. في ما يخص تصنيف الأذن، لقد عرضنا تقنيتين جديدتين: الأولى تتمثل في دمج عامل تصنيف محليان و مصنفين، أما التقنية الثانية فتعتمد على استخلاص الخصائص المحلية لصورة ما ثم دمج معاملات إعادة تكوين هذه الصورة مع بقايا تكوينها على الترتيب. أفضت التجارب المعمقة لتصنيف صور لأشخاص، مأخوذة من عدة قواعد بيانات عالمية، باستخدام التقنيات الجديدة المقترحة المذكورة أعلاه، إلى نتائج تبرهن فعالية هذه التقنيات، حيث أنه تم تحسين نسبة التصنيف وبشكل ملحوظ مقارنة مع غيرها من التقنيات المعروضة في أعمال البحث المنشورة حديثاً.

## Résumé:

Les travaux de recherche menés dans le cadre de cette thèse portent sur la classification des personnes basées sur des images bidimensionnelles. Nous avons adopté deux modalités, à savoir (i) le visage, et (ii) l'oreille. Ces deux modalités offrent plusieurs avantages par rapport aux modalités traditionnelles et elles sont complémentaires en ce sens qu'elles peuvent être combinées de façon multimodale. Précisément, deux composantes forment l'essence de notre travail. D'une part, nous avons utilisé les caractéristiques locales afin de bien décrire l'image face/oreille. D'autre part, nous exploitons le succès de la technique du Sparse Coding (SC) pour représenter un signal (image) de test donné en termes de seulement quelques atomes significatifs d'un ensemble d'entraînement donné et l'avons encore adapté à notre effort de classification. Précisément, en ce qui concerne la reconnaissance faciale, nous avons présenté une nouvelle variante au descripteur Local Binary Pattern (LBP) appelé Modified Local Binary Pattern (MLBP), couplé à un classificateur SC. En ce qui concerne la reconnaissance de l'oreille, nous avons proposé deux nouvelles techniques. La première consiste à fusionner deux descripteurs locaux et deux classificateurs, tandis que la seconde procède en extrayant les caractéristiques locales d'un sujet d'essai donné, puis on fusionne les coefficients de reconstruction (Alphats) avec leurs résidus respectifs. Des expériences approfondies ont été menées sur des ensembles de données de référence, concernant les deux modalités, et une robustesse remarquable avec des résultats plausibles a été notée, ce qui a entraîné des améliorations tangibles par rapport aux travaux récents.

## Abstract :

The research work carried out in this dissertation addresses the task of image-based person classification. We adopted two modalities, namely (i) face, and (ii) ear print, provided that they offer several advantages with respect to traditional modalities and they are complementary in the sense that they can be combined in a multimodal fashion. Precisely, two components form the essence of our work. On the one hand, we capitalized on local features in order to well describe the face/ear image. On the other hand, we exploit the success of Sparse Coding (SC) technique in representing a given test signal/image in terms of only a few significant atoms from a given training set, and further adapted it to our identification endeavor. Precisely, regarding face recognition, we put forth a novel variant to the well-known Local Binary Pattern (LBP) descriptor, called Modified Local Binary Pattern (MLBP), coupled with a SC classifier. Concerning ear recognition, we proposed two schemes. The first one consists in fusing two local descriptors and two classifiers, whilst the second one proceeds by extracting local features of a given test subject, followed by fusing the sparse representation codes along with their respective residuals. Extensive experiments have been conducted on benchmark datasets with respect to both modalities and notable robustness with plausible results have been scored, incurring tangible improvements as compared to recent works.