

Reconnaissance Biométrique par Fusion Multimodale du Visage et de l'Iris

Nicolas Morizet

▶ **To cite this version:**

Nicolas Morizet. Reconnaissance Biométrique par Fusion Multimodale du Visage et de l'Iris. domain_other. Télécom ParisTech, 2009. English. NNT : . pastel-00005811

HAL Id: pastel-00005811
https://pastel.hal.science/pastel-00005811

Submitted on 8 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



École Doctorale
d'Informatique,
Télécommunications
et Électronique de Paris

Thèse

présentée pour obtenir le grade de Docteur

de l'Ecole Nationale Supérieure des Télécommunications

Spécialité : Signal et Images

Nicolas MORIZET

Reconnaissance Biométrique par Fusion Multimodale du Visage et de l'Iris

Soutenue le 18 Mars 2009 devant le jury composé de :

M. Michel Terré	Professeur, CNAM, Paris	Rapporteur
M. Patrick Garda	Professeur, UPMC, Paris	Rapporteur
M. Yves Meyer	Professeur Émérite, ENS Cachan	Président, Examinateur
Mme Laurence Likforman	Docteur, ENST-TSI, Paris	Examinateur
M. Jérôme Gilles	Docteur, DGA, Arcueil	Examinateur
M. Thomas Ea	Docteur, ISEP, Paris	Examinateur
M. Amara Amara	Professeur, ISEP, Paris	Directeur de thèse

Je dédie ce mémoire à ma famille
et à mes amis proches...

Remerciements

Avant tout, je tiens à remercier tout particulièrement ma famille et mes amis proches.

Mon père Philippe qui a toujours su être là dans tous les moments, les bons comme les difficiles ; certaines idées sont nées de conversations que nous avons eues ensemble ; son soutien et sa générosité constante ont été fondamentaux, sa droiture et la conduite scientifique de ses projets ont toujours été un modèle pour moi.

Ma mère Ghislaine qui a toujours su m'écouter et pris le temps d'entendre mes histoires de mathématiques et de biométrie ; sa douceur, sa tendresse et sa volonté ont toujours mérité mon plus profond respect.

Ma soeur Émilie qui sait ce que c'est de devoir travailler avec du stress et qui est probablement l'une des rares personnes à pouvoir se coucher aussi tard que moi quand il le faut ; notre passion commune pour la musique a toujours donné lieu à des sessions d'écoute reposantes et à des découvertes fabuleuses.

Mon frère Matthieu qui a grandi particulièrement vite ces trois dernières années et avec qui j'affiche une grande complicité ; source de bonne humeur, il m'a toujours étonné par sa créativité et sa vivacité d'esprit.

Ma (toute petite) soeur Marie, un vrai petit ange à l'éveil remarquable, mais qui sait définitivement ce qu'elle veut.

Bastien Giraud, mon collègue de thèse, mais avant tout un véritable ami ; bien que ses travaux appartiennent au domaine de la microélectronique, nous avons eu d'innombrables échanges scientifiques très constructifs ; je le remercie pour ces grandes qualités humaines et de m'avoir supporté pendant ces trois années au travail et en dehors.

Je n'oublie pas non plus *Aurélie Leloup*, la seule femme qui me connaît par cœur après ma mère et qui a connu mes tous premiers tests de biométrie, ainsi que mes amis de classes préparatoires du *Lycée Janson de Sailly*, *Clément Martin* et *Norik Elchibegian*, qui vivent aussi une aventure formidable. De manière générale, tous mes amis qui m'ont apporté un grand soutient et permis de passer de très bons moments de joie et de détente qui ont été nécessaires.

Je souhaite à présent adresser mes sincères remerciements à toutes les personnes avec qui j'ai eu la chance de travailler ou que j'ai eu l'honneur de côtoyer avant et pendant ma thèse.

Monsieur Jean-Claude Boudenot qui a su me faire confiance dès ma deuxième année de cycle ingénieur et qui a toujours été d'excellents conseils ; je lui dois de m'avoir fait ren-

contrer des personnes formidables appartenant à la communauté scientifique ainsi que d'avoir partagé des discussions passionnantes. Sa grande disponibilité, sa gentillesse, son professionnalisme et son courage en ont fait un homme remarquable et un ami de grande valeur. Sa perte m'est très douloureuse ainsi que pour la communauté de la physique.

Monsieur Jacques Girard que j'ai rencontré pour la première fois grâce aux nanotechnologies et avec qui j'ai récemment vécu l'expérience de rédiger mon premier article de revue nationale sur les ondelettes, tout en partageant des réunions au sein de la SEE.

Monsieur le Professeur Michel Terré qui a immédiatement accepté d'être mon rapporteur de thèse ; à qui je dois très certainement ma passion pour le traitement du signal et sans qui je n'en serai sûrement pas là. Sa rigueur, ses compétences techniques et sa grande pédagogie sont un exemple.

Monsieur le Professeur Patrick Garda qui a également eu la gentillesse de bien vouloir être mon rapporteur de thèse et pris le soin de relire mon manuscrit avec beaucoup d'attention.

Monsieur le Professeur Olivier Durand, mon ancien maître de stage, avec qui j'ai eu un stage révélateur qui m'a donné le goût de la recherche et pendant lequel j'ai découvert les ondelettes. Il a su faire preuve d'une pédagogie et d'une gentillesse remarquables et j'espère que nous pourrons continuer nos travaux dans le futur. J'en profite également pour saluer toute l'équipe du laboratoire *NANOCARB*, dirigé par *le Docteur Pierre Legagneux*, ainsi que celle du *LAPA* de *Thales Research & Technology* avec qui j'ai partagé des moments formidables.

Je remercie également mon directeur de thèse, *Monsieur le Professeur Amara Amara* pour m'avoir donné un sujet passionnant et porteur.

Monsieur le Docteur Thomas Ea pour sa constante bonne humeur, son soutien et toutes les idées que nous avons échangées.

Monsieur Frédéric Amiel pour son aide précieuse concernant la partie implémentation matérielle de la thèse.

Monsieur le Docteur Sami Romdhani qui a eu la gentillesse de répondre à toutes mes questions avec une grande clarté lorsque j'ai mes travaux de thèse ont débuté.

Monsieur le Docteur Jérôme Gilles avec qui j'ai récemment commencé à travailler et avec qui j'ai partagé mes dernières idées de fusion multimodale, sa disponibilité et sa gentillesse ont été des qualités que j'ai vraiment appréciées.

Enfin, il existe des rencontres-clés dans la vie d'un doctorant ; ma rencontre avec *Monsieur le Professeur émérite Yves Meyer* en fait partie ; je tiens donc à lui adresser mes plus sincères remerciements pour m'avoir accueilli plusieurs fois dans son bureau au sein de l'*ENS Cachan* afin de travailler avec moi et répondre à mes interrogations. Sa disponibilité, sa grande générosité et son humilité lui font honneur et ce fut un grand privilège pour moi que d'avoir fait sa connaissance et d'apprendre de sa sagesse.

Résumé

La biométrie se réfère à la reconnaissance automatique des individus basée sur leurs caractéristiques physiologiques et/ou comportementales. Les systèmes biométriques unimodaux permettent de reconnaître une personne en utilisant une seule modalité biométrique, mais ne peuvent pas garantir avec certitude une bonne identification. De plus, ces systèmes sont sensibles au bruit introduit par l'unique capteur, à la non-universalité et au manque d'individualité de la modalité biométrique choisie ainsi qu'aux tentatives d'intrusion. La plupart de ces problèmes peuvent être réduits par la mise en place de systèmes biométriques multimodaux utilisant plusieurs signatures biométriques d'une même personne. Dans cette thèse, nous abordons plusieurs points importants concernant la biométrie multimodale. Tout d'abord, après avoir dressé un état de l'art en fusion multimodale, nous faisons le lien entre le fonctionnement du cerveau et certains algorithmes fondamentaux utilisés en reconnaissance faciale. Ensuite, nous mettons en avant l'utilisation des ondelettes à divers niveaux du système biométrique multimodal. Enfin, l'exploration de nouvelles techniques de fusion de signatures biométriques issues du visage (modalité naturelle et non intrusive) et de l'iris (une des modalités les plus précises) ainsi que des analyses statistiques à grande échelle des scores de similarité provenant de chaque modalité ont permis de mettre au point une méthode originale de fusion adaptative combinant l'utilisation des ondelettes et des moments statistiques.

Abstract

Biometrics refers to automatic recognition of individuals based on their physiological and/or behavioral characteristics. Unimodal biometric systems allow person recognition based on a single source of biometric information but cannot guaranty a perfect identification. Such systems are sensitive to noisy sensor data, non-universality and lack of individuality of the chosen biometric trait, and susceptible to spoof attacks. Most of those problems can be alleviated by using multimodal biometric systems that combine several biometric signatures. In this thesis, we address several important issues related to multimodal biometrics. First, after describing a state of the art in multimodal fusion, we establish the link between the brain processing and some basic face recognition algorithms. Then, we underline the use of wavelets in various levels of the multimodal biometric system. Lastly, the exploration of new fusion techniques of biometric signatures deriving from face (friendly and non-invasive) and iris (one of the most accurate biometrics) modalities and large-scale statistical analyses on match-scores deriving from both modalities have led to a novel adaptive fusion method combining wavelets and statistical moments.

Table des figures

1.1 Différentes modalités biométriques. 22
1.2 Analyse Zephyr : comparaison de différentes modalités selon quatre critères principaux : l'intrusivité, le pouvoir discriminant, le coût et l'effort. . 25
1.3 Enrôlement d'une personne dans un système biométrique. 26
1.4 Authentification d'un individu dans un système biométrique. 26
1.5 Identification d'un individu dans un système biométrique. 27
1.6 Illustration du FRR et du FAR. 28
1.7 Courbe ROC. 29
1.8 Courbes CMC du CSU System 5.0 pour le "FERET Probe Set FC" et pour différents algorithmes de reconnaissance faciale. 29

2.1 Scores de compatibilité pour différentes technologies biométriques dans un système MRTD. 32
2.2 Principe de fonctionnement de base d'un système de reconnaissance faciale. 32
2.3 Une classification des algorithmes principaux utilisés en reconnaissance faciale. 36
2.4 Sources de multiples éléments de preuve d'identité dans les systèmes biométriques multimodaux [Nand 05]. 40
2.5 Fusion au niveau score dans un système biométrique multimodal. 45
2.6 Normalisation QLQ. 49
2.7 Normalisation double sigmoïde. 50
2.8 Fonction d'influence de Hampel ($a = 0.7$, $b = 0.85$ et $c = 0.95$). 51
2.9 Résumé des approches de fusion dans les systèmes biométriques multimodaux (d'après [Jain 05]). 56

3.1 La Biométrie au croisement des TIC et de la Cognition. 57
3.2 Pyramide de la perception. 58
3.3 Illusion du triangle de Kanizsa. 60
3.4 Illusion de Thatcher. 61
3.5 Mooney Face Pictures. 62
3.6 Cartographie de l'hémisphère gauche (*E. Hubbard*) avec le gyrus fusiforme. 62
3.7 Illustration de la prosopagnosie. 63
3.8 Chemin visuel dans le cerveau humain. 64
3.9 La fovea : grossièrement située dans le centre de la rétine. 64
3.10 Adaptation d'un filtre de Gabor aux données correspondant aux mesures du comportement de la réponse d'une certaine cellule simple. 65
3.11 Fonction de transfert d'un filtre de Gabor de symétrie paire de bande passante élevée. 66

3.12 Fonction de transfert de Log-Gabor pour $k = 0.04$ (échelle fréquentielle linéaire à gauche, logarithmique à droite). 67
3.13 Modèle général de la représentation de données. 69

4.1 Système photographique d'enrôlement pour les images d'iris. 72
4.2 Caméra motorisée autofocus et réseau SONY SNC-RZ50P. 72
4.3 Pan/Tilt de la caméra SONY SNC-RZ50P. 73
4.4 Exemples d'images de visage prises avec la caméra SONY SNC-RZ50P (avec l'aimable autorisation de B.Giraud, L.Simon et D.Robin). 73
4.5 Quelques exemples d'images de la base de données FERET. 74
4.6 Caméra de capture de l'iris développée par CASIA. 74
4.7 Caméra de capture de l'iris par OKI (4.7.a). Une image d'iris de CASIA-IrisV3-Interval (4.7.b). 75
4.8 Exemples images d'iris de la base de données ISEP. Pupille dilatée (4.8.a), pupille contractée (4.8.b). 76
4.9 Deux architectures possibles pour la conception d'une base de données multimodale. 79

5.1 Le principe du bootstrap par ré-échantillonnage de données indépendantes avec remplacement. 84
5.2 Schéma de construction de la matrice de similarité FSM 85
5.3 FSM : Matrice de similarité du visage (FERET FA vs. FB). 86
5.4 Schéma de construction de la matrice de similarité ISM 86
5.5 ISM : Matrice de similarité de l'iris (CASIA-IrisV3-Interval). 87
5.6 Distribution des scores imposteurs bootstrappés pour le visage. 88
5.7 Distribution des scores imposteurs bootstrappés pour l'iris. 88
5.8 Autocorrélation des scores imposteurs bootstrappés normalisés pour le visage. 92
5.9 Autocorrélation des scores imposteurs bootstrappés normalisés pour l'iris. 92
5.10 Autocorrélation d'un bruit blanc gaussien (fonction *randn* sous MATLAB). 92

6.1 Illustration de deux gaborettes. 98
6.2 Illustration du principe d'Incertitude d'Heisenberg. 99
6.3 Trois ondelettes filles de Morlet : dilatée (gauche), échelle de base (milieu) et comprimée (droite). 100
6.4 Ondelette de Morlet. 101
6.5 Neuf premières notes de "La Lettre à Élise" de Ludwig Van Beethoven. . . 102
6.6 À gauche, la transformée de Fourier. En haut à droite, l'onde sonore de l'extrait musical. En bas à droite, la transformée en ondelettes avec en abscisses, le temps, en ordonnées, la fréquence. 103
6.7 Comparaison des plans temps-fréquence (transformée de Gabor), à gauche, et temps-échelle (transformée en ondelettes), à droite. 103
6.8 Scalogramme (à droite) d'un chirp linéaire. L'amplitude des coefficients va du plus faible (couleur claire), au plus fort (couleur foncée). 104
6.9 Transformée 2D discrète en ondelettes d'une image. 105
6.10 Analyse multirésolution. 106
6.11 Exemple d'un signal régulier par morceaux (512 points). 107

6.12 Image originale de 786 Ko (6.12.a). Image compressée avec un ratio de $75:1$ pour $10,6$ Ko (6.12.b), un ratio de $150:1$ pour $5,3$ Ko (6.12.c) et un ratio de $300:1$ pour $2,6$ Ko (6.12.d). 109
6.13 Influence de la compression par ondelettes sur le taux de reconnaissance faciale (tests FERET). 111
6.14 Principe de fusion d'images par ondelettes (WIF). 113
6.15 Focus sur le premier plan (à gauche). Focus sur l'arrière-plan (au milieu). Images fusionnées par ondelettes (à droite), le premier plan et l'arrière-plan sont nets simultanément. 113
6.16 Cadre de travail générique de schémas de fusion d'images. 114
6.17 Illustration de débruitage d'une image (Lena) par ondelettes. Image originale (6.17.a). Image bruitée (6.17.b), SNR=16.46 dB. Image débruitée avec Fourier (6.17.c), SNR=18.81 dB. Image débruitée par ondelettes (6.17.d), SNR=23.04 dB. 117
6.18 Bruit blanc gaussien (10^6 échantillons) et seuil universel de Donoho. . . . 118
6.19 Débruitage par ondelettes. Seuillage "doux" (6.19.a) et seuillage dur (6.19.b). 118

7.1 Illustration du FRR et du FAR. 119
7.2 Illustration du débruitage de score par ondelettes ($\kappa = 0.92$) sur les vecteurs de score du visage et de l'iris d'un utilisateur. Scores originaux du visage (7.2.a) et de l'iris (7.2.b). Scores débruités par ondelettes du visage (7.2.c) et de l'iris (7.2.d). 121
7.3 Distance de séparation pour un utilisateur donné (Δ). 123
7.4 Coefficient de dissymétrie (γ_1). 123
7.5 Coefficient d'aplatissement (γ_2). 123

8.1 Configuration du système avec la caméra. 126
8.2 L'interface graphique utilisateur FIRST (test FERET). 127
8.3 Spectre d'énergie propre du "FERET Training Set". 128
8.4 Architecture du système avec blocs IP. 129
8.5 Architecture interne du bloc IP de reconnaissance faciale. 130
8.6 Les trois sous-matrices FSM_1, FSM_2 et FSM_3 (8.6.a) et la matrice de similarité des scores d'iris (ISM) (8.6.b), utilisées dans nos tests de fusion. . 132
8.7 Ondelette de Daubechies d'ordre 19. 133
8.8 Cadre de travail de la technique de fusion (exemple de la deuxième ligne des matrices de similarité). 134
8.9 Illustration de la technique du "*Half-Point 2D Symmetric Padding*" pour gérer les effets de bord sur une matrice (2×3). 134
8.10 Courbes ROC moyennes (normalisation MinMax). 135
8.11 Courbes ROC pour la fusion MinMax-Min. 137
8.12 Densités de probabilité pour la fusion MinMax-Min. 137
8.13 Courbes ROC pour la fusion MinMax-Max. 138
8.14 Densités de probabilité pour la fusion MinMax-Max. 138
8.15 Courbes ROC pour la fusion MinMax-Somme. 139
8.16 Densités de probabilité pour la fusion MinMax-Somme. 139
8.17 Courbes ROC pour la fusion MinMax-WSDSF. 140
8.18 Densités de probabilité pour la fusion MinMax-WSDSF. 140
8.19 Courbes ROC pour la fusion ZScore-Min. 141

8.20 Densités de probabilité pour la fusion ZScore-Min. 141
8.21 Courbes ROC pour la fusion ZScore-Max. 142
8.22 Densités de probabilité pour la fusion ZScore-Max. 142
8.23 Courbes ROC pour la fusion ZScore-Somme. 143
8.24 Densités de probabilité pour la fusion ZScore-Somme. 143
8.25 Courbes ROC pour la fusion ZScore-WSDSF. 144
8.26 Densités de probabilité pour la fusion ZScore-WSDSF. 144
8.27 Courbes ROC pour la fusion QLQ-Min. 145
8.28 Densités de probabilité pour la fusion QLQ-Min. 145
8.29 Courbes ROC pour la fusion QLQ-Max. 146
8.30 Densités de probabilité pour la fusion QLQ-Max. 146
8.31 Courbes ROC pour la fusion QLQ-Somme. 147
8.32 Densités de probabilité pour la fusion QLQ-Somme. 147
8.33 Courbes ROC pour la fusion QLQ-WSDSF. 148
8.34 Densités de probabilité pour la fusion QLQ-WSDSF. 148
8.35 Courbes ROC pour la fusion TanH-Min. 149
8.36 Densités de probabilité pour la fusion TanH-Min. 149
8.37 Courbes ROC pour la fusion TanH-Max. 150
8.38 Densités de probabilité pour la fusion TanH-Max. 150
8.39 Courbes ROC pour la fusion TanH-Somme. 151
8.40 Densités de probabilité pour la fusion TanH-Somme. 151
8.41 Courbes ROC pour la fusion TanH-WSDSF. 152
8.42 Densités de probabilité pour la fusion TanH-WSDSF. 152
8.43 Courbes ROC pour la fusion Double Sigmoïde-Min. 153
8.44 Densités de probabilité pour la fusion Double Sigmoïde-Min. 153
8.45 Courbes ROC pour la fusion Double Sigmoïde-Max. 154
8.46 Densités de probabilité pour la fusion Double Sigmoïde-Max. 154
8.47 Courbes ROC pour la fusion Double Sigmoïde-Somme. 155
8.48 Densités de probabilité pour la fusion Double Sigmoïde-Somme. 155
8.49 Courbes ROC pour la fusion Double Sigmoïde-WSDSF. 156
8.50 Densités de probabilité pour la fusion Double Sigmoïde-WSDSF. 156
8.51 Comparaison des Courbes ROC des Meilleures Techniques de Fusion Retenues. 159
8.52 Tableau de synthèse des résultats obtenus en fusion multimodale visage/iris de 2003 à 2009, incluant notre méthode de fusion multimodale WSDSF. . . 160
8.53 Situation délicate : cas 1. 161
8.54 Situation délicate : cas 2. 162
8.55 Situation délicate : cas 2 bis. 163
8.56 Situation délicate : cas 3. 163
8.57 Situation délicate : cas 3 bis. 164

A.1 Représentation d'une sphère avec la distance euclidienne (A.1.a) et la distance City-Block (A.1.b). 172
A.2 Les deux vecteurs m et n dans l'espace de Mahalanobis 173

B.1 Passage d'une image vers un vecteur dans un espace vectoriel de grande dimension. Les coefficients $a_{i,j}$ représentent les valeurs des pixels en niveau de gris, codés de 0 à 255. 175

B.2 Image moyenne et les $15^{\text{ères}}$ eigenfaces. 177
B.3 Une version simplifiée de E_v illustrant les quatre résultats de la projection d'une image sur E_v. Dans ce cas, il y a deux vecteurs propres ($u1$ et $u2$) et trois classes d'individus connus ($\Omega 1$, $\Omega 2$, $\Omega 3$). 178
B.4 Illustration du principe de séparation optimale des classes par le LDA. Trois distributions 3D sont projetées sur deux sous-espaces 2D décris par les vecteurs $\mathbf{W_1}$ et $\mathbf{W_2}$. Puisque le LDA essaye de trouver la plus grande séparation parmi les classes, on voit bien que $\mathbf{W_1}$ est ici le vecteur optimal [Baht 03]. 181
B.5 Extraction de points caractéristiques dans l'algorithme EBGM. 185
B.6 Partir réelle d'un filtre de Gabor. Vue en 3D (B.6.a), vue en 2D de dessus (B.6.b). 185
B.7 Orientation de l'ondelette. 186
B.8 Fréquence centrale de l'ondelette. 186
B.9 Phase de l'ondelette. 187
B.10 Support temporel de l'ondelette. 187
B.11 Forme de l'enveloppe gaussienne. 187
B.12 Les 80 masques d'ondelettes de Gabor utilisés dans l'algorithme EBGM. . 188
B.13 Sélection de points caractéristiques (B.13.a) et création du treillis (B.13.b). . 188
B.14 Le "*Face Bunch Graph*" sert de représentation générale pour les visages. Chaque empilement de disques représente un Jet. À partir d'un ensemble de Jets relié à un nœud du treillis, seulement celui qui présente la plus grande similarité est sélectionné pour la reconnaissance (indiqué en ombre grisée à titre d'exemple). 189

Liste des tableaux

2.1 Résumé Des Techniques de Normalisation de Scores 52

4.1 Statistiques de la base de données CASIA-IrisV3. 75

5.1 Analyses Statistiques des Scores du Visage et d'Iris 90

8.1 Temps de reconnaissance pour les différentes architectures testées. 130
8.2 Taux d'Egale Erreur (EER) moyens (%) 135
8.3 EER moyens (%) et écart-type (%) 158
8.4 Séparation entre les distributions de scores imposteurs et authentiques. . . 158
8.5 Taux de reconnaissance moyens (%) à un FAR=0.001(%). 158

B.1 Les quatres possibilités qui apparaissent lors de la phase de reconnaissance. 179

Table des matières

Table des Figures **13**

Liste des Tableaux **15**

1 Introduction Générale **21**
1.1 Pourquoi la multimodalité ? . . . 23
1.2 Systèmes biométriques et modes de fonctionnements . . . 25
1.2.1 Caractérisation d'un système biométrique . . . 25
1.2.2 Modes de fonctionnement d'un système biométrique . . . 26
1.3 Mesure de la performance d'un système biométrique . . . 28

2 Reconnaissance du Visage, de l'Iris et Fusion Multimodale **31**
2.1 Introduction . . . 31
2.2 Reconnaissance du visage . . . 31
2.2.1 Méthodes globales . . . 33
2.2.2 Méthodes locales . . . 35
2.2.3 Méthodes hybrides . . . 36
2.3 Reconnaissance de l'iris . . . 37
2.4 Fusion multimodale . . . 38
2.4.1 Conception d'un système biométrique multimodal . . . 38
2.4.2 De multiples éléments de preuve d'identité . . . 39
2.5 Les différents niveaux de fusion . . . 41
2.5.1 Avant le Matching . . . 42
2.5.1.1 Niveau Capteur (*Sensor Level*) . . . 42
2.5.1.2 Niveau Caractéristiques (*Feature Level*) . . . 42
2.5.2 Après le Matching . . . 43
2.5.2.1 Niveau Décision (*Decision Level*) . . . 43
2.5.2.2 Niveau Rang (*Rank Level*) . . . 44
2.5.2.3 Niveau Score (*Score Level*) . . . 44
2.6 La fusion au niveau score . . . 44
2.7 Normalisation de score . . . 45
2.7.1 Pourquoi normaliser les scores ? . . . 46
2.7.2 Identification d'une technique de normalisation de scores . . . 47
2.7.3 Les différentes techniques de normalisation de scores . . . 47
2.8 Approche par classification de scores . . . 52
2.9 Approche par combinaison de scores . . . 53
2.10 Conclusion . . . 54

3 Perception Visuelle et Neurosciences **57**
3.1 Introduction . . . 57
3.2 Perception des images et Théorie de la Gestalt . . . 58
3.3 Perception des visages . . . 61
3.4 Le cortex visuel primaire . . . 63
3.5 Filtres de Gabor . . . 64
3.5.1 Champs réceptifs linéaires . . . 64
3.5.2 Limites des filtres de Gabor . . . 66
3.6 Filtres de Log-Gabor . . . 67
3.7 Liens entre le fonctionnement du cerveau et certains algorithmes utilisés en reconnaissance faciale . . . 68
3.7.1 Liens avec l'Analyse en Composantes Indépendantes (ICA) . . . 68
3.7.2 Liens avec l'Elastic Bunch Graph Matching (EBGM) . . . 69
3.7.3 Liens avec l'Analyse en Composantes Principales avec Filtres de Log-Gabor (LG-PCA) . . . 70
3.8 Conclusion . . . 70

4 Systèmes d'Acquisition en Environnement Réel & Bases de Données **71**
4.1 Introduction . . . 71
4.2 Systèmes d'acquisition . . . 72
4.2.1 Système de capture d'images d'iris . . . 72
4.2.2 Système de capture d'images de visage . . . 72
4.3 Bases de Données . . . 73
4.3.1 FERET . . . 73
4.3.2 CASIA . . . 74
4.3.3 ISEP . . . 76
4.3.4 Base de Données Multimodales . . . 76
4.3.4.1 A propos des bases de données "chimériques" et des utilisateurs "virtuels" . . . 76
4.3.4.2 Conception d'une base de données multimodale . . . 78
4.4 Conclusion . . . 78

5 Analyses Statistiques des Modalités Biométriques (Iris & Visage) **81**
5.1 Introduction et définition du problème . . . 81
5.2 Le Bootstrap . . . 82
5.2.1 Quand le nombre de données est faible . . . 83
5.2.2 Principe du Bootstrap : échantillonnage avec remplacement . . . 83
5.3 Modélisation mathématique des scores . . . 84
5.3.1 Matrices de similarité du visage et de l'iris . . . 85
5.3.2 Analyse des distributions des scores imposteurs "bootstrappés" . . . 88
5.3.3 Vers le bruit blanc gaussien . . . 91
5.4 Conclusion . . . 93

6 Intégration des Ondelettes dans le Système Multimodal **95**
6.1 Introduction . . . 95
6.2 Rappel de deux transformées classiques . . . 96
6.2.1 La transformée de Fourier : "un prisme mathématique" . . . 96
6.2.2 Vers une analyse de Fourier locale : la transformée de Fourier à fenêtre 97

6.3 Les ondelettes . . . 99
6.3.1 La naissance des ondelettes . . . 99
6.3.2 Qu'est-ce qu'une ondelette ? . . . 100
6.3.3 La transformée en ondelettes : un "microscope mathématique" . . . 101
6.3.3.1 Transformée continue 1D . . . 102
6.3.3.2 Transformée discrète 2D, Analyse Multirésolution (MRA) 105
6.3.3.3 Transformée en paquets d'ondelettes 2D . . . 106
6.4 Applications . . . 107
6.4.1 Les ondelettes : un outil polyvalent . . . 107
6.4.2 Compression . . . 108
6.4.3 Analyse . . . 112
6.4.4 Débruitage . . . 115
6.5 Conclusion . . . 117

7 Fusion Adaptative Combinant Débruitage par Ondelettes et Optimisation de Moments Statistiques 119
7.1 Introduction . . . 119
7.2 Explication de la méthode WSDSF . . . 120
7.2.1 Débruitage de Score par Ondelettes . . . 120
7.2.2 Distance de séparation (Δ) . . . 122
7.2.3 Coefficient de dissymétrie : "*Skewness*" (γ_1) . . . 122
7.2.4 Coefficient d'aplatissement : "*Kurtosis*" (γ_2) . . . 122
7.2.5 Règle de Fusion de la Méthode WSDSF . . . 124
7.3 Conclusion . . . 124

8 Résultats Expérimentaux 125
8.1 Introduction . . . 125
8.2 Interface Graphique *FIRST* . . . 125
8.3 Implémentation matérielle du PCA . . . 128
8.4 Résultats de fusion multimodale . . . 131
8.4.1 Fusion multimodale 2D . . . 132
8.4.2 Fusion multimodale 1D (*WSDSF*) . . . 136
8.4.3 Situations délicates . . . 161
8.5 Conclusion . . . 164

Conclusion Générale 165

Annexes 169

A Mesures de Distance 171
A.1 Distances Euclidiennes . . . 171
A.1.1 Distance City Block (L1) . . . 171
A.1.2 Distance Euclidienne (L2) . . . 171
A.2 Distances dans l'Espace de Mahalanobis . . . 172
A.2.1 De l'espace des images à l'espace de Mahalanobis . . . 172
A.2.2 Mahalanobis L1 (MahL1) . . . 173
A.2.3 Mahalanobis L2 (MahL2) . . . 173
A.2.4 Cosinus de Mahalanobis (*MahCosine*) . . . 173

B Algorithmes PCA, LDA, EBGM **175**
B.1 Analyse en Composantes Principales (PCA) 175
B.2 Analyse Discriminante Linéaire (LDA) 181
B.3 Elastic Bunch Graph Matching (EBGM) 185

C Preuves **191**
C.1 Preuve de l'approximation de la distribution binomiale par la distribution normale 191
C.2 Preuve du seuil universel de Donoho en utilisant le Lemme de Borel-Cantelli 195

Publications et Conférences **197**

Bibliographie **209**

Chapitre 1

Introduction Générale

La biométrie trouve ses origines dans des procédés de reconnaissance *anthropométrique*, le plus ancien étant l'analyse des empreintes digitales. L'empreinte du pouce servait déjà de signature lors d'échanges commerciaux à Babylone dans l'Antiquité et en Chine au 7ème siècle. Dans une époque beaucoup plus proche, au 19ème siècle, *Alphonse Bertillon*, grand criminologiste français, invente une méthode scientifique appelée "anthropologie judiciaire" permettant l'identification de malfrats d'après leurs mesures physiologiques. De nos jours, la puissance de calcul grandissante des ordinateurs peut être mise à contribution pour reconnaître des individus, grâce à des appareils couplés à des programmes informatiques complexes.

Depuis plusieurs années, des efforts importants sont fournis dans le domaine de la recherche en biométrie. Ce constat s'explique par la présence d'un contexte mondial dans lequel les besoins en sécurité deviennent de plus en plus importants et où les enjeux économiques sont colossaux. Les applications biométriques sont nombreuses et 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, casinos, etc.).

Nous allons commencer tout naturellement par introduire quelques définitions de base puis nous détaillerons les motivations et les objectifs de cette thèse qui porte sur *la reconnaissance biométrique par fusion multimodale*.

La première question à laquelle il nous faut répondre est la suivante : *qu'est-ce que la biométrie ?* Le mot biométrie désigne dans un sens très large l'étude quantitative des êtres vivants, mais dans notre contexte plus précis de reconnaissance et d'identification d'individus, il existe deux définitions principales qui se complètent :

1. *La biométrie est la science qui étudie à l'aide de mathématiques, les variations biologiques à l'intérieur d'un groupe déterminé* [1],
2. *Toute caractéristique physique ou trait personnel automatiquement mesurable, robuste et distinctif qui peut être employé pour identifier un individu ou pour vérifier l'identité qu'un individu affirme* [2].

[1] Selon le CLUSIF (CLUb de la Sécurité des systèmes d'Information Français.)

[2] Selon la RAND Public Safety and Justice (Woodward J.D. & al., *Biometrics, A Look at Facial Recognition*, Documented Briefing prepared for the Virginia State Crime Commission.)

Le terme biométrie regroupe en fait ce que l'on appelle des *modalités biométriques* ; contrairement à ce que l'on possède et que l'on peut donc perdre (une clé) ou ce que l'on sait et que l'on peut donc oublier (un mot de passe), les modalités biométriques représentent ce que l'on est et permettent de prouver notre identité.

Pour que des caractéristiques collectées puissent être qualifiées de modalités biométriques, elles doivent être :

- *universelles* (exister chez tous les individus),
- *uniques* (permettre de différencier un individu par rapport à un autre),
- *permanentes* (autoriser l'évolution dans le temps),
- *enregistrables* (collecter les caractéristiques d'un individu avec son accord),
- *mesurables* (autoriser une comparaison future).

L'empreinte digitale, la géométrie de la main, l'iris, la rétine, le visage, l'empreinte palmaire, la géométrie de l'oreille, l'ADN, la voix, la démarche, la signature ou encore la dynamique de frappe au clavier sont autant de modalités biométriques différentes (Fig. 1.1).

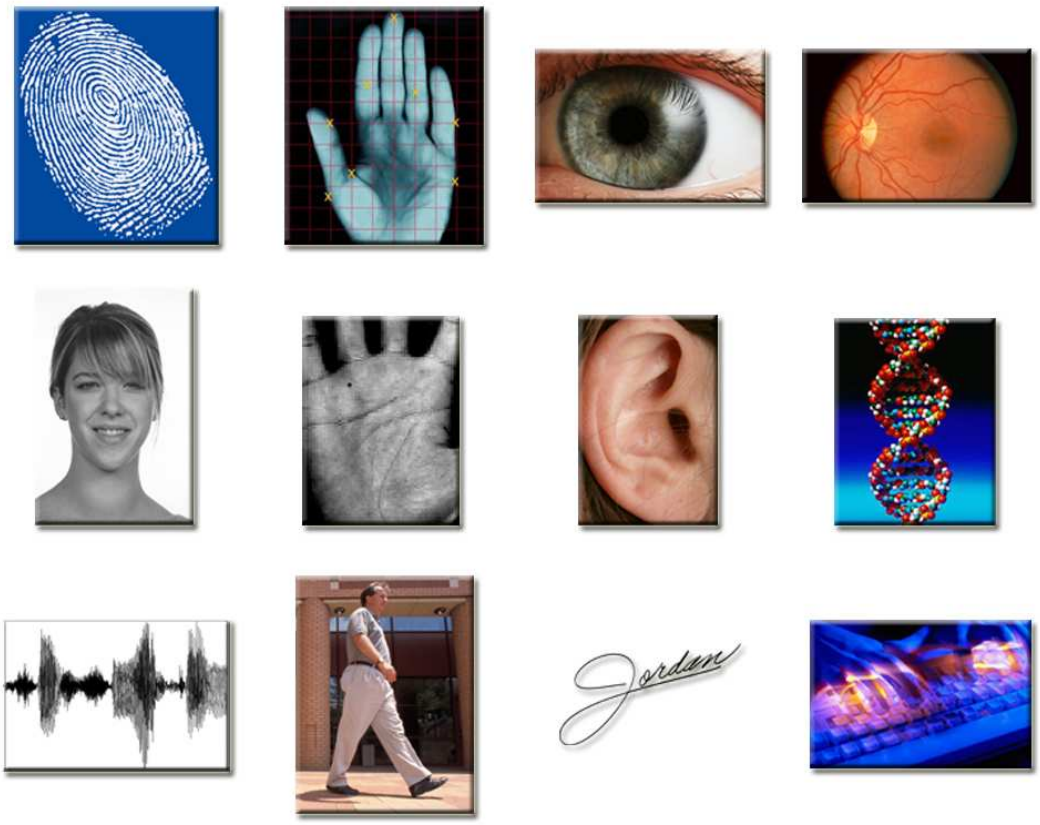


FIG. 1.1 – Différentes modalités biométriques.

1.1 Pourquoi la multimodalité ?

Bien que les techniques de reconnaissance biométrique promettent d'être très performantes, on ne peut garantir actuellement un excellent taux de reconnaissance avec des *systèmes biométriques unimodaux*, basés sur une unique signature biométrique. De plus, ces systèmes sont souvent affectés par les problèmes suivants [Jain 04] :

- *Bruit introduit par le capteur* : du bruit peut être présent dans les données biométriques acquises, ceci étant principalement dû à un capteur défaillant ou mal entretenu. Par exemple, l'accumulation de poussière sur un capteur d'empreintes digitales, un mauvais focus de caméra entraînant du flou dans des images de visage ou d'iris, etc. Le taux de reconnaissance d'un système biométrique est très sensible à la qualité de l'échantillon biométrique et des données bruitées peuvent sérieusement compromettre la précision du système [Chen 05],

- *Non-universalité* : si chaque individu d'une population ciblée est capable de présenter une modalité biométrique pour un système donné, alors cette modalité est dite *universelle*. Ce principe d'universalité constitue une des conditions nécessaires de base pour un module de reconnaissance biométrique. Cependant, toutes les modalités biométriques ne sont pas vraiment universelles. Le *National Institute of Standards and Technologies* (NIST) a rapporté qu'il n'était pas possible d'obtenir une bonne qualité d'empreinte digitale pour environ 2% de la population (personnes avec des handicaps liés à la main, individus effectuant de nombreux travaux manuels répétés, etc.) [NIST 02]. Ainsi, de telles personnes ne peuvent pas être enrôlées dans un système de vérification par empreinte digitale. De la même manière, des personnes ayant de très longs cils et celles souffrant d'anormalités des yeux ou de maladies oculaires (comme certains glaucomes et cataractes) ne peuvent fournir des images d'iris, ou de rétine, de bonne qualité pour une reconnaissance automatique. La non-universalité entraîne des erreurs d'enrôlement ("*Failure to Enroll*" ou FTE) et/ou des erreurs de capture ("*Failure to Capture*" ou FTC) dans un système biométrique,

- *Manque d'individualité* : les caractéristiques extraites à partir de données biométriques d'individus différents peuvent être relativement similaires. Par exemple, une certaine partie de la population peut avoir une apparence faciale pratiquement identique due à des facteurs génétiques (père et fils, vrais jumeaux, etc.). Ce manque d'unicité augmente le taux de fausse acceptation ("*False Accept Rate*" ou FAR) d'un système biométrique (cf. Section 1.3),

- *Manque de représentation invariante* : les données biométriques acquises à partir d'un utilisateur lors de la phase de reconnaissance ne sont pas identiques aux données qui ont été utilisées pour générer le modèle de ce même utilisateur lors de la phase d'enrôlement. Ceci est connu sous le nom de "*variations intra-classe*". Ces variations peuvent être dues à une mauvaise interaction de l'utilisateur avec le capteur (par exemple, changements de pose et d'expression faciale lorsque l'utilisateur se tient devant une caméra), à l'utilisation de capteurs différents lors de l'enrôlement et de la vérification, à des changements de conditions de l'environnement ambiant (par exemple, changements en éclairement pour un système de reconnaissance faciale)

ou encore à des changements inhérents à la modalité biométrique (par exemple, apparition de rides dues à la vieillesse, présence de cheveux dans l'image de visage, présence de cicatrices dans une empreinte digitale, etc.). Idéalement, les caractéristiques extraites à partir des données biométriques doivent être relativement invariantes à ces changements. Cependant, dans la plupart des systèmes biométriques, ces caractéristiques ne sont pas invariantes et, par conséquent, des algorithmes complexes sont requis pour prendre en compte ces variations. De grandes variations intra-classe augmentent généralement le taux de faux rejet ("*False Reject Rate*" ou FRR) d'un système biométrique (cf. Section 1.3),

– *Sensibilité aux attaques* : bien qu'il semble très difficile de voler les modalités biométriques d'une personne, il est toujours possible de contourner un système biométrique en utilisant des modalités biométriques usurpées. Des études [Mats 02] [Putt 00] ont montré qu'il était possible de fabriquer de fausses empreintes digitales en gomme et de les utiliser pour contrer un système biométrique. Les modalités biométriques comportementales telles que la signature et la voix sont plus sensibles à ce genre d'attaque que les modalités biométriques physiologiques.

Ainsi, à cause de tous ces problèmes pratiques, les taux d'erreur associés à des systèmes biométriques unimodaux sont relativement élevés, ce qui les rend inacceptables pour un déploiement d'applications critiques de sécurité. Pour pallier ces inconvénients, une solution est l'utilisation de *plusieurs modalités biométriques* au sein d'un même système, on parle alors de **système biométrique multimodal**.

Dans cette thèse, le choix a été fait de **combiner la modalité du visage avec celle de l'iris**. *Pourquoi ce choix ?* Tout d'abord, la modalité du visage est *non-intrusive*, cela veut dire qu'elle n'atteint pas l'intimité de l'individu ; elle constitue l'une des voies les plus naturelles pour reconnaître une personne, elle permet d'effectuer un travail à la volée et son coût de déploiement est relativement faible : une simple caméra reliée à un ordinateur peut suffire. Cependant, la reconnaissance faciale reste encore relativement sensible à l'environnement ambiant pour donner un taux de reconnaissance très élevé. D'autre part, la modalité de l'iris est certes plus intrusive, mais elle est actuellement considérée comme l'une des plus précises. Ce choix de combinaison de modalités est d'ailleurs confirmé par **l'analyse Zéphyr** (Fig. 1.2). De plus, un appareil de capture possédant une très forte résolution (environ 20 mégapixels) permettrait d'analyser simultanément la texture de l'iris et le visage d'une personne, sous réserve que la photographie soit prise à environ un mètre du sujet. Ce qui permettrait de travailler avec un unique système de capture au lieu de deux. Par ailleurs, les laboratoires de l'ISEP avaient déjà développé un algorithme de reconnaissance par l'iris qui donnait de très bons résultats sur une base de données interne ; il semblait donc intéressant de compléter cette modalité avec le visage afin de pouvoir effectuer des tests de fusion avec des bases de données officielles. Nous allons donc mettre en commun ces deux approches afin de créer un système multimodal, en fusionnant les signatures biométriques issues du visage et de l'iris. Le challenge que l'on doit relever est d'améliorer la performance d'un système de sécurité biométrique en trouvant une méthode de fusion originale telle que le taux de reconnaissance des modalités fusionnées soit supérieur au maximum des taux de reconnaissance des modalités prises séparément. On prendra également soin d'analyser la vitesse d'exécution de l'application ainsi que la complexité globale des calculs qui sont censées être respectivement plus lente et plus lourde que dans un système biométrique unimodal.

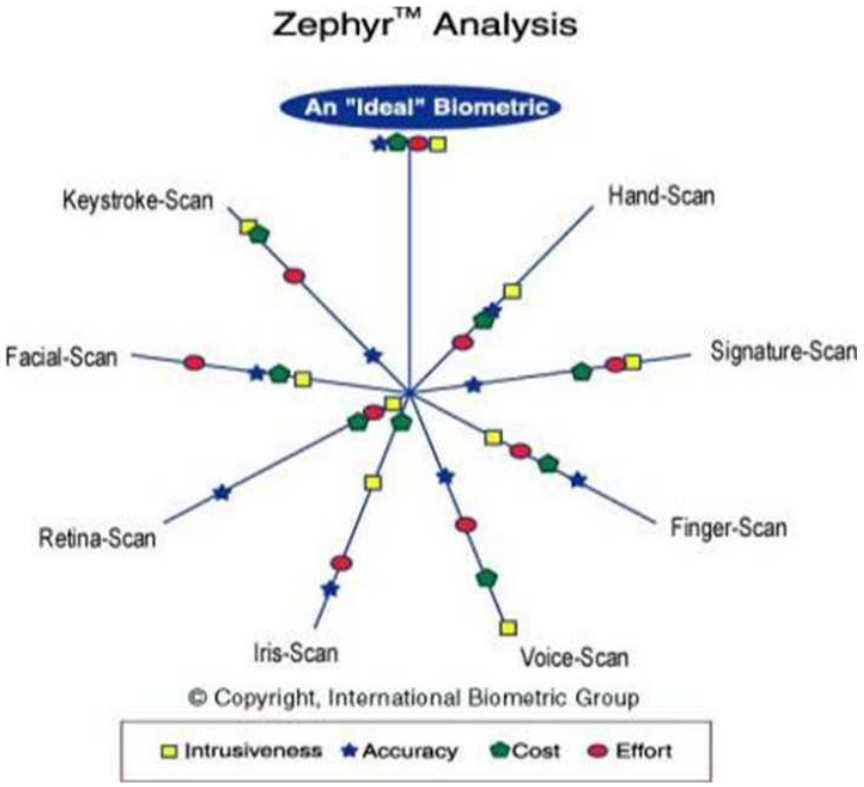


FIG. 1.2 – Analyse Zephyr : comparaison de différentes modalités selon quatre critères principaux : l'intrusivité, le pouvoir discriminant, le coût et l'effort.

1.2 Systèmes biométriques et modes de fonctionnements

1.2.1 Caractérisation d'un système biométrique

Un système biométrique typique peut être représenté par **quatre modules principaux** :

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.,
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. Idéalement, cette nouvelle représentation est censée être unique pour chaque personne et relativement invariante aux variations intra-classe,
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,
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).

1.2.2 Modes de fonctionnement d'un système biométrique

Les systèmes biométriques peuvent fournir *trois modes de fonctionnement*, à savoir, **l'enrôlement**, **l'authentification** (ou **vérification**) et **l'identification**. Dans ce qui suit, les figures illustreront l'exemple d'un système biométrique utilisant l'empreinte digitale comme modalité.

L'enrôlement (Fig. 1.3) est la première phase de tout système biométrique, il s'agit de l'étape pendant laquelle un utilisateur est enregistré dans le système pour la première fois et où une ou plusieurs modalités biométriques sont capturées et enregistrées dans une base de données. Cet enregistrement peut s'accompagner par l'ajout d'information biographique dans la base de données.

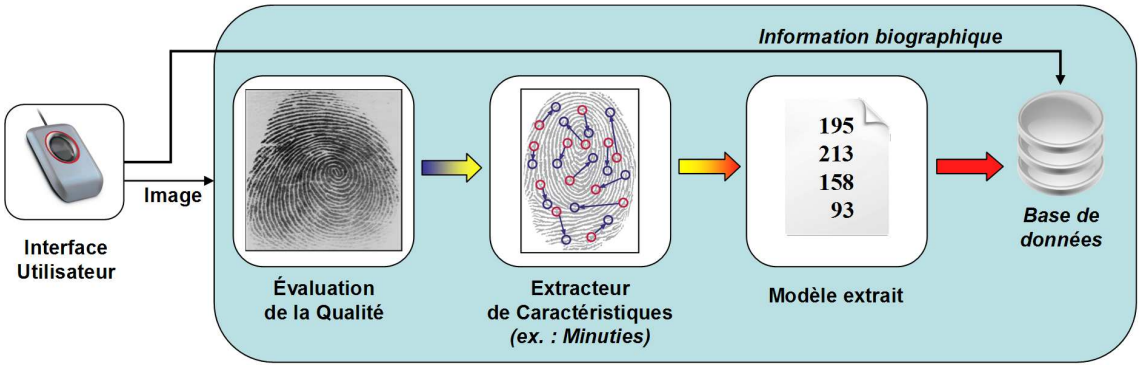


FIG. 1.3 – Enrôlement d'une personne dans un système biométrique.

Lorsqu'un système biométrique opère en **mode authentification** (Fig. 1.4), l'utilisateur affirme son identité et le système vérifie si cette affirmation est valide ou non.

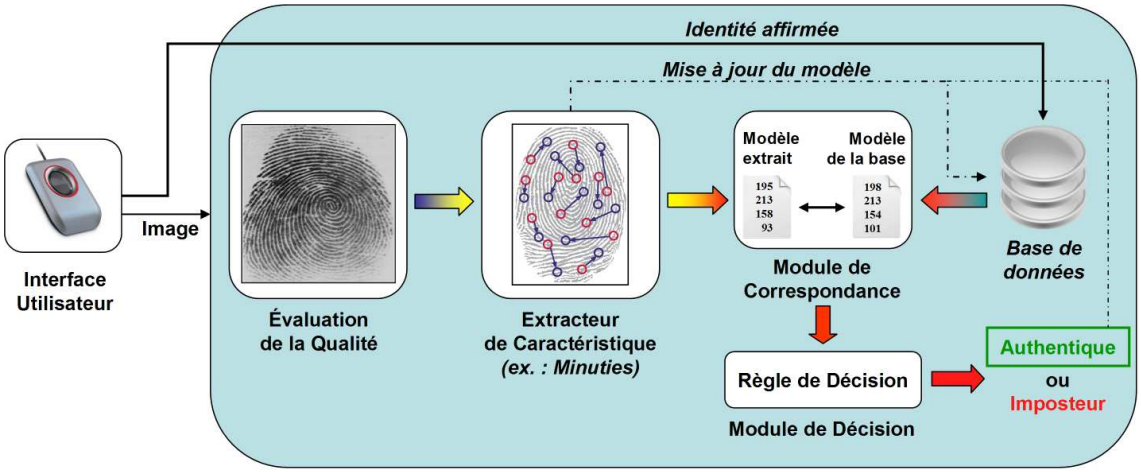


FIG. 1.4 – Authentification d'un individu dans un système biométrique.

Pour illustrer ce principe, prenons la situation où un utilisateur (*M. X*) souhaite retirer de l'argent à un distributeur de billets en entrant son code personnel d'identification (code PIN) et en présentant une modalité biométrique. Le système acquiert alors les

données biométriques et va les comparer **uniquement** avec le modèle enregistré correspondant à *M. X*. On parle alors de **correspondance 1:1**. Ainsi, si l'entrée biométrique de l'utilisateur et le modèle enregistré dans la base de données correspondant à l'identité affirmée possèdent un degré de similitude élevé, l'affirmation est validée et l'utilisateur est considéré comme étant un **authentique**. Dans le cas contraire, l'affirmation est rejetée et l'utilisateur est considéré comme étant un **imposteur**. En résumé, un système biométrique opérant en mode vérification répond à la question "*Suis-je bien M. X ?*".

Dans un système biométrique opérant en **mode identification** (Fig. 1.5), l'utilisateur ne dévoile pas explicitement son identité. Cependant, l'affirmation implicite faite par l'utilisateur est qu'elle est une des personnes déjà enrôlées par le système. Ainsi, l'échantillon biométrique de l'individu est comparé avec les modèles de **toutes les personnes** de la base de données. On parle alors de **correspondance 1:N**. La sortie du système biométrique est constituée par l'identité de la personne dont le modèle possède le degré de similitude le plus élevé avec l'échantillon biométrique présenté en entrée. Typiquement, si la plus grande similarité entre l'échantillon et tous les modèles est inférieure à un seuil de sécurité minimum fixé, la personne est **rejetée**, ce qui implique que l'utilisateur n'était pas une des personnes enrôlées par le système. Dans le cas contraire, la personne est **acceptée**.

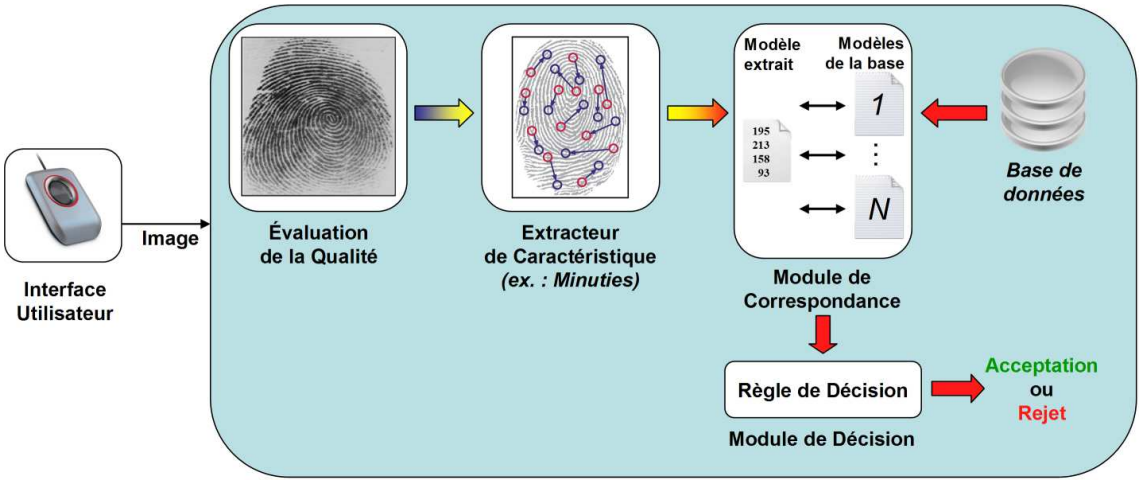


FIG. 1.5 – Identification d'un individu dans un système biométrique.

Un exemple de système opérant en mode identification serait l'accès à un bâtiment sécurisé : tous les utilisateurs qui sont autorisés à entrer dans le bâtiment sont enrôlés par le système ; lorsqu'un individu essaye de pénétrer dans le bâtiment, il doit d'abord présenter ses données biométriques au système et, selon la détermination de l'identité de l'utilisateur, le système lui accorde le droit d'entrée ou non. En résumé, un système biométrique opérant en mode identification répond à la question "*Suis-je bien connu du système ?*".

1.3 Mesure de la performance d'un système biométrique

Tout d'abord, afin de comprendre comment déterminer la performance d'un système biométrique, il nous faut définir clairement **trois critères principaux**, déjà évoqués dans la section 1.1 :

1. Le premier critère s'appelle le **taux de faux rejet** ("**F**alse **R**eject **R**ate" ou **FRR**). Ce taux représente le pourcentage de personnes censées être reconnues mais qui sont rejetées par le système,

2. Le deuxième critère est le **taux de fausse acceptation** ("**F**alse **A**ccept **R**ate" ou **FAR**). Ce taux représente le pourcentage de personnes censées ne pas être reconnues mais qui sont tout de même acceptées par le système,

3. Le troisième critère est connu sous le nom de **taux d'égale erreur** ("**E**qual **E**rror **R**ate" ou **EER**). Ce taux est calculé à partir des deux premiers critères et constitue un point de mesure de performance courant. Ce point correspond à l'endroit où $FRR = FAR$, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations.

La figure 7.1 illustre le FRR et le FAR à partir de distributions des scores authentiques et imposteurs tandis que l'EER est représenté sur la figure 1.7.

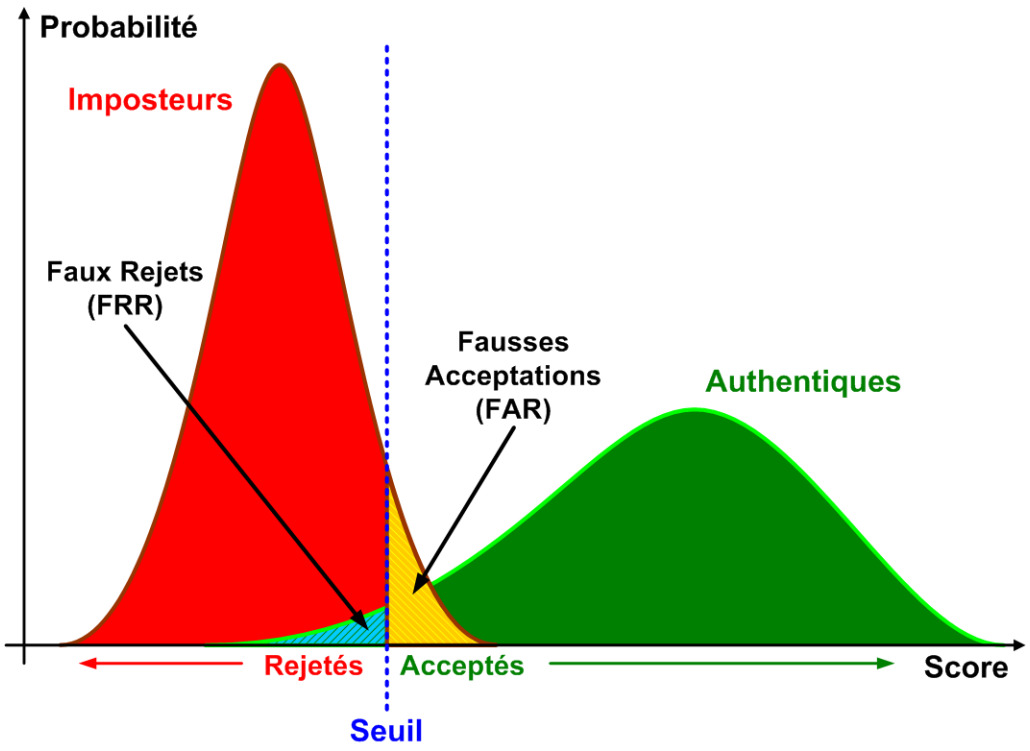


FIG. 1.6 – Illustration du FRR et du FAR.

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

– Lorsque le système opère en *mode authentification*, on utilise ce que l'on appelle une **courbe ROC** (pour "**R**eceiver **O**perating **C**haracteristic" en anglais). La courbe ROC (Fig. 1.7) trace le taux de faux rejet en fonction du taux de fausse acceptation [Perr 02]. Plus cette courbe tend à épouser la forme du repère, plus le système est performant, c'est-à-dire possédant un taux de reconnaissance global élevé,

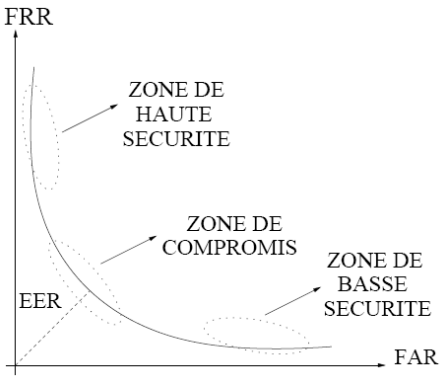


FIG. 1.7 – Courbe ROC.

– En revanche, dans le cas d'un système utilisé en *mode identification*, on utilise ce que l'on appelle une **courbe CMC** (pour "**C**umulative **M**atch **C**haracteristic" en anglais). La courbe CMC (Fig. 1.8) donne le pourcentage de personnes reconnues en fonction d'une variable que l'on appelle le **rang** [Beve 05]. On dit qu'un système reconnaît au rang 1 lorsqu'il choisit la plus proche image comme résultat de la reconnaissance. On dit qu'un système reconnaît au rang 2, lorsqu'il choisit, parmi deux images, celle qui correspond le mieux à l'image d'entrée, etc. On peut donc dire que **plus le rang augmente**, plus le taux de reconnaissance correspondant est lié à un **niveau de sécurité faible**.

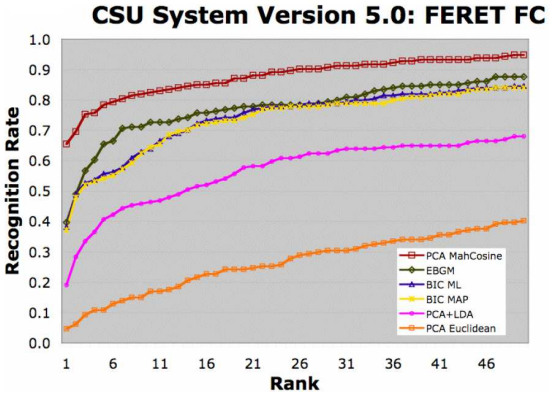


FIG. 1.8 – Courbes CMC du CSU System 5.0 pour le "FERET Probe Set FC" et pour différents algorithmes de reconnaissance faciale.

Enfin, il faut savoir que la courbe CMC n'est qu'une autre manière d'afficher la performance d'un système biométrique et peut également être calculée à partir du FAR et du FRR. Une étude comparative précisant le lien entre les courbes CMC et ROC peut être trouvée dans [Boll 05].

Tout au long de cette thèse, nous nous efforcerons d'être critiques et objectifs, en rapportant les difficultés qui ont été rencontrées avec certaines méthodes ou encore en proposant des solutions alternatives à certaines techniques. Cette thèse doit pouvoir montrer le cheminement scientifique qui s'est effectué depuis l'imprégnation de l'état de l'art jusqu'à la mise au point d'une nouvelle méthode de fusion multimodale.

Le reste de ce manuscrit est organisé de la façon suivante :

- Le **Chapitre 2** présente un état de l'art sur la fusion multimodale, en insistant d'abord sur la reconnaissance faciale et la reconnaissance par l'iris puisque nous avons choisi de fusionner ces deux modalités,

- Dans le **Chapitre 3**, nous nous concentrons sur la perception visuelle et les neurosciences afin de comprendre comment notre cerveau fonctionne pour reconnaître des visages et mieux appréhender les algorithmes de reconnaissance faciale,

- Le **Chapitre 4** traite de systèmes d'acquisition en environnement réel et des différentes bases de données qui ont été utilisés, en insistant tout particulièrement sur la conception de bases de données multimodales,

- Des analyses statistiques précises et à grande échelle des scores de similarité issus des modalités du visage et de l'iris sont détaillées dans le **Chapitre 5**, elle permettent de modéliser mathématiquement certaines distributions de données afin de donner les prérequis théoriques pour la technique de fusion,

- Le **Chapitre 6** attire notre attention sur l'intégration des ondelettes dans le système biométrique multimodal, il introduit la théorie des ondelettes et montre leur utilité à divers endroits de la chaîne biométrique concernant trois points fondamentaux en traitement du signal, à savoir la compression, l'analyse et le débruitage,

- **Une nouvelle méthode originale de fusion adaptative par combinaison de scores** est expliquée dans le **Chapitre 7** ; cette technique combine du débruitage par ondelettes avec l'optimisation d'une fonction de coût utilisant une distance de séparation et deux moments statistiques standards,

- Les résultats expérimentaux sont décrits dans le **Chapitre 8**,

Enfin, nous dresserons une conclusion globale sur ces travaux et présenterons les perspectives futures.

Chapitre 2

Reconnaissance du Visage, de l'Iris et Fusion Multimodale

2.1 Introduction

La recherche dans le domaine de la biométrie multimodale est relativement récente. Depuis 1995, date à laquelle les premiers travaux de fusion du visage et de la voix ont été menés [Brun 95], de nombreuses études ont été conduites en associant **différentes modalités**, en faisant **varier le niveau de fusion** des données et en testant **plusieurs règles de fusion** [Ross 06].

Notre système multimodal devant faire intervenir les modalités du visage et de l'iris, nous donnerons tout d'abord un rapide état de l'art en reconnaissance faciale et en reconnaissance de l'iris. Ensuite, nous analyserons plus en détail la fusion multimodale et les systèmes multimodaux avant d'expliquer les différents niveaux de fusion possibles. Enfin, nous porterons notre attention sur **la fusion au niveau score qui a été choisie pour la fusion de nos deux modalités**.

2.2 Reconnaissance du visage

La reconnaissance faciale est une tâche que les humains effectuent naturellement et sans effort dans leurs vies quotidiennes. La grande disponibilité d'ordinateurs puissants et peu onéreux ainsi que des systèmes informatiques embarqués ont suscité un énorme intérêt dans le traitement automatique des images et des vidéos numériques au sein de nombreuses applications, incluant l'identification biométrique, la surveillance, l'interaction homme-machine et la gestion de données multimédia. La reconnaissance faciale, 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 étant dû aux **avances rapides** dans des technologies telles que les appareils photo numériques, Internet et les dispositifs mobiles, le tout associé à des **besoins en sécurité** sans cesse en augmentation.

La reconnaissance faciale possède plusieurs avantages sur les autres technologies biométriques : elle est **naturelle**, **non intrusive** et **facile à utiliser**. Parmi les six attributs biométriques considérés par *Hietmeyer* [Hiet 00], les caractéristiques faciales marquent

un score de compatibilité le plus élevé dans un *système MRTD* ("*Machine Readable Travel Documents*") [Mach 08], ce score étant basé sur **plusieurs facteurs d'évaluation** tels que l'enrôlement, le renouvellement des données, les requis matériels et la perception des utilisateurs (Fig. 2.1).

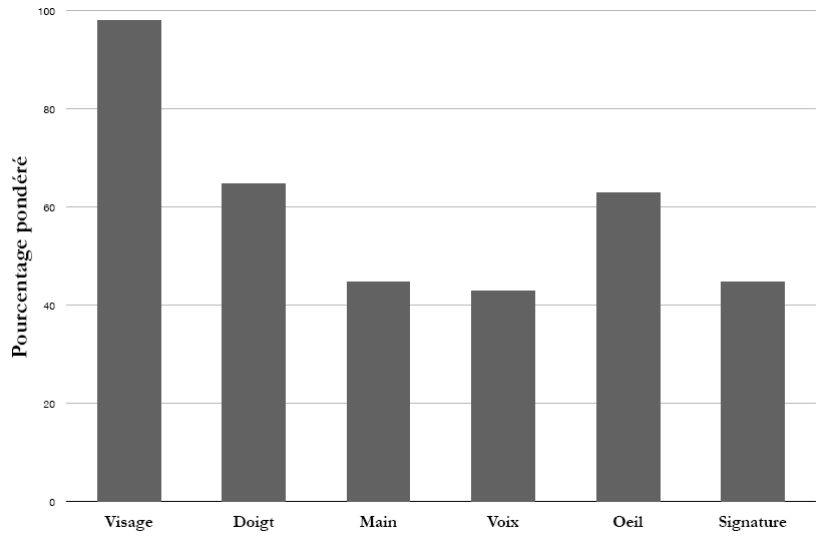


FIG. 2.1 – Scores de compatibilité pour différentes technologies biométriques dans un système MRTD.

Idéalement, un système de reconnaissance faciale doit pouvoir **identifier des visages** présents dans une image ou une vidéo **de manière automatique**. Le système peut opérer dans les deux modes suivants : authentification ou identification ; on peut également noter qu'il existe un autre type de scénario de reconnaissance faciale mettant en jeu une vérification sur une liste de surveillance ("*watch-list*"), où un individu est comparé à une liste restreinte de suspects. Le principe de fonctionnement de base d'un système de reconnaissance faciale (Fig. 2.2) peut être résumé en **quatre étapes** : les deux premières s'effectuent *en amont* du système (**détection** [Hori 06] [Kien 05] et **normalisation** du visage [Bolm 03]) et les deux dernières représentent la *reconnaissance* à proprement dit (**extraction et comparaison des caractéristiques**).

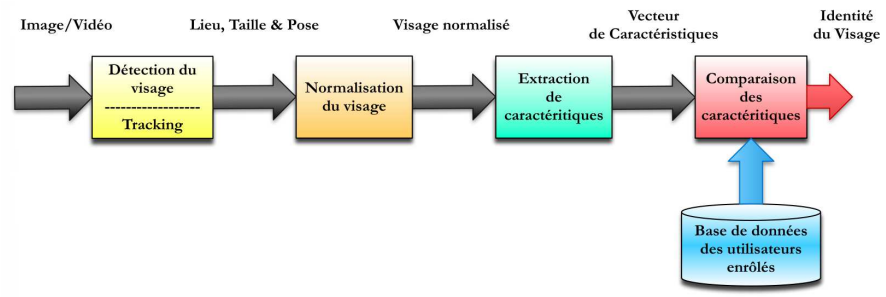


FIG. 2.2 – Principe de fonctionnement de base d'un système de reconnaissance faciale.

La performance des systèmes de reconnaissance faciale s'est significativement améliorée depuis les premiers travaux qui ont été menés dans les années 1960-70 [Bled 64] [Kell 70] [Kana 77] et de nombreux algorithmes de reconnaissance du visage ont été proposés depuis. Certaines méthodes se basent sur une **photographie (2D) du visage** tandis que d'autres méthodes prennent en compte des **informations 3D** de celui-ci. On peut également noter qu'il existe d'autres méthodes (parfois appelées 2.5D) qui effectuent la reconnaissance du visage en se basant la plupart du temps sur l'information de profondeur.

Cependant, ces dernières méthodes peuvent demander un **coup de déploiement élevé** dû à l'investissement de scanners 3D coûteux. Un autre **inconvénient majeur** concerne le **grand volume de données tridimensionnelles** qu'il est souvent **nécessaire de convertir** afin de pouvoir les traiter convenablement, ce qui implique une utilisation encore relativement inadaptée à des contraintes temps-réel, contrairement au traitement de photographies 2D. Enfin, à notre connaissance, **il n'existe pas de bases de données officielles 3D** comprenant un **nombre suffisamment élevé d'utilisateurs** pour pouvoir **évaluer** le rapport entre **la performance gagnée** par l'utilisation d'une troisième dimension et les coûts supplémentaires en termes de ressources et de temps de calcul. Par conséquent, **pour la partie concernant la reconnaissance faciale**, **nous privilégierons l'étude des méthodes 2D**.

Les méthodes de reconnaissance faciales peuvent être séparées en deux grandes familles, les **méthodes globales** (ou **holistiques**) et les **méthodes locales**, **basées sur des modèles**. Le choix a été fait de se concentrer sur ces deux types d'approches fondamentales et de n'aborder ni les *réseaux neuronaux* (NN) [Lin 97] (plus adaptés à la détection des visages), ni les *modèles cachés de Markov* (HMM) [Nefi 98] (plus utilisés en reconnaissance de la parole) car ces deux dernières techniques rencontrent des **problèmes lorsque le nombre d'individus augmente** (les **calculs deviennent très importants**) ; de plus elles **ne conviennent pas** pour les systèmes de reconnaissance basés sur une seule "image modèle" car **de nombreuses images par personne sont nécessaires pour entraîner les systèmes** afin de configurer leurs paramètres de façon "optimale". Cependant, on pourra se référer aux travaux des équipes françaises du *LE2I* [Smac 06] [Yang 03] et du *GIPSA* [Bail 06] [Giri 04], qui les ont respectivement appliquées aux images de visage et aux signaux de parole.

2.2.1 Méthodes globales

Les méthodes globales sont basées sur des **techniques d'analyse statistique bien connues**. **Il n'est pas nécessaire de repérer certains points caractéristiques du visage** (comme les centres des yeux, les narines, le centre de la bouche, etc.) **à part pour normaliser** les images. Dans ces méthodes, les images de visage (qui peuvent être vues comme des matrices de valeurs de pixels) **sont traitées de manière globale** et sont généralement transformées en vecteurs, plus faciles à manipuler.

L'avantage principal des méthodes globales est qu'elles sont **relativement rapides** à mettre en œuvre et que les calculs de base sont d'une **complexité moyenne**. En revanche, **elles sont très sensibles aux variations d'éclairement, de pose et d'expression faciale**. Ceci se comprend aisément puisque la moindre variation des conditions de l'environnement ambiant entraîne des changements inéluctables dans les valeurs des pixels qui sont traités directement.

Ces méthodes utilisent principalement une **analyse de sous-espaces de visages**. Cette expression repose sur un fait relativement simple : une classe de "*formes*" qui nous intéresse (dans notre cas, *les visages*) réside dans un sous-espace de l'espace de l'image d'entrée. Par exemple, considérons une petite image de 64×64, en niveaux de gris codés sur 8 bits (donc de 0 à 255). Cette dernière possède 4096 pixels qui peuvent exprimer un grand nombre de classes de formes, comme des arbres, des maisons ou encore des visages. Cependant, parmi les $256^{4096} > 10^{9864}$ "configurations" possibles, seulement une petite quantité peut correspondre à des visages. Ainsi, la représentation de l'image originale est très redondante et **la dimensionalité de cette représentation pourrait être grandement réduite** si l'on se concentre uniquement sur les formes qui nous intéressent. **L'utilisation de techniques de modélisation de sous-espace a fait avancer la technologie de reconnaissance faciale de manière significative**.

Nous pouvons distinguer **deux types de techniques** parmi les méthodes globales : les *techniques linéaires* et les *techniques non linéaires*.

Les techniques linéaires projettent linéairement les données d'un espace de grande dimension (par exemple, l'espace de l'image originale) sur un sous-espace de dimension inférieure. Malheureusement, ces techniques sont **incapables de préserver les variations non convexes** des *variétés* (géométriques donc au sens mathématique du terme) de visages afin de différencier des individus. Dans un sous-espace linéaire, les **distances euclidiennes** (*cf. Annexe A.1*) et plus généralement les **distances de Mahalanobis** (*cf. Annexe A.2*), qui sont normalement utilisées pour faire comparer des vecteurs de données, ne permettent pas une bonne classification entre les classes de formes "visage" et "non-visage" et entre les individus eux-mêmes. **Ce facteur crucial limite le pouvoir des techniques linéaires pour obtenir une détection et une reconnaissance du visage très précises**.

La technique linéaire la plus connue et sans aucun doute **l'analyse en composantes principales** (**PCA**), également appelée *transformée de Karhunen-Loeve*. Le PCA fut d'abord utilisé afin de représenter efficacement des images de visages humains [Siro 87] [Kirb 90]. **En 1991**, cette technique a été reprise dans le cadre plus spécifique de la reconnaissance faciale sous le nom de méthode des **Eigenfaces** [Turk 91] (on en trouvera une explication approfondie en *Annexe B.1*).

Cependant, le PCA classique nécessite que les images de visage soient mises sous formes de vecteurs, ce qui a pour effet de détruire la structure géométrique de l'image. Pour ne pas perdre les informations de voisinage lors du passage de l'image en vecteur, une méthode PCA bi-dimensionnelle (**2-D PCA**) [Jian 04] a été étudiée. Cette méthode prend en entrée des images et non plus des vecteurs.

Il existe d'autres techniques également construites à partir de *décompositions linéaires* comme **l'analyse discriminante linéaire** (**LDA**) [Belh 97] ou encore **l'analyse en composantes indépendantes** (**ICA**) [Bart 02].

Tandis que le PCA construit un sous-espace pour représenter de manière "optimale" (mathématiquement parlant) seulement "l'objet" visage, le LDA construit un sous-espace discriminant pour distinguer de façon "optimale" les visages de différentes personnes. Elle permet donc d'effectuer une véritable séparation de classes (une explication détaillée du LDA pourra être consultée en *Annexe B.2*). Des études comparatives (par exemple, [Belh 97]) montrent que les méthodes basées sur le LDA donne généralement de meilleurs résultats que les méthodes basées sur le PCA.

L'algorithme ICA, quant à lui, est une généralisation de l'algorithme PCA avec lequel il coïncide dans le cas de **données gaussiennes**. L'algorithme ICA est basé sur le concept intuitif de *contraste* et permet d'éliminer la redondance statistique des données de départ. Ce dernier a été rendu célèbre notamment avec l'expérience de la "*cocktail party*" [Brow 01] [Sagi 01] [Hayk 05] qui met en avant la résolution d'un problème de séparation (décorrélation) de sources audio. *Bartlett et al.* ont fourni *deux architectures différentes pour l'ICA* : une première architecture (**ICA I**) qui construit une **base d'images statistiquement indépendantes** et une deuxième architecture (**ICA II**) qui fournit une **représentation en code factoriel des données**.

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

Afin de pouvoir traiter ce problème de non-linéarité en reconnaissance faciale, de telles méthodes linéaires ont été étendues à des **techniques non linéaires** basées sur la notion mathématique de **noyau** ("*kernel*") comme le **Kernel PCA** [Scho 98] et le **Kernel LDA** [Mika 99]. Ici, une **projection non linéaire** (réduction de dimension) de l'espace de l'image sur l'espace de caractéristiques ("*feature space*") est effectuée ; les variétés présentes dans l'espace de caractéristiques résultant **deviennent simple**, de même que **les subtilités des variétés qui sont préservées**. Bien que les méthodes basées sur le noyau peuvent atteindre une bonne performance sur les données d'entraînement, il ne peut pas en être de même pour de nouvelles données en raison de leur plus grande flexibilité ; contrairement aux méthodes linéaires.

2.2.2 Méthodes locales

Les méthodes locales, **basées sur des modèles**, utilisent des **connaissances a priori** que l'on possède sur la **morphologie du visage** et s'appuient en général sur des **points caractéristiques** de celui-ci. Kanade présenta un des premiers algorithmes de ce type [Kana 73] en détectant certains points ou traits caractéristiques d'un visage puis en les comparant avec des paramètres extraits d'autres visages. Ces méthodes constituent une autre approche pour **prendre en compte la non-linéarité** en construisant un **espace de caractéristiques local** et en utilisant des filtres d'images appropriés, de manière à ce que les distributions des visages soient moins affectées par divers changements.

Les **approches Bayesiennes** (comme la méthode **BIC** [Liu 98]), les **machines à vecteurs de support** (**SVM**) [Guo 00], la méthode des **modèles actifs d'apparence** (**AAM**) [Coot 01] ou encore la méthode **"local binary pattern"** (**LBP**) [Ahon 04] ont été utilisées dans ce but.

Toutes ces méthodes ont l'avantage de pouvoir **modéliser plus facilement les variations de pose, d'éclairage et d'expression par rapport aux méthodes globales**. Toutefois, elles sont plus lourdes à utiliser puisqu'il faut souvent placer manuellement un assez grand nombre de points sur le visage alors que les méthodes globales ne nécessitent de connaître que la position des yeux afin de normaliser les images, ce qui peut être fait automatiquement et de manière assez fiable par un algorithme de détection [Arca 06].

2.2.3 Méthodes hybrides

Les méthodes hybrides permettent d'**associer les avantages des méthodes globales et locales** en combinant **la détection de caractéristiques géométriques** (ou structurales) **avec l'extraction de caractéristiques d'apparence locales**. Elles permettent d'**augmenter la stabilité de la performance** de reconnaissance lors de changements de pose, d'éclairement et d'expressions faciales.

L'analyse de caractéristiques locales (**LFA**) [Pene 96] et les **caractéristiques extraites par ondelettes de Gabor** (comme l'**Elastic Bunch Graph Matching**, **EBGM** [Wisk 97], dont on trouvera une présentation plus détaillée en *Annexe B.3*), sont des algorithmes hybrides typiques.

Plus récemment, l'algorithme **Log Gabor PCA** (**LG-PCA**) [Perl 05] effectue une convolution avec des **ondelettes de Gabor orientées** autour de certains points caractéristiques du visage afin de créer des vecteurs contenant **la localisation et la valeur d'amplitudes énergétiques locales** ; ces vecteurs sont ensuite envoyés dans un **algorithme PCA** afin de **réduire la dimension des données**.

La figure 2.3 fournit une classification des algorithmes principaux de reconnaissance faciale.

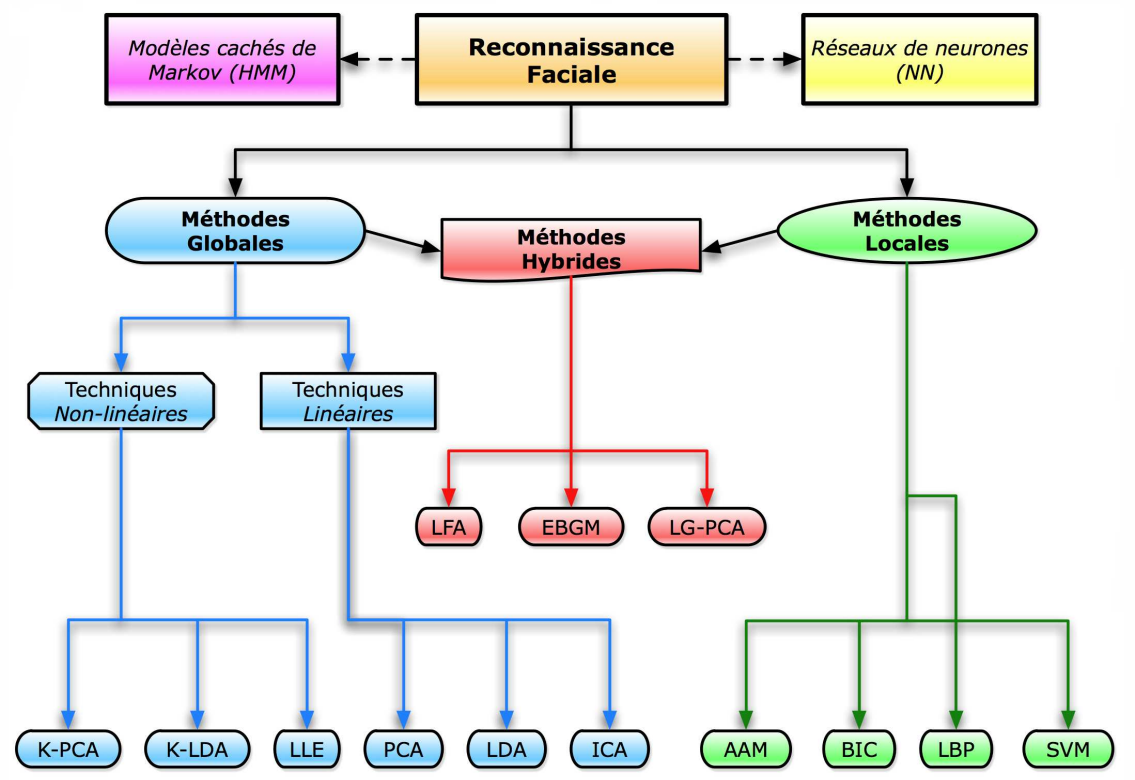


FIG. 2.3 – Une classification des algorithmes principaux utilisés en reconnaissance faciale.

2.3 Reconnaissance de l'iris

La première personne à s'être officiellement rendu compte des possibilités qu'offre la texture de l'iris en tant qu'outil d'identification était l'ophtalmologiste américain *Frank Burch*, qui a proposé cette méthode lors d'une conférence pour l'*American Academy of Ophthalmology*, en **1936**. Avant cela, l'iris n'était principalement considéré que pour sa couleur. Pendant plusieurs années, l'idée de Burch a été reproduite dans les manuels d'ophtalmologie mais il n'y avait que très peu de recherche effectuée, avec un manque tout particulier dans le domaine de l'ingénierie.

En 1987, deux ophtalmologistes américains, *Leonard Flom* et *Aron Safir*, ont tenté de transposer l'idée de Burch en voulant développer un système d'identification opérationnel. Malheureusement, ils n'ont pas été capables de mettre au point une méthode de classification des différences entre les iris.

Le Professeur *John Daugman*, alors professeur à l'Université Harvard (et maintenant à l'Université de Cambridge), est venu les aider en mettant au point une approche mathématique pour l'analyse des motifs aléatoires de l'iris. La collaboration entre Daugman, Sarin et Flom a abouti à un **prototype fonctionnel breveté en 1994**. Ce système a depuis largement été amélioré et, de nos jours, plusieurs sociétés utilisent, sous licence, ces algorithmes dans leurs applications. Le brevet est maintenant la propriété de la société *Iridian*. Le travail de Daugman est basé sur **l'analyse par ondelettes complexes de Gabor non orthogonales**. Pour simplifier, nous dirons que des filtres à valeur complexe sont appliqués sur la texture de l'iris en deux dimensions et l'information de phase est extraite pour former la signature [Daug 04] [Daug 06].

En 1997, *Wildes* présente un système basé sur une **représentation pyramidale multirésolution des textures d'iris** [Wild 97]. L'idée qui se cache derrière les pyramides d'image est la suivante : une image peut être représentée comme une approximation sous-échantillonnée et une ou plusieurs images résiduelles à différentes résolutions. Le système de *Wildes* utilise quatre sous-images à partir d'une pyramide multirésolution comme signature et une mesure de corrélation normalisée pour la classification.

En 1998, *Boles* et *Boashash* ont proposé un algorithme basé sur les "**zero-crossings**" **d'une transformée unidimensionnelle par ondelettes orthogonales** [Bole 98]. Le signal qui est transformé est également unidimensionnel et est acquis en enregistrant les valeurs en niveaux de gris d'un ou plusieurs cercles concentriques de l'iris. Le rayon des cercles est normalisé par rapport au rayon de l'iris, assurant que les mêmes points de la texture de l'iris sont échantillonnés indépendamment de la taille de l'iris dans l'image.

En 2000, *Zhu et al.* ont fait des **études comparatives** entre les ondelettes réelles de Gabor non orthogonales et l'analyse par ondelettes orthogonales de Daubechies d'ordre 4 ("db4") [Zhu 00]. **Des ondelettes de Gabor multidirectionnelles et de différentes fréquences sont utilisées pour séparer la texture de l'iris en différentes sous-bandes de fréquence**. La moyenne et l'écart-type sont utilisés comme signature. Dans le cas de l'analyse par ondelettes orthogonale, la moyenne et l'écart-type des sous-bandes générées à partir d'une transformée sur cinq niveaux sont utilisés comme signature. **Le résultat est très nettement en faveur de l'analyse par ondelettes de Gabor**. Dans des recherches

plus récentes [Ma 02], une variation symétrique circulaire de l'ondelette réelle de Gabor est utilisée.

En 2001, *Lim et al.* utilisent une **transformée par ondelettes orthogonales avec une simple ondelette de Haar** [Lim 01]. **Quatre niveaux de transformation sont calculés** et la signature est composée des coefficients des détails diagonaux à l'échelle la plus basse et de chaque valeur moyenne des coefficients de détails diagonaux des trois autres sous-bandes aux niveaux supérieurs. Les coefficients de signature sont tous de moyenne nulle, et sont tronqués à seulement un bit en fonction de leur signe. **Une approche par réseau de neurones** est utilisée pour faire correspondre deux signatures.

En 2002, *Tisse et al.* utilisent la **transformée de Hilbert** [Tiss 02]. La "fréquence émergente" et la phase instantanée (combinaison du signal original et de la transformée de Hilbert) servent à générer la signature ; la **distance de Hamming est utilisée** pour comparer deux signatures.

En 2004, *Rydgren et al.* utilisent une **transformée en paquets d'ondelettes sur trois niveaux** en l'appliquant à l'image déroulée de l'iris [Rydg 04] afin d'extraire une signature codée sur 1.5 bits (-1, 0 ou 1).

En 2005, *Noh et al.* utilisent une méthode appelée **multirésolution ICA** [Noh 05]. En calculant la valeur en niveaux de gris moyenne comme une fonction de rayon, ils construisent une signature d'iris unidimensionnelle qui sert d'entrée à l'algorithme multirésolution ICA.

Enfin, on trouvera une **présentation de quelques méthodes actuelles en reconnaissance de l'iris** dans [Daug 07] ainsi qu'une très récente étude globale qui couvre le développement historique ainsi que l'état de l'art actuel dans la compréhension du traitement d'images en reconnaissance de l'iris dans [Bowy 08].

2.4 Fusion multimodale

Les humains se reconnaissent entre eux à partir de **plusieurs caractéristiques biométriques** (physiques ou comportementales) **associées à de nombreux détails contextuels de l'environnement**. Comme il a été annoncé dans l'introduction concernant les systèmes unimodaux, chaque modalité en soi ne peut pas toujours être utilisée de manière fiable pour effectuer la reconnaissance. Cependant, la consolidation d'informations présentées par les différentes modalités peut permettre une authentification (ou vérification) précise de l'identité.

On peut donc s'attendre à ce que les systèmes biométriques multimodaux soient **plus performants** et ceci grâce à la présence de **multiples éléments de preuve** [Hong 99], un peu à la manière d'une *enquête policière*.

2.4.1 Conception d'un système biométrique multimodal

Un schéma de fusion d'informations dans le contexte de la biométrie soulève plusieurs questions concernant la conception du système. Le premier point important est

la **conception d'une Interface Homme-Machine (IHM) adéquate** afin de permettre une **acquisition efficace des informations biométriques** d'un individu, tout en **minimisant l'inconvenance** pour l'utilisateur [Ovia 03]. Le passage d'un système unimodal à un système multimodal impose que l'on prenne un **soin tout particulier à la conception de l'IHM**.

Il existe également d'**autres facteurs qui impactent** sur le design et la structure d'un système biométrique multimodal :

- **Les avantages liés au coût** : quel est le compromis entre le coût supplémentaire et l'amélioration de la performance du système ? Le coût est fonction du nombre de capteurs déployés, du temps pris pour acquérir les données biométriques, des exigences de stockage, du temps de traitement de l'algorithme et du degré de commodité perçue par l'utilisateur,

- **La détermination des sources d'information biométrique** : Quelles sont les différentes sources d'information biométrique à utiliser et quelles en sont les plus pertinentes pour un type d'application donné ?

- **L'acquisition et la séquence de traitement** : les modalités biométriques doivent-elles être acquises simultanément ou en série ? De la même manière, les informations acquises doivent-elles être traitées en même temps ou bien séquentiellement ?

- **Le type d'information** : quels types d'informations ou d'attributs (caractéristiques, scores de correspondance, décision, etc.) doivent être fusionnés ? Quel est l'impact de la corrélation parmi les sources d'information sur la performance du système de fusion ?

- **La méthode de fusion** : Quel schéma de fusion choisir pour combiner les informations présentées par de multiples sources biométriques ? Est-il possible de prédire le gain de performance obtenu en utilisant différentes méthodes de fusion dans le but d'en déterminer une optimale ?

Ce que l'on peut retenir : alors qu'il est tout a fait possible de quantifier le coût additionnel des capteurs et l'augmentation du temps de reconnaissance, **il est relativement difficile de quantifier la capacité d'un système multimodal à dissuader d'éventuels imposteurs de lancer une attaque frauduleuse** ("*spoof attacks*"). Ensuite, **l'avantage d'un système biométrique multimodal** est souvent évalué à partir de **sa précision de reconnaissance**, du **nombre d'utilisateurs qui peuvent être pris en compte par le système**, du **coût résultant par l'ajout de nouveaux capteurs** et du **temps supplémentaire requis pour acquérir et traiter plusieurs modalités pendant la phase d'enrôlement et de reconnaissance**. Enfin, il faut noter qu'en général, la fusion de deux classifieurs faibles décorrélés (ou corrélés négativement) peut donner de **meilleurs résultats** que la fusion de deux classifieurs plus forts corrélés positivement [Ross 06].

2.4.2 De multiples éléments de preuve d'identité

Quelles sont les *sources* d'information qui peuvent être considérées dans un système biométrique multimodal ? En fait, lorsque l'on parle de **système multimodal au sens large**, il existe de **nombreux scénarios possibles** (Fig. 2.4) :

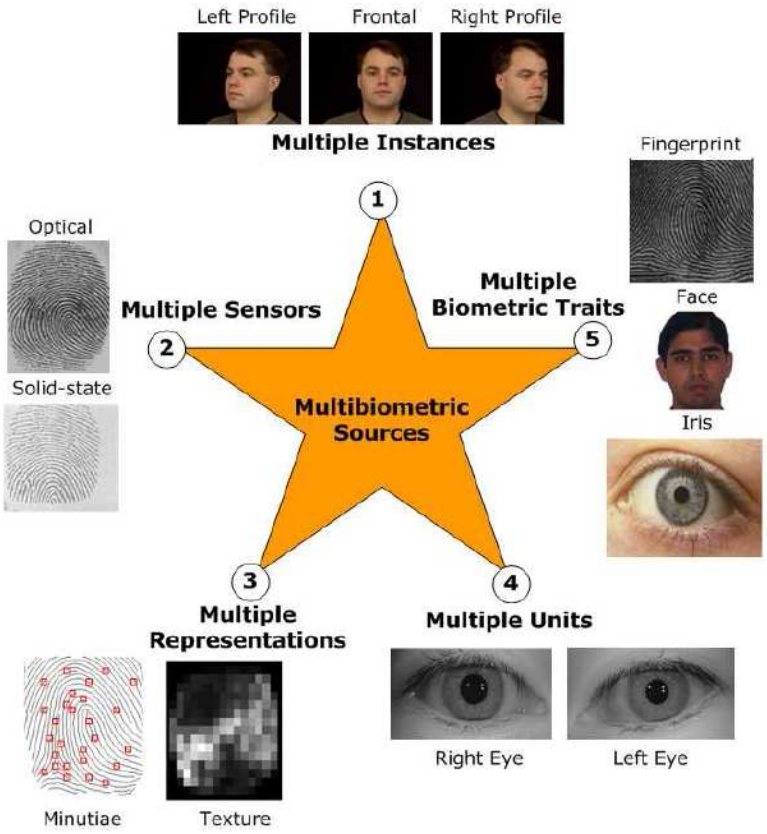


FIG. 2.4 – Sources de multiples éléments de preuve d'identité dans les systèmes biométriques multimodaux [Nand 05].

1. **Systèmes multi-échantillons** : Un unique capteur peut être utilisé pour acquérir plusieurs échantillons du même trait biométrique dans le but de prendre en compte les variations qui peuvent se produire au sein de ce trait, ou pour obtenir une représentation plus complète du caractère sous-jacent. Par exemple, un système de reconnaissance faciale peut capturer (et enregistrer) le profil frontal du visage d'une personne ainsi que les profils gauches et droits afin de tenir compte des variations de la pose faciale,

2. **Systèmes multi-capteurs** : dans ces systèmes, un même trait biométrique est analysé à l'aide de plusieurs capteurs afin d'extraire diverses informations provenant de l'enregistrement des images. Par exemple, un système peut enregistrer le contenu de la texture 2D du visage d'une personne avec une caméra CCD et la forme de la surface 3D du visage avec une autre gamme de capteurs dans le but de procéder

à la reconnaissance. Dans ce cas, c'est l'introduction des capteurs 3D servant à mesurer la variation de la surface du visage qui est responsable de l'augmentation du coût du système biométrique multimodal,

3. **Systèmes multi-algorithmes** : dans ces systèmes, les mêmes données biométriques sont traitées à travers plusieurs algorithmes. Par exemple, des algorithmes d'analyse de texture et de minuties peuvent être associés pour traiter la même image d'empreinte digitale afin d'extraire diverses caractéristiques qui peuvent améliorer la performance du système [Ross 03]. Ainsi, ce genre de système ne nécessite pas de capteurs supplémentaires et n'oblige pas l'utilisateur à interagir avec de multiples capteurs, d'où l'amélioration de la commodité d'utilisation,

4. **Systèmes multi-instances** : ces systèmes utilisent tout simplement plusieurs instances d'un même trait biométrique. Par exemple, les iris gauches et droits d'un individu peuvent être utilisés afin de vérifier son identité. Ces systèmes ne nécessitent généralement pas l'introduction de nouveaux capteurs, pas plus qu'ils n'entraînent le développement de nouveaux algorithmes d'extraction de caractéristiques ou de reconnaissance et sont, par conséquent, rentables. A titre d'information, les systèmes automatisés d'identification d'empreintes digitales ("Automated Fingerprint Identification Systems", AFIS) tirent profit de capteurs capables d'acquérir rapidement les empreintes des dix doigts d'un utilisateur,

5. **Systèmes multimodaux** : au **sens strict** du terme, **ces systèmes vont particulièrement attirer toute notre attention** car ceux sont eux qui **permettent de combiner les preuves présentées par différentes modalités biométriques afin d'établir l'identité d'un individu**. Par exemple, l'un des premiers systèmes biométriques multimodaux utilise les caractéristiques du visage de et la voix [Brun 95]. On s'attend à ce que des **traits biométriques décorrélés** (comme les empreintes digitales et l'iris) **fournissent une nette amélioration de la performance d'un système que des traits biométriques corrélés** (comme la voix et les mouvements des lèvres). Le **coût de déploiement** de ce genre de systèmes est plus dû à l'**introduction de nouveaux capteurs** et, par conséquent, au **développement d'interfaces utilisateur appropriées**. La **précision en reconnaissance peut significativement être améliorée en utilisant un nombre croissant de traits biométriques**, bien que le **phénomène problématique de la dimensionalité grandissante** ("*curse-of-dimensionality*") devrait imposer une limite à ce nombre. Ce problème de dimensionalité limite le nombre d'attributs (ou de caractéristiques) utilisés dans un système de classification de formes lorsque l'on possède seulement un faible nombre d'échantillons d'entraînement [Jain 04]. **Le nombre de traits biométriques utilisés dans une application spécifique est également limité par des considérations pratiques** comme le coût de déploiement, le temps d'enrôlement, le temps de retour ("*throughput time*") ou encore le taux d'erreur attendu.

Enfin, on peut noter que l'on utilise le terme de *systèmes hybrides* [Chan 05] pour se référer aux systèmes qui intègrent un sous-ensemble des 5 scénarios que nous venons d'évoquer. On trouvera également un **historique complet de nombreux systèmes biométriques multimodaux qui ont été développés** dans [Ross 06].

2.5 Les différents niveaux de fusion

Dans un système typique de reconnaissance de formes, **la quantité d'informations disponible devient compressée à mesure que l'on progresse du module de capture vers le module de décision**. Dans un système biométrique multimodal, la fusion peut se faire en utilisant l'information disponible dans n'importe quel de ces modules. Nous allons maintenant détailler ces niveaux de fusion que l'on peut répartir en deux grandes familles, la fusion avant la correspondance ("matching") et la fusion après la correspondance [Sand 02].

2.5.1 Avant le Matching

Avant le matching, l'intégration d'informations peut avoir lieu soit au *niveau capteur*, soit au *niveau caractéristiques*.

2.5.1.1 Niveau Capteur (*Sensor Level*)

Les données brutes ("*raw data*") provenant des capteurs sont combinées par fusion au niveau capteur [Lyen 95]. La fusion au niveau capteur peut se faire uniquement si les diverses captures sont des instances du même trait biométrique obtenu à partir de plusieurs capteurs compatibles entre eux ou plusieurs instances du même trait biométrique obtenu à partir d'un seul capteur. De plus, les captures doivent être compatibles entre elles et la correspondance entre les points dans les données brutes doit être connue par avance. Par exemple, les images de visage obtenues à partir de plusieurs caméras peuvent être combinées pour former un modèle 3D du visage. Un autre exemple de fusion au niveau capteur consiste à mettre en mosaïque plusieurs images d'empreintes digitales afin de former une image d'empreinte digitale finale plus complexe [Ross 02] [Moon 04]. La fusion au niveau capteur n'est généralement pas possible si les instances des données sont incompatibles (par exemple, il est peut être difficile de fusionner des images de visages provenant de caméras ayant des résolutions différentes).

2.5.1.2 Niveau Caractéristiques (*Feature Level*)

La fusion au niveau caractéristiques consiste à **combiner différents vecteurs de caractéristiques** ("*feature vectors*") qui sont obtenus à partir d'une des sources suivantes : plusieurs capteurs du même trait biométrique, plusieurs instances du même trait biométrique, plusieurs unités du même trait biométrique ou encore plusieurs traits biométriques.

Quand les vecteurs de caractéristiques sont **homogènes** (par exemple, plusieurs images d'empreinte digitale du doigt d'un utilisateur), un unique vecteur de caractéristiques résultant peut être calculé comme une somme pondérée des vecteurs de caractéristiques individuels.

Lorsque les vecteurs de caractéristiques sont **hétérogènes** (par exemple, des vecteurs de caractéristiques de différentes modalités biométriques comme le visage et la géométrie de la main), nous pouvons les concaténer pour former un seul vecteur de caractéristiques. Cependant, la concaténation n'est pas possible lorsque les ensembles de caractéristiques sont incompatibles. Par exemple, les minuties d'empreintes digitales et les coefficients de visage issus du PCA ("*eigen-face coefficients*").

Les tentatives de *Kumar et al.* [Kuma 03] qui ont combiné des caractéristiques de l'empreinte palmaire et de la géométrie de la main d'une part, et *Ross et Govindarajan* [Ross 05] qui ont combiné des caractéristiques du visage avec celles de la géométrie de la main d'autre part, n'ont rencontré qu'un **succès limité**.

Les systèmes biométriques qui intègrent l'information à une étape **en amont** du traitement sont **censés être plus efficaces** que les systèmes qui opèrent une fusion à un niveau **plus abstrait**. Puisque les caractéristiques issues d'une entrée biométrique sont supposées contenir une information plus riche qu'un score de correspondance ou la décision d'un matcher (module de reconnaissance) biométrique, la fusion au niveau caractéristiques devrait fournir de meilleurs résultats de reconnaissance que les autres niveaux d'intégration.

Cependant, la fusion au niveau caractéristiques est **difficile à atteindre en pratique** à cause des raisons suivantes :

1. **La relation entre les espaces de caractéristiques** ("*feature spaces*") de différents systèmes biométriques **n'est pas forcément connue**. Dans le cas ou la relation est connue par avance, on doit prendre soin d'éliminer les caractéristiques qui sont fortement corrélées. Cela requiert l'application d'algorithmes de sélection de caractéristiques avant l'étape de classification,

2. **La concaténation de deux vecteurs de caractéristiques peut engendrer un vecteur de caractéristiques ayant une grande dimension**, menant au fameux problème de la "*malédiction de la dimensionalité*" (cf. Chapitre 2.4.2). Bien que ce soit un problème général dans la plupart des applications de reconnaissance de forme, cela est encore plus marquant dans les applications biométriques à cause du temps, de l'effort et du coût impliqués dans la collecte de grandes quantités de données biométriques,

3. **La plupart des systèmes biométriques commerciaux ne fournissent pas l'accès aux vecteurs de caractéristiques qui sont utilisés dans leurs produits**. Ainsi, **très peu de chercheurs ont étudié la fusion au niveau caractéristiques** et la plupart d'entre eux se tournent généralement vers les **schémas de fusion après le matching**.

2.5.2 Après le Matching

Les schémas d'intégration de l'information après l'étape de la classification ou de correspondance peuvent être divisés en quatre catégories : **sélection dynamique de classifieurs**, **fusion au niveau décision**, **fusion au niveau rang** et **fusion au niveau score**.

Un schéma de sélection dynamique de classifieurs choisit les résultats d'un classifieur qui est le plus à même de donner la décision correcte d'un modèle d'entrée spécifique [Wood 97]. Ceci est aussi connu sous le nom d'approche "*winner-take-all*" et le système qui effectue cette sélection est appelée "*associative switch*" [Chen 97].

2.5.2.1 Niveau Décision (*Decision Level*)

L'intégration d'information au niveau abstrait ou au niveau décision peut être mis en place lorsque chaque matcher biométrique décide individuellement de la meilleure correspondance possible selon l'entrée qui lui est présentée. Les méthodes comme le "*majority voting*" [Lam 97], le "*behavior knowledge space*" [Lam 95], le "*weighted voting*" basé sur la *théorie Dempster-Shafer* [Xu 92], *les règles ET et OU* [Daug 98], etc. peuvent utilisées afin d'arriver à la décision finale.

2.5.2.2 Niveau Rang (*Rank Level*)

Quand la sortie de chaque "matcher" (module de reconnaissance) biométrique est un sous-ensemble de correspondances possibles triées dans un ordre décroissant de confiance, la fusion peut se faire au niveau *rang*. *Ho et al.* [Ho 94] décrivent trois méthodes pour combiner les rangs assignés par différents matchers. Dans la technique "*highest rank method*", on assigne à chaque correspondance possible le meilleur (minimum) rang calculé par différents matchers. En cas d'égalité, on en retient un seul au hasard afin d'arriver à un ordre de rang strict et la décision finale est prise selon les rangs combinés. La méthode "*Borda count*" utilise la somme des rangs assignés par les matchers individuels afin de calculer les rangs combinés. La méthode de *régression logistique* est une généralisation de la méthode "*Borda count*" où une somme pondérée des rangs individuels est calculée et les poids sont déterminés par régression logistique.

2.5.2.3 Niveau Score (*Score Level*)

Après les vecteurs de caractéristiques, **les scores** (de correspondance) **donnés en sortie par les matchers contiennent l'information la plus riche à propos du modèle d'entrée**. En fait, la fusion au niveau score (Fig. 2.5) donne le **meilleur compromis** entre la **richesse d'information** et la **facilité d'implémentation**. Aussi, il est relativement **facile d'accéder et de combiner les scores générés par les différents matchers**. En conséquence, l'intégration d'information au **niveau score** est **l'approche la plus courante dans les systèmes biométriques multimodaux**. Nous allons maintenant expliquer plus en détail ce niveau de fusion.

2.6 La fusion au niveau score

Il existe **deux approches pour combiner les scores** obtenus par différents matchers. La première approche est de voir cela comme un problème de *classification*, tandis que l'autre approche est de traiter le sujet comme un problème de *combinaison*. **Il est important de noter que *Jain et al.* ont montré que les approches par combinaison sont plus performantes que la plupart des méthodes de classification [Jain 05] [Ross 03].**

Dans l'approche par **classification**, un vecteur de caractéristiques est construit en utilisant les scores de correspondance donnés en sortie par les matchers individuels ; ce vecteur est ensuite attribué à une des deux classes : "accepté" (utilisateur authentique ou "*genuine user*") ou "rejeté" (utilisateur imposteur ou "*impostor user*"). En général, le classifieur utilisé pour cette opération est capable d'apprendre la frontière de décision sans tenir compte de la manière dont le vecteur de caractéristiques a été généré. Ainsi, les

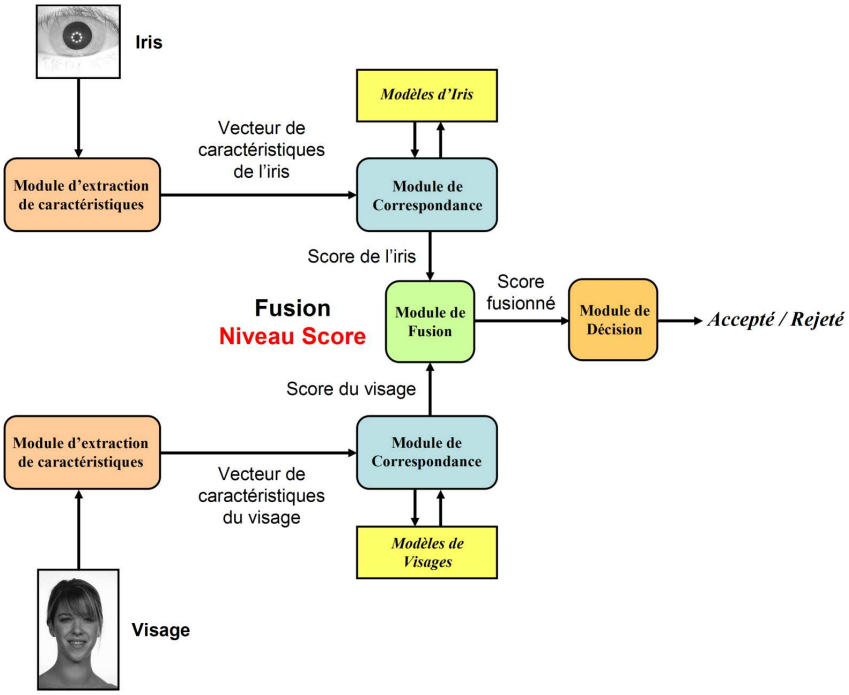


FIG. 2.5 – Fusion au niveau score dans un système biométrique multimodal.

scores en sortie de différentes modalités peuvent être non-homogènes (mesure de distance ou de similarité, différents intervalles de valeurs prises, etc.) et aucun traitement n'est requis avant de les envoyer dans le classifieur.

Dans l'approche par **combinaison**, les scores de correspondance individuels sont combinés de manière à former un unique score qui est ensuite utilisé pour prendre la décision finale. **Afin de s'assurer que la combinaison de scores provenant de différentes modalités soit cohérente, les scores doivent d'abord être transformés dans un domaine commun** : on parle alors de *normalisation de score*.

2.7 Normalisation de score

Considérons un système de vérification biométrique multimodale qui adopte une approche de fusion par combinaison, au niveau score. Le cadre de travail théorique développé par *Kittler et al.* [Kitt 98] utilise une **approche probabiliste** qui peut être appliquée à ce système seulement si la sortie de chaque modalité est de la forme $P\{authentique|X\}$, c'est-à-dire la probabilité a posteriori qu'un utilisateur soit "authentique" sachant l'échantillon d'entrée biométrique X. En pratique, la plupart des systèmes biométriques donnent en sortie un score de correspondance s. *Verlinde et al.* [Verl 99] ont proposé que ce score de correspondance s soit lié à $P\{authentique|X\}$ comme suit (2.1) :

$$s = f(P\{authentique|X\}) + \eta(X), \tag{2.1}$$

où f est une *fonction monotonique* et η est l'erreur faite par le système biométrique qui dépend de l'échantillon d'entrée biométrique X.

Cette erreur η peut être due au bruit introduit par le capteur pendant l'acquisition du signal biométrique et les erreurs faites par l'extraction de caractéristiques et les processus de correspondance.

Si l'on suppose que $\eta = 0$, il est raisonnable d'approximer $P\{\textit{authentique}|X\}$ par $P\{\textit{authentique}|s\}$. Dans ce cas, le problème revient à calculer $P\{\textit{authentique}|s\}$ et cela requiert l'estimation des densités conditionnelles $p(s|\textit{authentique})$ et $p(s|\textit{imposteur})$.

Les techniques de fusion de *Snelick et al.* [Snel 03] utilisent l'approche probabiliste de *Kittler et al.* en se servant d'un ensemble de données d'entraînement composé de 100 personnes afin d'estimer les probabilités a posteriori des authentiques $P\{\textit{authentique}|s\}$; Pour ce faire, ils utilisent la moyenne et la variance des scores authentiques et imposteurs provenant de cet ensemble de données d'entraînement en supposant une distribution normale pour leurs densités conditionnelles respectives $p(s|\textit{authentique})$ et $p(s|\textit{imposteur})$. Cependant, leur approche possède **deux inconvénients majeurs**. Tout d'abord, **l'hypothèse d'une distribution normale pour les scores n'est pas valide dans certains cas**, en particulier pour les scores authentiques. D'autre part, cette approche **n'utilise pas les probabilités a priori** des utilisateurs authentiques et imposteurs qui pourraient être disponibles au sein du système. En effet, la formule que *Snelick et al.* utilisent pour estimer $P\{\textit{authentique}|s\}$ [Snel 03] **n'est vraie que dans le cas où l'on considère les probabilités a priori** (c'est-à-dire les fréquences des classes) **équiprobables** ; dans le cas contraire, on doit directement appliquer le *Théorème de Bayes* avec des densités de probabilités (2.2) :

$$P\{\textit{authentique}|s\} = \frac{p(s|\textit{authentique}) * P\{\textit{authentique}\}}{p(s|\textit{authentique}) * P\{\textit{authentique}\} + p(s|\textit{imposteur}) * P\{\textit{imposteur}\}}. \tag{2.2}$$

Pour ces raisons, une méthode d'estimation de densité par fenêtre de Parzen a été proposée dans [Duda 01] ; cette technique permet d'estimer la véritable densité conditionnelle des scores authentiques et imposteurs.

Ainsi, bien que la technique d'estimation de densité par fenêtre de Parzen réduise de manière significative l'erreur sur l'estimation de $P\{\textit{authentique}|s\}$ (en particulier lorsque les densités conditionnelles ne sont pas gaussiennes), l'estimation de densité possède toujours des inexactitudes dues à la taille finie de l'ensemble des données d'entraînement et aux problèmes liés au choix de la largeur de fenêtre optimale pendant le processus d'estimation de densité. De plus, l'hypothèse selon laquelle la valeur de η dans l'équation (2.1) est nulle n'est pas valide, d'un point de vue pratique, dans la plupart des systèmes biométriques. Puisque η dépend de l'échantillon d'entrée biométrique X, il est possible d'estimer η seulement si le système biométrique donne une **mesure de confiance** (qui prend en compte la nature de l'entrée X) **sur le score de correspondance** en plus du score lui-même. En l'absence de cette mesure de confiance, la valeur calculée de $P\{\textit{authentique}|s\}$ n'est pas un bon estimateur de $P\{\textit{authentique}|X\}$ et cela peut amener à une **faible performance** de reconnaissance du système multimodal.

En conclusion, lorsque les sorties des modalités individuelles sont des scores de correspondance **sans aucune mesure** qui quantifierait la confiance de ces scores, **il est préférable de combiner directement ces scores de correspondance sans les convertir en probabilités**.

2.7.1 Pourquoi normaliser les scores ?

Trois problèmes importants ont besoin d'être considérés avant même de combiner les scores de correspondance en un seul et unique score. Tout d'abord, les scores de correspondance au niveau des sorties des matchers individuels **peuvent ne pas être homogènes**. Par exemple, un matcher peut donner en sortie une mesure de distance (**dissimilarité**) pendant qu'un autre donne en sortie une mesure de proximité (**similarité**). Ensuite, les sorties des matchers individuels **ne sont pas nécessairement inclus dans le même intervalle**. Enfin, les scores de correspondance en sortie des matchers **peuvent suivre différentes distributions statistiques**. A cause de ces raisons, la **normalisation** de score est **essentielle** pour transformer les scores des matchers individuels dans un domaine commun avant de les combiner. La normalisation de score est une **étape critique** dans la conception d'un schéma de combinaison pour la fusion au niveau score.

2.7.2 Identification d'une technique de normalisation de scores

La normalisation de score consiste à changer les paramètres de position (moyenne) et d'échelle (écart-type) des distributions de scores de correspondance en sortie des matchers individuels, de manière à ce que les scores de correspondance de différents matchers soient transformés dans un domaine commun. Quand les paramètres utilisés pour la normalisation sont déterminés en utilisant un ensemble de données d'entraînement fixé, on parle de **normalisation de score fixée** [Brun 95]. Dans ce cas, la distribution des scores de correspondance de l'ensemble des données d'entraînement est examinée et un modèle cohérent est choisi pour "coller" à la distribution. A partir de ce modèle, les paramètres de normalisation sont déterminés. Dans la **normalisation de score adaptative**, les paramètres de normalisation sont estimés en se basant sur le vecteur de caractéristiques actuel. Cette approche à la faculté de s'adapter aux variations de la donnée en entrée. **Pour avoir un bon schéma de normalisation**, les estimateurs des paramètres de position et d'échelle de la distribution de score de correspondance doivent être **robustes** et **efficaces**. La *robustesse* se réfère à l'insensibilité à la présence de valeurs aberrantes ("*outliers*"). L'*efficacité* se réfère à la proximité de l'estimateur obtenu par rapport à l'estimateur optimal lorsque la distribution des données est connue. *Huber* [Hube 81] explique les concepts de robustesse et d'efficacité de procédures statistiques. Finalement, bien que de nombreuses techniques peuvent être utilisées pour la normalisation de score, **le défi réside dans l'identification d'une technique qui serait à la fois robuste et efficace**.

2.7.3 Les différentes techniques de normalisation de scores

La technique de normalisation la plus simple est la **normalisation Min-Max**. Elle est la plus adaptée dans le cas où les bornes (valeurs minimales et maximales) des scores produits par un matchers sont connues. Dans ce cas, on peut facilement translater les scores minimums et maximums respectivement vers 0 et 1. Cependant, même si les scores de correspondance ne sont pas bornés, on peut estimer les valeurs minimales et maximales pour un jeu de scores de correspondance donné et appliquer ensuite la normalisation Min-Max. Soit s_{ij} le $j^{\text{ème}}$ score de correspondance de sortie de la $i^{\text{ème}}$ modalité, où $i = 1, 2, ..., R$ et $j = 1, 2, ..., M$ (R est le nombre de modalités et M le nombre de scores de correspondance disponibles dans l'ensemble de données d'entraînement). Le score normalisé Min-Max pour le score de test s_{ik} est donné par :

$$s'_{ik} = \frac{s_{ik} - \min(\{s_{i.}\})}{\max(\{s_{i.}\}) - \min(\{s_{i.}\})}, \tag{2.3}$$

où $\{s_{i.}\} = \{s_{i1}, s_{i2}, ..., s_{iM}\}$. Quand les valeurs minimales et maximales sont estimées à partir du jeu d'entraînement de scores donné, cette méthode **n'est pas robuste** (c'est-à-dire que cette méthode est fortement sensible aux valeurs aberrantes dans les données utilisées pour l'estimation). La normalisation Min-Max conserve la distribution de scores originale à un facteur d'échelle près et transforme tous les scores dans l'intervalle $[0, 1]$. Les scores relatifs à des mesures de distance peuvent être transformés en des scores de similarité en soustrayant le score normalisé à 1.

La méthode de "**decimal scaling**" peut être appliquée lorsque les scores de différents matchers évoluent selon une échelle logarithmique. Par exemple, si un matcher a des scores dans l'intervalle $[0, 1]$ et l'autre matcher a des scores dans l'intervalle $[0, 100]$, la normalisation suivante peut être appliquée :

$$s'_{ik} = \frac{s_{ik}}{10^n}, \tag{2.4}$$

où $n = \log_{10} \max(\{s_{i.}\})$. Les problèmes avec cette approche sont le manque de robustesse et l'hypothèse selon laquelle les scores de différents matchers varient d'un **facteur logarithmique**.

La technique de normalisation de score la plus employée est certainement la **Z-Score** qui utilise la moyenne arithmétique et l'écart-type des données. On peut s'attendre à ce que cette méthode fonctionne bien si on a une connaissance a priori du score moyen et des variations de score d'un matcher. Si on n'a pas de connaissance a priori sur la nature de l'algorithme de reconnaissance, nous devons alors estimer la moyenne et l'écart-type des scores à partir d'un jeu de scores de correspondance donné. Les scores normalisés sont donnés par :

$$s'_{ik} = \frac{s_{ik} - \mu}{\sigma}, \tag{2.5}$$

où μ est la moyenne arithmétique et σ l'écart-type des données. Cependant, la moyenne et l'écart-type sont tous les deux sensibles aux valeurs aberrantes et donc cette méthode n'est pas robuste. De plus, la normalisation *Z-Score* ne garantit pas un intervalle commun pour les scores normalisés provenant de différents matchers. Si la distribution des scores n'est pas gaussienne, la normalisation Z-Score ne conserve pas la distribution d'entrée en sortie. Cela est simplement dû au fait que la moyenne et l'écart-type sont les paramètres de position et d'échelle optimaux seulement pour une distribution gaussienne. Pour une distribution arbitraire, la moyenne et l'écart-type sont respectivement des estimateurs raisonnables de position et d'échelle, mais ne sont pas optimaux.

La médiane et l'écart absolu médian (MAD) sont insensibles aux valeurs aberrantes et aux points aux extrémités d'une distribution. Ainsi, une méthode de normalisation utilisant la médiane et la MAD serait robuste et est donnée par :

$$s'_{ik} = \frac{s_{ik} - median}{MAD}, \tag{2.6}$$

où $MAD = median(\{|s_{i.} - median(\{s_{i.}\})|\})$. Cependant, les estimateurs issus de la médiane et de la MAD ont une faible efficacité comparée aux estimateurs issus de la moyenne et de l'écart-type, c'est-à-dire que lorsque la distribution de score n'est pas gaussienne, la médiane et la MAD sont de pauvres estimateurs des paramètres de position et d'échelle. Ainsi, cette technique de normalisation ne conserve pas la distribution d'entrée et ne transforme pas les scores dans un intervalle commun.

Snelick et al. [Snel 05] utilisent une **fonction quadratique-linéaire-quadratique** (**QLQ**) pour normaliser des scores au préalable transformés dans l'intervalle $[0, 1]$ (s_{MM}) grâce à une normalisation *MinMax*. Cette normalisation QLQ prend comme paramètres le centre c et la largeur w de la zone de recouvrement des distributions des scores imposteurs et authentique (Fig. 2.6).

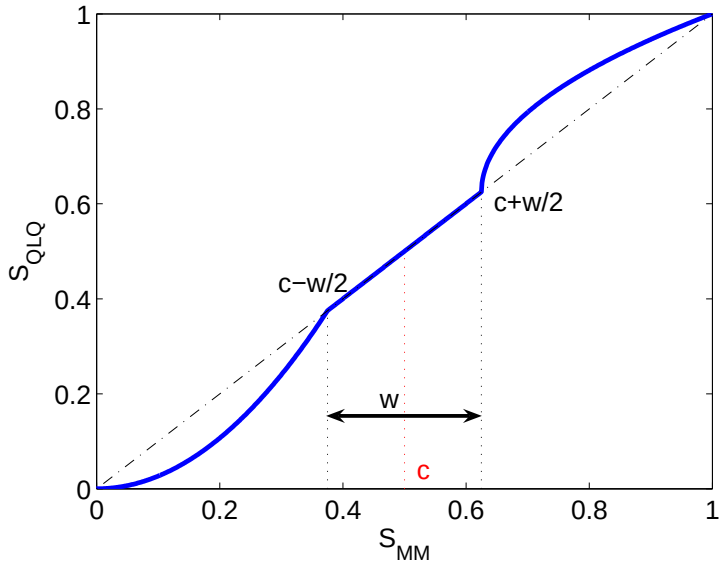


FIG. 2.6 – Normalisation QLQ.

La zone de recouvrement reste inchangée tandis que les autres régions sont transformées à l'aide de deux fonctions quadratiques par segment. Le score normalisé est donné par (2.7) :

$$s_{QLQ} = \begin{cases} \frac{1}{(c-\frac{w}{2})} s_{MM}^2, & \text{si } s_{MM} \leq (c - \frac{w}{2}) \\ s_{MM}, & \text{si } (c - \frac{w}{2}) < s_{MM} \leq (c + \frac{w}{2}) \\ (c + \frac{w}{2}) + \sqrt{(1 - c - \frac{w}{2})(s_{MM} - c - \frac{w}{2})}, & \text{sinon.} \end{cases} \quad (2.7)$$

Cappelli et al. [Capp 00] ont utilisé une **fonction double sigmoïde** pour la normalisation de score dans un système biométrique multimodal qui combine différents matchers d'empreinte digitale. Le score normalisé est donné par (2.8) :

$$s'_{ik} = \begin{cases} \frac{1}{1+\exp(-2(\frac{s_{ik}-t}{r_1}))} & \text{si } s_k < t, \\ \frac{1}{1+\exp(-2(\frac{s_{ik}-t}{r_2}))} & \text{sinon,} \end{cases} \quad (2.8)$$

où t est le point de fonctionnement de référence et r_1 et r_2 sont respectivement les bords gauche et droit de la région dans laquelle la fonction est linéaire, c'est-à-dire que la fonction double sigmoïde montre des caractéristiques linéaires dans l'intervalle $(t - r_1, t - r_2)$. La figure 2.7 montre un exemple d'une normalisation double sigmoïde, où les scores étant dans l'intervalle $[0, 300]$ sont transformés dans l'intervalle $[0, 1]$, avec $t = 200$, $r_1 = 20$ et $r_2 = 30$.

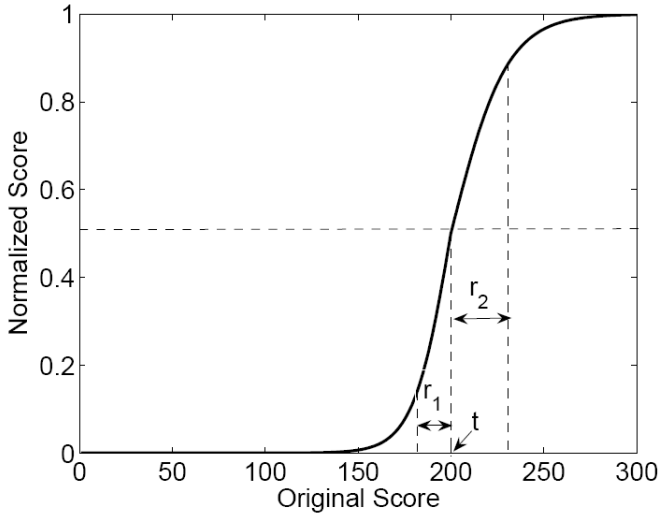


FIG. 2.7 – Normalisation double sigmoïde.

Ce schéma transforme les scores dans l'intervalle $[0, 1]$. Mais il demande de paramétrer prudemment t, r_1 et r_2 afin d'obtenir une bonne efficacité. En général, t est choisi de façon à être une valeur appartenant à la région de recouvrement entre les distributions des scores authentiques et imposteurs, et r_1 et r_2 sont fixés de telle sorte qu'ils correspondent respectivement aux extrémités de recouvrement entre les deux distributions vers la gauche et vers la droite de t (on peut également voir cela comme les valeurs correspondant respectivement au minimum des scores de similarité des authentiques et au maximum des scores de similarité des imposteurs). Ce schéma de normalisation fournit donc une transformation linéaire des scores dans la région de recouvrement, tandis que les scores à l'extérieur de cette région sont transformés de manière non linéaire. La normalisation double sigmoïde est très similaire à la normalisation Min-Max suivie par l'application d'une fonction de deux quadriques (QQ) ou d'une fonction de régression logistique (LG) comme suggérée par *Snelick et al.* [Snel 05]. Lorsque r_1 et r_2 sont grands, la normalisation double sigmoïde est assez proche de la normalisation QQ-Min-Max. D'autre part, on peut faire tendre la normalisation double sigmoïde vers la normalisation LG-Min-Max en assignant des petites valeurs à r_1 et r_2. Enfin, on peut noter que la normalisation double sigmoïde ne peut s'appliquer dans le cas où il y a plusieurs intervalles de régions de recouvrement entre les scores authentiques et imposteurs ; elle ne conserve pas la forme de la distribution originale en entrée.

Les estimateurs tanh (pour **tan**gente **h**yperbolique) introduits par *Hampel et al.* sont robustes et très efficaces [Hamp 86]. La normalisation est donnée par :

$$s'_{ik} = \frac{1}{2}\{tanh(0.01(\frac{s_{ik} - \mu_{GH}}{\sigma_{GH}})) + 1\}, \tag{2.9}$$

où μ_{GH} et σ_{GH} sont respectivement les estimateurs de la moyenne et de l'écart-type de la distribution des scores authentiques, tels qu'ils sont donnés par les estimateurs de Hampel. Les estimateurs de Hampel sont basés sur la fonction d'influence (ψ) suivante :

$$\psi(u) = \begin{cases} u & 0 \leq |u| < a, \\ a * sign(u) & a \leq |u| < b, \\ a * sign(u) * (\frac{c-|u|}{c-b}) & b \leq |u| < c, \\ 0 & |u| \geq c, \end{cases} \tag{2.10}$$

La figure 2.8 illustre la **fonction d'influence de Hampel**. Cette dernière réduit l'influence des points aux extrémités d'une distribution (identifiés par a, b et c) pendant l'estimation des paramètres de position et d'échelle. Ainsi, cette méthode n'est pas sensible aux valeurs aberrantes. Si plusieurs points constituant une extrémité d'une distribution ne sont plus pris en compte, l'estimateur est robuste mais pas efficace (optimal). D'autre part, si tous les points constituant l'extrémité d'une distribution sont considérés, l'estimateur n'est pas robuste mais son efficacité augmente. Par conséquent, les paramètres a, b, et c doivent être soigneusement choisis selon la quantité de robustesse exigée, ce qui dépend alternativement de l'évaluation de la quantité de bruit dans l'ensemble des données d'entraînement disponible.

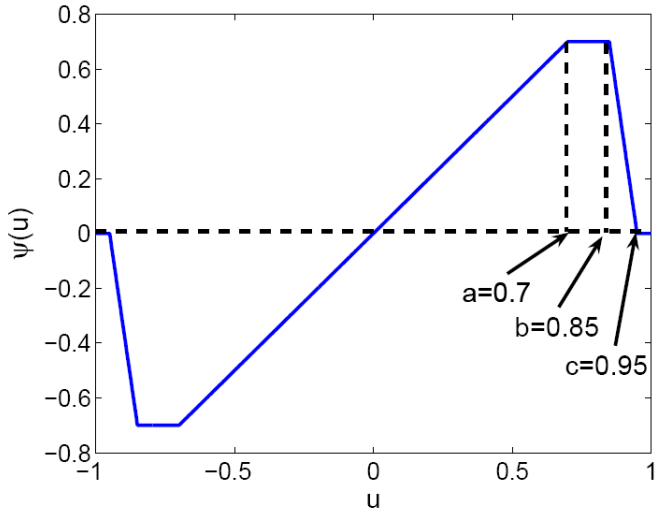


FIG. 2.8 – Fonction d'influence de Hampel ($a = 0.7$, $b = 0.85$ et $c = 0.95$).

On remarque que les scores authentiques s_a doivent être d'abord transformés dans l'intervalle $[-1, 1]$, ce qui peut être fait grâce à une transformation de score qui est très proche de la normalisation Min-Max (2.11) :

$$s'_{ak} = \frac{2 * s_{ak} - (\max(\{s_{a.}\}) + \min(\{s_{a.}\}))}{\max(\{s_{a.}\}) - \min(\{s_{a.}\})}, \tag{2.11}$$

Mosteller et Tukey [Most 77] ont introduit les **estimateurs "biweight"** pour estimer les paramètres de position et d'échelle. Cette méthode est robuste et efficace. Mais les estimateurs biweight sont itératifs par nature (une estimation initiale est choisie et cette estimation est mise à jour, basée sur des scores issus d'un jeu d'entraînement) et ne sont applicables que pour des données gaussiennes.

Les caractéristiques des différentes techniques de normalisation que nous venons de détailler sont résumées dans le tableau 2.1 en termes de robustesse et d'efficacité qui ont été définis dans le paragraphe 2.7.2 :

Technique de Normalisation	**Robustesse**	**Éfficacité**
MinMax	Non	N/A
Decimal Scaling	Non	N/A
Z-Score	Non	Elevée (optimale pour des données gaussiennes)
Médiane et MAD	Oui	Modérée
QLQ	Oui	Élevée
Double Sigmoïde	Oui	Élevée
Estimateurs tanh	Oui	Élevée
Estimateurs biweight	Oui	Élevée

TAB. 2.1 – Résumé Des Techniques de Normalisation de Scores

2.8 Approche par classification de scores

Plusieurs classifieurs ont été utilisés pour combiner les scores de correspondance afin d'arriver à une décision.

Wang et al. [Wang 03] considèrent les scores provenant de modules de reconnaissance faciale et de reconnaissance par l'iris comme un vecteur de caractéristiques à deux dimensions. Une analyse discriminante linéaire de Fisher (LDA) et un classifieur par réseau de neurones combiné à une fonction de base radiale (RBF) sont ensuite utilisés pour la classification.

Verlinde et Chollet [Verl 99] combinent les scores provenant de deux modules de reconnaissance faciale et d'un module de reconnaissance de la parole avec l'aide de trois classifieurs : un premier classifieur utilisant la méthode des "k plus proches voisins" ("k-nearest neighbor algorithm", "k-NN") avec une quantification vectorielle, un deuxième classifieur basé sur un arbre décisionnel et un dernier classifieur basé sur un modèle de régression logistique.

Chatzis et al. [Chat 99] utilisent une méthode de regroupement ("clustering") appelée "fuzzy k-means" et une quantification vectorielle floue ("fuzzy vector quantization"), couplée à un classifieur de réseau de neurones de RBF médiane pour fusionner les scores obtenus à partir de systèmes biométriques basés sur des caractéristiques visuelles (visage) et acoustiques (voix).

Sanderson et al. [Sand 02] utilisent un classifieur basé sur une Machine à Vecteurs de Support (SVM) pour combiner les scores issus d'un module de reconnaissance faciale et d'un module de reconnaissance de la parole. Ils montrent que la performance d'un tel classifieur se détériore en la présence de conditions de bruit en entrée. Afin de surmonter ce problème, ils implémentent des classifieurs résistants au bruit structurel comme un classifieur linéaire définit par morceau ("piece-wise linear classifier") et un classifier Bayésien modifié.

Ross et Jain [Ross 03] utilisent un arbre décisionnel et des classifieurs discriminants linéaires pour combiner les scores des modalités du visage, d'empreinte digitale et de géométrie de la main.

2.9 Approche par combinaison de scores

Kittler et al. [Kitt 98] ont développé un cadre théorique pour combiner les informations d'identification obtenues à partir de plusieurs classifieurs en utilisant des schémas tels que la règle somme ("*sum rule*"), la règle "produit" ("*product rule*"), la règle maximum ("*max rule*"), la règle minimum ("*min rule*") et la règle médiane ("*median rule*"). Afin d'employer ces schémas, les scores de correspondance doivent être convertis en probabilités a posteriori conformément à un utilisateur authentique ou imposteur. Ils considèrent le problème de classer un modèle d'entrée X dans l'une des m classes possibles (dans un système de vérification, $m = 2$) d'après l'information d'identification fournie par R classifieurs ou matchers différents.

Soit $\overrightarrow{x}_i$ un vecteur de caractéristiques (provenant du modèle d'entrée X) présenté au $i^{\text{ème}}$ matcher. Considérons les sorties des matchers individuels $P(w_j|\overrightarrow{x}_i)$, c'est-à-dire, la probabilité a posteriori de la classe w_j sachant $\overrightarrow{x}_i$. Soit $c \in \{w_1, w_2, ..., w_m\}$ la classe finalement assignée à X. Les règles suivantes peuvent alors être utilisées pour estimer c :

– **La Règle Produit** (*Product Rule*) : Cette règle est basée sur l'hypothèse d'indépendance statistique des représentations $\overrightarrow{x}_i$, $\overrightarrow{x}_2$, ..., $\overrightarrow{x}_R$. Le modèle d'entrée est assignée à la classe c telle que :

$$c = \arg\max_j \prod_{i=1}^{R} P(w_j|\overrightarrow{x}_i) \tag{2.12}$$

En général, les différents traits biométriques d'un individu (par exemple, visage, empreinte digitale, géométrie de la main, etc.) sont mutuellement indépendants. Ceci nous permet donc d'utiliser la règle produit dans un système biométrique multimodal basé sur cette hypothèse d'indépendance statistique.

– **La Règle Somme** (*Sum Rule*) : La règle somme est plus efficace que la règle produit lorsque il y a un fort niveau de bruit menant à une certaine ambiguïté dans le problème de classification. La règle somme assigne la classe c au modèle d'entrée X comme suit :

$$c = \arg\max_j \sum_{i=1}^{R} P(w_j|\overrightarrow{x}_i) \tag{2.13}$$

– **La Règle Maximum** (*Max Rule*) : La règle maximum approxime la moyenne des probabilités a posteriori par la valeur maximum. Dans ce cas, on assigne la classe c au modèle d'entrée X de la manière suivante :

$$c = \arg\max_j \max_i P(w_j|\overrightarrow{x}_i) \tag{2.14}$$

– **La Règle Minimum** (*Min Rule*) : La règle minimum est obtenue en bornant le produit des probabilités a posteriori. On assigne alors la classe c au modèle d'entrée X de la façon suivante :

$$c = \arg\max_j \min_i P(w_j|\overrightarrow{x}_i) \tag{2.15}$$

Prabhakar et Jain [Prab 01] affirment que **l'hypothèse d'indépendance statistique** des ensembles de caractéristiques ("*feature sets*") **peut être non valide** dans le cas d'un système biométrique multimodal qui utilise différentes représentations de caractéristiques et différents algorithmes de correspondance ("*matching algorithms*") du **même trait biométrique**. Ils proposent un schéma basé sur une estimation non paramétrique de densité pour combiner les scores provenant de quatre algorithmes de correspondance d'empreinte digitale et utilisent un test de ratio de vraisemblance pour prendre la décision finale. Ils montrent que leur méthode est optimale au sens de la décision de *Neyman-Pearson*, lorsque il y a suffisamment de données d'entraînement disponibles pour estimer les densités jointes.

L'utilisation de **statistiques Bayésiennes** dans la combinaison de scores de différents matchers biométriques a été démontrée par *Bigun et al.* [Bigu 97]. Ils ont proposé un nouvel algorithme pour le module de fusion d'un système biométrique multimodal qui prend en compte la précision estimée des classifieurs individuels pendant le processus de fusion. ils ont montré que leur système multimodal utilisant des données visuelles et de la parole fournissent de meilleurs résultats de reconnaissance que les modalités individuelles.

Les scores de correspondance combinés peuvent aussi être calculés comme une somme pondérée de scores de correspondance des matchers individuels [Ross 03] [Wang 03]. *Jain et Ross* [Jain 02] ont proposé l'utilisation de poids spécifiques à chaque utilisateur afin de calculer la somme pondérée de scores provenant de différentes modalités. La motivation qui se cache derrière cette idée est qu'une petite partie de la population ne peut pas fournir de manière fiable certains traits biométriques. Par exemple, on ne peut pas obtenir d'empreintes digitales de bonne qualité à partir de personnes possédant des doigts abîmés par des opérations manuelles répétées. Pour de tels utilisateurs, assigner un poids plus faible au score d'empreinte digitale et un poids plus élevé aux scores d'autres modalités réduit leurs probabilités d'être faussement rejetés par le système biométrique. Cette méthode requiert un apprentissage de poids spécifiques à chaque utilisateur à partir des scores d'entraînement disponibles pour chaque utilisateur. Dans [Jain 02], une méthode basée sur l'utilisation de seuils spécifiques à chaque utilisateur a également été suggérée.

2.10 Conclusion

La figure 2.9 présente un résumé de type haut-niveau des techniques de fusion que nous venons de présenter. La plupart des travaux de recherche en fusion dans les systèmes biométriques multimodaux se sont concentrés sur la fusion au niveau score, en particulier, l'approche par combinaison de scores a reçu une attention considérable et nous avons vu qu'elle se révélait être plus performante que la plupart des approches par classification [Jain 05] [Ross 03]. Lorsqu'il n'est pas possible d'obtenir une mesure de la qualité de l'échantillon biométrique présenté au système, on effectue les opérations de fusion directement sur les scores de correspondance et non plus sur des probabilités. Cependant, il y a toujours de nombreuses questions ouvertes qui sont laissées sans réponse. Il n'existe aucune technique standard ni pour convertir les scores en probabilités ni pour normaliser les scores obtenus à partir de plusieurs algorithmes de correspondance. De même, une évaluation méthodique de différentes techniques de normalisation n'est pas disponible. De plus, la plupart des techniques de fusion au niveau score peuvent être appliquées seulement si les modalités individuelles peuvent fournir une performance de reconnaissance raisonnablement bonne.

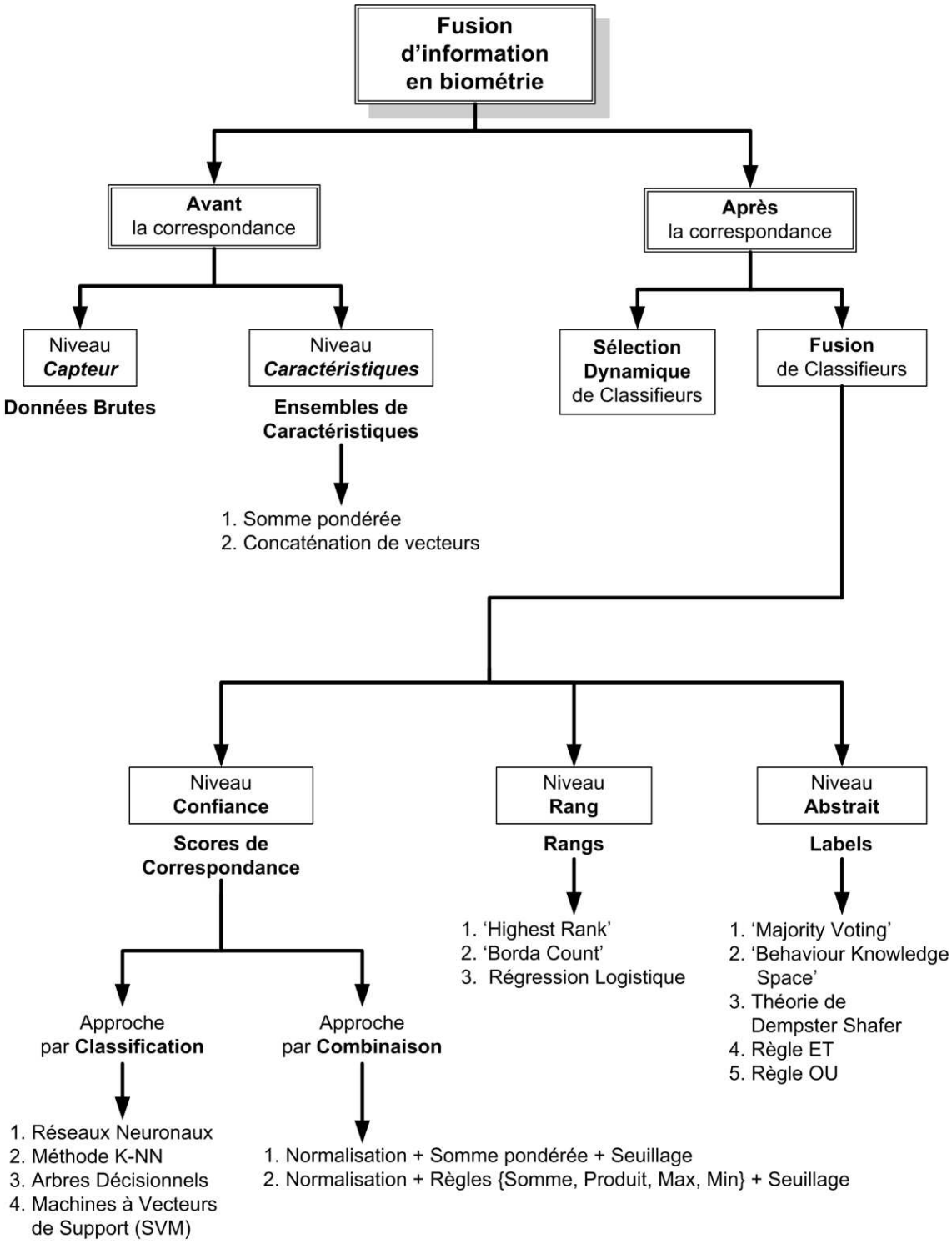


FIG. 2.9 – Résumé des approches de fusion dans les systèmes biométriques multimodaux (d'après [Jain 05]).

Chapitre 3

Perception Visuelle et Neurosciences

3.1 Introduction

Ce chapitre introduit **quelques concepts de base en perception des images et des visages** dans un contexte **neuroscientifique**. Il doit permettre de mieux comprendre les processus de reconnaissance faciale par les humains et par ordinateur. Nous devrions alors être en mesure de nous orienter vers l'algorithme de reconnaissance du visage qui semble le plus prometteur et qui sera utilisé dans notre système multimodal. Par ailleurs, les **T**echnologies de l'**I**nformation et de la **C**ommunication (**TIC**) regroupent les ressources nécessaires pour manipuler de l'information tandis que **la cognition** rassemble les divers processus mentaux qui vont de l'analyse perceptive de l'environnement à la commande motrice et qui fait appel à l'informatique et aux neurosciences. La biométrie et tout particulièrement **la reconnaissance faciale** peut donc être vue comme étant au **croisement des TIC et de la cognition** (Fig. 3.1).

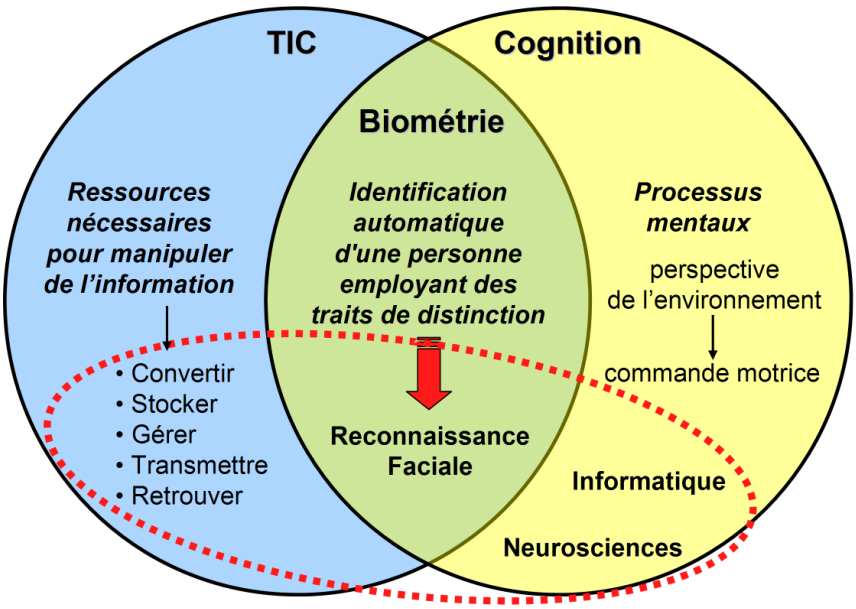


FIG. 3.1 – La Biométrie au croisement des TIC et de la Cognition.

Il semble alors intéressant de se poser les questions suivantes :

- *Comment fonctionne notre cerveau pour identifier et reconnaître une personne par son visage ?*
- *Quelles sont les parties du cerveau qui sont impliquées dans ce processus complexe mais à la fois tellement rapide ?*
- *Dans quelles mesures peut-on comparer le fonctionnement de notre cerveau à certains algorithmes utilisés en reconnaissance faciale par ordinateur ?*

Pour tenter d'y répondre, nous allons donc nous intéresser au phénomène de perception des images pour ensuite étudier la perception des visages. Enfin, nous mettrons en avant les liens qui peuvent exister entre le fonctionnement de notre cerveau et certains algorithmes fondamentaux utilisés en reconnaissance faciale.

3.2 Perception des images et Théorie de la Gestalt

La perception d'une image est le **produit d'une chaîne d'opérations** que l'on peut représenter par une **pyramide** (Fig. 3.2). Chaque étage de cette pyramide représente un niveau de perception. La base de la pyramide correspond au premier (bas) niveau de la perception où s'opèrent la détection et le codage des **structures élémentaires** et des formes simples. Les niveaux supérieurs font appel à la cognition où des **tâches plus complexes** sont accomplies. Il est important de noter que l'information circule de manière **bilatérale** (avec des "*feedbacks*") le long de cette pyramide et qu'elle est "calculée" de proche en proche selon ce que l'on appelle des **algorithmes pyramidaux** qui permettent de "fabriquer" des approximations successives et de plus en plus grossières de l'image.

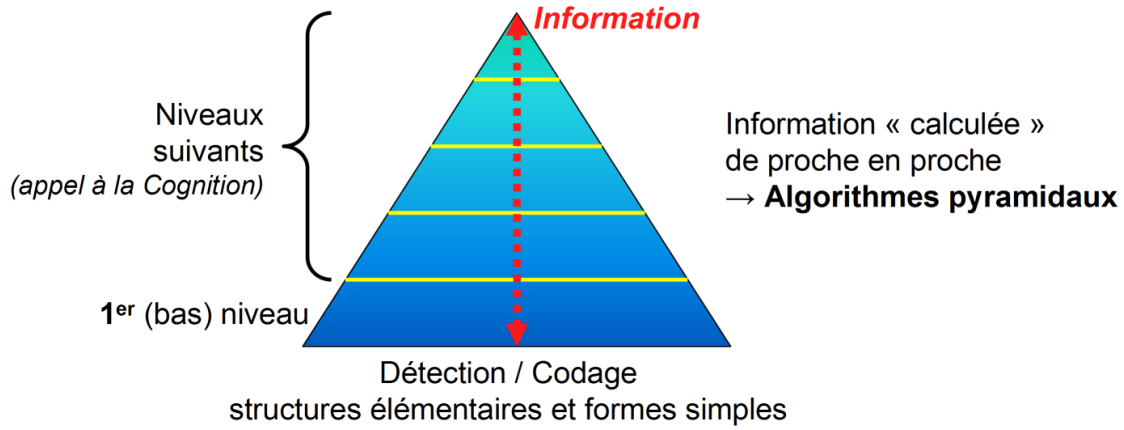


FIG. 3.2 – Pyramide de la perception.

Selon le philosophe allemand *Ernst Cassirer* (1874 - 1945), "*La perception d'une image naturelle est basée sur la reconnaissance des formes ou des structures sous-jacentes qui y figurent*".

La **Théorie de la Gestalt** illustre très bien les propos d'Ernst Cassirer. Elle relève du domaine de la psychophysiologie et trouve son origine dans quelques idées de *Goethe*.

Aux 19ème et 20ème siècles, ce sont *Ernst Mach*, *Wolfgang Köhler*, *Kurt Koffka* et *Kurt Lewin* qui la développent et surtout *Christian von Ehrenfels* et *Max Wertheimer* qui font office de précurseurs. Ehrenfels n'était pas un psychologue gestaltiste mais on peut l'inclure car il étudiait les phénomènes de perception de musique et rendait compte de la façon dont on perçoit une mélodie : quand on transpose un morceau (changement d'octave par exemple), on ne change pas la forme de la mélodie ; quand on change juste quelques notes, on modifie la forme de la mélodie. Il proposa l'idée selon laquelle on ne peut pas percevoir la mélodie comme une sensation élémentaire ; **pour décrire une forme, il faut considérer les éléments et les relations entre les éléments**. En ce qui concerne Wertheimer, on lui attribut les débuts de la gestalt. En 1912, il travaille sur le mouvement apparent. L'expérience consiste à allumer, par alternance dans l'obscurité, deux points lumineux légèrement distants. On voit apparaître successivement chacun des deux points en deux lieux distincts dans un rythme lent ; si on augmente le rythme on a l'impression de va et vient d'un seul point, d'où la perception (qui est différente de sensation). **Pour Ehrenfels et Wertheimer**, **la perception correspond à une organisation indépendante de la sensation**.

Le mot allemand *gestalt* est traduit par "forme" (ainsi, *Gestalttheorie* signifie "théorie de la forme"). Bien qu'il s'agisse en réalité de quelque chose de beaucoup plus complexe, elle dit à peu près la chose suivante : "**A priori, une image possède une structure géométrique simple**". Le verbe *gestalten* signifie "mettre en forme, donner une structure signifiante". Le résultat, la "gestalt", est donc une forme structurée, complète et prenant sens pour nous. Par exemple, lorsqu'on regarde les étoiles, chacune d'elle est un stimulus visuel, pourtant on peut facilement les organiser en constellations, en ensemble formé de stimuli. Ainsi, l'image mentale que nous avons en tête est une forme, et peut être évaluée par notre esprit en tant que telle, par exemple en la nommant : "la Grande Ourse".

On constate ainsi que **le tout est différent de la somme de ses parties**, un des principes phares de la Théorie de la Gestalt. La théorie souligne aussi qu'une partie dans un tout est autre chose que cette même partie isolée ou incluse dans un autre tout, puisqu'elle tire des propriétés particulières de sa place et de sa fonction dans chacun d'entre eux. **Pour comprendre un comportement ou une situation**, il importe donc, non seulement de les analyser, mais surtout, d'en avoir une vue synthétique, de les percevoir dans l'ensemble plus vaste du contexte global, avoir un regard non pas plus "pointu" mais plus large : pour pouvoir apprécier un tableau de peinture, **il faut prendre du "recul"** !

En fait, il existerait des **processus internes** qui guideraient les perceptions. Ces principes sont connus sous le nom de **lois de l'organisation perceptive** qui s'appliquent soit à la perception d'une figure elle-même, soit à la ségrégation figure/fond :

1. *Principe de proximité* : dans une figure, les éléments proches sont groupés,
2. *Principe de similitude* : les aspects de l'objet qui se ressemblent tendent à être regroupés. Ce principe est plus fort que celui de proximité. Il faut que les éléments soient suffisamment similaires et distincts d'un autre groupe perceptif,
3. *Principe du destin ou du sort commun* : les éléments qui semblent se mouvoir sont considérés comme appartenant au même groupe perceptif,
4. *Principe de bonne continuation* : on tend à préserver une continuité plutôt que d'introduire des changements brusques,

5. *Principe de fermeture* : on a tendance à voir des figures fermées plutôt qu'ouvertes quand les deux sont possibles,
6. *Principe de taille relative, d'environnement, d'orientation et de symétrie* :
toutes choses étant égales par ailleurs, la plus petite des deux surfaces tend à être perçue comme la figure du fond ; d'autre part, les régions orientées horizontalement ou verticalement tendent aussi à être perçues comme des figures,
7. *Principe de bonne forme* : quand plusieurs organisations sont possibles, c'est la forme la plus simple et la plus stable qui émerge. On met en évidence la corrélation entre le jugement de tridimensionnalité et le nombre d'angles de la figure. Une figure qui contient beaucoup d'angles est perçue de manière plus simple.

Une des justifications de la Théorie de la Gestalt est fournie par les illusions visuelles dont certaines montrent que **notre cerveau est précâblé** pour trouver (extraire) de la géométrie dans une image. Un exemple bien connu est une illusion d'optique cognitive publiée en 1955 par *Gaetano Kanizsa*, **le triangle de Kanizsa** (Fig. 3.3).

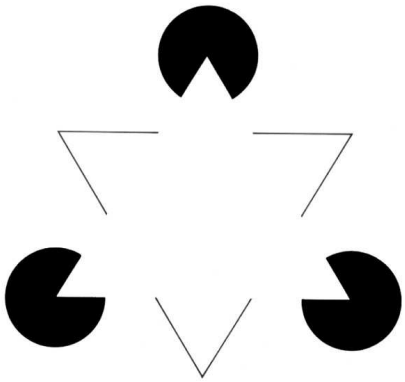


FIG. 3.3 – Illusion du triangle de Kanizsa.

Dans la figure 3.3, un mécanisme cognitif s'appuie sur le relief apparent du dessin produisant l'illusion. La figure subjective semble se trouver en avant des autre figures, les cachant en partie à notre vue. Ainsi, un triangle blanc semble se découper au-dessus des cercles noirs supports des sommets et des trois pointes de triangle situés entre les cercles. Cet effet est connu sous le nom de *contour subjectif*. Il existe peut-être un autre mécanisme cognitif qui joue un rôle important : lorsque nous percevons une figure illusoire, nous l'interprétons inconsciemment en fonction de nos expériences antérieures.

En fait, dès notre naissance, la première "forme" importante que nous reconnaissions est une Gestalt : c'est **le visage de notre mère**. Le nouveau-né n'en perçoit pas encore les détails, mais la forme globale est "signifiante" pour lui. Nos perceptions obéissent à un certain nombre de lois : ainsi, une symphonie est autre chose qu'une simple succession de notes de même qu'une totalité (dans notre précédent exemple, un visage humain) ne peut se réduire à la simple somme des stimuli perçus. Ces dernières remarques nous amènent donc tout naturellement à parler de la *perception des visages*.

3.3 Perception des visages

Nathalie George, chargée de recherche au CNRS dans le laboratoire de neurosciences cognitives et d'imagerie cérébrale résume parfaitement le caractère spécial d'un visage. Elle explique : "*[...] Du point de vue physique, les visages constituent une catégorie d'objets très homogène [...] avec une symétrie verticale globale. Cependant, les visages humains sont tous différents (sauf peut-être de prime abord ceux des jumeaux) et nous, êtres humains, sommes capables de discriminer des milliers de visages, d'en mémoriser des centaines, d'en identifier beaucoup. Nous pouvons nous souvenir du visage d'une personne rencontrée une seule fois, mais aussi reconnaître une personne familière à travers les ans, et cela, malgré différents changements d'apparence. Le visage est donc un "objet spécial" [...]. C'est aussi sans doute l'un des objets visuels les plus importants de notre environnement, auquel nous sommes exposés depuis notre naissance.*".

La perception des visages est un processus cognitif par lequel le cerveau analyse une image pour y détecter et identifier un visage. C'est une faculté très développée chez l'être humain. Il s'agit d'une aptitude très spécifique qui repose sur des mécanismes neurocognitifs complexes **en partie innés** dont certains sont présents chez le nourrisson dès la naissance. On parle alors de **précâblage couplé à un apprentissage**. Il existe d'ailleurs une analogie très intéressante avec le langage chez les nourrissons : en effet, sans la notion de précâblage, comment expliquer le fait que n'importe quel nourrisson est capable de parler n'importe quelle langue ? En fait, il semblerait que certains neurones vont se spécialiser pour évoluer (apprentissage) vers la langue qu'entend le plus souvent le nourrisson, les autres neurones vont dégénérer.

De manière analogue aux illusions d'optique relatives à la perception des images, il existe deux illusions d'optique bien connues *concernant les visages* qui sont respectivement les illusions de Thatcher (Fig. 3.4) et les "*Mooney Face Pictures*" (Fig. 3.5).



FIG. 3.4 – Illusion de Thatcher.

Dans l'illusion de Thatcher, on renverse les deux photos de la partie supérieure, qui *nous semblent très voisines*, cependant, *après retournement*, la photo en bas à droite est plus impressionnante que prévu !

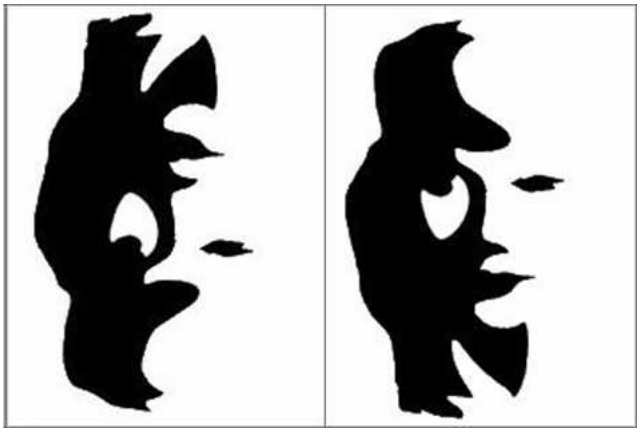


FIG. 3.5 – Mooney Face Pictures.

Quant aux *Mooney Face Pictures*, il s'agit d'un exemple d'images utilisées par le psychologue cognitif *Craig Mooney* : on perçoit aisément un visage dans l'image de droite mais pas dans l'image de gauche, bien que cette dernière soit seulement l'image de droite renversée verticalement ; on peut également faire appel à la cognition pour apporter de la sémantique à l'image et voir l'image de droite comme un saxophoniste qui ferait la manche en lisant sa partition.

Au sein du système visuel, un certain nombre de régions sont particulièrement impliquées dans la perception des visages, notamment l'**aire fusiforme des visages** ("**F**usiform **F**ace **A**rea", **FFA**) [Tong 00] qui constitue une partie du gyrus fusiforme au niveau de la jonction des **lobes temporaux et occipitaux** (Fig. 3.6).

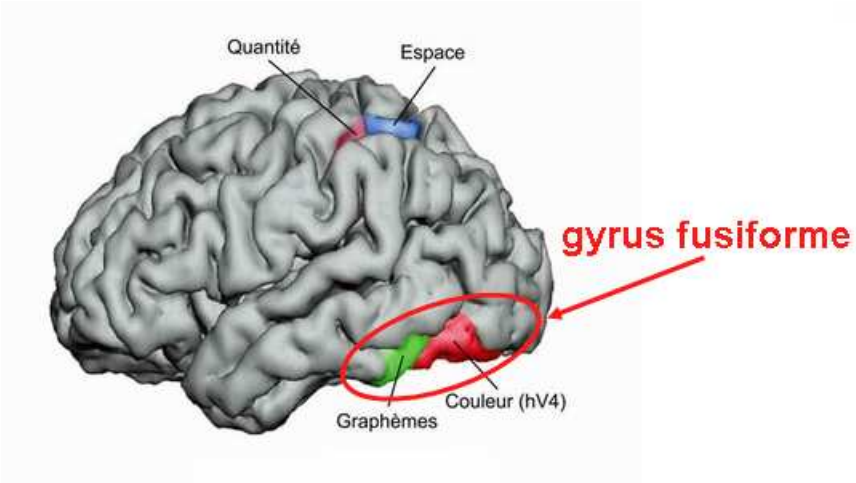


FIG. 3.6 – Cartographie de l'hémisphère gauche (*E. Hubbard*) avec le gyrus fusiforme.

Une **lésion cérébrale** de ces régions et en particulier de la FFA, peut entraîner une *prosopagnosie* (Fig. 3.7), c'est-à-dire une **incapacité à reconnaître les visages** (avec une vision normale par ailleurs).

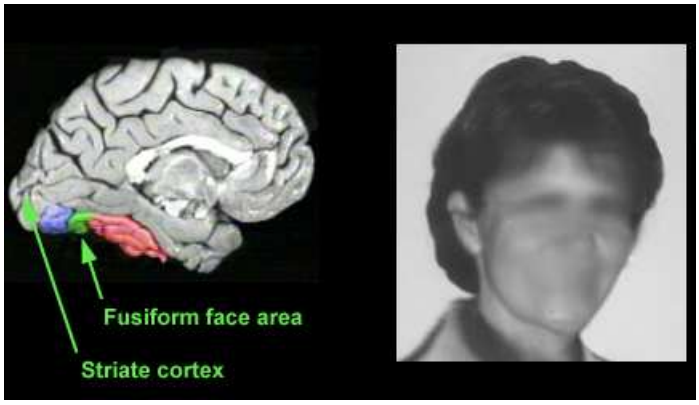


FIG. 3.7 – Illustration de la prosopagnosie.

En ce qui concerne la **modélisation du fonctionnement du système visuel**, *David Marr* suppose que le traitement bas-niveau nous donne une traduction de l'information fournie par **la rétine** en ce qu'il appelle une *représentation* qu'il définit de la façon suivante :

"*A representation, therefore, is not a foreign idea at all- we all use representations all the time. However, the notion that one can capture some aspect of reality by making a description of it using a symbol and that to do so can be useful seems to me a fascinating and powerful idea. But even the simple examples we have discussed introduce some rather general and important issues that arise whenever one chooses to use one particular representation. For example, if one chooses the Arabic numeral representation, it is easy to discover whether a number is a power of 10 but difficult to discover whether it is a power of 2. If one chooses the binary representation, the situation is reversed. Thus, there is a trade-off ; any particular representation makes certain information explicit at the expense of information that is pushed into the background and may be quite hard to recover.*"

Par ailleurs, **une image possède une information gigantesque** et l'**on ne perçoit finalement qu'une partie infime** d'une image que l'on regarde ! Certains neurophysiologistes ont émis l'hypothèse de travail suivante : si les **cellules du cerveau du cortex visuel primaire** sont affectées à des **tâches spécifiques de reconnaissance de certaines structures géométriques**, c'est peut-être parce que **cette solution biologique au problème de lecture d'une image est optimale en terme d'analyse et de compression des données**. Nous allons donc étudier d'un peu plus près ce fameux cortex visuel primaire.

3.4 Le cortex visuel primaire

Le cortex visuel primaire, aussi appelé **V1** (Fig. 3.8), est l'endroit du cerveau où la reconnaissance d'objet est la plus efficace [Logo 99]. *Pourquoi dans le V1* ? Elle se fait de manière **très rapide** (*de 100ms à 200ms*) et reste **très robuste** aux variations en éclairement, pose, taille et angle de vue. L'information circule le long d'une sorte de "nappe" en partant d'abord de la rétine avant de passer par le noyau géniculaire latéral (**LGN**) pour enfin atteindre le V1. Le **flux dorsal** est le chemin qui indique "**où**" sont les objets tandis que le **flux ventral** indique "**quoi**", donc **la nature des objets**. Le V1 représente le premier niveau de traitement dans le cortex visuel et reste "le plus facile" à caractériser

notamment grâce à la présence de ce que l'on appelle des **cellules "simples"** [Jone 87]. Voyons comment nous pouvons modéliser ces cellules.

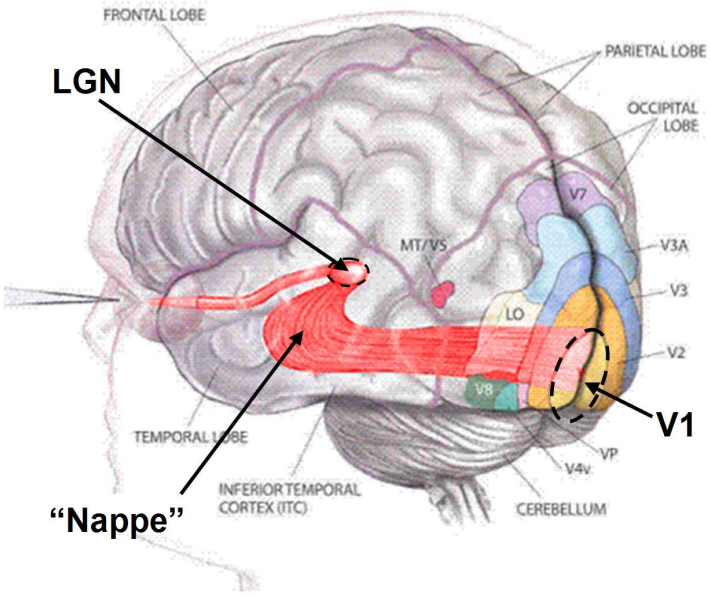


FIG. 3.8 – Chemin visuel dans le cerveau humain.

3.5 Filtres de Gabor

3.5.1 Champs réceptifs linéaires

Les réponses de ces cellules "simples" dans le V1 s'appellent **les champs réceptifs linéaires** (**RFs**) [Marc 80]. Ces RFs sont restreints par de petites régions de l'espace, hautement structurés et appartiennt à une zone discrète qui est relative à la *fovea* (Fig. 3.9), où la présentation ou le retrait d'un stimulus visuel provoque une **activation cellulaire**.



FIG. 3.9 – La fovea : grossièrement située dans le centre de la rétine.

En 1981, *Pollen* et *Ronner* [Poll 81] examinent la **relation de phase de cellules adjacentes dans le cortex visuel de chats**. Ils en viennent à la conclusion que **les cellules d'une paire de cellules adjacentes** (définies par des fréquences similaires, c'est-à-dire des amplitude et des spécificités de direction similaires) **possèdent certaines symétries**. La première cellule, de symétrie paire, peut être modélisée par une fonction **cosinus** et la deuxième cellule, de symétrie impaire, par une fonction **sinus**. Ces résultats permettent déjà la possibilité de **modéliser les RFs** d'une telle paire de cellules par une **fonction à valeur complexe**, dont nous parlerons un peu plus tard.

En 1987, Dans les expériences de *Jones* et *Palmer* [Jone 87], des cellules simples sont mesurées avec une micro-électrode. Le RF d'une certaine cellule a été précisément mesuré en projetant un stimulus sous forme de point sur un écran homogène que l'oeil correspondant regarde. Un exemple des réponses résultantes est donné sur la Fig. 3.10. *Jones* et *Palmer* suggèrent alors de modéliser la forme des RFs par un **filtre 2D de Gabor**, c'est-à-dire un **plan d'onde restreint par une gaussienne**.

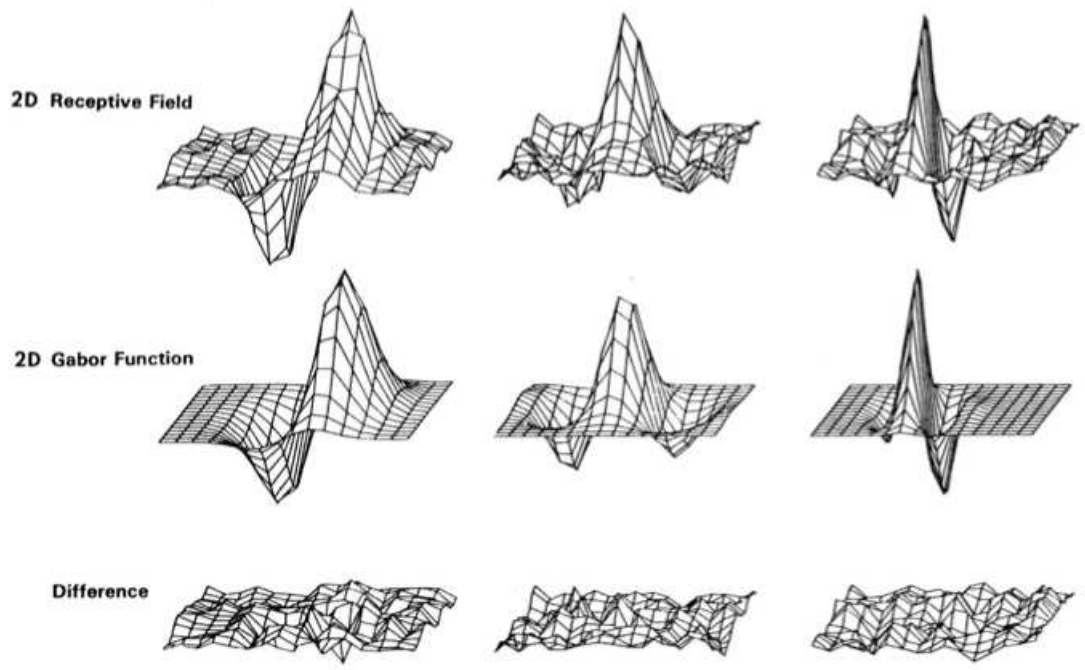


FIG. 3.10 – Adaptation d'un filtre de Gabor aux données correspondant aux mesures du comportement de la réponse d'une certaine cellule simple.

Sur la première ligne de la Fig. 3.10, on peut voir le **RF expérimental** et sur la deuxième ligne, le **RF modélisé par des ondelettes 2D de Gabor**. On peut apprécier la **très grande similitude** entre les deux entités par leur **très faible différence** représentée sur la dernière ligne.

3.5.2 Limites des filtres de Gabor

Les filtres de Gabor constituent un choix traditionnel pour obtenir une information fréquentielle localisée. Ils offrent la meilleure localisation simultanée en temps et en fréquence (cf. *Principe d'Incertitude d'Heisenberg*[1], Chapitre 6.2.2). Les filtres de Gabor complexes utilisés dans les traitements de signaux unidimensionnels sont définis, dans le domaine temporel, comme le produit d'une enveloppe gaussienne par une sinusoïde complexe (3.1) :

$$g(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{t^2}{2\sigma^2}\right)} \times e^{j(2\pi f_0 t + \phi)} \tag{3.1}$$

où σ représente l'écart type de l'enveloppe gaussienne, f_0 la fréquence centrale du filtre et ϕ la phase de la sinusoïde.

Nous pouvons voir ce filtre de Gabor complexe comme l'association de deux filtres déphasés, correspondants à la partie réelle (filtre de symétrie paire, (3.2)) et à la partie imaginaire (filtre de symétrie impaire, (3.3)) d'une fonction complexe :

$$g_r(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{t^2}{2\sigma^2}\right)} \times \cos\left(2\pi f_0 t + \phi\right) \tag{3.2}$$

$$g_i(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{t^2}{2\sigma^2}\right)} \times \sin\left(2\pi f_0 t + \phi\right) \tag{3.3}$$

Cependant, ils possèdent **deux principales limitations**. La **bande passante maximale** d'un filtre de Gabor est **limitée approximativement à une octave** et les filtres de Gabor **ne sont pas optimaux** si l'on cherche à avoir une information spectrale large avec une localisation spatiale maximale. En effet, il est impossible de construire des fonctions de Gabor avec une **bande passante arbitrairement large tout en maintenant une composante continue (moyenne du signal dans le domaine fréquentiel) raisonnablement faible dans le filtre de symétrie paire** (3.2) (Fig. 3.11).

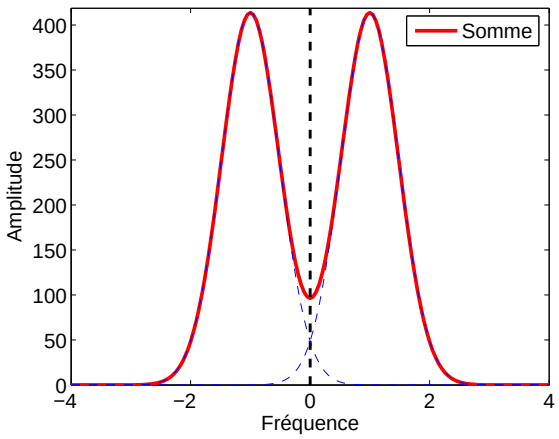


FIG. 3.11 – Fonction de transfert d'un filtre de Gabor de symétrie paire de bande passante élevée.

[1] "Exhibit on Heisenberg and the uncertainty principle", from the American Institute of Physics. Ressource Internet : `http://www.aip.org/history/heisenberg/`

La fonction de transfert est la somme de deux gaussiennes centrées à plus et moins la fréquence centrale. Si l'écart-type σ de ces gaussiennes devient supérieur à un tiers de la fréquence centrale, ces deux gaussiennes vont commencer à se recouvrir de manière significative à l'origine, entraînant une composante continue non nulle. Dans un contexte de filtrage, cette composante continue non nulle constitue un inconvénient puisque une même fréquence peut alors être analysée par deux filtres différents, entraînant des artefacts dans le spectre. À la situation critique où la fréquence centrale est égale à 3σ, la bande passante sera approximativement égale à une octave. Cela peut être vu de la manière suivante : pour une gaussienne, les points où sa valeur chute de la moitié du maximum sont situés approximativement à $+\sigma$ et $-\sigma$, ces points définissant les fréquences de coupures. Ainsi, les fréquences de coupures inférieures et supérieures seront respectivement situées à environ 4σ et 2σ, donnant une **bande passante d'une octave**. Cette limitation sur la bande passante signifie que **nous avons besoin de beaucoup de filtres de Gabor pour obtenir une large couverture spectrale**.

Une alternative à la fonction de Gabor est la **fonction de Log-Gabor** proposée par *Field* en **1987** [Fiel 87]. Les filtres de Log-Gabor peuvent être construits avec une bande passante arbitraire qui **peut être optimisée** afin de produire un filtre avec une **extension spatiale minimale**.

3.6 Filtres de Log-Gabor

Field suggère que **les images naturelles sont mieux codées** par des filtres qui ont des fonctions de transfert gaussiennes vues depuis une **échelle fréquentielle logarithmique** (Fig. 3.12). Si l'on choisit une échelle fréquentielle linéaire, la fonction de Log-Gabor a une fonction de transfert de la forme suivante (3.4) :

$$G(w) = e^{\left(\frac{-\left(\log\left(\frac{w}{w_0}\right)\right)^2}{2\left(\log\left(\frac{k}{w_0}\right)\right)^2}\right)} \tag{3.4}$$

où w_0 est la *fréquence centrale du filtre*.

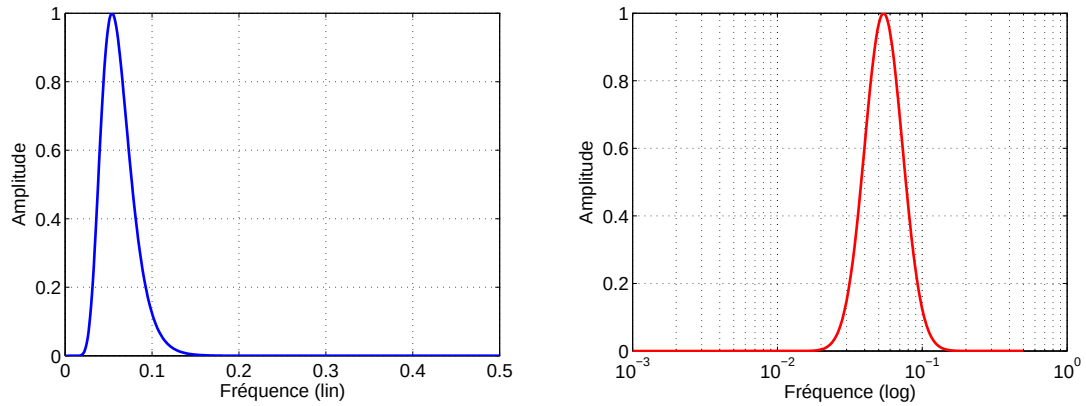


FIG. 3.12 – Fonction de transfert de Log-Gabor pour $k = 0.04$ (échelle fréquentielle linéaire à gauche, logarithmique à droite).

Afin d'obtenir des filtres de ratio de forme constante, le rapport $\frac{k}{w_0}$ doit rester constant lorsque w_0 varie. Par exemple, lorsque $\frac{k}{w_0} = 0.74$, la bande passante du filtre est d'environ une octave, 0.55 donnera une bande passante de deux octaves et 0.41 produira trois octaves.

Il est intéressant de noter qu'un filtre de Log-Gabor ayant une bande passante de trois octaves possède la même largeur spatiale qu'un filtre de Gabor d'une octave, ce qui montre la capacité des filtres de Log-Gabor de capturer une information spectrale large tout en ayant un filtre spatial compact.

Ensuite, il y a **deux caractéristiques importantes à noter**. Tout d'abord, les fonctions de Log-Gabor, par définition, **ne possèdent jamais de composante continue**, ensuite, la fonction de transfert de la fonction de Log-Gabor possède une **extrémité étendue aux hautes fréquences**.

Les études de *Field* sur les statistiques des images naturelles [Fiel 87] indiquent que les images naturelles ont une **amplitude spectrale qui décroît approximativement en** $1/w$. Pour encoder des images ayant de telles caractéristiques spectrales, on devrait utiliser des filtres ayant un spectre similaire. Field suggère que les fonctions de Log-Gabor, possédant des extrémités étendues, devrait être capable d'encoder des images naturelles de manière **plus efficace que des fonctions de Gabor ordinaires**, qui sur-représenteraient les composantes basses-fréquences et sous-représenteraient les composantes hautes-fréquences dans n'importe quel encodage.

Un autre point en faveur de la fonction de Log-Gabor est qu'**elle est cohérente avec des mesures effectuées au sein des systèmes visuels des mammifères** qui indiquent que nous avons des réponses de cellules qui sont symétriques, vues selon une échelle fréquentielle logarithmique. Cette section constitue une version française corrigée de la page web de *Peter Kovesi* [Kove 00].

3.7 Liens entre le fonctionnement du cerveau et certains algorithmes utilisés en reconnaissance faciale

3.7.1 Liens avec l'Analyse en Composantes Indépendantes (ICA)

l'ICA permet de résoudre le problème de séparation de sources dont une illustration classique est le problème de la soirée cocktail ("*cocktail party problem*") [Brow 01] [Sagi 01] [Hayk 05]. Ce problème rentre dans le cadre plus général de la *séparation aveugle de sources* qui consiste à retrouver un certain nombre de sources à partir des observations d'un mélange de celles-ci. Le terme "aveugle" traduit simplement le fait que l'on ignore la façon dont les sources se mélangent, ainsi que le nombre de sources que l'on doit retrouver.

Supposons que nous représentions les données observées (Fig. 3.13) par un vecteur aléatoire à p dimensions noté x, le problème revient donc à trouver une fonction $\mathcal{F}$ (appelée *fonction de mélange*) d'un certain nombre n de "sources primitives" qui sont aussi considérées comme un vecteur aléatoire $s = (s_1, s_2, \ldots, s_n)^T$, telles que $x = \mathcal{F}(s)$.

Dans le cadre du problème de la "soirée cocktail", chacune des p dimensions représente un capteur (microphone par exemple) et les n sources sont les voix des convives et les autres bruits ambiants, on demande également à ce que $p \geq n$. De manière générale, il

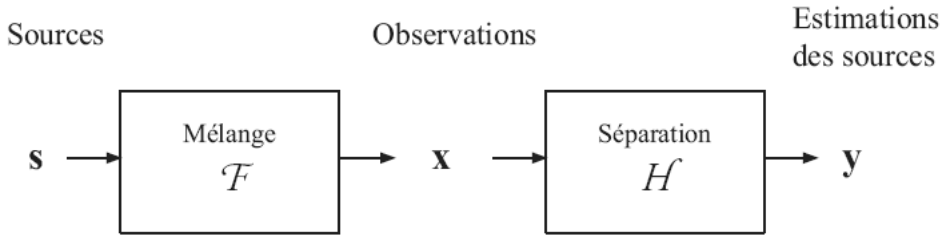


FIG. 3.13 – Modèle général de la représentation de données.

s'agit de trouver la meilleure façon de représenter les données x comme transformées des variables s au moyen de la fonction $\mathcal{F}$. Cela revient donc à trouver un nouvel espace de représentation des données, la *meilleure* façon dépend bien entendu de la manière dont on veut comprendre les données, donc des hypothèses formulées dans un cadre applicatif déterminé. **Dans le cas de l'ICA, l'hypothèse sous-jacente permettant la meilleure représentation des données est que les sources sont statistiquement indépendantes entre elles**. On trouvera un cadre mathématique complet dans [Card 02].

En 1961, dans un contexte neurobiologique, *H.B. Barlow* fait l'hypothèse selon laquelle **le rôle des neurones sensoriels primaires est de retirer la redondance statistique de l'entrée sensorielle** [Barl 89]. Des variantes de cette hypothèse de "*codage efficace*" ont été formulées par de nombreux autres auteurs [Bart 07]. Selon *D.J. Field* et *B.A. Olshausen*, **l'ICA modéliserait le travail effectué par les cellules rétiniennes** [Olsh 96] [Olsh 97].

L'algorithme **ICA** [Bart 02] répond à ce dernier défi d'**éliminer la redondance statistique**. L'ICA est un algorithme global, basé sur le **concept intuitif de "contraste"**. En effet, pour extraire une information pertinente d'un ensemble riche de données complexes et non structurées, il faut optimiser le contraste, c'est-à-dire disposer de **différents points de vue**, à partir de directions les plus éloignées les unes des autres.

Ainsi, en reconnaissance faciale, l'ICA minimise les dépendances statistiques d'ordre élevé des données d'entrée (et non pas seulement du second ordre, comme c'est le cas dans le PCA), en tentant de trouver les bases sur lesquelles les données projetées sont statistiquement indépendantes entre-elles. Rappelons enfin que, dans le cas gaussien, l'ICA coïncide avec le PCA, qui est un des algorithmes les plus connus en reconnaissance faciale.

3.7.2 Liens avec l'Elastic Bunch Graph Matching (EBGM)

L'algorithme EBGM [Wisk 97] tente de surmonter les difficultés causées par les variations de l'image en termes d'éclairement, position, taille, expression faciale et pose. Cet algorithme local extrait une **description concise d'un visage** sous la forme d'un graphe d'image dans lequel des points caractéristiques [Arca 06] (centres des yeux, du nez, de la bouche, etc.) sont décrits par un **jeu de coefficients d'ondelettes de Gabor** appelés "**Jet**". **Un Jet peut alors être vu comme la réponse d'une cellule "simple" du V1**. En effet, la réponse α d'une telle cellule est la *corrélation* d'une *image d'entrée* $I(x)$ avec le *RF modélisé* par une ondelette 2D de Gabor $\Psi(x)$ (3.5) :

$$\alpha_j(x_0) = \int I(x)\Psi_j(x - x_0)dx \tag{3.5}$$

Enfin, On pourra également se référer aux nombreux travaux du *Professeur Jeanny Hérault* (LIS, Université Joseph Fourier, Grenoble) sur la modélisation du système visuel, les réseaux de neurones, la transformée de Gabor et l'ICA [Héra 05].

3.7.3 Liens avec l'Analyse en Composantes Principales avec Filtres de Log-Gabor (LG-PCA)

L'algorithme hybride LG-PCA [Perl 05] qui a été présenté dans l'état de l'art peut être vu comme la combinaison de l'extraction de caractéristiques locales fournies par l'algorithme EBGM et la réduction de dimension offerte par l'action de l'algorithme PCA (cas particulier de l'algorithme ICA). Il a également le grand avantage d'utiliser les filtres de Log-Gabor que nous avons présenté dans le Chapitre 3.6. **L'algorithme LG-PCA modéliserait donc à la fois les réponses des cellules simples du V1 ainsi que le travail effectué par les cellules rétiniennes**, tout en **encodant les images naturelles de manière plus performante** grâce aux filtres de Log-Gabor, le **rapprochant encore plus du fonctionnement de notre cerveau**. C'est pourquoi nous avons porté toute notre attention sur cet algorithme et que **nous l'avons implémenté pour les tests de fusion**.

3.8 Conclusion

Il existe deux niveaux de perception : le bas-niveau (V1) dont certaines cellules possèdent des tâches spécifiques (reconnaissance de certaines structures géométriques élémentaires) et les niveaux suivants (V2 à V7) faisant appel à la cognition. D'autre part, la perception des visages suppose un précâblage couplé à un apprentissage et il existe dans le système visuel, une zone spécialisée dans la reconnaissance faciale (FFA). Par ailleurs, l'étude du cortex visuel primaire a montré qu'il était possible de modéliser les réponses de certaines cellules "simples" par des filtres 2D de Gabor mais que l'on pouvait faire encore mieux en utilisant des filtres de Log-Gabor. Ainsi, certains algorithmes fondamentaux en reconnaissance faciale (ICA, EBGM et LG-PCA) tentent de reproduire le fonctionnement de notre cerveau. l'ICA modéliserait le travail effectué par les cellules rétiniennes en éliminant la redondance statistique des images naturelles. De plus, l'EBGM modéliserait les réponses des cellules corticales simples du V1 grâce à la convolution de certains points caractéristiques d'une image de visage avec des ondelettes 2D de Gabor. Enfin, le LG-PCA permettrait de combiner les avantages de modélisation des deux algorithmes précédents en améliorant le codage des images naturelles grâce à l'utilisation de filtres de Log-Gabor. Ce chapitre est le fruit de *trois mois d'études en neurosciences* qui ont aboutit à la conférence "*Le Croisement des TIC et de la Cognition : L'exemple de la Biométrie*", pendant la journée "*NBIC pour les TIC*" (FING), donnée à l'Université Paris Descartes, le 25 Avril 2007.

Chapitre 4

Systèmes d'Acquisition en Environnement Réel & Bases de Données

4.1 Introduction

Dans ce chapitre, nous allons décrire nos dispositifs de capture ("*sensors*") qui ont permis d'acquérir des images de visage et d'iris. Nous détaillerons ensuite certaines bases de données qui seront utilisées, plus tard, pour nos tests (Chapitre 8).

Il existe deux manières de tester des algorithmes de reconnaissance biométrique :

- La première façon consiste à utiliser des **bases de données biométriques publiques** (gratuites) ou privées (payantes). C'est une solution relativement facile et certaines bases de données publiques sont devenues des standards (par exemple, FERET pour les visages, CASIA pour les iris), ce qui permet de comparer les résultats de reconnaissance des algorithmes développés à ceux des algorithmes existants. Cependant, dans ces bases de données, nous n'avons accès qu'aux images en sortie des divers dispositifs de capture ; la réalité de l'acquisition des données en environnement réel est alors mise de côté,

- La deuxième méthode consiste donc à établir un **système biométrique multimodal réel**, en utilisant des dispositifs de capture dédiés. La phase d'acquisition est la première étape dans un système biométrique et la nature de l'environnement (éclairement, reflets, etc.) peut avoir un certain impact sur les performances d'un système biométrique. C'est pourquoi il est recommandé d'effectuer des tests en environnement réel en plus des tests avec les bases de données, même si le nombre d'images acquises est moins important que celui que peut fournir une base de données officielles contenant souvent plusieurs centaines d'individus.

4.2 Systèmes d'acquisition

Le système d'acquisition de biométrie multimodale qui a été mis en place est divisé en deux parties. Un premier dispositif permet de prendre en photo l'iris d'une personne tandis qu'un deuxième appareil est consacré à la capture d'une image de visage.

4.2.1 Système de capture d'images d'iris

Les iris ont été photographiés avec un système adapté (*Miles Research*) (Fig. 4.1). L'appareil photo est un **Nikon D70** associé à une lentille de 105mm. Un **guide** est fixé à l'objectif et permet d'amener le **flash sur quatre spots** en sortie d'objectif. Un **support** permet d'ajuster la position de l'utilisateur et de faire la mise au point sans modifier les réglages de l'appareil. Les photos ont été prises dans une pièce sombre afin d'éviter les reflets. Un éclairement au préalable de la pupille à l'aide d'une *led* permet de modifier la dilatation de celle-ci.

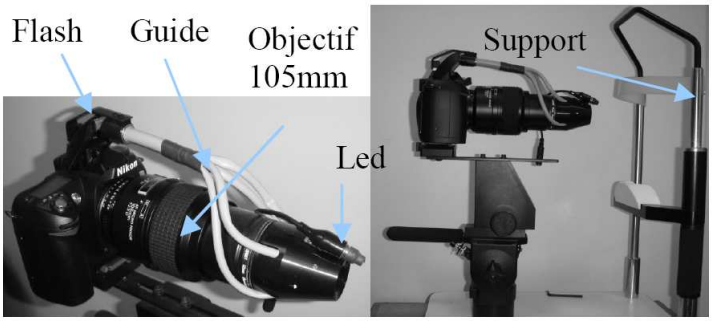


FIG. 4.1 – Système photographique d'enrôlement pour les images d'iris.

4.2.2 Système de capture d'images de visage

Les images de visage ont été acquises par une caméra motorisée autofocus et réseau **SONY SNC-RZ50P** (Fig. 4.2) ; elle est montée sur trépied, à hauteur du visage, soit environ 1m70. Elle est dotée d'une fonctionnalité PTZ, la plage horizontale (Pan) est de $340°$, la plage verticale (Tilt) est de $115°$ (Fig. 4.3) et possède un puissant zoom optique $\times 26$ et numérique $\times 12$ ($\times 312$ au total). Cette caméra permet de prendre des images avec une résolution maximale de (640×480) pixels à un taux de 25 images/seconde (Fig. 4.4). Enfin, ses fonctions d'alarme avancées et de sorties électroniques la rendent parfaitement adaptée aux domaines de la sécurité et de la biométrie.



FIG. 4.2 – Caméra motorisée autofocus et réseau SONY SNC-RZ50P.

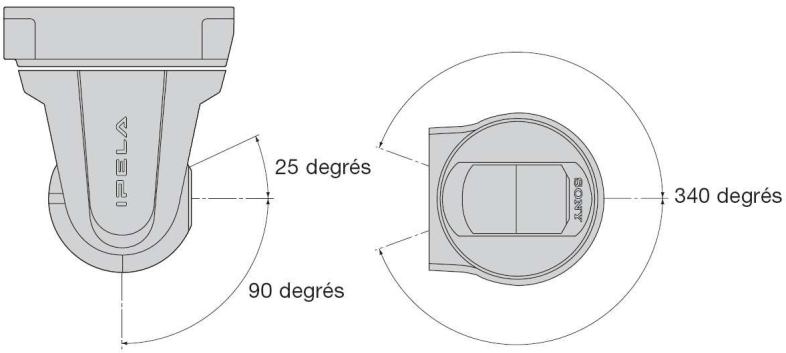


FIG. 4.3 – Pan/Tilt de la caméra SONY SNC-RZ50P.



4.4.a 4.4.b 4.4.c

FIG. 4.4 – Exemples d'images de visage prises avec la caméra SONY SNC-RZ50P (avec l'aimable autorisation de B.Giraud, L.Simon et D.Robin).

4.3 Bases de Données

4.3.1 FERET

Le programme **FERET** (pour "***FacE RE**cognition **T**echnology*") [Phil 00] s'est déroulé de 1993 à 1997. Sponsorisé par le "*Department of Defense's Counterdrug Technology Development Program*" et la "*Defense Advanced Research Products Agency*" (**DARPA**), sa mission première était de développer des solutions de reconnaissance faciale automatique qui pourraient être employées pour aider la sécurité, les renseignements et les forces de l'ordre dans l'exercice de leurs fonctions.

Le corpus d'images FERET [FERE 01] a été assemblé pour faciliter les tests et l'évaluation d'algorithmes de reconnaissance faciale gouvernementaux américains, en utilisant des tests et des procédures standardisés. Le corpus final rassemble un total de **14051 images en niveaux de gris** (codées sur 8 bits) de visages humains combinant des vues frontales et des profils gauches et droits. La figure 4.5 illustre quelques exemples d'images de la base de données FERET qui est constituée des ensembles suivants (pour les tests standards) :

- **Gallery (FA)** : ensemble de **1196 images** de personnes qui sont enrôlées (connues) par le système,
- **Training Set** : sous-ensemble de la *Gallery*, contenant **501 images**. Le *Training Set*, comme son nom l'indique, permet "d'entraîner" un système à partir d'un échantillon de personnes connues du système,

- **Probe Set (FB)** : ensemble de test contenant **1195 images**. Il contient les images des mêmes personnes que la *Gallery* mais avec des **expressions faciales différentes**,
- **Probe Set (FC)** : ensemble de test contenant **194 images**. Il contient les images de certaines personnes de la *Gallery* mais avec des **variations en éclairement**,
- **Probe Set (DUP1)** : ensemble de test contenant **722 images**. Ces images ont été prises **entre 0 et 1031 jours** après l'image correspondante de la *Gallery*. La médiane est de 72 et la moyenne est de 251 jours,
- **Probe Set (DUP2)** : sous-ensemble de test de *DUP1*, contenant **234 images**. Ces images ont été prises **entre 540 et 1031 jours** après l'image correspondante de la *Gallery*. La médiane est de 569 et la moyenne est de 627 jours.

Les ensembles *DUP1* et *DUP2* servent donc à effectuer des tests de **changements dans le temps**.



FIG. 4.5 – Quelques exemples d'images de la base de données FERET.

4.3.2 CASIA

La base de données **CASIA-IrisV3** a été collectée par le "***C**hinese **A**cademy of **S**ciences **I**nstitute of **A**utomation*" (Fig. 4.6 et 4.7.a). Il s'agit de la plus grande base de données publique d'images d'iris avec laquelle il est relativement difficile de travailler car les images présentent beaucoup d'occlusions (recouvrements de l'iris par les paupières).

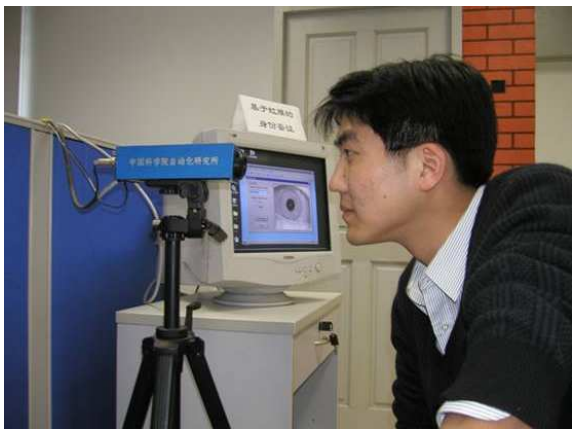
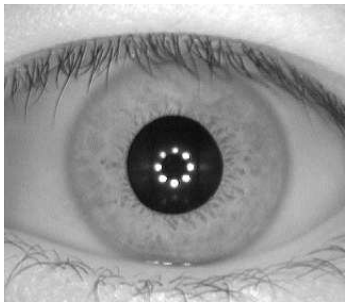


FIG. 4.6 – Caméra de capture de l'iris développée par CASIA.

La base de données *CASIA-IrisV3* [CASI 05] comprend 3 sous-ensembles appelés *CASIA-IrisV3-Interval* (Fig. 4.7.b), *CASIA-IrisV3-Lamp* et *CASIA-IrisV3-Twins*. *CASIA-IrisV3* contient un total de **22051 images d'iris** provenant de plus de **700 personnes**. Toutes les images d'iris sont en niveaux de gris, codées sur 8 bits au format JPEG et ont été prises sous un éclairage en **proche infra-rouge**. Quelques statistiques et caractéristiques de chaque sous-ensemble sont présentées dans Tab.4.1. Pratiquement tous les sujets sont chinois exceptés une petite quantité dans *CASIA-IrisV3-Interval*. Étant donné que les 3 ensembles de données ont été collectés à différents moments, seuls *CASIA-IrisV3-Interval* et *CASIA-IrisV3-Lamp* possèdent un petit recouvrement concernant les sujets.



4.7.a



4.7.b

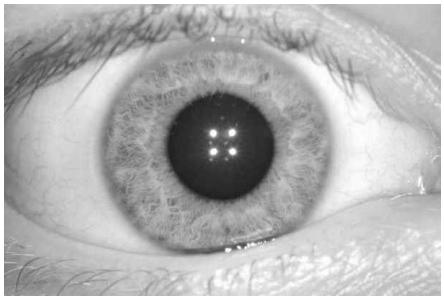
FIG. 4.7 – Caméra de capture de l'iris par OKI (4.7.a). Une image d'iris de CASIA-IrisV3-Interval (4.7.b).

	CASIA-IrisV3-Interval	CASIA-IrisV3-Lamp	CASIA-IrisV3-Twins
Capteur	Propriétaire	OKI's IRISPASS-h	OKI's IRISPASS-h
Environnement	Intérieur	Intérieur avec lampe on/off	Extérieur
Session	La plupart des images ont été capturées sur deux sessions, avec au moins un mois d'écart.	une	une
N. de sujets	249	411	200
N. de classes	396	819	400
N. d'images	2655	16213	3183
Résolution	320×280	640×480	640×480
Caractéristiques	Très bonne qualité d'image avec des détails de texture de l'iris extrêmement clairs.	Présence de déformations non linéaires dues à des variations de l'éclairage dans le domaine visible.	Le premier ensemble de données publique d'images d'iris de jumeaux.
Total	22051 images d'iris, plus de 700 sujets et 1500 yeux.		

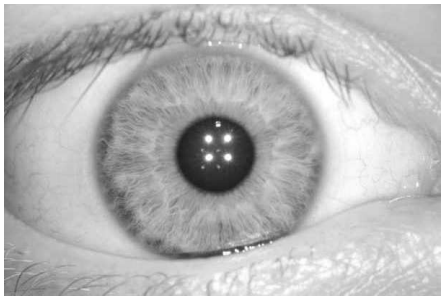
TAB. 4.1 – Statistiques de la base de données CASIA-IrisV3.

4.3.3 ISEP

Les images de notre base ont été prises en interne dans notre laboratoire. La base de données contient actuellement **1373 images d'iris** pour **312 individus** après suppression de tous les cas présentant des occlusions de paupières trop importantes. Les images sont de résolution **600×400**. Elles sont de bonne qualité, sans reflet et avec un taux d'occlusion de paupières acceptable.



4.8.a



4.8.b

FIG. 4.8 – Exemples images d'iris de la base de données ISEP. Pupille dilatée (4.8.a), pupille contractée (4.8.b).

Le protocole mis en place (Fig. 4.1) nous permet d'obtenir **2 images d'iris avec pupilles dilatées (Fig. 4.8.a) et 2 images avec pupilles non dilatées (Fig. 4.8.b) pour chaque individu**, soit un total de **quatre photographies par personne**. Par ailleurs, une *quarantaine d'individus* ont fourni **4 images d'iris par œil** dont 2 avec une pupille dilatée et 2 avec une pupille contractée.

Nous avons utilisé une partie de cette base de données interne (228 personnes) afin d'effectuer quelques expériences préliminaires de fusion (simple concaténation de vecteurs de caractéristiques) et nous avons obtenus d'excellents résultats (taux de reconnaissance proche de 100%). Cependant, notre base de données contient des images d'excellente qualité, avec très peu d'occlusions de paupière et n'est bien évidemment pas officielle. C'est pourquoi nous utiliserons la base *CASIA*, plus difficile à traiter ; mais nous permettant de comparer nos résultats à d'autres algorithmes existants.

4.3.4 Base de Données Multimodales

4.3.4.1 A propos des bases de données "chimériques" et des utilisateurs "virtuels"

Le problème majeur auquel nous sommes confrontés lorsqu'il s'agit de travailler en biométrie multimodale est le **manque de bases de données d'utilisateurs "réels"**. Le mot *réel* signifiant ici que les **modalités biométriques** proviennent de le **même personne**. A notre connaissance, il n'existe que deux grandes bases de données multimodales constituées d'utilisateurs réels combinant les modalités du visage et de l'iris : la première est bien sûr la base de données *Biosecure*, extrêmement complète comprenant un grand nombre d'utilisateurs, malheureusement, elle ne sera disponible qu'à partir de fin 2008 ! La deuxième est la toute récente base de données **MBGC** (“***M**ultiple **B**iometric **G**rand **C**hallenge*”) [MBGC 08] disponible seulement depuis Mai 2008.

Cependant, nous pouvons toujours utiliser les bases de données *FERET* et *CASIA* que nous venons de présenter pour effectuer des expérimentations multimodales, mais cela implique alors la *combinaison de modalités biométriques provenant de différentes base de données*. Cette opération provoque la création de ce que l'on appelle des **utilisateurs virtuels**, crées par combinaison de traits biométriques de différents sujets [Dori 06] [Garc 05]. Une telle assignation arbitraire est une pratique courante dans la littérature de la biométrie multimodale et a été profondément discutée lors du *2003 Workshop on Multimodal User Authentication* [Duge 03].

Au moins **deux arguments** (le premier est d'ordre *technique*, le second est d'ordre *éthique*) peuvent justifier de l'utilisation de sujets virtuels :

- **L'hypothèse d'indépendance des modalités** : en général, plusieurs modalités biométriques d'une même personne sont souvent considérées comme étant indépendantes entre-elles [Hami 94]. Une démonstration utilisant une matrice de corrélation construite à partir de classifieurs de visage et de la parole peut être trouvée dans [Kitt 98].

- **Les problèmes liés à la vie privée** : les personnes qui participent à des expériences de biométrie, en particulier à l'élaboration d'une base de données multimodale, ne souhaitent souvent pas que les organismes et les instituts puissent rassembler et conserver une trop grande quantité de leurs informations personnelles au même endroit.

Par ailleurs, il a été montré [Poh 05b] qu'un opérateur de fusion provenant d'une database chimérique **n'améliore ni ne dégrade la performance d'un système biométrique** (sur des utilisateurs réels) par rapport à son "entraînement" sur une base de données de sujets réels. Cependant, l'avantage est qu'une **plus grande quantité de données d'entraînement peuvent être générées artificiellement**, permettant ainsi de surmonter les problèmes relatifs au manque de données multimodales d'entraînement.

Plus récemment, les expérimentations et les résultats provenant de [Sche 07] montrent que, dans la plupart des cas, l'utilisation de sujets virtuels conduit à une précision inférieure que celle que l'on obtiendrait avec l'utilisation de personnes réelles en termes de taux d'égale erreur (*EER*) ; cela signifie donc que l'**on minorerait** la performance globale d'un système par l'utilisation de base de données virtuelles et que l'on peut s'attendre à ce que **le même système utilisant des utilisateurs réels ait une performance légèrement supérieure**. Enfin, une conclusion importante de [Sche 07] est que la **création de bases de données multimodales utilisant des données officielles pour des évaluations biométriques est fortement encouragée**.

4.3.4.2 Conception d'une base de données multimodale

Un autre point important concerne la construction d'une base de données multimodale. Dès lors que l'on travaille avec plusieurs modalités, **il est nécessaire que la base de données soit structurée par utilisateur**. Au moins **deux architectures** sont possibles (Fig. 4.9) :

1. **Architecture I** : dans ce cas, il existe une **seule base de données** comprenant **un dossier (et son identifiant) par personne**. Chaque dossier rassemble les données biométriques d'un utilisateur, provenant de différentes modalités biométriques,

2. **Architecture II** : dans cette configuration, il existe **plusieurs bases de données** (une base par modalité biométrique). Ici, chaque base de données comprend toujours un dossier (et son identifiant) par utilisateur mais **chaque dossier ne contient que les informations biométriques de la modalité correspondante à la base de données**. On comprend alors qu'il est nécessaire que **les dossiers** contenant les informations biométriques relatives à une **même personne** doivent être **ordonnés**, c'est-à-dire avoir le **même identifiant**.

l'architecture I à l'avantage d'être **très pratique à mettre en place**. Cependant, cela signifie que la totalité des informations biométriques des utilisateurs est rassemblée à la **même place**, ce qui peut rendre le système plus sensible aux cas de *panne* (ou de *dysfonctionnement*) ou encore aux *tentatives d'intrusion*.

L'architecture II est **plus lourde à gérer** mais elle autorise par exemple plusieurs organisations ou instituts à stocker les informations biométriques des utilisateurs sur différents serveurs (dans le cas où les données sont envoyées à distance) en évitant de concentrer toutes les données biométriques au même endroit. Elle permet également d'**augmenter la probabilité de toujours faire fonctionner au moins une modalité biométrique**.

Pour la création de notre base multimodale, nous avons opté pour *l'architecture I* et utilisé les bases de données *FERET* et *CASIA*. Nous donnerons quelques détails sur la façon dont nous avons pu générer 338 classes pour la modalité de l'iris ainsi que l'utilisation de l'ensemble de données le plus grand de FERET ("*FB Probe Set*") pour la modalité du visage dans le chapitre suivant.

4.4 Conclusion

Dans ce chapitre, nous avons détaillé les deux dispositifs de capture en environnement réel que nous avons utilisé pour acquérir des images de visage et d'iris ainsi que les différentes bases de données officielles que nous utiliserons dans nos expérimentations. Ensuite, nous avons insisté sur une question fondamentale concernant les bases de données "chimériques" et les utilisateurs "virtuels" en biométrie multimodale et donné quelques raisons qui nous permettent de les utiliser. Enfin, nous avons présenté deux types d'architectures possibles de structures de base de données multimodales afin de pouvoir conduire des expérimentations cohérentes.

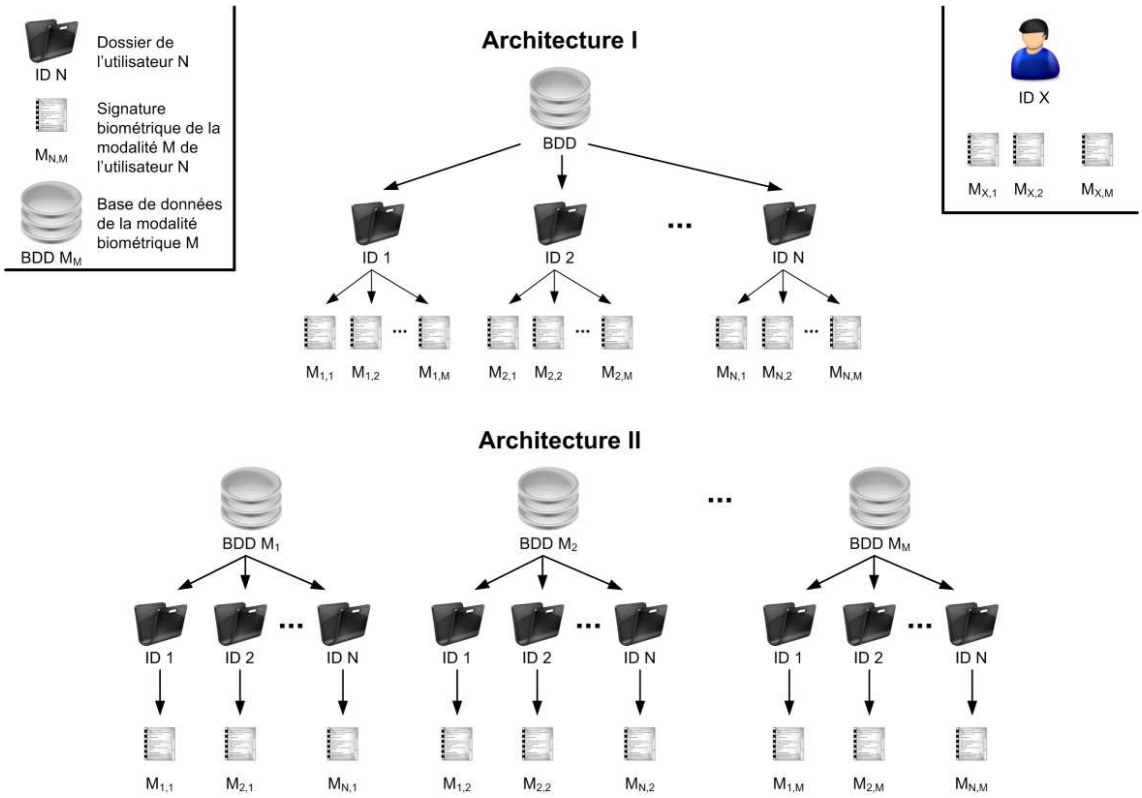


FIG. 4.9 – Deux architectures possibles pour la conception d'une base de données multimodale.

Chapitre 5

Analyses Statistiques des Modalités Biométriques (Iris & Visage)

5.1 Introduction et définition du problème

L'objectif principal de ces analyses statistiques est d'étudier les distributions des scores en sortie des modules de reconnaissance du visage et de l'iris, afin de pouvoir les modéliser mathématiquement. Nous souhaitons montrer que, sous certaines conditions, les scores imposteurs peuvent être assimilés à du bruit blanc gaussien.

Une manière courante de travailler au niveau score est l'utilisation d'une **matrice de similarité**. Un telle matrice contient les scores provenant d'un module de reconnaissance biométrique et exprime la similarité entre un **ensemble-cible** ("*Target Set*") $\mathcal{T}$, c'est-à-dire les **sujets connus du système** (la base de données) et un **ensemble de requêtes** ("*Query Set*") $\mathcal{Q}$, c'est-à-dire **les individus qui doivent être testés par le système**.

Ainsi, soit T le nombre de personnes de $\mathcal{T}$ et Q le nombre d'individus de $\mathcal{Q}$. Soit $s_{q,t}$ la réalisation d'une variable aléatoire S correspondant au score de similarité fourni par un module de reconnaissance biométrique, après comparaison du $q^{\text{ème}}$ individu de $\mathcal{Q}$ avec la $t^{\text{ème}}$ personne de $\mathcal{T}$, $q = 1, ..., Q$; $t = 1, ..., T$. Soit $M_{Q,T}$ une matrice de taille $(Q \times T)$ et d'élément générique $s_{q,t}$. $M_{Q,T}$ est appelée **matrice de similarité** telle que :

$$M_{Q,T} = \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,T} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,T} \\ \vdots & \vdots & s_{q,t} & \vdots \\ s_{Q,1} & s_{Q,2} & \cdots & s_{Q,T} \end{pmatrix} \tag{5.1}$$

En pratique, l'étude d'une matrice de similarité doit s'effectuer en **mode fermé** ("*Closed Set*"), c'est-à-dire que les **mêmes personnes** doivent être dans les deux ensembles (donc $Q = T$) mais avec des **instances différentes de leur signature biométrique**, ce qui aboutit à une matrice de similarité **carrée**. Ceci se fait généralement en extrayant respectivement de $\mathcal{T}$ et $\mathcal{Q}$, **deux sous-ensembles** arbitraires de **même taille** appelés **galerie** ("*Gallery Set*") $\mathcal{G}$ et **ensemble-test** ("*Probe Set*") $\mathcal{P}$.

Nous pouvons alors identifier **deux types de scores** : les **scores authentiques** (“*genuine scores*”) qui résultent de la comparaison d'éléments de la **même personne** (ces scores apparaissent le long de la diagonale principale de la matrice de similarité) et les **scores imposteurs** (“*impostor scores*”) qui sont le résultat de comparaisons de **personnes différentes** ; ces scores sont souvent supposés être *indépendants et identiquement distribués* (i.i.d.), avec des valeurs de moyenne μ et de variances σ^2 finies, pour des raisons d'*inférence statistique*[1] [Indo 03] [Nand 08].

Considérons maintenant une matrice de similarité carrée M_N de taille $(N \times N)$. M_N possède alors $N \times (N-1)$ scores imposteurs et N scores authentiques. Les estimateurs statistiques de ces distributions de scores peuvent être erronés si N est trop petit, conduisant à de mauvaises modélisations mathématiques. Il convient alors d'utiliser des outils statistiques précis, possédant une faible marge d'erreur, afin d'améliorer la précision de ces modélisations.

Dans ce qui suit, nous présentons un de ces outils statistiques, le *Bootstrap*, qui a été utilisé pour analyser des distributions de scores de similarité puis nous montrons comment modéliser mathématiquement les scores imposteurs. Enfin, nous expliquons dans quelles conditions les scores imposteurs issus des modalités du visage et de l'iris peuvent être assimilés à du bruit blanc gaussien.

5.2 Le Bootstrap

En 2007, nous fêtions "l'anniversaire d'argent" du *bootstrap*. En effet, cela fait 30 ans que son inventeur, *Bradley Efron*, a donné son premier cours sur le bootstrap à l'Université Stanford en 1977, publié deux ans plus tard dans [Efro 79] et repris en 1993 par *Efron et Tibshirani* [Efro 93]. Le terme *bootstrap* est souvent associé au conte du *Baron Von Münchhausen* qui, à un moment du récit, se sauve d'un marécage en tirant uniquement sur ses boucles de bottes (“*bootstraps*”). De manière nettement moins imagée, ce passage souligne le fait que le Baron se serait sorti d'une situation inextricable par ses propres moyens, ce qui se traduit en anglais par la phrase idiomatique “*Pull yourself up by your bootstraps*”. Cette analogie peut suggérer que le bootstrap est capable d'effectuer "l'impossible" et qu'il a parfois donné lieu à des résultats inattendus, notamment dans le traitement de données réelles. Clairement, le bootstrap n'est pas une technique magique qui fournit une solution à tous les problèmes d'inférence statistique, mais il a le pouvoir de substituer des axiômes analytiques fastidieux par des calculs par ordinateur.

La première question que l'on peut se poser est alors la suivante : "**A quoi sert le bootstrap ?**" Nous pouvons répondre de manière générale par "*le bootstrap est un outil de calculs utilisé pour résoudre des problèmes d'inférence statistique*". La deuxième question qui suit logiquement la première est : "**Quand pouvons-nous utiliser le bootstrap ?**" Une réponse simple serait "*Quand nous avons peu d'informations à propos des statistiques des données ou que nous avons une faible quantité de données qui nous empêchent d'utiliser des résultats asymptotiques*."

[1] L'inférence statistique consiste à induire les caractéristiques inconnues d'une population à partir d'un échantillon issu de cette population.

5.2.1 Quand le nombre de données est faible

Supposons que nous ayons collecté n mesures dans le vecteur $x = \{x_1, x_2, ..., x_n\}$, ces dernières étant des réalisations d'une variable aléatoire $X = \{X_1, X_2, ..., X_n\}$, provenant d'une distribution inconnue F_X. Soit $\hat{\theta} = \hat{\theta}(X)$ un estimateur d'un paramètre quelconque θ de F_X, qui peut être, par exemple, la moyenne $\theta = \mu_X$ de F_X estimé à partir de la moyenne empirique $\hat{\theta} = \hat{\mu}_X = 1/n \sum_{i=1}^{n} X_i$.

Le but est de trouver les caractéristiques de $\hat{\theta}$ à partir de sa distribution. Parfois, l'estimateur $\hat{\theta}$ est calculé à partir d'une collection de n données indépendantes et identiquement distribuées (i.i.d) $X_1, X_2, ..., X_n$. Si la fonction de distribution F_X est connue (ou supposée être connue) et étant donné que la fonction $\hat{\theta}(X)$ est relativement simple, alors il est possible d'évaluer exactement la distribution de l'estimateur $\hat{\theta}$.

Dans de nombreuses applications pratiques, soit la distribution F_X est inconnue soit l'estimateur $\hat{\theta}(x)$ est trop compliqué pour que sa distribution puisse être représentée de manière simple. Viennent alors les questions suivantes : "*A quel point l'estimateur $\hat{\theta}$ est-il fiable en termes de biais et de variance ?*" ou encore "*Comment pourrions-nous, par exemple, savoir si le paramètre θ est significativement différent d'une quelconque valeur nominale* ?" Nous pourrions utiliser des arguments asymptotiques et approximer la distribution de $\hat{\theta}$. Dans le cas de la moyenne statistique $\hat{\mu}_X$ introduite plus haut, nous voudrions appliquer le *théorème de la limite centrale* en supposant que la distribution de $\hat{\mu}_X$ est gaussienne, ce qui permettrait de répondre aux questions d'inférence statistique.

Mais alors comment peut-on procéder si le *théorème de la limite centrale* ne s'applique pas parce que n est faible et que nous ne pouvons pas répéter l'expérience ? **C'est à ce moment précis qu'intervient le bootstrap**.

5.2.2 Principe du Bootstrap : échantillonnage avec remplacement

Dans un contexte d'analyse de données, le mot "*bootstrap*" signifie qu'un ensemble de n observations disponibles peut donner naissance à beaucoup d'autres vecteurs de mesure par ré-échantillonnage. Tandis que l'objectif original de la méthode de "*cross-validation*" [Koha 95] est de vérifier la reproductibilité des résultats et que celle de la méthode "*jackknife*" [Thom 82] est de détecter les valeurs aberrantes, *Efron* a développé la méthode du bootstrap à des fins d'inférence statistique.

La méthode du bootstrap donne des estimateurs de l'erreur-type[2] de la médiane en prélevant de manière répétée des échantillons de bootstrap ("*bootstrap samples*") à partir des données originales, en réévaluant la médiane pour chaque échantillon de bootstrap et en estimant l'erreur-type de la médiane originale par la variation observée dans les médianes du bootstrap.

Dans le bootstrap, **les données originales peuvent être dupliquées autant de fois que les ressources de l'ordinateur le permettent** ; cet ensemble de données supplémentaire est traité comme une partie d'une population virtuelle. Ensuite, les échantillons sont pris parmi cette population afin de vérifier les estimateurs. De plus, contrairement aux

[2]L'erreur-type est l'écart-type estimé de l'erreur de cette estimation. C'est donc une estimation de l'écart-type entre les valeurs mesurées ou estimées et les vraies valeurs d'une distribution.

méthodes de "*cross-validation*" et "*jackknife*", le bootstrap emploie de **l'échantillonnage avec remplacement** (Fig. 5.1). En effet, le fait de ré-échantillonner avec remplacement dans un bootstrap est **plus précis** que de ré-échantillonner sans remplacement en termes de simulation de chance de prendre tel ou tel échantillon. Par ailleurs, dans les méthodes de "*cross-validation*" et "*jackknife*", la valeur n dans l'ensemble dupliqué est plus petit que celui dans les données originales, mais **dans le bootstrap, chaque échantillon de bootstrap possède le même nombre d'observations que dans les données originales**. Ainsi, la méthode du bootstrap a **l'avantage de modéliser les impacts concernant la taille des données étudiées** [Fan 96].

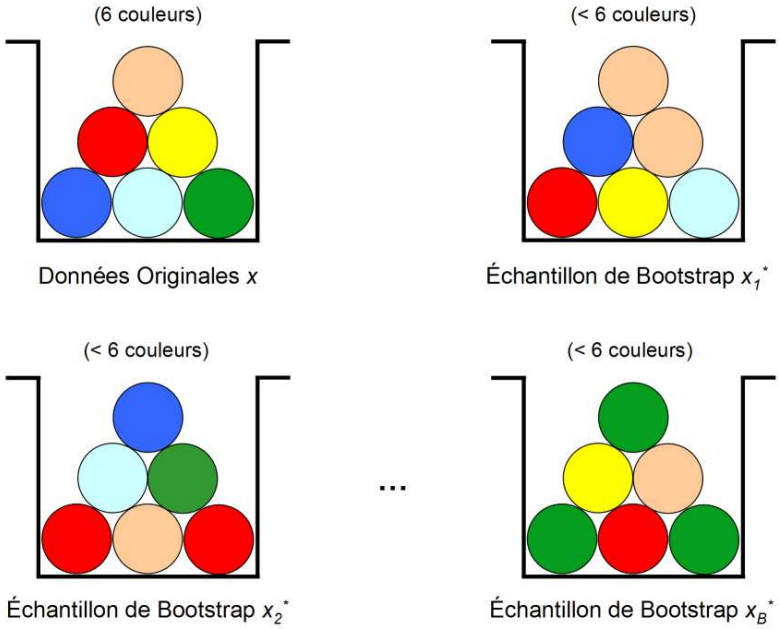


FIG. 5.1 – Le principe du bootstrap par ré-échantillonnage de données indépendantes avec remplacement.

5.3 Modélisation mathématique des scores

Afin de modéliser mathématiquement les scores issus des deux modalités biométriques de l'iris et du visage, nous allons appliquer la méthode du bootstrap avec remplacement aux scores imposteurs issus de matrices de similarité. Cette méthode va nous permettre de générer un grand nombre d'échantillons selon un protocole bien précis et nous pourrons alors essayer de trouver un modèle statistique qui correspond le mieux aux distributions de ces scores.

Un problème courant et important dans l'utilisation de la méthode du bootstrap consiste à trouver le **nombre de répétitions de bootstraps**, noté B. Afin de déterminer une valeur convenable pour B, nous allons suivre le protocole décrit dans [Andr 00] : nous spécifions une limite de pourcentage d'écart, notée *pdb*, et nous demandons que le pourcentage d'écart *réel* soit inférieur à cette limite avec une probabilité fixée, notée $1-\tau$,

proche de 1. Ce protocole prend comme arguments *pdb* et τ et fournit une méthode dépendante des données afin de déterminer une certaine valeur de *B*, notée *B**, de manière à atteindre le niveau de précision souhaité. Par exemple, lorsque $(pdb, \tau) = (10, 0.05)$, on obtient une valeur de *B** de sorte que le pourcentage d'écart soit inférieur à 10% avec une probabilité d'approximation de 95%.

5.3.1 Matrices de similarité du visage et de l'iris

La matrice de similarité du visage (FSM) a été construite à partir de la base de données *FERET* [FERE 01], en utilisant le *Gallery Set (FA)* et le *Probe Set (FB)* (Fig. 5.2).

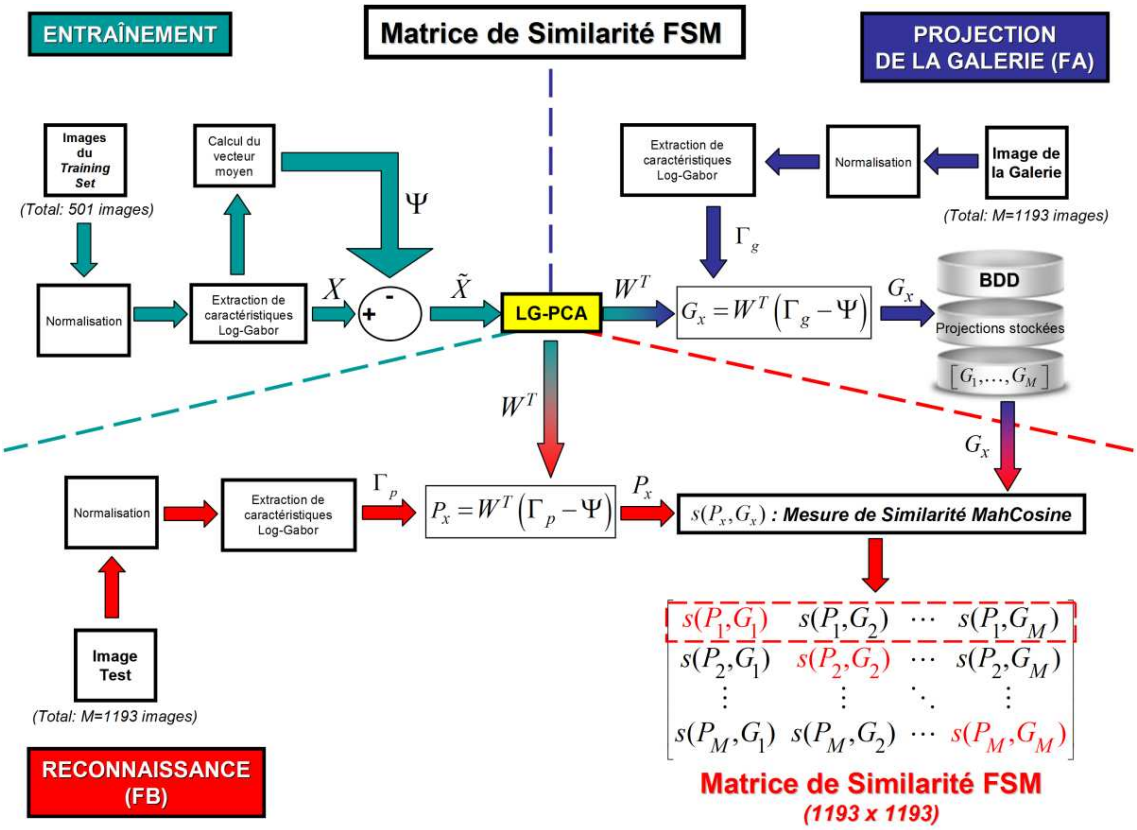


FIG. 5.2 – Schéma de construction de la matrice de similarité FSM .

Afin de pouvoir opérer en *mode fermé*, certains conflits d'identifiants, relevés par le programme FERET, ont entraîné l'enlèvement de deux scores dans le *Probe Set* et trois scores dans le *Gallery Set*. Le prétraitement des images de visages (alignement et normalisation) a été effectué en s'inspirant du *CSU System 5.0* [Bolm 03]. Le processus de calcul des scores de similarité pour l'algorithme LG-PCA [Perl 05] est le suivant : des vecteurs de caractéristiques "Log-Gabor" Γ_g sont d'abord extraits d'images d'apprentissage normalisées et l'algorithme PCA leur est ensuite appliqué ; les vecteurs de données G_x issus de la projection des vecteurs de caractéristiques sur le "*Face Space*" sont alors stockés dans une base de données afin de servir d'éléments de comparaison avec d'autres projections P_x issues d'images de test normalisées. Pour comparer deux projections, la mesure de similarité *MahCosine* [Beve 03] (*cf. Annexe A.2.4*) $s(P_x, G_x)$ a été utilisée. Ainsi, la matrice FSM est de taille (1193×1193) (Fig. 5.3).

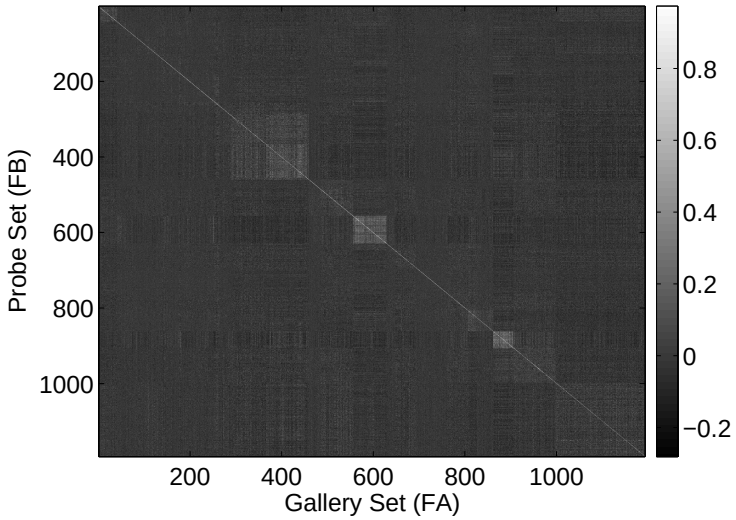


FIG. 5.3 – FSM : Matrice de similarité du visage (FERET FA vs. FB).

La matrice de similarité de l'iris (**ISM**), a été construite à partir de la base de donnée *CASIA* [CASI 05] en utilisant l'ensemble de données *CASIA-IrisV3-Interval* (Fig. 5.4).

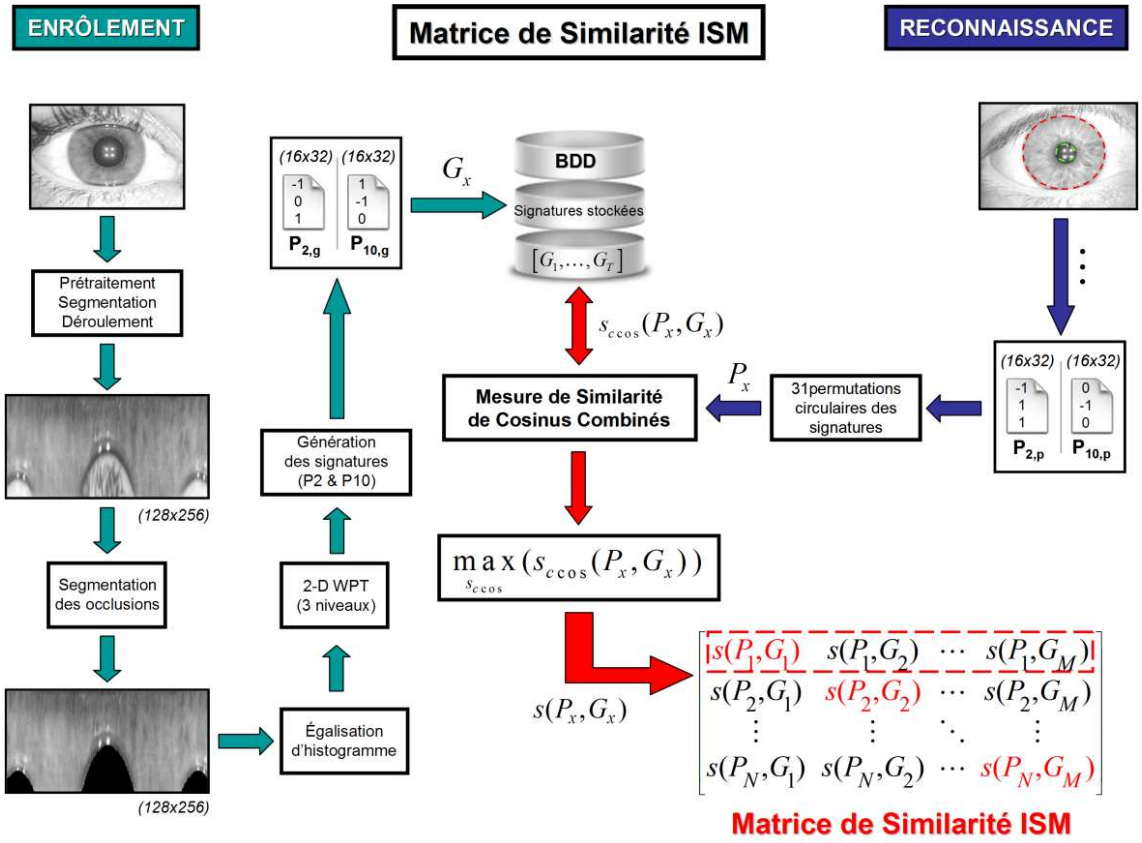


FIG. 5.4 – Schéma de construction de la matrice de similarité ISM .

Les signatures biométriques de l'iris sont générés à partir de l'algorithme qui a été développé dans les laboratoires de l'ISEP [Rydg 04] et qui utilise une transformée en paquets d'ondelettes sur trois niveaux de l'image déroulée de l'iris (notre méthode diffère de celle de *J. Daugman* [Daug 04] qui utilise une des ondelettes de Gabor). Cependant, une **modification a été apportée** dans l'étape de codage des coefficients d'ondelettes qui utilisait jusqu'alors un seuil fixe (trop sensible aux changements de conditions de capture et variations entre les différentes bases de données) ; il a été **remplacé par un seuil dynamique** prenant en compte 98.5% **de l'énergie des coefficients d'ondelettes**. Ainsi, 338 classes différentes ont été générées (les deux yeux d'un même individu pouvant compter pour deux classes différentes grâce à *l'hypothèse d'indépendance statistique*), cinq iris par classe dont trois pour le *Target Set* (donc $338 \times 3 = 1014$ signatures) et deux pour le *Query Set* (donc $338 \times 2 = 676$ signatures).

Lorsque les signatures d'iris sont générées, on ne peut pas savoir précisément le degré d'inclinaison de la tête et donc de la rotation des yeux. Pour **corriger cet effet de rotation**, il est nécessaire d'effectuer les comparaisons entre toutes les paires de signatures d'iris sur une **échelle raisonnable de mouvements relatifs**. Dans notre algorithme, nous prenons en compte des rotations allant de -15^o à $+15^o$, donc 31 permutations circulaires soit une permutation par unité d'angle. Les permutations circulaires sont effectuées sur chaque signature d'iris de requête $\overrightarrow{Q}$ avant de calculer une mesure de similarité de cosinus combinés (5.2) avec toutes les signatures d'iris cible $\overrightarrow{T}$.

$$S_{ccos}(\overrightarrow{Q}, \overrightarrow{T}) = \frac{1}{2} \prod_{i=\{2,10\}} \left(\frac{\overrightarrow{Q}_{P_i} \cdot \overrightarrow{T}_{P_i}}{\|\overrightarrow{Q}_{P_i}\| \times \|\overrightarrow{T}_{P_i}\|} \right) , \tag{5.2}$$

où $\overrightarrow{Q}_{P_i}$ et $\overrightarrow{T}_{P_i}$ représentent respectivement les signatures d'iris de requête et les signatures d'iris cible du $i^{\text{ème}}$ paquet d'ondelettes P_i et $\|.\|$ la norme euclidienne [Rydg 04].

L'effet de rotation des yeux est corrigé en conservant la meilleure (plus grande) mesure de similarité de cosinus combinés parmi toutes les permutations circulaires comme score de similarité final.

Finalement, nous obtenons d'abord une matrice de similarité de taille (676×1014) puis nous gardons aléatoirement une ligne sur deux (individu de requête aléatoire) et une colonne sur trois (correspondance-cible aléatoire) afin d'obtenir une matrice ISM finale de taille (338×338) (Fig. 5.5).

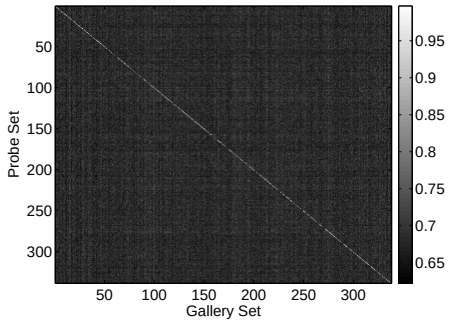


FIG. 5.5 – ISM : Matrice de similarité de l'iris (CASIA-IrisV3-Interval).

5.3.2 Analyse des distributions des scores imposteurs "bootstrappés"

Pour chaque matrice de similarité générée précédemment (FSM et ISM), une matrice *d'entraînement* plus petite (40% de la taille de la matrice de similarité) est extraite aléatoirement. Ensuite, les scores imposteurs de cette matrice d'entraînement sont "boostrappés" avec les paramètres $(pdb, \tau) = (2, 0.01)$. C'est-à-dire que le nombre de bootstraps est estimé de sorte que le pourcentage d'écart soit inférieur à 2% avec une probabilité approximative de 99%, nous garantissant une grande précision dans l'utilisation du bootstrap. Les figures 5.6 et 5.7 illustrent respectivement les distributions des scores imposteurs bootstrappés pour le visage et l'iris, provenant des matrices aléatoires d'entraînement.

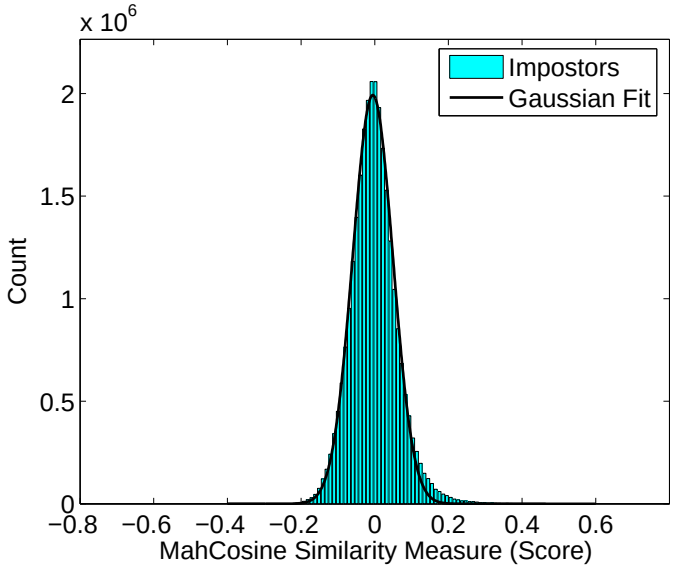


FIG. 5.6 – Distribution des scores imposteurs bootstrappés pour le visage.

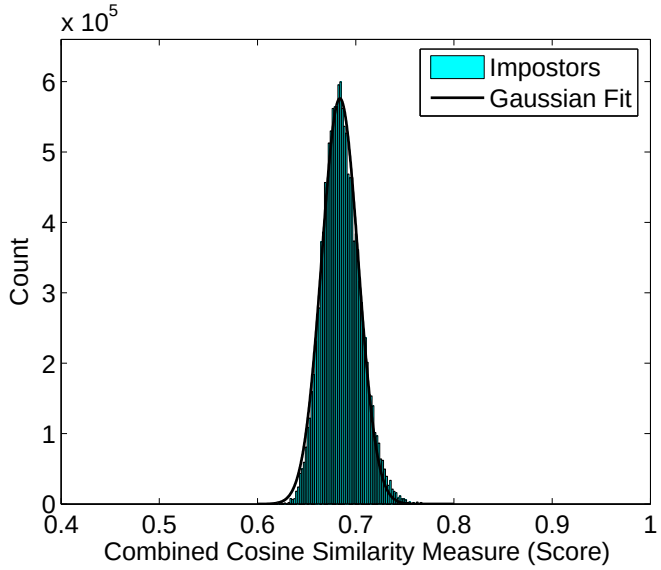


FIG. 5.7 – Distribution des scores imposteurs bootstrappés pour l'iris.

Avant de poursuivre notre raisonnement et d'analyser ces deux distributions issues du bootstrap, il convient de faire deux remarques importantes :

- Tout d'abord, en pratique, les distributions des scores imposteurs ne sont pas censées être gaussiennes et ce, quelles que soient les modalités considérées, même sans l'utilisation de valeurs extrêmes [Wu 07]. Cependant, on peut retenir de [Poh 05a] la chose suivante : bien que l'hypothèse gaussienne ne soit pas toujours validée, les **1186 expériences** conduites sur la base de données multimodale **BANCA** (visage et voix) [Bail 03] montrent que les points des taux d'égale erreur (EER) théoriques (prédits grâce à l'hypothèse gaussienne) correspondent approximativement aux points d'EER calculés directement de manière empirique à partir des scores de similarité. **Ce dernier résultat constitue une forte indication sur le fait que l'hypothèse gaussienne peut être acceptable en pratique**.

- Ensuite, nous avons précisé qu'afin de pouvoir corriger l'effet de rotation des yeux, des permutations circulaires des signatures d'iris avaient été effectuées avant de calculer les scores d'iris et de garder le score maximum comme score final. Ce schéma de sélection du score maximum a également été utilisé par *J. Daugman* [Daug 06], un des pionniers de la reconnaissance par l'iris. Les scores imposteurs qu'obtient *J. Daugman* décrivent une *distribution de valeurs extrêmes* que ce dernier approche par une distribution binomiale de valeurs extrêmes, possédant une asymétrie orientée vers les scores de similarité les plus élevés. Le choix de son approximation s'explique par le fait qu'il utilise une distance de Hamming pour comparer deux "*IrisCodes*" et que ces comparaisons sont alors considérés comme des essais de Bernoulli [Daug 06] ; ce qui diffère de notre cas puisque l'on utilise une mesure de similarité de cosinus combinés (5.2). De plus, l'analogie avec des essais de Bernoulli est discutable car un "*IrisCode*" possède des corrélations internes provenant des caractéristiques de l'iris et des corrélations introduites par les filtres d'ondelette 2-D de Gabor (en amplitude à cause de la partie passe-bas et en phase à cause de la partie passe-bande) [Daug 06]. Par ailleurs, si les scores imposteurs d'iris doivent être générés à l'aide de la mesure de distance de Hamming (cas standard), on peut toujours approximer la distribution binomiale des scores imposteurs par une distribution normale sous réserve de remplir les conditions suivantes :

  1. Le nombre d'échantillons n doit être assez grand (la technique du bootstrap peut aider à remplir cette première condition,
  2. L'asymétrie de la courbe ne doit pas être très prononcée,
  3. Une correction de continuité convenable[3] doit être utilisée.

Si ces trois conditions sont remplies, alors une très bonne approximation de la distribution binomiale $B(n, p)$ est donnée par la distribution normale $N(np, np(1-p))$ [Zven 83] (*preuve en Annexe C.1*).

[3]Une correction de continuité doit s'appliquer lorsqu'on approche une loi discrète par une loi continue, il faut réécrire les probabilités de la fonction de masse $p(X = k)$ sous la forme d'une probabilité d'intervalle. Lorsque les valeurs du support de X sont des nombres entiers consécutifs, comme c'est le cas pour la loi binomiale, la probabilité $p(X = k)$ doit se réécrire $p(k - 0,5 < X < k + 0,5)$ pour que l'on puisse effectuer le calcul de l'aire correspondante dans le modèle continu.

	Visage (FSM)	Iris (ISM)
Taille originale des matrices de similarité	(1193×1193)	(338×338)
Taille des matrices aléatoires d'entraînement	(477×477)	(135×135)
Nombre de répétitions de bootstraps (B*)	121	763
Nombre de scores imposteurs bootstrappés	$27,473,292$	$13,802,670$
Paramètres des courbes gaussiennes (μ, σ)	$(-0.005, 0.054)$	$(0.683, 0.018)$
Coefficient de corrélation linéaire de Pearson (R)	0.998	0.996
Estimateurs de la moyenne et de l'écart-type après normalisation Z-Score $(\widehat{\mu}, \widehat{\sigma})$	$(-0.012, 0.976)$	$(0.038, 0.920)$

TAB. 5.1 – Analyses Statistiques des Scores du Visage et d'Iris

L'observation des figures 5.6 et 5.7 nous montre que nous avons en effet une *légère asymétrie* dans les deux distributions (un *skewness* de 0.5 pour le visage et de 0.4 pour l'iris). Malgré ce phénomène, elles *semblent tout de même proches d'une distribution gaussienne*. Pour vérifier cette hypothèse, les deux distributions ont été approchées par des **courbes gaussiennes** (5.3) en utilisant une approximation au sens des **moindres-carrés**, ce qui est représenté par les **traits continus** sur les figures 5.6 et 5.7.

$$y(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5.3}$$

De plus, afin de quantifier la qualité de cette approximation, nous avons calculé le **coefficient de corrélation linéaire de Pearson** (5.4) [Edwa 76], noté R :

$$R = \sqrt{\frac{(\sum xy - n\bar{x}\bar{y})^2}{(\sum x^2 - n\bar{x}^2)(\sum y^2 - n\bar{y}^2)}}\ , \tag{5.4}$$

où $R \in [0, 1]$ ($R = 1$ correspondant à une parfaite corrélation) et n le nombre de points de données (x, y) centrés par rapport à leurs moyennes respectives $\bar{x}$ et $\bar{y}$.

Les résultats de ces premières analyses statistiques sont résumés dans le tableau 5.1. On peut tout d'abord noter que les **nombres de bootstraps nécessaires** pour obtenir la précision que nous nous sommes fixés **sont très différents** selon les matrices de similarité ; cela confirme bien que le bootstrap *s'adapte* aux données originales qui sont moins nombreuses dans le cas de l'iris et qui requièrent donc un nombre de bootstraps plus élevé.

De plus, **les nombres de scores imposteurs bootstrappés qui ont été générés sont très élevé** (**plusieurs millions**), ce qui nous place dans la configuration dans laquelle on peut considérer que notre nombre n est suffisamment grand afin de pouvoir effectuer une hypothèse gaussienne. Cette dernière peut d'ailleurs être validée par la très forte valeur du coefficient linéaire de Pearson (très proche de 1).

La dernière ligne de ce tableau va être expliquée dans la sous-section suivante qui apporte les dernières analyses statistiques nécessaires à la fin de notre raisonnement.

5.3.3 Vers le bruit blanc gaussien

Nous venons de voir que les distributions des scores imposteurs bootstrappés pour les modalités du visage et de l'iris **suivent assez bien des distributions gaussiennes**. Nous allons donc enfin pouvoir vérifier si nous pouvons les assimiler à du **bruit blanc gaussien** [Hami 94].

Nous devons tout d'abord nous assurer que nos distributions de scores imposteurs bootstrappés possèdent une moyenne nulle et un écart-type unitaire ; ceci est rendu possible en leur appliquant la normalisation *Z-Score*, qui est optimale dans le cas de données suivant une loi gaussienne [Jain 05]).

D'autre part, nous devons calculer pour chaque modalité, l'autocorrélation des scores imposteurs bootstrappés et normalisés afin de la comparer à l'autocorrélation d'un "vrai" bruit blanc gaussien, en évaluant $\forall t \neq \tau, E[\varepsilon_t \varepsilon_\tau]$ (la moyenne des coefficients d'autocorrélation, excepté le coefficient central). Cette comparaison se fait de manière objective, en calculant l'erreur relative.

Pour ce faire, nous avons respecté, pour chaque modalité, le protocole suivant :

1. Extraction d'une matrice aléatoire d'entraînement (40% de la taille de la matrice de similarité originale),
2. Bootstrap des scores imposteurs avec les paramètres du tableau 5.1,
3. Normalisation Z-Score des scores imposteurs bootstrappés,
4. Extraction aléatoire de 10^6 des scores imposteurs bootstrappés et normalisés (pour des raison de ressources CPU en ce qui concerne l'étape suivante),
5. Calcul de l'autocorrélation des 10^6 scores,
6. Calcul de l'autocorrélation d'un "vrai" bruit blanc gaussien constitué de 10^6 échantillons (fonction *randn* sous MATLAB),
7. Calcul de l'erreur relative.

Étant donné que les étapes 1, 4 et 6 font intervenir un tirage aléatoire, **nous avons répété ce protocole de test 100 fois** afin d'obtenir des résultats plus représentatifs. **Nous avons ainsi obtenu une erreur relative de 3.07% pour le visage et une erreur relative de 3.10% pour l'iris, ce qui est bien acceptable en pratique**. Les figures 5.8, 5.9 et 5.10 illustrent respectivement les autocorrélations des scores imposteurs bootstrappés normalisés pour le visage et pour l'iris ainsi que l'autocorrélation d'un bruit blanc gaussien généré par la fonction *randn* (10^6 échantillons) sous MATLAB.

En pratique, dans un système biométrique multimodal dont la fusion s'opère au niveau score, l'étape de normalisation de score doit également être effectuée sur les scores authentiques. La dernière ligne du tableau 5.1 donne les estimateurs de la moyenne et de l'écart-type des deux distributions de scores issues des matrices de similarité totales (avec prise en compte des scores authentiques), après application de la normalisation Z-Score. On se rend alors compte que l'impact des scores authentiques ($N - 1$ fois moins nombreux que les scores imposteurs) sur les estimateurs statistiques est relativement faible et que les valeurs obtenues sont tout a fait acceptables en pratique, si l'on compare aux valeurs théoriques $\mu = 0$ et $\sigma = 1$ que l'on devrait obtenir pour remplir la condition de distribution normale.

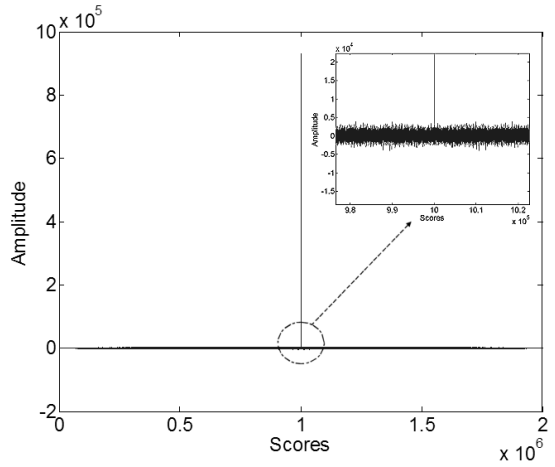


FIG. 5.8 – Autocorrélation des scores imposteurs bootstrappés normalisés pour le visage.

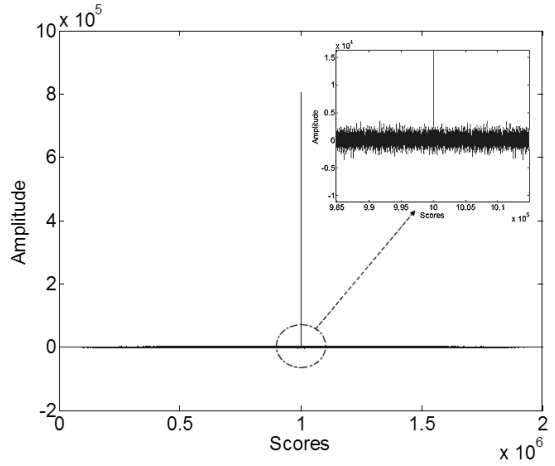


FIG. 5.9 – Autocorrélation des scores imposteurs bootstrappés normalisés pour l'iris.

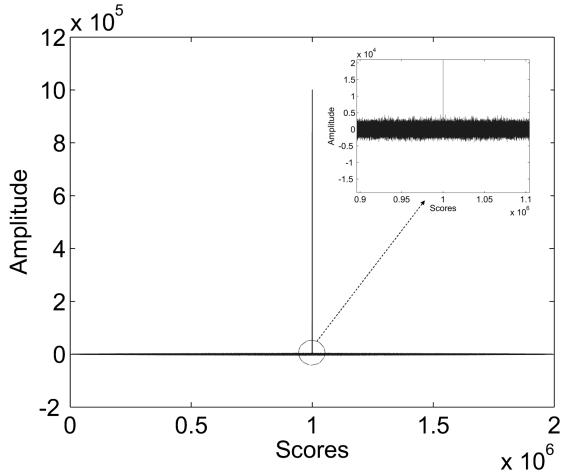


FIG. 5.10 – Autocorrélation d'un bruit blanc gaussien (fonction *randn* sous MATLAB).

5.4 Conclusion

Après avoir présenté le bootstrap, nous avons conduit des analyses statistiques à grande échelle sur les scores imposteurs bootstrappés issus des matrices de similarité du visage (base de données FERET) et de l'iris (base de données CASIA). Nous avons ensuite montré que **si l'on applique une normalisation Z-Score aux scores de similarité, les scores imposteurs peuvent être considérés comme du bruit blanc gaussien**, ce qui est confirmé par l'observation des autocorrélations des distributions de scores normalisés. **Ces résultats intermédiaires sont importants car ils vont nous permettre d'appliquer un théorème de débruitage par ondelettes, qui va contribuer à l'amélioration de la performance du système biométrique multimodal**. Enfin, nous pouvons faire une dernière remarque : notre étude s'est portée sur l'analyse des scores imposteurs mais il pourrait être intéressant de savoir comment se comporte la distribution des scores authentiques. En fait, il est connu que *la distribution des scores authentiques ne suit généralement pas une distribution connue* et peut même posséder *plusieurs modes*. Dans ce cas, on peut la modéliser de manière très précise en utilisant une *distribution discrète* [Dass 05].

Chapitre 6

Intégration des Ondelettes dans le Système Multimodal

6.1 Introduction

"Les choses devraient être rendues aussi simples que possible, mais pas plus simples." (*Albert Einstein*). Cette remarquable citation du physicien allemand est intéressante pour deux raisons qui nous concernent ici. D'une part, elle reflète assez bien le concept fondamental de l'analyse en traitement du signal, selon lequel on cherche à décomposer des signaux complexes en éléments simples. D'autre part, elle résume tout à fait le défi pédagogique que l'on doit s'imposer : il s'agit de pouvoir relier la théorie des ondelettes à des exemples concrets et des analogies simples, sans pour autant négliger d'approfondir certains points qui montrent la richesse et le formidable potentiel des ondelettes.

Les ondelettes ont permis à de nombreux spécialistes de sujets mathématiques et physiques qui s'ignoraient mutuellement, de se rencontrer, de parler un langage commun et de former une véritable "*Communauté des ondelettes*". L'analyse de Fourier est incontournable dans le traitement des signaux continus. Néanmoins, elle possède une certaine faiblesse dans le traitement des signaux comprenant des zones de non stationnarité. Le problème réside dans le fait que ces signaux transitoires sont largement représentés dans le monde physique qui nous entoure : dans le domaine biomédical (électroencéphalogramme (EEG), électrocardiogramme (ECG)), en géophysique et séismologie (détection de séismes), analyse de turbulences, etc. Les ondelettes, de part leur nature, sont devenues nécessaires pour analyser ces phénomènes non stationnaires et constituent un outil formidable en traitement du signal et de l'image.

Nous ferons tout d'abord une présentation du contexte qui a donné naissance à la transformée en ondelettes, en rappelant deux transformées classiques. Ensuite, nous étudierons de plus près différentes transformées en ondelettes (1D et 2D). Enfin, nous montrerons que les ondelettes sont capables de fournir des solutions intéressantes à divers endroits d'un système biométrique multimodal, en traitant des problèmes de compression, d'analyse et de débruitage.

6.2 Rappel de deux transformées classiques

Nous commencerons par redonner les notions fondamentales de la transformée de Fourier pour ensuite étudier la transformée de Gabor. Tout au long de ces rappels, nous insisterons sur deux phases essentielles en traitement du signal : **l'analyse** et la **synthèse**. L'analyse consiste à déployer un signal dans un espace transformé ; elle est, en général, facile tandis que la synthèse résulte plus souvent d'un "tour de force". La synthèse consiste à reconstruire le signal d'origine à partir des coefficients obtenus lors de l'analyse. Par analogie, on peut penser à la synthèse de l'urée (*Friedrich Wöhler*, 1828) qui a joué un rôle fondamental, car cela a permis d'établir un pont entre le monde de la chimie et celui du vivant.

6.2.1 La transformée de Fourier : "un prisme mathématique"

On ne peut pas parler de l'analyse par ondelettes sans commencer par parler de **l'analyse de Fourier dont elle découle**. Nous devons l'analyse de Fourier au Mathématicien et Physicien Français *Jean Baptiste Joseph Fourier* (1768-1830), considéré comme le fondateur de la physique mathématique. **De 1802 à 1809**, à Grenoble, il travaille sur la théorie de la propagation de la chaleur en modélisant l'évolution de la température avec des séries trigonométriques, et rédige son "*Mémoire sur la propagation de la chaleur dans les corps solides*". **En 1822**, il publie son article fondateur "*Théorie analytique de la chaleur*" qui donnera naissance à l'analyse de Fourier.

L'analyse de Fourier est **un des acquis majeurs de la physique et des mathématiques**. Elle est **incontournable en théorie et en traitement des signaux continus** pour de nombreuses raisons :

- La première est certainement **l'universalité du concept de fréquence** sur lequel elle est fondée : une description fréquentielle fournit un complément indispensable à la seule information temporelle (fournie par la sortie de capteurs par exemple) généralement utilisée en premier lieu pour l'analyse de signaux,

- La seconde raison tient à la **structure mathématique** elle-même de l'analyse de Fourier qui se prête très naturellement à des transformations de filtrage linéaire, en les traduisant de manière particulièrement simple,

- Une troisième raison enfin, plus pratique, est que l'ensemble de ces avantages a conduit au **développement de nombreux algorithmes et logiciels**, conférant à celle-ci la **notoriété d'usage** qu'on lui reconnaît aujourd'hui.

La transformée de Fourier (6.1) est une opération mathématique qui consiste à **décomposer une fonction selon ses fréquences de même qu'un prisme décompose la lumière**. Elle transforme une fonction originale $s(t)$ qui dépend du temps, en une nouvelle fonction $S(f)$ qui dépend de la fréquence. C'est une représentation fréquentielle **globale** du signal.

$$S(f) = \int_{-\infty}^{+\infty} s(t) e^{-2i\pi f t} dt \tag{6.1}$$

Dans la transformée de Fourier, on décompose le signal à étudier $s(t)$ selon une **base d'exponentielles complexes**, c'est-à-dire des ondes éternelles parfaitement **délocalisées dans le temps**. Une chose essentielle à retenir est que l'on peut, **en sens inverse**, **reconstruire le signal** $s(t)$ à partir de sa transformée de Fourier $S(f)$. On a en fait :

$$s(t) = \int_{-\infty}^{+\infty} S(f) e^{2i\pi ft} df \tag{6.2}$$

Cependant, cette formule de reconstruction est parfois **numériquement instable** (on peut penser au sinus cardinal), car les **phases** des nombres complexes $S(f)$ jouent un rôle essentiel et demandent à être connues avec une **très grande précision** pour que la formule de reconstruction (6.2) puisse donner une information sur le temps. De plus, mesurer une localisation temporelle d'un signal monochromatique en analysant la phase d'un terme de la transformée de Fourier est très difficile.

Pour illustrer ce phénomène, *Barbara Burke Hubbard* écrit dans [Hubb 00] : "*Théoriquement, la phase ne semble pas compliquée. En réalité, il est impossible de déterminer les phases avec assez de précision pour en extirper l'information sur le temps. La fréquence du la du diapason, par exemple, est* 440 *hertz. Peut-on savoir, à partir de la transformée de Fourier de l'enregistrement d'une symphonie, si le la est joué 20 minutes après l'ouverture ? Pour cela, il faudrait connaître sa phase avec une incertitude inférieure à* $1/(20 \times 60 \times 440)$*, c'est-à-dire* $1/528000$*. La phase reste identique durant toute la symphonie, mais les erreurs d'arrondi s'additionnent cycle après cycle. Pour savoir où l'on en est dans le cycle, on doit diviser ce nombre par au moins cinq, ce qui donne* $1/2640000$*. Connaître la phase d'une seule fréquence équivaut à mesurer un kilomètre à moins d'un demi millimètre près*".

La transformée de Fourier possède une **très bonne résolution fréquentielle** puisque la fonction analysante est une sinusoïde de fréquence précise, et quand on la multiplie par le signal, le coefficient obtenu ne **se réfère qu'à cette fréquence**. Cependant, elle ne permet pas d'analyser le **contenu fréquentiel local**, ni la régularité locale d'un signal. La question que l'on peut alors se poser est la suivante : "*Est-il possible de rendre cette analyse locale* ?".

6.2.2 Vers une analyse de Fourier locale : la transformée de Fourier à fenêtre

En 1947, le Physicien Hongrois Denis Gabor (1900-1979) qui a reçu le Prix Nobel en 1971 pour l'invention de l'holographie, suggère de **rendre locale** l'analyse de Fourier en s'aidant de **fenêtres**.

Qu'est-ce qu'une fenêtre ? Une fenêtre (ou enveloppe) est une fonction $g(t)$, régulière, lentement variable et bien localisée temporellement (c'est-à-dire qui s'annule en dehors d'une certaine zone, que l'on appelle son support). Sa représentation graphique est un morceau de courbe qui délimite une zone contenant des oscillations. En général, on la choisit symétrique et réelle. Lorsque la fenêtre est **gaussienne**, on parle de **transformée de Gabor**. Une *gaborette* (Fig. 6.1) est l'association d'une **fenêtre** et d'**oscillations internes**.

La **transformée de Fourier à fenêtre** (6.3) (aussi connue sous le nom de *transformée de Fourier à court terme*) permet d'analyser un signal à la fois en temps et en fréquence. On parle alors de **transformée temps-fréquence** [Flan 98]. L'idée consiste à décomposer

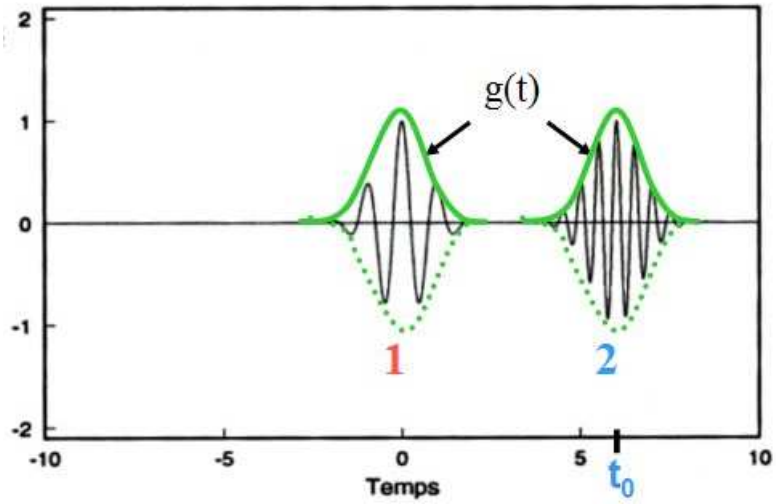


FIG. 6.1 – Illustration de deux gaborettes.

un signal $x(t)$ en fréquences, **intervalle par intervalle** : de cette manière, on limite la plage de temps analysée. La taille de la fenêtre définit la taille de l'intervalle analysé. Elle ne change pas pendant l'analyse, mais on la remplit successivement d'oscillations de fréquences différentes (Fig. 6.1).

$$W(t_0, f_0) = \int_{-\infty}^{+\infty} x(t) g(t - t_0) e^{-2i\pi f_0 t} dt \tag{6.3}$$

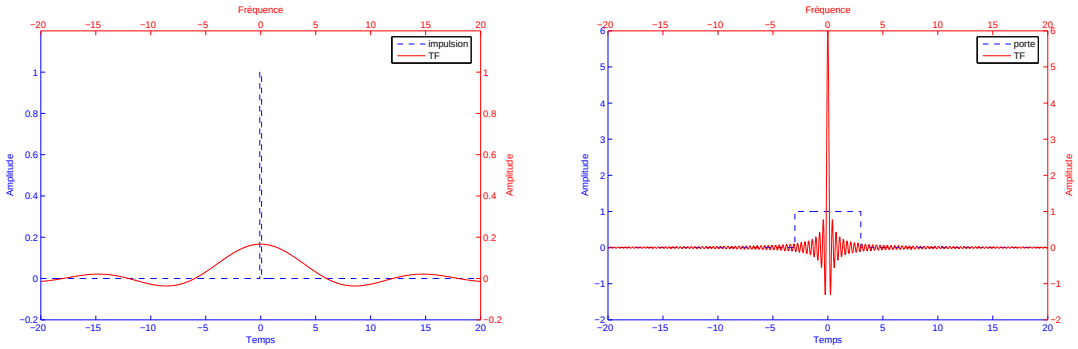
De la même manière que pour la transformée de Fourier, il est possible de reconstruire le signal original $x(t)$ à l'aide des coefficients obtenus dans l'analyse. Ceci étant, la formule d'inversion (6.4) (synthèse) est immédiate et s'écrit :

$$x(t) = c \int \int g(t - t_0) e^{2i\pi f_0 t} W(t_0, f_0) dt_0 df_0 \tag{6.4}$$

où $c > 0$ est une constante numérique dont la valeur n'a pas d'importance ici.

Tandis que la transformée de Fourier classique compare le signal entier à des sinusoïdes de durée infinie et de diverses fréquences, la transformée de Fourier à fenêtre compare un segment du signal à des **morceaux de courbes oscillantes de différentes fréquences**. Quand on a analysé un segment (c'est-à-dire lorsque l'on a effectué une transformée de Fourier à l'intérieur de la fenêtre), on répète l'opération en faisant "glisser" la fenêtre le long du signal, pour en analyser un autre.

Cependant, le fait de garder constante la taille de la fenêtre d'analyse implique un **sérieux compromis**. Avec une **fenêtre étroite**, on localise plutôt bien les **composantes transitoires** de hautes fréquences mais on devient alors aveugle aux composantes de longue durée, donc de basse fréquence, car la période du phénomène observé est trop grande pour rentrer dans une petite fenêtre. A l'inverse, quand une **fenêtre** est **large**, **on ne peut préciser l'instant où se produit un changement brutal dans le signal** (pic ou discontinuité) : cette information est noyée dans la totalité de l'information correspondant à l'intervalle de temps sélectionné par la fenêtre.



(6.2.a) (6.2.b)

FIG. 6.2 – Illustration du principe d'Incertitude d'Heisenberg.

Ce compromis peut être relié au **principe d'Incertitude d'Heisenberg** concernant la dualité onde-corpuscule. Dans son papier "*Uncertainty paper*" (1927), *Heisenberg* écrit qu'une particule élémentaire n'a pas simultanément une position et une quantité de mouvement précises. Cet énoncé, rapporté au domaine du traitement du signal, signifie qu'**un signal n'a pas simultanément une localisation précise en temps et en fréquence**.

La figure 6.2 permet de mieux comprendre le principe d'Incertitude d'Heisenberg. Dans la figure 6.2.a, on considère l'impulsion $s(t) = \mathbb{1}[-a/2; +a/2]$, $a = 1/6$, qui est **bien localisée en temps**. On s'aperçoit alors que sa transformée de Fourier $S(f) = a * sin_c(af)$ est **délocalisée en fréquence**. Dans la figure 6.2.b, on "élargit" le signal $s(t)$ en prenant $a = 6$, de manière à construire une porte un peu large, **moins bien localisée en temps**. Cette fois-ci, la transformée de Fourier est quasiment centrée en 0, donc **mieux localisée en fréquence** !

L'analyse par gaborettes a pour objet de déployer le signal dans le **plan temps-fréquence**. Cela résout le problème de la dictée musicale (qui consiste à écrire la partition en entendant la musique), mais toutes les notes ont alors la **même durée** puisque la **taille de la fenêtre d'analyse reste constante**.

Nous venons donc de comprendre que garder constante la taille de la fenêtre durant l'analyse d'un signal **ne nous donnait pas la totalité des informations** : il faut faire un choix entre la localisation des hautes fréquences ou la localisation des basses fréquences. Il a donc fallu trouver un outil induisant une **méthode de reconstruction qui soit indépendante de l'échelle d'analyse**. Ce nouvel outil s'appelle **les ondelettes**.

6.3 Les ondelettes

6.3.1 La naissance des ondelettes

Yves Meyer[1] a recensé au moins *15 origines distinctes* de la théorie des ondelettes, quelques-unes remontant jusqu'en **1930**. La communauté des physiciens de l'époque

[1] Yves Meyer est Professeur Émérite à l'École Normale Supérieure de Cachan et Membre de l'Académie des Sciences depuis 1993. Spécialiste de l'analyse harmonique, il a découvert les ondelettes orthogonales.

était intuitivement consciente de l'existence des ondelettes, depuis un article datant de **1971**, de *Kenneth Wilson*, lauréat du prix Nobel, qui décrit la renormalisation. En mathématiques, on employait les ondelettes sous le nom de "*décompositions atomiques*" (sortes de briques de construction), pour étudier différents espaces fonctionnels. D'autres chercheurs ont développé des ondelettes - sous le nom de "*fonctions autosimilaires de Gabor*" - pour **modéliser le système visuel humain**.

Cependant, on peut prendre comme point de départ l'année **1975** et les travaux de l'ingénieur géophysicien Français *Jean Morlet*, qui travaillait pour Elf-Aquitaine et qui créa les ondelettes pour chercher du pétrole sous terre. **En 1960**, la méthode standard de détection du pétrole sous terre, consiste à envoyer des vibrations dans la terre et à analyser les échos. De cette manière, on tente de déterminer la profondeur des diverses couches de pétrole, leur épaisseur et leur constitution, les échos de hautes fréquences correspondant grossièrement aux couches minces.

Dans ses travaux, Jean Morlet est souvent confronté à de très nombreuses couches : tous les signaux de réflexion **interfèrent entre eux** et il lui est donc **difficile de les séparer**. Jusque là, il utilise l'analyse de Fourier et des ordinateurs de plus en plus puissants. Puis, il utilise des fenêtres de différentes tailles, avec pour objectif d'essayer d'avoir une définition locale plus fine, de manière à accéder à des informations sur des couches d'épaisseurs différentes, **sans succès**... **Vers 1975**, Morlet s'inspire de l'analyse de Fourier à fenêtre proposée par Gabor trente ans plus tôt. Malheureusement, cette représentation demeure **imprécise sur le temps aux hautes fréquences**.

C'est alors que Morlet a une **idée géniale**. Au lieu de garder fixe la taille de la fenêtre et de faire varier le nombre d'oscillations à l'intérieur de cette fenêtre, il fait l'inverse : il garde **constant le nombre d'oscillations** et fait **varier la taille de la fenêtre**, l'*étirant* et la *comprimant* telle un accordéon (Fig. 6.3).



FIG. 6.3 – Trois ondelettes filles de Morlet : dilatée (gauche), échelle de base (milieu) et comprimée (droite).

Étirer la fenêtre a pour effet d'étirer les oscillations, donc de baisser leur "fréquence", la comprimer a pour effet de comprimer les oscillations, donc d'augmenter leur "fréquence". **Les ondelettes étaient nées**.

6.3.2 Qu'est-ce qu'une ondelette ?

Selon *Yves Meyer*, "*c'est le signal transitoire le plus simple que l'on puisse imaginer*". Une ondelette est un objet mathématique relativement simple : il s'agit d'une fonction Ψ, localisée (par exemple, nulle hors d'un certain intervalle) et oscillante (Fig. 6.4). On demande au moins que l'intégrale de Ψ soit nulle, mais on souhaite souvent que $\int t\Psi(t)dt = ... = \int t^{N-1}\Psi(t)dt = 0$, pour un certain entier $N \geq 2$.

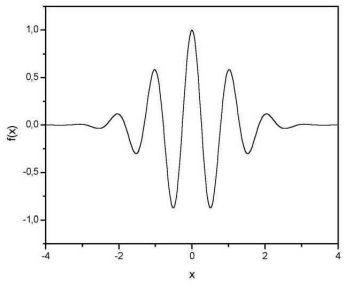


FIG. 6.4 – Ondelette de Morlet.

Il faut savoir qu'**il existe de nombreuses ondelettes**. Certaines ont des formules mathématiques explicites alors que d'autres sont construites à partir de propriétés mathématiques plus complexes.

Chaque ondelette possède des qualités bien spécifiques, comme par exemple :

- La **symétrie** : utile pour éviter le déphasage,
- Le **nombre de moments nuls** : c'est-à-dire le nombre d'oscillations, utile pour la compression,
- La **régularité** : utile pour obtenir des signaux reconstruits lisses et réguliers.

Par défaut, lorsqu'on emploie le terme d'ondelette, il faut entendre **ondelette mère** Ψ car elle va servir de **prototype de base** pour générer toute une famille d'ondelettes $\Psi_{a,b}$ (les *ondelettes filles*), aussi appelées **atomes de base**. Ce sont des **copies dilatées**, **comprimées** et **translatées** dans le temps de l'ondelette mère. D'où la formule suivante d'un atome de base :

$$\Psi_{a,b} = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right), a \in \mathbb{R}^{*+}, b \in \mathbb{R} \tag{6.5}$$

Donnons maintenant quelques explications concernant la formule (6.5) : nous voyons apparaître deux paramètres a et b. Le paramètre $a > 0$ est le **paramètre d'échelle**, c'est lui qui va agir sur la compression ou la dilatation de l'atome de base. Lorsque $0 < a < 1$, l'atome de base est *compressé* et lorsque $a > 1$, l'atome de base est *dilaté*. Le paramètre b est le **paramètre de position**, il permet de faire translater notre atome de base **à gauche** ($b < 0$) ou **à droite** ($b > 0$), le long du signal à analyser . Reste le facteur en $\frac{1}{\sqrt{a}}$ qui permet de **conserver la même énergie de l'atome de base à chaque échelle** a. En fait, il convient de préciser que *l'inverse d'une échelle est une fréquence* (au niveau des ordres de grandeur).

Ainsi plus l'ondelette fille est **dilatée** (a devient donc de plus en plus grand), plus son **amplitude et sa fréquence diminuent**. À l'inverse, plus l'ondelette fille est **compressée** (a devient donc de plus en plus petit), plus son **amplitude et sa fréquence augmentent** (Fig. 6.4).

6.3.3 La transformée en ondelettes : un "microscope mathématique"

Une manière pratique d'introduire la transformée en ondelettes consiste à établir une **analogie avec la musique**. De la même manière que les notes de musique possèdent une certaine hauteur et une durée, **les ondelettes oscillent à une certaine fréquence**

et pendant un temps limité. Prenons l'exemple d'un extrait d'une partition musicale représentant les premières notes de *La Lettre à Élise* de *Ludwig Van Beethoven* (Fig. 6.5).



FIG. 6.5 – Neuf premières notes de "La Lettre à Élise" de Ludwig Van Beethoven.

L'expérience consiste à **comparer la transformée de Fourier et la transformée en ondelettes de l'extrait musical**, afin de mettre en avant quelques remarques préliminaires concernant l'apport de l'analyse par ondelettes. Sur la figure 6.6, les raies de la transformée de Fourier nous indiquent quelles sont les notes qui sont présentes dans l'extrait musical, en nous renseignant précisément sur leurs fréquences ; on ne sait cependant pas l'ordre dans lequel les notes sont jouées. Ces raies verticales se retrouvent à droite dans le *scalogramme*, sous la forme de bandes horizontales (parce que les fréquences sont maintenant en ordonnées). En outre, ces fréquences sont localisées en temps (ce qui est un "plus"). Mais le prix à payer est la perte de localisation fréquentielle. **Ce qu'on gagne d'un côté est perdu de l'autre** (Fig. 6.7). Enfin, il est très intéressant de remarquer que le scalogramme qui résulte de la transformée en ondelettes **tend vers la représentation idéale qu'est la partition musicale**. D'une certaine manière, les ondelettes peuvent être assimilées aux "*notes mathématiques*" d'un morceau de musique.

En observant la figure 6.7, on peut se rendre compte de deux choses : d'une part, en ce qui concerne la transformée de Gabor, la résolution fréquentielle ne diminue pas lorsque la fréquence varie. D'autre part, dans la transformée en ondelettes, on possède une bonne résolution fréquentielle aux basses fréquences (au détriment d'une perte de localisation en temps) et qu'à l'inverse, on possède une bonne résolution temporelle aux hautes fréquences (au détriment d'une perte de localisation en fréquence). Ce *pavage du plan temps-fréquence* permet d'identifier ce que l'on appelle les **boîtes d'Heisenberg**.

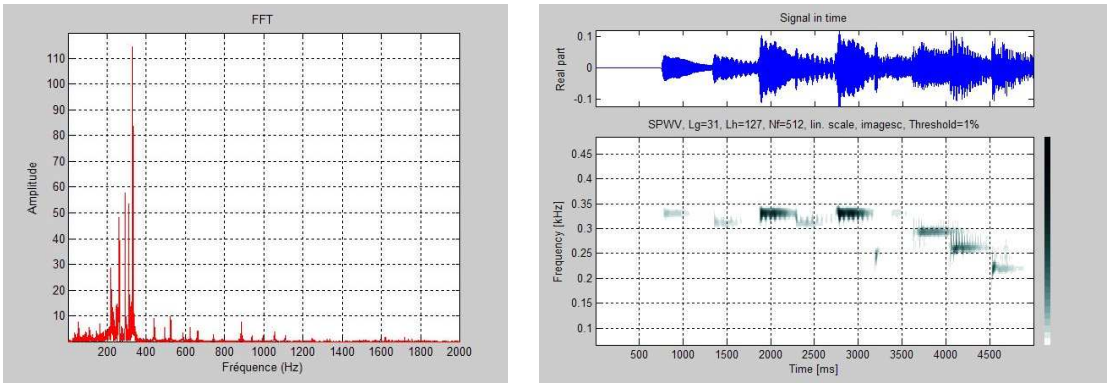
6.3.3.1 Transformée continue 1D

La transformée continue en ondelettes (6.6) est une opération mathématique qui consiste à transformer une fonction originale $f(t)$ qui dépend du temps, en une nouvelle fonction $C_f(a,b)$ qui *dépend à la fois de l'échelle et du temps*.

$$C_f(a,b) = \int_{\mathbb{R}} f(t)\overline{\Psi_{a,b}(t)}dt \tag{6.6}$$

Dans la transformée en ondelettes, on décompose le signal à étudier $f(t)$ selon une **base de fonctions d'ondelettes localisées dans le temps**. On obtient alors des coefficients numériques $C_f(a,b)$: les **coefficients d'ondelettes**. Le calcul de ces coefficients est analogue à celui que l'on effectue lors d'une transformée de Fourier : on multiplie le signal par la fonction analysante puis on calcule l'intégrale du produit. On notera que l'on prend le conjugué de l'ondelette afin de pouvoir opérer une *reconstruction exacte lors de la transformée inverse* (6.7). Cette *synthèse* s'effectue par la formule suivante :

$$f(t) = c\int_0^{+\infty}\int_{-\infty}^{+\infty} C_f(a,b)\Psi_{a,b}(t)db\frac{da}{a^2} \tag{6.7}$$



(6.6.a) (6.6.b)

FIG. 6.6 – À gauche, la transformée de Fourier. En haut à droite, l'onde sonore de l'extrait musical. En bas à droite, la transformée en ondelettes avec en abscisses, le temps, en ordonnées, la fréquence.

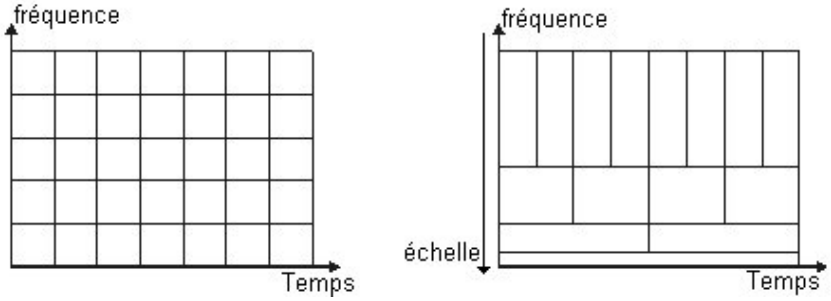


FIG. 6.7 – Comparaison des plans temps-fréquence (transformée de Gabor), à gauche, et temps-échelle (transformée en ondelettes), à droite.

où $c > 0$ est une constante qui ne dépend que de Ψ.

Un coefficient d'ondelette a une amplitude d'autant plus grande que l'ondelette "ressemble" au signal sur la portion analysée. Tous ces coefficients sont représentés dans ce que l'on appelle un **scalogramme** (équivalent du spectrogramme pour la transformée de Fourier), avec le temps en abscisses et la fréquence (ou l'échelle) en ordonnées.

La figure 6.8 illustre l'analyse d'un *chirp linéaire* (signal dont la fréquence augmente linéairement avec le temps), elle permet de comparer la *densité spectrale d'énergie* avec le **scalogramme de Morlet** d'un chirp linéaire dont la fréquence normalisée (*au sens de Shannon*) varie linéairement de 0.1 à 0.4. Le caractère fréquentiel linéaire du chirp n'est bien visible qu'avec la transformée en ondelettes.

Contrairement à la transformée de Fourier, la transformée en ondelettes est une **analyse locale de résolution temporelle variable des signaux**. Il faut comprendre par là qu'elle permet de décrire le comportement local des signaux à différentes échelles de temps. On parle alors d'**analyse multirésolution**. Les ondelettes *s'adaptent* aux différentes composantes du signal. Lorsque la fenêtre est étroite, on peut observer les hautes fréquences et lorsque la fenêtre est plus large, on observe les basses fréquences. La transfor-

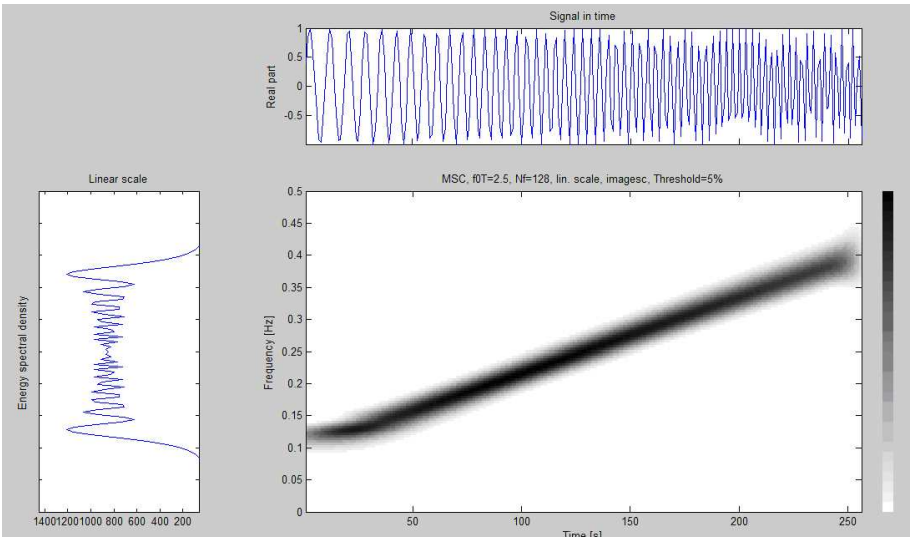


FIG. 6.8 – Scalogramme (à droite) d'un chirp linéaire. L'amplitude des coefficients va du plus faible (couleur claire), au plus fort (couleur foncée).

mée en ondelettes a donc l'aptitude de pouvoir "*zoomer*" sur les zones de non stationnarité d'un signal. *Marie Farge* parle de "*microscope mathématique*".

Néanmoins, les ondelettes de Morlet ne constituent que l'un des outils à l'intérieur d'une **immense boîte à outils** dont fait partie l'analyse de Fourier. En fait, la détection optimale d'un chirp linéaire se fait à l'aide de ce qu'on appelle l'**analyse de Wigner-Ville**[2] (datant des années 40 !). Ceci montre que **l'outil doit être adapté à la classe des signaux que l'on souhaite analyser**.

En traitement audio, une belle réponse est donnée par *Stéphane Molla* et *Bruno Torrésani* [Moll 04]. Les auteurs utilisent *deux bases différentes* : les **bases trigonométriques locales** (pour représenter les parties stationnaires par morceaux, les voyelles par exemple) et **les ondelettes** (pour représenter les *consonnes plosives*).

En traitement de l'image, on peut grossièrement distinguer les objets (délimités par les bords) ayant une structure géométrique simple et les textures (par exemple, un champ de blé est une texture différente d'un champ de seigle). Bien entendu, **les ondelettes de Morlet sont incapables d'analyser les textures** (de distinguer le seigle du blé). Finalement, en ce qui concerne les images, **les ondelettes conviennent aux objets ayant des structures géométriques** mais **pas aux textures**, où l'on préférera utiliser d'une **transformée en paquets d'ondelettes**, qui sera abordée un peu plus loin dans la sous-section suivante.

[2]La transformée de Wigner-Ville fournit une décomposition temps-fréquence sans aucune restriction sur les résolutions temporelles et fréquentielles. Elle est définie par : $s\left(t+\frac{\tau}{2}\right) s\left(t-\frac{\tau}{2}\right) e^{-2i\pi f t} dt$. Cette formule représente l'énergie d'un signal s au temps t et à la fréquence f. Malheureusement, la non-linéarité de cette transformée a des conséquences désastreuses qui se manifestent par l'apparition d'interférences. Une solution consiste à adoucir localement la rigueur de la résolution. On parle alors de *transformée de Wigner-Ville lissée*.

6.3.3.2 Transformée discrète 2D, Analyse Multirésolution (MRA)

Stéphane Mallat propose un **algorithme rapide de décomposition et de reconstruction pour la transformée discrète en ondelettes** à la fin des années **1980** [Mall 89]. Ceci a permis de développer grandement les applications en traitement de signal et de l'image. Grâce à cette transformée, on décompose une fonction au moyen de **filtres à réponse impulsionnelle finie**.

Jean Morlet suggérait de prendre $a = \alpha 2^j$, $b = k\beta 2^j$ où j (la résolution) et k sont des entiers relatifs, les pas d'échantillonnage α et β étant positifs. Cependant, on peut faire beaucoup mieux et *Yves Meyer* a établi que l'on peut construire une fonction Ψ appartenant à la *classe de Schwartz* de sorte que les fonctions $2^{j/2}\Psi(2^j x - k), j, k \in \mathbb{Z}$ constituent une base orthonormée de $\mathcal{L}^2(\mathbb{R})$ composée d'ondelettes. La discrétisation des coefficients se faisant par des puissances de deux, on parle de *transformée dyadique*.

Dans le cas 2D, l'analyse consiste à décomposer une image en **plusieurs sous-bandes de fréquence** (Fig. 6.9).

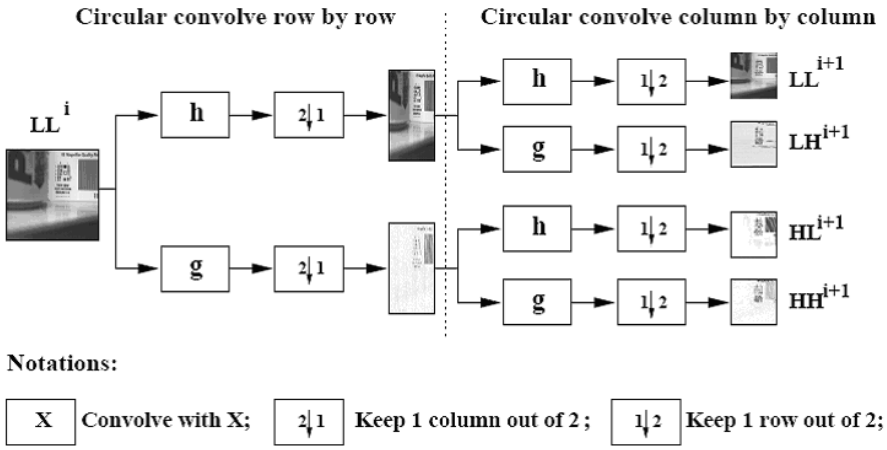


FIG. 6.9 – Transformée 2D discrète en ondelettes d'une image.

On commence par analyser l'image originale en appliquant deux filtres le long des lignes, l'un passe-haut (g) l'autre passe-bas (h), les coefficients de ces filtres étant propres à l'ondelette utilisée. Ensuite, on ne considère qu'une colonne sur deux auxquelles on applique de la même façon un passe-haut, et un passe-bas, et au final on ne considère qu'une ligne sur deux. On obtient alors **quatre images de tailles réduites au premier niveau de décomposition** :

- Celle obtenue après filtrage par deux passe-bas (LL_1) constitue une **approximation de l'image de départ à une résolution inférieure** (échelle supérieure),
- Après passage par un filtre passe-bas et un filtre passe-haut (LH_1), on obtient une image qui met en avant les **détails horizontaux** de l'image originale,
- Après passage par un filtre passe-haut puis un filtre passe-bas (HL_1), on obtient une image qui met en avant les **détails verticaux** de l'image originale,
- Enfin, après passage par deux filtres passe-haut, l'image (HH_1) met en avant les **détails diagonaux** de l'image originale.

On applique de nouveau cette décomposition à l'approximation de l'image au premier niveau de décomposition, etc. (Fig. 6.10).

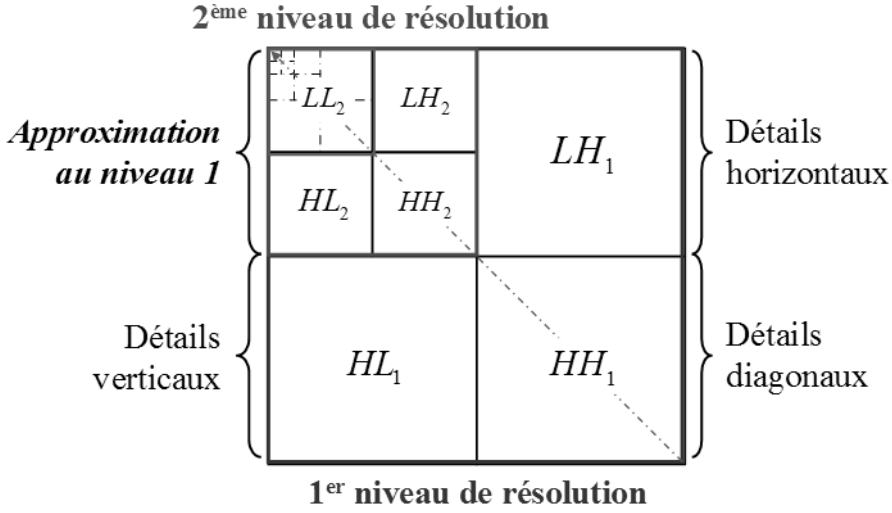


FIG. 6.10 – Analyse multirésolution.

Ce principe est souvent représenté par une **pyramide** avec différents étages. Le **sommet** de cette pyramide représente la **résolution la plus grossière** de l'image originale (pixel moyen de l'image entière). On associe à chaque étage inférieur une résolution supérieure. Si l'étage inférieur correspond à la résolution $j+1$ alors la résolution de l'étage supérieur est j. On parle alors d'**analyse multirésolution** [Mall 00].

6.3.3.3 Transformée en paquets d'ondelettes 2D

Des signaux compliqués, comme des enregistrements de la parole, incluent des structures temps-fréquence très différentes. Il est donc **nécessaire de trouver des bases dont les propriétés temps-fréquence peuvent être adaptées au signal**. Les bases d'ondelettes sont une famille particulière de bases qui représentent efficacement les signaux réguliers par morceaux (Fig. 6.11). D'autres bases sont construites pour **mieux représenter des signaux ayant des propriétés différentes**, notamment des signaux très oscillants.

Les **bases de paquets d'ondelettes orthonormées** utilisent des *filtres miroirs conjugués* pour diviser l'axe fréquentiel en intervalles de tailles diverses. Un signal discret de taille N se décompose sur plus de $2^{N/2}$ bases de paquets d'ondelettes par un algorithme rapide de banc de filtres qui ne nécessite que $O(Nlog_2(N))$ opérations.

Si les propriétés du signal varient dans le temps, il est préférable d'**isoler différents intervalles de temps par des fenêtres translatées**. Les bases de cosinus locaux [Ausc 92] sont construites en multipliant ces fenêtres par des fonctions cosinus. Les bases de paquets d'ondelettes et de cosinus locaux sont des **familles de bases duales**. **Les paquets d'ondelettes découpent l'axe fréquentiel et sont uniformément translatés en temps** alors que les bases de cosinus locaux divisent l'axe temporel et sont uniformément translatées en fréquence.

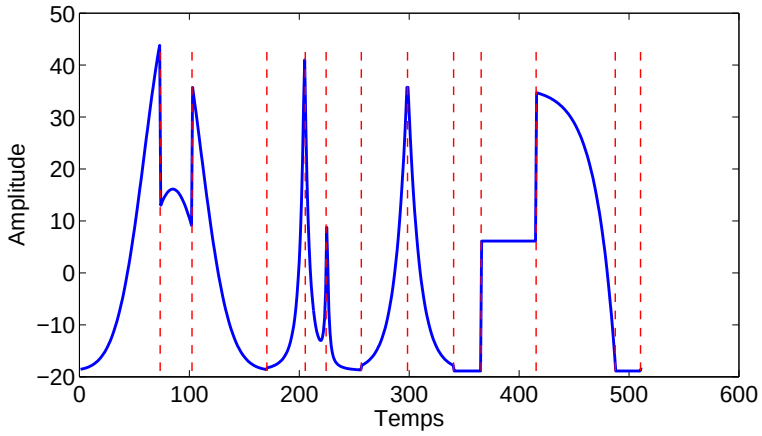


FIG. 6.11 – Exemple d'un signal régulier par morceaux (512 points).

Les paquets d'ondelettes ont été définis par *Coifman, Meyer et Wickerhauser* [Coif 92] en généralisant la relation entre les approximations multirésolutions et les ondelettes. Un espace $\mathbf{V}_j$ d'une approximation multirésolution se décompose en un espace de résolution plus basse $\mathbf{V}_{j+1}$ et un espace de détails $\mathbf{W}_{j+1}$. Cela se fait en divisant la base orthogonale $\{\phi_j(t-2^j n)\}_{n\in\mathbb{Z}}$ de $\mathbf{V}_j$ en deux nouvelles bases orthogonales :

$$\{\phi_{j+1}(t-2^{j+1}n)\}_{n\in\mathbb{Z}} \text{ de } \mathbf{V}_{j+1} \quad \text{et} \quad \{\psi_{j+1}(t-2^{j+1}n)\}_{n\in\mathbb{Z}} \text{ de } \mathbf{W}_{j+1}$$

6.4 Applications

6.4.1 Les ondelettes : un outil polyvalent

Les ondelettes sont employées dans un **grand nombre de domaines**. On peut citer, entre autres, la géophysique, l'astrophysique, l'imagerie dans tous ses aspects et l'imagerie médicale en particulier, le codage de signaux vidéo, la modélisation du trafic dans les réseaux de communication comme Internet, l'analyse de la turbulence atmosphérique ou celle de souffleries [Abry 97], l'optoélectronique [Dura 06] et la liste est encore longue...

La variété des secteurs concernés et le nombre de problèmes traités sont surprenants [Misi 03]. La question que l'on peut tout de suite se poser est la suivante : "*Comment se fait-il qu'un même outil fonctionne pour tant d'applications* ?" Puisque qu'il est impossible de trouver une unique réponse qui soit réellement satisfaisante, nous allons donner quelques raisons qui paraissent plausibles :

– Tout d'abord, **la méthode des ondelettes est relativement récente en traitement du signal**. Elle apporte des innovations techniques et permet la construction d'un *nouveau dictionnaire de formes*, reliant des caractéristiques du signal à celles des transformées en ondelettes, permettant d'interpréter des propriétés des signaux à partir des structures des coefficients. On sait par exemple, que les coefficients nuls dans une décomposition discrète indiquent que le signal est lisse et que rien ne change,

– Ensuite, **la technique des ondelettes constitue en outre un outil d'analyse locale**, en quoi elle diffère fortement d'une analyse de Fourier. L'ondelette sait porter son regard localement en inspectant les voisinages d'un point. Les informations codées dans les coefficients sont alors entièrement déterminées par les valeurs du signal situées dans le champ de l'ondelette. Ce travail local est complété par un travail de synthèse qui réalise la comparaison des résultats obtenus en différents points. **On globalise en quelque sorte les analyses locales**,

– De plus, **les ondelettes analysent le signal échelle par échell**e. Elles servent de véritable "microscope mathématique" en regardant le signal à différentes résolutions. Chaque point est ausculté, très finement à petite échelle dans une petite zone et plus grossièrement à grande échelle,

– Enfin, **les ondelettes permettent de caractériser certains espaces de signaux un peu ignorés jusque là parce que difficiles à manipuler**. Ces espaces contiennent des fonctions qui peuvent présenter des parties moins régulières que les fonctions dérivables, séparées par des discontinuités. Ces signaux sont difficiles à aborder par les autres outils actuellement disponibles.

Dans ce qui suit, nous allons enfin voir comment **intégrer les ondelettes dans plusieurs modules appartenant à la chaîne biométrique multimodale**. Nous montrerons également que l'utilisation des ondelettes peut répondre de manière **très efficace** à trois problèmes majeurs en traitement du signal et de l'image, à savoir *la compression*, *l'analyse* et *le débruitage*.

6.4.2 Compression

Partant du signal à comprimer, on le décompose d'abord sur une base orthogonale d'ondelettes au moyen de la transformée discrète. Ensuite, on sélectionne une partie des coefficients par seuillage dur, en conservant intact les coefficients d'approximation de niveau convenablement choisi. Enfin, à partir des coefficients seuillés, on reconstruit le signal original en leur appliquant la transformée discrète inverse. Le signal ainsi obtenu est le signal comprimé.

Le standard de compression d'image bien connu qu'est le JPEG est basé sur une transformée de Fourier discrète. Le JPEG analyse l'image par bloc de 8×8 pixels ce qui conduit aux fameux effets de mosaïques que nous avons tous vu au moins une fois, et qui altère considérablement la qualité de l'image compressée. Ce défaut est d'autant plus visible que le taux de compression est élevé. Le nouveau standard à venir en compression d'image est le **JPEG 2000** [JPEG] (*Yves Meyer*) et qui est basé sur la transformée en ondelettes. Cette fois-ci, l'image est analysée de manière globale et la qualité reste très bonne même à un taux de compression élevé. A titre d'exemple, si l'on considère une image originale de 512×512 pixels (codée en 24 bits RGB) et d'une taille de 786 Ko, il est très difficile de voir la différence avec la même image compressée par ondelettes d'un ratio de $75 : 1$ et dont la taille atteint seulement 10.6 Ko (Fig. 6.12). Grâce au *JPEG 2000*, il est non seulement possible de **prévoir la taille du fichier compressé** mais aussi de permettre un **affichage progressif** de l'image. On commence par les détails grossiers de l'image puis les hautes fréquences sont progressivement ajoutées, ce qui a pour effet d'ajouter les détails de plus en plus fin, jusqu'à l'image finale.



(6.12.a)



(6.12.b)



(6.12.c)



(6.12.d)

FIG. 6.12 – Image originale de 786 Ko (6.12.a). Image compressée avec un ratio de $75 : 1$ pour $10,6$ Ko (6.12.b), un ratio de $150 : 1$ pour $5,3$ Ko (6.12.c) et un ratio de $300 : 1$ pour $2,6$ Ko (6.12.d).

Nous allons maintenant montrer **comment utiliser la compression par ondelettes dans le cadre de la biométrie**. L'utilisation de l'analyse en composantes principales (PCA) peut demander beaucoup de ressources CPU et de temps de calcul. Ceci est principalement dû à la phase d'extraction des vecteurs propres d'une grande matrice de covariance. Bien que l'on puisse contourner en partie ce problème grâce à l'astuce mathématique décrite dans l'article original de *M.A. Turk et A.P. Pentland* [Turk 91], nous sommes contraints, à la base, de travailler en fonction de la résolution des images.

Un article très intéressant [Sinh 06], montre dix-neuf résultats relatifs à la reconnaissance faciale par les humains. Le premier résultat qui y est exposé nous concerne tout particulièrement ici :

Résultat 1 : *Les humains peuvent reconnaître des visages familiers à très basse résolution.*

Faisons alors une analogie avec un système biométrique : les "visages familiers" seraient représentés par les personnes qui ont déjà été enrôlées par le système (donc connues de ce dernier). Cela signifie qu'un système de reconnaissance faciale bien conçu (par exemple, imitant le fonctionnement du cerveau) devrait être capable de reconnaître les personnes enregistrées, même à très faible résolution.

Cependant, avec l'amélioration constante de la résolution que peut fournir une caméra ou un appareil photo, il est très tentant d'utiliser toute cette quantité d'information potentielle et de détails supplémentaires pour la reconnaissance faciale. Il faut bien comprendre qu'autant **la reconnaissance par l'iris requiert une grande résolution d'image** pour pouvoir analyser correctement la texture de l'iris, autant **la reconnaissance faciale ne demande pas nécessairement une forte résolution des images de visage** ; en fait la montée en résolution permet d'apporter une discrimination entre les individus basée sur des différences très subtiles dans leurs caractéristiques faciales. Par ailleurs, ces schémas de reconnaissance basés sur des images très détaillées ne sont pas toujours disponibles ; ce qui est particulièrement vrai dans les situations où des individus doivent être reconnus à distance.

Ainsi, dans le but de construire des **systèmes robustes à la dégradations des images**, on peut prendre comme inspiration le système visuel humain. L'identification à distance des visages est une tâche à laquelle nous sommes confrontés tous les jours où nous devons extraire de l'information pertinente à partir d'images à faible résolution. On peut se demander "*Comment la performance d'identification des visages change en fonction de la résolution des images* ?"

Les premiers travaux en reconnaissance faciale à basse résolution ont été effectués par *L.D. Harmon et B. Julesz* [Harm 73a] [Harm 73b]. En travaillant à partir d'images moyennes par bloc de visages familiers, ils ont obtenu des taux de reconnaissance élevés même avec images contenant des blocs de 16×16. Plus récemment, *Yip et Sinha* [Yip 02] se sont aperçus que des sujets pouvaient reconnaître plus de la moitié d'un ensemble d'images de visages familiers qui ont été rendues floues de manière à obtenir une résolution d'image équivalente d'environ 7×10 pixels ; la performance de reconnaissance ayant atteint son niveau le plus haut à une résolution de 19×27 pixels.

On peut alors effectuer un test simple pour se rendre encore mieux compte de l'impact de la résolution des images de visage sur le taux de reconnaissance faciale. On considère tous les ensembles d'images de la base de données *FERET* (FB, FC, DUP1 et DUP2) auxquels nous appliquons l'algorithme PCA associé à la mesure de similarité *MahCosine* (*cf. Annexe A.2.4*). On trace alors les différentes courbes CMC (Fig. 6.13) à partir des différentes matrices de similarité. D'un côté, les images normalisées originales de taille 140×120 (traits pointillés cerclés - "Original"), de l'autre, les images normalisées compressées deux fois par ondelettes bi-orthogonales "bior1.3" (traits pleins étoilés - "WComp") et dont la taille atteint 38×33 ! Il est assez remarquable de constater que **le taux de reconnaissance faciale est non seulement très stable mais que dans trois cas sur quatre, il est sensiblement meilleur lorsque les images ont été compressées par ondelettes**.

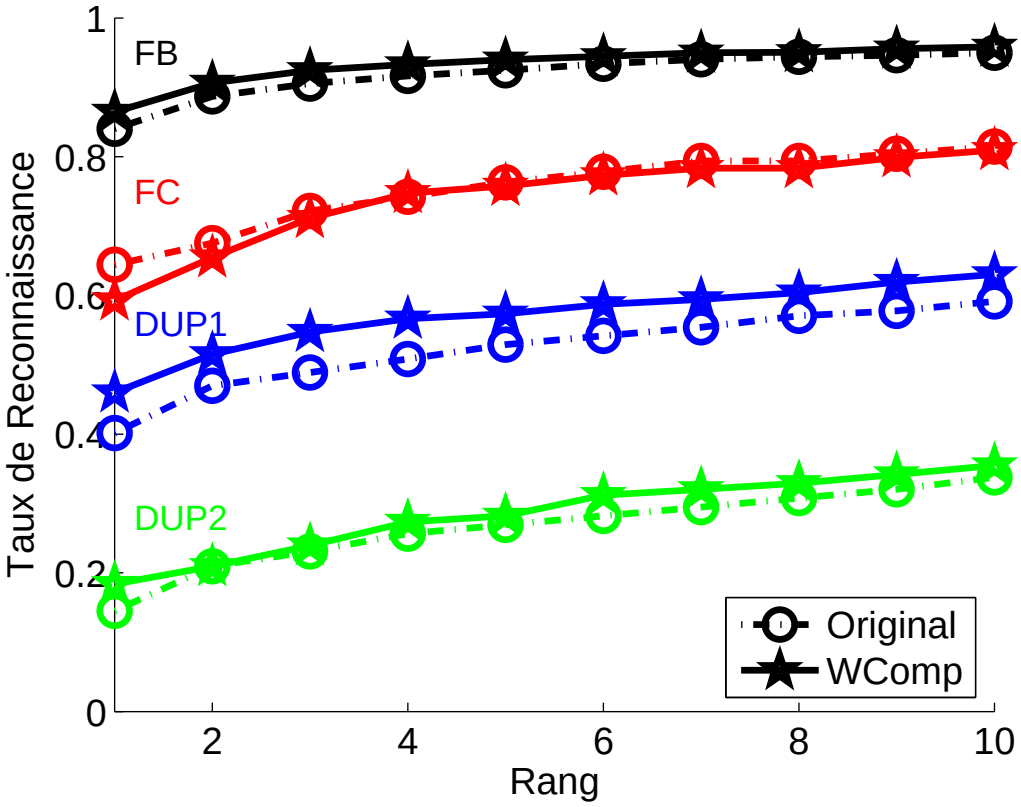


FIG. 6.13 – Influence de la compression par ondelettes sur le taux de reconnaissance faciale (tests FERET).

Ainsi, pour en revenir au PCA, la double compression par ondelettes des images permet d'obtenir une **matrice de covariance beaucoup plus petite** et donc de pouvoir extraire les vecteurs et les valeurs propres **beaucoup plus facilement**, conduisant à un **temps de reconnaissance beaucoup plus rapide et très stable, voire amélioré**. A titre d'information, le système réel avec la caméra de sécurité et l'association PCA-MahCosine que nous avons développé permet de vérifier une personne enregistrée parmi la base de données *FERET* (≈ 1200 **personnes**) en **un peu moins de 200ms** !

Tandis que la remarquable tolérance du système visuel humain à la réduction de résolution est indiscutable, il faut bien avouer que nous n'avons toujours *pas d'idée précise concernant la réalisation de ce processus*. Enfin, ce résultat démontre que **les détails caractéristiques fins ne sont pas nécessaires pour obtenir de bons taux de reconnaissance faciale**.

6.4.3 Analyse

L'aspect analyse par ondelettes intervient à la fois dans la **modalité de l'iris et dans celle du visage**. La conclusion de 6.3.3.1 nous dit que **si nous voulons analyser des textures**, il vaut mieux utiliser une **transformée en paquets d'ondelettes** qui est plus adaptée, plutôt qu'une transformée classique en ondelettes, plutôt réservée à l'extraction de contours. Ainsi, comme nous l'avons vu dans le chapitre précédent concernant la construction de la matrice de similarité des scores d'iris, **une signature biométrique de l'iris est extraite en appliquant une transformée en paquets d'ondelettes sur trois niveaux à partir de l'image déroulée de l'iris**. On quantifie l'énergie de chaque paquet d'ondelettes et on "code" ceux qui sont les plus discriminants (les plus informatifs). Dans notre algorithme, ce sont les paquets 2 et 10 [Rydg 04] qui sont utilisés pour générer la signature de l'iris. Sans l'apport des ondelettes, il serait probablement plus difficile d'extraire une description concise et efficace de la texture de l'iris et ce n'est pas un hasard si *J. Daugman*, pionnier de la reconnaissance de l'iris, utilise des ondelettes en codant leurs phases pour générer un "code-barre" de l'iris.

De plus, dans le chapitre sur les neurosciences, nous avons vu que les résultats des expériences de *Jones et Palmer* [Jone 87] montrent très clairement que les réponses des cellules "simples" du cortex visuel primaire peuvent être modélisées avec des **ondelettes 2D de Gabor**. Cette étonnante découverte à été mise à profit dans l'algorithme LG-PCA que nous utilisons pour la partie reconnaissance faciale. En effet, ce dernier utilise un **banc d'ondelettes 2D de Gabor paramétrées** afin d'extraire des caractéristiques du visage, avant d'effectuer une réduction de dimension fournie par l'action de l'analyse en composantes principales.

Par ailleurs, **l'analyse par ondelettes permet également de fusionner des images de manière très intéressante**. De nombreuses techniques de fusion d'images en utilisant une transformée en ondelettes ont été proposées [Li 95] [Zhan 99] [Niko 01] [Hill 02] [Sant 03] [Chao 07], on trouvera d'ailleurs un récent état de l'art dans [Gosh 07]. Le principal avantage de fusionner des images dans le domaine des ondelettes est que l'on peut **traiter divers domaines de fréquence de manière différente**.

Considérons deux images différentes $I_1(x, y)$ et $I_2(x, y)$ qui doivent être fusionnées en une seule (Fig. 6.14). Tout d'abord, une transformée discrète 2D en ondelettes (2D-DWT, notée W pour simplifier l'écriture), sur L niveaux (cf. Chapitre 6.3.3.2), est effectuée sur les deux images d'entrée. Les images transformées sont ensuite **combinées dans le domaine des ondelettes** en utilisant une règle de fusion ϕ, puis une transformée discrète 2D inverse en ondelettes (2D-IDWT, notée W^{-1} pour simplifier l'écriture) est appliquée aux images combinées afin de retourner dans le domaine spatial, pour donner l'image finale fusionnée $I(x, y)$ (6.8) (Fig. 6.15).

$$I(x, y) = W^{-1}\{\phi[W(I_1(x, y)), W(I_2(x, y))]\} \tag{6.8}$$

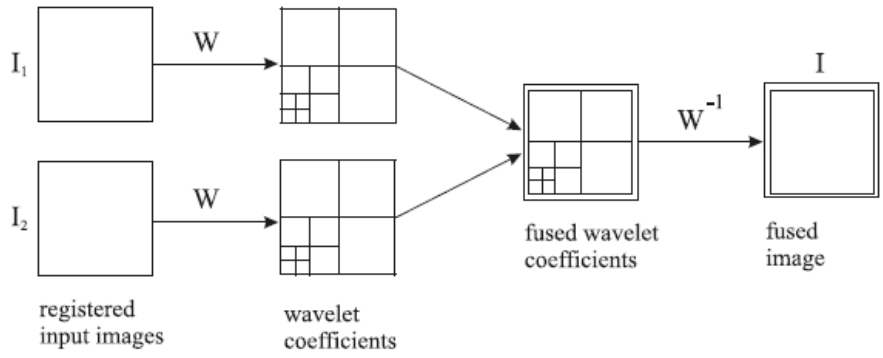


FIG. 6.14 – Principe de fusion d'images par ondelettes (WIF).



FIG. 6.15 – Focus sur le premier plan (à gauche). Focus sur l'arrière-plan (au milieu). Images fusionnées par ondelettes (à droite), le premier plan et l'arrière-plan sont nets simultanément.

En plus de l'analyse MRA (cf. Chapitre 6.3.3.2) fournie par la 2D-DWT, un point fondamental dans les techniques de fusion d'images concerne la façon dont on doit construire la représentation MRA fusionnée à partir des représentations MRA des images sources. Le processus qui permet d'accomplir cet objectif est ce que l'on appelle une règle de fusion. Quelques solutions générales pour établir une règle de fusion sont illustrées sur la figure 6.16.

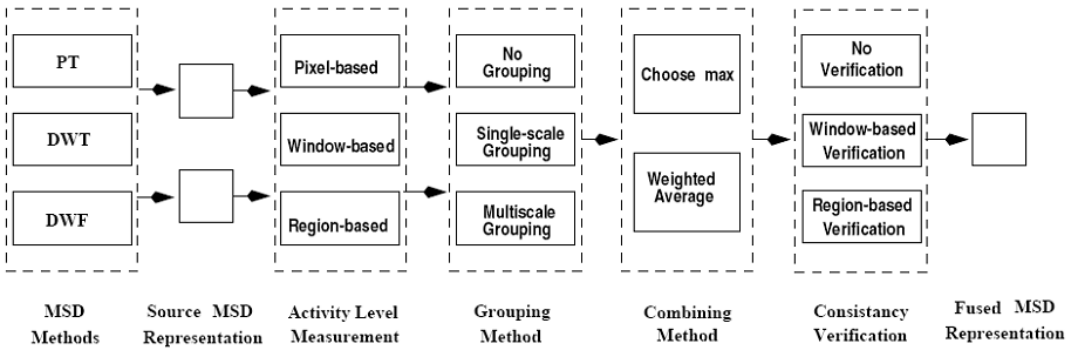


FIG. 6.16 – Cadre de travail générique de schémas de fusion d'images.

Ces schémas de fusion incluent **plusieurs choix** :

- **Mesure du niveau d'activité** : le niveau d'activité de coefficients issus d'une analyse MRA représente l'information (énergie) locale d'une image. Cette mesure peut être effectuée au niveau pixel, fenêtre ou région (Fig. 6.16),

- **Méthode de groupement de coefficients** : la façon dont on associe les coefficients MRA entre eux. Nous pouvons, par exemple, décider de ne pas les grouper ou bien de les grouper sur une seule ou plusieurs échelles d'analyse (Fig. 6.16),

- **Méthode de combinaison de coefficients** : la façon de produire une représentation MRA composite à partir des représentations MRA sources,

- **Méthode de vérification de cohérence** : la vérification de cohérence tente d'exploiter l'idée selon laquelle il est très probable qu'une bonne méthode de fusion génère des coefficients voisins dans la représentation MRA composite de la même manière. Par exemple, pour la méthode de combinaison par Choix Maximum ("*Choose Max*", **CM**), la vérification de cohérence est particulièrement simple ; elle permet de s'assurer qu'un coefficient MRA composite ne provient pas d'une image source différente ou de l'ensemble des coefficients voisins.

Au moment de construire des décisions de fusion, une méthode courante est de sélectionner le coefficient MRA possédant le plus grand niveau d'activité. Ce choix se base sur l'hypothèse selon laquelle un niveau d'activité plus grand implique plus d'informations, ce qui doit être manipulé avec soin car cela n'est pas toujours vrai.

Afin d'expliquer les différentes alternatives disponibles pour former une règle de fusion, nous nous plaçons dans le cas où il y a juste deux images sources (X et Y) et une image fusionnée Z. On peut noter que toutes les méthodes qui sont décrites peuvent très bien être étendues à des cas comportant plus de deux images sources ; ce qui, dans notre cas, correspondrait à l'utilisation de plus de deux modalités biométriques.

En général, pour une image I, on note D_I la représentation MRA et A_I le niveau d'activité. Ainsi, dans notre cas, nous avons D_X, D_Y, D_Z, A_X et A_Y. Soit $\vec{p} = (m, n, k, l)$ l'indice correspondant à un coefficient MRA particulier, où (m, n) est la position spatiale du coefficient dans une bande de fréquence donnée, k le niveau de décomposition et l la bande de fréquence de la représentation MRA. Par conséquent, $D_I(\vec{p})$ et $A_I(\vec{p})$ sont respectivement la valeur de la représentation MRA et le niveau d'activité du coefficient correspondant.

Les mesures d'activité basées sur les coefficients ("*Coefficient-Based Activity*", **CBA**) considèrent chaque coefficient séparément. Le niveau d'activité est décrit comme la valeur absolue du coefficient correspondant dans la représentation MRA.

$$A_I(\vec{p}) = |D_I(\vec{p})| \,, \tag{6.9}$$

Nous allons maintenant présenter quelques schémas de sélection dans le domaine des ondelettes :

– **Le schéma de Sélection du Minimum (MinS)** : ce simple schéma prend les coefficients d'ondelette dans chaque sous-bande possédant **la plus petite amplitude** :

$$D_Z(\vec{p}) = \min(|D_X(\vec{p})|, |D_Y(\vec{p})|) \ , \tag{6.10}$$

– **Le schéma de Sélection du Maximum (MaxS)** : ce simple schéma prend les coefficients d'ondelette dans chaque sous-bande possédant **la plus grande amplitude** :

$$D_Z(\vec{p}) = \max(|D_X(\vec{p})|, |D_Y(\vec{p})|) \ , \tag{6.11}$$

– **Le schéma de Sélection de la Moyenne (AveS)** : ce simple schéma prend **la moyenne de chaque paire de coefficients d'ondelette** dans chaque sous-bande :

$$D_Z(\vec{p}) = \frac{D_X(\vec{p}) + D_Y(\vec{p})}{2} \ , \tag{6.12}$$

– **Le schéma de Sélection de la Moyenne Pondérée (WaS)** : ce schéma, développé par *Burt et Kolczynski* [Burt 93], utilise une corrélation normalisée entre les sous-bandes de deux images sur une petite région locale. Le coefficient résultant pour la reconstruction est calculé à partir de cette mesure par une moyenne pondérée des coefficients des deux images,

– **Le schéma de Vérification par Fenêtre (WBV)** : ce schéma, développé par *Li et al.* [Li 95] crée une carte de décision binaire afin de choisir entre chaque paire de coefficients en utilisant un filtre majoritaire.

Nous montrerons dans le Chapitre 8 quelques résultats concernant une application de fusion multimodale en s'appuyant sur cette théorie de fusion d'images par ondelettes.

6.4.4 Débruitage

Les ondelettes fournissent des **moyens simples et efficaces d'estimer de façon non linéaire des signaux, en présence de bruit**. La procédure est simple : on transforme le signal en ondelettes ; on élimine, à toutes les résolutions, les coefficients supérieurs à une valeur seuil et on reconstruit le signal avec les coefficients qui restent grâce à une transformée inverse en ondelettes. La chose qui est surprenante est que **cette méthode exige bien peu de connaissances sur le signal**. Traditionnellement, il faut connaître assez précisément le signal pour pouvoir l'extirper du bruit. En particulier, avec les méthodes traditionnelles, on était censé deviner le type de régularité du signal.

Avec les ondelettes, ceci n'est plus nécessaire : cette performance vient du fait que la transformation en ondelettes orthogonales[3] **comprime l'énergie du signal en un nombre**

[3]Si l'on doit calculer, en temps réel, la transformée de Fourier d'un signal de longueur N ou d'une image

assez restreint de gros coefficients, comme si elle rangeait le signal dans quelques compartiments. Précisons enfin que le débruitage par ondelettes est très performant pour des signaux constitués à la fois de parties lisses et de zones présentant des irrégularités et des variations brusques. **Les méthodes classiques sont incapables d'une telle adaptation en temps** (Fig. 6.17).



(6.17.a)



(6.17.b)



(6.17.c)



(6.17.d)

FIG. 6.17 – Illustration de débruitage d'une image (Lena) par ondelettes. Image originale (6.17.a). Image bruitée (6.17.b), SNR=16.46 dB. Image débruitée avec Fourier (6.17.c), SNR=18.81 dB. Image débruitée par ondelettes (6.17.d), SNR=23.04 dB.

composée de 5 megapixels, on doit utiliser un algorithme rapide. Sinon, le calcul de la transformée de Fourier de l'image demanderait 25000 milliards d'opérations ! L'algorithme rapide mais exact (découvert par Cooley et Tuckey en 1965) s'appelle la FFT (Fast Fourier Transform). En ce qui concerne la transformée en ondelettes, l'algorithme rapide s'appelle la FWT (Fast Wavelet Transform) et repose sur l'existence de bases orthonormées d'ondelettes.

Donoho et Johnstone ont proposé une méthode [Dono 94] pour reconstruire une fonction inconnue f sur $[0, 1]$ à partir de données bruitées :

$$d_i = f(t_i) + \sigma z_i, \quad i = 0, ..., n-1 \ , \qquad (6.13)$$
$$t_i = i/n, \quad z_i \overset{iid}{\sim} N(0,1) \ ,$$

où $N(0, 1)$ est du bruit blanc gaussien, σ le niveau de bruit.

La reconstruction $\hat{f}_n^*$ est définie dans le domaine des ondelettes en translatant tous les coefficients d'ondelette empiriques (C_w) de d d'une quantité $T = \sigma\sqrt{2log(n)}$, appelé **seuil universel de Donoho**, où σ et n sont respectivement le niveau de bruit et la longueur totale des échantillons de bruit. Ce résultat peut être expliqué de manière synthétique à l'aide d'une simple illustration (Fig. 6.18). En fait, le bruit blanc gaussien est bien **encadré par l'intervalle** $[-T, +T]$. On trouvera une *démonstration théorique* du seuil universel de *Donoho*, utilisant le *Lemme de Borel-Cantelli* en *Annexe C.2*.

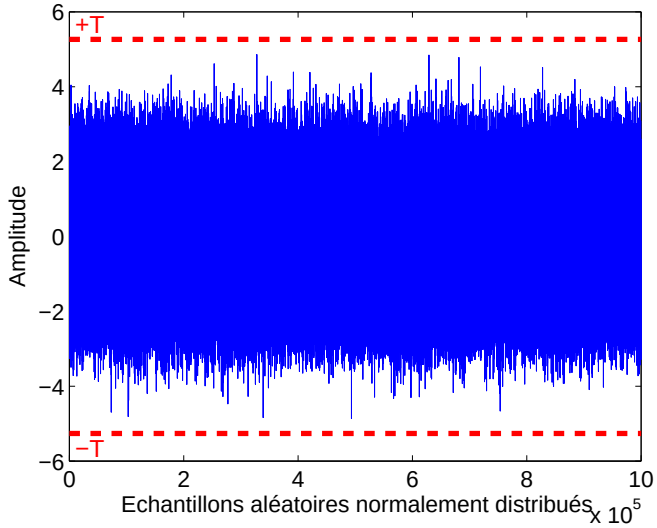
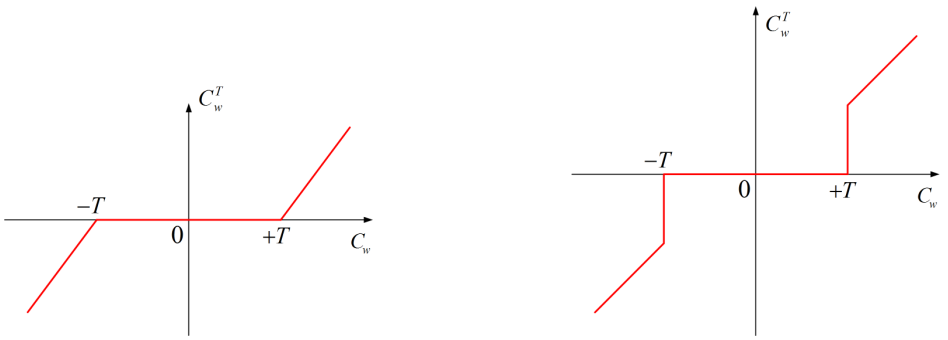


FIG. 6.18 – Bruit blanc gaussien (10^6 échantillons) et seuil universel de Donoho.

En pratique, nous avons besoin d'estimer correctement le niveau de bruit σ. Une façon courante de procéder est l'utilisation de l'**écart absolu médian** (MAD) [Sach 84]. Cependant, lorsque la distribution de données n'est pas parfaitement gaussienne, la médiane et le MAD ne sont pas de bons estimateurs de position et d'échelle [Jain 05] en termes de biais et de variance. Ainsi, de meilleurs estimateurs peuvent être obtenus en utilisant l'écart-type usuel. Enfin, les coefficients d'ondelettes seuillés (C_w^T) peuvent être obtenus soit par la méthode du **seuillage "doux"** ("*soft-thresholding*") (Fig. 6.19.a), soit par la méthode du **seuillage "dur"** ("*hard-thresholding*") (Fig. 6.19.b) [Merr 05].

Nous pouvons alors utiliser la théorie de débruitage par ondelettes pendant l'étape de fusion de système biométrique multimodal. En effet, nous avons vu dans le chapitre précédent que la distribution des scores imposteurs (pour les deux modalités) était très proche d'une distribution gaussienne et que nous pouvions donc, sous réserve d'une normalisation de score adéquate, assimiler ces scores à du bruit blanc gaussien. Dans le but d'améliorer la performance globale de notre système, une solution consisterait donc à pouvoir **débruiter par ondelettes nos vecteurs de scores** de manière à baisser l'amplitude des scores imposteurs tout en conservant au maximum l'amplitude des scores



(6.19.a) (6.19.b)

FIG. 6.19 – Débruitage par ondelettes. Seuillage "doux" (6.19.a) et seuillage dur (6.19.b).

authentiques. Cette technique sera dévoilée plus en détails dans le chapitre suivant, lors de l'explication de la méthode de fusion originale qui a été inventée.

6.5 Conclusion

Nous avons introduit un outil de traitement de signal relativement récent que sont les ondelettes puis nous avons montré les solutions qu'elles pouvaient apporter concernant notre système biométrique multimodal. Les ondelettes, grâce à leur **compression efficace** des images de visages, permettent d'**augmenter la vitesse de calcul** dans le cas d'un système de reconnaissance faciale utilisant le PCA classique tout en fournissant une **très grande stabilité** (voire une **amélioration**) **du taux de reconnaissance**. Par ailleurs, la **transformée en paquets d'ondelettes** fournie une **technique d'analyse de texture discriminante** conduisant à la construction de signatures biométriques de l'iris. Enfin, les ondelettes offrent une **approche élégante à la résolution du problème de débruitage de données**, en ne nécessitant que de **très peu de connaissances sur le signal** à débruiter et en procédant de **manière dynamique**.

Chapitre 7

Fusion Adaptative Combinant Débruitage par Ondelettes et Optimisation de Moments Statistiques

7.1 Introduction

Le but fondamental de tout système biométrique opérant au niveau score, est de pouvoir séparer au maximum les distributions de score des imposteurs et des authentiques (Fig. 7.1). En minimisant la zone de recouvrement entre ces deux distributions, on améliore la performance globale du système en augmentant le taux de reconnaissance.

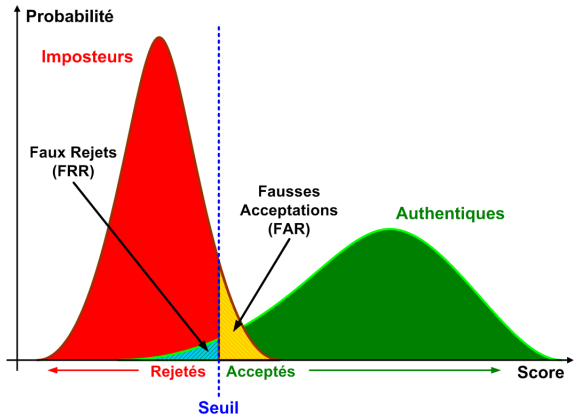


FIG. 7.1 – Illustration du FRR et du FAR.

De nombreuses règles de fusion ont été proposées au niveau score en ce qui concerne les approches par combinaison (cf. Chapitre 2.9). Les premières règles de simple combinaison de scores (min, max, somme, produit) [Ross 06] ne sont pas assez robustes pour faire face à la grande diversité de cas que l'on peut rencontrer dans un système multimodal.

Afin de pouvoir s'adapter à l'utilisateur et de prendre en compte les performances individuelles de plusieurs modalités biométriques, des approches adaptatives, plus prometteuses, ont vu le jour. Ces méthodes utilisent par exemple des seuils ou des règles de fusion spécifiques à chaque utilisateur [Jain 02] [Indo 03], ou encore une somme pondérée de scores issus de matchers biométriques individuels [Ross 03] [Wang 03].

Dans ce qui suit, nous allons ici présenter une nouvelle technique de fusion adaptative appelée "***W**avelet **S**core **D**enoising **S**tatistical **F**usion*" (**WSDSF**) qui combine du débruitage de score par ondelettes et des moments statistiques standards.

7.2 Explication de la méthode WSDSF

Rappelons tout d'abord que si l'on considère une matrice de similarité de taille $N \times N$, chaque ligne (donc chaque utilisateur) est un vecteur de taille N composé de $N-1$ scores imposteurs et d'un seul score authentique.

Notre méthode effectue tout d'abord un débruitage par ondelettes des vecteurs de score relatifs à chaque utilisateur. Ensuite, elle trouve, pour chaque utilisateur, la combinaison de poids idéale (grâce à une première somme pondérée) pour les différentes modalités qui maximise la séparation entre les distributions de scores imposteurs et authentiques ; elle s'appuie sur l'analyse de la forme des distributions de scores en optimisant une fonction de coût à trois paramètres : une distance de séparation ainsi que les coefficients de dissymétrie ("*skewness*") et d'aplatissement ("*kurtosis*") [Pres 92] des distributions des scores imposteurs. Une fois que la combinaison de poids idéale est trouvée, la somme pondérée finale correspondante est appliquée aux scores de chaque utilisateur.

7.2.1 Débruitage de Score par Ondelettes

L'obtention d'un score de similarité peut être vu comme la réalisation d'un événement d'une variable aléatoire. De plus, nous avons montré dans le chapitre précédent que les distributions des scores imposteurs pour les modalités du visage et de l'iris peuvent être assimilés à du bruit blanc gaussien, sous réserve de l'application d'une normalisation Z-Score.

L'idée qui se cache derrière l'utilisation de la théorie du débruitage par ondelettes sur les scores de similarité est de pouvoir effectuer un pré-traitement sur les vecteurs de score de chaque utilisateur, en diminuant l'amplitude des scores imposteurs tout en conservant au maximum l'amplitude des scores authentiques.

Deux conclusions intéressantes sont issues de [Gill 06] : la première est que le seuil universel de Donoho que nous avons introduit dans le chapitre précédent donne, en général, de meilleurs résultats lorsqu'il est associé à la méthode de "hard-thresholding"". La deuxième conclusion est que ce seuil peut être modifié par $T = \kappa\sigma\sqrt{log(n)}$, où κ est un coefficient d'ajustement. En effet, le seuil universel peut être amélioré selon l'application et des seuils inférieurs sont meilleurs si l'on souhaite mesurer la performance par l'erreur quadratique moyenne [Coif 95].

La phase de débruitage de score par ondelettes peut se décomposer en 4 étapes :

1. Une transformée 1D en ondelettes à 2 niveaux (ondelette de “haar”) est effectuée sur chaque ligne de la matrice de similarité,
2. À chaque échelle de décomposition, le niveau de bruit σ est estimé à l'aide de l'écart-type usuel,
3. Le seuil universel de Donoho modifié $T = \kappa\sigma\sqrt{log(n)}$ est calculé,
4. La méthode du “hard thresholding” est utilisée et la transformée 1D inverse en ondelettes est appliquée.

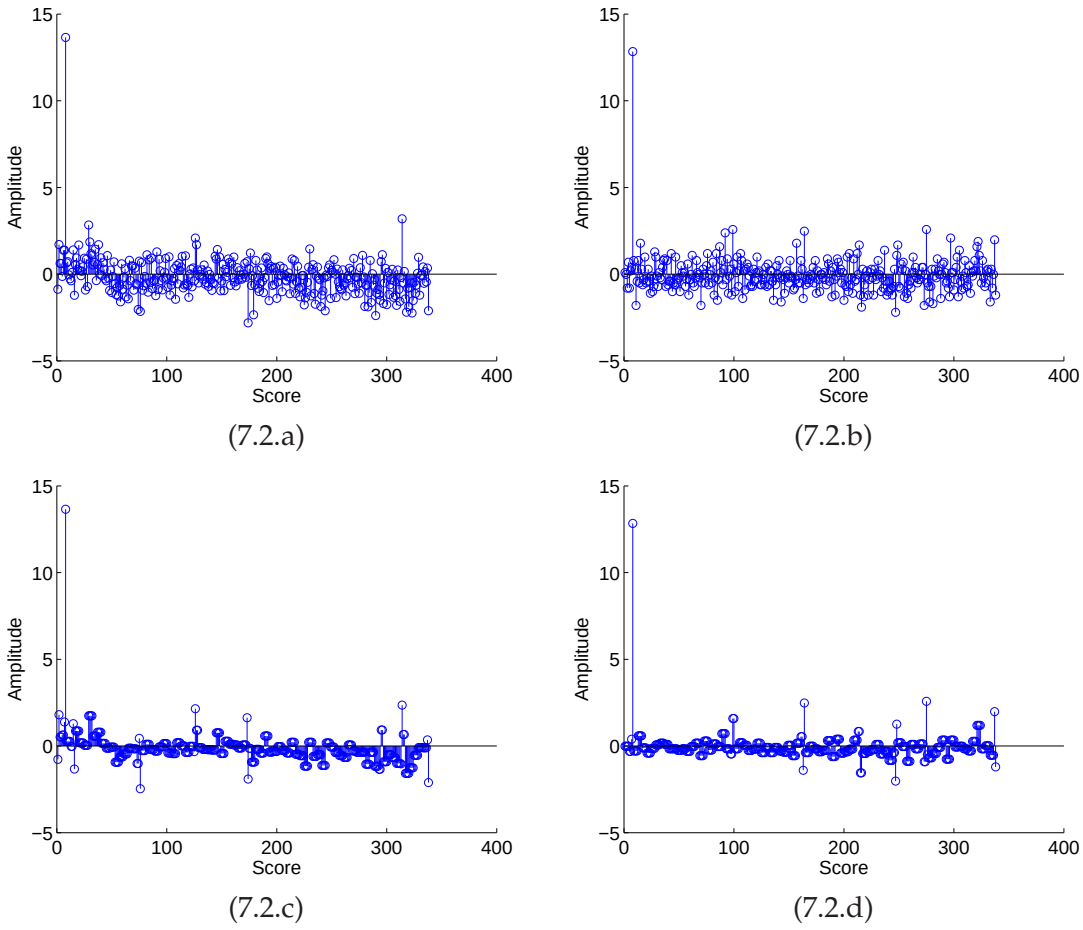


FIG. 7.2 – Illustration du débruitage de score par ondelettes ($\kappa = 0.92$) sur les vecteurs de score du visage et de l'iris d'un utilisateur. Scores originaux du visage (7.2.a) et de l'iris (7.2.b). Scores débruités par ondelettes du visage (7.2.c) et de l'iris (7.2.d).

On peut apprécier la conséquence de la méthode du “hard thresholding” sur la figure 7.2 qui montre clairement que l'amplitude du score authentique est très bien conservée tandis que l'amplitude des scores imposteurs diminue fortement.

7.2.2 Distance de séparation (Δ)

Ce premier paramètre est simplement la distance entre le plus grand score imposteur et le score authentique d'un utilisateur donné (Fig. 7.3).

Maximiser cette distance de séparation contribue à minimiser la zone de recouvrement précédemment introduite.

7.2.3 Coefficient de dissymétrie : "*Skewness*" (γ_1)

Le coefficient de dissymétrie est le *troisième moment statistique standard*. C'est une mesure de la dissymétrie d'une distribution de probabilité d'une valeur aléatoire réelle x, ayant une moyenne μ et un écart-type σ (7.1) :

$$\gamma_1 = Skew(X) = \frac{E[(X-\mu)^3]}{\sigma^3} \tag{7.1}$$

Si l'extrémité de gauche d'une distribution est plus prononcée que l'extrémité de droite, la fonction possède un coefficient de dissymétrie négatif. Si l'opposé est vrai, la fonction possède alors un coefficient de dissymétrie positif. Si les deux sont égales, le coefficient de dissymétrie est nul (Fig. 7.4).

D'après la figure 7.1, nous voudrions que la distribution des scores imposteurs soit **la plus élancée possible vers la gauche**. Cette dernière doit donc, selon la figure 7.4, avoir un coefficient de dissymétrie positif le plus grand possible. Ceci peut être obtenu en maximisant γ_1.

7.2.4 Coefficient d'aplatissement : "*Kurtosis*" (γ_2)

Le coefficient d'aplatissement est le *quatrième moment statistique standard*. Ce coefficient mesure l'aspect "étroit" d'une distribution de probabilité d'une valeur aléatoire réelle ayant une moyenne μ et un écart-type σ (7.2) :

$$\gamma_2 = Kurt(X) = \frac{E[(X-\mu)^4]}{\sigma^4} \tag{7.2}$$

Une distribution qui possède un coefficient d'aplatissement élevé aura un pic plus étroit et des extrémités plus allongées, tandis qu'une distribution qui possède un faible coefficient d'aplatissement aura un pic plus arrondi et des "épaules" plus larges (Fig. 7.5).

D'après la figure 7.1, nous voudrions que la distribution des scores imposteurs soit **la plus étroite possible**. Cette dernière doit donc, selon la figure 7.5, posséder un coefficient d'aplatissement positif le plus grand possible. Cela peut être obtenu en maximisant γ_2.

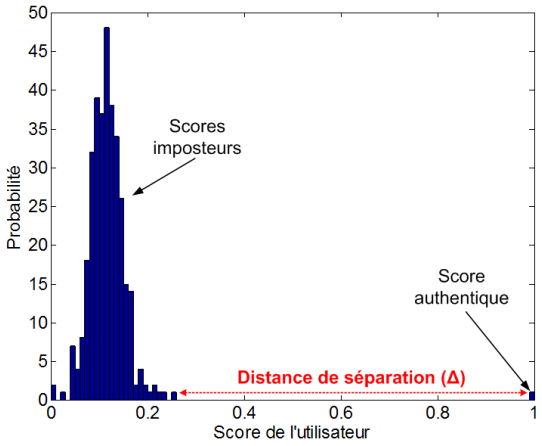


FIG. 7.3 – Distance de séparation pour un utilisateur donné (Δ).

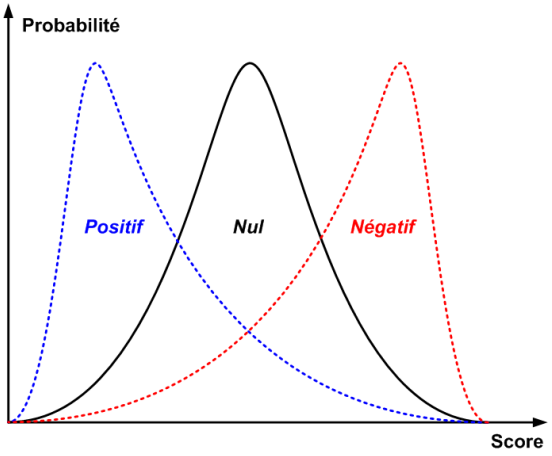


FIG. 7.4 – Coefficient de dissymétrie (γ_1).

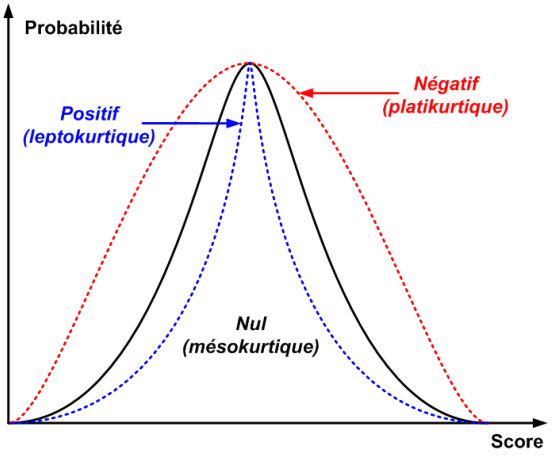


FIG. 7.5 – Coefficient d'aplatissement (γ_2).

7.2.5 Règle de Fusion de la Méthode WSDSF

Soit $S_{visage,i}$ et $S_{iris,i}$, les vecteurs de scores débruités par ondelettes provenant respectivement des matrices de similarité du visage et de l'iris, pour le $i^{\text{ème}}$ utilisateur. Soit $p_{1,i}$ and $p_{2,i}$ les poids des modalités du visage et de l'iris, respectivement, pour le $i^{\text{ème}}$ utilisateur. Nous faisons varier les poids $p_{1,i}$ et $p_{2,i}$ dans l'intervalle $[0, 1]$, avec un pas de $\frac{1}{100}$ et avec la contrainte $p_{1,i} + p_{2,i} = 1$.

Pour le i^{eme} utilisateur et pour chaque paire de poids $\{p_{1,i}, p_{2,i}\}$:

1. Nous calculons la première somme pondérée des vecteurs de score (7.3) :

$$\Sigma_{(p_{1,i},p_{2,i})} = p_{1,i} \times S_{visage,i} + p_{2,i} \times S_{iris,i} \ , \tag{7.3}$$

2. Ensuite, nous calculons la distance de séparation entre le score authentique et le score imposteur le plus élevé (7.4). Puisque nous travaillons avec des scores de similarité, le score authentique est supposé être égal à $\max(\Sigma_{(p_{1,i},p_{2,i})})$, les scores restant étant les scores imposteurs :

$$\Delta_i = \max(\Sigma_{(p_{1,i},p_{2,i})}) - \max(\text{scores imposteurs}_i) \ , \tag{7.4}$$

3. Nous estimons également $\gamma_{1,i}$ (7.1) et $\gamma_{2,i}$ (7.2) à partir de la distribution des scores imposteurs du $i^{\text{ème}}$ utilisateur.

On répète ensuite les étapes 1 à 3 jusqu'à ce que toutes les paires de poids aient été utilisées. Ensuite, nous trouvons l'indice $k_i \in [0, 1]$, définissant la meilleure paire de poids $P_{1,i}$ et $P_{2,i}$ (7.5) telle que :

$$k_i = \arg\max \left(\Delta_i * \gamma_{1,i} * \gamma_{2,i}\right) \ , \tag{7.5}$$

Enfin, le $i^{\text{ème}}$ vecteur de score fusionné est exprimé comme suit (7.6) :

$$S_{fusion,i} = P_{1,i} \times S_{visage,i} + P_{2,i} \times S_{iris,i} \ . \tag{7.6}$$

Ce nouveau vecteur $S_{fusion,i}$ constitue donc la $i^{\text{ème}}$ ligne de la matrice de similarité de fusion.

7.3 Conclusion

Cette technique adaptative de combinaison de scores utilise la théorie de débruitage par ondelettes sur les scores de similarité issus des modalités du visage et de l'iris. Elle permet de maximiser la séparation entre les scores imposteurs et les scores authentiques, pour chaque utilisateur, en prenant en compte la forme des distributions des scores imposteurs à l'aide de moments statistiques standards. Cette technique s'adapte à chaque utilisateur et permet d'optimiser de manière dynamique l'influence de chaque modalité sur la fusion multimodale. On peut noter qu'elle peut s'appliquer à plus de deux modalités, sous réserve que les scores imposteurs correspondants suivent des distributions gaussiennes et soient normalisés par une normalisation Z-Score.

Chapitre 8

Résultats Expérimentaux

8.1 Introduction

Les résultats expérimentaux que nous allons présenter sont divisés en deux catégories. Nous donnerons tout d'abord les résultats qui ont été obtenus concernant la reconnaissance faciale à travers la conception d'une interface graphique couplée à notre caméra, fonctionnant en environnement réel ; nous donnerons également des résultats concernant une étude comparative de plusieurs implémentations matérielles de l'algorithme PCA. La deuxième partie des résultats se concentre tout particulièrement sur la fusion multimodale en exploitant deux techniques de fusion différentes.

8.2 Interface Graphique *FIRST*

L'interface graphique utilisateur (GUI) **FIRST** ("**F**ace and **I**ris **R**ecognition **S**ys**T**em") a été conçue sous MATLAB avec l'outil *GUIDE* (Fig. 8.2) afin de pouvoir valider la performance et la rapidité du système unimodal de reconnaissance faciale avant de réfléchir à des méthodes de fusion. En effet, il est intéressant de posséder des algorithmes unimodaux performants si l'on veut avoir un système multimodal efficace. La *GUI FIRST* est capable de travailler avec une base de données, en particulier la base de données FERET, mais elle permet surtout de travailler en environnement réel lorsque elle est reliée en réseau à la caméra de sécurité SONY SNC-RZ50P (Fig. 8.1) le contrôle total de la caméra se faisant par une interface web.

L'interface *FIRST* est constituée des modules suivants :

1. **Menu** : il permet de générer ou de charger un *Face Space* à partir de données d'entraînement et de la galerie. Le menu permet également de retirer une personne de la base de données ou encore de tracer des courbes de performance telles que les courbes CMC ou ROC,
2. **Enrôlement** : ce module permet tout simplement d'enregistrer une personne dans la base de données (en pratique, elle est incorporée à la base de données FERET),
3. **Test** : ce module permet de confronter l'image de visage d'un individu au système de reconnaissance faciale. On peut soit choisir une image de l'un des ensembles de

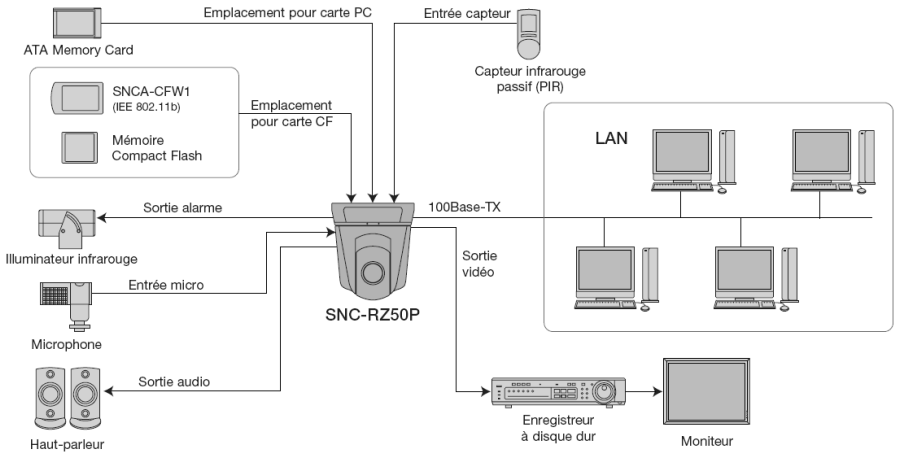


FIG. 8.1 – Configuration du système avec la caméra.

tests de la base de FERET (grâce au bouton parcourir (“Browse”)), soit prendre une image de la caméra (bouton Acquérir (“Acquire”)) ou encore tester une image de visage provenant d’une source quelconque,

4. **Reconnaissance** : cette partie affiche l’image de visage de la personne que le système a reconnue (il s’agit d’une image de la base de données),
5. **Affichage des scores** : ce module affiche les scores de similarité de la personne testée avec la base de données. Le trait en pointillé représente le seuil de sécurité du système qui permet de prendre la décision d’accepter ou de rejeter un individu dans le cas où l’on opère en mode vérification, etc.,
6. **Affichage des informations** : ce module affiche quelques informations importantes comme ce que doit faire l’utilisateur qui dirige l’interface durant les différentes opérations, le temps d’exécution, si la personne est connue du système ou non, une fausse acceptation ou un faut rejet,
7. **Affichage des correspondances suivantes** : cette partie permet de voir les 7 autres personnes qui "ressemblent" le plus à la personne identifiée (donc du rang 2 à 8).

En moyenne, pour reconnaître **un individu parmi 1200 personnes** de la base de données FERET, il faut un peu moins de **200ms**. Précisons que l’algorithme utilisé est le PCA modifié (avec double compression par ondelettes des images de visage normalisées), couplé avec la mesure de similarité MahCosine (*cf. Annexe A.2.4*) ; enfin la machine utilisée est un ordinateur portable *ASUS A6J* avec processeur *Intel* double-coeur **T2400** cadencé à **1.83 Ghz**, muni de **2 Go de RAM** sous l’environnement *Windows XP*.

Les récents travaux du stagiaire *Than Chaleenukul* (*Sirindannhorn International Institute of Technology* (SIIT), *Thailand*) ont permis de s’affranchir de l’interface web en mettant en place une série de commandes executables directement depuis MATLAB, le but était de pouvoir automatiser les actions de la caméra au lieu de faire des opérations manuelles de déplacement, de zoom et de capture d’image.

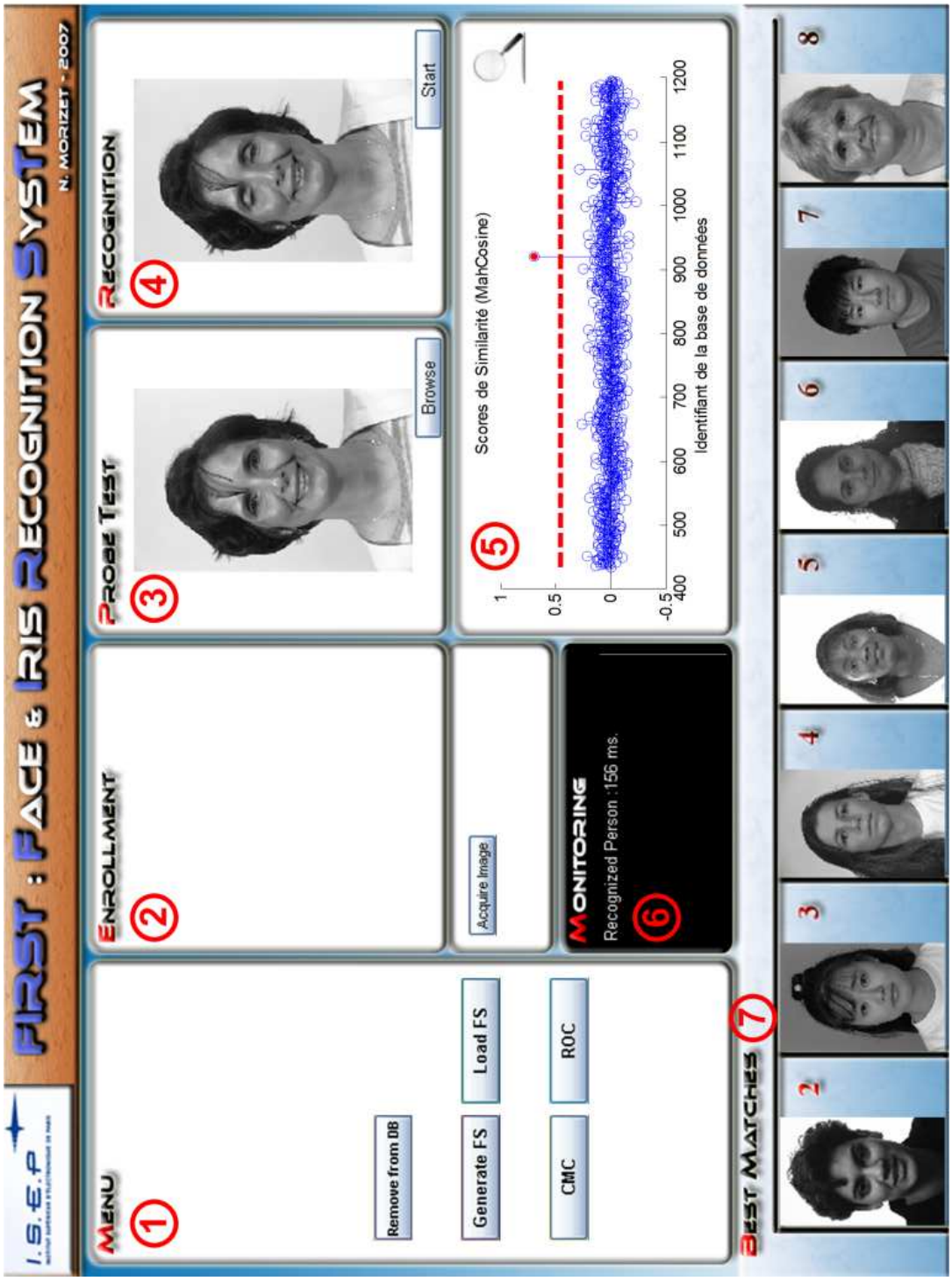


FIG. 8.2: L'interface graphique utilisateur FIRST (test FERET).

Une longue étude du protocole propriétaire *VISCA* de SONY ainsi que la création d'une librairie de fonctions en *langage C* ont permis de faire faire à la caméra un panoramique horizontal automatique sur 180^o. La caméra s'arrête brièvement tous les 18^o afin de prendre une image, on appelle ensuite une librairie de lien dynamique (*dll*) de détection de visage dont l'algorithme est basé sur des machines à vecteurs de support [Kien 05].

8.3 Implémentation matérielle du PCA

Nous avons effectué une étude comparative [Mori 07b] concernant une implémentation de la phase de reconnaissance de notre algorithme PCA modifié (cf. Chapitre 6.4.2) sur différentes architectures matérielles. Le but étant de pouvoir choisir la meilleure solution pour concevoir un système de reconnaissance faciale en temps réel, à partir d'un banc de tests qui permet de comparer un ordinateur, un DSP, un FPGA et un bloc IP.

La phase d'apprentissage de notre algorithme PCA modifié (prenant en entrée des images compressées deux fois par ondelettes), utilise la base de données d'images de visage FERET : l'ensemble des données d'entraînement ("*FERET Training Set*", 501 personnes) et la galerie ("*FERET Gallery*", 1196 personnes) ont servi à générer les valeurs propres et les vecteurs propres (Fig. 8.3) constituant le "*Face Space*". Nous avons retenu les 240 premiers vecteurs propres, ce qui correspond à 98.5% de l'énergie totale, soit 48% du nombre total de vecteurs propres (pour rester conforme avec les tests du *CSU System 5.0* [Beve 05]). Ainsi, chaque vecteur propre est de taille (1254×1).

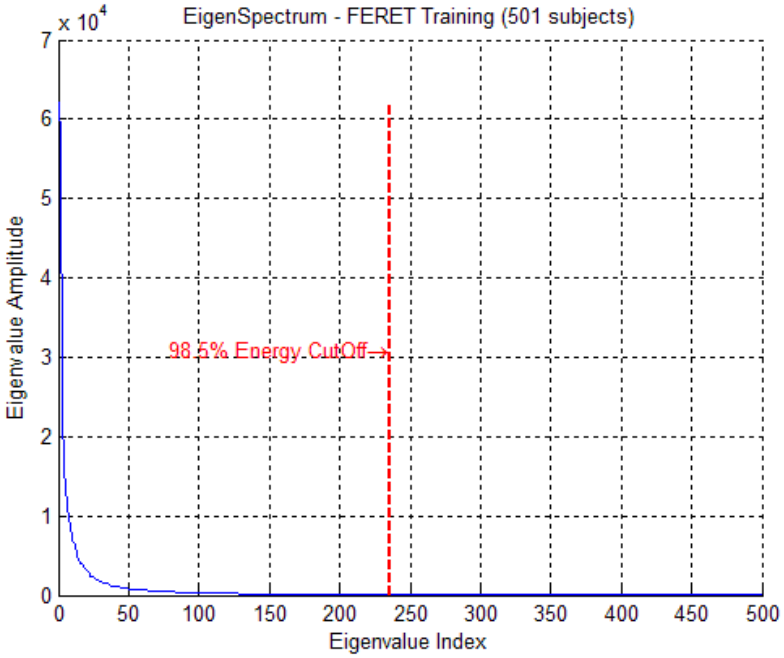


FIG. 8.3 – Spectre d'énergie propre du "FERET Training Set".

La phase de reconnaissance utilise l'ensemble de test de changements en expression faciale ("*FERET FB Probe Set*", 1195 personnes). La mesure de distance utilisée est la **distance euclidienne au carré** (*SED*) ; nous n'avons pas utilisé la racine carrée d'une distance euclidienne standard afin de gagner un peu en temps de calcul pour le FPGA, ce choix n'a de toute façon **aucune influence sur le taux de reconnaissance du système**.

Pour nos expériences, l'identité de la personne testée est affirmée et nous choisissons de prendre **le minimum** des SEDs comme décision finale. Par conséquent, si l'identifiant d'un individu test est I_t et que l'indice du minimum des SEDs trouvé est I_g, l'individu est accepté si $I_t = I_g$, sinon il est rejeté du système.

Nous décidons d'implémenter l'algorithme PCA sur des plateformes qui n'effectuent pas de calculs en virgule flottante. Nous devons alors étudier, pour tous les calculs, la précision en virgule fixe requise. La plateforme finale n'implémente que la seconde partie de l'algorithme PCA, à savoir la projection de l'image sur le *Face Space*, le calcul de distance (SED) et la détection des minimums des SEDs. Ces étapes ont été codées en *langage C* (variables de précision "doubles").

Nous avons choisi de tester notre système sur différentes plateformes :

1. **Ordinateur** : pour établir et vérifier l'algorithme. Nous n'utiliserons pas cette plateforme qui ne doit servir que de référence en termes de temps de calcul car nous voulons une plateforme embarquée et à faible consommation,
2. **DSP** : nous avons injecté notre programme sur un processeur DSP TI6713 [Texa 08] (avec calculs en virgule flottante),
3. **FPGA** : nous avons lancé le programme sur un processeur *softcore* (Altera NIOS II [Alte 08]). Une architecture avec des blocs IP a également été conçue et "*mappée*" sur le FPGA (Fig. 8.4).

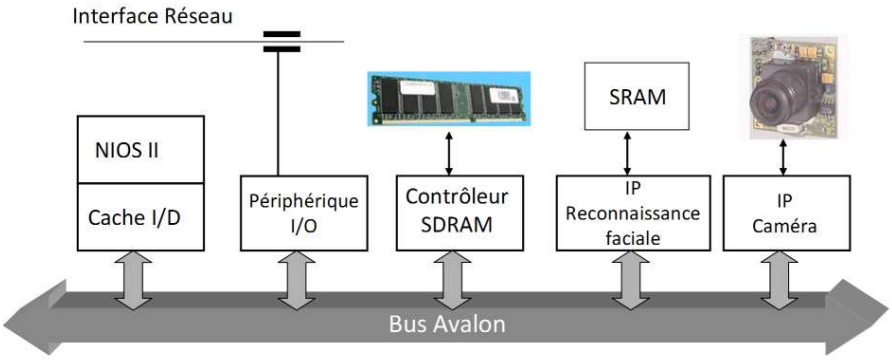


FIG. 8.4 – Architecture du système avec blocs IP.

Par ailleurs, plusieurs versions du NIOS II ont été configurées :

– Le mode "*Soft floating point*" : émulation logicielle de virgule flottante,
– Le mode "*Hard floating point*" : version matérielle de virgule flottante, avec instructions personnalisées,
– Le mode "*Fixed point*" : calculs en virgule fixe, avec des entiers larges sur 32 bits en natif.

Le bloc IP que nous avons conçu sur le FPGA (Fig. 8.5) n'utilise qu'un seul opérateur MAC (multiplications/additions) avec un séquenceur spécifique afin d'extraire les données de la mémoire. Ainsi, cette architecture prend très peu d'espace et possède une précision suffisante. La mémoire est alternativement partagée entres les différents blocs.

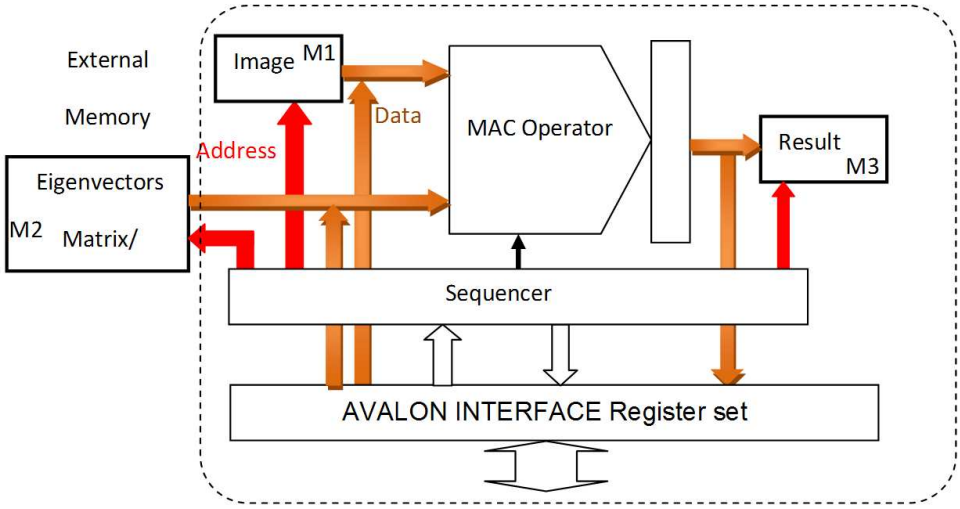


FIG. 8.5 – Architecture interne du bloc IP de reconnaissance faciale.

Plateforme	**Temps de reconnaissance**
PC 3 GHz	*3ms*
DSP 225 MHz	**5ms**
NIOS, mode "*soft floating point*"	62s
NIOS, mode "*hard floating point*"	28s
NIOS, mode "*fixed point*"	15s
IP 50 MHz	**60ms**

TAB. 8.1 – Temps de reconnaissance pour les différentes architectures testées.

Le tableau 8.1 montre les résultats de temps de reconnaissance pour les différentes architectures testées. Les résultats montrent une assez mauvaise performance du *NIOS* en mode "*soft core*" même lorsque les calculs sont effectués en virgule fixe. En revanche, le DSP donne un temps de reconnaissance très acceptable mais le bloc IP offre, pour le FPGA, un bon compromis entre les performances et le matériel additionnel requis. De plus, il est possible d'augmenter la fréquence de fonctionnement du bloc IP (qui n'est, ici, "que" de 50MHz) et d'en implémenter plusieurs en parallèle, ce qui peut représenter un sérieux concurrent à la solution non embarquée que représente l'utilisation de l'ordinateur.

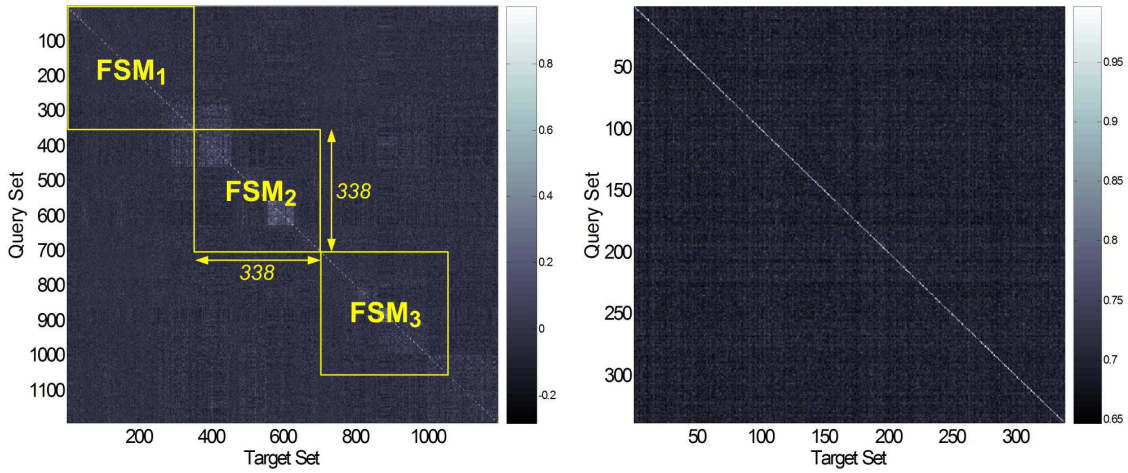
Je tiens à remercier *Frédéric Amiel* (ISEP), ma stagiaire *Insaf Dris-Hamed* (*Escuela Técnica Superior de Ingenieros* (ETSI), *Sevilla*) et Thomas Ea (ISEP) pour leurs collaborations.

8.4 Résultats de fusion multimodale

Avant de présenter nos deux techniques de fusion multimodale, nous pouvons montrer un premier résultat concernant l'hypothèse d'indépendance de modalité qui nous permet de pratiquer les tests sur des bases de données chimériques. Pour cela, nous nous référons à la Définition 1 (Chapitre 4.3.4.1) : soit respectivement X et Y les matrices de similarité centrées (de moyenne nulle) des scores du visage et de l'iris ; nous obtenons $E(XY) - E(X)\,E(Y) \approx 0.0001$. Ceci prouve que l'on peut très bien, en pratique, considérer les matrices de similarité comme étant statistiquement indépendantes (au sens faible) lorsqu'on décide d'effectuer une fusion au niveau score et que nous pouvons donc les fusionner pour créer des individus "virtuels".

Par ailleurs, il existe un protocole très précis permettant de générer des matrices de similarité et de conduire des évaluations de méthodes de normalisation de scores et de techniques de fusion. Ce protocole se décompose en huit étapes et permet d'évaluer la performance d'un système biométrique multimodal ; il a été mis au point par le *National Institute of Standards and Technology* (*NIST*) [Snel 03] :

1. **Assembler deux ensembles de signatures biométriques** ("Target Set" et "Query Set"). En pratique, l'intersection de ces deux ensembles ne doit pas être nulle,
2. **Obtenir un score pour chaque paire de signatures** "Query/Target" et les enregistrer dans une matrice de similarité dont la taille est : taille(Query Set) $\times$ taille(Target Set),
3. **Extraire un "Gallery Set"** (n'importe quel sous-ensemble arbitraire du Target Set) ainsi qu'un **"Probe Set"** (n'importe quel sous-ensemble arbitraire du Query Set), afin de pouvoir effectuer des expériences sur un sous-ensemble de la population,
4. **Répéter les étapes 1-3 pour chaque modalité biométrique**,
5. **Assembler et "aligner" les matrices de similarité de l'étape 2** : cela inclus la conversion des données vers un format commun (measure de distance ou score de similarité), la formation de sous-ensembles afin d'obtenir des **matrices de même taille** et la correspondance des données pour créer des individus réels ou virtuels, en se basant sur l'hypothèse selon laquelle les modalités individuelles concernées sont statistiquement indépendantes l'une de l'autre et peuvent donc être assignées de manière arbitraire (mais de manière systématique) afin de former des ensembles de sujets virtuels à des fins de tests,
6. **Normaliser les scores** : transformer les scores des matrices de similarité dans un intervalle commun,
7. **Fusionner les scores** : fusionner les ensembles des matrices de similarité normalisées en une seule matrice de similarité de fusion,
8. **Calculer les statistiques de performance** : courbes ROC ou CMC.



(8.6.a) (8.6.b)

FIG. 8.6 – Les trois sous-matrices FSM_1, FSM_2 et FSM_3 (8.6.a) et la matrice de similarité des scores d'iris (ISM) (8.6.b), utilisées dans nos tests de fusion.

Les matrices qui ont été utilisées proviennent des bases de données FERET et CASIA, selon la méthode décrite dans le Chapitre 5.3.1. La matrice de similarité des scores du visage (FSM) est de taille (1193×1193) et celle des scores d'iris (ISM) est de taille (338×338). Aussi, afin de respecter la cinquième étape du protocole d'évaluation et de tests multimodaux qui impose de fusionner des matrices de similarité de même taille, nous avons partitionné la matrice FSM en trois sous-matrices disjointes (FSM_1,FSM_2 et FSM_3) de taille (338×338), en gardant bien évidemment les scores authentiques le long de la diagonale principale (Fig. 8.6.a). Cela signifie donc que les 179 derniers individus de la base de données FERET ne sont pas utilisés pour nos tests. Cela implique également que tous les résultats que nous allons donner sont le fruit de la fusion des trois sous-matrices de similarité de scores du visage avec la matrice de similarité ISM, qui elle, reste fixe (Fig. 8.6.b).

Tous les résultats que nous allons donner sont donc des résultats moyennés sur 1014 individus chimériques (cf. Chapitre 4.3.4.1).

8.4.1 Fusion multimodale 2D

Notre approche [Mori 07a] est née de l'observation de techniques de fusion d'image par ondelettes et s'appuie sur la théorie décrite dans le Chapitre (6.4.3). Ici, **les vecteurs de score des matrices de similarité sont transformés en images qui seront alors fusionnées dans le domaine des ondelettes**. On peut remarquer que l'on fait en quelque sorte l'opération inverse de l'algorithme PCA qui transforme des images de visage en vecteurs-colonne pour le traitement des données, la seule différence ici est que nous opérons au niveau score. Cette transformée peut se faire grâce à un simple "*mapping*" 2D [Mori 07a]. Par ailleurs, puisque nous travaillons avec des scores de similarité, le fait d'améliorer le taux de reconnaissance de notre système multimodal en augmentant la séparation entre les distributions des scores imposteurs et authentiques peut être vu comme une **opération d'amélioration de contraste**, sur des matrices de score de taille réduite.

Puisque nous allons effectuer une MRA sur nos images, notre attention s'est portée sur les **ondelettes temps-échelle**. Nous avons retenu l'ondelette de Daubechies qui est orthogonale à support compact et qui possède le plus grand nombre de moments nuls (cf. Chapitre 6.3.2) pour une largeur de support donnée. Étant donné que plus le nombre de moments nuls est élevé, plus l'amélioration de contraste est importante, une ondelette de Daubechies d'ordre élevé (Fig. 8.7) devrait être adaptée à notre application.

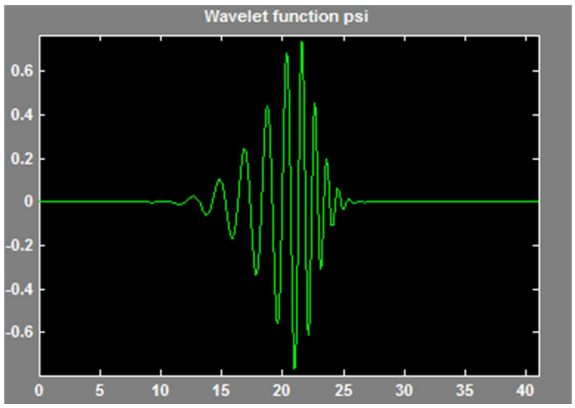


FIG. 8.7 – Ondelette de Daubechies d'ordre 19.

Résumons maintenant le cadre de travail de cette technique de fusion (Fig. 8.8). Pour chaque modalité, les vecteurs de scores de similarité d'un individu sont transformées en matrices que nous allons désormais considérer comme des images en niveaux de gris. Notons que nous leur appliquons la méthode du "*zero-padding*" pour s'assurer d'avoir des matrices carrées, ceci est d'autant plus utile que nous travaillons avec des scores de similarité et qu'aucune information n'est ajoutée pendant ce processus. Nous avons également pris soin de gérer les effets de bords qui constituent un problème courant en traitement d'image en utilisant la technique du "*Half-Point 2D Symmetric Padding*" (**2D-HPSP**) (Fig. 8.9). On applique alors à chaque "image" une transformée discrète 2D en ondelettes : l'ondelette choisie est l'ondelette de *Daubechies* d'ordre 19 (ce choix correspond à la taille de la matrice carrée) et la transformée est effectuée sur huit niveaux ($2^8 < 338$, où 338 est le taille de la matrice carrée de similarité de l'iris).

Ensuite, on regroupe les coefficients d'ondelettes différemment selon les bandes de fréquences et à tous les niveaux de décomposition ("*multiscale grouping*") : le schéma de sélection *MinS* a été choisi pour traiter les coefficients d'*approximation* (basses fréquences) et le schéma *MaxS* pour traiter les coefficients de *détails* (hautes fréquences), car rappelons que nous voulons simultanément diminuer encore plus les faibles scores imposteurs tout en consolidant l'amplitude des scores authentiques. Cette phase de combinaison de coefficients permet d'obtenir une seule matrice de fusion. On applique alors à cette dernière une transformée discrète 2D en ondelettes inverse pour se retrouver dans le domaine des "scores" et on applique la transformée inverse qui nous a permis de passer du vecteur en image au tout début. Enfin, on raccourci le vecteur résultant ("*vector cropping*") en enlevant des "scores-artefacts" qui étaient à la même place que les zéros introduit lors de l'étape de "*zero-padding*". On s'attend alors à ce que le vecteur résultant possède des scores imposteurs d'amplitude plus faibles et un score authentique d'amplitude conservée ou plus grande.

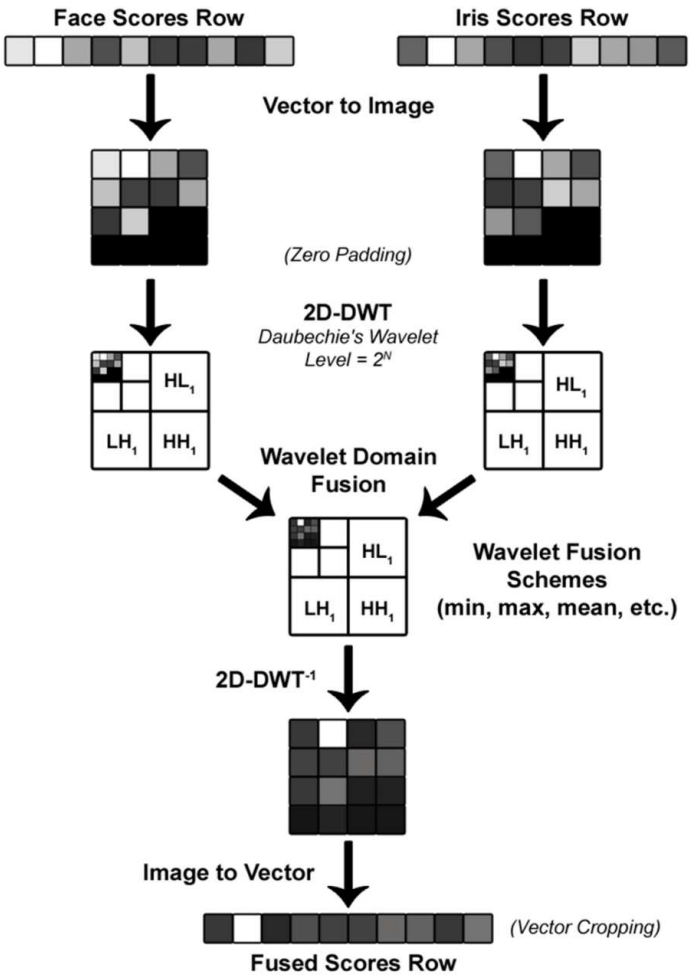


FIG. 8.8 – Cadre de travail de la technique de fusion (exemple de la deuxième ligne des matrices de similarité).

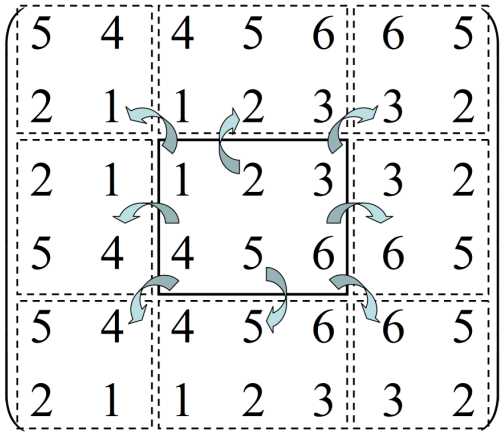


FIG. 8.9 – Illustration de la technique du *"Half-Point 2D Symmetric Padding"* pour gérer les effets de bord sur une matrice (2×3).

Les résultats sont représentés sous forme de courbes ROC (Fig. 8.10) moyennes et de taux d'égale erreur (EER) moyens (Tab. 8.2) pour différentes méthodes de normalisation de scores et de règles de fusion.

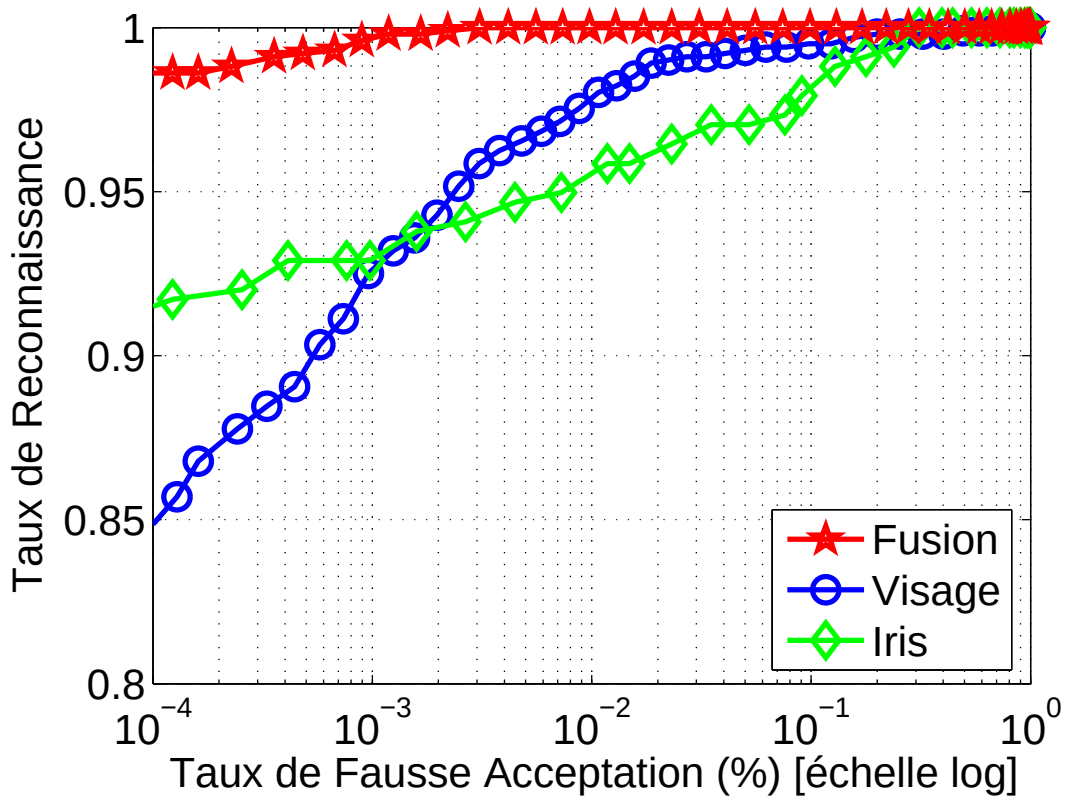


FIG. 8.10 – Courbes ROC moyennes (normalisation MinMax).

Méthode de Normalisation	*Règle de Fusion*			
	Min	Max	Somme	Proposée
MinMax	2.96	0.28	0.04	**0.02**
Z-Score	1.43	0.22	**0.03**	0.09
QLQ	1.88	0.19	0.11	**0.08**

TAB. 8.2 – Taux d'Egale Erreur (EER) moyens (%)

On peut tout d'abord remarquer que la courbe ROC de notre méthode de fusion surpasse largement celles des modalités du visage et de l'iris prises individuellement. De plus, elle donne les meilleurs résultats lorsqu'elle est combinée avec la normalisation *MinMax* avec un EER moyen de 0.02% améliorant respectivement les méthodes MinMax/Somme, Z-Score/Somme et QLQ/Somme de 50.0%, 33.3% et 81.8% ; ces dernières étant considérées comme des techniques de combinaison de score déjà efficaces [Ross 06].

Bien que les résultats en termes de taux de reconnaissance et de taux d'égale erreur soient intéressants, *on peut apporter quelques remarques à cette technique de fusion 2D*. Tout d'abord, la transformée du vecteur de score en matrice (que nous avons vu sous forme d'image) est arbitraire et peut, après tout, être effectuée de différentes façons (celle que nous avons choisie minimise la surface du carrée ainsi obtenu) mais doit être identique aux différentes matrices de similarité de manière à conserver le score authentique à la même place (cf. Chapitre 4.3.4.2). Il est alors assez difficile de formaliser convenablement ce "*mapping*", on peut cependant voir ce processus comme un sous-échantillonnage par bloc du vecteur de score. Par ailleurs, l'utilisation d'une décomposition multi-échelle peut être discutable : en effet, les matrices de score obtenues après le "*mapping*" ne possèdent pas de structure géométrique particulière, ce qui serait différent si l'on devait fusionner directement des matrices de similarité où l'on verrait alors apparaître de forts coefficients (scores authentiques) le long de la diagonale principale.

8.4.2 Fusion multimodale 1D (*WSDSF*)

La méthode de fusion 2D de score dans le domaine des ondelettes semble prometteuse mais les conclusions de la section précédente nous incitent à nous tourner vers un type d'approche similaire mais plus facile à formaliser. Pourquoi ne pas traiter directement les vecteurs de score en une dimension ? Les résultats, issus de la technique *WSDSF* (Chapitre 7), que nous allons maintenant présenter s'appuient sur les résultats intermédiaires des analyses statistiques du Chapitre 5.

Nous avons montré que nous pouvions assimiler les scores imposteurs des deux modalités à du bruit blanc gaussien lorsqu'ils étaient normalisés par la normalisation Z-Score. Le but de cette technique [Mori 08] est donc de débruiter les scores imposteurs relatifs à chaque individu et d'optimiser les scores fournis par les deux modalités, en ce sens que les scores du visage et d'iris n'ont pas obligatoirement le même poids selon les individus. Par exemple, on devrait pouvoir attribuer à un score de similarité issu d'une image d'iris de mauvaise qualité un poids inférieur par rapport à un score de similarité issu d'une image de visage prise dans de bonnes conditions et, au contraire, un poids supérieur à un score d'iris lorsque celui du visage est perturbé par une trop forte variation en éclairement, pose ou expression faciale.

Dans les résultats que nous allons présenter, il est important de comprendre que le taux d'égale erreur (EER), qui ne constitue qu'un point de fonctionnement, ne suffit pas pour évaluer la performance d'une technique de fusion. En effet, il est nécessaire de posséder, en plus, une vue d'ensemble et d'observer, par exemple, la vitesse à laquelle la courbe ROC correspondante tend vers un taux de reconnaissance de 100%. Cela signifie qu'une technique A peut avoir un meilleur EER moyen qu'une technique B sans qu'elle soit globalement meilleure. C'est pourquoi nous présenterons nos résultats sous la forme de deux graphiques pour chaque permutation "méthode de normalisation de score/règle de fusion" : une première figure illustrant les densités de probabilité des scores de fusion accompagnées de la distance de séparation globale entre les deux distributions ; cette dernière est représentée par l'intervalle compris entre les deux traits en pointillés (une distance de séparation négative indique un recouvrement des distributions tandis qu'une distance de séparation positive indique un écartement des distributions), un deuxième graphique montrant la courbe ROC de fusion comparée à celles des modalités individuelles du visage et de l'iris.

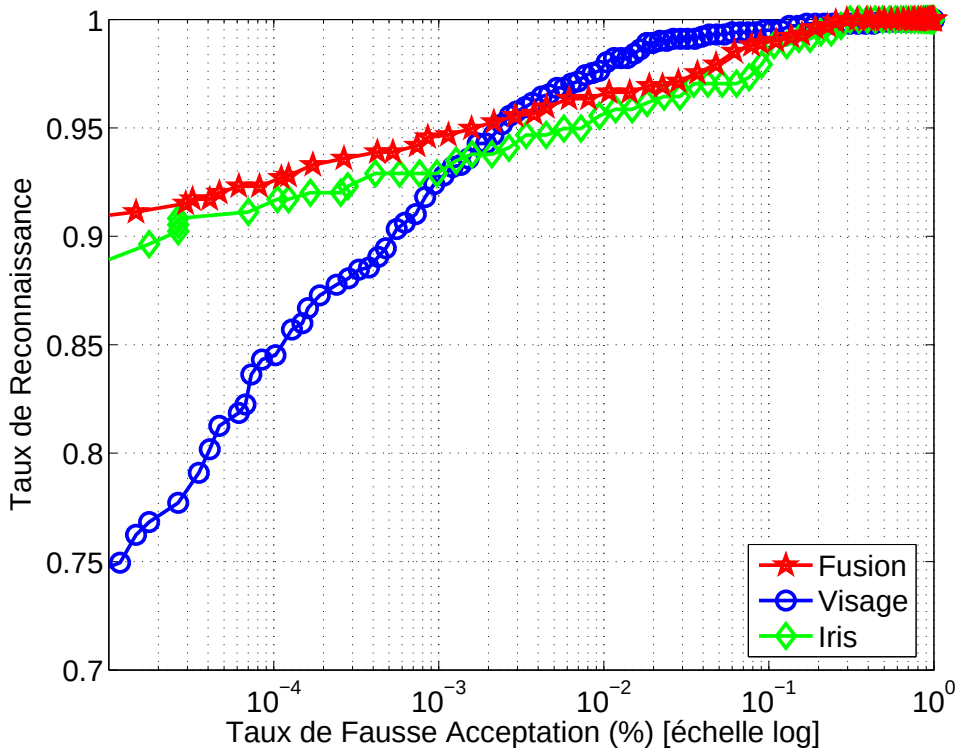


FIG. 8.11 – Courbes ROC pour la fusion MinMax-Min.

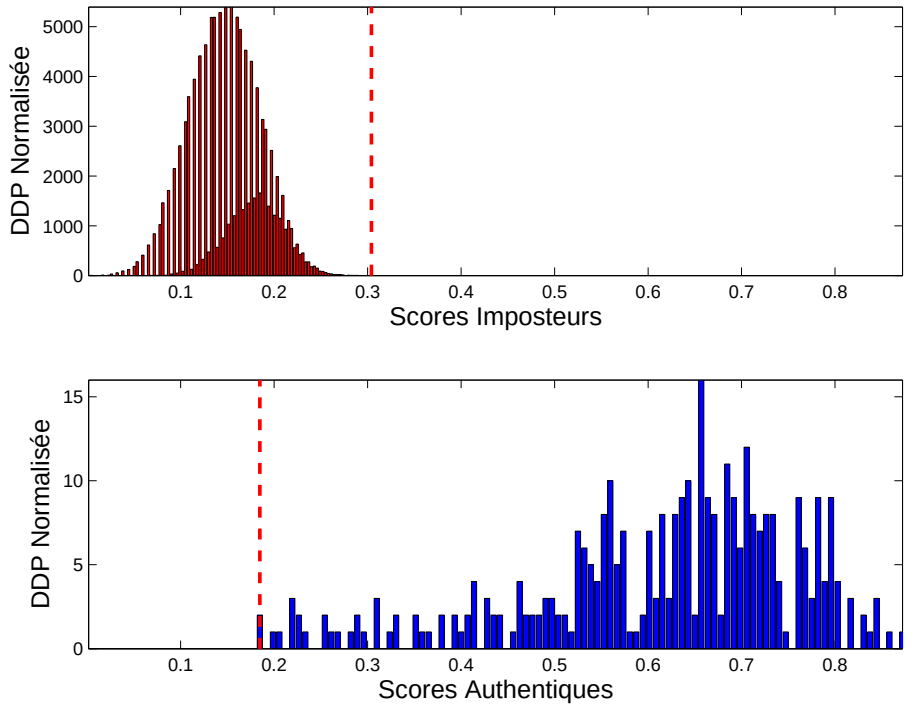


FIG. 8.12 – Densités de probabilité pour la fusion MinMax-Min.

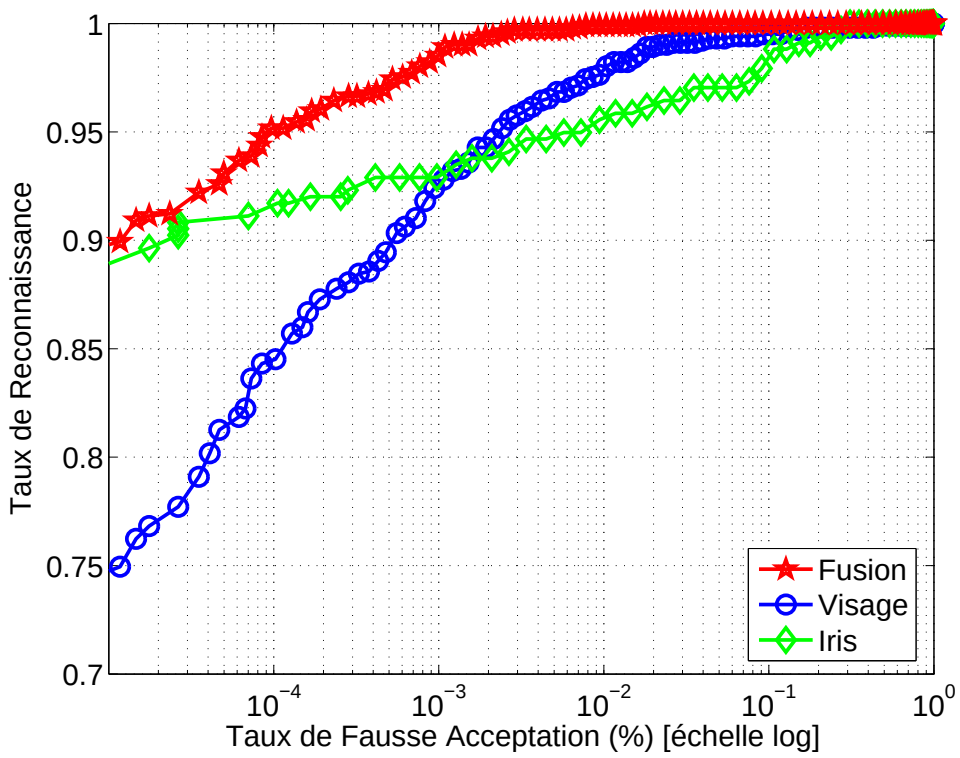


FIG. 8.13 – Courbes ROC pour la fusion MinMax-Max.

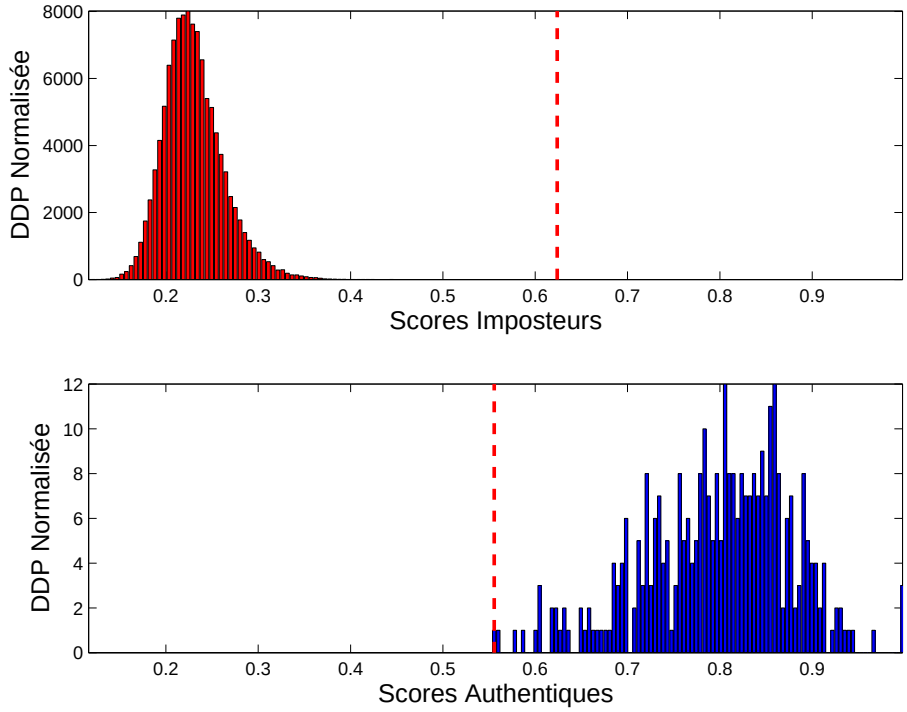


FIG. 8.14 – Densités de probabilité pour la fusion MinMax-Max.

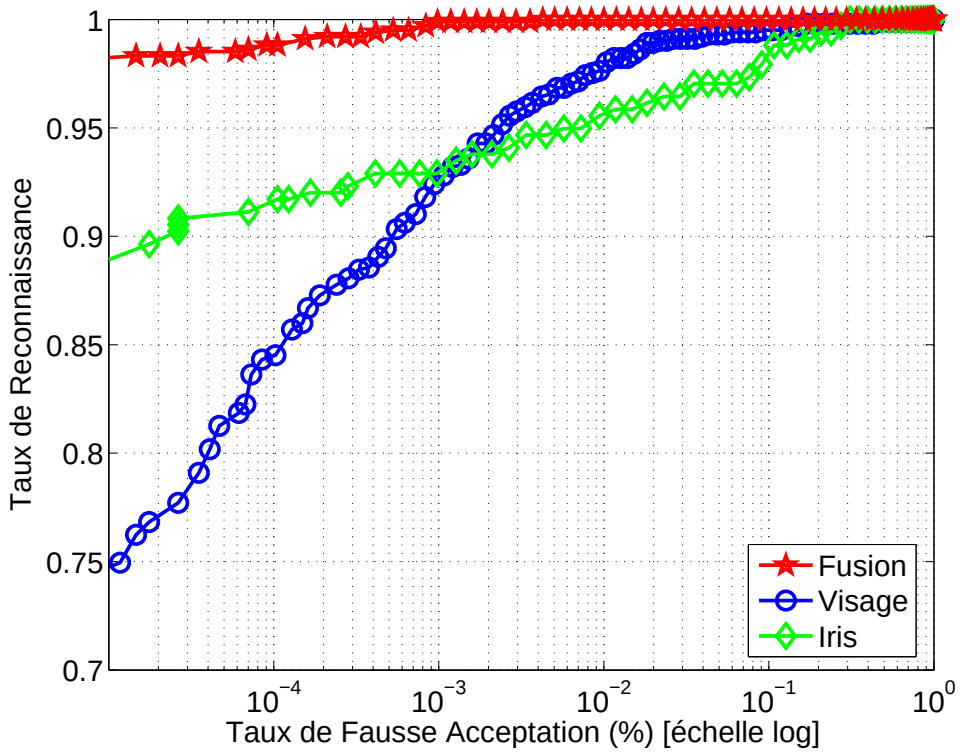


FIG. 8.15 – Courbes ROC pour la fusion MinMax-Somme.

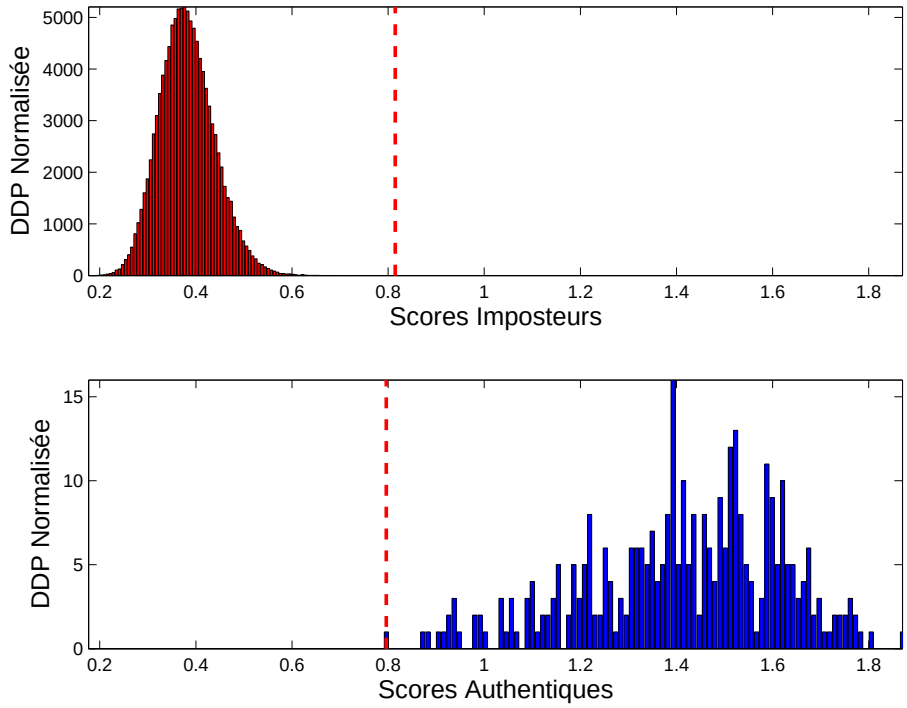


FIG. 8.16 – Densités de probabilité pour la fusion MinMax-Somme.

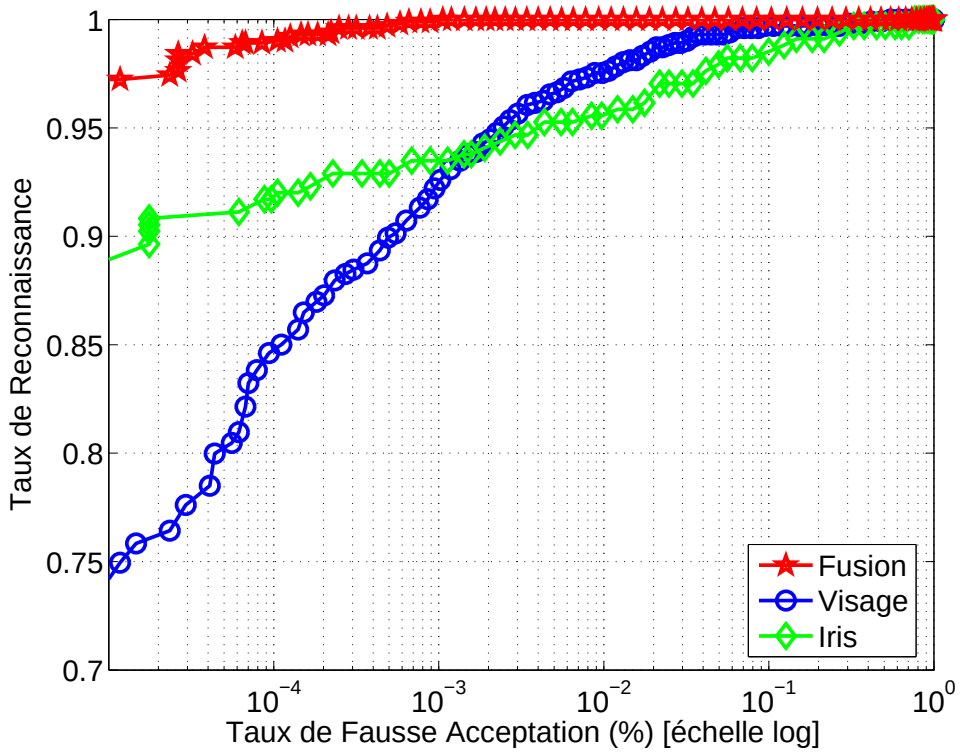


FIG. 8.17 – Courbes ROC pour la fusion MinMax-WSDSF.

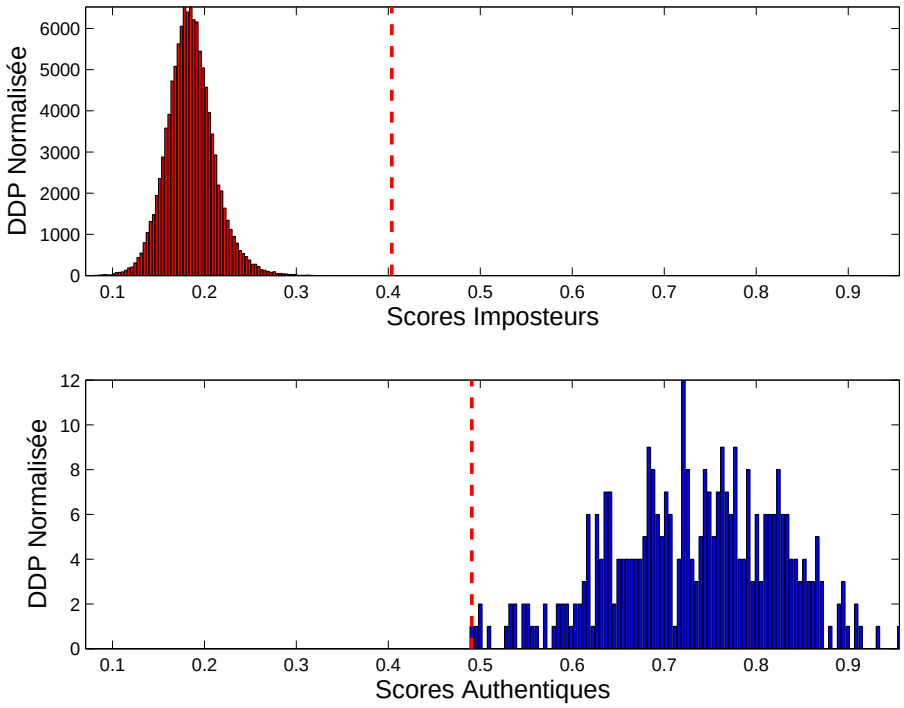


FIG. 8.18 – Densités de probabilité pour la fusion MinMax-WSDSF.

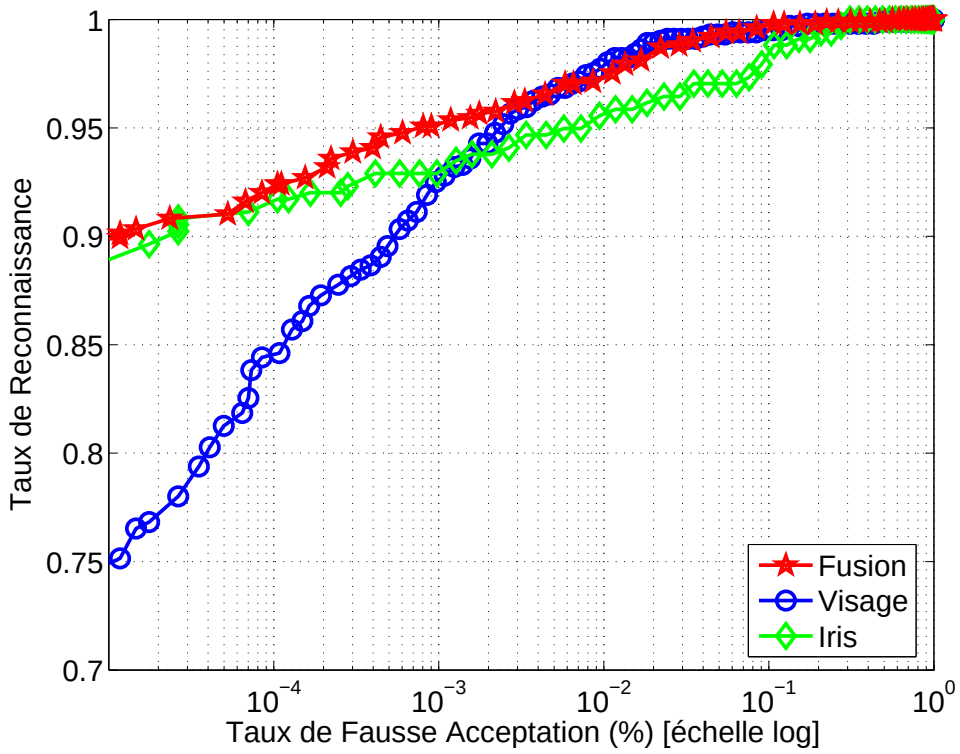


FIG. 8.19 – Courbes ROC pour la fusion ZScore-Min.

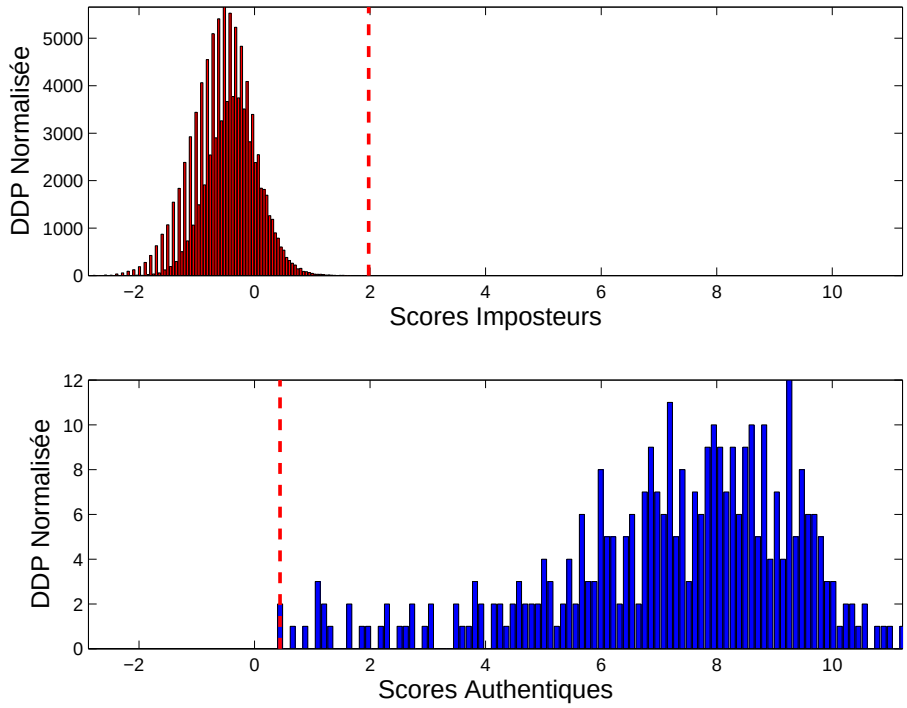


FIG. 8.20 – Densités de probabilité pour la fusion ZScore-Min.

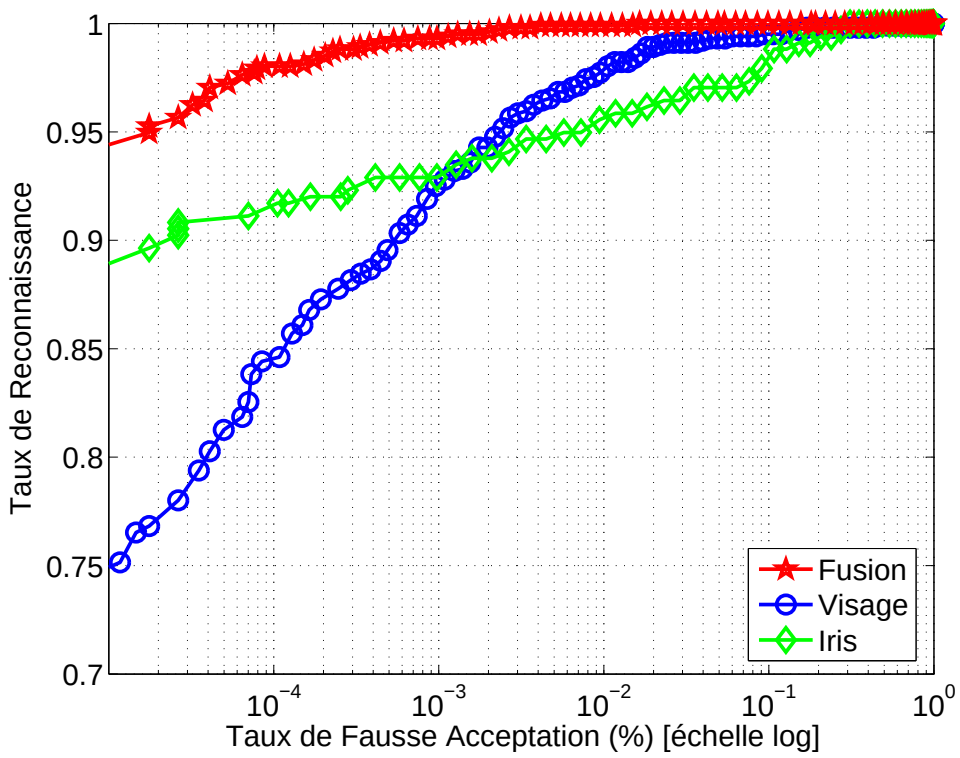


FIG. 8.21 – Courbes ROC pour la fusion ZScore-Max.

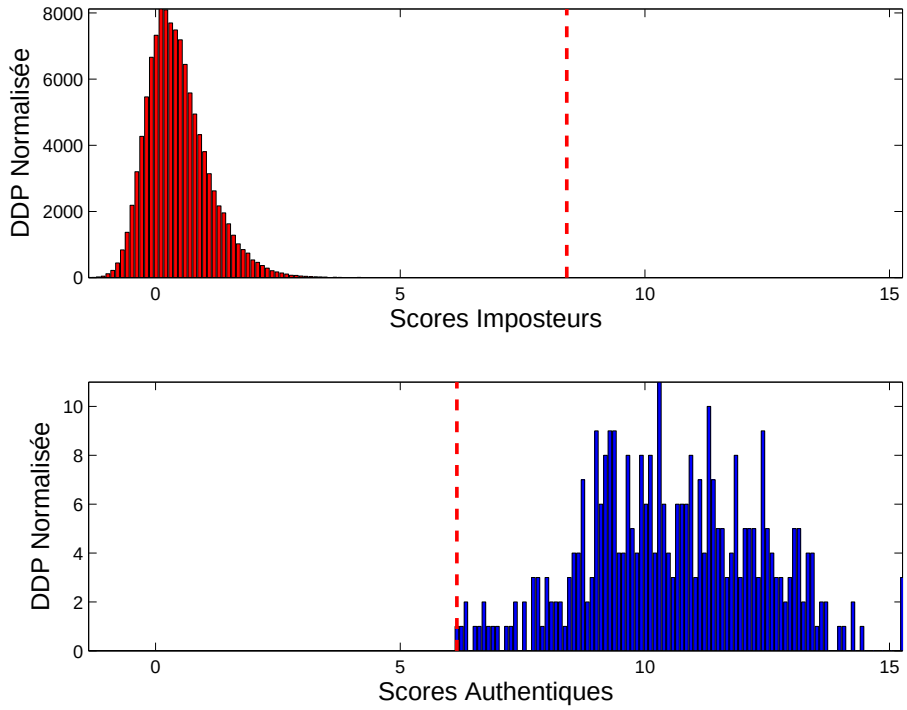


FIG. 8.22 – Densités de probabilité pour la fusion ZScore-Max.

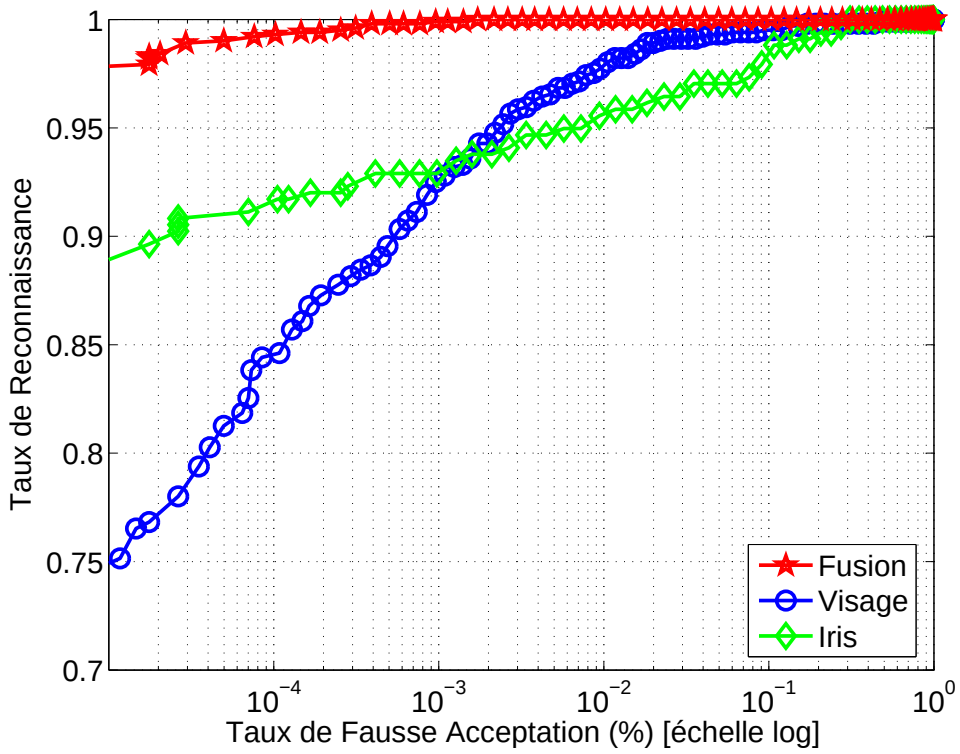


FIG. 8.23 – Courbes ROC pour la fusion ZScore-Somme.

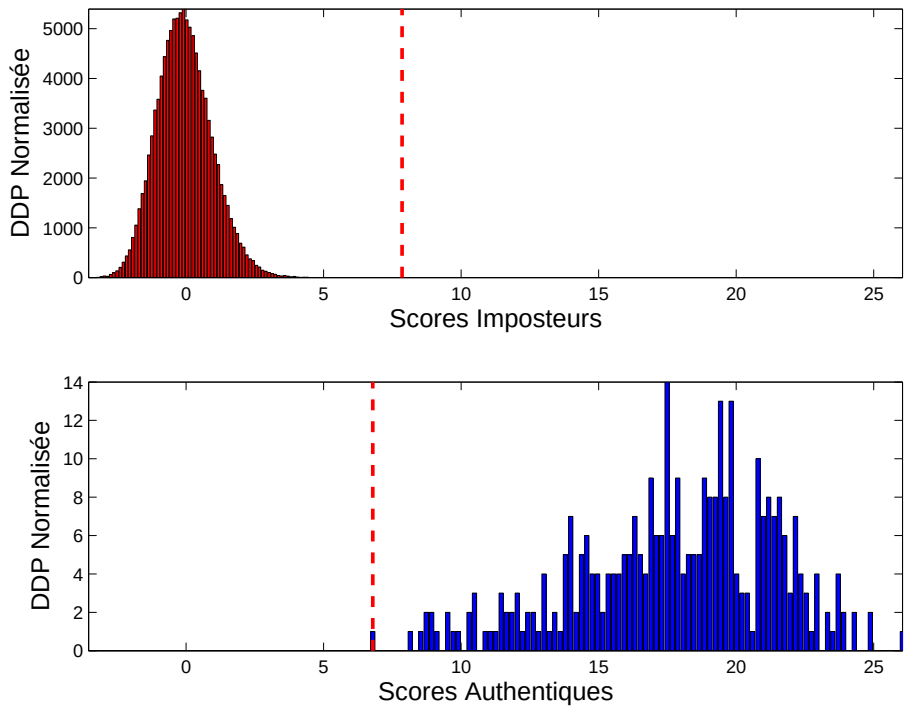


FIG. 8.24 – Densités de probabilité pour la fusion ZScore-Somme.

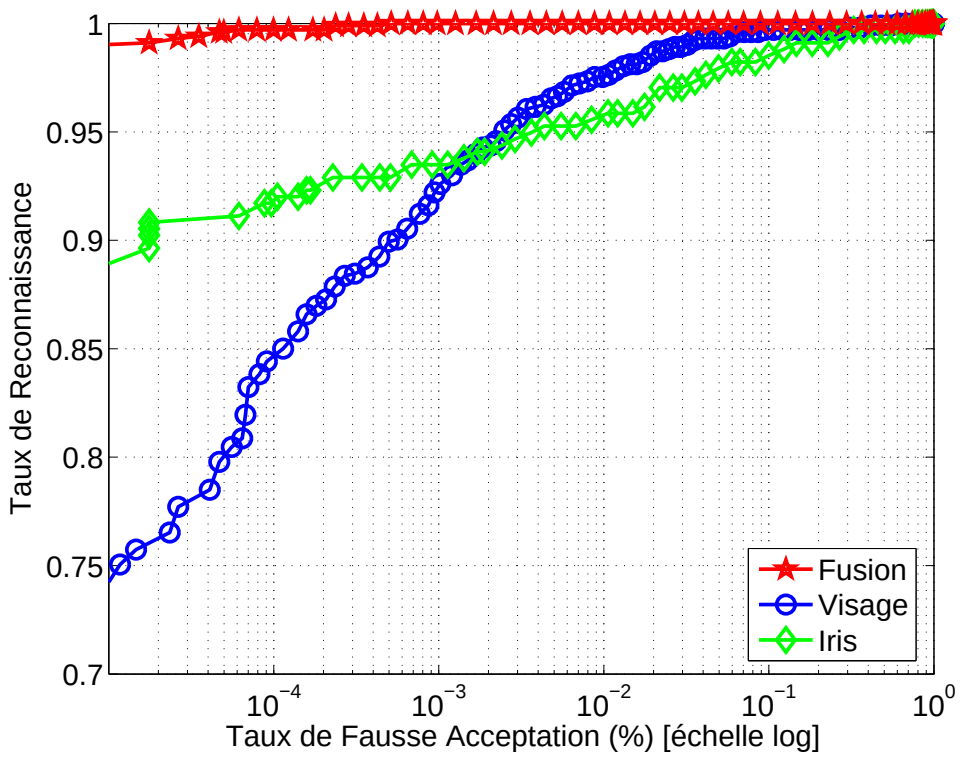


FIG. 8.25 – Courbes ROC pour la fusion ZScore-WSDSF.

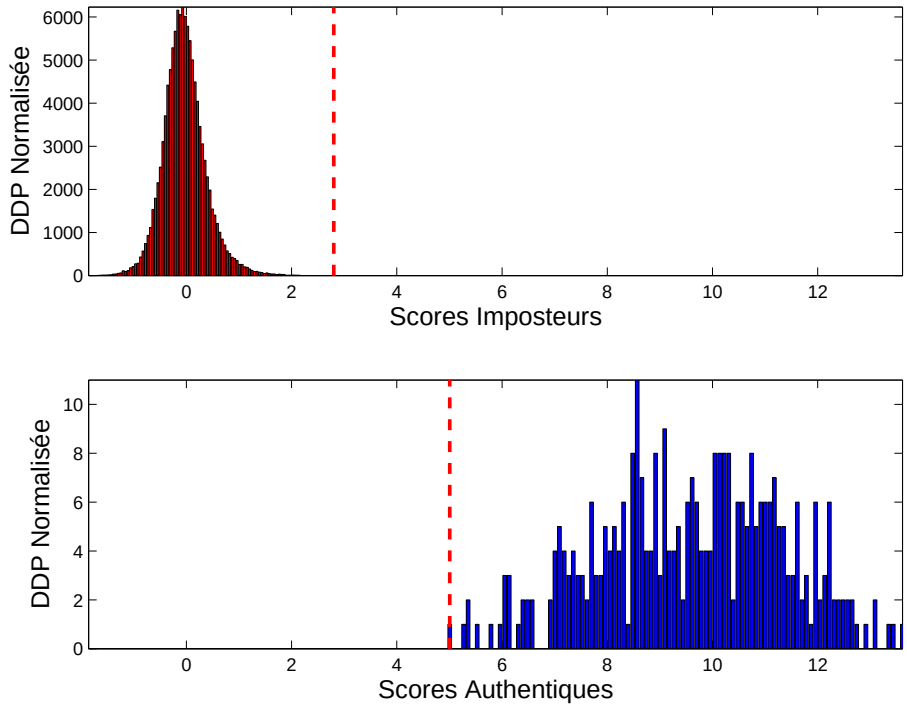


FIG. 8.26 – Densités de probabilité pour la fusion ZScore-WSDSF.

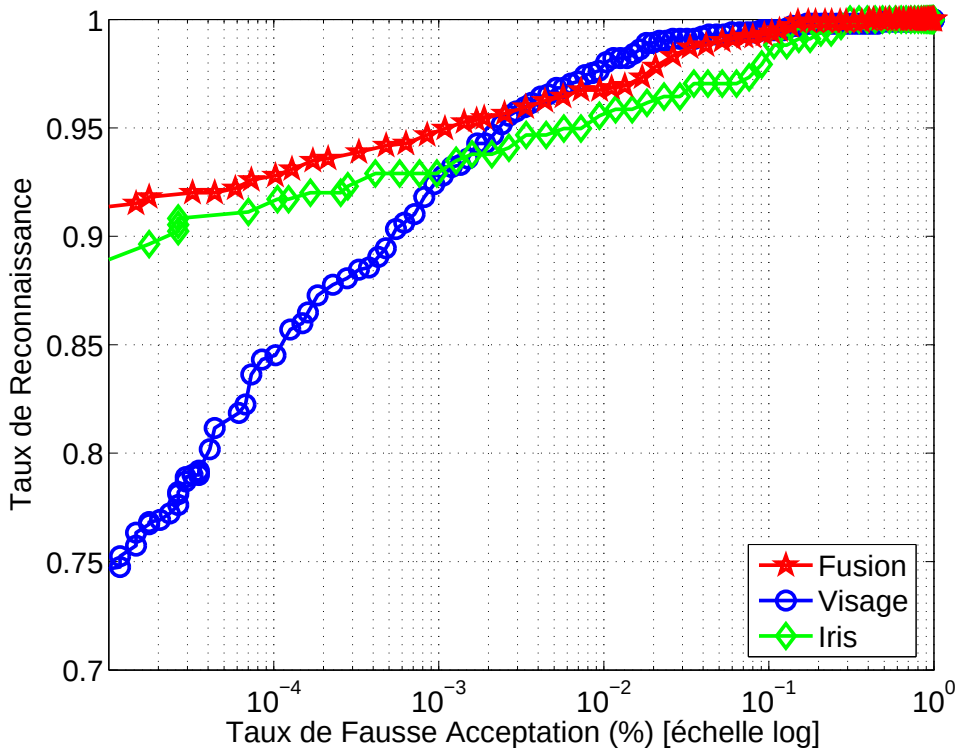


FIG. 8.27 – Courbes ROC pour la fusion QLQ-Min.

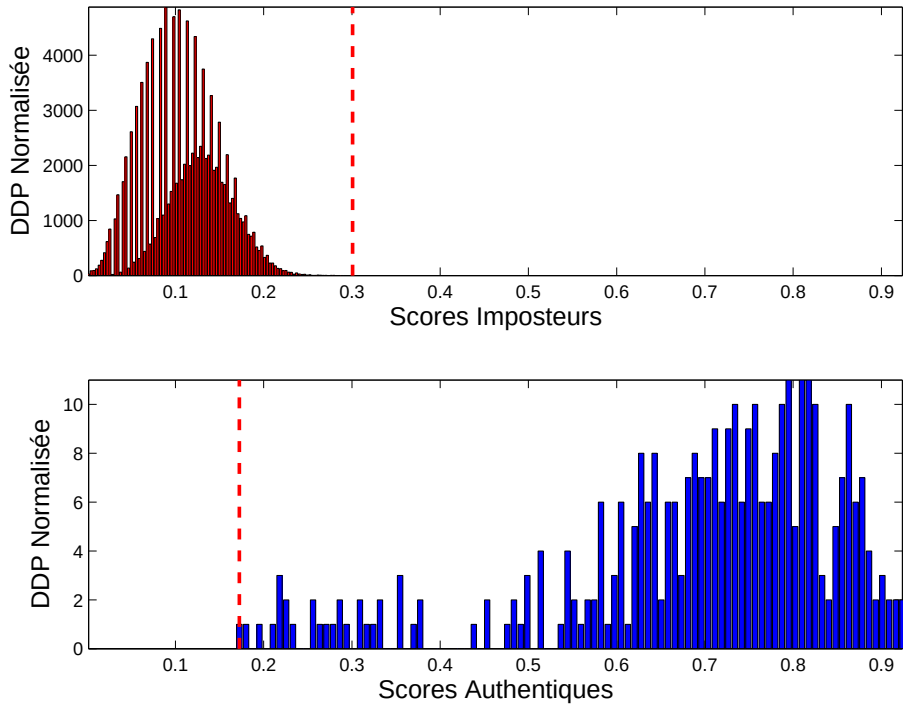


FIG. 8.28 – Densités de probabilité pour la fusion QLQ-Min.

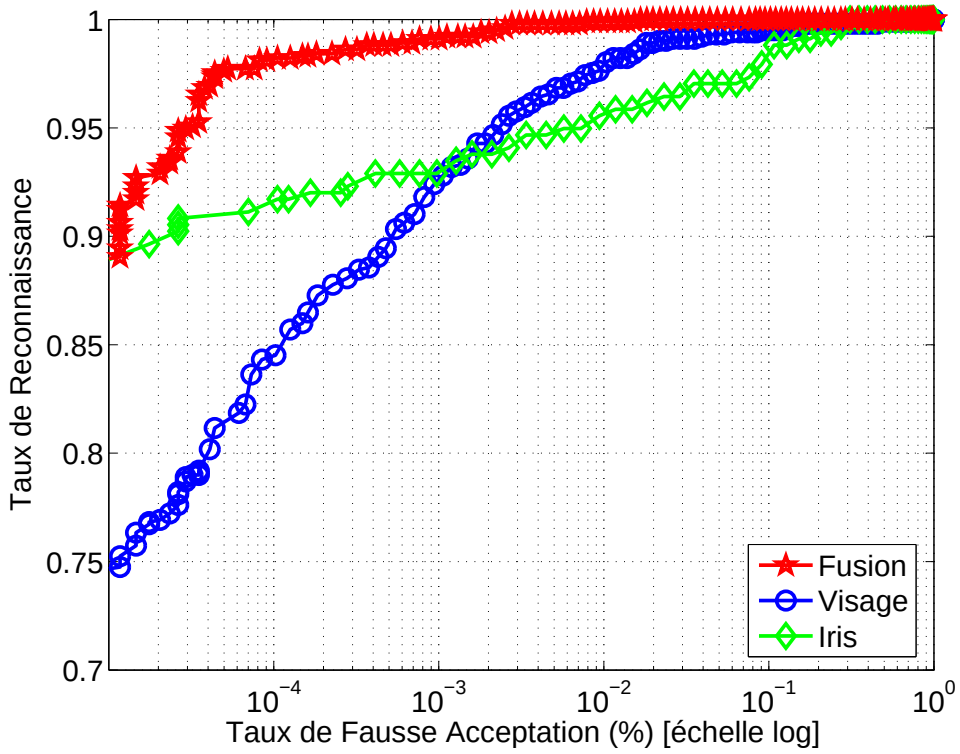


FIG. 8.29 – Courbes ROC pour la fusion QLQ-Max.

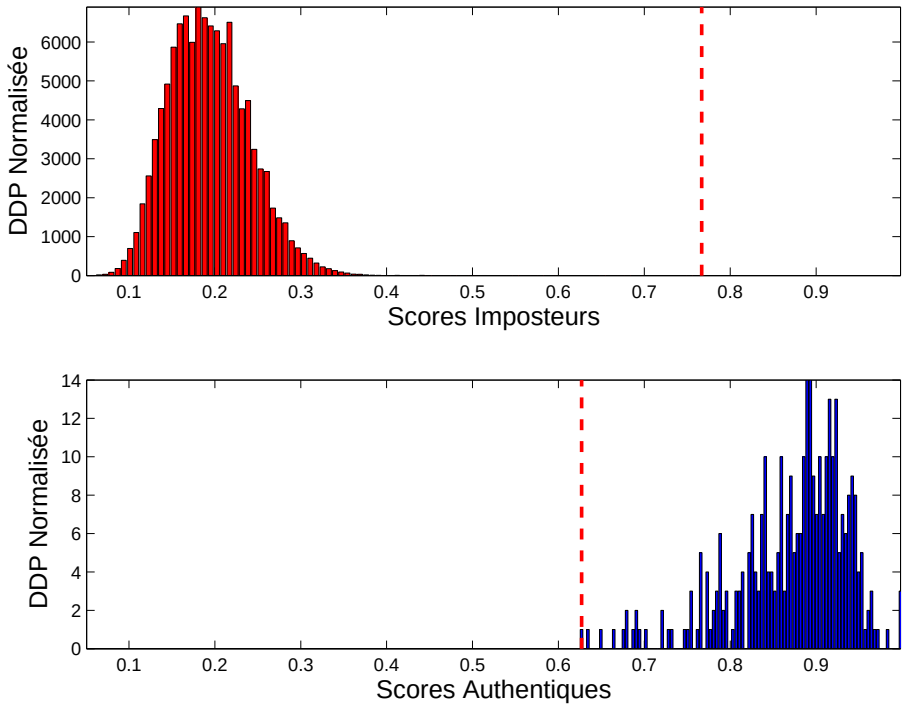


FIG. 8.30 – Densités de probabilité pour la fusion QLQ-Max.

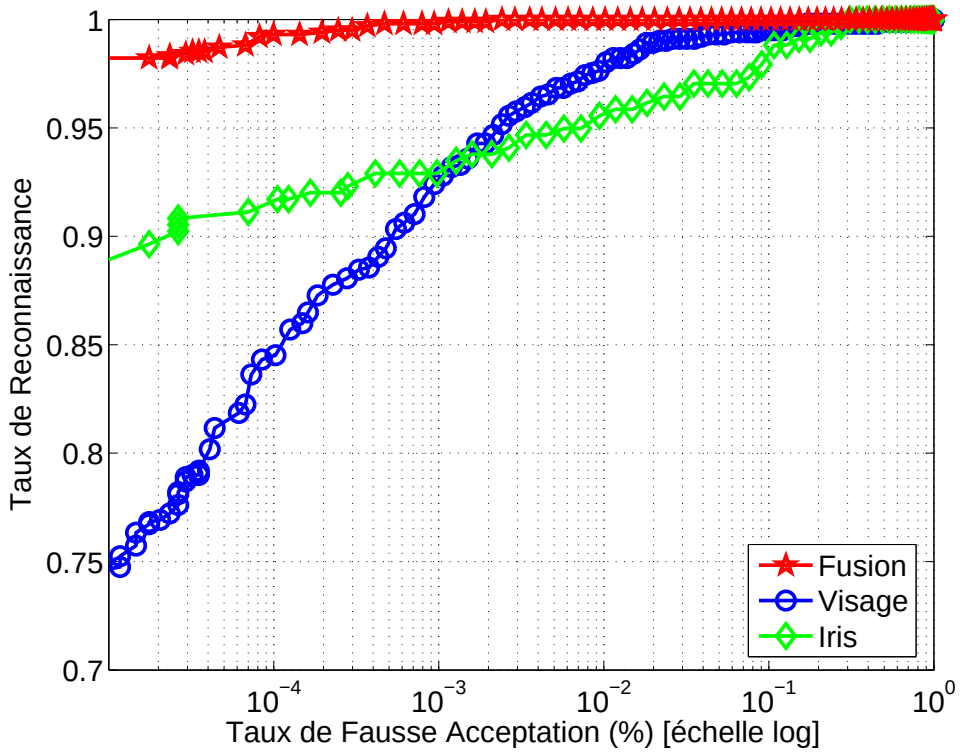


FIG. 8.31 – Courbes ROC pour la fusion QLQ-Somme.

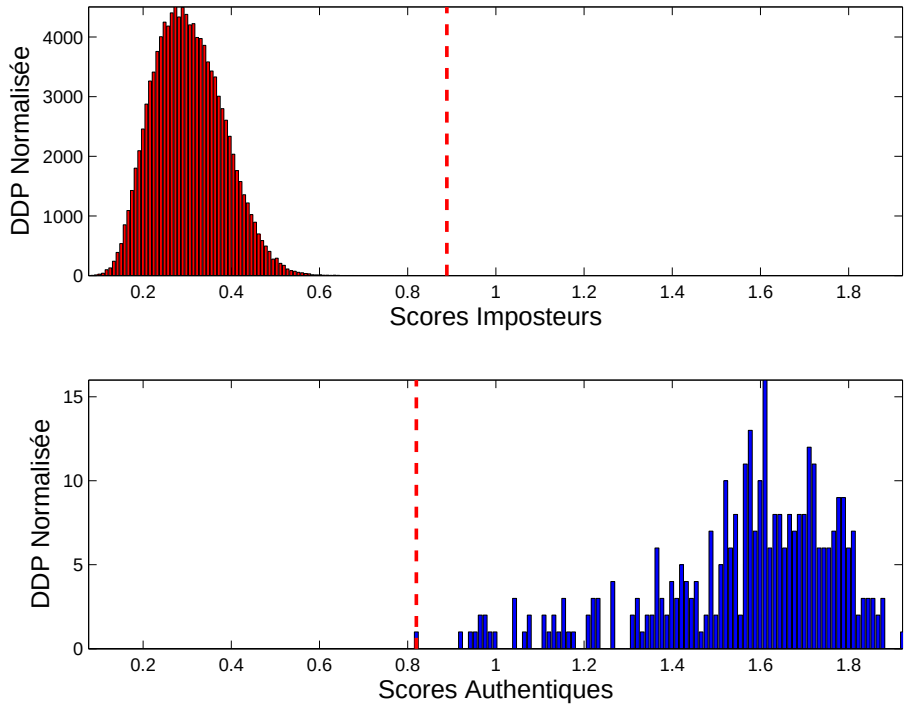


FIG. 8.32 – Densités de probabilité pour la fusion QLQ-Somme.

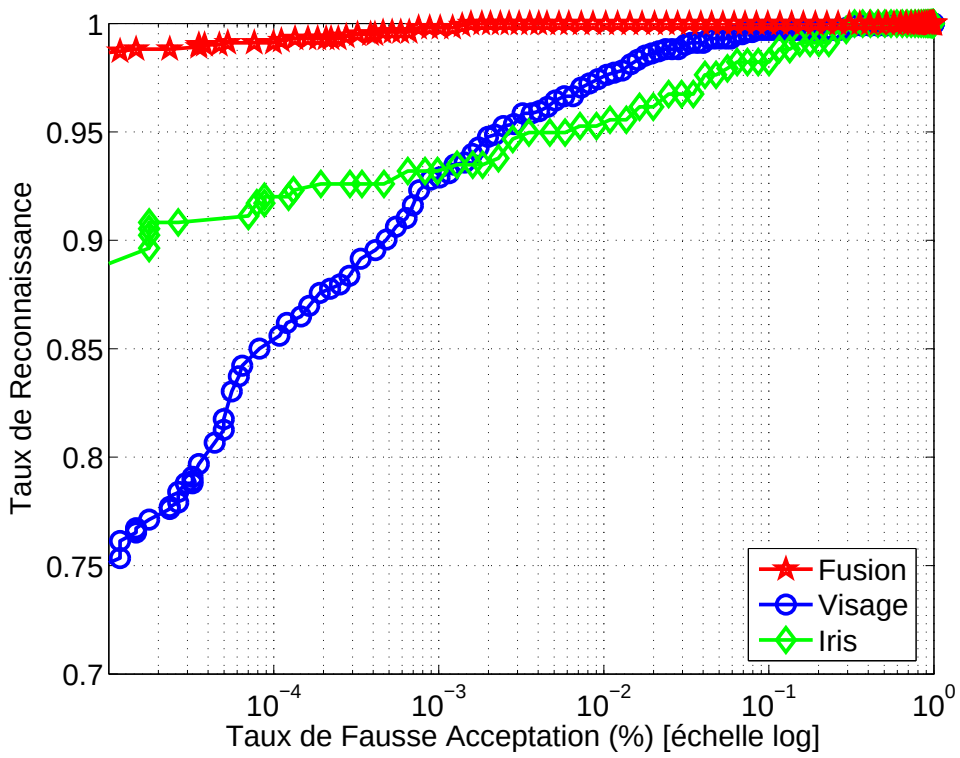


FIG. 8.33 – Courbes ROC pour la fusion QLQ-WSDSF.

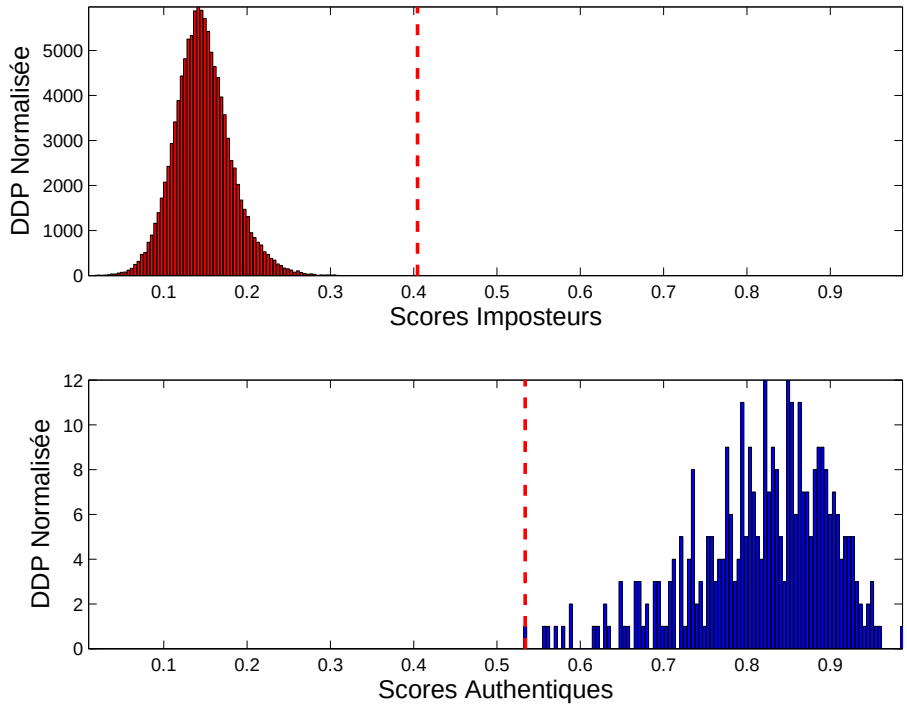


FIG. 8.34 – Densités de probabilité pour la fusion QLQ-WSDSF.

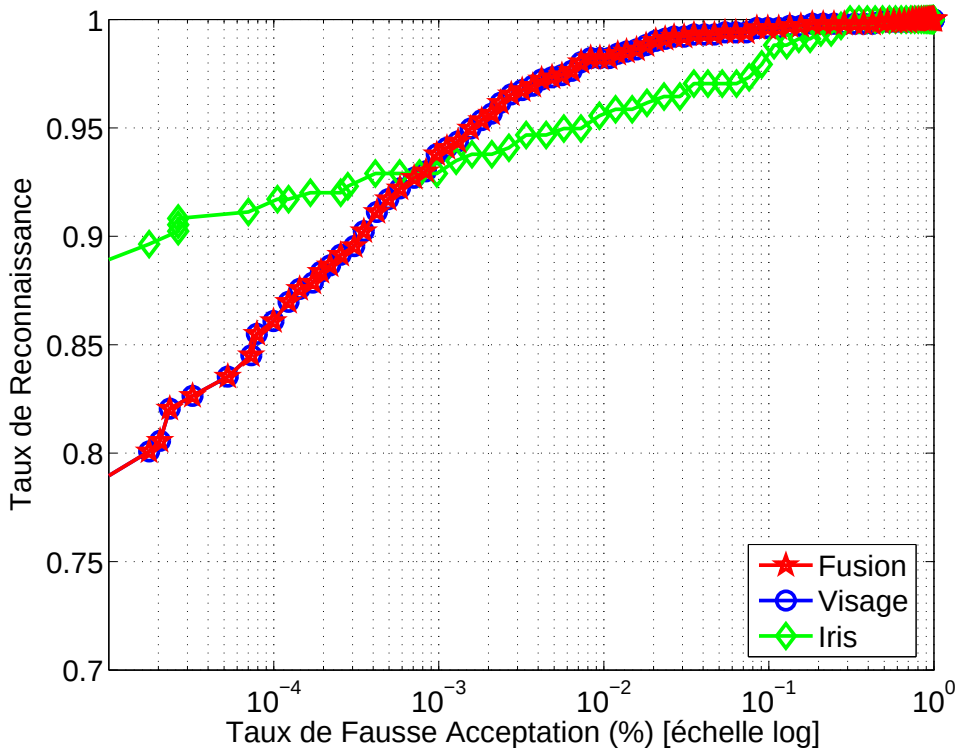


FIG. 8.35 – Courbes ROC pour la fusion TanH-Min.

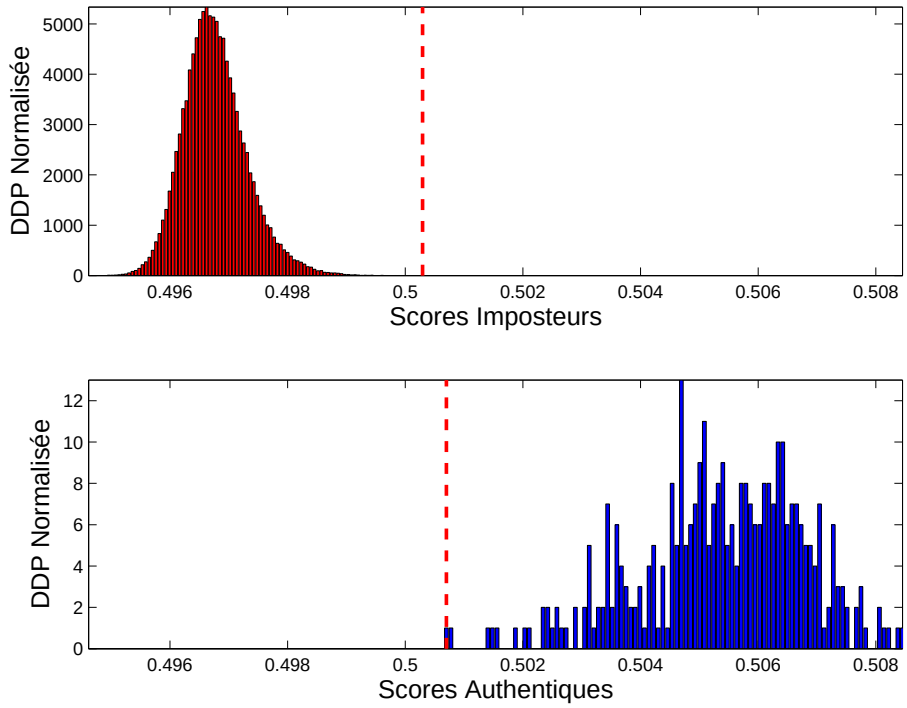


FIG. 8.36 – Densités de probabilité pour la fusion TanH-Min.

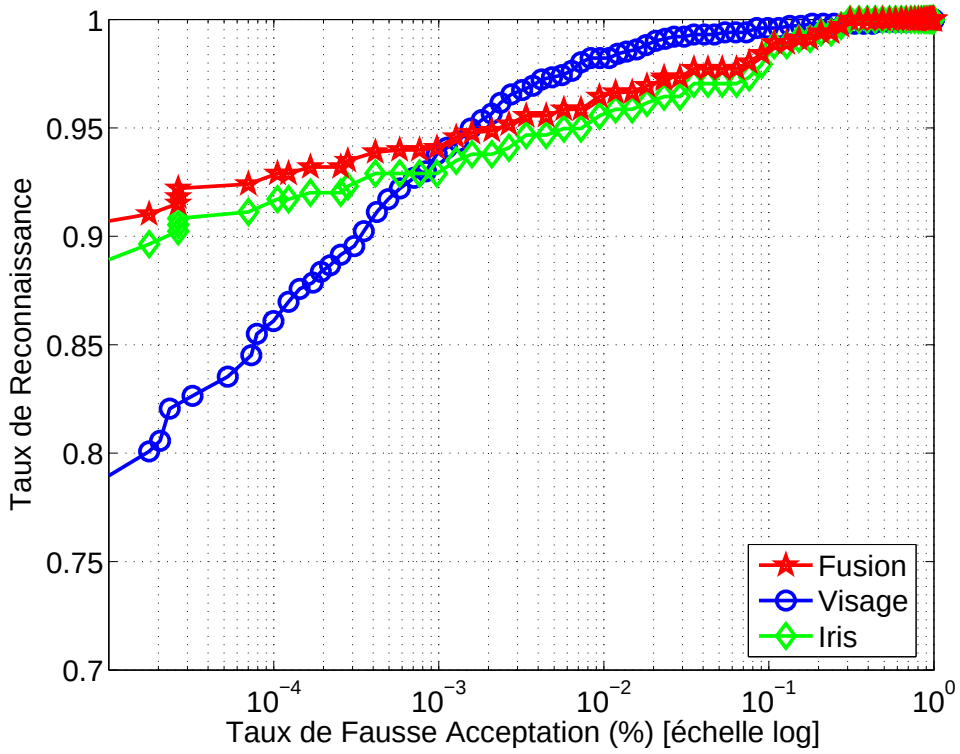


FIG. 8.37 – Courbes ROC pour la fusion TanH-Max.

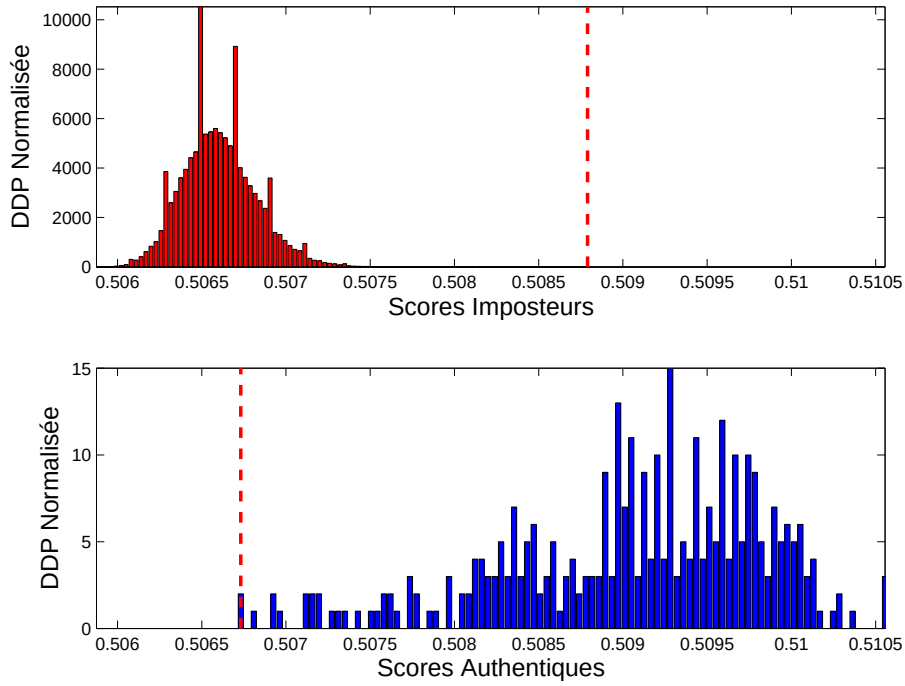


FIG. 8.38 – Densités de probabilité pour la fusion TanH-Max.

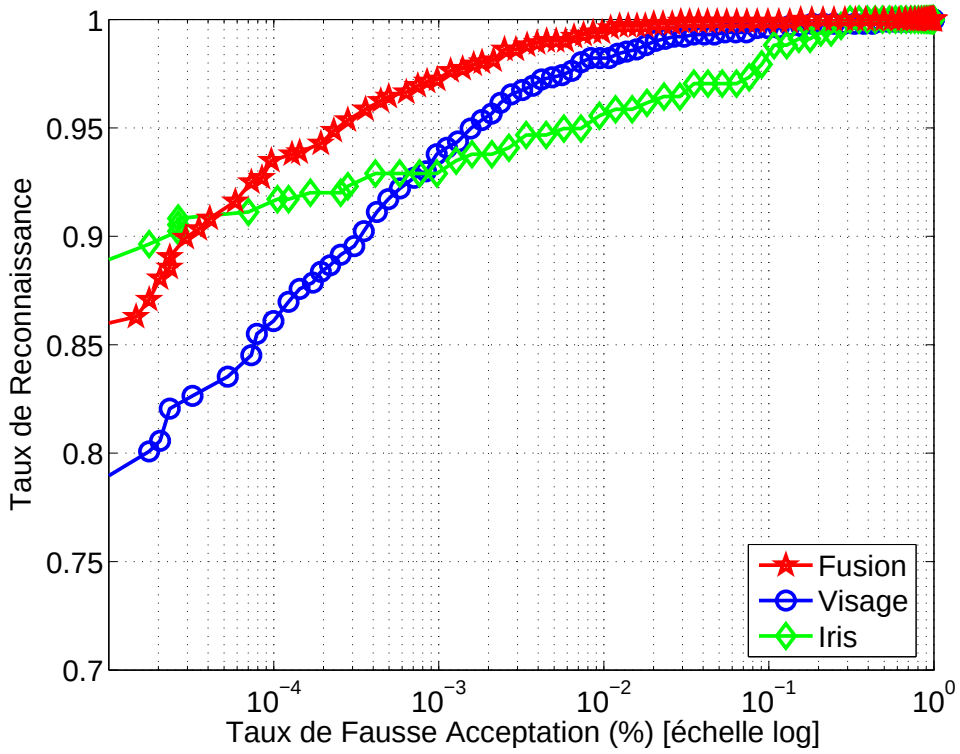


FIG. 8.39 – Courbes ROC pour la fusion TanH-Somme.

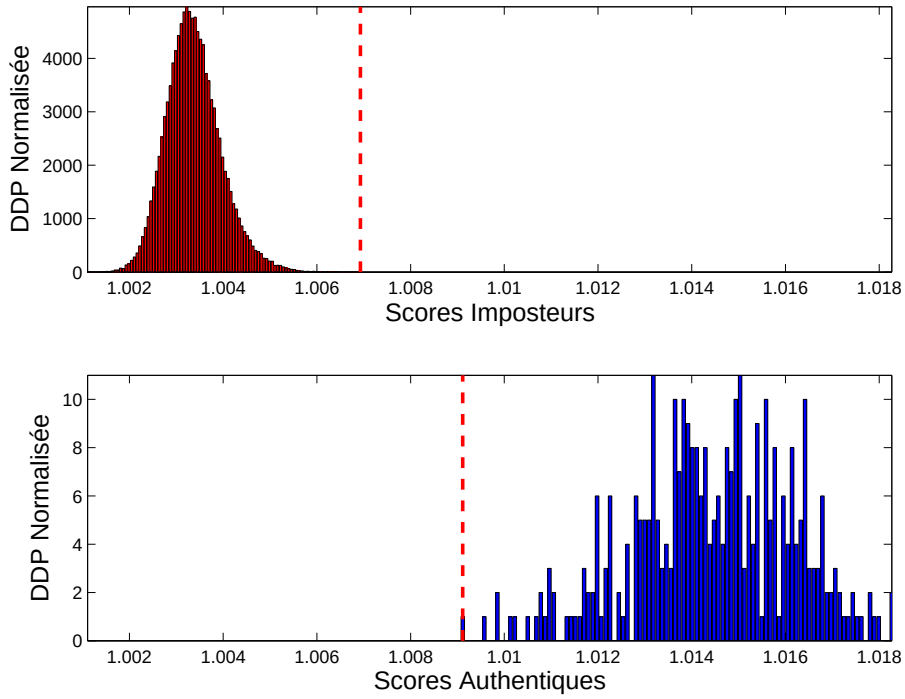


FIG. 8.40 – Densités de probabilité pour la fusion TanH-Somme.

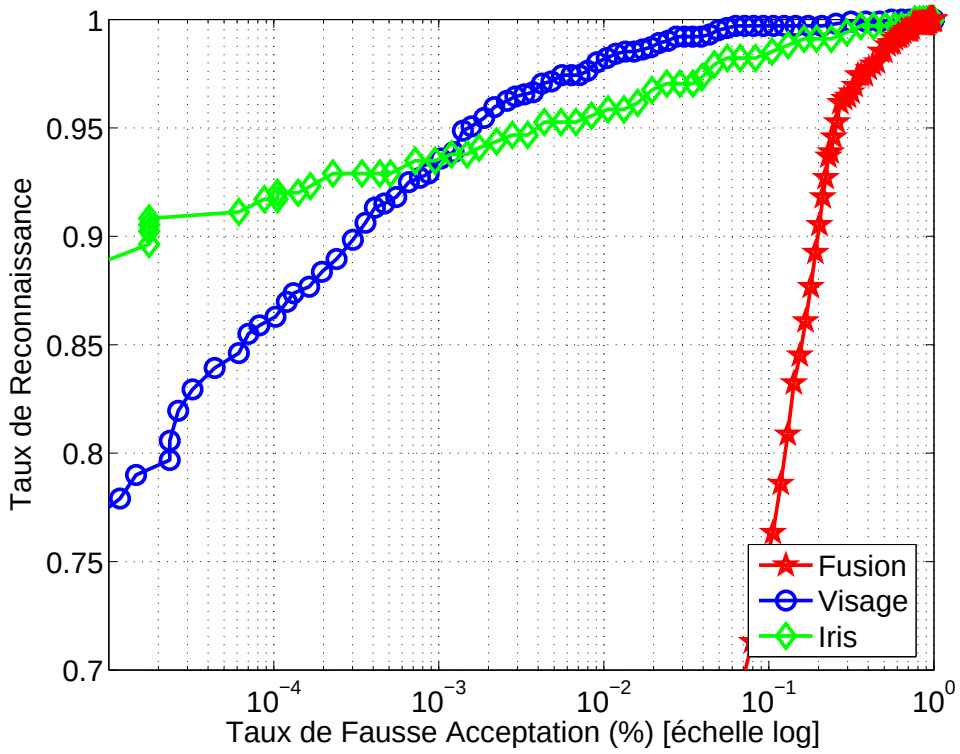


FIG. 8.41 – Courbes ROC pour la fusion TanH-WSDSF.

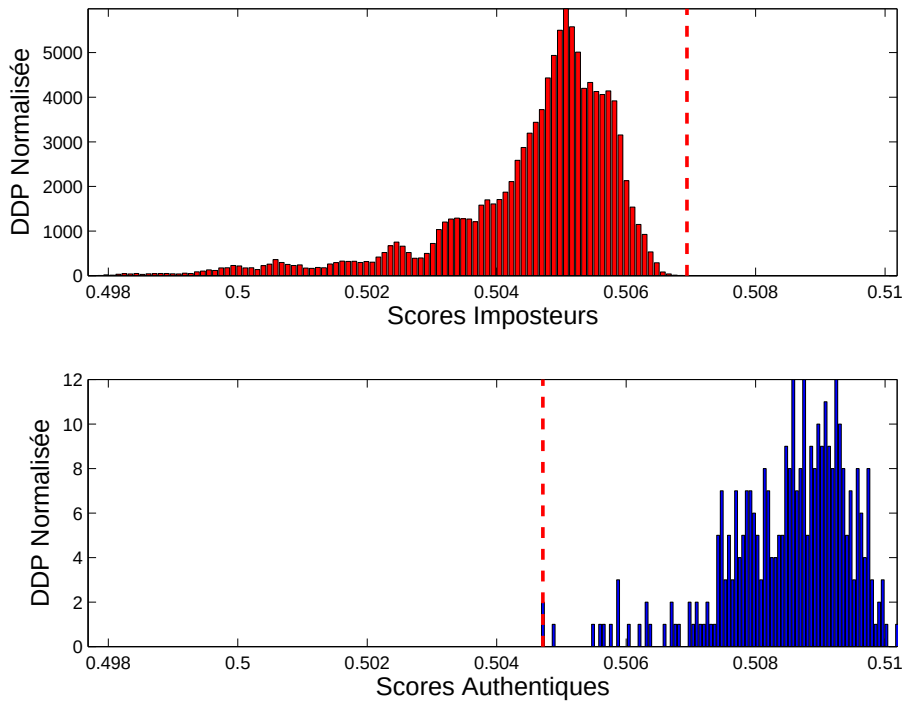


FIG. 8.42 – Densités de probabilité pour la fusion TanH-WSDSF.

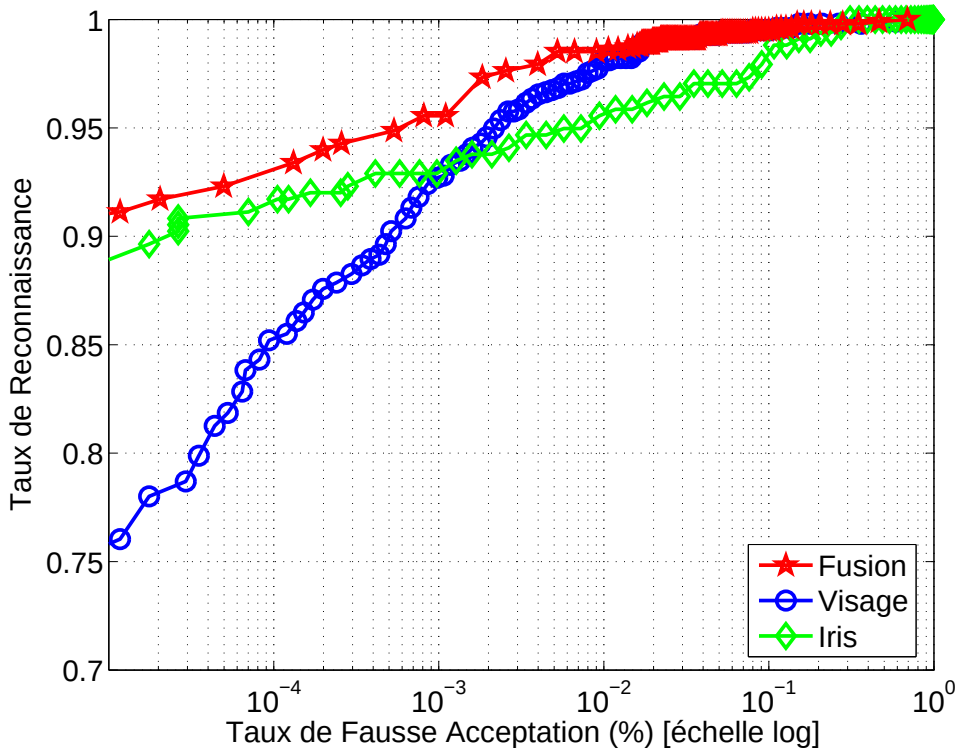


FIG. 8.43 – Courbes ROC pour la fusion Double Sigmoïde-Min.

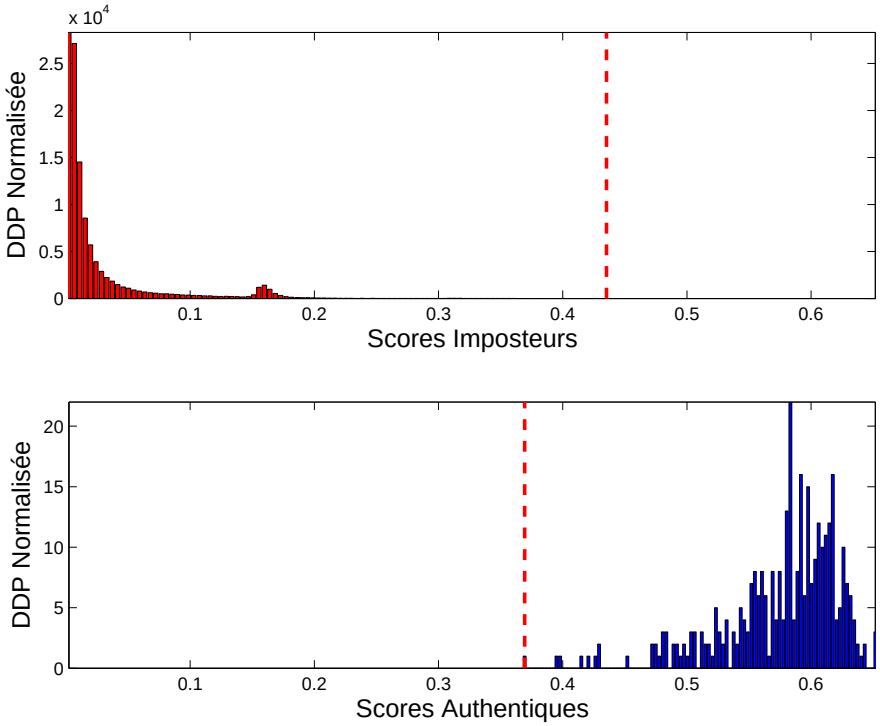


FIG. 8.44 – Densités de probabilité pour la fusion Double Sigmoïde-Min.

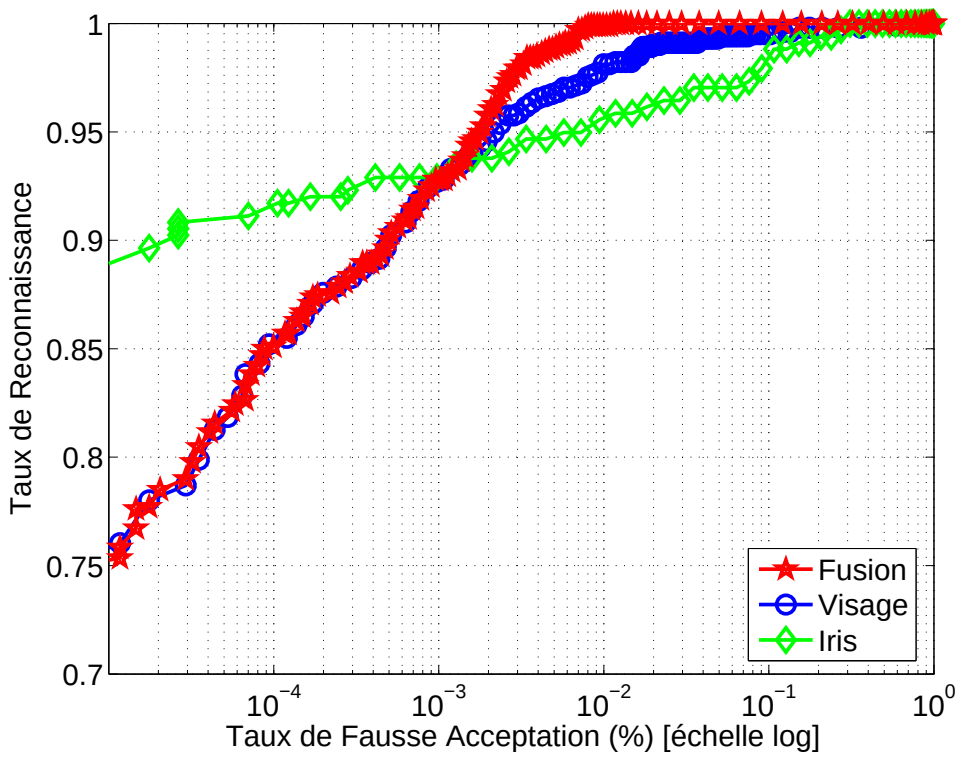


FIG. 8.45 – Courbes ROC pour la fusion Double Sigmoïde-Max.

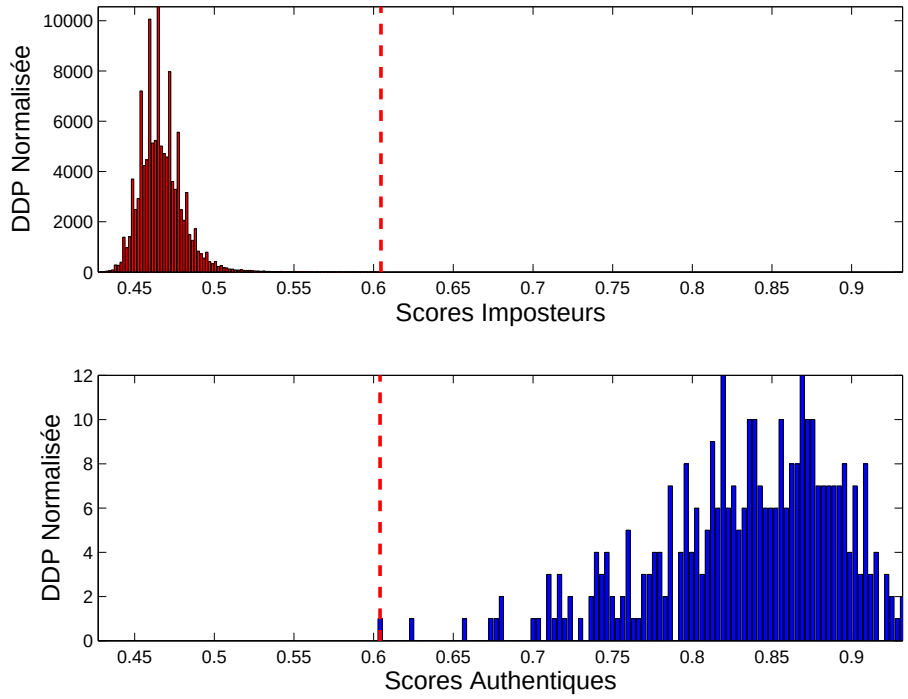


FIG. 8.46 – Densités de probabilité pour la fusion Double Sigmoïde-Max.

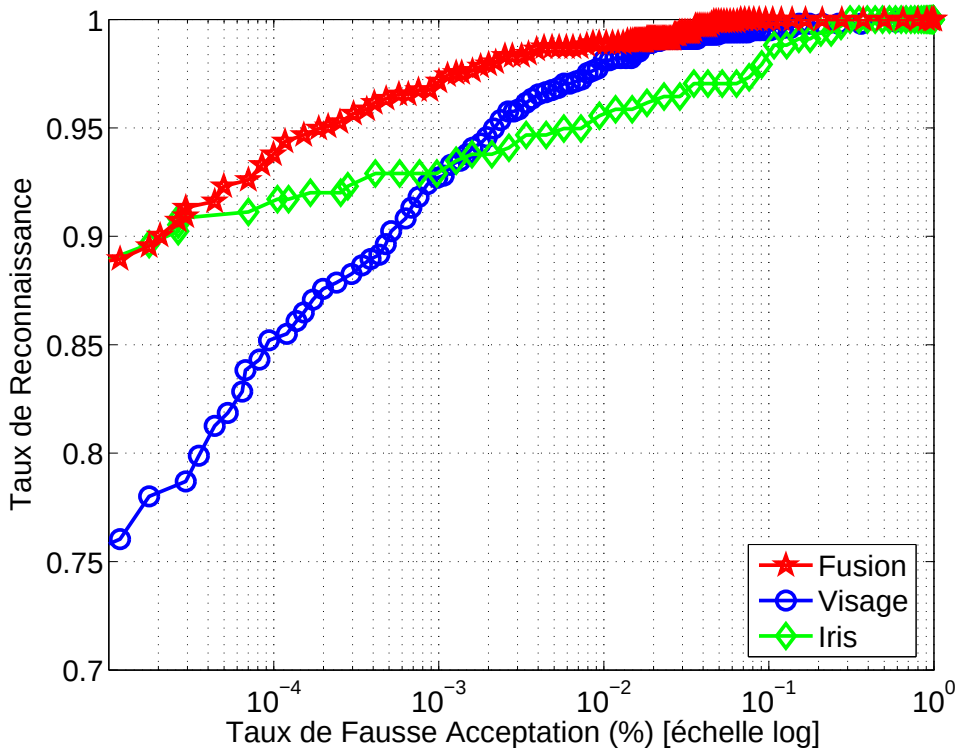


FIG. 8.47 – Courbes ROC pour la fusion Double Sigmoïde-Somme.

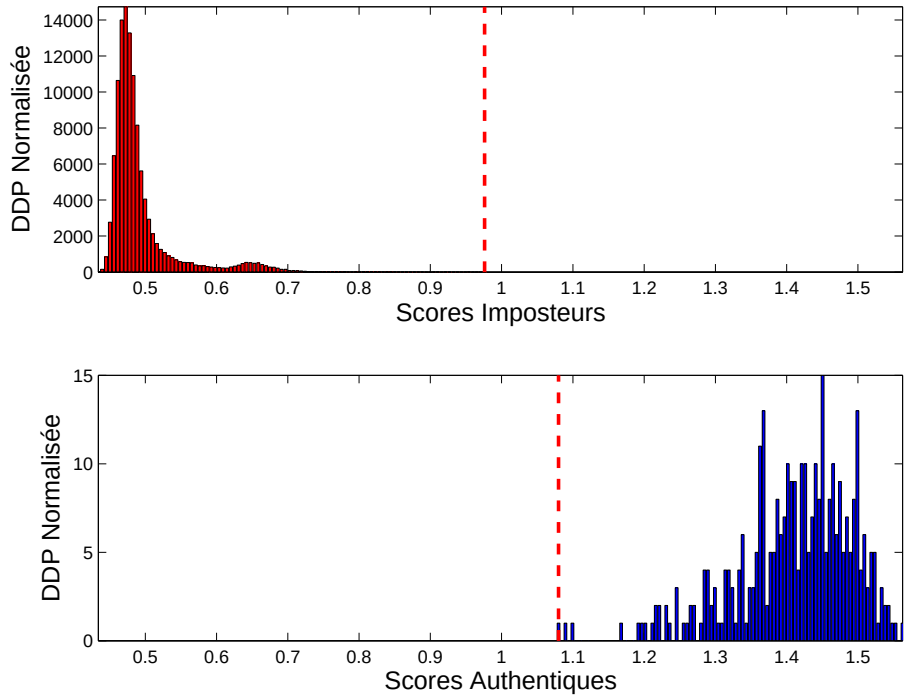


FIG. 8.48 – Densités de probabilité pour la fusion Double Sigmoïde-Somme.

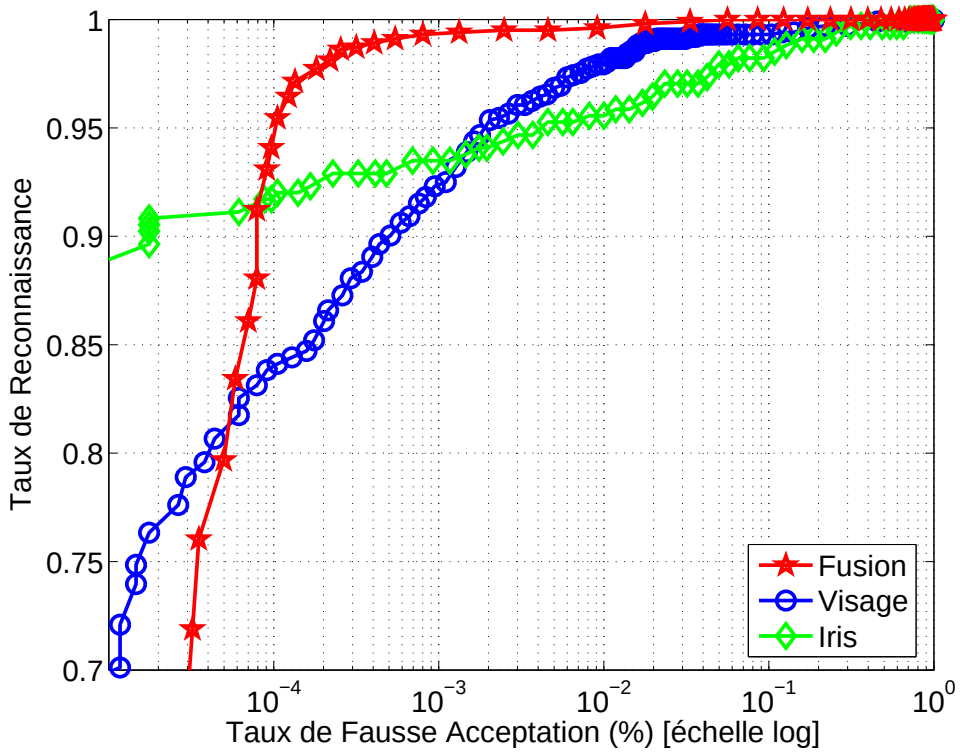


FIG. 8.49 – Courbes ROC pour la fusion Double Sigmoïde-WSDSF.

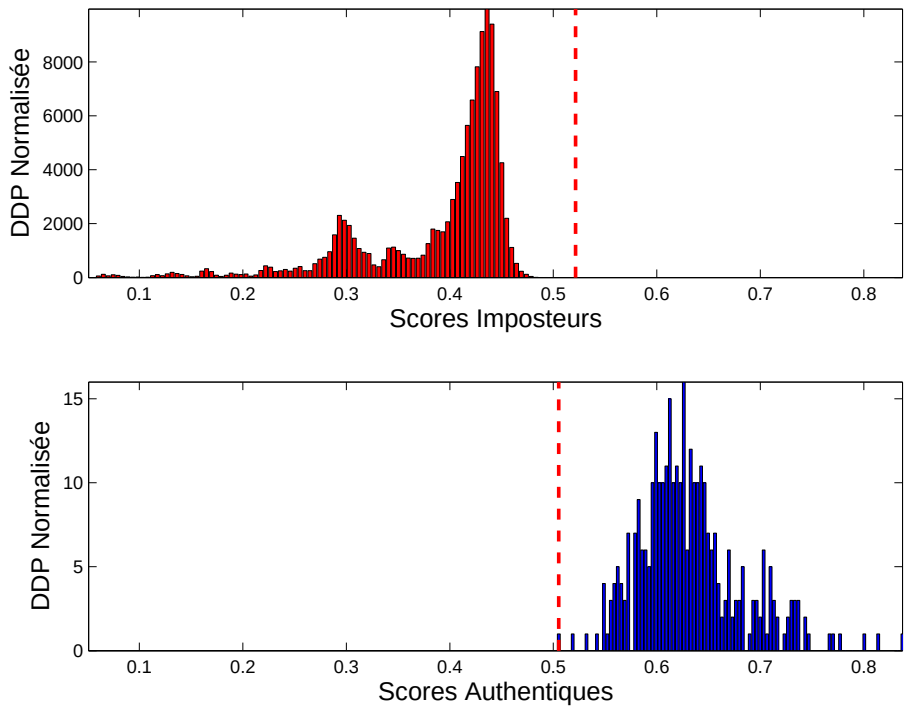


FIG. 8.50 – Densités de probabilité pour la fusion Double Sigmoïde-WSDSF.

Tout d'abord, précisons les paramètres de certaines méthodes de normalisation de score qui ont permis générer certains graphiques. Par analogie à l'étape des analyses statistiques (Chapitre 5), nous avons extrait, pour chaque modalité, une matrice aléatoire d'entraînement de taille 40% la taille de la matrice de similarité originale correspondante. Les scores sont ensuite bootstrappés et les paramètres de normalisation sont ensuite déduits des distributions de score (cf. Chapitre 2.7.3). Il est important de préciser que ces paramètres sont générés en amont de la fusion avec des données d'entraînement et qu'ils sont ensuite fixés une fois pour toutes lors des tests de fusion.

Pour la *normalisation QLQ*, nous obtenons pour les scores du visage $c = 0.46$ et $w = 0.32$, pour les scores d'iris $c = 0.29$ et $w = 0.19$.

Pour la *normalisation tanH*, les paramètres standards ont été utilisés pour les deux modalités, à savoir $a = 0.70$, $b = 0.80$ et $c = 0.95$. Bien entendu, les résultats correspondant à cette méthode de normalisation peuvent être légèrement améliorés si nous les choisissons plus convenablement.

Pour la *normalisation Double Sigmoïde*, nous avons pour les scores du visage $r_1 = 0.075$, $r_2 = 0.500$ et $t = 0.222$, pour les scores de l'iris $r_1 = 0.710$, $r_2 = 0.790$ et $t = 0.750$.

Cette première série de courbes doit permettre de comparer les performances de chacune des techniques de fusion en les comparant avec celles des modalité du visage et de l'iris prises individuellement. De manière générale, la règle de fusion *Min* donne d'assez mauvais résultats et possède un écart-type élevé ; de plus les courbes ROC de fusion correspondantes possèdent une vitesse de convergence lente et tendent souvent vers de celle de l'iris ou vers celle du visage. La règle de fusion *Max* donne de meilleurs résultats, sauf pour la normalisation *TanH* et *Double Sigmoïde*, mais possède encore des lacunes, notamment en ce qui concerne la vitesse de convergence et le recouvrement des distributions. En revanche, la règle de fusion *Somme* donne de bons résultats et commence à pouvoir séparer les distributions des scores imposteurs et authentiques. Par ailleurs, notre règle de fusion *WSDSF* permet d'effectuer une fusion véritablement efficace, en particulier lorsque la normalisation *Z-Score* est utilisée, où la séparation entre les distributions est la plus grande. Enfin, on notera deux choses intéressantes : d'une part, la normalisation *tanH* n'est pas du tout adaptée à notre technique pour la simple raison que cette normalisation dégrade beaucoup trop la forme des scores imposteurs qui ne suit plus du tout une distribution gaussienne ; d'autre part, la combinaison *DSig/WSDSF* offre une vitesse de convergence de la courbe ROC la plus rapide malgré un petit recouvrement des distributions.

En ce qui concerne les taux d'égale erreur (EER) moyens (Tab. 8.3), les résultats pour la règle de fusion *Min* confirme les observations des courbes. La règle de fusion *Max* donne globalement de meilleurs résultats mais le concurrent le plus sérieux pour notre méthode de fusion WSDSF est la règle *Somme*. Cette dernière possède déjà un EER moyen (sur toutes les méthodes de normalisation testées) de 0.315%. Cependant, la méthode de fusion **Z-Score/WSDSF** possède un EER moyen de 0.011, qui s'avère être le meilleur EER moyen, toutes combinaisons "méthode de normalisation/règle de fusion" confondues. Cela correspond à une amélioration de plus de 65% par rapport à la deuxième méthode *Z-Score/Somme* et une amélioration d'environ 68% par rapport aux méthodes *MinMax/Somme* et *QLQ/Somme*. Rappelons que ces trois dernières techniques sont considérées comme étant des approches par combinaison de score performantes [Ross 06].

Méthode de Normalisation	*Règle de Fusion*			
	Min	Max	Somme	**WSDSF**
MinMax	2.936 (0.20)	0.285 (0.00)	0.035 (0.00)	0.081 (0.10)
Z-Score	1.431 (0.50)	0.211 (0.20)	0.032 (0.00)	**0.011 (0.00)**
QLQ	2.066 (0.20)	0.286 (0.00)	0.035 (0.00)	0.099 (0.20)
TanH	1.427 (0.80)	2.680 (0.40)	0.392 (0.00)	3.947 (0.60)
Double-Sigmoïd	1.159 (0.80)	0.542 (0.30)	1.082 (0.90)	0.299 (0.20)

TAB. 8.3 – EER moyens (%) et écart-type (%)

En ce qui concerne la séparation globale (normalisée) des distributions de score de fusion (Tab. 8.4), elle est positive (donc une véritable séparation des distributions) pour seulement 6 des 20 permutations méthodes de normalisation/règles de fusion proposées ; les plus faibles correspondent aux techniques *TanH/Min* avec 0.029, *Dsig/Somme* avec 0.092 et *MM/WSDSF* avec 0.098, les plus fortes corespondent aux techniques *TanH/Somme* avec 0.127, *QLQ/WSDSF* avec 0.132 tandis que la plus grande séparation des distributions est obtenue pour la combinaison *Z-Score/WSDSF* avec 0.143.

Méthode de Normalisation	*Règle de Fusion*			
	Min	Max	Somme	**WSDSF**
MinMax	-0.137	-0.077	-0.011	0.098
Z-Score	-0.109	-0.135	-0.036	**0.143**
QLQ	-0.139	-0.148	-0.037	0.132
TanH	0.029	-0.440	0.127	-0.178
Double-Sigmoïd	-0.102	-0.001	0.092	-0.021

TAB. 8.4 – Séparation entre les distributions de scores imposteurs et authentiques.

Enfin, nous présentons les résultats des taux de reconnaissance pour un taux de fausse acceptation (FAR) de 0.001% qui constitue un point de mesure courant (Tab. 8.5).

Méthode de Normalisation	*Règle de Fusion*			
	Min	Max	Somme	**WSDSF**
MinMax	94.63	98.57	99.90	99.97
Z-Score	95.18	99.33	99.90	**100.00**
QLQ	94.85	99.10	99.83	99.80
TanH	93.85	94.13	97.28	40.81
Double-Sigmoïd	95.56	92.80	97.13	99.35

TAB. 8.5 – Taux de reconnaissance moyens (%) à un FAR=0.001(%).

D'après toutes les analyses précédentes, nous avons retenu les meilleures techniques de fusion afin de les comparer entre-elles. Les résultats sont illustrées sur la figure 8.4.2.

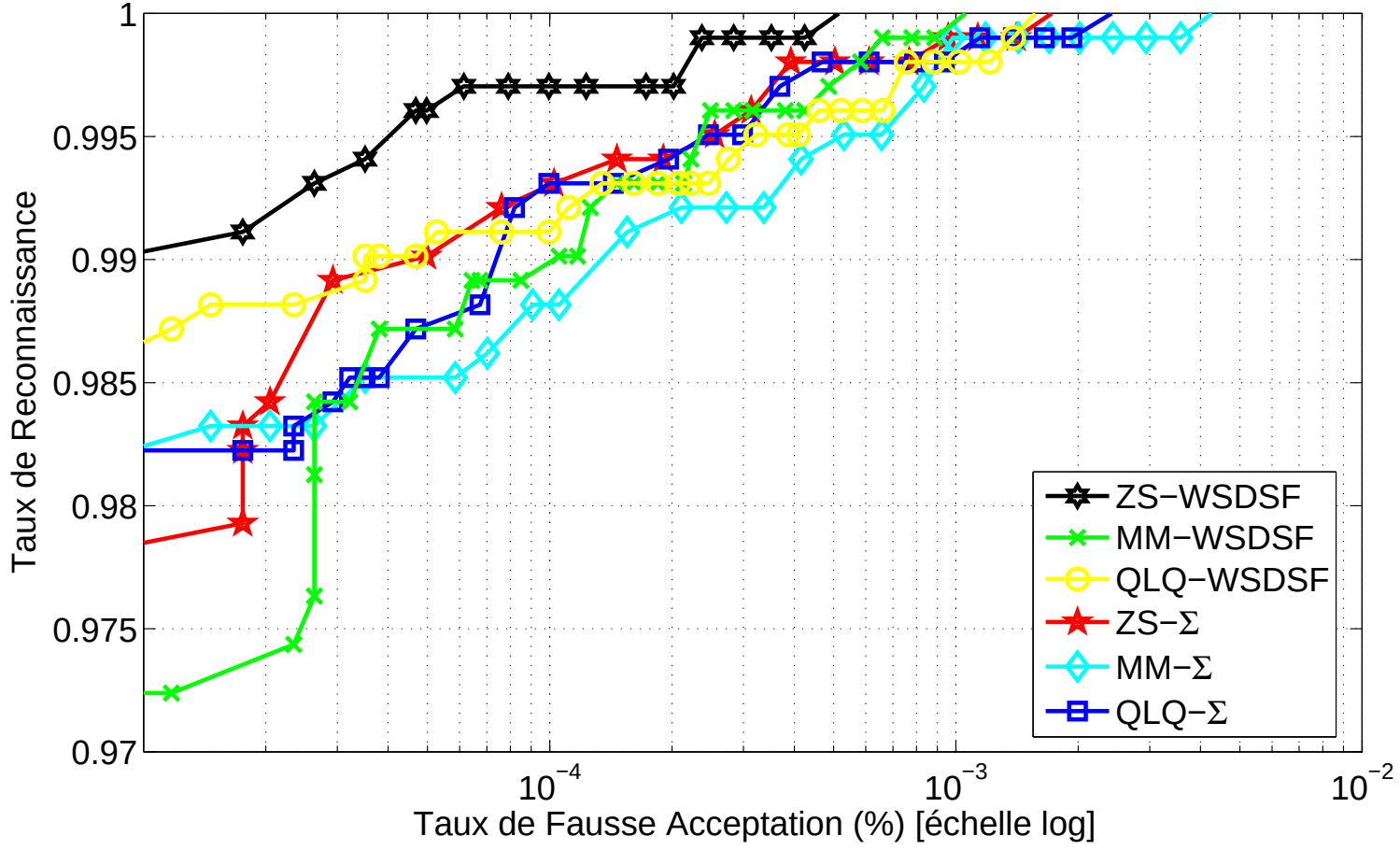


FIG. 8.51: Comparaison des Courbes ROC des Meilleures Techniques de Fusion Retenues.

Ce dernier graphique montre bien la haute performance de notre méthode de fusion *WSDSF*. Ce résultat est d'autant plus intéressant que cette technique est la seule à atteindre 100% de taux de reconnaissance pour un $FAR = 7.10^{-4}\%$, tandis que les autres techniques ne peuvent l'atteindre qu'à partir d'un $FAR = 10^{-3}\%$. Un excellent taux de reconnaissance à un *FAR* très faible est très important car il précise que l'on garantie une très grande sécurité au système biométrique, en ne faisant que très peu d'erreurs concernant l'acceptation d'individus qui ne devraient pas l'être.

En conclusion de cette série de résultats, un tableau de synthèse (Fig.8.52) permet de comparer notre méthode de fusion multimodale WSDSF à tous les travaux connus en fusion multimodale visage/iris depuis six ans.

Date	Auteurs	Niveau de fusion	Stratégies	Utilisateurs testés	Bases de données	Meilleurs résultats
2003	*Wang et al.*	Score	Σ, WΣ, FLD, NN	90	ORL - MIT - Yale - NLPR	*(NN)* EER=0.24%
2005	*Chen et Chu*	Caractéristiques	WPNN	100	IIS - ORL / interne	EER=0.33%
2006	*Son et Lee*	Caractéristiques	RJFV DLDA	400	IIS - ORL / interne	GAR=99.98% *(FAR=N/A, EER= N/A)*
2007	*Zhang et al.*	Score	MinMax, Σ, $\prod$	112	interne	*(MM, Σ)* GAR=99.75% *(FAR=10^{-3}%)*
2008	*Rattani et Tistarelli*	Caractéristiques	Concaténation SIFT	513	Equinox / CASIA	EER=0.050%
2009	Morizet	Score	WSDSF	1014	FERET / CASIA	EER=0.011% GAR=100% @FAR=7.10^{-4}%

FIG. 8.52 – Tableau de synthèse des résultats obtenus en fusion multimodale visage/iris de 2003 à 2009, incluant notre méthode de fusion multimodale WSDSF.

Nous pouvons remarquer que notre méthode de fusion multimodale WSDSF, opérant au niveau score, a été testée sur le plus grand nombre d'utilisateurs bimodaux possibles (1014) provenant de bases de données officielles que sont *FERET* et *CASIA*. Ce nombre d'utilisateurs, qui est un paramètre très important pour juger de la robustesse d'un système, est deux fois plus élevé que celui issu des derniers travaux de *Rattani et Tistarelli* (513), ces dernier utilisant une base de données d'images de visage (*Equinox*) non officielle. De plus, notre taux d'égale-erreur est environ cinq fois meilleur que celui obtenu en 2008, pour un taux de reconnaissance parfait à un taux de faux-rejet très faible ($FAR = 7.10^{-4}\%$), en deçà du taux habituellement utilisé de $10^{-3}\%$. Enfin, non seulement la plupart des travaux concurrents ont utilisé au moins une base de données interne (ce qui ne devrait plus être retenu pour annoncer des résultats futurs) mais surtout, il est rare que les résultats soient présentés de manière aussi complète que nous l'avons fait : uniquement les taux d'égale erreur (2003, 2005 et 2008) ou les taux de reconnaissance (2006 et 2007).

8.4.3 Situations délicates

Souvent, le but n'est pas tant de baisser de manière importante des scores imposteurs déjà faibles et qui ne présentent donc qu'un faible risque d'erreur de décision (bien que cela améliore significativement l'EER global du système), mais bien de mieux séparer des scores proches. Nous allons donc illustrer, quelques cas où la prise de décision semble plus délicate et montrer comment notre méthode de fusion *WSDSF* est capable de gérer ce genre de situations.

Dans les tests que nous avons pratiqués, les scores imposteurs des deux modalités ont été générés à partir de deux gaussiennes, d'après les paramètres du modèle (Tab. 5.1). Nous pouvons ainsi créer des vecteurs de scores de taille quelconque (ici, 1000). Les vecteurs de scores fusionnés seront représentés pour les méthodes de fusion "Somme" et *WSDSF*. Les scores auront été transposés dans $[0;1]$ grâce à la normalisation *MinMax*, afin de pouvoir comparer les méthodes de fusion.

Nous pouvons alors identifier *trois scénarios*, ordonnés par difficulté de prise de décision croissante :

1. Cas #1 : Un des scores authentiques d'une modalité est relativement faible, proche des scores imposteurs les plus élevés,

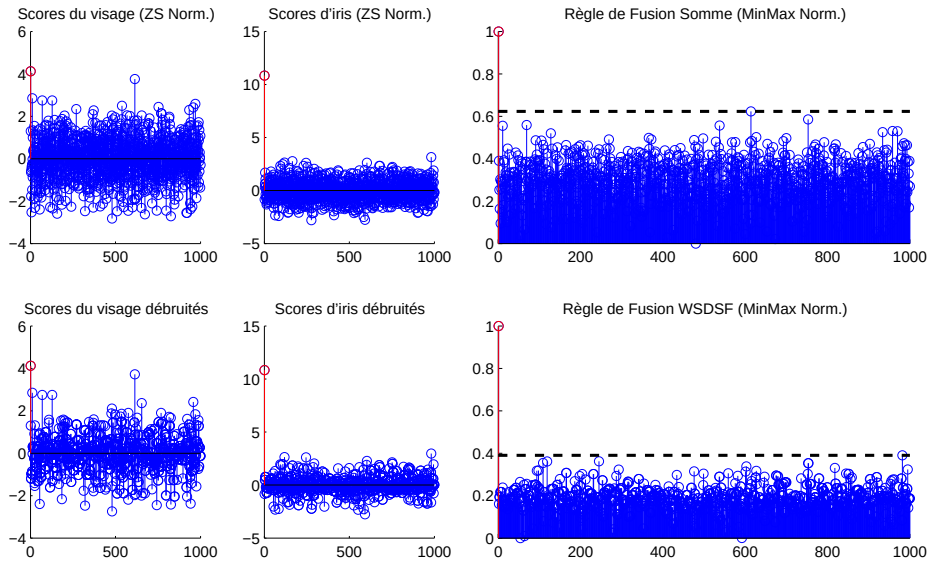


FIG. 8.53 – Situation délicate : cas 1.

La situation de la figure 8.53 montre l'utilisateur $\#1$ dont le score authentique du visage pose problème puisqu'il est très proche des scores imposteurs (le score authentique d'iris ne posant pas de problème par ailleurs). La règle de fusion "Somme" donne un bon résultat de reconnaissance mais montre que le maximum des scores imposteurs de fusion est à 0.6 alors que la règle de fusion WSDSF fournit un maximum des scores imposteurs de fusion à 0.4, soit une amélioration d'environ 33%.

2. Cas #2 : Un des scores authentique d'une modalité n'est pas le maximum, un score imposteur est plus élevé,

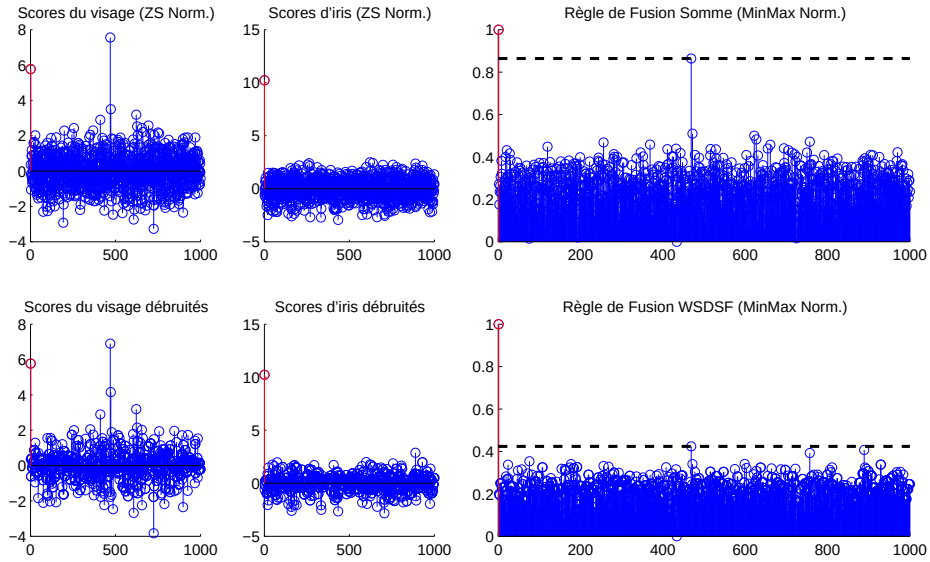


FIG. 8.54 – Situation délicate : cas 2.

Ce scénario (Fig. 8.54), pour l'utilisateur $\#1$, est un peu plus compliqué à gérer car une des deux modalités (ici, le visage) donne un mauvais résultat de reconnaissance. En effet, le score authentique du visage (0.6) est plus faible de 25% par rapport au maximum des scores imposteurs (0.8). La règle de fusion "Somme" fournit un bon résultat mais montre ses limites : le maximum des scores imposteurs de fusion (0.85) se rapproche dangereusement du score authentique de fusion tandis que la règle de fusion WSDSF contrôle mieux la situation en laissant le maximum des scores imposteurs de fusion à 0.40, soit une amélioration d'environ 53%.

Le cas 2 bis (Fig. 8.55) présente une difficulté encore plus grande dans la prise de décision car le score authentique d'iris de l'utilisateur $\#1$ (étant égal à -13) est très nettement inférieur aux scores imposteurs, le score authentique du visage ne posant pas de problème par ailleurs. Cette fois-ci, la règle de fusion "Somme" ne parvient pas à donner un résultat de reconnaissance valide puisque le score authentique de fusion est inférieur à une vingtaine de scores imposteurs de fusion. La règle de fusion WSDSF donne un bon résultat de reconnaissance, avec un maximum des scores imposteurs égal à 0.62.

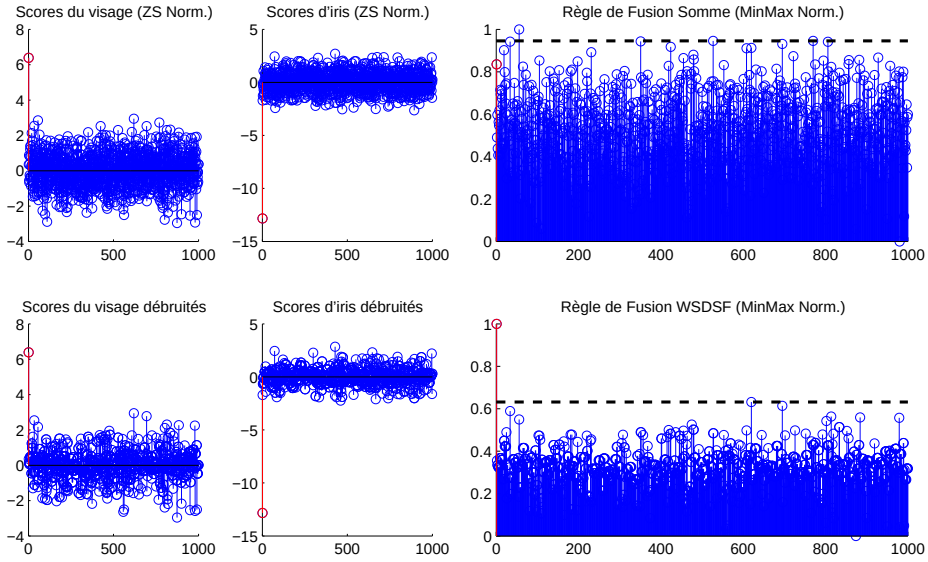


FIG. 8.55 – Situation délicate : cas 2 bis.

3. Cas #3 : Les scores authentiques des deux modalités ne sont pas les maximums, cela veut dire que les systèmes biométriques unimodaux du visage et de l'iris feraient une erreur sans l'aide de la fusion.

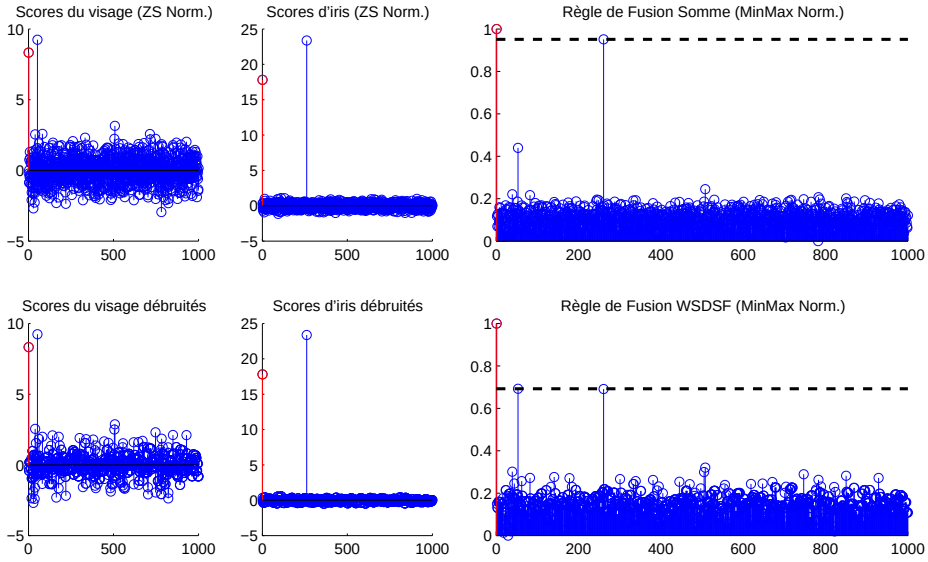


FIG. 8.56 – Situation délicate : cas 3.

La complexité augmente encore avec le cas de la figure 8.56. Ici, les scores authentique du visage et de l'iris de l'utilisateur #1 ne sont pas les maximums. Cette situation, peu courante, signifie tout simplement que les système unimodaux (visage

et iris) ne peuvent pas fournir un résultat de reconnaissance valide et que la fusion devient nécessaire. La règle de fusion "Somme" permet tout juste de donner un bon résultat de reconnaissance avec un maximum des scores imposteurs de fusion égal à 0.95, là où la règle de fusion WSDSF laisse le maximum des scores imposteurs de fusion à 0.70, soit une amélioration de 26.3%.

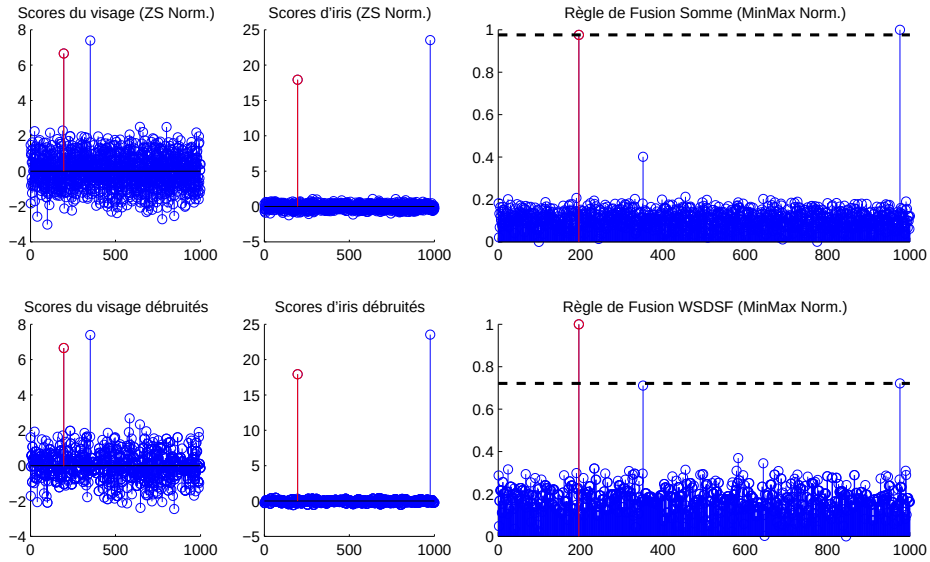


FIG. 8.57 – Situation délicate : cas 3 bis.

La cas de la figure 8.57 montre des scores authentiques de l'utilisateur $\#200$ dans des conditions similaires à celles de la figure 8.56. Cette fois-ci, la règle de fusion "Somme" se trompe dans le résultat final de reconnaissance. En revanche, la règle de fusion WSDSF ne laisse aucun doute quant à la prise de décision finale et donne un bon résultat de reconnaissance puisque le score authentique de fusion est nettement au dessus du maximum des scores imposteurs (0.72).

8.5 Conclusion

En suivant un protocole de test et d'évaluation multimodale très précis, nous avons d'abord exploré une méthode de fusion inspirée de concepts de traitement d'image, en particulier la fusion d'image par ondelettes. Cette méthode a donné des résultats prometteurs et a permis de nous orienter vers une méthode de fusion unidimensionnelle, plus rapide et plus facile à formaliser théoriquement. Cette technique de fusion, que nous avons appelée *WSDSF*, a montré à travers une étude comparative complète, qu'elle pouvait fournir d'excellents résultats en termes de taux d'égale erreur (EER), de taux de reconnaissance et de séparation globale des distributions des imposteurs et authentiques. Notons que pour pouvoir appliquer cette technique, il est important de faire des analyses statistiques préliminaires afin de vérifier l'allure des distributions des scores imposteurs pour une modalité donnée, qui doit relativement bien suivre une distribution gaussienne, ce qui est très fréquent en général [Bail 03]. Si cette condition est remplie, notre méthode

WSDSF peut parfaitement prendre en compte plusieurs modalités afin de tenter d'augmenter encore plus la sécurité d'un système biométrique multimodal. Enfin, quelques simulations ont montré que notre méthode *WSDSF* est capable de gérer des situations délicates, en particulier lorsque les systèmes unimodaux ne permettent pas d'effectuer une bonne reconnaissance, justifiant ainsi la nécessité de fusionner plusieurs modalités biométriques.

Conclusion Générale

Après avoir introduit les concepts généraux en biométrie, nous avons présenté un état de l'art en reconnaissance faciale, en reconnaissance de l'iris et en fusion multimodale où nous avons détaillé les différents niveaux de fusion et de nombreuses techniques de fusion possibles dans un système biométrique multimodal. Nous avons également montré les liens qui peuvent exister entre le fonctionnement du cerveau pour identifier et reconnaître des personnes et les algorithmes fondamentaux en reconnaissance faciale. Ensuite, nous avons présenté les systèmes d'acquisition d'images de visage et d'iris ainsi que les bases de données officielles qui ont servi à pratiquer nos tests de fusion, tout en proposant deux architectures possibles pour la création d'une base de données multimodale. Des analyses statistiques à grande échelle, utilisant la technique du bootstrap, ont permis de modéliser mathématiquement les scores imposteurs des deux modalités visage et iris. Elles ont montré que nous pouvions assimiler ces scores imposteurs à du bruit blanc gaussien, sous réserve de l'application d'une normalisation Z-Score sur les scores de similarité. Ces analyses ont constitué un résultat préliminaire important pour posséder les prérequis nécessaires à la nouvelle méthode de fusion qui a été mise en place. Par la suite, nous avons introduit les ondelettes et détaillé les solutions qu'elles apportent à divers niveaux de notre chaîne biométrique multimodale. En effet, les ondelettes permettent de compresser les images de visages normalisées avant d'utiliser l'algorithme PCA, l'avantage principal est que nous pouvons traiter un plus grand volume de données, plus rapidement et avec des taux de reconnaissance faciale très stables, voire améliorés, par rapport aux taux de reconnaissance obtenus avec des images non compressées par ondelettes (tests FERET). En ce qui concerne l'aspect analyse, les ondelettes ne sont pas adaptées pour analyser les textures, c'est pourquoi nous avons appliqué une transformée en paquets d'ondelettes aux images déroulées d'iris ; cette transformée a permis de construire les signatures biométriques de l'iris, codées sur 1.5 bits (-1, 0 ou 1). De plus, nous avons vu que les ondelettes offrent une solution élégante au problème de débruitage de données, elles ont donc été utilisées en prétraitement des scores de similarité avant d'appliquer notre méthode originale de fusion. Cette méthode apporte une nouvelle approche au domaine de fusion par combinaison de scores, elle agit de manière dynamique en fonction de chaque utilisateur et maximise une fonction de coût combinant trois paramètres : une distance de séparation entre la distribution des scores imposteurs et le score authentique de chaque utilisateur ainsi que deux moments statistiques standards (les coefficients de dissymétrie et d'aplatissement) calculés à partir des distributions des scores imposteurs de chaque utilisateur. Enfin, nous avons présenté les expérimentations et les résultats de fusion multimodale à partir d'un protocole précis permettant de travailler au niveau score en utilisant des matrices de similarité.

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 des individus, nous obligeant à travailler dans un **contexte de très grande diversité**. Cette diversité se retrouve également dans le nombre considérable d'algorithmes qui ont été proposés en reconnaissance faciale et de l'iris.

Les systèmes biométriques unimodaux existent déjà depuis quelques années mais ils sont plutôt adaptés à un niveau de sécurité moyen. En fait, **plus le niveau de sécurité est élevé, plus on tendra vers l'utilisation de systèmes multimodaux, plus performants et plus sûrs**.

Cependant, la biométrie a encore du mal à s'imposer en Europe ; ceci est principalement dû à un problème éthique en relation directe avec la CNIL (Commission Nationale de l'Informatique et des Libertés). Il est d'ailleurs intéressant de voir que, dans le domaine de la recherche actuelle en biométrie, la modalité dominante varie d'un pays à l'autre : le visage et l'empreinte digitale pour les États-Unis, l'iris pour la Chine et l'empreinte digitale pour la France.

Parmi les divers niveaux de fusion existant, nous avons choisi de travailler au niveau score car il offre **le meilleur compromis entre la richesse d'information et la facilité d'implémentation**. De plus, ce niveau de fusion est **complètement indépendant des algorithmes qui sont utilisés en amont** pour générer des scores de similarité. Ceci offre donc une grande souplesse d'intégration dans des systèmes existants qui peuvent être mis à jour sans interférer avec la technique de fusion à mettre en place.

Ces travaux ont également montré que les futures méthodes de fusion qui doivent être mises au point ont tout intérêt à **prendre en compte chaque utilisateur en utilisant des méthodes adaptatives** et non plus des simples techniques de combinaison de scores dont la règle reste identique pour chaque individu.

En revanche, il n'existe pas encore de critères universels ni de méthodologie standard permettant de comparer des algorithmes de fusion entre eux ; de même qu'il est encore difficile de trouver des bases de données multimodales composées d'utilisateurs "réels" en nombre élevé.

Cette thèse apporte donc une **contribution triple** au domaine de la biométrie multimodale :

1. Tout d'abord, elle montre les **liens** qu'il existe entre certains **algorithmes fondamentaux de reconnaissance faciale** et le **fonctionnement de notre cerveau**,

2. Ensuite, elle apporte des **solutions efficace aux problèmes de compression, d'analyse et de débruitage grâce aux ondelettes** qui sont exploitées afin d'**améliorer les performances du système biométrique multimodal**,

3. Enfin, cette thèse apporte une **nouvelle approche adaptative à la fusion multimodale par combinaison de scores, en associant l'utilisation des ondelettes à des outils statistiques**.

Il existe quelques perspectives intéressantes en biométrie : par exemple, il serait intéressant d'étudier des techniques pour *crypter des données biométriques* car si elles sont "interceptées" et enregistrées par une personne malveillante, elle pourrait alors s'en servir ultérieurement afin de se faire passer pour une personne enregistrée dans la base de données. Ensuite, un des grands challenges serait de pouvoir *mieux maîtriser les variations d'environnement*, qui perturbent encore trop les systèmes de reconnaissance. En ce qui concerne les techniques de fusion, on pourrait par exemple *fusionner* les vecteurs de score de chaque modalité en utilisant des représentations très condensées comme les *arbres de décomposition en paquets d'ondelettes* (dont la transformée est parfaitement inversible), selon des critères d'entropie (au sens de Shannon) à chaque nœud. Par ailleurs, il est vivement conseillé de comparer, dans la mesure du possible, les performances d'un système biométrique multimodal utilisant des bases de données chimériques avec celles d'un système se servant de bases de données d'utilisateurs "réels" (la toute récente base de données *MBGC* peut constituer un bon point de départ).

Annexe A

Mesures de Distance

Lorsqu'on souhaite comparer deux vecteurs de caractéristiques issus du module d'extraction de caractéristiques d'un système biométrique, on peut soit effectuer une mesure de similarité (ressemblance), soit une mesure de distance (divergence).

La première catégorie de distances est constituée de distances Euclidiennes et sont définies à partir de la *distance de Minkowski d'ordre* p dans un espace euclidien $\mathbf{R}^N$ (N déterminant la dimension de l'espace euclidien).

Considérons deux vecteurs $X = (x_1, x_2, ..., x_N)$ et $Y = (y_1, y_2, ..., y_N)$, la *distance de Minkowski d'ordre* p notée $\mathbf{L}_p$ est définie par :

$$\mathbf{L}_p = \left(\sum_{i=1}^{N} |x_i - y_i|^p \right)^{1/p} \tag{A.1}$$

Nous allons présenter quelques mesures de distance dans l'espace original des images puis dans *l'espace de Mahalanobis*.

A.1 Distances Euclidiennes

A.1.1 Distance City Block (L1)

pour $p = 1$, on obtient la distance *City-Block* (ou distance de *Manhattan*) :

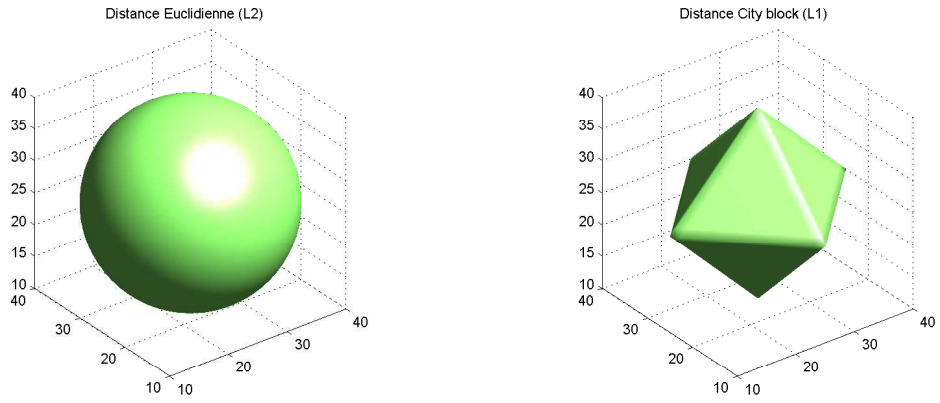
$$\mathbf{L}_1(x, y) = \sum_{i=1}^{N} |x_i - y_i| \tag{A.2}$$

A.1.2 Distance Euclidienne (L2)

Pour $p = 2$, on obtient la distance *euclidienne* :

$$\mathbf{L}_2(x, y) = \sqrt{\sum_{i=1}^{N} |x_i - y_i|^2} \tag{A.3}$$

Les objets peuvent alors apparaître de façon très différentes selon la mesure de distance choisie (Fig. A.1).



A.1.a A.1.b

FIG. A.1 – Représentation d'une sphère avec la distance euclidienne (A.1.a) et la distance City-Block (A.1.b).

A.2 Distances dans l'Espace de Mahalanobis

A.2.1 De l'espace des images à l'espace de Mahalanobis

Avant de pouvoir effectuer des mesures de distance dans l'espace de Mahalanobis, il est essentiel de bien comprendre comment l'on passe de l'espace des images $\mathfrak{I}_m$ à l'espace de Mahalanobis $\mathcal{E}_{Mah}$.

En sortie de l'algorithme PCA, nous obtenons des vecteurs propres associés à des valeurs propres (représentant la variance selon chaque dimension). Ces vecteurs propres définissent une rotation vers un espace dont la covariance entre les différentes dimensions est nulle. **L'espace de Mahalanobis est un espace où la variance selon chaque dimension est égale à 1**. On l'obtient à partir de l'espace des images $\mathfrak{I}_m$ en divisant chaque vecteur propre par son écart-type correspondant.

Soit u et v deux vecteurs propres de $\mathfrak{I}_m$, issus de l'algorithme PCA, et m et n deux vecteurs de $\mathcal{E}_{Mah}$. Soit λ_i les valeurs propres associées aux vecteurs u et v, et σ_i l'écart-type, alors on définit $\lambda_i = \sigma_i^2$. Les vecteurs u et v sont reliés aux vecteurs m et n à partir des relations suivantes :

$$m_i = \frac{u_i}{\sigma_i} = \frac{u_i}{\sqrt{\lambda_i}} \quad et \quad n_i = \frac{v_i}{\sigma_i} = \frac{v_i}{\sqrt{\lambda_i}} \tag{A.4}$$

A.2.2 Mahalanobis L1 (MahL1)

Cette distance est exactement la même que la distance *City-Block* sauf que les vecteurs sont projetés dans l'espace de Mahalanobis. Ainsi, pour des vecteurs propres u et v de projections respectives m et n sur l'espace de Mahalanobis, la distance Mahalanobis L1 est définie par :

$$Mah_{\mathbf{L}_1}(u,v) = \sum_{i=1}^{N} |m_i - n_i| \tag{A.5}$$

A.2.3 Mahalanobis L2 (MahL2)

Cette distance est identique à la distance euclidienne à part qu'elle est calculée dans l'espace de Mahalanobis. Ainsi, pour des vecteurs propres u et v de projections respectives m et n sur l'espace de Mahalanobis, la distance Mahalanobis L2 est définie par :

$$Mah_{\mathbf{L}_2}(u,v) = \sqrt{\sum_{i=1}^{N} |m_i - n_i|^2} \tag{A.6}$$

Par défaut, lorsqu'on parle de distance de Mahalanobis, c'est à cette distance que l'on doit se référer.

A.2.4 Cosinus de Mahalanobis (*MahCosine*)

Il s'agit tout simplement du cosinus de l'angle entre les vecteurs u et v, une fois qu'ils ont été projetés sur $\mathcal{E}_{Mah}$ et normalisées par des estimateurs de la variance (Fig. A.2).

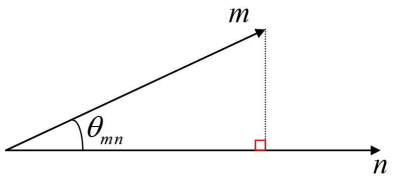


FIG. A.2 – Les deux vecteurs m et n dans l'espace de Mahalanobis

Nous avons donc par définition :

$$S_{MahCosine}(u,v) = \cos(\theta_{mn})$$

De plus, on peut écrire :

$$\cos(\theta_{mn}) = \frac{|m||n| \times \cos(\theta_{mn})}{|m||n|}$$

D'où la formule finale de la mesure de similarité *MahCosine* :

$$S_{MahCosine}(u,v) = \frac{m \cdot n}{|m||n|}, D_{MahCosine}(u,v) = -S_{MahCosine}(u,v) \tag{A.7}$$

où $D_{MahCosine}(u,v)$ est la mesure de distance équivalente. On peut enfin noter qu'il s'agit principalement de la covariance entre les vecteurs dans l'espace de Mahalanobis.

Annexe B

Algorithmes PCA, LDA, EBGM

B.1 Analyse en Composantes Principales (PCA)

L'algorithme PCA est né des travaux de *M. A. Turk* et *A. P. Pentland* au MIT Media Lab, en 1991. Il est aussi connu sous le nom de *Eigenfaces* car il utilise des vecteurs propres et des valeurs propres (respectivement *Eigenvectors* et *Eigenvalues* en anglais).

L'*idée principale* consiste à exprimer M images de départ selon une base de vecteurs orthogonaux particuliers - les vecteurs propres - contenant des informations indépendantes d'un vecteur à l'autre. Les données originales seront donc exprimées d'une manière plus appropriée à la reconnaissance du visage.

En *théorie de l'information*, nous voulons extraire l'information caractéristique d'une image de visage, pour l'encoder aussi efficacement que possible afin de la comparer à une base de données de modèles encodés de manière similaire.

En *termes mathématiques*, cela revient à trouver les vecteurs propres de la matrice de covariance formée par les différentes images de notre base d'apprentissage.

Une image $\mathbf{I}_{i_{(m,n)}}$ est traitée comme un vecteur $\mathbf{\Gamma}_{i_{(m\times n,1)}}$ dans un espace vectoriel de grande dimension $(N = m \times n)$, par concaténation des colonnes (Fig. B.1).

$$\mathbf{I_1} = \begin{pmatrix} a_{1,1} & \dots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,m} \end{pmatrix} \quad \Rightarrow \quad \mathbf{\Gamma_1} = \begin{pmatrix} a_{1,1} \\ \vdots \\ a_{n,1} \\ \vdots \\ a_{1,m} \\ \vdots \\ a_{n,m} \end{pmatrix}$$

FIG. B.1 – Passage d'une image vers un vecteur dans un espace vectoriel de grande dimension. Les coefficients $a_{i,j}$ représentent les valeurs des pixels en niveau de gris, codés de 0 à 255.

Après avoir rassemblé les M images dans une unique matrice, nous obtenons une *matrice d'images* Γ, où chaque colonne représente une image Γ_i :

$$\mathbf{\Gamma} = \begin{pmatrix} a_{1,1} & b_{1,1} & \dots & z_{1,1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n,1} & b_{n,1} & \dots & z_{n,1} \\ \vdots & \vdots & \dots & \vdots \\ a_{1,m} & b_{1,m} & \dots & z_{1,m} \\ \vdots & \vdots & \dots & \vdots \\ a_{n,m} & b_{n,m} & \dots & z_{n,m} \end{pmatrix}$$

On calcule ensuite l'image moyenne Ψ de toutes les images collectées. Cette image peut être vu comme le *centre de gravité* du jeu d'images :

$$\Psi = \frac{1}{M} \sum_{i=1}^{M} \Gamma_i \tag{B.1}$$

On ajuste ensuite les données par rapport à la moyenne. L'image moyenne est alors soustraite de chaque image avec la formule suivante :

$$\Phi_i = \Gamma_i - \Psi, i = 1...M \tag{B.2}$$

On calcule ensuite la matrice de covariance du jeu de données. Cette matrice peut être vue comme une *matrice de moments d'ordre 2* :

$$\mathbf{C} = \sum_{i=1}^{M} \Phi_i \Phi_i^T = AA^T, A = [\Phi_1 \Phi_2 ... \Phi_M] \tag{B.3}$$

La prochaine étape consiste à calculer les vecteurs propres et les valeurs de cette matrice de covariance $\mathbf{C}$ de taille $(N \times N)$, c'est-à-dire *de l'ordre de la résolution d'une image*. Le problème est que cela peut parfois être très difficile et très long ! En effet, si $N > M$ (si la résolution est supérieure au nombre d'images), il y aura seulement $M - 1$ vecteurs propres qui contiendront de l'information (les vecteurs propres restants auront des valeurs propres associées *nulles*). Par exemple, pour 100 images de résolution 320×240, nous pourrions résoudre une matrice $\mathbf{L}$ de 100×100 au lieu d'une matrice de 76800×76800 pour ensuite prendre les combinaisons linéaires appropriées des images Φ_i. Le gain de temps de calcul serait considérable ! Typiquement, nous passerions d'une complexité de l'ordre du nombre de pixels dans une image à une complexité de l'ordre du nombre d'images.

Voici comment procéder pour accélérer les calculs :

Considérons les vecteurs propres e_i de $\mathbf{C} = AA^T$, associés aux valeurs propres λ_i :

$$\mathbf{C}e_i = \lambda_i e_i \tag{B.4}$$

Les vecteurs propres v_i de $\mathbf{L} = A^T A$, associés aux valeurs propres μ_i sont tels que :

$$\mathbf{L}v_i = \mu_i v_i$$

Soit :

$$A^T A v_i = \mu_i v_i$$

En multipliant *à gauche* par A des deux côtés de l'égalité, nous obtenons :

$$A A^T A v_i = A \mu_i v_i$$

Puisque $\mathbf{C} = AA^T$, nous pouvons simplifier :

$$\mathbf{C}(Av_i) = \mu_i (Av_i) \tag{B.5}$$

De (B.4) et (B.5), nous voyons que Av_i et μ_i sont respectivement les vecteurs propres et les valeurs propres de $\mathbf{C}$:

$$\begin{cases} e_i = Av_i \\ \lambda_i = \mu_i \end{cases} \tag{B.6}$$

Nous pouvons donc trouver les valeurs propres de cette énorme matrice $\mathbf{C}$ en trouvant les valeurs propres d'une matrice $\mathbf{L}$ beaucoup plus petite. Pour trouver les vecteurs propres de $\mathbf{C}$, il suffit juste de pré-multiplier les vecteurs propres de $\mathbf{L}$ par la matrice A. Les vecteurs propres trouvés sont ensuite ordonnés selon leurs valeurs propres correspondantes, de manière décroissante. Plus une valeur propre est grande, plus la variance capturée par le vecteur propre est importante. Cela implique que la majeure partie des informations est contenue dans les premiers vecteurs propres.

Une part de la grande efficacité de l'algorithme PCA vient de l'étape suivante qui consiste à ne sélectionner que les k meilleurs vecteurs propres (ceux avec les k plus grandes valeurs propres). On définit alors un espace vectoriel engendré par ces k vecteurs propres, que l'on appelle *l'espace des visages* E_v (“*Face Space*”).

Les images originales peuvent être reconstituées par combinaison linéaire de ces vecteurs propres. Les représentations graphiques de ces vecteurs rappellent un peu des images fantômes, chacune mettant en avant une partie du visage, on les appelle *eigenfaces* (Fig. B.2).

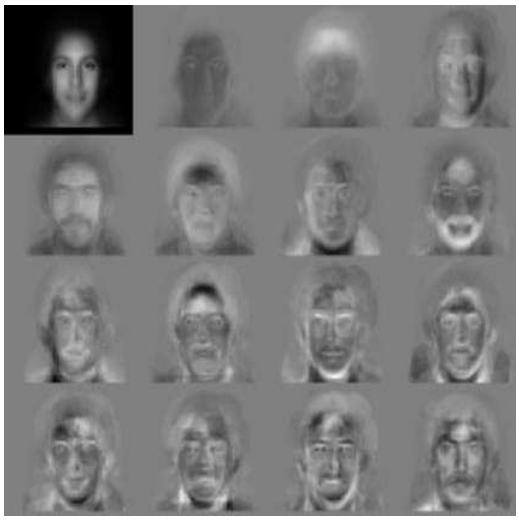


FIG. B.2 – Image moyenne et les $15^{\text{ères}}$ eigenfaces.

Nous allons maintenant projeter nos images de départ sur E_v. Une image Γ_i est alors transformée en ses composantes eigenfaces par une simple opération de projection vectorielle :

$$\omega_k = {e_k}^T(\Gamma_i - \Psi), k = 1, ..., M' \tag{B.7}$$

Les vecteurs ω_k sont appelés *poids* et forment une matrice $\Omega^T = [\omega_1, \omega_2, ..., \omega_{M'}]$ qui décrit la contribution de chaque eigenface dans la représentation de l'image d'entrée. La matrice Ω^T est alors utilisée pour trouver quelle est, parmi un nombre pré-défini de classes, celle qui décrit le mieux une image d'entrée.

La méthode la plus simple pour déterminer quelle classe de visage fournit la meilleure description d'une image d'entrée est de trouver la classe de visage k qui minimise la distance euclidienne.

$$\varepsilon_k^2 = \|\Omega - \Omega_k\|^2 \tag{B.8}$$

où Ω_k est un vecteur qui décrit la k^e classe de visage.

Un visage appartient à une classe k quand le minimum ε_k est en dessous d'un certain seuil θ_ε. Dans le cas contraire, le visage est classé comme étant *inconnu* et peut éventuellement être utilisé pour créer une nouvelle classe de visage. La création de la matrice de poids Ω^T est équivalent à la projection du visage original sur E_v. Puisque la distance ε entre l'image de visage et E_v est simplement le carré de la distance entre l'image d'entrée réajustée par rapport à la moyenne $\Phi = \Gamma - \Psi$ et $\Phi_f = \sum_{i=1}^{M'} \omega_i e_i$, sa projection sur E_v est : $\varepsilon^2 = \|\Phi - \Phi_f\|^2$.

Il y a alors **quatre possibilités** pour une image d'entrée d'être reconnue ou non (Fig. B.3).

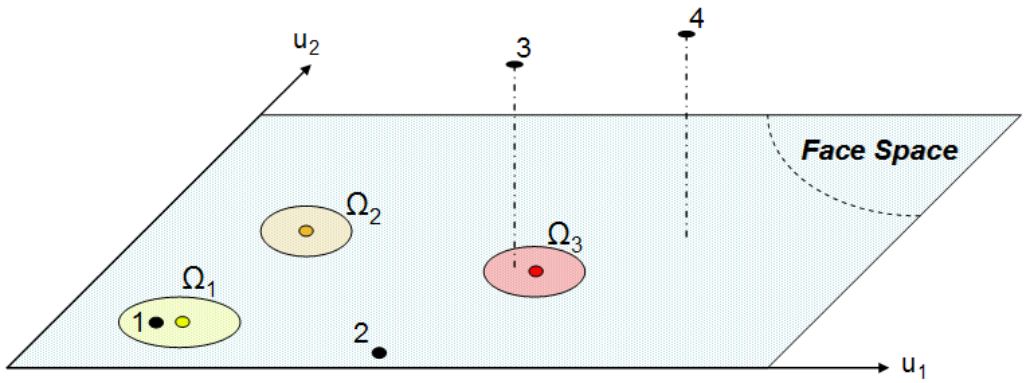


FIG. B.3 – Une version simplifiée de E_v illustrant les quatre résultats de la projection d'une image sur E_v. Dans ce cas, il y a deux vecteurs propres ($u1$ et $u2$) et trois classes d'individus connus ($\Omega1$, $\Omega2$, $\Omega3$).

Dans le cas 1, un individu est reconnu et identifié. Dans le cas 2, un individu inconnu est présent. Les deux derniers cas (3 et 4) indiquent que l'image n'est pas une image de visage. Pour le cas 3 , l'image est éloigné de E_v mais la projection est proche d'une classe connue, on parle alors de *fausse acceptation*. Ces remarques sont résumées dans le tableau B.1

	Espace des Visages (E_v)	Classe de Visage
Cas 1	proche	proche
Cas 2	proche	éloigné
Cas 3	éloigné	proche
Cas 4	éloigné	éloigné

TAB. B.1 – Les quatres possibilités qui apparaissent lors de la phase de reconnaissance.

En résumé, l'algorithme PCA est une méthode globale utilisant en premier lieu les niveaux de gris des pixels d'une image. Sa simplicité à mettre en oeuvre contraste avec une forte sensibilité aux changements d'éclairement, de pose et d'expression faciale.
Néanmoins, le PCA ne nécessite aucune connaissance à priori sur l'image et se révèle plus efficace lorsqu'il est couplé à la mesure de distance MahCosine.
Le principe selon lequel on peut construire un sous-espace vectoriel en ne retenant que les « meilleurs »vecteurs propres, tout en conservant beaucoup d'information utile, fait du PCA un algorithme efficace et couramment utilisé en réduction de dimensionnalité.
Enfin, l'étude théorique de l'algorithme PCA est très pédagogique et permet d'acquérir de solides bases pour la reconnaissance 2D du visage. C'est un algorithme incontournable !

B.2 Analyse Discriminante Linéaire (LDA)

L'algorithme LDA est né des travaux de *Belhumeur et al.* de l'Université Yale (New Haven, USA), en 1997. Il est aussi connu sous le nom de *Fisherfaces*.

Contrairement à l'algorithme PCA, l'algorithme LDA effectue une véritable *séparation de classes* (Fig. B.4).
Pour pouvoir l'utiliser, il faut donc au préalable organiser la base d'images d'apprentissage en plusieurs classes : une classe par personne et plusieurs images par classe. Le LDA analyse les vecteurs propres de la matrice de dispersion des données, avec pour objectif de maximiser les variations inter-classes tout en minimisant les variations intra-classes.

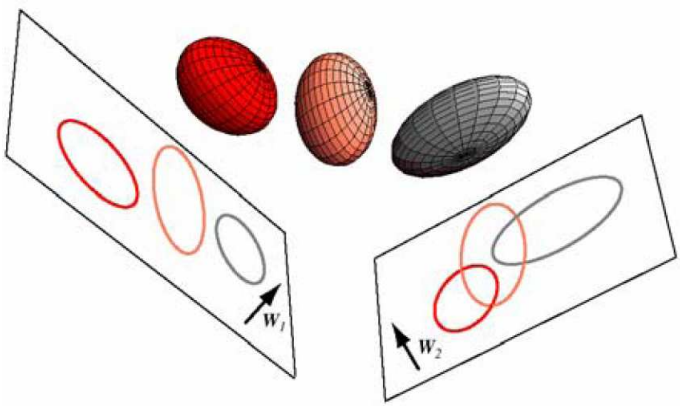


FIG. B.4 – Illustration du principe de séparation optimale des classes par le LDA. Trois distributions 3D sont projetées sur deux sous-espaces 2D décris par les vecteurs $\mathbf{W_1}$ et $\mathbf{W_2}$. Puisque le LDA essaye de trouver la plus grande séparation parmi les classes, on voit bien que $\mathbf{W_1}$ est ici le vecteur optimal [Baht 03].

Tout comme dans le PCA, on rassemble les images de la base d'apprentissage dans une grande matrice d'images Γ où chaque colonne représente une image Γ_i, puis on calcule l'image moyenne Ψ.
Ensuite, pour chaque classe C_i, on calcule l'image moyenne Ψ_{C_i} :

$$\Psi_{C_i} = \frac{1}{q_i} \sum_{k=1}^{q_i} \Gamma_k \tag{B.9}$$

avec q_i, le nombre d'images dans la classe C_i.

Chaque image Γ_i de chaque classe C_i est ensuite recentrée par rapport à la moyenne. On obtient alors une nouvelle image Φ_i :

$$\Phi_i = \Gamma_i - \Psi_{C_i} \tag{B.10}$$

Vient ensuite le calcul de nos différentes matrices de dispersion. On notera c le nombre total de classes (i.e. le nombre d'individus), q_i le nombre d'images dans la classe C_i et M le nombre total d'images.

1. La Matrice de Dispersion Intra-Classe ($\mathbf{S}_w$)

$$\mathbf{S}_w = \sum_{i=1}^{c} \sum_{\Gamma_k \in C_i} (\Gamma_k - \Psi_{C_i})(\Gamma_k - \Psi_{C_i})^T \tag{B.11}$$

2. La Matrice de Dispersion Inter-Classe ($\mathbf{S}_b$)

$$\mathbf{S}_b = \sum_{i=1}^{c} q_i \, (\Psi_{C_i} - \Psi)(\Psi_{C_i} - \Psi)^T \tag{B.12}$$

3. La Matrice de Dispersion Totale ($\mathbf{S}_T$)

$$\mathbf{S}_T = \sum_{i=1}^{M} (\Gamma_i - \Psi)(\Gamma_i - \Psi)^T \tag{B.13}$$

Une fois ces matrices calculées, nous devons trouver une projection optimale W qui minimise la dispersion intra-classe, relative à la matrice $\mathbf{S}_w$, tout en maximisant la dispersion inter-classe, relative à la matrice $\mathbf{S}_b$.
En d'autres termes, nous devons trouver W qui maximise le *critère d'optimisation de Fisher* $J(T)$:

$$W = \arg\max_T \left(J(T)\right)$$

$$\Rightarrow \quad \max(J(T)) = \frac{|T^T \mathbf{S}_b T|}{|T^T \mathbf{S}_w T|} | T = W \tag{B.14}$$

W peut alors être trouvée en résolvant le problème généralisé aux valeurs propres [Golu 04] :

$$\mathbf{S}_b W = \lambda_w \mathbf{S}_w W \tag{B.15}$$

Une fois W trouvée, le même schéma que le PCA concernant la projection des images apprises ainsi que la projection d'une image test est appliqué.
Ainsi, la projection vectorielle d'une image apprise réajustée par rapport à la moyenne Φ_i est définie par :

$$g(\Phi_i) = W^T \Phi_i$$

La phase de reconnaissance d'une image test Φ_t s'effectue en projetant Φ_t sur W^T :

$$g(\Phi_t) = W^T \Phi_t$$

Enfin, on effectue une mesure de distance entre l'image test et l'image projetée sur l'espace vectoriel engendré par W^T. Par exemple, pour la distance euclidienne, on calcule la distance d_{ti} :

$$d_{ti} = \|g(\Phi_t) - g(\Phi_i)\|$$

d'où :

$$d_{ti} = \sqrt{\sum_{k=1}^{c} (g(\Phi_t) - g(\Phi_i))^2} \tag{B.16}$$

Finalement, une image test est dans la classe dont la distance est minimale par rapport à toutes les autres distances de classe.

En résumé, l'algorithme LDA permet d'effectuer une véritable séparation de classes, selon un critère mathématique qui minimise les variations entre les images d'un même individu (variations intra-classe) tout en maximisant les variations entre les images d'individus différents (variations inter-classes). Cependant, pour des problèmes « sous-échantillonnés »en reconnaissance du visage, c'est-à-dire lorsque le nombre d'individus à traiter est plus faible que la résolution de l'image, il est difficile d'appliquer le LDA qui peut alors faire apparaître des matrices de dispersions singulières (non inversibles). Afin de contourner ce problèmes, certains algorithmes basés sur le LDA ont récemment été mis au point (par exemple, les algorithmes ULDA, OLDA et NLDA).

B.3 Elastic Bunch Graph Matching (EBGM)

En 1997, *Wiskott et al.* de la *Southern California University* (USC, USA) et de la *Rurh University* (Allemagne) proposent d'utiliser un modèle d'*Elastic Bunch Graph Matching* (EBGM) pour la reconnaissance faciale. À partir d'une image de visage, on localise des points caractéristiques (coins des yeux, de la bouche, nez, etc.). Cette localisation peut se faire manuellement ou automatiquement à l'aide d'un algorithme (Fig. B.5) [Arca 06].

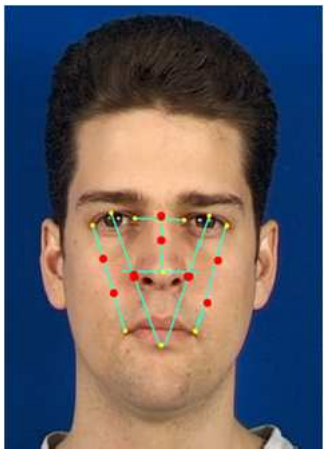
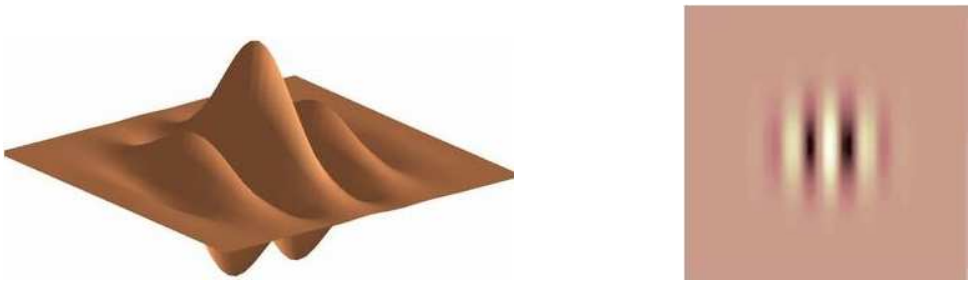


FIG. B.5 – Extraction de points caractéristiques dans l'algorithme EBGM.

Un *treillis élastique virtuel* est ensuite appliqué sur l'image de visage à partir de ces points. Chaque point représente un *nœud labélisé* auquel on associe un jeu de coefficients d'ondelettes complexes de Gabor, appelés *Jet*.

Pour effectuer une reconnaissance avec une image test, on fait une mesure de similarité entre les différents Jets et les longueurs des segments du treillis de deux images.

Un *Jet* est basé sur une *transformée en ondelettes* [Hubb 00], défini comme la convolution d'une image avec une famille de *noyaux de Gabor*. Ces noyaux de Gabor peuvent être assimilés à des ondes localisées dans le temps, modulées par une gaussienne (Fig. B.6). On peut parler de transformée en ondelettes car la famille des noyaux de Gabor est générée à partir d'une *ondelette mère* par dilatation et rotation.



(B.6.a) (B.6.b)

FIG. B.6 – Partir réelle d'un filtre de Gabor. Vue en 3D (B.6.a), vue en 2D de dessus (B.6.b).

Dans l'EBGM, les ondelettes de Gabor sont des fonctions de type $f(\theta, \lambda, \phi, \sigma, \gamma)$. Voici

une explication des cinq paramètres :

1. **L'orientation de l'ondelette** (θ) : ce paramètre fait pivoter l'ondelette autour de son centre. L'orientation de l'ondelette (Fig. B.7) détermine l'angle des contours ou des lignes de l'image auxquelles l'ondelette va être sensible,

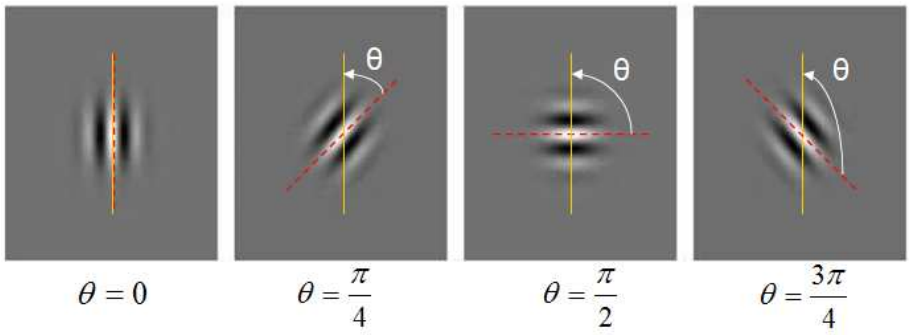


FIG. B.7 – Orientation de l'ondelette.

2. **La fréquence centrale de l'ondelette** (λ) : ce paramètre spécifie la longueur d'onde du cosinus ou inversement la fréquence centrale de l'ondelette (Fig. B.8). Les ondelettes avec une grande longueur d'onde seront sensibles à des changements progressifs d'intensité dans une image. Les ondelettes avec une petite longueur d'onde seront sensibles à des contours et des lignes abruptes,

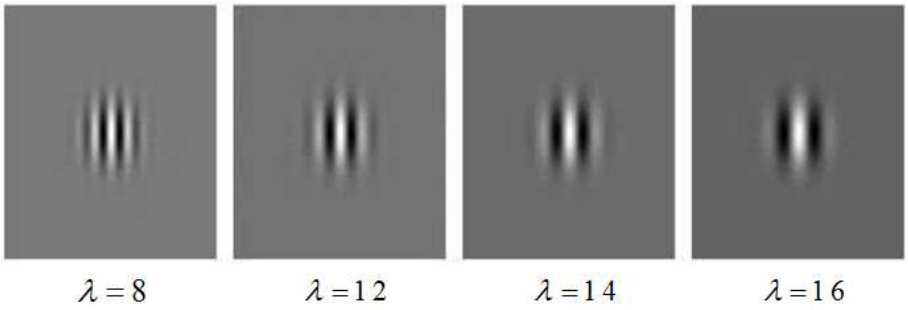


FIG. B.8 – Fréquence centrale de l'ondelette.

3. **La phase de la sinusoïde** (ϕ) : on utilise à la fois la partie réelle et la partie imaginaire de l'ondelette complexe de Gabor. Ce qui nous donne, en quelque sorte, deux ondelettes : une ondelette paire et une ondelette impaire. La convolution (incluant alors deux phases) donne un coefficient complexe basé sur deux ondelettes qui sont déphasées de $\frac{\pi}{2}$ (Fig. B.9),

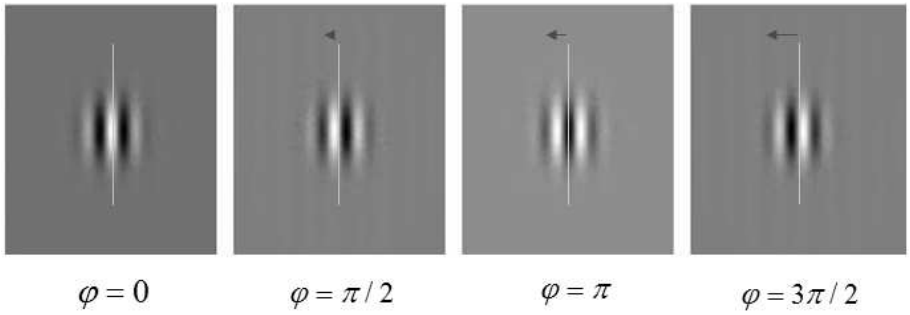


FIG. B.9 – Phase de l'ondelette.

4. **Le support temporel de l'ondelette** (σ) : ce paramètre spécifie le rayon de la gaussienne (Fig. B.10). La taille de la gaussienne détermine la quantité de pixels de l'image qui vont être pris en compte dans la convolution,

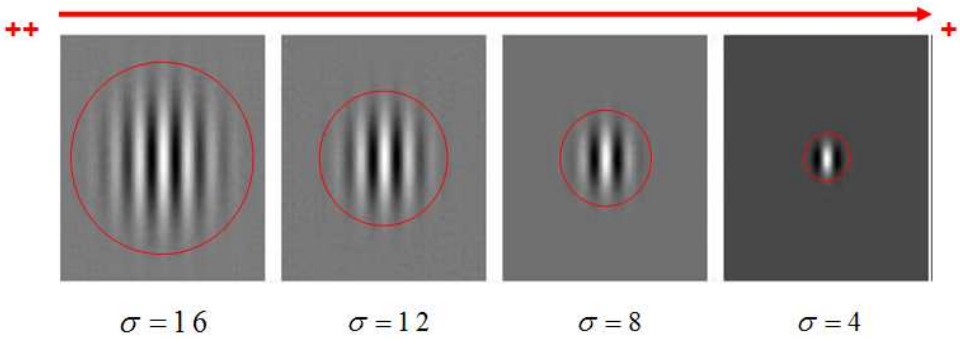


FIG. B.10 – Support temporel de l'ondelette.

5. **L'enveloppe de la gaussienne** (γ) : ce paramètre agit sur la forme de l'enveloppe gaussienne, en l'étirant spatialement (Fig. B.11). Ce paramètre a été inclus de manière à ce que les ondelettes puissent approximer certains modèles biologiques. La plupart des ondelettes testées avec l'algorithme EBGM du CSU System 5.0 [Beve 03] utilisent une enveloppe gaussienne avec $\gamma = 1$.

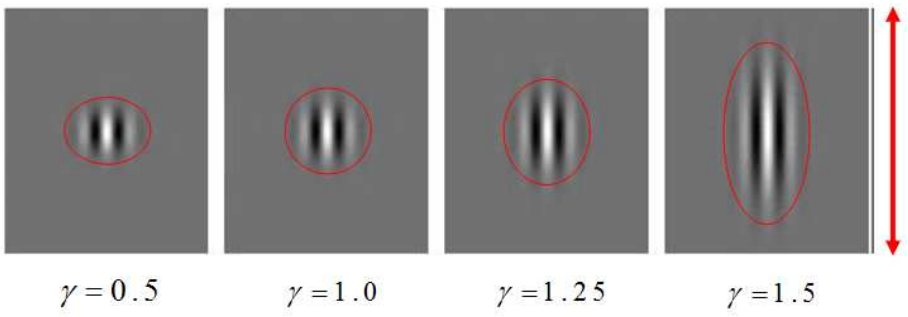


FIG. B.11 – Forme de l'enveloppe gaussienne.

En utilisant cinq fréquences différentes, huit orientations différentes, et deux phases différentes. On obtient un total de 80 masques d'ondelettes de Gabor différents (Fig. B.12).

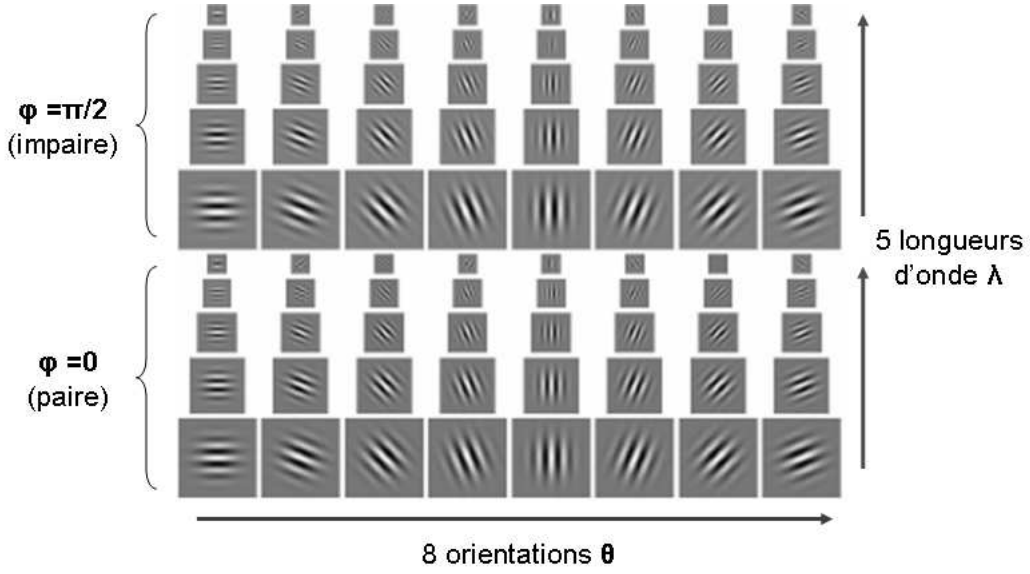


FIG. B.12 – Les 80 masques d'ondelettes de Gabor utilisés dans l'algorithme EBGM.

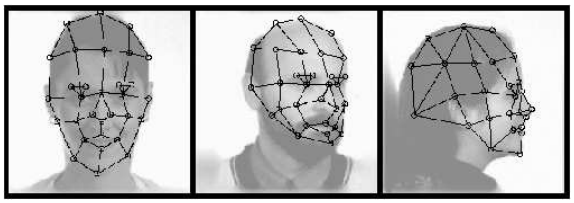
Maintenant que nous avons vu comment paramétrer l'ondelette de Gabor, nous allons résumer les différentes étapes de l'algorithme EBGM à l'aide de quelques illustrations :

1. **Sélection de points caractéristiques** : des points caractéristiques d'une image de visage sont sélectionnés (manuellement ou avec un algorithme) (Fig. B.13.a),

2. **Création du treillis** : un treillis (Fig. B.13.b) est construit en reliant les points caractéristiques précédemment trouvés,



(B.13.a)



(B.13.b)

FIG. B.13 – Sélection de points caractéristiques (B.13.a) et création du treillis (B.13.b).

3. **Calcul des Jets** : à chaque nœud du treillis correspond un point caractéristique et contient un jeu de coefficients complexes d'ondelettes de Gabor : **le Jet**. Les différents Jets sont calculés en convoluant l'image autour des points caractéristiques avec plusieurs ondelettes de Gabor paramétrées. Un Jet peut être écrit comme un ensemble de coefficients complexes $\mathcal{J}_j = a_j \exp(i\phi_j)$ avec une amplitude a_j qui varie peu avec la position, et une phase ϕ_j, dont la variation en rotation est plus importante. La représentation obtenue est appelée "*Face Bunch Graph*"(FBG) (Fig. B.14),

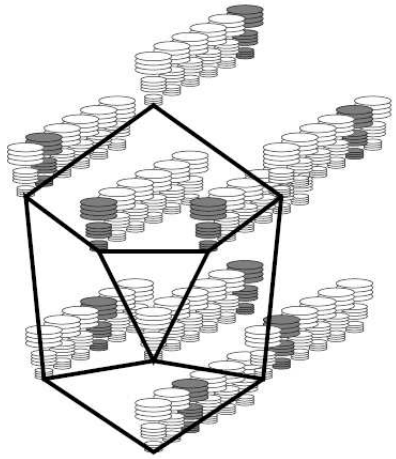


FIG. B.14 – Le "*Face Bunch Graph*" sert de représentation générale pour les visages. Chaque empilement de disques représente un Jet. À partir d'un ensemble de Jets relié à un nœud du treillis, seulement celui qui présente la plus grande similarité est sélectionné pour la reconnaissance (indiqué en ombre grisée à titre d'exemple).

4. **Calcul de similarité de deux images** : une fois la structure du FBG trouvée, l'algorithme va effectuer un calcul de similarité entre une image de la base d'apprentissage et une image test. Pour cela, les points caractéristiques de l'image test sont trouvés, le treillis est mis en place et les nouveaux Jets calculés. Précisons une dernière fois que l'on associe au treillis non seulement la localisation des points caractéristiques mais aussi les différents Jets. La similarité de deux images est alors une fonction de la correspondance des treillis. La reconnaissance finale se fait en maximisant cette fonction.

En résumé, l'algorithme EBGM trouve ses fondements dans les neurosciences, en imitant le fonctionnement de certaines cellules spécialisées localisées dans le cortex visuel primaire. C'est un algorithme local (il ne traite pas directement les valeurs de niveaux de gris des pixels d'une image de visage), ce qui lui confère une plus grande robustesse aux changements d'éclairement, de pose et d'expression faciale. Cependant il est plus difficile à implémenter que les méthodes globales PCA et LDA précédemment exposées, et le temps de prétraitement s'en retrouve augmenté. Enfin, une partie de son originalité provient du fait que l'EBGM utilise des ondelettes entièrement paramétrables pour générer des coefficients complexes qui vont être utilisés lors de la phase de reconnaissance.

Annexe C

Preuves

C.1 Preuve de l'approximation de la distribution binomiale par la distribution normale

La distribution binomiale donne la distribution discrète de probabilité $P_p(n|N)$ d'obtenir exactement n succès parmi N *essais de Bernoulli* (où le résultat de chaque essai de *Bernoulli* est vrai avec la probabilité p et faux avec la probabilité $q = 1 - p$).
La distribution binomiale est donc donnée par :

$$P_p(n|N) = C_N^n p^n q^{N-n} \tag{C.1}$$

$$= \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \tag{C.2}$$

Pour N assez grand, une approximation de la distribution binomiale peut être obtenue en effectuant un développement limité de P au voisinage de $\tilde{n}$ où $P(n)$ est maximum, c'est-à-dire où $\frac{d}{dn}P(n) = 0$. Puisque la fonction $\ln$ est monotone, on peut également effectuer le développement limité de $\ln[P(n)]$. Soit $n \equiv \tilde{n} + \eta$, nous avons :

$$\ln[P(n)] = \ln[P(\tilde{n})] + B_1\eta + \frac{1}{2}B_2\eta^2 + \frac{1}{3!}B_3\eta^3 + \dots \tag{C.3}$$

où

$$B_k \equiv \left[\frac{d^k \ln[P(n)]}{dn^k}\right]_{n=\tilde{n}}$$

Or, nous développons au voisinage du maximum, donc, par définition :

$$B_1 = \left[\frac{d \ln[P(n)]}{dn}\right]_{n=\tilde{n}} = 0 \tag{C.4}$$

Ce qui signifie aussi que B_2 est négatif et que l'on peut donc écrire $B_2 = -|B_2|$.
En prenant le logarithme de (C.2), nous avons :

$$\ln[P(n)] = \ln N! - \ln n! - \ln(N-n)! + n \ln p + (N-n) \ln q \tag{C.5}$$

Pour n et $N - n$ assez grands, nous pouvons utiliser l'approximation de Stirling :

$$\ln[n!] \simeq n \ln n - n \tag{C.6}$$

Ainsi

$$\frac{d[\ln(n!)]}{dn} \approx (\ln n + 1) - 1 \tag{C.7}$$

$$= \ln n \tag{C.8}$$

$$\frac{d[\ln(N-n)!]}{dn} \approx \frac{d}{dn}[(N-n)\ln(N-n) - (N-n)] \tag{C.9}$$

$$= \left[-\ln(N-n) + (N-n)\frac{-1}{N-n} + 1\right] \tag{C.10}$$

$$= -\ln(N-n) \tag{C.11}$$

et

$$\frac{d\ln[P(n)]}{dn} \simeq -\ln n + \ln(N-n) + \ln p - \ln q \tag{C.12}$$

Pour trouver $\tilde{n}$, nous devons résoudre (pour n) l'expression précédente égale à zéro :

$$\ln\left(\frac{N-\tilde{n}}{\tilde{n}}\frac{p}{q}\right) = 0 \tag{C.13}$$

$$\frac{N-\tilde{n}}{\tilde{n}}\frac{p}{q} = 1 \tag{C.14}$$

$$(N-\tilde{n})p = \tilde{n}q \tag{C.15}$$

$$\tilde{n}(q+p) = \tilde{n} = Np \tag{C.16}$$

Nous pouvons maintenant trouver les termes B_i du développement limité :

$$B_2 \equiv \left[\frac{d^2\ln[P(n)]}{dn^2}\right]_{n=\tilde{n}} \tag{C.17}$$

$$= -\frac{1}{\tilde{n}} - \frac{1}{N-\tilde{n}} \tag{C.18}$$

$$= -\frac{1}{Npq} \tag{C.19}$$

$$= -\frac{1}{Np(1-p)} \tag{C.20}$$

De même, on trouve :

$$B_3 \equiv \left[\frac{d^3\ln[P(n)]}{dn^3}\right]_{n=\tilde{n}} \tag{C.21}$$

$$= \frac{1-2p}{N^2p^2(1-p)^2} \tag{C.22}$$

$$B_4 \equiv \left[\frac{d^4\ln[P(n)]}{dn^4}\right]_{n=\tilde{n}} \tag{C.23}$$

$$= \frac{2(3p^2-3p+1)}{N^3p^3(1-p)^3} \tag{C.24}$$

Nous considérons maintenant les distributions dans le domaine continu :

$$\lim_{N\to+\infty}\sum_{n=0}^{N} p(n) \simeq \int P(n)dn = \int_{-\infty}^{+\infty} P(\tilde{n}+\eta)d\eta = 1 \tag{C.25}$$

De plus, puisque chaque terme B_i est de l'ordre de $\frac{1}{N} \sim \frac{1}{\sigma^2}$ plus petit que B_{i-1}, on peut négliger les termes supérieurs à B_2. Ainsi :

$$P(n) = P(\tilde{n}) e^{\frac{-|B_2|\eta^2}{2}} \tag{C.26}$$

Par ailleurs, la probabilité devant être normalisée, on a :

$$\int_{-\infty}^{+\infty} P(\tilde{n}) e^{\frac{-|B_2|\eta^2}{2}} d\eta = P(\tilde{n}) \sqrt{\frac{2\pi}{|B_2|}} = 1 \tag{C.27}$$

et

$$P(n) = \sqrt{\frac{|B_2|}{2\pi}} e^{\frac{-|B_2|(n-\tilde{n})^2}{2}} \tag{C.28}$$

$$= \frac{1}{\sqrt{2\pi Npq}} \exp\left[-\frac{(n-Np)^2}{2Npq}\right] \tag{C.29}$$

En définissant $\tilde{n} \equiv Np$ et $\sigma^2 \equiv Npq$:

$$P(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(n-\tilde{n})^2}{2\sigma^2}\right] \tag{C.30}$$

qui est une distribution normale.

La distribution binomiale peut donc être approximée par une distributionn normale pour n'importe quel p fixé (même faible) et pour $N \to +\infty$.

Remarque : Si $N \to +\infty$ et $p \to 0$ de telle sorte que $Np \to \lambda$, alors la distribution binomiale converge vers une distribution de Poisson de moyenne λ.

C.2 Preuve du seuil universel de Donoho en utilisant le Lemme de Borel-Cantelli

Nous allons tout d'abord donner le **Lemme de Borel-Cantelli** (en *deux parties*) puis le théorème qui nous permet de justifier la formule du **seuil universel de Donoho**, avant d'en donner une démonstration.

Lemme 1 (Premier Lemme de Borel-Cantelli) *Soit (E_n) une suite d'événements dans un espace de probabilité quelconque. Si la somme des probabilités de (E_n) est finie ($\sum_{n=1}^{\infty} P(E_n) < \infty$) alors la probabilité qu'une infinité d'entre eux se réalisent simultanément est nulle. C'est-à-dire que $P(\limsup_{n\to+\infty} E_n) = 0$.*

Lemme 2 (Second Lemme de Borel-Cantelli) *Si les événements de (E_n) sont indépendants et que la somme des probabilités de la suite (E_n) diverge vers l'infini, alors la probabilité qu'une infinité d'entre eux se réalisent simultanément est 1.*

Théorème 1 *Soit $g_n(w)$, $w \in \Omega$, $n \geq 1$, une séquence d'événements indépendants et identiquement distribués (i.i.d.) $N(0,1)$. Nous avons, presque sûrement : $\overline{\lim}_{n\to+\infty} \frac{g_n(w)}{\sqrt{2log(n)}} = 1$ et $\underline{\lim} \frac{g_n(w)}{\sqrt{2log(n)}} = -1$.*

$\forall \varepsilon > 0$, on considère l'ensemble $E_n(\varepsilon)$ des $w \in \Omega$, tel que : $g_n(w) > (1+\varepsilon)\sqrt{2log(n)}$. Ainsi, nous avons :

$$P(E_n(\varepsilon)) = \frac{1}{\sqrt{2\pi}} \int_{(1+\varepsilon)\sqrt{2log(n)}}^{\infty} e^{-t^2/2}\, dt.$$

Mais nous avons également :

$$\int_x^{\infty} e^{-t^2/2}\, dt = \sqrt{2} \int_{x^2/2}^{\infty} e^{-u}\, \frac{du}{\sqrt{u}}\ ,$$

et :

$$\int_x^{\infty} e^{-u} u^{\alpha}\, du = e^{-x} x^{-\alpha} \left(1 + O\left(\frac{1}{x}\right)\right) \text{ (Integration par parties).}$$

Dans notre cas, nous obtenons :

$$\int_x^{\infty} e^{-t^2/2}\, dt = e^{-x^2/2} \left(\frac{1}{x} + O\left(\frac{1}{x^3}\right)\right)\ ,$$

et :

$$P(E_n(\varepsilon)) = \frac{1}{\sqrt{2\pi}} \frac{1}{(1+\varepsilon)\sqrt{2log(n)}} e^{-(1+\varepsilon)^2 log(n)} = n^{-(1+\varepsilon)^2} \varepsilon_n \quad \text{avec} \quad \varepsilon_n \to 0\ .$$

Il s'agit d'une suite finie. Nous pouvons donc appliquer le Lemme de Borel-Cantelli et $\limsup E_n(\varepsilon) = \emptyset$.

Maintenant, on considère l'ensemble (F_n) défini par $g_n(w) > \sqrt{2log(n)}$. Ces ensembles sont indépendants et $P(F_n)$ a pour équivalent :

$$\frac{1}{\sqrt{2\pi}\sqrt{2log(n)n}}\ ,$$

qui est le terme général d'une série divergente. Ainsi, $\limsup F_n(\varepsilon) = \Omega$.
C'est-à-dire que pour presque tous les $w \in \Omega$, nous avons $g_n(w) > \sqrt{2log(n)}$ pour une infinité de valeurs de n tandis que $g_n(w) < (1+\varepsilon)\sqrt{2log(n)}$ pour n assez grand.

Finalement, nous avons :

$$\lim_{n\to+\infty} \frac{g_n(w)}{\sqrt{2log(n)}} = 1 \ ,$$

De la même manière, on montre que :

$$\underline{\lim} \frac{g_n(w)}{\sqrt{2log(n)}} = -1 \ .$$

Publications et Conférences

N. Morizet et J. Gilles, "*A New Adaptive Combination Approach to Score Level Fusion for Face and Iris Biometrics Combining Wavelets and Statistical Moments*". ISVC 2008, 4th International Symposium on Visual Computing, December 1-3 2008, Las Vegas, Nevada, USA.

N. Morizet, I.D. Hamed, F. Amiel et T. Ea, "*A Comparative Implementation of PCA Face Recognition Algorithm*". ICECS 2007, 14th IEEE International Conference on Electronics, Circuits and Systems, December 11-14 2007, Marrakech, Morocco.

N. Morizet, "*Wavelet Domain Match Score Level Fusion for Multimodal Biometrics*". COGIS 2007, 3rd Conference on COGnitive systems with Interactive Sensors, November 26-27 2007, Stanford University, California, CA, USA.

N. Morizet, "*Le Croisement des TIC et de la Cognition : l'exemple de la biométrie*". Conférence FING, 25 Avril 2007, Université Paris Descartes, France.

N. Morizet, "*Revue des Algorithmes PCA, LDA et EBGM utilisés en reconnaissance 2D du visage pour la biométrie*". Tutoriel tenu pendant la conférence MajecSTIC'06, 24 Novembre 2006, Lorient, France.

N. Morizet, "*Initiation aux Ondelettes*". Article invité de la Revue de l'Electricité et de l'Electronique (REE). **Prix Jean-Marie Bergès 2007, meilleur article de la REE 2006**.

N. Morizet, "*Ondelettes et Applications*". Conférence invitée donnée à Thales Land & Joint Systems, 21 Novembre 2008.

Bibliographie

[Abry 97] P. Abry and P. Flandrin. *Ondelettes et turbulences. Multirésolutions, algorithmes de décomposition, invariance d'échelle et signaux de pression*. Diderot Éditeur, Arts & Sciences, 1997.

[Ahon 04] T. Ahonen, A. Hadid, and M. Pietikainen. *Face Recognition with Local Binary Patterns*. 2004.

[Alte 08] "Altera. FPGAs, CPLDs & Structure ASICs". 2008. `http://www.altera.com`.

[Andr 00] D. Andrews and M. Buchinsky. "A Three-Step Method for Choosing the Number of Bootstrap Repetitions". *Econometrica*, Vol. 68, No. 1, pp. 23–52, January 2000.

[Arca 06] S. Arca, P. Campadelli, and R. Lanzarotti. "A Face Recognition System Based On Automatically Determined Facial Fiducial Points". *Pattern Recognition*, Vol. 39, No. 3, pp. 432–443, 2006.

[Ausc 92] W. G. Auscher, P. and V. Wickerhauser. "Local sine and cosine bases of Coifman and Meyer and the construction of smooth wavelets". pp. 237–256, 1992.

[Baht 03] G. Bahtiyar. *Holistic Face Recognition by Dimension Reduction*. Master's thesis, Department of Electrical and Electronics Engineering, Graduate School of Natural and Applied Sciences of the Middle East Technical University, September 2003.

[Bail 03] E. Bailly-Bailliére, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruiz, and J.-P. Thiran. "The BANCA Database and Evaluation Protocol". In : *Proceedings of Fourth International Conference on Audio- and Video-Based Authentication (AVBPA)*, pp. 625–638, 2003.

[Bail 06] G. Bailly, F. Elisel, P. Badin, and C. Savariaux. "Degrees of freedom of facial movements in face-to-face conversational speech". In : *International Workshop on Multimodal Corpora*, pp. 33–36, 2006.

[Barl 89] H. Barlow. "Unsupervised Learning". *Neural Computation*, Vol. 1, pp. 295–311, 1989.

[Bart 02] M. Bartlett, J. Movellan, and T. Sejnowski. "Face recognition by independent component analysis". *IEEE Transactions on Neural Networks*, Vol. 13, No. 6, pp. 1450–1464, 2002.

[Bart 07] M. Bartlett. "Information maximization in face processing". *Neurocomputing*, Vol. 70, No. 13-15, pp. 2204–2217, 2007.

[Belh 97] P. Belhumeur, J. Hespanha, and D. Kriegman. "Eigenfaces vs. Fisherfaces : Recognition Using Class Specific Linear Projection". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 711–720, 1997.

[Beve 03] J. Beveridge, D. Bolme, M. Teixeira, and B. Draper. "The CSU Face Identification Evaluation System User's Guide : Version 5.0". Tech. Rep., Colorado State University, May 2003.

[Beve 05] R. Beveridge and M. Kirby. "Biometrics and Face Recognition". *IS&T Colloquium*, p. 25, 2005.

[Bigu 97] E. Bigun, J. Bigun, B. Duc, and S. Fischer. "Expert Conciliation for Multimodal Person Authentication Systems using Bayesian Statistics". In : *Proceedings of First International Conference on Audio- and Video-Based Authentication (AVBPA)*, pp. 291–300, Crans-Montana, Switzerland, March 1997.

[Bled 64] W. Bledsoe. "The model method in facial recognition". Tech. Rep. PRI :15, Stanford University, Palo Alto, CA, USA, 1964.

[Bole 98] W. Boles and B. Boashash. "A human identification technique using images of the iris and wavelet transform". In : *IEEE Transactions on Signal Processing*, pp. 1185–1188, April 1998.

[Boll 05] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior. "The Relation between the ROC Curve and the CMC". In : *Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 15–20, 2005.

[Bolm 03] D. Bolme, J. Beveridge, M. Teixeira, and B. Draper. "The CSU Face Identification Evaluation System : Its Purpose, Features, and Structure". In : *Proceedings of the 3rd International Conference on Computer Vision Systems (ICVS)*, pp. 304–313, 2003.

[Bowy 08] K. Bowyer, K. Hollingsworth, and P. Flynn. "Image understanding for iris biometrics : A survey". *Computer Vision and Image Understanding*, Vol. 110, No. 2, pp. 281–307, 2008.

[Brow 01] G. Brown, S. Yamada, and T. Sejnowski. "Independent Component Analysis at the Neural Cocktail Party". *Trends in Neuroscience*, Vol. 24, pp. 54–63, 2001.

[Brun 95] R. Brunelli and D. Falavigna. "Person identification using multiple cues". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, pp. 955–966, 1995.

[Burt 93] P. Burt and R. Kolczynski. "Enhanced image capture through fusion". In : *Proceedings of the 4th International Conference on Computer Vision (ICCV)*, pp. 173–182, May 1993.

[Capp 00] R. Cappelli, D. Maio, and D. Maltoni. "Combining Fingerprint Classifiers". In : *Proceedings of the First International Workshop on Multiple Classifier Systems*, pp. 351–361, Springer-Verlag, London, UK, 2000.

[Card 02] J.-F. Cardoso. "Analyse en Composantes Indépendantes". In : *Actes des XXXIVèmes Journées de Statistique (JSBL), Conférence invitée*, Bruxelles, Belgique, 2002.

[CASI 05] CASIA. 2005. CASIA-IrisV3 Database. Available at `http://www.cbsr.ia.ac.cn/IrisDatabase.htm`.

[Chan 05] K. Chang, K. Bowyer, and P. Flynn. "An Evaluation of Multimodal 2D+3D Face Biometrics". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 4, pp. 619–624, 2005.

[Chao 07] W. Chao and Y. Zhong-Fu. "Perceptual Contrast-Based Image Fusion : A Variational Approach". *Acta Automatica Sinica*, Vol. 33, No. 2, pp. 132–137, 2007.

[Chat 99] V. Chatzis, A. Bors, and I. Pitas. "Multimodal decision-level fusion for person authentication". *IEEE Transactions on Systems, Man and Cybernetics, Part A : Systems and Humans*, Vol. 29, No. 6, pp. 674–681, November 1999.

[Chen 05] Y. Chen, S. Dass, and A. Jain. "Fingerprint Quality Indices for Predicting Authentication Performance". In : *Proceedings of Fifth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pp. 160–170, New York, NY, USA, July 2005.

[Chen 97] K. Chen, L. Wang, and H. Chi. "Methods of combining multiple classifiers with different features and their applications to text-independent speaker identification". *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 11, No. 3, pp. 417–445, 1997.

[Coif 92] R. Coifman, Y. Meyer, and V. Wickerhauser. "Wavelet analysis and signal processing". In : *Wavelets and their Applications*, pp. 153–178, 1992.

[Coif 95] R. Coifman and D. Donoho. "Translation-invariant de-noising". Tech. Rep., Stanford University, 1995.

[Coot 01] T. Cootes, G. Edwards, and C. Taylor. "Active Appearance Models". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, pp. 681–685, June 2001.

[Dass 05] S. Dass, K. Nandakumar, and A. Jain. "A principled approach to score level fusion in multimodal biometric systems". In : *Proceedings of Fifth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pp. 1049–1058, 2005.

[Daug 04] J. Daugman. "How iris recognition works". *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, pp. 21–30, 2004.

[Daug 06] J. Daugman. "Probing the Uniqueness and Randomness of IrisCodes : Results From 200 Billion Iris Pair Comparisons". In : *Proceedings of the IEEE*, pp. 1927–1935, November 2006.

[Daug 07] J. Daugman. "New Methods in Iris Recognition". In : *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, pp. 1167–1175, October 2007.

[Daug 98] J. Daugman. "Combining Multiple Biometrics". 1998. Available at `http://www.cl.cam.ac.uk/~jgd1000/combine/combine.html`.

[Dono 94] D. Donoho and I. Johnstone. "Ideal spatial adaptation by wavelet shrinkage". *Biometrika*, Vol. 81, No. 3, pp. 425–455, 1994.

[Dori 06] B. Dorizzi, S. Garcia-Salicetti, and L. Allano. "Multimodality In Biosecure : Evaluation On Real Vs. Virtual Subjects". In : *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. V–V, 2006.

[Duda 01] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley & Sons, 2001.

[Duge 03] J.-L. Dugelay, J.-C. Junqua, K. Rose, and M. Turk. In : *Workshop on Multimodal User Authentication (MMUA)*, Santa Barbara, CA, USA, 2003.

[Dura 06] O. Durand and N. Morizet. "Fourier-inversion and wavelet-transform methods applied to X-ray reflectometry and HRXRD profiles from complex thin-layered heterostructures". *Current trends in optical and X-ray metrology of advanced materials for nanoscale devices. Symposium P of the E-MRS spring meeting 2005*, Vol. 253, No. 1, pp. 133–137, 2006.

[Edwa 76] A. Edwards. *An Introduction to Linear Regression and Correlation*, Chap. 4, The Correlation Coefficient, pp. 33–46. W. H. Freeman and Co Ltd, April 1976.

[Efro 79] B. Efron. "Bootstrap Methods : Another Look at the Jackknife". *Annals of Statistics*, Vol. 7, No. 1, pp. 1–26, 1979.

[Efro 93] B. Efron and R. Tibshirani. *An introduction to the Bootstrap*. Vol. 57 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, 1993.

[Fan 96] X. Fan and L. Wang. "Comparability of jackknife and bootstrap results : An investigation for a case of canonical correlation analysis". *The Journal of experimental education*, Vol. 64, No. 2, pp. 173–189, 1996.

[FERE 01] FERET. 2001. The FERET Database. Available at `http://www.itl.nist.gov/iad/humanid/feret/`.

[Fiel 87] D. Field. "Relations between the statistics of natural images and the response properties of cortical cells". *Journal of Optical Society of America*, Vol. 4, No. 12, pp. 2379–2394, December 1987.

[Flan 98] P. Flandrin. *Temps-fréquence. Collection traitement du signal*, Hermes Sciences Publications, 2ème édition Ed., June 1998.

[Garc 05] S. Garcia-Salicetti, M. Anouar, L. Allano, and B. Dorizzi. "A Generic Protocol for Multibiometric Systems Evaluation on Virtual and Real Subjects". In : *Proceedings of Fifth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pp. 494–502, 2005.

[Gill 06] J. Gilles. *Décomposition et détection de structures géométriques en imagerie*. PhD thesis, École Normale Supérieure de Cachan, June 2006.

[Giri 04] L. Girin. "Joint matrix quantization of face parameters and LPC coefficients for low bit rate audiovisual speech coding". *IEEE Transactions on Speech and Audio Processing*, Vol. 12, No. 3, pp. 265–276, 2004.

[Golu 04] G. Golub. "The Generalized Eigenvalue Problem". In : *Lectures on Matrix Computation, Ph.D. program of the Dipartimento di Matematica "Istituto Guido Castelnuovo"*, Roma, 2004.

[Gosh 07] A. Goshtasby and S. Nikolov. "Guest editorial : Image fusion : Advances in the state of the art". *Information Fusion*, Vol. 8, No. 2, pp. 114–118, 2007.

[Guo 00] G. Guo, S. Li, and K. Chan. "Face Recognition by Support Vector Machines". In : *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 196–201, 2000.

[Hami 94] J. Hamilton. *Time Series Analysis*. Princeton University Press, January 1994.

[Hamp 86] F. Hampel, P. Rousseeuw, E. Ronchetti, and W. Stahel. *Robust Statistics : The Approach Based on Influence Functions*. John Wiley & Sons, 1986.

[Harm 73a] L. Harmon and B. Julesz. "Masking in visual recognition : effects of two-dimensional filtered noise". *Science*, Vol. 180, pp. 1194–1197, June 1973.

[Harm 73b] L. Harmon and B. Julesz. "The recognition of faces". *Scientific American*, Vol. 229, pp. 70–83, 1973.

[Hayk 05] S. Haykin and Z. Chen. "The Cocktail Party Problem". *Neural Computation*, Vol. 17, No. 9, pp. 1875–1902, 2005.

[Hiet 00] R. Hietmeyer. "Biometric identification promises fast and secure processing of airline passengers". *The International Civil Aviation Organization Journal*, Vol. 17, No. 9, pp. 10–11, 2000.

[Hill 02] P. Hill, N. Canagarajah, and D. Bull. "Image fusion using complex wavelets". In : *Proceedings of the 13th British Machine Vision Conference (BMVC)*, pp. 487–496, Cardiff, U.K., 2002.

[Ho 94] T. Ho, J. Hull, and S. Srihari. "Decision Combination in Multiple Classifier Systems". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 1, pp. 66–75, 1994.

[Hong 99] L. Hong, A. Jain, and S. Pankanti. "Can multibiometrics improve performance ?". In : *Proceedings of the IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 59–64, 1999.

[Hori 06] Y. Hori, M. Kusaka, and T. Kuroda. "A 0.79mm2 29mW Real-Time Face Detection Core". *Symposium on VLSI Circuits Digest of Technical Papers*, pp. 188–189, June 2006.

[Hubb 00] B. Hubbard. *Ondes et ondelettes. La Saga d'un outil mathématique*. Pour la Science, July 2000.

[Hube 81] P. Huber. *Robust Statistics*. John Wiley & Sons, 1981.

[Héra 05] J. Hérault. 2005. Page personnelle du Pr. Jeanny Hérault disponible à l'adresse suivante : `http://www.lis.inpg.fr/pages_perso/herault/`.

[Indo 03] M. Indovina, U. Uludag, R. Snelick, A. Mink, and A. Jain. "Multimodal Biometric Authentication Methods : A COTS Approach". In : *Proceedings of the 2003 Workshop on Multimodal User Authentication (MMUA)*, pp. 99–106, 2003.

[Jain 02] A. Jain and A. Ross. "Learning User-specific Parameters in a Multibiometric System". In : *Proceedings of International Conference on Image Processing (ICIP)*, pp. 57–60, New York, NY, USA, September 2002.

[Jain 04] A. K. Jain and A. Ross. "Multibiometric systems". *Communications of the ACM, special issue on multimodal interfaces*, Vol. 47, No. 1, pp. 34–40, January 2004.

[Jain 05] A. Jain, K. Nandakumar, and A. Ross. "Score normalization in multimodal biometric systems". *Pattern Recognition*, Vol. 38, No. 12, pp. 2270–2285, December 2005.

[Jian 04] Y. Jian, D. Zhang, A. Frangi, and J.-Y. Yang. "Two-Dimensional PCA : A New Approach to Appearance-Based Face Representation and Recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 1, pp. 131–137, January 2004.

[Jone 87] J. Jones and L. Palmer. "An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex". *Journal of Neurophysiology*, Vol. 58, No. 6, pp. 1233–1258, 1987.

[JPEG] JPEG2000. Site de l'EPFL sur le JPEG 2000, `http://sawww.epfl.ch/SIC/SA/publications/FI01/fi-3-1/3-1-page1.html`.

[Kana 73] T. Kanade. "Picture Processing System by Computer Complex and Recognition of Human Faces". In : *Doctoral dissertation, Kyoto University*, November 1973.

[Kana 77] T. Kanade. "Computer Recognition of Human Faces". *Interdisciplinary Systems Research*, Vol. 47, 1977.

[Kell 70] M. Kelly. "Visual identification of people by computer". Tech. Rep. AI-130, Stanford University, Stanford University, Palo Alto, CA, USA, 1970.

[Kien 05] W. Kienzle, G. Bakir, M. Franz, and B. Schölkopf. "Face Detection – Efficient and Rank Deficient". In : *Advances in Neural Information Processing Systems*, 2005.

[Kirb 90] M. Kirby and L. Sirovich. "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, 1990.

[Kitt 98] J. Kittler, M. Hatef, R. Duin, and J. Matas. "On Combining Classifiers". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, pp. 226–239, 1998.

[Koha 95] R. Kohavi. "A study of cross-validation and bootstrap for accuracy estimation and model selection". In : *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pp. 1137–1143, 1995.

[Kove 00] P. Kovesi. "What Are Log-Gabor Filters and Why Are They Good ?". School of Computer Science & Software Engineering, The University of Western Australia, 2000. Available from : `http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/PhaseCongruency/Docs/convexpl.html`.

[Kuma 03] A. Kumar, D. Wong, H. Shen, and A. Jain. "Personal Verification Using Palmprint and Hand Geometry Biometric". In : *Proceedings of Fourth International Conference on Audio- and Video-Based Authentication (AVBPA)*, pp. 668–678, Guildford, U.K., June 2003.

[Lam 95] L. Lam and C. Suen. "Optimal combinations of pattern classifiers". *Pattern Recognition Letters*, Vol. 16, No. 9, pp. 945–954, 1995.

[Lam 97] L. Lam and C. Suen. "Application of majority voting to pattern recognition : an analysis of its behavior and performance". *IEEE Transactions on Systems, Man and Cybernetics, Part A*, Vol. 27, No. 5, pp. 553–568, September 1997.

[Li 95] H. Li, B. Manjunath, and S. Mitra. "Multisensor image fusion using the wavelet transform". *Graphical Models and Image Processing*, Vol. 57, pp. 235–245, 1995.

[Lim 01] S. Lim, K. Lee, O. Byeon, and T. Kim. "Efficient Iris Recognition through Improvement of Feature Vector and Classifier". *ETRI Journal*, Vol. 23, No. 2, pp. 61–70, 2001.

[Lin 97] S. Lin, S. Kung, and L. Lin. "Face Recognition/Detection by Probabilistic Decision-Based Neural-Network". *IEEE Transactions on Neural Networks*, Vol. 8, No. 1, pp. 114–132, January 1997.

[Liu 98] C. Liu and H. Wechsler. "A Unified Bayesian Framework for Face Recognition". In : *Proceedings of the IEEE International Conference on Image Processing*, pp. 151–155, 1998.

[Logo 99] N. Logothetis. "Vision : A window on consciousness". *Scientific American*, Vol. 281, pp. 44–51, 1999.

[Lyen 95] S. Lyengar, L. Prasad, and H. Min. *Advances in Distributed Sensor Technology*. 1995.

[Ma 02] L. Ma, Y. Wang, and T. Tan. "Iris recognition using circular symmetric filters". In : *Proceedings of the 16th International Conference on Pattern Recognition*, pp. 414–417, 2002.

[Mach 08] "Machine Readable Travel Documents (MRTD)". 2008. `http://www2.icao.int/en/mrtd/Pages/default.aspx`.

[Mall 00] S. Mallat. *Une exploration des signaux en ondelettes*. Les Éditions de l'École Polytechnique, September 2000.

[Mall 89] S. Mallat. "A Theory for Multiresolution Signal Decomposition : The Wavelet Representation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pp. 674–693, 1989.

[Marc 80] S. Marcelja. "Mathematical description of the responses of simple cortical cells". *Journal of Optical Society of America*, Vol. 70, No. 11, pp. 1297–1300, 1980.

[Mats 02] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. "Impact of Artificial "Gummy" Fingers on Fingerprint Systems". In : *Proceedings of SPIE : Optical Security and Counterfeit Deterrence Techniques IV*, pp. 275–289, January 2002.

[MBGC 08] MBGC. 2008. Multiple Biometric Grand Challenge (MBGC) Database. Available at `http://face.nist.gov/mbgc`.

[Merr 05] R. Merry. "Wavelet Theory and Applications - A Literature Study". Tech. Rep., Eindhoven University of Technology, 2005.

[Mika 99] S. Mika, G. Ratsch, J. Weston, B. Schölkopf, and K.-R. Müller. "Fisher Discriminant Analysis With Kernels". In : *Neural Networks for Signal Processing IX*, pp. 41–48, 1999.

[Misi 03] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi. *Les ondelettes et leurs applications*. *Traité IC2, série Traitement du signal et de l'image*, Hermes Science Publications, February 2003.

[Moll 04] S. Molla and B. Torrésani. "Determining local transientness in Audio signals". In : *IEEE Signal Processing Letters*, pp. 625–628, 2004.

[Moon 04] Y. Moon, H. Yeung, K. Chan, and S. Chan. "Template synthesis and image mosaicking for fingerprint registration : An experimental study". In : *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 409–412, 2004.

[Mori 07a] N. Morizet. "Wavelet Domain Match Score Level Fusion for Multimodal Biometrics". In : *International Conference on COGnitive systems with Interactive Sensors (COGIS)*, Stanford University, Palo Alto, CA, USA, November 2007.

[Mori 07b] N. Morizet, F. Amiel, I. Dris Hamed, and T. Ea. "A Comparative Implementation of PCA Face Recognition Algorithm". *14th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pp. 865–868, December 2007.

[Mori 08] N. Morizet and J. Gilles. "A New Adaptive Combination Approach to Score Level Fusion for Face and Iris Biometrics Combining Wavelets and Statistical Moments". In : *Proceedings of the 4th International Symposium on Visual Computing (ISVC)*, Las Vegas, Nevada, USA, December 2008.

[Most 77] F. Mosteller and J. Tukey. *Data Analysis and Regression : A Second Course in Statistics*. Addison-Wisley, 1977.

[Nand 05] K. Nandakumar. *Integration of Multiple Cues in Biometric Systems*. Master's thesis, Michigan State University, May 2005.

[Nand 08] K. Nandakumar. *Multibiometric Systems : Fusion Strategies And Template Security*. PhD thesis, 2008.

[Nefi 98] A. Nefian and M. Hayes. "Face Detection and Recognition Using Hidden Markov Models". In : *International Conference on Image Processing (ICIP)*, pp. 141–145, 1998.

[Niko 01] S. Nikolov, P. Hill, D. Bull, and N. Canagarajah. *Wavelets for Image Fusion*, Chap. 1, Wavelets in signal and image analysis, from theory to practice. Kluwer Academic Publishers, 2001.

[NIST 02] "NIST report to the United States Congress. Summary of NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability". November 2002. Available at `ftp://sequoyah.nist.gov/pub/nist_internal_reports/NISTAPP_Nov02.pdf`.

[Noh 05] S.-I. Noh, K. Bae, K. Park, and J. Kim. "A New Iris Recognition Method Using Independent Component Analysis". *IEICE - Transactions on Information and Systems*, Vol. E88-D, No. 11, pp. 2573–2581, 2005.

[Olsh 96] B. Olshausen and D. Field. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". *Nature*, Vol. 381, No. 6583, pp. 607–609, June 1996.

[Olsh 97] B. Olshausen and D. Field. "Sparse coding with an overcomplete basis set : a strategy employed by V1". *Vision Research*, Vol. 37, pp. 3311–3325, 1997.

[Ovia 03] S. Oviatt. "Advances in Robust Multimodal Interface Design". *IEEE Computer Graphics and Applications*, Vol. 23, No. 5, pp. 62–68, 2003.

[Pene 96] P. Penev and J. Atick. "Local features analysis : A general statistical theory for object representation". *Neural Systems*, Vol. 7, No. 3, pp. 477–500, 1996.

[Perl 05] V. Perlibakas. "Face recognition using Principal Component Analysis and Log-Gabor Filters". March 2005.

[Perr 02] F. Perronnin and J.-L. Dugelay. "Introduction à la biométrie - Authentification des individus par traitement audio-vidéo". *Traitement du signal*, Vol. 19, No. 4, 2002.

[Phil 00] P. Phillips, H. Moon, P. Rauss, and S. Rizvi. "The FERET evaluation methodology for face recognition algorithms". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, October 2000.

[Poh 05a] N. Poh and S. Bengio. "How Do Correlation and Variance of Base-Experts Affect Fusion in Biometric Authentication Tasks ?". *IEEE Transactions on Signal Processing*, Vol. 53, No. 11, pp. 4384–4396, November 2005.

[Poh 05b] N. Poh and S. Bengio. "Using Chimeric Users to Construct Fusion Classifiers in Biometric Authentication Tasks : An Investigation". Tech. Rep., 2005.

[Poll 81] D. Pollen and S. Ronner. "Phase relationship between adjacent simple cells in the visual cortex". *Science*, Vol. 212, No. 4501, pp. 1409–1411, 1981.

[Prab 01] S. Prabhakar and A. Jain. "Decision-Level Fusion in Fingerprint Verification". In : *Proceedings of the Second International Workshop on Multiple Classifier Systems*, pp. 88–98, Springer-Verlag, London, UK, 2001.

[Pres 92] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in FORTRAN : The Art of Scientific Computing, 2nd ed.*, Chap. 14.1, Moments of a Distribution : Mean, Variance, Skewness, and So Forth, pp. 604–609. Cambridge University Press, 1992.

[Putt 00] T. Putte and J. Keuning. "Don't Get Your Fingers Burned". In : *Proceedings of IFIP TC8/WG8.8 Fourth Working Conference on Smart Card Research and Advanced Applications*, pp. 289–303, 2000.

[Ross 02] A. Ross and A. Jain. "Fingerprint Mosaicking". In : *Proceedings of the IEEE International Conference on Acoustic Speech and Signal Processing*, May 2002.

[Ross 03] A. Ross and A. Jain. "Information fusion in biometrics". *Pattern Recognition Letters*, Vol. 24, No. 13, pp. 2115–2125, 2003.

[Ross 05] A. Ross and R. Govindarajan. "Feature Level Fusion Using Hand and Face Biometrics". In : *Proceedings of SPIE Conference on Biometric Technology for Human Identification*, pp. 196–204, Florida, USA, March 2005.

[Ross 06] A. Ross, K. Nandakumar, and A. Jain. *Handbook of Multibiometrics*. Springer-Verlag New York, Inc., 2006.

[Rydg 04] E. Rydgren, T. Ea, F. Amiel, F. Rossant, and A. Amara. "Iris features extraction using wavelet packets". *International Conference on Image Processing (ICIP)*, Vol. 2, pp. 861–864, October 2004.

[Sach 84] L. Sachs. *Applied Statistics : A Handbook of Techniques*, p. 253. Springer-Verlag, 1984.

[Sagi 01] B. Sagi, S. Nemat-Nasser, R. Kerr, R. Hayek, D. C.C., and R. Hecht-Nielsen. "A Biologically Motivated Solution to the Cocktail Party Problem". *Neural Computation*, Vol. 13, No. 7, pp. 1575–1602, 2001.

[Sand 02] C. Sanderson and K. Paliwal. "Information fusion and person verification using speech and face information". Tech. Rep. IDIAP-RR 02-33, IDAIP, September 2002.

[Sant 03] M. Santos, G. Pajares, M. Portela, and J. De La Cruz. "A new wavelet image fusion strategy". In : *Proceedings of the First Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, pp. 919–926, Puerto de Andratx, Mallorca, Spain, June 2003.

[Sche 07] T. Scheidat and C. Vielhauer. "Analyzing a multimodal biometric system using real and virtual users". In : *Proceedings of SPIE Security, Steganography, and Watermarking of Multimedia Contents IX*, February 2007.

[Scho 98] B. Schölkopf, A. Smola, and K.-R. Müller. "Nonlinear component analysis as a kernel eigenvalue problem". *Neural Computation*, Vol. 10, No. 5, pp. 1299–1319, 1998.

[Sinh 06] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. "Face Recognition by Humans : Nineteen Results all Computer Vision Researchers should know about". In : *Proceedings of the IEEE*, pp. 1948–1962, 2006.

[Siro 87] L. Sirovich and M. Kirby. "Low-dimensional procedure for the characterization of human faces". *Journal of Optical Society of America*, Vol. 4, No. 3, pp. 519–524, 1987.

[Smac 06] F. Smach, M. Atri, J. Miteran, and M. Abid. "Design of a Neural Networks Classifier for Face Detection". *Journal of Computer Science, Sciences Publications*, Vol. 2, No. 3, pp. 261–268, 2006.

[Snel 03] R. Snelick, M. Indovina, J. Yen, and A. Mink. "Multimodal biometrics : issues in design and testing". In : *Proceedings of the 5th international conference on Multimodal Interfaces (ICMI)*, pp. 68–72, ACM, New York, NY, USA, 2003.

[Snel 05] R. Snelick and M. Indovina. "Large-Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 3, pp. 450–455, 2005.

[Texa 08] "Texas Instruments (TI). Analog Technologies, Semiconductors, Digital Signal Processing". 2008. `http://www.ti.com.`

[Thom 82] D. Thomson. "Spectrum estimation and harmonic analysis". In : *Proceedings of the IEEE*, pp. 1055–1096, 1982.

[Tiss 02] M. L. Tisse, C.-L., L. Torres, and M. Robert. "Person identification technique using human iris recognition". In : *Proceedings of Vision Interface*, pp. 294–299, 2002.

[Tong 00] F. Tong, K. Nakayama, M. Moscovitch, O. Weinrib, and N. Kanwisher. "Response Properties of the Human Fusiform Face Area". *Cognitive Neuropsychology*, Vol. 17, pp. 257–279, 2000.

[Turk 91] M. Turk and A. Pentland. "Eigenfaces for recognition". *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71–86, 1991.

[Verl 99] P. Verlinde, P. Druyts, G. Cholet, and M. Acheroy. "Applying Bayes based classifiers for decision fusion in a multi-modal identity verification system". In : *Proceedings of International Symposium on Pattern Recognition "In Memoriam Pierre Devijver"*, Brussels, Belgium, February 1999.

[Wang 03] Y. Wang, T. Tan, and A. Jain. "Combining face and iris biometrics for identity verification". In : *Proceedings of Fourth International Conference on Audio- and Video-Based Authentication (AVBPA)*, pp. 805–813, Guildford, U.K., June 2003.

[Wild 97] R. Wildes. "Iris recognition : an emerging biometric technology". In : *Proceedings of the IEEE*, pp. 1348–1363, September 1997.

[Wisk 97] L. Wiskott, J.-M. Fellous, N. Krüger, and C. Von Der Malsburg. "Face Recognition by Elastic Bunch Graph Matching". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 775–779, 1997.

[Wood 97] K. Woods, W. Kegelmeyer, and K. Bowyer. "Combination of Multiple Classifiers Using Local Accuracy Estimates". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 4, pp. 405–410, 1997.

[Wu 07] J. Wu and C. Wilson. "Nonparametric analysis of fingerprint data on large data sets". *Pattern Recognition*, Vol. 40, No. 9, pp. 2574–2584, 2007.

[Xu 92] L. Xu, A. Krzyzak, and C. Suen. "Methods of combining multiple classifiers and their applications tohandwriting recognition". *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 22, No. 3, pp. 418–435, 1992.

[Yang 03] F. Yang and M. Paindavoine. "Implementation of a RBF neural network on embedded systems : Real time face tracking and identity verification". *IEEE Transactions on Neural Networks*, Vol. 14, No. 5, pp. 1162–1175, September 2003.

[Yip 02] A. Yip and P. Sinha. "Role of color in face recognition". *Perception*, Vol. 31, pp. 995–1003, 2002.

[Zhan 99] Z. Zhang and R. Blum. "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application". In : *Proceedings of the IEEE*, pp. 1315–1326, 1999.

[Zhu 00] Y. Zhu, T. Tan, and Y. Wang. "Biometric personal identification based on iris patterns". In : *Proceedings of the 15th International Conference on Pattern Recognition*, pp. 801–804, 2000.

[Zven 83] A. Zvenigorosky and J. Milliard. "Approximation de la distribution binomiale par la distribution normale ou par la distribution de Poisson". *Revue de Statistique Appliquée*, Vol. 31, No. 1, pp. 63–73, 1983.