

N° d'ordre : 40181

THÈSE

présentée par

Ahmed CHAARI

pour obtenir les grades de

Docteur de l'Université Lille 1 et Docteur de l'Université de Sfax

Discipline : Automatique, Génie informatique, Traitement du signal et des images

Reconnaissance de visages par réseaux d'ondelettes de Gabor

Soutenue publiquement le 08 décembre 2009 devant le jury composé de :

M. Christophe ROSENBERGER	Professeur à l'ENSICAEN	Rapporteur
Mme Najoua ESSOUKRI BEN AMARA	Professeure à l'ENISO, Université de Sousse	Rapporteuse
M. Mohamed DAOUDI	Professeur à Télécom Lille 1	Examinateur
M. Lionel PREVOST	MCF-HDR à l'Université Paris 6	Examinateur
Mme Amel BEN AZZA	Professeure à SUP'COM, Univ. 7 novembre	Examinatrice
M. François CABESTAING	Professeur à l'Université Lille 1	Co-Directeur
M. Nabil DERBEL	Professeur à l'ENIS, Université de Sfax	Co-Directeur
Mme Dorra SELLAMMI MASMOUDI	MCF-HDR à l'ENIS, Université de Sfax	Co-encadrante

Thèse préparée conjointement au Laboratoire d'Automatique, Génie Informatique et Signal (LAGIS UMR-CNRS-8146) et dans l'unité de recherche Intelligent Control, design and Optimisation of complex Systems (ICOS 02-UR-1101)

*À mes parents, à Hana,
à toute ma famille, à mes amis...*

Remerciements

Je voudrai tout d'abord témoigner mes sincères remerciements et ma profonde reconnaissance à Monsieur François Cabestaing, Professeur à l'Université Lille 1, pour m'avoir accueilli au sein de l'équipe Vision & Image du LAGIS et m'avoir encadré durant la durée de réalisation de ce travail. Monsieur Cabestaing a été toujours présent pour écouter mes propositions et répondre à mes questions. Il m'a orienté constamment vers la bonne direction. Je le remercie beaucoup pour la qualité de son encadrement, son soutien moral, ses conseils et ses qualités humaines. Je lui serai toujours reconnaissant pour l'effort qu'il a investi dans cette thèse.

Mes remerciements s'adressent également à Monsieur Nabil Derbel, Professeur à l'Université de Sfax, de m'avoir permis d'intégrer l'unité de recherche ICOS à l'ENIS. Je le remercie pour son aide précieuse et son encouragement. Je remercie Madame Dorra Sellama-Masmoudi, MCF-HDR à l'Université de Sfax, pour m'avoir proposé le sujet de recherche, pour l'aide et l'encadrement qu'elle m'a apportés durant cette thèse.

Je voudrai remercier Monsieur Christophe Rosenberger, Professeur à l'ENSICAEN, et Madame Najoua Essoukri Ben Amara, Professeur à l'Université de Sousse, qui ont eu la gentillesse d'accepter la tâche importante d'être rapporteurs. Je les remercie pour l'intérêt qu'ils ont porté à mon travail et les conseils avisés qui m'ont aidé à améliorer la version finale du manuscrit.

Je tiens aussi à remercier Madame Amel Ben Azza, Professeure à SUP'COM, Monsieur Mohamed Douadi, Professeur à Telecom'Lille1, et Monsieur Lionel Prevost, MCF-HDR à l'Université Paris 6, d'avoir accepté d'être examinateurs de ma thèse.

Cette thèse n'aurait pas pu aboutir sans le support financier que j'ai reçu sous forme de contrats d'enseignement et sous forme de bourse du gouvernement tunisien, bourse du gouvernement français et bourse EGIDE. Je remercie le personnel du service universitaire

et scientifique de l'IFC-Tunisie, le personnel du service de coopération du MESRST-Tunisie et les responsables de l'équipe VI au LAGIS.

Merci à toutes les personnes avec qui j'ai travaillé en enseignement, je remercie tout particulièrement Monsieur Ludovic Macaire, Professeur à l'Université Lille 1, d'avoir aménagé son emploi de temps de façon à permuter nos heures d'enseignement pour que je puisse me concentrer sur la rédaction de ce manuscrit.

Un grand merci s'adresse à tous les membres de l'équipe VI à Lille et de l'équipe ICOS à Sfax. Je remercie en particulier Monsieur Olivier Losson, MCF à l'Université de Lille 1, pour l'outil BibTeX dont il est responsable. Cet outil m'a beaucoup aidé à gérer mes références bibliographiques. Durant ces années de préparation de thèse, j'ai partagé des moments exceptionnels avec tous les collègues. Je leur exprime ma profonde sympathie et leur souhaite beaucoup de succès. Une pensée spéciale à Christophe Vieren pour nos discussions enrichissantes. Je souhaite beaucoup de courage à tous les collègues doctorants qui n'ont pas soutenu leur thèse encore, en particulier à : Mariam, Yara, Halima, Molka, Imène, Hachem, Tarik, Gaël, Yann, Rostom et Mohamed.

Je ne peux terminer mes remerciements sans exprimer ma sincère gratitude à mes parents, à mes frères et à ma soeur pour leur affection et leurs encouragements. Je remercie du profond de mon coeur ma chérie Hana pour son soutien, son amour, pour ces années de compréhension, de patience et d'efforts communs.

Table des matières

Dédicaces	iii
Remerciements	v
Table des figures	ix
Table des notations	xi
Introduction générale	1
1 Reconnaissance de visages	7
1.1 Introduction	7
1.2 La reconnaissance de visages chez l'Homme	8
1.3 La reconnaissance automatique de visages	10
1.3.1 Évaluation des performances et bases de données	12
1.4 Détection de visages et pré-traitements	15
1.4.1 Détection de visages	16
1.4.2 Techniques de pré-traitement	19
1.5 Méthodes d'extraction de signatures	20
1.5.1 Méthodes géométriques	21
1.5.2 Méthodes statistiques	23
1.5.2.1 Eigenface	24
1.5.2.2 FisherFaces	27
1.5.2.3 Analyse en composantes indépendantes	30
1.5.3 Réseaux de neurones	32
1.6 Méthodes d'identification et de décision	35

1.6.1	Calcul des distances	36
1.6.2	Réseaux de neurones	38
1.6.3	Séparateurs à vastes marges	40
1.7	Conclusion	42
2	Modélisation paramétrique d'une image	45
2.1	Introduction	45
2.2	Modélisation paramétrique de données	47
2.3	Modélisation paramétrique par transformation	49
2.3.1	Transformation de Fourier	50
2.3.2	Transformation en cosinus	53
2.3.3	Transformation en ondelettes	55
2.4	Modélisation itérative de la fonction image	60
2.4.1	RNA pour la modélisation d'une image	61
2.4.2	Réseau d'ondelettes	63
2.4.3	Ondelettes de Gabor et réseau d'ondelettes de Gabor	64
2.4.3.1	Ondelettes de Gabor	64
2.4.3.2	Réseau d'ondelettes de Gabor 2D	67
2.4.4	Paramétrisation par déformation de graphe	68
2.4.4.1	EGM : Elastic Graph Matching	69
2.4.4.2	EBGM : Elastic Bunch Graph Matching	71
2.5	Conclusion	72
3	Reconnaissance de visages par réseaux d'ondelettes de Gabor	73
3.1	Introduction	73
3.2	Principe de l'approche proposée	74
3.3	Pré-traitement	76
3.3.1	Recadrage des images	76
3.3.2	Filtrage	77
3.4	Phase d'enrôlement	79
3.4.1	Initialisation	80
3.4.2	Apprentissage avec plusieurs images	82

3.4.3	Analyse des données d'apprentissage	83
3.4.4	Indice de confiance associé à une ondelette	86
3.4.5	Masque des confiances	88
3.5	Phase d'identification	90
3.5.1	Calcul direct du vecteur poids	90
3.5.2	Projection / reconstruction	91
3.5.3	Calcul des distances	93
3.5.4	Classification et décision	94
3.6	Conclusion	94
4	Résultats, évaluations et discussions	95
4.1	Introduction	95
4.2	Base de données YaleFaces	96
4.3	Protocole expérimental 1	99
4.4	Protocole expérimental 2	102
4.5	Protocole expérimental 3	103
4.6	Performances en terme de temps de calcul	107
4.7	Comparaison avec d'autres méthodes de reconnaissance	109
4.8	Conclusion	111
	Conclusion générale	113
	Bibliographie	117
	Résumé - abstract	135

Table des figures

I.1	Évolution du marché de la biométrie selon le rapport de l'IBG	2
I.2	Pourcentages des revenus associés aux principales techniques de biométrie en 2009 selon l'IBG	3
I.3	Organes utilisés dans la biométrie par analyse morphologique	3
I.4	Nombre de publications sur le domaine de la reconnaissance de visages, référéncées par le moteur de recherche IEEE Xplore	4
1.1	Schéma descriptif des fonctions du système visuel humain (extrait de [V08])	8
1.2	Les aires cérébrales dans le cerveau	9
1.3	Fonctionnement d'un système de reconnaissance automatique de visages .	11
1.4	Courbe ROC (extraite de [Hiz09])	14
1.5	Synoptique du module de détection et de pré-traitement	16
1.6	Image brute avec visages détectés	16
1.7	Imagettes des visages détectées dans l'image brute	17
1.8	Processus de normalisation de visages	20
1.9	Visage de la première personne de la base de données YaleFaces	21
1.10	Visage accompagné de mesures géométriques (image extraite de [Zai08]) .	22
1.11	Mesures géométriques réalisées sur le visage (image extraite de [JK08b]) .	22
1.12	Mesures géométriques utilisées pour l'extraction de signatures	23
1.13	Les différents modules d'un système de reconnaissance de visage basé sur l'ACP (extraite de [MP01])	26
1.14	Projection de deux classes sur un axe ACP et un axe ADF (extraite de [Vis05])	29
1.15	Modèle général du mélange et de la séparation de sources	30
1.16	Modèles utilisés pour la reconnaissance de visages (extraite de [BMS02]) .	31

1.17	Analogie entre le neurone biologique et le neurone formel	33
1.18	RNA de type diabolo pour l'extraction de signatures	34
1.19	RNA discriminant pour la reconnaissance de visages	39
1.20	Séparation de deux classes de données	41
2.1	Synoptique du module d'extraction de signatures	45
2.2	Deux formes de représentation d'une image numérique	49
2.3	Ondelettes mère oscillantes et à support compact	57
2.4	Architecture du RNA pour l'approximation d'une image	61
2.5	Graphes de l'ondelette de Gabor	65
2.6	Allure des champs réceptifs 2D expérimentaux et modélisés (extraite de [Dau85])	66
2.7	Architecture du réseau d'ondelettes de Gabor 2D	67
2.8	Graphe d'EGM rectangulaire sur une image de visage comparée à une image de référence (extraite de [SKC07])	70
3.1	Synoptique général d'un système de reconnaissance de visage	74
3.2	Images originales et images reconstruites par un ROG	74
3.3	Schéma synoptique d'extraction de la signature d'un individu	75
3.4	Synoptique du module de pré-traitement	76
3.5	Recadrage d'une image	77
3.6	Recadrage de toutes les images d'un même individu	77
3.7	Filtrage des images	78
3.8	Intérêt du filtrage des images	79
3.9	Synoptique de la phase d'enrôlement	79
3.10	Précision du modèle en fonction du nombre d'ondelettes	80
3.11	Positions des ondelettes sur l'image avant et après apprentissage	81
3.12	Zoom sur les zones stables et variables du visage pour différentes expressions faciales	84
3.13	Variation de d_x en fonction du numéro d'itération	85
3.14	Variations des positions des centres d'ondelettes particulières durant l'apprentissage	85

3.15	Fonction sigmoïde pour deux valeurs du méta-paramètre τ	87
3.16	Allure d'une fonction $\psi_{\mathbf{p}}^{g'}(x, y)$ avec des valeurs initiales de \mathbf{p} et w	89
3.17	Modèle du visage et masque associé	89
3.18	Images initiales et après projection / reconstruction	92
4.1	Images de l'individu 1 de la base YaleFaces	97
4.2	Image 1 de chaque individu de la base YaleFaces	97
4.3	Synoptique du système de validation de notre méthode	98
4.4	Images utilisées dans le protocole expérimental 1	99
4.5	Distances de corrélation entre différentes images (extraite de [Kru00])	101
4.6	Présentation des images des individus 8 et 13	101
4.7	Images utilisées en mode enrôlement dans le protocole 3	104
4.8	Images utilisées en mode identification dans le protocole 3	105
4.9	Pondération des images par le masque des confiances	106
4.10	Application du masque sur les images	107

Table des notations

(x, y)	Coordonnées d'un pixel dans le domaine spatial
Y	Nombre de lignes d'une image
X	Nombre de colonnes d'une image
$I_{m,n}(x, y)$	Image originale numéro n de l'individu m
$\hat{I}_{m,n}(x, y)$	Image reconstruite numéro n de l'individu m
$J_{m,n}$	Image originale numéro n de l'individu m , sous forme de vecteur colonne
M	Nombre d'individus dans la base de données
N	Nombre d'images originales par individu
β	Base de données de visages, $\beta = \{I_{m,n}(x, y), 1 \leq m \leq M, 1 \leq n \leq N\}$
s_k	Élément k d'une signature (scalaire)
K	Nombre d'éléments dans une signature
\mathbf{s}	Signature (vecteur colonne $\mathbf{s} = (s_1 \ s_2 \ \cdots \ s_K)^T$)
\mathbf{S}	Base de données des signatures
λ	Valeur propre
ν	Vecteur propre
C	Matrice de covariance
$\psi(\cdot)$	Fonction élémentaire (atome d'une transformation)
p	Paramètre d'une ondelette
\mathbf{p}	Vecteur de paramètres d'une ondelette
(t_x, t_y)	Translations d'une ondelette
(d_x, d_y)	Dilatations d'une ondelette
θ	Angle d'orientation d'une ondelette
ω	Poids d'une ondelette

$\psi_{\mathbf{p}}(\cdot)$	Ondelette paramétrée par \mathbf{p}
L	Nombre d'ondelettes ou d'atomes dans une modélisation
Ψ	Ensemble d'ondelettes, $\Psi = \{\psi_{\mathbf{p}_l}, 1 \leq l \leq L\}$
$I^c(x, y)$	Masque des confiances
ROG	Réseau d'ondelettes de Gabor
Var	Variance
τ	Écart type
$\psi_g(\cdot)$	Ondelette de Gabor
$\psi'_g(\cdot)$	Ondelette de Gabor modifiée
$\tilde{\psi}_g(\cdot)$	Ondelette de Gabor duale

Introduction générale

Contexte

Selon la définition de Pierson, la biométrie permet de déterminer l'identité d'un individu par l'analyse mathématique de ses caractéristiques biologiques [Pie07]. C'est donc l'art de reconnaître des individus par leurs traits physiologiques ou comportementaux, à savoir les empreintes digitales, le visage, l'iris, le sourire, etc. Ces traits sont retenus comme caractéristiques biométriques car ils sont très différents d'une personne à une autre et ils présentent l'avantage d'être difficilement falsifiables. En revanche, l'exploitation de ces traits physiologiques ou comportementaux pour identifier une personne nécessite de disposer d'équipements sophistiqués intégrant un savoir-faire très pointu.

La biométrie existe en fait depuis des siècles. Par exemple, dans les échanges commerciaux à Babylone, l'empreinte du pouce servait comme signature caractéristique d'un individu. Au début du XIX^{ème} siècle, Alphonse Bertillon a proposé la première méthode de biométrie qu'on peut qualifier de scientifique : le Bertillonage permettait d'identifier des malfaiteurs grâce à plusieurs mesures physiologiques [Cha01]. Au début du XX^{ème} siècle, on a redécouvert le fait que les empreintes digitales sont des caractéristiques infalsifiables d'un individu. Dès lors, cette relation univoque entre une empreinte digitale et son propriétaire a été largement utilisée par les services de police. Dans les années 80 [Par02], des terminaux d'acquisition des empreintes digitales ont été mis en vente, visant un marché bien plus vaste que celui de la police scientifique. C'est à cette époque que le marché de la biométrie a commencé à voir le jour.

Ce marché, qui ne cesse d'évoluer depuis son apparition, croît à un rythme exponentiel. La figure I.1 montre les prévisions d'évolution du chiffre d'affaires du marché mondial de la biométrie dans les cinq prochaines années. Ces prévisions ont été publiées dans un

rapport édité par l'IBG¹ (International Biometric Group).

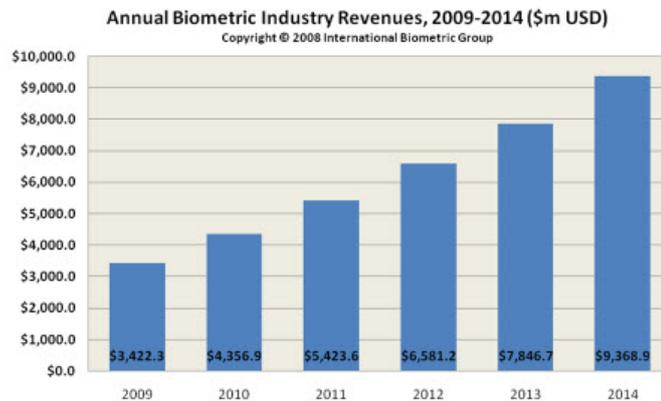


FIGURE I.1 : Évolution du marché de la biométrie selon le rapport de l'IBG

Le développement du marché de la biométrie repose principalement sur les nombreuses innovations technologiques réalisées durant les deux dernières décennies. Ce développement est également rendu possible par des évolutions notables dans les domaines social et juridique. A titre d'exemple en France, la Commission Nationale de l'Informatique et des Libertés² a autorisé pour la première fois, en février 2008, l'utilisation de deux nouvelles technologies biométriques, à savoir la reconnaissance vocale et la reconnaissance du réseau veineux du doigt de la main.

Du point de vue de la répartition du marché, on constate une grande disparité entre les pays. Outre les critères de population et de niveau de développement économique, le volume du marché de la biométrie dépend également des facteurs culturels et psychologiques de chaque société. Les états unis d'Amérique prennent la position de leader dans ce domaine, ce pays représentant à lui seul 40% du marché global, viennent ensuite le marché Japonais et le marché Européen.

S'agissant des techniques utilisées dans les systèmes biométriques opérationnels, les empreintes digitales occupent toujours la première place, soit plus de la moitié du marché mondial. La reconnaissance de visages vient en deuxième position, en occupant 11.4% de ce même marché. La figure I.2 présente l'importance relative des principales technologies biométriques, selon le pourcentage des revenus commerciaux associés.

1. <http://www.biometricgroup.com/>

2. <http://www.cnil.fr/>

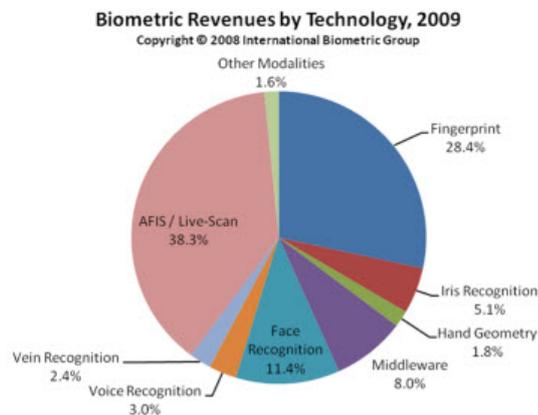


FIGURE I.2 : Pourcentages des revenus associés aux principales techniques de biométrie en 2009 selon l'IBG

Les techniques biométriques actuellement fonctionnelles sont très nombreuses. Elles peuvent cependant être répertoriées dans trois grandes familles :

- l'analyse morphologique : les empreintes digitales, l'iris de l'oeil, la géométrie de la main, les traits du visage, le réseaux veineux de la rétine, etc.
- l'analyse des marqueurs biologiques : l'ADN, le sang, la salive, l'urine, etc.
- l'analyse comportementale : dynamique de frappe au clavier, dynamique de signature, démarche, sourire, etc.

Nous présentons dans la figure suivante les principaux organes utilisés actuellement dans l'analyse morphologique.

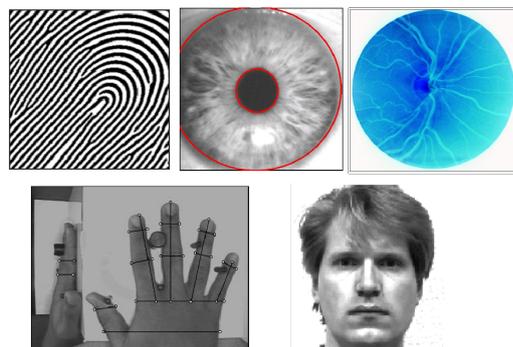


FIGURE I.3 : Organes utilisés dans la biométrie par analyse morphologique

Pour explorer plus en détails la biométrie, ses domaines d'application, ses principaux acteurs, son marché et naturellement les techniques associées, le lecteur peut se référer à

l'ouvrage de Pierson : *La biométrie : l'identification par le corps* [Pie07] et au portail web *Biometrie-Online*³.

Objectif de cette thèse

Comme nous l'avons signalé auparavant, la reconnaissance de visages est classée parmi les principales techniques de biométrie. Depuis les années 80, elle ne cesse de susciter l'intérêt de la communauté scientifique. Cette technique, qui se veut efficace, sans contact et sans aucune coopération de l'utilisateur, représente maintenant un axe de recherche très important pour les groupements de recherche académiques, mais aussi et surtout pour les entreprises privées spécialisées dans la biométrie et la sécurité. La figure I.4 présente l'évolution du nombre de publications sur le thème de la reconnaissance de visages référencées dans la base IEEE Xplore⁴. Les titres des articles comptabilisés contiennent au moins l'un des mots clés suivants : Face Recognition, Face Identification, Face Authentication, Face Representation ou Face Verification.

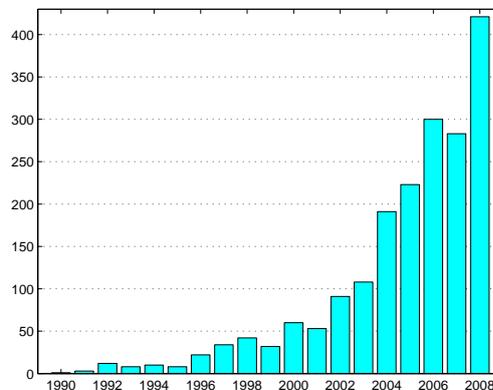


FIGURE I.4 : Nombre de publications sur le domaine de la reconnaissance de visages, référencées par le moteur de recherche IEEE Xplore

Tous les systèmes de reconnaissance de visages qui ont été développés intègrent globalement la même procédure de traitement composée de trois étapes : 1) l'acquisition et le pré-traitement des images ; 2) l'apprentissage et l'extraction de signatures (phase

3. <http://www.biometrie-online.net>

4. <http://ieeexplore.ieee.org>

d' enrôlement appelée aussi *hors-ligne*) ; 3) enfin la reconnaissance et la prise de décision (phase d'identification appelée aussi *en-ligne*).

Plusieurs catégories de méthodes ont été proposées et décrites dans la littérature en vue de reconnaître une image de visage, à savoir les techniques statistiques, les techniques géométriques, ou encore les techniques reposant sur une modélisation tri-dimensionnelle du visage. Parmi les techniques jugées efficaces, figurent en bonne place celles qui modélisent la fonction image, notamment par l'intermédiaire de fonctions élémentaires comme les ondelettes. Dans cette catégorie, les réseaux approchant la fonction image par une combinaison linéaire d'ondelettes de Gabor ont prouvé leur grande efficacité et leur robustesse vis-à-vis des changements qui peuvent survenir lors de l'acquisition des images. En effet, l'apparence de l'image du visage peut être modifiée par des facteurs intrinsèques (grimaces, expressions faciales variées) ou des facteurs extrinsèques (port de lunettes, éclairage, vêtements, etc.).

Les travaux menés dans le cadre de cette thèse concernent le développement d'une nouvelle technique d'extraction de signatures basée sur les réseaux d'ondelettes de Gabor. Cette méthode intervient à la fois dans les phases d' enrôlement et d'identification. Dans la phase d' enrôlement, elle permet de modéliser le visage à partir de plusieurs images correspondant à différentes expressions faciales d'un même individu. Dans la phase d'identification, elle permet de reconnaître de façon fiable une personne grâce à une seule image faciale.

L'ondelette de Gabor a été choisie de part ses caractéristiques particulièrement adaptées à la modélisation d'une fonction image. On peut relever notamment sa grande efficacité quant à la détection de détails et le bon compromis qu'elle réalise concernant la représentation des données dans les domaines spatial et fréquentiel. Ces caractéristiques alliées à l'efficacité des réseaux de neurones en ce qui concerne l'apprentissage ont donné naissance à un nouveau type de réseau adapté aux applications de traitement des images de visages. Ces méthodes ont notamment été utilisées pour le suivi de visages dans une séquence vidéo, la détermination de l'orientation des visages visibles dans une image et la reconnaissance de visages.

Organisation du manuscrit

Ce manuscrit de thèse comporte quatre chapitres et une conclusion, chaque partie présentant plus précisément un volet de nos travaux.

Le chapitre 1 présente un état de l'art des systèmes de reconnaissance de visages exploitant une image de la face de l'individu. Nous décrivons d'abord le système visuel humain, qui s'avère excessivement performant pour réaliser cette tâche, puis arrivons aux systèmes automatiques. Une étude détaillée est menée sur les trois modules constituant la chaîne de traitements intégrés dans les approches standard. Nous présentons les caractéristiques théoriques et pratiques des principales techniques qui ont été développées dans chacun des modules et les meilleurs résultats décrits dans la littérature.

Dans le chapitre 2, nous nous concentrons sur les méthodes paramétriques utilisées dans le module d'extraction de signatures. Nous décrivons les méthodes de calcul directes et itératives des paramètres caractéristiques. Les réseaux de neurones et d'ondelettes, en particulier celles de Gabor, sont présentés ainsi que leur utilisation dans l'application de reconnaissance de visages.

Notre méthode originale d'apprentissage d'un modèle du visage est décrite dans le chapitre 3. La modélisation réalisée durant la phase d'enrôlement et la reconnaissance réalisée durant la phase d'identification sont détaillées, ces descriptions étant accompagnées de justifications de l'approche proposée. La première originalité consiste à exploiter plusieurs images correspondant à différentes expressions faciales du même individu pour réaliser l'apprentissage du réseau. La deuxième réside dans l'association de coefficients de confiance aux ondelettes constituant le modèle de visage. Ces coefficients sont notamment utilisés pour constituer un masque qui définit les parties du visage les plus significatives pour la reconnaissance durant la phase d'identification.

Des résultats expérimentaux calculés sur les images de la base de données YaleFaces⁵ sont présentés dans le chapitre 4. Nous commentons les performances de notre nouvelle technique et tirons des conclusions sur l'utilisation des réseaux d'ondelettes de Gabor pour la reconnaissance faciale.

Un dernier chapitre conclut le mémoire en résumant nos principaux apports et présente les différentes perspectives qu'ouvrent nos travaux.

5. <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>

Chapitre 1

Reconnaissance de visages

1.1 Introduction

Avec l'évolution impressionnante des outils technologiques, notamment en électronique et en informatique, l'Homme se trouve actuellement capable de produire des processus automatiques de très haute performance. Pour développer les systèmes d'intelligence artificielle, les scientifiques s'inspirent généralement du comportement naturel de l'être humain. De nos jours, il arrive parfois que la performance des systèmes artificiels dépasse largement la performance humaine.

Dans le cas de la reconnaissance de visages, les scientifiques ne sont pas encore arrivés à une performance comparable à celle de l'Homme, mais plusieurs pistes sont en cours d'exploration et semblent prometteuses. Différentes études ont été développées sur ce thème de recherche et publiées sur différents supports. On note en particulier le livre de poche écrit par Li et Jain [LJ04] et le site web¹ créé et mis à jour par Grgic et Delac de l'université de Zagreb.

Le sujet a également fait l'objet de plusieurs travaux de recherche portant sur le développement d'une nouvelle technique de reconnaissance de visages, et ce dans le cadre de projets de Master [Zha07, Jou06] ou de thèses de doctorat [Vis05, Sam07, Zai08]. On trouve également des revues bibliographiques qui synthétisent parfaitement la grande majorité des méthodes existantes [ANRS07, ZCRP03].

Les travaux de recherche effectués depuis de nombreuses années ont montré que les

1. <http://www.face-rec.org/>

difficultés rencontrées sont quasiment les mêmes pour toutes les méthodes. Elles résident essentiellement dans la complexité du traitement, vu la multitude de facteurs qui interviennent lors de la prise de vue. Ce traitement se réalise d'une manière très naturelle chez l'Homme et ne demande pas d'effort mental particulier. En revanche, pour une machine, il demande la mise en place d'un ensemble d'algorithmes sophistiqués capables de résoudre les problèmes posés par la grande variabilité de la qualité des images capturées.

Dans ce premier chapitre, nous reviendrons en détails sur les plus importantes méthodes développées pour la reconnaissance automatique de visages, mais auparavant nous proposons une brève description du système humain et de son comportement vis-à-vis de cette tâche mentale particulière.

1.2 La reconnaissance de visages chez l'Homme

Par définition, le système visuel est l'ensemble des organes participant à la perception visuelle, de la rétine aux différentes parties du cortex visuel. Son rôle est de percevoir et d'interpréter les deux images en deux dimensions acquises par nos yeux, mais également l'image en trois dimensions de la scène reconstruite en intégrant des connaissances *a priori*.

En supposant qu'un schéma simple puisse décrire approximativement le système visuel, il comporterait les 5 modules principaux représentés sur la figure 1.1. Ces derniers assurent les fonctions d'acquisition, de traitement et d'interprétation des images.

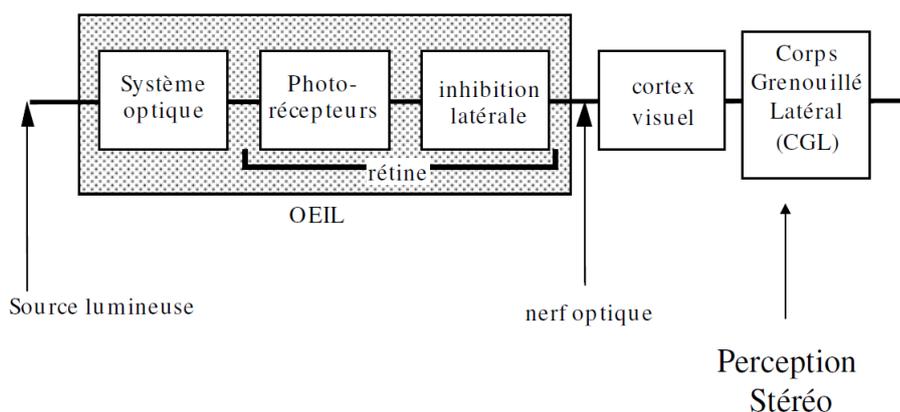


FIGURE 1.1 : Schéma descriptif des fonctions du système visuel humain (extrait de [V08])

La reconnaissance de visages, en tant que tâche indépendante des autres tâches de perception visuelle, constitue une des capacités les plus étonnantes du système visuel

humain. En effet, l'être humain est capable de discriminer et de reconnaître un très grand nombre de visages, même s'ils constituent une catégorie de stimulus homogènes partageant un ensemble de traits similaires et une structure commune (yeux, nez, bouche) [Geo01].

Plusieurs chercheurs de domaines différents — neuropsychologues, neurophysiologistes et neuroscientifiques — ont posé l'hypothèse que la reconnaissance de visages bénéficie d'un traitement particulier dans le cerveau. En effet, ils ont montré que l'observation de visages engendre des réponses cérébrales spécifiques dans des zones qui semblent dédiées à ce traitement.

La figure 1.2 présente les aires cérébrales couramment décrites par les neurophysiologistes. Dans l'aire visuelle, un certain nombre de régions sont particulièrement impliquées dans la perception des visages, notamment l'aire fusiforme des visages (Fusiform Face Area, FFA) qui constitue une partie du gyrus fusiforme au niveau de la jonction des lobes temporaux et occipitaux. Cette zone est responsable de la détection d'un visage et du traitement de ses traits pour reconnaître l'identité de la personne observée. Une lésion cérébrale de ces régions, en particulier de la FFA, peut entraîner une prosopagnosie, c'est à dire une incapacité spécifique à reconnaître les visages tout en conservant des capacités visuelles normales pour les autres types d'objets observés [DHHR07].

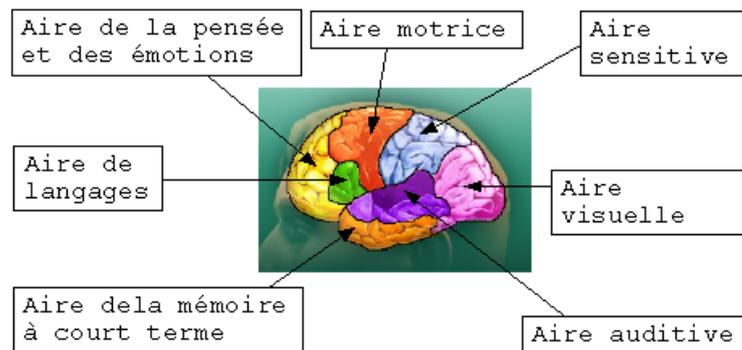


FIGURE 1.2 : Les aires cérébrales dans le cerveau

En se basant sur les caractéristiques des potentiels évoqués visuels (PEV), des études ont montré l'existence d'une composante positive, appelée P2 vertex ou VPP (Vertex Positive Potential), qui apparaît préférentiellement quand l'individu observe un visage plutôt que d'autres catégories de stimulus visuels. Cette composante atteint son maximum d'amplitude avec une latence de 150 à 210 ms [Geo01].

Des études plus récentes ont montré la présence d'une composante négative carac-

téristique de l'observation d'un visage, à savoir la N170 (ce qui indique une latence de 170 ms). Cette composante est supposée refléter des processus spécifiques à l'encodage structurel des visages. L'encodage semble reposer de façon cruciale sur les relations qui existent entre les traits faciaux (traitement configural), contrairement aux traitements de nombreux autres objets, encodés à partir des éléments qui les composent (traitement analytique) [Geo01].

Les latences de ces différentes composantes des PEV sont très faibles, ce qui indique que notre système visuel peut discerner très rapidement, mais pourtant très efficacement, un visage dans une scène même si elle est complexe.

Pour plus de détails concernant ces aspects physiologiques, le lecteur peut se référer au livre de Barbeau *et coll.* "*Traitement et reconnaissance des visages : du percept à la personne*", qui décrit les automatismes de reconnaissance de visages dans le cerveau ainsi que le codage neural correspondant [BJF09].

1.3 La reconnaissance automatique de visages

Tout comme le système visuel humain, un système automatique de reconnaissance de visages doit intégrer une étape d'apprentissage durant laquelle il associe l'allure du visage à l'identité d'une personne. Cette étape est réalisée chez les êtres humains d'une façon spontanée et évolutive. Dans un système artificiel, cette étape permet de construire une base de données des personnes connues, stockant des images étiquetées des identités.

Pour ce faire, un système automatique comporte deux modes de fonctionnement : un mode enrôlement et un mode identification. Le premier mode sert à extraire pour chaque personne les éléments caractéristiques et les met sous la forme d'un vecteur caractéristique, appelé par la suite *signature*. Cette dernière, associée à une étiquette d'identité, sera stockée dans une base de données dédiée. Le mode d'identification permet de reconnaître une personne à partir de son image faciale, c'est à dire de retrouver l'identité associée à l'image.

Nous présentons dans la figure 1.3 une illustration générale d'un système de reconnaissance de visages dans ces deux modes de fonctionnement. Les méthodes appliquées dans les modules formant le système dépendent essentiellement de la technique d'acqui-

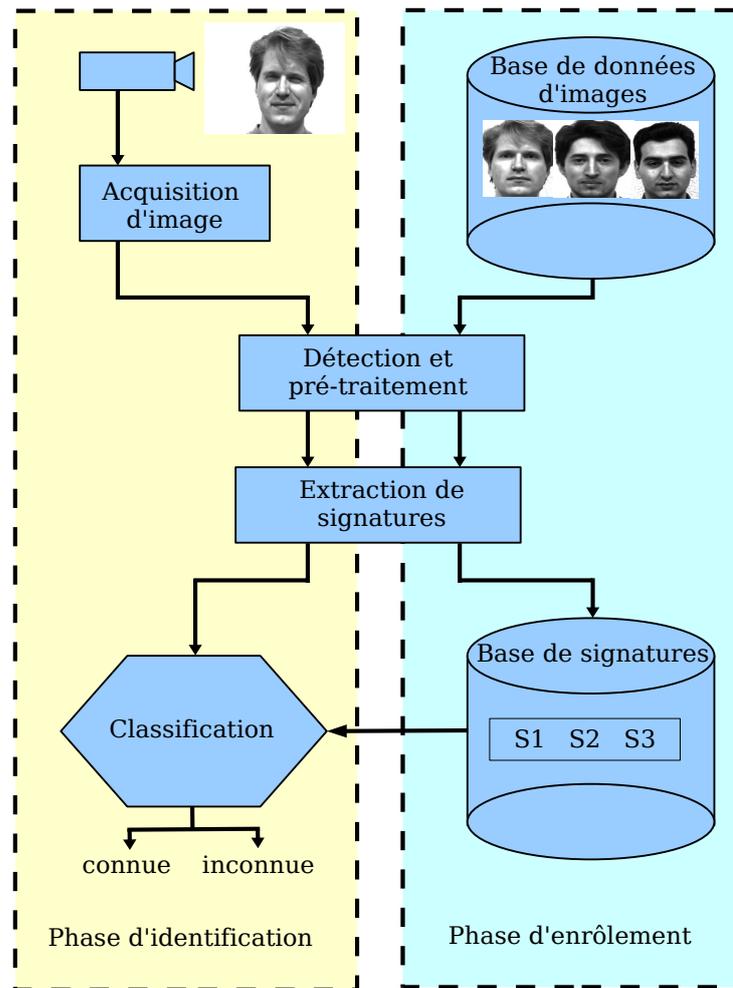


FIGURE 1.3 : Fonctionnement d'un système de reconnaissance automatique de visages

tion d'images de visage. En effet, s'il s'agit d'un scanner tridimensionnel ou d'un système d'acquisition stéréoscopique, la détection de visages et l'extraction de signatures reposent sur des techniques de traitement 3D [Sam07]. Cette thématique est actuellement en pleine expansion. Elle met en évidence l'information de profondeur qui enrichit les données utilisées aussi bien dans la phase d'enrôlement que dans la phase d'identification. Toutefois, elle est limitée par le prix élevé de l'appareil d'acquisition (s'il s'agit d'un scanner 3D) et la difficulté d'installation (s'il s'agit d'un système stéréoscopique). Cette limite laisse le traitement 2D des images de visages plus accessible et encore très exploré, tant dans le domaine académique que dans le milieu industriel. Les travaux de recherche menés dans cette thèse portent sur l'utilisation des images 2D de visages. Le système comporte deux modules qui appliquent les mêmes traitements dans les deux modes de fonctionnement.

Le premier module permet de détecter le visage dans l'image brute captée par le système d'acquisition, il fournit ainsi au second module une image contenant seulement le visage. À partir de celle-ci, le second module extrait la signature discriminante. Elle sera soit stockée soit utilisée pour la classification. En mode hors-ligne (enrôlement), elle sera étiquetée et stockée dans une base de données dédiée. En mode en-ligne (identification), elle servira comme entrée à un troisième module s'occupe de la classification de cette signature et de fournir la décision finale : personne *inconnue* ou personne *connue* et *qui*.

Nous présentons dans ce chapitre les principales méthodes proposées dans la littérature pour constituer chaque module. Nous décrivons également leur utilisation selon le mode de fonctionnement du système et leurs performances. Concernant ce dernier point, avant de passer à la présentation de l'état de l'art, nous décrivons rapidement les critères de performance utilisés dans la littérature et les principales bases de données sur lesquelles les tests sont réalisés.

1.3.1 Évaluation des performances et bases de données

Afin de mesurer les performances d'un système de reconnaissance de visages, les scientifiques ont établi un certain nombre de règles communes permettant de disposer des mêmes critères d'évaluation. Ces critères s'appliquent sur des bases de données également communes et partagées par l'ensemble de la communauté scientifique. Une base de données regroupe plusieurs images de plusieurs personnes. L'ensemble des images d'une personne est dénommé classe ou individu. Il se compose de N images prises généralement dans des conditions différentes. Une base regroupe M individus.

Pour le module de détection, le critère le plus important est le taux de détection, c'est à dire le pourcentage de visages correctement détectés dans une image complexe. Le deuxième critère de performance est le taux de faux positifs, c'est à dire le taux de zones marquées dans l'image alors qu'elles ne contiennent pas de visage. Un système de détection de visages est jugé fiable s'il donne un taux de détection élevé et un taux de faux positifs faible.

Selon Garcia et Delakis, les bases de données les plus utilisées pour l'évaluation des méthodes de détection sont CMU et MIT [GD04]. CMU est la base constituée par l'Insti-

tut de Robotique² au sein de l'Université Carnegie Mellon³. Elle contient 41.368 images de 68 personnes. CMU-125, version compacte de la base mère CMU, est également très couramment utilisée pour la validation et les tests d'une nouvelle méthode. La base de données MIT a été développée dans le centre CBCL⁴ (Center for Biological & Computational Learning) au sein du Massachusetts Institute of Technology, elle contient plusieurs images de 10 personnes. MIT-20 est la version compacte de cette base.

Pour les modules d'extraction de signatures et de classification, on trouve plusieurs bases de données d'images de visages qui ont été construites pour mesurer les performances générales d'un système de reconnaissance de visages. Les critères d'évaluation d'un tel système sont le taux de fausses acceptations (FAR : False Accept Rate) et le taux de faux rejets (FRR : False Reject Rate). Ce sont les critères utilisés dans les évaluations des prototypes participant au concours FRVT⁵ (Face Recognition Vendor Tests) dans le cadre du programme FRGC (Face Recognition Grand Challenge).

À partir de ces deux mesures, un score de moitié des taux d'erreurs totales (HTER : Half Total Error Rate) est calculé selon la moyenne du FAR et du FRR. Ce score reflète les performances globales d'un système de reconnaissance de visages. En agissant sur le seuil d'acceptation et de rejet, on peut tracer une courbe qui représente les variations du FAR en fonction du FRR. La figure 1.4 présente un exemple de cette courbe appelée courbe ROC (Receiver Operation Characteristic). Le point particulier qui correspond à un FAR et un FRR égaux est appelé EER (signifiant Equal Error Rate). Le seuil choisi correspond à la valeur d'EER. Il permet ainsi de définir le HTER pour les valeurs de FAR et FRR correspondantes [Hiz09].

Toutefois, on peut évaluer un système par des critères plus simples qui sont le taux de reconnaissance et le taux d'erreur du classifieur. Le premier présente le pourcentage d'individus correctement reconnus sur la totalité des images testées, le second est le pourcentage d'erreurs de reconnaissance commises sur toute la base.

L'interprétation de ces approches d'évaluation varie selon l'application de système. En effet, un système de reconnaissance de visages peut être utilisé selon deux applications,

2. <http://www.ri.cmu.edu/>

3. <http://www.cmu.edu>

4. <http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database>

5. <http://www.frvt.org/>

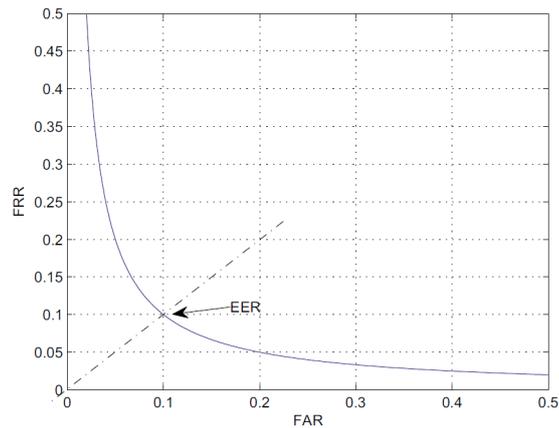


FIGURE 1.4 : Courbe ROC (extraite de [Hiz09])

à savoir pour l'identification d'individu ou la vérification d'identité. Dans la première application, le système doit fournir l'identité d'un individu à partir de son image faciale. Cette application demande un parcours complet de la base de données de signatures. Des seuils d'acceptation et de rejet sont généralement utilisés dans ce cas pour délivrer la première décision : *connue* ou *inconnue*. Si c'est *connue*, le système s'appuie sur une technique de classification pour déterminer l'identité. Dans la deuxième application, le système vérifie la compatibilité d'une identité donnée par l'utilisateur et son image faciale captée à l'instant. Le système n'a pas besoin dans ce cas de parcourir toute la base de signatures, mais plutôt utilise des critères de validation définis *a priori*.

Les bases de données les plus célèbres concernant la partie reconnaissance de l'application sont FERET⁶, YaleFaces⁷, AT&T⁸ (anciennement ORL) et AR⁹.

La base de données FERET (Facial Recognition Technology) a été développée dans le cadre d'un programme de recherche visant à établir une référence pour la validation des techniques de reconnaissance de visages. L'acquisition des images a été opérée au sein de l'Université George Masson¹⁰ dans un environnement semi-contrôlé et durant plusieurs années. Cette base est considérée parmi les plus anciennes et les plus volumineuses, du fait qu'elle regroupe 14.126 images de 1.199 individus.

Développée dans le cadre d'un projet de reconnaissance de visages au sein du centre

6. <http://face.nist.gov/colorferet/>

7. <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>

8. <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>

9. <http://cobweb.ecn.purdue.edu/>

10. <http://www.gmu.edu/>

SVRG¹¹ (Speech, Vision and Robotics Group) de l'Université de Cambridge¹², la base AT&T contient 10 images de 40 individus. Les conditions de prise de vue sont différentes d'une image à une autre, en revanche elles ont en commun le fond et la résolution (92×112).

La base de données YaleFaces contient 165 images en niveaux de gris de 15 individus (soit 11 images par individu). L'acquisition des images a été opérée au centre CVC (Center for Computational Vision and Control) de l'université Yale dans un environnement entièrement contrôlé. Les images contiennent des variations de luminosité et d'expressions faciales, elles ont été acquises toutes sur un fond blanc et avec une résolution de 320×243 .

La base de données de visages AR a été créée par Martinez et Benavente dans le centre CVC¹³(Computer Vision Center) à l'Université Autonome de Barcelone¹⁴, elle contient 3.120 images en couleur correspondant aux visages de 120 personnes (65 hommes et 55 femmes). La prise de vue a été effectuée selon deux modes d'éclairage. Chacun comporte une acquisition de 7 images avec variations des expressions faciales et 6 images avec occultations. Le fond et la résolution (255×255) étaient fixes quelles que soient les conditions de prise de vue.

1.4 Détection de visages et pré-traitements

Le module de détection de visages permet de traiter l'image brute afin de fournir en sortie une image du visage isolé du reste de la scène et prête à être traitée. Il comporte deux sous-modules : un pour la localisation et un autre pour le pré-traitement. La figure 1.5 présente un exemple d'entrée et de sortie de ce premier module de la chaîne de traitement.

L'image d'entrée comporte un nombre indéfini de visages visibles à l'avant d'un fond qui peut être complexe. À la sortie, le module doit séparer chaque visage dans une *image* normalisée selon un format standard défini au préalable.

11. <http://mi.eng.cam.ac.uk/milab.html>

12. <http://www.cam.ac.uk/>

13. <http://www.cvc.uab.es/>

14. <http://www.uab.es>

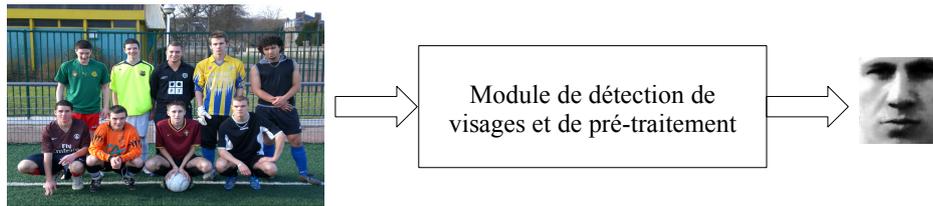


FIGURE 1.5 : Synoptique du module de détection et de pré-traitement

1.4.1 Détection de visages

La détection de visages dans l'image est un traitement indispensable avant la phase de reconnaissance. En effet, le processus de reconnaissance de visages ne pourra jamais devenir intégralement automatique s'il n'a pas été précédé par une étape de détection efficace. Le traitement consiste à rechercher dans une image brute la position des visages et de les extraire sous la forme d'un ensemble d'images afin de faciliter leur traitement ultérieur.



FIGURE 1.6 : Image brute avec visages détectés

Selon la définition proposée dans [GD04], un visage est considéré correctement détecté si la taille d'image extraite ne dépasse pas 20% de la taille réelle de la région faciale, et qu'elle contient essentiellement les yeux, le nez et la bouche. Toutefois, après la détection de visage, un élargissement de fenêtre peut être opéré pour que l'image extraite prenne en compte toute la surface de la tête. Elle sera ensuite affinée par un pré-traitement qui sera décrit dans le paragraphe 1.4.2.



FIGURE 1.7 : Imagettes des visages détectées dans l'image brute

Les problèmes rencontrés dans ce processus sont liés principalement aux conditions de prise de vue. Il s'agit essentiellement de la variabilité concernant l'orientation du visage, de l'occultation des quelques parties de ce dernier, ou encore des changements d'expression faciale et/ou d'éclairage de la scène.

Au cours de notre étude bibliographique, nous avons constaté que les solutions proposées pour traiter cette étape et résoudre les éventuels problèmes se regroupent en quatre catégories : des solutions basées sur des règles, des solutions basées sur des caractéristiques invariantes, des solutions basées sur la mise en correspondance et des solutions basées sur des techniques d'apprentissage automatique [Ngu05, YKA02].

Dans la première catégorie, c'est à dire celle des méthodes exploitant des règles, Yang et Huang proposent dans [YH94] un système de détection de visages qui se base sur une technique de recherche conditionnée. Chaque région candidate est considérée comme un visage uniquement si elle répond parfaitement à des règles. Parmi les règles utilisées, on peut citer par exemple :

- la partie centrale du visage possède des valeurs d'intensité uniformes,
- la différence entre la moyenne des valeurs d'intensité de la partie centrale et celle de la partie supérieure est importante,
- la détection des yeux, du nez et de la bouche,
- la vérification de la symétrie des yeux.

D'autres méthodes utilisent comme règles des contraintes imposées sur les distances entre les organes du visage. On trouve en particulier le calcul des distances entre les yeux, le nez et la bouche [SRH03]. L'application des règles se fait généralement de façon hiérarchique pour garantir la bonne détection et réduire le taux de faux positifs.

D'autres techniques appliquent des projections horizontale et verticale de l'image, en vue de simplifier l'information visuelle [KP97]. La projection sert surtout à la recherche

des sourcils, des yeux, du nez, etc. Cette technique n'est pas très efficace pour la détection de plusieurs visages sur une image de fond complexe.

Les solutions basées sur les caractéristiques visent à trouver les propriétés structurelles invariantes du visage. Même si les conditions de prise d'images changent, ces propriétés ne varient pas. Les caractéristiques souvent utilisées sont la forme du visage, la texture et la couleur de peau [SP96].

La mise en correspondance consiste à comparer chaque partie de l'image avec un modèle de visage enregistré dans une base de connaissances. Cette technique utilise des modèles de visage à de multiples résolutions et intégrant d'éventuelles déformations [Sid07].

Les méthodes basées sur des techniques d'apprentissage automatique créent des modèles variés et les utilisent pour trouver les visages dans les images. Ces méthodes traitent la détection comme un problème de classification et séparent les voisinages extraits de l'image en deux classes : *visage* et *non-visage*. Plusieurs techniques ont été proposées dans cette catégorie, notamment les réseaux de neurones artificiels [MWJI08, Row99, SL07] et les séparateurs à vaste marge [LYYL08].

Outre ces quatre catégories, on trouve également des méthodes hybrides qui utilisent conjointement — simultanément ou successivement — deux techniques de catégories différentes. [MWJI08] présente l'utilisation d'une méthode d'apprentissage (réseaux de neurones) associée à la méthode des règles (la sélection, peau ou non-peau, s'effectue selon la texture ou la couleur). Cette méthode hybride extrait les caractéristiques de chaque région candidate par une transformée en cosinus puis la classifie par un réseau de neurones artificiels.

La même technique de sélection selon la couleur a été utilisée par Lin *et coll.*, mais avec une classification par les séparateurs à vastes marges [LYYL08]. Cette technique de classification sera détaillée dans le paragraphe 1.6.3 de ce chapitre. Le principe de base utilisé dans [LYYL08] est presque identique à celui utilisé dans [MWJI08], à part que le premier utilise une couche de classification supplémentaire. Cette couche améliore la performance de la méthode initiale pour les taux de détection et de faux positifs qui passent respectivement de 88.26% à 96.37% et de 5.81% à 3.63%. Ces résultats ont été obtenus sur 1.817 images couleur arbitraires de tailles variées et contenant 2.615 visages. Elles provenaient de sources différentes : téléchargées du web, scannées, acquises par un

appareil photo numérique, etc.

Nous récapitulons dans le tableau 1.1 les meilleurs résultats obtenus sur les bases de données standard présentées dans la section 1.3.1.

	CMU	CMU-125	MIT	MIT-20	CMU+MIT
[GD04]	90.3%	90.5%	90.1%	90.2%	90.2%
[JK08a]	–	–	–	–	94.7%

TABLE 1.1 : Taux de détection de la méthode de convolution de réseaux de neurones sur les bases CMU et MIT

La méthode proposée dans [GD04] est basée sur une architecture de convolution opérée par des réseaux de neurones, il s’agit de réseaux comportant 6 couches qui reçoivent une image d’entrée de taille 32x36 pixels et qui la classifie dans la classe des *visages* ou dans la classe des *non visages*. Les 4 premières couches alternent une opération de convolution avec une opération de sous-échantillonnage, la cinquième couche contient des neurones dont les fonctions d’activation sont des sigmoïdes afin de fournir à la dernière couche la décision de classification.

Jang et Im proposent une structure intégrant des algorithmes évolutifs placés en cascade et en amont d’une architecture d’apprentissage AdaBoost (Adaptive Boosting). Cette structure permet de supprimer les classifieurs peu performants tout en conservant une bonne qualité de détection. La réduction du nombre des classificateurs permet également de diminuer le temps de calcul [JK08a].

1.4.2 Techniques de pré-traitement

La phase de pré-traitement permet de préparer l’image de visage de telle sorte qu’elle soit exploitable dans la phase d’extraction de signatures. On l’appelle aussi phase de normalisation puisqu’elle ramène à un format prédéfini toutes les images extraites de l’image brute. Elle consiste généralement en une rotation du visage, un centrage et une élimination des zones non informatives.

Pour garantir la bonne performance du système de reconnaissance de visages, il est important que toutes les images soient de taille identique, à la même échelle et au même format concernant les couleurs (par exemple, les images couleur sont parfois

converties en niveaux de gris). Ceci améliore incontestablement le fonctionnement de l'étape d'extraction de signatures et par conséquent la qualité de cette dernière.

Sur certains systèmes, on remarque l'application d'un algorithme d'égalisation de l'histogramme dans l'étape de pré-traitement [RCM07]. Cet algorithme permet d'harmoniser la répartition des niveaux de luminosité de l'image, de manière à tendre vers un même nombre de pixels pour chacun des niveaux de gris de l'histogramme [BJL03].

Prenons le cas du visage numéro 4 dans la figure 1.7, le pré-traitement consiste à ce que l'imagette extraite de l'image originale soit parfaitement alignée sur les normes fixées par les processus d'extraction de signatures et de reconnaissance. Elle subit donc une rotation de -8 degrés, un découpage de la zone non informative, un redimensionnement, un changement de format (couleur vers niveaux de gris) et enfin une égalisation de l'histogramme.

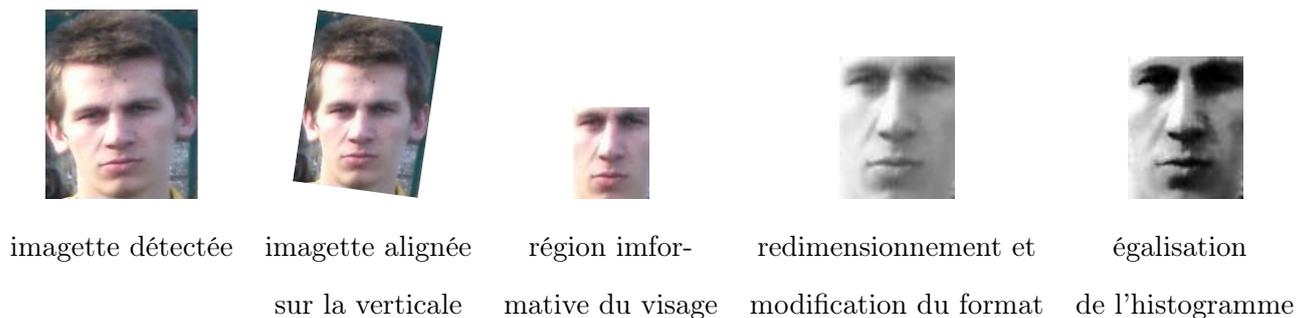


FIGURE 1.8 : Processus de normalisation de visages

1.5 Méthodes d'extraction de signatures

Comme nous l'avons souligné précédemment, l'extraction de signatures est une phase d'importance cruciale dans un système de reconnaissance de visages. On entend ici par signature l'ensemble des caractéristiques propres à un visage. Cette dernière doit être très différente d'une personne à une autre et invariante pour les images d'une même personne. La définition d'une signature est un problème très complexe, du fait que les images d'une même personne peuvent être différentes en fonction des expressions faciales et/ou des différentes conditions de prise de vue.

Les facteurs modifiant l'apparence de l'image du visage d'une même personne sont qualifiés d'intrinsèques ou d'extrinsèques. Les facteurs intrinsèques sont des facteurs in-

dépendants du contexte dans lequel évolue la personne. Il s'agit essentiellement des changements de son expression faciale lors de la prise de vue, et ce pour différentes raisons : les émotions, le bavardage, l'ouverture de la bouche, etc. Les facteurs extrinsèques sont des facteurs ayant un effet sur l'apparence du visage et indépendants de l'état émotionnel de la personne. Ces facteurs peuvent correspondre à un masquage de traits importants du visage, par des lunettes ou un cache-col, ou à une modification des conditions de prise de vue, par exemple lorsque la source d'éclairage est modifiée. La figure 1.9 présente trois différentes conditions de prise de vue qui fournissent des images très différentes d'une même personne.

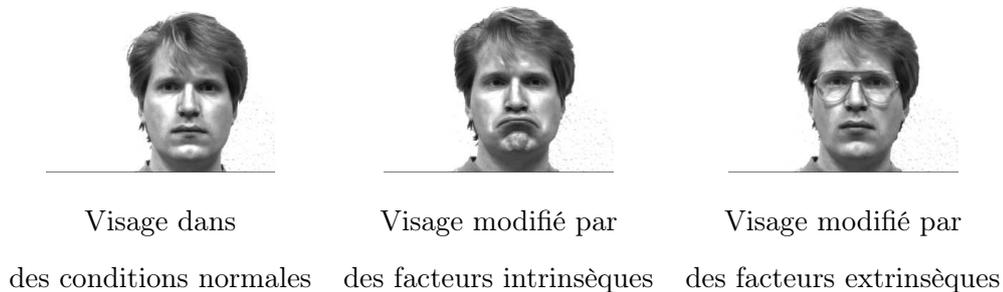


FIGURE 1.9 : Visage de la première personne de la base de données YaleFaces

Dans la bibliographie scientifique, les méthodes d'extraction de signatures se regroupent en trois grandes familles : les méthodes géométriques, les méthodes statistiques et les méthodes basées sur un modèle. Nous présentons dans ce chapitre les principales techniques employées dans les deux premières familles. La troisième famille sera détaillée dans le chapitre suivant de ce manuscrit.

1.5.1 Méthodes géométriques

Les méthodes géométriques se basent généralement sur la mesure des traits distinctifs du visage, de leur position dans l'image et de leurs positions les uns par rapport aux autres. Parmi les mesures importantes, on trouve notamment : la position et la largeur du nez, la position et la largeur de la bouche, l'épaisseur des sourcils et leurs positions verticales, etc.

Pour traiter plusieurs personnes, on stocke les mesures prises sur les images de chaque personne pendant la phase d'enrôlement. Elles seront comparées à celles prises sur l'image de test pendant la phase d'identification. La classification et la décision sont généralement

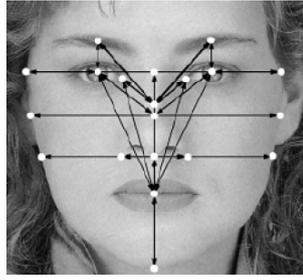


FIGURE 1.10 : Visage accompagné de mesures géométriques (image extraite de [Zai08])

réalisées selon le degré de ressemblance maximale entre les deux séries de mesures. En effet, deux visages se ressemblent si la distance calculée entre les deux ensembles de mesures est minimale. Bien évidemment, si toutes les distances calculées sont supérieures à un seuil, le système considère que la personne est *inconnue*.

Il existe plusieurs techniques de prises de mesures, mais leur point commun est qu'elles se concentrent toutes sur la mesure de trois organes importants, à savoir : les yeux, le nez et la bouche. Ben Jemaa et Khanfir extraient par exemple un vecteur de 8 mesures, correspondant notamment à des distances entre les organes [JK08b] :

$$V = [D_{center-eye}, D_{eye}, D_{interior-eye}, D_{nose}, D_{eye-nose}, D_{mouth}, D_{nose-mouth}] \text{ (cf. figure 1.11)}$$

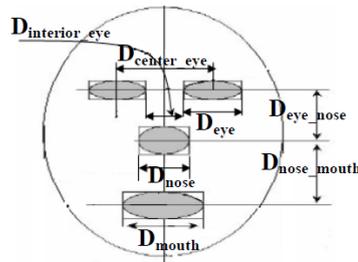
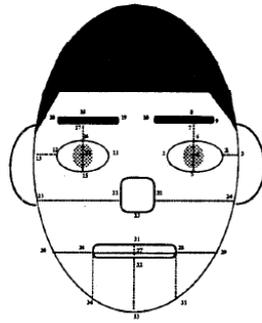


FIGURE 1.11 : Mesures géométriques réalisées sur le visage (image extraite de [JK08b])

D'autres techniques utilisent des mesures plus nombreuses afin de constituer la signature d'un visage. Dans [CGY96] et [BP93], les auteurs proposent d'utiliser jusqu'à 35 mesures. La figure 1.12 montre la répartition de ces indices géométriques et des mesures associées sur le visage.

En dépit de sa rapidité de calcul et de son utilisation très réduite d'espace mémoire, l'approche géométrique reste peu performante du fait qu'elle est très sensible à la présence de facteurs extrinsèques de modification de l'image du visage. Les comparaisons menées



Mesures utilisées dans [CGY96]

(extraite de [CGY96])



Mesures utilisées dans [BPT92]

(extraite de [BPT92])

FIGURE 1.12 : Mesures géométriques utilisées pour l'extraction de signatures

sur les mêmes données en opposant une méthode géométrique à d'autres méthodes ont montré clairement les limitations et les faiblesses de la première approche vis-à-vis des conditions de prise de vue. Brunelli et Poggio ont déduit cette conclusion en opposant leur méthode géométrique à une technique modulaire [BP93].

Plus récemment, [JK08b] a confirmé cette conclusion. Il a montré que la méthode géométrique (cf. figure 1.11) associée à une classification par réseau de neurones donne 84.2% de taux de reconnaissance sur une base de 47 personnes, alors que la méthode d'extraction de signatures par transformée en ondelettes de Gabor associée à la même architecture de réseau de neurones pour la reconnaissance atteint un taux de reconnaissance de 99.4%.

1.5.2 Méthodes statistiques

Une image 2D numérique est un tableau dans lequel sont stockées les valeurs des luminosités de pixels selon leurs coordonnées *ligne* et *colonne*. Grâce à cet arrangement, on peut tenir compte des valeurs associées aux pixels, mais également des corrélations spatiales existant entre ces valeurs. Dans certains cas, on peut cependant considérer l'image comme une donnée sans s'intéresser directement à son contenu visuel. Dès lors, une image initialement représentée par un tableau de Y lignes et X colonnes est ramenée à un vecteur regroupant $X \times Y$ valeurs scalaires.

Plusieurs techniques statistiques ont été employées afin d'extraire une signature des images faciales en considérant une telle représentation des données. Elles consistent à

appliquer une transformation sur les données disponibles pour chaque personne, cette transformation permettant de conserver certaines propriétés de la distribution initiale des valeurs des pixels. Dans la plupart des cas, la transformation vise à réduire la dimension de l'espace contenant les vecteurs représentant une image de visage. Ainsi, la signature associée à un visage est simplement le nouveau vecteur résultant de la transformation appliquée à la donnée initiale.

On distingue deux familles de traitements statistiques : les méthodes linéaires et les méthodes non linéaires. Pour la première famille, le traitement consiste essentiellement à projeter l'espace d'entrée dans un autre espace de plus faible dimension et à ne conserver que les données considérées comme significatives (selon des critères de sélection définis *a priori*). En revanche, la dimension de l'espace final de représentation est fixée *a posteriori*, après la sélection. Quant à la deuxième famille, elle regroupe les méthodes non linéaires qui ont pour but de trouver des espaces de séparation capables de représenter les données sous la forme de classes indépendantes, chaque classe correspondant à un individu. Ces méthodes sont généralement itératives avec une dimension de l'espace de représentation fixée *a priori*.

Nous présentons dans ce manuscrit les méthodes les plus connues de ces deux familles, à savoir l'Eigenface (basée sur l'ACP, Analyse en Composantes Principales), FisherFaces (basée sur l'ADL, Analyse Discriminante Linéaire) et l'ACI (Analyse en Composantes Indépendantes) pour la famille linéaire, puis les réseaux de neurones pour la famille non-linéaire. Cette dernière technique est également utilisée dans le module de classification et de décision.

1.5.2.1 Eigenface

Eigenface, introduite par Turk et Pentland en 1991 [TP91], est l'une des premières techniques fonctionnelles de reconnaissance de visages. Elle repose sur l'analyse en composantes principales (ACP) d'un ensemble d'images de visage stockées dans une base de données suffisamment complète.

L'ACP consiste à rechercher une transformation linéaire de l'espace initial dans lequel sont représentées les données aboutissant à un nouvel espace dans lequel l'information est concentrée sur un nombre limité de directions orthogonales. L'ACP est connue pour

son efficacité en terme de réduction de dimensions et de compression de données tout en préservant le maximum d'informations utiles [AVE97, Liu04, RS08].

Pour déterminer la transformation linéaire optimale, on recherche les axes principaux d'inertie du nuage de points constitué par les données représentant tous les visages de la base. Pour ce faire, on calcule les vecteurs propres associés aux valeurs propres maximales de la matrice de covariance de l'ensemble des données disponibles dans la base. La matrice de covariance intègre l'information concernant la variabilité des données représentatives de tous les visages disponibles.

Plus précisément, la transformation est déterminée de la façon suivante. La structure bidimensionnelle initiale de l'image n'étant pas prise en compte explicitement dans cette méthode, les images $I(x, y)$ sont tout d'abord transformées en vecteurs colonne $J = [I(1, 1) \cdots I(x, y) \cdots I(X, Y)]^T$, à $X \times Y$ composantes. On calcule ensuite la moyenne \bar{J} de toutes les images de la base, selon l'expression :

$$\bar{J} = \frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N J_{m,n} ,$$

dans laquelle $J_{m,n}$ est l'image (vecteur colonne) numéro n de l'individu numéro m de la base, M et N désignant respectivement le nombre total d'individus dans la base et le nombre d'images disponibles pour chaque individu. Chaque image est ensuite normalisée en lui soustrayant cette image moyenne : $\hat{J}_{m,n} = J_{m,n} - \bar{J}$. Toutes les images normalisées sont regroupées afin de constituer une matrice de taille $(X \times Y) \times (M \times N)$ contenant toutes les données image :

$$A = \left[\hat{J}_{1,1} \ \hat{J}_{1,2} \ \cdots \ \hat{J}_{1,N} \ \hat{J}_{2,1} \ \cdots \ \hat{J}_{m,n} \ \cdots \ \hat{J}_{M,N} \right] .$$

On calcule ensuite la matrice de covariance inter-pixels qui permet de mesurer la dispersion des vecteurs image autour de leur moyenne (nulle du fait de la normalisation préalable) :

$$C = \frac{1}{M \times N} A \cdot A^T . \quad (1.1)$$

L'étape suivante consiste à calculer les vecteurs propres ($\mathbf{v} = [\nu_1 \ \nu_2 \ \cdots \ \nu_{X \times Y}]$) de C et à les classer selon l'ordre décroissant de leurs valeurs propres ($\lambda_1 > \lambda_2 > \cdots > \lambda_{X \times Y}$). Il faut noter que le rang de la matrice C est inférieur au nombre total $M \times N$ d'images disponibles et qu'il existe donc des méthodes efficaces de calcul évitant de traiter inutilement des matrices de très grande dimension [MP01].

Afin de déterminer la transformation linéaire qui permet de calculer la signature d'une image, on conserve enfin les K vecteurs propres correspondant aux plus grandes valeurs propres. Il faut souligner que dans cette procédure la dimension K de l'espace des signatures est choisie *a posteriori*.

Lors de la phase de reconnaissance, le traitement consiste à projeter l'image d'entrée J sur l'espace des signatures. Cette projection est réalisée après normalisation de l'image, c'est à dire après lui avoir soustrait la moyenne \bar{J} . La comparaison de la signature obtenue avec toutes les signatures des images de la base permet de déterminer si l'image correspond ou non à un individu connu. Cette partie du processus de reconnaissance sera décrite dans la section 1.6.1.

L'implantation de la méthode Eigenface a été testée sur différentes architectures. [MAHE07] a établi une comparaison entre 3 implémentations différentes, à savoir sur une architecture séquentielle de type PC, sur un processeur numérique de signal (*DSP-Digital Signal Processor*) et sur un circuit logique programmable (*FPGA-Field Programmable Gate Array*). Les performances ont été mesurées sur la base de données FERET, elles ont montré que le DSP offre un bon compromis entre le temps de traitement et la facilité de mise en oeuvre (5ms de temps de traitement sur un DSP 225 MHz).

Différentes méthodes dérivant de Eigenface et exploitant une ACP ont été décrites dans la littérature [MP01, Yam00, RS08]. Moon et Phillips ont proposé dans [MP01] une utilisation générique de l'ACP dans un système modulaire de reconnaissance de visages dont la disposition des 3 modules est décrite dans la figure 1.13.

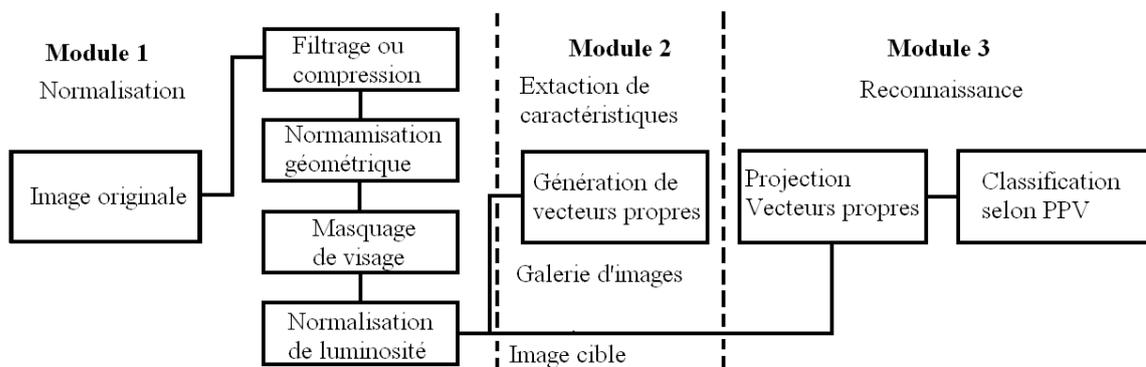


FIGURE 1.13 : Les différents modules d'un système de reconnaissance de visage basé sur l'ACP (extraite de [MP01])

Une deuxième déclinaison de cette architecture modulaire a été utilisée dans [GA04]. Gottumukkal et Asari ont proposé de diviser l'image en sous-images et d'appliquer ensuite l'ACP sur chacune d'entre elles. La reconnaissance d'une image revient alors à calculer la distance entre la signature obtenue par projection de l'image cible et celle obtenue lors de l'apprentissage. Cette architecture a permis d'améliorer le taux de reconnaissance à 88% contre 70% obtenus avec une ACP classique (YaleFaces a été utilisée comme base de données de test).

La méthode Eigenface a également été améliorée par l'introduction d'une version bidimensionnelle de l'ACP, qui préserve mieux les propriétés structurelles des images. Pour ce faire, Yang a proposé de supprimer l'étape de transformation de l'image $I(x, y)$ en vecteur J (indispensable pour Eigenface) en décrivant une technique de traitement de l'image sous sa forme matricielle [YZFyY04]. Cette méthode a été validée sur deux bases de données (AT&T et AR) et a prouvé son efficacité, non seulement pour le taux de reconnaissance mais également pour l'optimisation du temps de calcul (cf. tableau 1.2).

	ACP	ACP2D
AT&T	93.5 % / 304.61 s	96.0 % / 14.03 s
AR	94.7 % / 130.42 s	96.1 % / 07.25 s

TABLE 1.2 : Comparaison entre ACP et ACP2D sur les bases AT&T et AR [YZFyY04] (*Taux de reconnaissance / Temps d'exécution par individu*)

1.5.2.2 FisherFaces

FisherFaces est une méthode similaire à l'Eigenface dans le sens qu'elle vise à réduire la taille de l'espace contenant les images afin de constituer des signatures. Cependant, elle tend à construire un sous-espace plus approprié pour différencier les différents individus connus dans la base initiale [EC97].

En effet, Eigenface tend à déterminer une transformation permettant de réduire la dimension de n'importe quelle image de visage afin de constituer une signature. Durant la phase d'enrôlement, toutes les images disponibles dans la base sont traitées simultanément et ceci indifféremment des individus auxquels elles sont associées.

En revanche, dans la méthode FisherFaces, l'étiquette associée à chaque image de la

base, c'est à dire l'individu qu'elle représente, est exploitée lors de la phase d'enrôlement. Les images de la base sont regroupées en *classes*, chacune regroupant toutes les images correspondant à un même individu. La méthode vise à déterminer une transformation linéaire pour chaque classe de la base. Dans l'espace résultant de cette transformation, les projections des visages de la classe seront clairement séparées des projections des visages correspondant à d'autres individus.

L'analyse discriminante de Fisher (ADF) est utilisée pour déterminer les transformations associées à toutes les classes de la base [ZCK98, ZKC⁺98]. Son principe consiste à projeter les vecteurs J représentant les images sur un axe particulier de l'espace initial. Après projection sur un axe, chaque vecteur se trouve transformé en une valeur unique. L'ADF consiste à rechercher l'axe sur lequel toutes les valeurs correspondant à une classe de visages (donc à un unique individu) sont à la fois proches les unes des autres et le plus éloignées possible des valeurs résultant de la projection sur le même axe de toutes les autres images.

La comparaison entre l'analyse discriminante linéaire et l'analyse en composantes principales a fait l'objet de très nombreuses études de recherche tant théoriques que pratiques. Parmi les comparaisons simples et explicites, on trouve l'étude du cas binaire, clairement illustré par la figure 1.14. Sur ces données en deux dimensions, l'ACP détermine l'axe principal d'inertie (vecteur propre avec la plus grande valeur propre) qui permet de conserver le maximum d'information après réduction de dimension. Ici, la réduction de dimension par l'ACP consiste à projeter les données initiales sur l'axe bleu. En revanche, l'ADF permet de déterminer l'axe qui sépare au mieux les données des deux classes après projection, ici l'axe représenté en rouge. Sur les données projetées, maintenant mono-dimensionnelles, on constate aisément qu'il est plus facile de différencier les deux classes avec l'ADF qu'avec l'ACP.

En terme de performance pour la reconnaissance de visages, [MK01] présente une comparaison directe entre les deux techniques. Les résultats expérimentaux obtenus sur la base de données de visages AR montrent que globalement la technique de l'ADF est mieux adaptée au problème de reconnaissance de visages. Rappelons tout de même que cette base contient 4000 images couleur pour 126 personnes.

Différentes variantes de l'ADF ont été développées afin d'améliorer la performance et

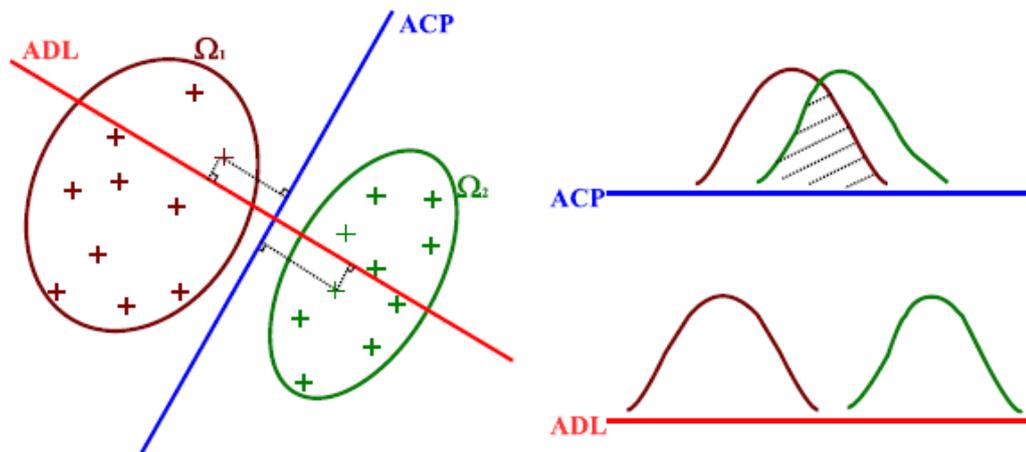


FIGURE 1.14 : Projection de deux classes sur un axe ACP et un axe ADL (extraite de [Vis05])

l'efficacité de cette méthode pour la reconnaissance de visages. On trouve par exemple l'analyse fractionnelle (FLDA : Fractional-step Linear Discriminant Analysis) inspirée de [LK00] et appliquée par [LPV03]. Cette étude compare deux variantes de l'ADF, à savoir l'ADF fractionnelle et l'ADF directe. La FLDA permet de trouver une direction de projection même quand le nombre de données de la base est inférieur au nombre de dimensions de l'espace (problème connu sous le nom de "malédiction de la dimension").

Le tableau 1.3 présente les performances de l'Eigenface, de FisherFaces et de la méthode FLDA sur les images de la base de données AT&T.

Méthodes	Eigenface	FisherFaces	FD-LDA
Taux de reconnaissance	74.18%	38.51%	80.03%

TABLE 1.3 : Résultats obtenus sur la base de données AT&T

Visani a présenté dans [VGJ05] et dans [VGJ04] des améliorations du concept de l'ADF inspirées de l'ACP-Bidimensionnelle de [YZFyY04]. Elle décrit dans [VGJ04] les bases d'une analyse bidimensionnelle linéaire orientée et dans [VGJ05] le concept d'une analyse discriminante bilinéaire pour la représentation et la reconnaissance de visages. Pour plus de détails sur ces deux techniques, se référer au manuscrit de sa thèse de doctorat [Vis05].

1.5.2.3 Analyse en composantes indépendantes

Les techniques de réduction de dimension présentées auparavant peuvent également être considérées comme visant à résoudre un problème inverse. Ce problème se formule de la façon suivante. Supposons que chaque individu est effectivement caractérisé par une signature unique \mathbf{s} et qu'il existe une transformation directe (et inconnue) F qui permet de calculer une image J de son visage à partir de cette signature. La variabilité de l'image du visage peut être modélisée en considérant que du bruit se superpose soit à la signature, soit au résultat de la transformation de cette signature. Retrouver la signature à partir du visage consiste donc à déterminer l'inverse de la transformation, d'où l'appellation problème inverse.

Dans ce formalisme, chaque élément s du vecteur signature \mathbf{s} est considéré en tant que source d'information [Bor04]. Dans ce cas, la transformation directe évoquée précédemment est appelée *fonction de mélange*, car elle combine en fait linéairement les sources d'information afin de constituer le vecteur J accessible à l'observation en tant qu'image. Le problème inverse est celui de l'estimation des sources à partir des observations. Cette modélisation du problème est représentée par le synoptique de la figure 1.15, dans lequel F désigne la fonction de mélange et H la fonction de séparation qui permet d'obtenir une estimation des sources (ici la signature) à partir des observations (ici les images).

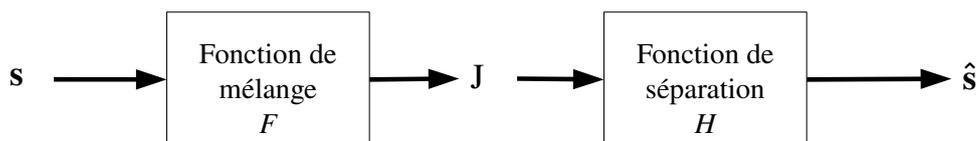


FIGURE 1.15 : Modèle général du mélange et de la séparation de sources

L'analyse en composantes indépendantes (ACI) a été introduite par les spécialistes du traitement de signal afin de trouver une solution au problème de séparation des sources lorsque la fonction de mélange F est inconnue. Le traitement consiste à extraire les composantes linéaires d'une observation multivariée afin qu'elles soient aussi indépendantes que possible. Outre le traitement de signal, cette technique a été utilisée dans d'autres domaines, en l'occurrence les télécommunications et le traitement des signaux biomédicaux. Elle sert généralement à analyser les signaux issus de multiples capteurs pour lesquels la

nature exacte des sources est inconnue, d'où vient son appellation de séparation aveugle de sources [Car02].

Dans le cas considéré en reconnaissance de visages, la fonction de séparation H permet d'obtenir une estimation \hat{s} de la signature d'une image J selon $\hat{s} = H(J)$. Pour déterminer la fonction H , comme pour l'ACP, on part de la matrice de covariance C des images de la base calculée en utilisant l'équation 1.1. Ensuite, plutôt que de diagonaliser cette matrice comme dans l'ACP, on cherche à la factoriser sous une forme qui autorise que les axes autour desquels se concentre l'information ne soient pas forcément orthogonaux [Com94, LW99]. Cette factorisation, obtenue de façon itérative par des algorithmes variés, détermine conjointement la fonction de séparation H .

Cette approche a été utilisée par [BMS02] pour la reconnaissance de visages. Il propose deux modèles différents de mélange des sources aboutissant à la formation de l'image J à partir des signatures (cf. figure 1.16). Dans chaque cas, il utilise une analyse en composantes indépendantes pour déterminer les fonctions de séparation. Les deux méthodes ont été testées sur la base de données FERET et ont donné de bons résultats (88% de taux de reconnaissance). Les tests effectués sur les bases YaleFaces et AT&T ont donné également des résultats satisfaisants (respectivement 96.36 % et 99.00 % de taux de reconnaissance) et ainsi prouvé l'efficacité de la méthode [KCA02].

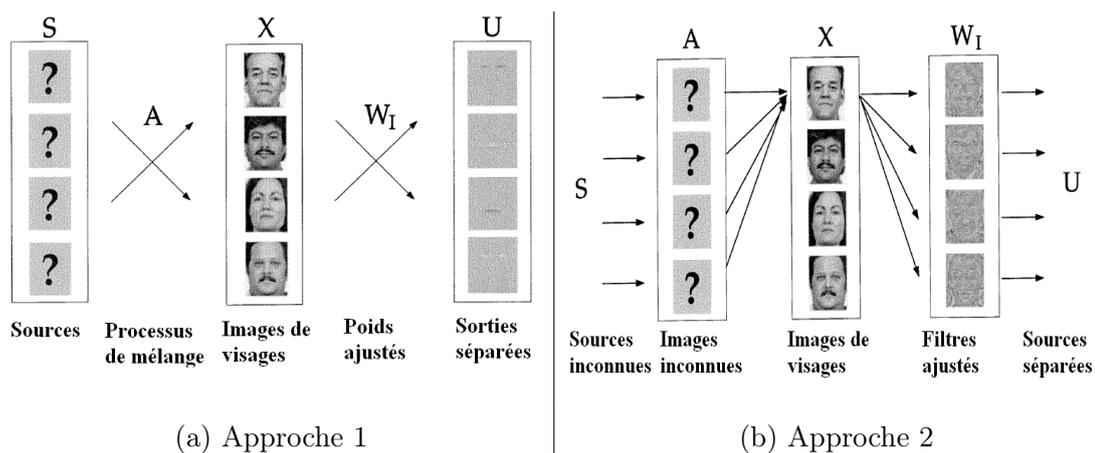


FIGURE 1.16 : Modèles utilisés pour la reconnaissance de visages (extraite de [BMS02])

Il faut également noter que de nombreuses variantes de l'algorithme d'ACI ont été proposées récemment. On trouve en particulier la technique rapide (Fast ICA) [Mc08, LWL07], ou aussi FICA (FLD ICA) : la combinaison entre l'analyse en composantes indépendantes et l'analyse discriminante de Fisher [KP07]. Ces améliorations ont permis de

remédier au problème du temps de calcul observé avec l'ACI classique. On trouve également des variantes basées sur la façon de traiter l'image originale, telle que la technique basée sur l'ACI par blocs [ZGD07]. Il s'agit ici de subdiviser l'image en ensemble de blocs et d'appliquer une ACI à chacun.

En terme de comparaison, il a été démontré que l'ACI, ainsi que ses différentes variantes, est très performante par rapport à l'ACP [LHZL08]. En effet, les composantes recherchées dans l'espace de représentation des images ne sont pas forcément orthogonales pour l'ACI, alors qu'elles le sont pour l'ACP.

1.5.3 Réseaux de neurones

Parmi les techniques non-linéaires d'extraction de signatures qui ont été largement utilisées pour la reconnaissance de visages, on trouve celles qui reposent sur un réseau de neurones artificiels (RNA). Ces derniers ont été initialement inspirés de la physiologie du système nerveux si parfaitement créée et conçue. Le neurone formel, introduit par J. Mc Culloch et W. Pitts dans les années quarante, constitue la base de l'architecture des RNA.

En se basant sur la théorie de l'analogie, le modèle de neurone formel a été introduit pour tenter de reproduire le fonctionnement des cellules neuronales. Dans le système nerveux biologique, on considère que le neurone est l'élément de base assurant à la fois la propagation et le traitement de l'information. Chaque neurone reçoit, au niveau des dendrites attachées à son corps cellulaire via des synapses ou liaisons inter-neurones, des stimulations électriques ou chimiques venant d'autres neurones. Lorsque le neurone a reçu par ce biais un nombre suffisant de stimulations, il transmet à son tour une information vers des neurones voisins, grâce à un influx nerveux qui se propage le long de son axone (cf. figure 1.17(a)).

Un neurone formel est une fonction algébrique non-linéaire (paramétrée et à valeurs bornées) de variables réelles appelées entrées, qui tente de reproduire ce mode de fonctionnement. Un neurone formel effectue une somme pondérée des signaux d'entrée qui lui parviennent. Cette somme pondérée sert de paramètre à une fonction, souvent non-linéaire, qui la transforme en nouveau signal transmis à la sortie (cf. figure 1.17(b)). En référence au fonctionnement du neurone biologique, la fonction intervenant après la

sommation des entrées est appelée fonction d'activation.

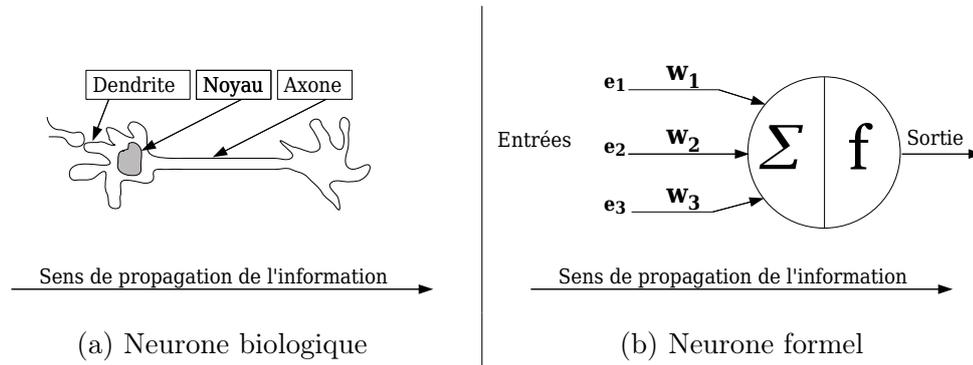


FIGURE 1.17 : Analogie entre le neurone biologique et le neurone formel

Les RNA ont été utilisés dans nombreuses applications, particulièrement pour la classification de données, la modélisation de processus complexes et le traitement non-linéaire des signaux. Ces recherches ont engendré une panoplie d'architectures de réseaux dont chacune répond parfaitement à une application donnée. On note en particulier l'architecture MLP (Multi-Layer Perceptron, Perceptron Multi-Couches), l'architecture RBF (Radial Basis Function, Fonctions à base radiale) et l'architecture SOM (Self-Organizing Maps, Cartes auto organisatrices de Kohonen).

Dans le cas de la reconnaissance de visages, on constate que les réseaux de neurones ont été employés dans tous les modules intervenant dans la chaîne de traitements. Ils sont utilisés pour la détection de visages, pour l'extraction de signatures et pour la classification. Le tableau 1.4 regroupe par module les références bibliographiques que nous avons recensées.

Module	Détection	Extraction de signatures	Classification et décision
Références biblio.	[Row99] [MWJI08] [GD04]	[Wan08] [JK08b] [CSB07]	[Mc08] [SG08] [PSKK05] [LWL07] [ECW05] [SL07]

TABLE 1.4 : Références bibliographiques sur les RNA pour la reconnaissance de visages

Naturellement chacun de ces travaux introduit une originalité scientifique, mais généralement les architectures utilisées se ressemblent et possèdent plusieurs points communs. Nous présentons ici l'architecture diabolo qui est la plus utilisée pour extraire une signature à partir d'un ensemble d'images d'un même individu. Les techniques de classification

reposant également sur un RNA seront décrites dans la section 1.6.2. L'architecture diabololo est présentée sur le synoptique de la figure 1.18.

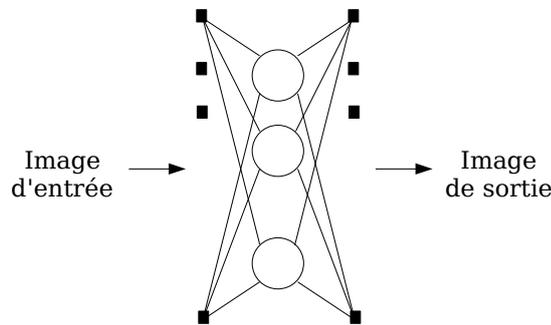


FIGURE 1.18 : RNA de type diabololo pour l'extraction de signatures

Dans son principe, la méthode d'extraction de signatures par un RNA diabololo présente des similarités avec celle mise en oeuvre par les techniques d'ACI présentées auparavant. L'ACI vise à déterminer une fonction de séparation H de telle sorte que la signature estimée (sources reconstruites) soit la plus proche possible de la signature réelle (sources initiales). L'image constitue l'observation, c'est à dire la seule information permettant de déterminer H .

Dans le cas du RNA diabololo, la fonction H est en fait codée dans les paramètres des neurones de la couche cachée. Les sorties de ces mêmes neurones constituent la signature de l'image placée en entrée. La couche de sortie permet de reconstruire une version approchée de l'image aussi proche que possible de celle placée en entrée. Cette couche remplit donc un rôle similaire à la fonction F de mélange des sources. Les neurones de la couche cachée opèrent ainsi une réduction de dimension et ceux de la couche de sortie une reconstruction de l'image à partir de la signature.

Durant la phase d'apprentissage, les paramètres du RNA diabololo sont adaptés afin d'obtenir en définitive une estimation simultanée des fonctions de séparation et de mélange. Pour ce faire, on présente l'une des images d'un individu en entrée et on calcule la différence quadratique entre cette image et celle reconstruite en sortie. La différence quadratique mesure l'erreur entre l'image initiale et l'image reconstruite par le réseau. En ajustant les paramètres du réseau de façon itérative afin de minimiser l'erreur, la plupart du temps par rétro-propagation du gradient de l'erreur, on obtient en définitive un réseau qui peut extraire la signature la plus appropriée pour une classe donnée de visages.

Nous présenterons plus en détail cette approche dans le chapitre 2, principalement pour les méthodes dans lesquelles le RNA est utilisé pour l'approximation de la fonction image afin d'en extraire une signature paramétrique.

1.6 Méthodes d'identification et de décision

Nous décrivons dans cette section le troisième module d'un système de reconnaissance de visages. Après avoir détecté le visage dans l'image et extrait ses attributs caractéristiques, nous présentons ici les méthodes qui permettent de le classifier et de fournir la décision finale : *personne inconnue* ou *personne connue* et *qui*. Les approches proposées dans la littérature pour résoudre ce problème relèvent de la classification automatique de données, champ de recherche largement exploré durant les dernières décennies pour de nombreuses applications.

Les méthodes pouvant être appliquées dans cette étape dépendent principalement de la technique utilisée lors de l'étape d'extraction de signatures. Certaines n'interviennent que dans la phase d'identification, c'est à dire pour reconnaître l'individu auquel appartient le visage. Dans cette catégorie, on distingue notamment les méthodes qui sont basées sur le calcul des distances entre les signatures stockées dans la base (hors-ligne) et celle qui a été calculée sur le visage en cours de traitement (en-ligne). Dans ce cas, les frontières entre classes dans l'espace des signatures sont déterminées *a posteriori*.

D'autres méthodes interviennent d'une part dans la phase d'enrôlement afin de déterminer *a priori*, dans l'espace des signatures, les frontières entre les classes correspondant aux différents individus. Cette phase d'enrôlement est qualifiée d'apprentissage. Il s'agit ici d'un apprentissage supervisé du fait que les étiquettes des visages de la base de données sont connues. Une technique de discrimination associée intervient ensuite durant la phase d'identification, assurant la classification du visage cible à partir des connaissances acquises durant l'apprentissage. Parmi ces méthodes, nous décrivons les deux plus utilisées, à savoir les classifieurs exploitant un réseau de neurones et les séparateurs à vastes marges (SVM).

Il est par ailleurs important de noter que le temps de calcul pour l'étape d'identification doit être minimal. En effet, la réponse concernant l'identification de la personne

cible doit être rapidement disponible pour que le système soit pleinement opérationnel comme tout autre système de contrôle biométrique. Par contre, pour la phase hors-ligne d'apprentissage (enrôlement), le temps de calcul n'est en général pas critique.

1.6.1 Calcul des distances

La méthode de calcul des distances s'applique pour déterminer des degrés de ressemblance. Selon le cas, la distance peut être calculée entre deux signatures, ou entre une signature et un ensemble de signatures. La première option s'applique dans les méthodes qui assignent une signature unique à chaque individu durant la phase d'enrôlement, indépendamment du nombre d'images disponibles pour cet individu. La deuxième option s'applique lorsqu'une signature est assignée à chaque image de la base et non à chaque individu durant la phase d'enrôlement.

Dans le domaine de la reconnaissance de visages, les distances entre deux signatures les plus utilisées sont les suivantes (K désigne le nombre d'éléments du vecteur signature) :

- distance directe (basée sur la norme L_1) :

$$d(\mathbf{s}, \mathbf{s}') = \sum_{k=1}^K |s_k - s'_k| \quad (1.2)$$

- distance euclidienne (basée sur la norme L_2) :

$$d(\mathbf{s}, \mathbf{s}') = \sum_{k=1}^K (s_k - s'_k)^2 \quad (1.3)$$

- mesure de l'angle entre les deux signatures :

$$d(\mathbf{s}, \mathbf{s}') = \frac{\mathbf{s} \cdot \mathbf{s}'}{\|\mathbf{s}\| \|\mathbf{s}'\|} \quad (1.4)$$

En partant de ces différentes expressions, on peut définir simplement la distance entre une signature \mathbf{s} et un ensemble S de signatures par le minimum des distances calculées entre \mathbf{s} et tous les éléments de S :

$$d(\mathbf{s}, S) = \mathbf{inf} \{d(\mathbf{s}, \mathbf{s}') \mid \mathbf{s}' \in S\} . \quad (1.5)$$

La distance de Mahalanobis est également très utilisée pour déterminer la similarité entre une signature et un ensemble de signatures calculées durant la phase d'enrôlement.

Dans ce cas, on utilise les statistiques calculées sur les signatures de l'ensemble S , à savoir la moyenne $\bar{\mathbf{s}}$ et la matrice de covariance C_S , lors du calcul de la distance. De cette façon, on tient compte explicitement de la dispersion dans l'espace des signatures constituant l'ensemble S . La distance de Mahalanobis se calcule selon l'expression suivante :

$$d(\mathbf{s}, S) = \sqrt{(\mathbf{s} - \bar{\mathbf{s}}) \cdot C_S^{-1} \cdot (\mathbf{s} - \bar{\mathbf{s}})^T} . \quad (1.6)$$

Dans une version simplifiée de la distance de Mahalanobis, à savoir la distance euclidienne pondérée, on ne tient compte que des variances des éléments de la signature. Cela revient à considérer que la matrice de covariance C_S est diagonale. En notant σ_k^2 la variance du $k^{\text{ème}}$ élément de la signature, calculée pour toutes les signatures de S , la distance euclidienne pondérée se calcule selon l'expression :

$$d(\mathbf{s}, S) = \sqrt{\sum_{k=1}^K \frac{(s_k - \bar{s}_k)^2}{\sigma_k^2}} . \quad (1.7)$$

Une fois que la distance a été choisie, le problème de la reconnaissance d'un visage cible à partir des signatures stockées dans la base est résolu par la méthode du plus proche voisin. On suppose qu'à chaque individu m connu dans la base on a associé une signature unique \mathbf{s}_m ou un ensemble S_m contenant N signatures (nombre d'images de visage disponibles pour cet individu). On calcule ensuite toutes les distances d_m entre la signature \mathbf{s} du visage cible et les signatures ou ensembles de signatures des individus de la base :

$$d_m = d(\mathbf{s}, \mathbf{s}_m), \text{ ou } d_m = d(\mathbf{s}, S_m) , \quad (1.8)$$

et on considère que le visage cible correspond à l'individu pour lequel la distance d_m est minimale.

Avec cette technique basique, le visage cible est toujours reconnu, car il existe toujours un individu de la base pour lequel la distance est minimale. Pour pallier ce problème, on peut introduire un seuil de rejet, c'est à dire une borne supérieure pour la valeur de la distance d_m . Si la distance minimale déterminée en balayant la base est inférieure à ce seuil, on considère que l'image de visage a été reconnue, sinon on considère qu'elle ne correspond pas à un individu de la base.

La distance euclidienne et la distance de Mahalanobis figurent parmi les plus utilisées pour reconnaître un visage. Elles ont été employées conjointement avec toutes les méthodes d'extraction de signatures proposées dans la littérature, notamment les méthodes

géométriques. Dans ce cas, elles calculent en fait la distance entre les deux vecteurs de mesures obtenus sur les visages [CGY96, BP93].

Ces mêmes distances sont employées dans la méthode Eigenface, après projection de l'image sur l'espace des signatures [MAHE07, MP01, Yam00, GA04, YZFyY04]. Quant à la technique de FisherFaces, la distance euclidienne est la plus utilisée pendant la phase de reconnaissance [EC97, ZCK98, LK00, LPV03, VGJ05, YZFyY04, MK01].

Avec une signature extraite par analyse en composantes indépendantes, la plupart des auteurs utilisent également les distances de Mahanalobis [LW99, KP07] ou euclidienne [BMS02, ZGD07].

Lorsque les signatures sont déterminées par des RNA de type diablo, ce sont les réseaux eux-mêmes qui permettent de calculer implicitement une distance entre signatures [CSB07]. Dans ce cas, durant la phase d'enrôlement, le système a adapté les paramètres d'autant de RNA diablo qu'il y a d'individus différents dans la base de données. Lorsqu'il s'agit d'identifier un visage cible, on présente ce dernier en entrée de tous les RNA diablo. On peut alors déterminer une distance pour chaque RNA en calculant la différence quadratique entre l'image reconstruite par ce dernier et l'image cible. La décision est ensuite prise d'une façon similaire à celle décrite précédemment, en intégrant ou non un seuil pour éliminer les individus inconnus.

1.6.2 Réseaux de neurones

Nous avons précédemment décrit la façon par laquelle un RNA de type diablo peut servir à extraire la signature associée à un visage. Dans cette partie, nous décrivons une autre application des RNA dans le domaine de la reconnaissance de visages, à savoir la classification de données multidimensionnelles grâce à une architecture discriminante. Cette architecture est présentée sur la figure 1.19.

Durant la phase d'enrôlement, un apprentissage permet d'adapter les paramètres du réseau afin qu'il puisse par la suite discriminer les signatures présentées sur ses entrées. Pour ce faire, on procède de façon itérative. On sélectionne un visage de la base d'apprentissage dont on calcule la signature qu'on place en entrée du réseau. On calcule les sorties du réseau pour cette valeur de ses entrées, puis on évalue une fonction d'erreur en considérant que les sorties idéales sont toutes nulles sauf celle qui correspond à la classe

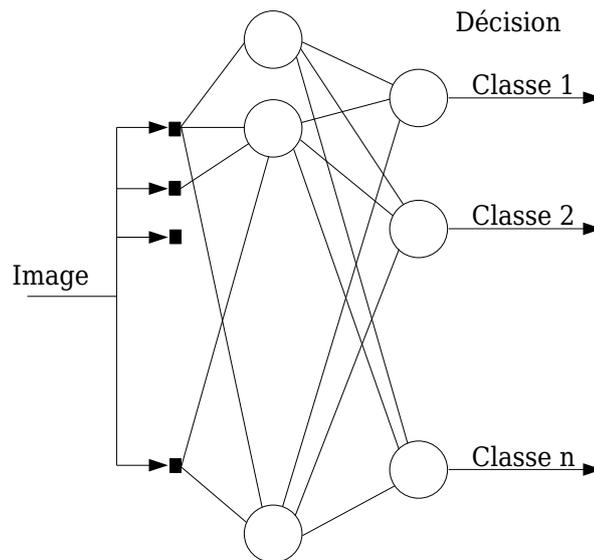


FIGURE 1.19 : RNA discriminant pour la reconnaissance de visages

dont fait partie l'individu. Pour ajuster les paramètres du réseau, on utilise la technique standard de rétro-propagation de l'erreur. Cette procédure est itérée en présentant à l'entrée, plusieurs fois et dans un ordre aléatoire, toutes les signatures extraites des images de la base.

Durant la phase d'identification, on calcule la signature du visage cible et on l'applique en entrée du RNA discriminant. La sortie du réseau qui prend la valeur maximale indique à quelle classe appartient le visage cible. L'introduction d'un seuil sur la valeur de la sortie maximale permet de déterminer si l'image cible ne correspond à aucun individu connu de la base.

Lorsqu'on utilise un RNA discriminant comme technique de reconnaissance, l'ajout d'un individu dans la base de visages nécessite de recommencer tout l'apprentissage. Le temps de calcul est important durant la phase d'enrôlement, en revanche il est très court durant la phase d'identification, du fait qu'il s'agit simplement de calculer la réponse du RNA à un stimulus particulier (ici la signature).

L'architecture discriminante RNA a été utilisée par de nombreux auteurs. Mazloom et Ayat ont développé une architecture multi-couches pour classifier des signatures issues d'une analyse en composantes principales [MA08]. Sabrina et Laskri emploient cette même architecture pour classifier une signature composée de sept éléments, lesquels sont calculés d'après les moments géométriques [SL07]. D'autres chercheurs ont utilisé un RNA discri-

minant dont les fonctions d'activation sont à base radiale (réseau RBF), sur des signatures extraites par une ACI [Mc08, LWL07], par une transformée en ondelettes [PSKK05], ou encore par une transformée en cosinus [ECW05].

1.6.3 Séparateurs à vastes marges

Un SVM (Support Vector Machine, ou Séparateur à Vastes Marges en français) est une méthode de classification supervisée, particulièrement bien adaptée pour traiter des données de grande dimension. Les SVM ont été introduits par Vapnik en 1995 [Vap95], puis ont fait l'objet d'un brevet d'invention déposé en 1997. La technique était initialement développée pour les problèmes de classification binaire où les deux groupes sont linéairement séparables, puis elle a bénéficié d'importantes améliorations pour traiter le cas des données non séparables linéairement et permettre de résoudre les problèmes multi-classes. D'après Guo *et coll.*, les SVM constituent la technique la plus efficace pour l'identification implicite des modèles des distributions sous-jacentes à la répartition des données dans l'espace des attributs [GLC00].

Considérons d'abord le cas binaire, dans lequel on dispose de deux sous-ensembles de données étiquetées supposées appartenir à deux classes distinctes. Un SVM vise à déterminer l'hyperplan optimal (OSH : Optimal Separating Hyperplane) qui sépare les deux classes de points (cf. figure 1.20). L'OSH est calculé de telle sorte que la distance entre ce dernier et les points les plus proches dans chaque classe soit maximale. Lorsque l'OSH a été trouvé, les points les plus proches dans chacune des classes sont appelés *vecteurs support* et la distance entre ces points et l'OSH est appelée *marge*.

Dans l'exemple présenté, les données sont linéairement séparables, c'est à dire qu'il est possible de trouver un hyperplan qui les sépare parfaitement en respectant les classes. L'intérêt des SVM est qu'ils permettent également de traiter les cas plus complexes dans lesquels les classes sont plus imbriquées, en définissant une hypersurface de séparation qui n'est plus un hyperplan. Pour ce faire, grâce à une fonction *noyau*, on augmente artificiellement la dimension de l'espace des attributs de telle sorte que les données représentées dans le nouvel espace soient linéairement séparables. Pour optimiser la position de l'OSH dans ce nouvel espace, on doit calculer des distances donc disposer d'un produit scalaire. Un choix adéquat de la fonction noyau permet de calculer simplement ce produit scalaire

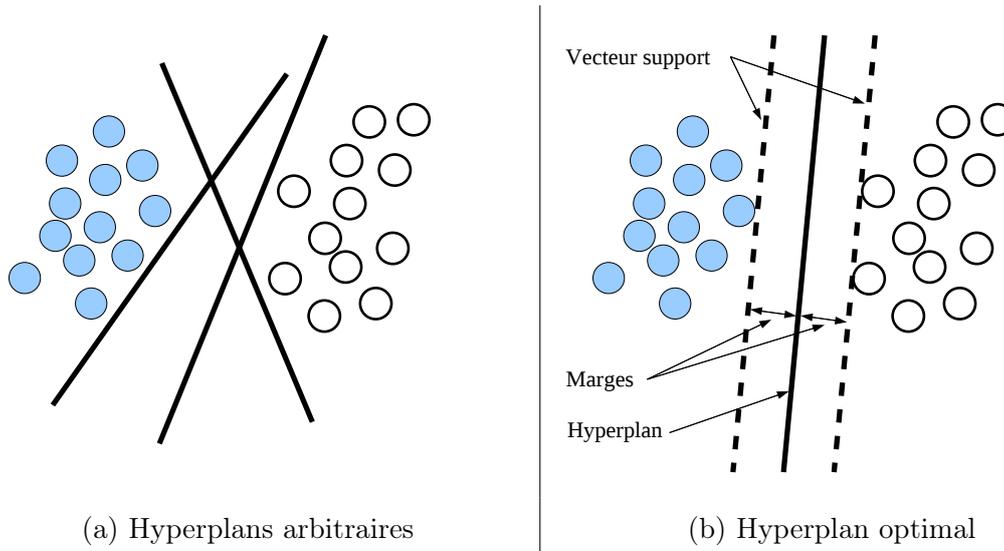


FIGURE 1.20 : Séparation de deux classes de données

de façon implicite (technique connue sous le nom de *kernel trick*).

Le SVM, intrinsèquement binaire, a été étendu au cas multi-classes par la technique du “un contre tous” (one versus all). Cela consiste en fait à déterminer autant d’hypersurfaces de séparation qu’il y a de classes dans le problème. Pour calculer chaque hypersurface, on considère d’une part tous les points de la classe considérée et on regroupe d’autre part toutes les autres données disponibles quelle que soit leur classe. De cette façon, on assimile le problème multi-classes à un ensemble de problèmes binaires.

L’application des SVM à des problématiques liées au traitement des images de visages a été présentée pour la première fois par [OFG97]. Les résultats obtenus ayant montré l’intérêt de cette approche, de nombreux auteurs ont continué à explorer cette piste par la suite. D’autres études ont été publiées dans la foulée, [HSW98] pour la détection de pose sur les images de la base de données FERET, [HHP01] pour la détection de visages dans une image (classification “visage” ou “non-visage” de régions extraites des images).

Si on considère uniquement le traitement réalisé dans le module de reconnaissance, la détermination des hypersurfaces séparant les classes est réalisée durant la phase d’enrôlement. Durant la phase d’identification, un visage cible est assigné à un individu selon la position de sa signature dans l’espace partitionné en régions distinctes par les hypersurfaces SVM [GLC00].

L’efficacité de la discrimination par les SVM a été prouvée par de nombreuses études. Huang a obtenu un taux de reconnaissance de 100% sur les visages de la base FE-

RET [HSW98]. Avec la même méthode, mais sur d'autres protocoles de test (1.355 images issues des bases de données AR, AT&T, YaleFaces, UMIST, Berne University et MPEG-7 News), les taux varient selon le nombre de classes. Pour 10 classes, les SVM atteignent des taux de reconnaissance, de faux négatifs et de faux positifs valant respectivement : 98.89%, 1.85% et 0%. Pour 60 classes, ces taux deviennent respectivement : 95.57%, 2.71% et 0.74% [PKB05].

Des SVM à noyau non-linéaire ont également été testés pour la reconnaissance de visages. On peut notamment citer les travaux de Rioux et Simon, qui ont utilisé un noyau gaussien du fait qu'ils avaient constaté que les signatures n'étaient pas linéairement séparables dans leur application [RS08].

1.7 Conclusion

Dans ce chapitre, nous avons mis en lumière l'ensemble du processus automatique de reconnaissance de visages allant de la détection à la prise de décision. Cette présentation de l'état de l'art a été précédée par une brève description du processus équivalent chez l'être humain.

Nous avons présenté l'architecture globale d'un système de reconnaissance de visages automatique, en mettant systématiquement l'accent sur le fonctionnement distinct des deux phases de traitement, à savoir la phase d'enrôlement et la phase d'identification. Nous avons également présenté le traitement contenu dans chacune des phases avec un état de l'art des méthodes existantes.

Nous avons considéré que la chaîne de traitements permettant la reconnaissance de visages peut être décomposée en trois modules agissant de façon séquentielle : un module de détection, un module d'extraction de signatures et un module de classification et de décision. Ce dernier module n'intervient que dans la phase d'identification, il est remplacé dans la phase d'enrôlement par une simple opération d'extraction et de stockage des signatures. Toutefois, les deux premiers modules doivent être semblables dans les deux phases.

Une description détaillée a été présentée pour chacun des modules avec une revue des principales techniques utilisées. Les synthèses bibliographiques des méthodes ont concerné

aussi bien l'aspect scientifique que méthodologique. Nous avons présenté pour chaque méthode les meilleurs résultats relatés dans la littérature afin de donner au lecteur un aperçu des performances qui peuvent actuellement être atteintes.

Parmi ces trois modules, celui chargé de l'extraction des signatures a une influence notable sur la performance globale d'un système de reconnaissance automatique. Cela est dû à la complexité des données dont on dispose d'une part (en effet un visage regroupe un ensemble d'éléments visuels à la fois très semblables au niveau de l'aspect général et très discriminants au niveau des détails) et à la diversité des techniques utilisées dans ce module d'autre part.

Ce module doit donc viser à déterminer une signature qui n'inclut que les éléments discriminants. Pour choisir une signature il faut également réaliser un compromis entre sa performance discriminante et le temps nécessaire à son calcul, notamment durant la phase d'identification. Les méthodes paramétriques décrites dans le chapitre suivant sont reconnues pour satisfaire correctement ce compromis.

Chapitre 2

Modélisation paramétrique d'une image

2.1 Introduction

Dans le chapitre précédent, nous avons montré qu'extraire la signature caractéristique d'une image de visage est une tâche particulièrement importante dans tout système de reconnaissance. La signature constitue l'unique caractéristique de la personne, qui servira à la reconnaître durant la phase de classification (mode identification). Les signatures extraites des images doivent être semblables pour une même personne et très différentes d'une personne à une autre. Le défi dans cette étape est donc d'obtenir une signature indépendante des facteurs intrinsèques et extrinsèques qui peuvent modifier l'apparence de l'image du visage.

Le schéma synoptique du module d'extraction de signatures est présenté dans la figure 2.1, il prend l'imagette fournie par le module précédent en entrée, y applique un traitement et donne en sortie un vecteur de paramètres, qui est la signature discriminante de la personne.

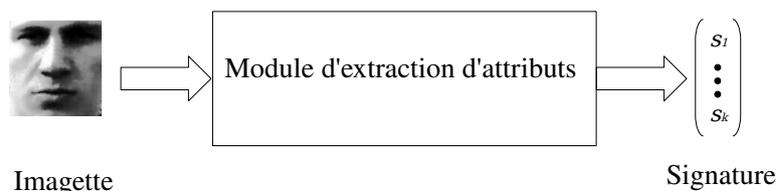


FIGURE 2.1 : Synoptique du module d'extraction de signatures

Dans le premier chapitre, nous avons présenté les principales familles de méthodes

d'extraction de signatures. Nous avons commenté leurs performances, leurs avantages et inconvénients ainsi que leurs conditions optimales d'utilisation. Nous consacrons maintenant ce chapitre à une présentation détaillée des méthodes d'extraction de signatures reposant sur un modèle paramétrique de la fonction image. Cela nous permet de décrire plus précisément le cadre dans lequel s'insère notre travail, dont la présentation fera l'objet du chapitre suivant.

Par définition, la modélisation est une technique de conception et de paramétrage d'une représentation théorique reflétant les caractéristiques et le comportement d'un système. Dans le cas de l'image d'un visage, il ne s'agit pas à proprement parler d'un système, mais d'un ensemble structuré de données. Ainsi, le modèle doit pouvoir représenter fidèlement les caractéristiques les plus discriminantes de cet ensemble de données.

Les paramètres du modèle décrivant une image constituent alors sa signature. Le paramétrage du modèle décrivant une image est réalisé dans une phase d'apprentissage (mode enrôlement). Pendant la phase de reconnaissance et de décision (mode identification), la méthode de classification exploitera les signatures afin de traiter l'image cible. Bien évidemment, les techniques utilisées pour l'extraction de signatures durant la phase d'enrôlement et celles utilisées pour la reconnaissance en-ligne doivent exploiter le même modèle paramétrique.

Nous présentons dans ce chapitre les plus importantes méthodes d'extraction de la signature d'une image par modélisation paramétrique. Dans la première partie, nous décrivons le cadre général permettant d'obtenir une modélisation de la fonction image. Il s'agit ici de mettre en évidence les points communs aux différentes méthodes que nous présenterons par la suite. En effet, calculer un modèle d'une fonction image, ou d'un ensemble de fonctions image, consiste à représenter les données par combinaison d'un ensemble de fonctions élémentaires prédéfinies.

Suivant ce formalisme, nous considérons que les méthodes exploitant une transformation font partie des techniques de modélisation paramétrique. Les principales transformations seront présentées dans la deuxième partie de ce chapitre, à savoir la transformée de Fourier, la transformée en cosinus et la transformée en ondelettes. Nous mettrons l'accent sur cette dernière du fait que les ondelettes servent également de fonctions de référence dans les réseaux d'ondelettes, lesquels sont largement exploités dans la méthode de mo-

délisation que nous proposons par la suite.

La troisième partie du chapitre sera consacrée à l'approximation de fonctions par réseaux de neurones et notamment par réseaux d'ondelettes. Un cas particulier de ce dernier type de réseau sera détaillé, à savoir les réseaux d'ondelettes de Gabor.

En dernière partie, nous présentons une revue bibliographique sur la modélisation par déformation de graphes. Les deux principales méthodes relevant de cette approche seront présentées : l'EGM (Elastic Graph Matching) et l'EBGM (Elastic Bunch Graph Matching).

2.2 Modélisation paramétrique de données

La modélisation paramétrique de données consiste à reconstruire un jeu de données grâce à une combinaison d'un ensemble de fonctions élémentaires dont on peut ajuster les paramètres¹. Les données initiales sont supposées disponibles sous la forme d'un ensemble ordonné de valeurs. On peut alors, sans imposer de condition supplémentaire, supposer que ces données sont les valeurs prises par une fonction $g(\cdot)$ sur un intervalle borné de \mathbb{N} débutant en 0 :

$$\{g(x) \mid x \in \mathbb{N}, 0 \leq x < X\} . \quad (2.1)$$

La modélisation paramétrique des données consiste à rechercher les paramètres, regroupés dans un vecteur colonne \mathbf{p} , d'une fonction $\hat{g}_{\mathbf{p}}(x)$ qui approche au mieux la fonction $g(x)$ sur tout son intervalle de définition. En pratique, dans la plupart des méthodes décrites dans ce chapitre, la fonction $\hat{g}_{\mathbf{p}}(x)$ est une combinaison linéaire de fonctions élémentaires $\psi_{\mathbf{p}_l}(\cdot)$ — ou *atomes* —, chacune dépendant de x et d'un paramètre \mathbf{p}_l :

$$\hat{g}_{\mathbf{p}}(x) = \sum_l \omega_l \psi_{\mathbf{p}_l}(x) , \quad (2.2)$$

les ω_l représentant les poids des atomes dans la combinaison. Le vecteur paramètre global \mathbf{p} regroupe tous les poids ω_l de la combinaison ainsi que tous les paramètres \mathbf{p}_l des atomes.

Le but est de sélectionner correctement l'ensemble des atomes $\psi_{\mathbf{p}_l}(\cdot)$, de telle sorte à pouvoir approcher le mieux possible la fonction $g(\cdot)$. On définit généralement la mesure de

1. Les méthodes de régression, qui peuvent être décrites en utilisant un formalisme similaire, sont un cas particulier de modélisation paramétrique d'un ensemble de données

l'écart² entre la fonction réelle et le modèle par l'erreur quadratique suivante, ou *résidu* :

$$\rho(\mathbf{p}) = \sum_{x=0}^{X-1} (g(x) - \hat{g}_{\mathbf{p}}(x))^2 . \quad (2.3)$$

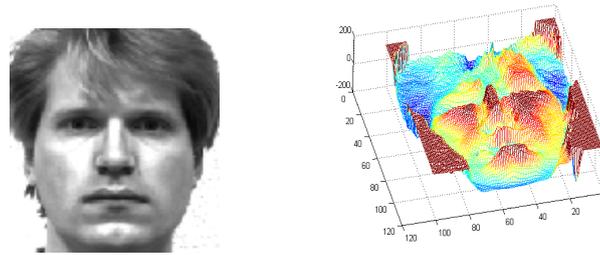
L'ajustement des paramètres du modèle consiste, pour un ensemble fixé d'atomes, à déterminer les poids ω_l et les paramètres \mathbf{p}_l qui permettent de minimiser le résidu $\rho(\mathbf{p})$, voire de l'annuler. Dans certains cas, il suffit de déterminer les poids, du fait que les paramètres \mathbf{p}_l n'apparaissent que de façon implicite dans les expressions algébriques des atomes sélectionnés pour modéliser les données.

Si on revient au problème initial, à savoir l'extraction d'une signature, la modélisation paramétrique ne présente un intérêt que si le vecteur paramètre \mathbf{p} , qui deviendra la signature, possède moins de composantes que le nombre initial de données. Si c'est le cas, la modélisation paramétrique permet de réduire la dimension des données initiales. Afin de satisfaire cette condition, on accepte dans la plupart des cas que le résidu soit non nul, mais simplement le plus petit possible. On peut alors parler d'approximation paramétrique de la fonction, plutôt que de modélisation. Pour simplifier le discours, nous n'utiliserons par la suite que le terme modélisation.

Des formalismes similaires peuvent être utilisés pour décrire la modélisation paramétrique d'ensembles bidimensionnels (ou multidimensionnels) de valeurs, comme les images numériques. Une première approche consiste à balayer dans un ordre prédéfini l'ensemble des valeurs pour se ramener au cas monodimensionnel. En général, le fait de ne plus tenir compte de la topologie initiale rend la modélisation moins précise. Dans le cas des images numériques, on cherche souvent à modéliser directement la fonction image sous son expression bidimensionnelle (cf. figure 2.2).

Les méthodes de modélisation paramétrique peuvent être classées dans deux grandes familles, selon que les paramètres sont déterminés par un calcul direct à partir des valeurs de la fonction $g(\cdot)$ ou par un processus itératif. Les méthodes par transformation sont les techniques les plus utilisées pour le calcul direct de paramètres. Dans cette catégorie, on trouve les transformées de Fourier, en cosinus et en ondelettes, qui sont décrites dans la partie suivante. Quant aux méthodes nécessitant un calcul itératif, nous décrirons la déformation de graphes, en l'occurrence l'EGM (Elastic Graph Matching) et l'EBGM

2. dans la littérature, cet écart est appelé aussi performance ou erreur, qui est la différence entre la fonction initiale et son modèle



(a) Niveaux de gris (b) Graphe de la fonction image

FIGURE 2.2 : Deux formes de représentation d'une image numérique

(Elastic Bunch Graph Matching), les réseaux de neurones et les réseaux d'ondelettes. Il est important de noter que le calcul itératif permet de surmonter le problème posé par la modélisation paramétrique dans le cas où les atomes ne sont pas des fonctions orthogonales.

2.3 Modélisation paramétrique par transformation

Les méthodes par transformation sont des techniques de modélisation visant à exprimer les données sous la forme d'une somme de fonctions paramétriques possédant des propriétés remarquables. La plus connue des transformations est celle de Fourier qui consiste à représenter les données par des sommes de fonctions exponentielles complexes. Les transformées en cosinus et en ondelettes sont également très connues et utilisées dans plusieurs domaines des sciences pour l'ingénieur.

Le principal intérêt d'une transformation réside dans le fait qu'elle est inversible. Cela signifie que la modélisation est parfaite, c'est à dire que le résidu de modélisation est nul. De ce fait, si on connaît uniquement le vecteur paramètre, on peut retrouver les valeurs exactes de toutes les données initiales au moyen d'une transformation inverse.

Quand les atomes utilisés dans la transformation possèdent la propriété d'orthogonalité, on peut déterminer les paramètres de la modélisation (également appelée *transformée*) par un calcul direct. Pour définir la propriété d'orthogonalité, il faut supposer que toutes les fonctions ($g(\cdot)$ et les atomes $\psi_{\mathbf{p}}(\cdot)$) appartiennent à un espace vectoriel sur lequel est défini un produit scalaire $\langle \cdot, \cdot \rangle$. Dans ce cas, deux atomes $\psi_{\mathbf{p}_l}(\cdot)$ et $\psi_{\mathbf{p}_l'}(\cdot)$ sont orthogonaux si leur produit scalaire est nul. Dans toutes les transformations usuelles, les atomes constituent

une base orthonormale de l'espace vectoriel des fonctions, c'est à dire qu'ils vérifient :

$$\langle \psi_{\mathbf{p}_l}(\cdot), \psi_{\mathbf{p}_{l'}}(\cdot) \rangle = \begin{cases} 1 & \text{si } \mathbf{p}_l = \mathbf{p}_{l'} \\ 0 & \text{sinon} \end{cases} . \quad (2.4)$$

Cette propriété permet un calcul direct des poids des atomes dans la modélisation paramétrique, du fait que :

$$\begin{aligned} \langle \psi_{\mathbf{p}_l}(\cdot), g(\cdot) \rangle &= \langle \psi_{\mathbf{p}_l}(\cdot), \sum_j \omega_j \psi_{\mathbf{p}_j}(\cdot) \rangle \\ &= \sum_j \omega_j \langle \psi_{\mathbf{p}_l}(\cdot), \psi_{\mathbf{p}_j}(\cdot) \rangle \\ &= \omega_l , \end{aligned} \quad (2.5)$$

puisque tous les produits scalaires sont nuls sauf $\langle \psi_{\mathbf{p}_l}(\cdot), \psi_{\mathbf{p}_l}(\cdot) \rangle$ qui vaut 1. Le poids correspondant à un atome $\psi_{\mathbf{p}_l}(\cdot)$ est donc calculé par le produit scalaire de cet atome par la fonction qu'on cherche à modéliser.

Dans l'application de reconnaissance de visages, les transformations ont été employées dans différents modules, en l'occurrence pour la préparation des données et/ou l'extraction de signatures. Nous présentons dans les sections suivantes les transformations les plus utilisées et le rôle que certains auteurs leur ont fait jouer pour la reconnaissance de visages.

2.3.1 Transformation de Fourier

La transformation de Fourier a été introduite dans sa version originale dans un mémoire publié par Joseph Fourier en 1807. Ce dernier affirma qu'il est possible, dans certaines conditions, de décomposer une fonction périodique f sous la forme d'une somme infinie de signaux sinusoïdaux. Pendant très longtemps cette transformée est restée un outil purement théorique, les puissances de calcul disponibles n'étant pas suffisantes pour l'implémenter efficacement. La transformée de Fourier s'est réellement développée avec l'introduction de sa version rapide (FFT : Fast Fourier Transform), proposée par Rader et Brenner [RB76]. Le livre de Gasquet et Witomski décrit à la fois les aspects théoriques et pratiques liés à l'analyse de Fourier [GW04].

La fonction $g(\cdot)$ qu'il s'agit de modéliser est échantillonnée, de part sa définition, et elle peut être considérée comme périodique sur \mathbb{Z} puisque son ensemble de définition est un intervalle de largeur finie de \mathbb{N} . C'est donc la version discrète de la transformée

de Fourier qui s'applique dans notre cas. En suivant le formalisme présenté auparavant, La transformée de Fourier discrète 1D est une modélisation paramétrique qui utilise les atomes :

$$\psi_u(x) = \frac{1}{X} \exp\left(\frac{2i\pi u}{X}x\right), \quad (2.6)$$

lesquels dépendent d'un seul paramètre u . Nous rappelons que dans cette équation X désigne le nombre de données à modéliser. Afin de constituer une base orthogonale des fonctions échantillonnées sur X points, il faut disposer de X atomes, qui sont obtenus en sélectionnant également le paramètre u dans l'intervalle $[0 \dots X - 1]$.

Dans ces conditions, la modélisation paramétrique de la fonction $g(x)$ s'exprime sous la forme d'une combinaison linéaire de X termes, le vecteur paramètre \mathbf{p} étant constitué uniquement par l'ensemble des coefficients de pondération $(\omega_0 \omega_1 \dots \omega_{X-1})^T$:

$$\hat{g}_{\mathbf{p}}(x) = \sum_{u=0}^{X-1} \omega_u \psi_u(x), \quad (2.7)$$

qui est égale à $g(x)$ du fait que le résidu est nul. Cette équation correspond en fait à la transformation de Fourier inverse, puisqu'elle permet de calculer les données initiales à partir des paramètres de Fourier. La transformation de Fourier directe, qui détermine les coefficients de pondération ω_u à partir d'un produit scalaire, est donnée par :

$$\omega_u = \sum_{x=0}^{X-1} g(x) \exp\left(-\frac{2i\pi u}{X}x\right), \text{ pour } 0 \leq u < X. \quad (2.8)$$

L'extension bidimensionnelle est immédiate, du fait des propriétés particulières des atomes qui sont des exponentielles complexes. Les atomes de Fourier pour la modélisation paramétrique d'une fonction 2D, définie sur $[0 \dots X - 1] \times [0 \dots Y - 1]$, sont obtenus par le produit des atomes de la modélisation paramétrique 1D :

$$\psi_{u,v}(x, y) = \psi_u(x)\psi_v(y) = \frac{1}{XY} \exp\left(2i\pi\left(\frac{xu}{X} + \frac{yv}{Y}\right)\right), \quad (2.9)$$

lesquels dépendent dans ce cas de deux paramètres u et v . Le choix de cette expression pour les atomes permet de disposer d'une transformation séparable, dans ce sens que la transformation 2D peut être obtenue par application successive de deux séries de transformations 1D.

La transformation de Fourier d'une image numérique $I(x, y)$ consiste à déterminer les coefficients de pondération $\omega_{u,v}$ de ces atomes de Fourier 2D dans une combinaison

linéaire. Les coefficients sont alors calculés grâce à l'expression :

$$\omega_{u,v} = \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} I(x,y) \exp\left(-2i\pi\left(\frac{xu}{X} + \frac{yv}{Y}\right)\right), \text{ pour } 0 \leq u < X \text{ et } 0 \leq v < Y. \quad (2.10)$$

Pour extraire une signature d'une image numérique, il faut réduire la dimension de l'espace de représentation. Du fait que les coefficients de pondération sont des nombres complexes, la modélisation paramétrique d'une fonction par des atomes de Fourier entraîne une augmentation de dimension d'un facteur 2. Pour obtenir en définitive une réduction de dimension, il faut ainsi conserver dans la signature moins d'un coefficient sur deux.

De nombreuses études ont été menées sur les propriétés de la transformée de Fourier d'une image numérique [GW04], indiquant notamment quelles sont les caractéristiques de l'image qui sont codées par des coefficients de Fourier particuliers. Les composantes de fréquences spatiales élevées, modélisées par les atomes pour lesquels les paramètres u et v prennent des valeurs importantes, décrivent les détails et les textures contenus dans l'image. Si ces informations ne sont pas importantes pour une application particulière, il suffit de ne pas tenir compte des atomes correspondant dans la modélisation paramétrique, en conservant par exemple comme vecteur paramètre :

$$\mathbf{p} = (\omega_{0,0} \cdots \omega_{0,Y'-1} \ \omega_{1,0} \cdots \omega_{1,Y'-1} \cdots \omega_{X'-1,Y'-1})^T \quad (2.11)$$

avec $X' < X$ et $Y' < Y$. Dans ce cas, le résidu de modélisation $\rho(\mathbf{p})$ n'est plus nul.

Dans l'application de reconnaissance de visages, la modélisation paramétrique de Fourier été utilisée et comparée avec d'autres techniques fréquentielles dans le cadre du concours de reconnaissance de visages FRGC³ (Face Recognition Grand Challenge) [KSC06]. Elle a également été utilisée dans [DR07] pour la sélection des basses fréquences comme attributs discriminants d'une image faciale, ou pour le filtrage de l'image dans [ZCMC06].

Outre ces diverses variantes d'utilisation pour l'élimination de certaines fréquences spatiales, l'analyse de Fourier a permis d'implanter de nombreuses méthodes hybrides de reconnaissance de visages. On la trouve par exemple associée à l'analyse en composantes principales et aux réseaux de neurones [EAEK00]. Cette combinaison a permis d'atteindre un bon taux de reconnaissance des visages.

3. <http://www.frvt.org/frgc/>

2.3.2 Transformation en cosinus

La transformation en cosinus discrète (DCT) est très similaire à la transformation de Fourier discrète, la différence résidant dans les atomes utilisés pour modéliser les fonctions. En effet, comme son nom l'indique, la première utilise des atomes intégrant une fonction cosinus plutôt qu'une exponentielle complexe. De ce fait, les coefficients de pondération intervenant dans la modélisation paramétrique sont des valeurs réelles plutôt que complexes.

La découverte de la DCT date de 1974 [ANR74]. Cela a constitué une avancée notable dans le domaine du traitement du signal, notamment après que furent publiés les travaux d'implémentation entrepris par Chen, Smith et Fralick en 1977 [RY90]. Cette transformation est à la base de nombreuses techniques et technologies, notamment pour la compression avec perte des signaux temporels (comme l'audio dans le MP3), des images fixes (comme le JPEG) ou les vidéos (différentes normes MPEG).

Les atomes servant de base à la DCT 1D⁴ sont définis par l'expression suivante :

$$\psi_u(x) = \sqrt{\frac{2}{X}} b(u) \cos\left(\frac{\pi u}{2X}(2x+1)\right), \quad (2.12)$$

dans laquelle $b(u)$ vaut 1 sauf pour $b(0)$ qui vaut $2^{-1/2}$. Chaque atome dépend donc du seul paramètre réel u .

Comme pour la transformation de Fourier, l'extension en 2D est immédiate, les atomes de la transformation bidimensionnelle étant simplement obtenus en multipliant deux atomes de la transformation 1D :

$$\psi_{u,v}(x,y) = \psi_u(x)\psi_v(y) = \frac{2}{\sqrt{XY}} b(u)b(v) \cos\left(\frac{\pi u}{2X}(2x+1)\right) \cos\left(\frac{\pi v}{2Y}(2y+1)\right), \quad (2.13)$$

u et v prenant respectivement leurs valeurs dans les intervalles $[0 \dots X-1]$ et $[0 \dots Y-1]$. Cette expression des atomes confère également la propriété de séparabilité à la transformation 2D.

Avec la DCT 2D, les coefficients de pondération des atomes dans la modélisation paramétrique d'une fonction image $I(x,y)$ sont obtenus par le calcul direct :

$$\omega_{u,v} = \frac{2}{\sqrt{XY}} b(u)b(v) \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} I(x,y) \cos\left(\frac{\pi u}{2X}(2x+1)\right) \cos\left(\frac{\pi v}{2Y}(2y+1)\right). \quad (2.14)$$

4. Différentes expressions ont été proposées dans la littérature pour définir les atomes. Celle que nous avons retenue, qui semble la plus utilisée, est référencée sous l'appellation transformation de type II.

La DCT est souvent utilisée pour déterminer une modélisation paramétrique de petits blocs carrés (8×8 ou 16×16) extraits de l'image. Dans ce cas, les coefficients de la modélisation paramétrique sont également représentés par un bloc carré, dans lequel ceux qui correspondent aux fréquences spatiales basses sont localisés dans le coin supérieur gauche.

Afin de diminuer la dimension de l'espace de représentation, les coefficients correspondant aux fréquences spatiales élevées sont soit tout simplement annulés, soit codés avec une précision plus limitée, de telle sorte à diminuer l'espace mémoire dans lequel ils sont stockés. Cette dernière approche est celle qui est utilisée dans la plupart des standards de compression de données avec perte (JPEG, MPEG).

En dépit des efforts employés pour développer des algorithmes rapides de calcul de la DCT, il a fallu attendre jusqu'à 2001 pour la voir appliquée à la problématique de la reconnaissance de visages. Sa première utilisation a été relatée dans [HL01], pour comparer ses performances à celles obtenues avec la transformation KLT (Karhunen-Loeve Transform).

Après ce premier succès, la DCT a attiré plus directement l'intérêt des scientifiques, d'autres études et comparaisons ont suivi pour analyser plus précisément ses performances vis-à-vis des variations géométriques et/ou d'illumination des visages durant l'acquisition des images. Elle a été associée à d'autres types de transformations, comme dans [SAI05] où les auteurs ont combiné les transformations en ondelettes et en cosinus pour autoriser le traitement d'images de visages acquises avec une rotation pouvant atteindre 20 degrés.

La transformée en cosinus a également été utilisée pour réduire les dimensions de l'image originale avant l'apprentissage [ECW05]. Cette étape permet de supprimer les zones de grandes variations de luminosité en annulant les coefficients des basses fréquences spatiales. L'extraction de signatures est ensuite opérée par une ADL et la classification assurée par un réseau de neurones de type RBF (Radial Basis Function). Dans le même contexte, Yin *et coll.* ont choisi de réduire les dimensions de l'image par une DCT suivie d'une compression des paramètres de la modélisation [YFQ07]. Pour ce faire, seules les basses fréquences spatiales sont sélectionnées et analysées par ADF.

Une autre utilisation a été proposée par Abouzar *et coll.* [AYS07], laquelle consiste à utiliser deux méthodes de transformation comme pré-traitements avant la phase d'ex-

traction de signatures. Il s'agit de la transformée en ondelettes (Daubechies 4) et de la DCT. La première sert à la réduction de dimensions et la deuxième à supprimer les fréquences élevées dans l'image de taille réduite. L'ACP est ensuite appliquée pour extraire la signature et les SVM pour la classification.

Les différentes comparaisons qui ont opposé l'analyse par DCT à d'autres techniques de reconnaissance de visages ont prouvé sa grande performance, et ce quel que soit son rôle dans le processus de reconnaissance. Lu *et coll.* résumant bien ce constat [LLA06], les auteurs ayant relevé les différentes utilisations de la DCT dans les trois modules, afin d'assurer un traitement :

- sans association à une autre méthode ;
- associée à la SVD (Singular Value Decomposition) ;
- associée à l'ACP 2D.

Après avoir effectué une série de tests sur deux bases de données (AMP et Yale), les auteurs ont conclu que la combinaison la plus stable est celle du duo DCT+ACP du fait que ces deux analyses combinent parfaitement les représentations dans les domaines fréquentiel et spatial.

2.3.3 Transformation en ondelettes

L'analyse par ondelettes a été introduite par Morlet pour améliorer la solution de fenêtrage que Gabor avait proposée pour remédier aux limitations de la transformation de Fourier. En effet, la transformation de Fourier présente quelques inconvénients qui peuvent être résumés en trois points :

- elle ne permet pas de caractériser avec précision un signal simultanément dans le domaine fréquentiel et le domaine spatial ;
- le résultat obtenu par transformation fait disparaître toute information concernant le domaine spatial : le début ou la fin d'une composante du signal ne sont plus localisables dans le modèle paramétrique ;
- la fréquence spatiale d'un signal étant inversement proportionnelle à sa période, si on veut obtenir des informations sur un signal basse fréquence, l'intervalle d'observation doit être grand (et inversement).

La solution de fenêtrage consiste à utiliser des fonctions oscillantes dont l'étendue spatiale est limitée par l'intermédiaire d'une fenêtre. Cette fenêtre, qualifiée de *glissante* est déplacée afin d'analyser indépendamment les différentes portions de la série initiale de données. Le cas le plus simple de fenêtrage correspond à la transformation de Fourier à court-terme (Short-Term Fourier Transform : STFT) dans lequel les atomes sont simplement des exponentielles complexes dont le support est d'étendue limitée.

La modélisation paramétrique par les ondelettes introduit la notion d'échelle. Il s'agit d'utiliser comme atomes des fonctions toutes similaires, qui sont des copies les unes des autres mais à des facteurs d'échelle variés. En 1D, les atomes sont une famille de fonctions $\psi_{d,t}$ construite à partir d'une fonction ψ de l'espace de Hilbert $L^2(\mathbb{R})$, à valeurs réelles ou complexes, laquelle est appelée *ondelette mère*, ou *ondelette analysante* :

$$\psi_{d,t}(x) = \sqrt{d}\psi(d(x-t)), \quad (2.15)$$

dans laquelle les paramètres d et t correspondent respectivement au paramètre de *dilatation*⁵ et au paramètre de *translation*. Le coefficient de normalisation (\sqrt{d}) n'est utile que lorsque l'ondelette est utilisée dans une transformation, afin d'obtenir une base orthonormale.

L'ondelette mère, fonction qui engendre les autres ondelettes de la famille, doit respecter certaines propriétés analytiques [Chu92, Dau92, Mal98]. Nous présentons dans ce paragraphe les plus importantes propriétés afin de souligner les particularités des ondelettes par rapport des autres familles de fonctions mathématiques [Ous98, Zai08] :

- la localisation : l'ondelette mère doit présenter une décroissance rapide aux limites de son domaine de définition, c'est-à-dire que son énergie doit être contenue dans un intervalle fini. Cette énergie doit elle-même être finie (fonction de carré sommable). On l'appelle également fonction à *support compact*.
- l'oscillation : pour qu'elle puisse estimer les fréquences spatiales, l'ondelette mère doit être suffisamment oscillante, comme les atomes des transformations de Fourier et de la transformation en cosinus. D'autre part, l'ondelette mère doit avoir une

5. Dans ce manuscrit, nous avons choisi de conserver les notations des auteurs auxquels se réfère notre travail, bien que leur choix soit ambigu. En effet, le coefficient d (dilatation en anglais) intervenant au numérateur, les ondelettes obtenues en augmentant d sont moins étendues spatialement, et donc compressées au lieu d'être dilatées.

intégrale nulle dans le cas continu, ou la somme de ses valeurs sur l'intervalle de définition doit être nulle dans le cas discret.

Ces propriétés sont illustrées sur les ondelettes mère représentées sur la figure 2.3, qui montrent clairement l'aspect oscillant et la compacité du support.

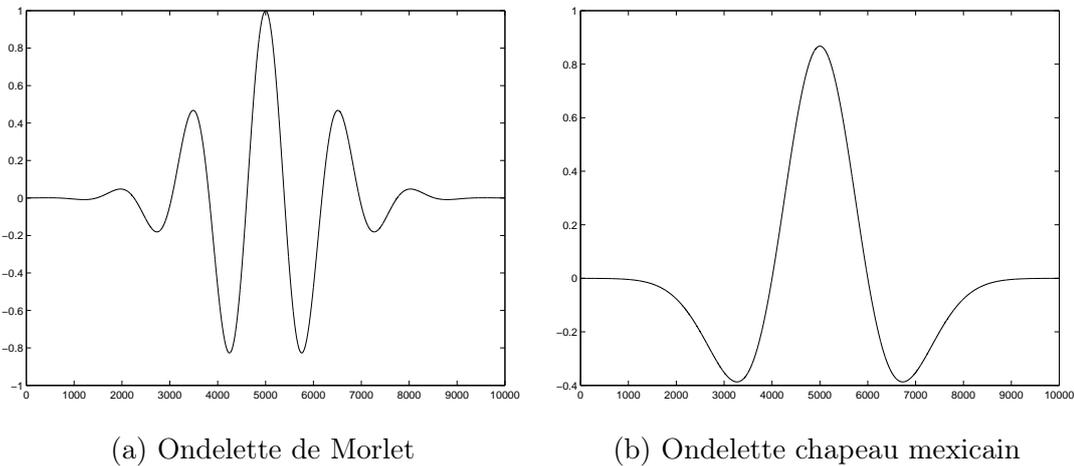


FIGURE 2.3 : Ondelettes mère oscillantes et à support compact

Outre ces propriétés indispensables, d'autres sont souhaitables pour garantir des propriétés particulières des atomes utilisés pour la modélisation paramétrique. L'intérêt de ces propriétés varie beaucoup selon l'application. Nous présentons donc une liste non exhaustive de propriétés complémentaires d'une ondelette mère, étant entendu qu'elle peut difficilement posséder simultanément toutes ces propriétés :

- la régularité, qui assure une meilleure localisation fréquentielle. Cette propriété est surtout essentielle dans les applications de compression d'images. Quand certains coefficients sont annulés dans la représentation paramétrique pour compresser les données, si l'ondelette n'est pas régulière des artéfacts très visibles apparaissent sur les images reconstruites.
- l'orthogonalité. Comme nous l'avons indiqué précédemment, la propriété d'orthogonalité permet d'obtenir un calcul direct des coefficients de la modélisation paramétrique. Cette propriété permet également d'assurer l'existence d'une transformée inverse et d'une décomposition non redondante.
- la symétrie. Cette propriété, importante principalement dans le domaine du traitement des signaux numériques, assure que ces derniers ne seront pas distordus après un filtrage suivi d'une reconstruction.

- expression analytique : cette propriété est notamment indispensable pour construire un réseau d'ondelettes. En effet, un réseau d'ondelettes possède la même architecture que celle d'un réseau de neurones mais intègre une ondelette en tant que fonction d'activation. Si cette ondelette n'a pas d'expression analytique, le calcul de la sortie d'un neurone est plus complexe.

Les familles d'ondelettes les plus utilisées sont les suivantes :

- ondelettes brutes : symétriques et non orthogonales. Exemples : ondelette gaussienne, ondelette de Morlet, ondelette chapeau mexicain.
- ondelettes régulières infinies : infiniment dérivables, orthogonales, mais non nulles en dehors de l'intervalle de définition (support non compact). Elles peuvent être utilisées à la fois dans l'analyse continue ou discrète. Exemple : ondelette de Meyer.
- ondelettes orthogonales à support compact : ces ondelettes ne sont pas régulières, mais sont orthogonales et nulles en dehors de leur support de définition.
- ondelettes bi-orthogonales à support compact. La bi-orthogonalité est une condition moins stricte que l'orthogonalité, qui permet simplement de garantir que la transformation inverse existe. Cette transformation inverse est obtenue à partir d'atomes dérivés d'une autre ondelette mère, appelée *ondelette mère duale*. Les ondelettes de ce type sont en général symétriques et à support compact. Exemple : ondelettes B-splines.
- ondelettes complexes : à valeurs complexes, non orthogonales mais symétriques. Exemples : ondelette gaussienne complexe, ondelette de Morlet complexe, ondelette de Shannon complexe.

Pour aboutir à une transformation en ondelettes discrète, donc à une modélisation paramétrique inversible, il faut encore définir l'ensemble dans lequel les paramètres d et t prennent leurs valeurs. De nombreuses propositions ont été réalisées à ce sujet, notamment par Daubechies [Dau90, Dau92], et leur description détaillée n'aurait pas d'intérêt dans ce manuscrit. Citons simplement l'approche qui est la plus utilisée dans le cas de la transformation en ondelettes discrètes, qui consiste à choisir des coefficients de dilatation d en progression géométrique de raison $1/2$ et des coefficients de translation t à valeurs entières. Dans ce cas, l'expression générale des atomes est la suivante :

$$\psi_{a,b}(x) = 2^{-(a/2)}\psi(2^{-a}(x - b)) , \quad (2.16)$$

avec a et b entiers relatifs. Avec certaines ondelettes mère, les coefficients de cette transformation discrète peuvent être calculés de façon efficace par dichotomie [Dau92].

La transformation en ondelettes discrète, étendue en 2D pour traiter des images, a été utilisée dans la phase d'extraction des signatures [MSP05, XSK05]. Cependant, le faible nombre de références bibliographiques indique que son utilisation dans le domaine de la reconnaissance de visages est beaucoup moins courante que celle des autres transformations usuelles.

Dans le cas général, contrairement aux transformations de Fourier et en cosinus, l'extension bidimensionnelle ou multidimensionnelle de l'analyse par ondelettes n'est pas triviale. La transformation n'est pas forcément séparable et les atomes ne sont pas forcément le produit de plusieurs atomes 1D. C'est cependant cette solution qui a été retenue par Zhang et Benveniste [ZB92], qui proposent de déterminer l'ondelette mère $\psi^K(\cdot)$ dans un espace de dimension K par un produit d'ondelettes mère 1D :

$$\psi^K(\mathbf{x}) = \prod_{k=1}^K \psi(x_k) , \quad (2.17)$$

dans laquelle \mathbf{x} est le vecteur colonne $(x_1 \ x_2 \ \cdots \ x_K)^T$. Les auteurs ne précisent pas s'il est possible de déduire des propriétés particulières de l'ondelette mère multidimensionnelle à partir de celles de l'ondelette mère 1D. Pour simplifier les notations quand il n'y a pas d'ambiguïté, nous omettrons par la suite l'exposant K dans la notation d'une ondelette mère multidimensionnelle.

Pour définir les atomes, donc une paramétrisation d'une ondelette dans un espace de dimension K à partir d'une ondelette mère multidimensionnelle, considérons la série de transformations affines suivantes :

- une homothétie définie par la matrice de transformation $D = \text{diag}(d_1, d_2, \dots, d_K)$;
- une rotation, de matrice R , constituée par la combinaison de $K(K-1)/2$ rotations élémentaires définies par des matrices de rotation de Givens⁶ ;
- une translation de vecteur $T = (t_1 \ t_2 \ \cdots \ t_K)^T$.

Tous les paramètres de ces transformations géométriques peuvent être regroupés dans un vecteur \mathbf{p} unique, qui comporte donc $K(K+3)/2$ composantes (K pour l'homothé-

6. Les matrices de rotation de Givens permettent de généraliser les angles d'Euler à un espace à K dimensions. Pour définir une rotation quelconque dans un espace à K dimensions il faut $K(K-1)/2$ paramètres angulaires.

tie, $K(K - 1)/2$ pour la rotation et K pour la translation) [ZB92]. En appliquant une combinaison des transformations affines paramétriques définies précédemment, on obtient l'expression d'un atome de paramètre \mathbf{p} construit sur l'ondelette mère $\psi(\cdot)$ grâce à l'expression suivante :

$$\psi_{\mathbf{p}}(\mathbf{x}) = \psi(D \cdot R(\mathbf{x} - T)) . \quad (2.18)$$

Dans le cas bidimensionnel, le vecteur paramètre \mathbf{p} comporte cinq composantes : deux pour les facteurs de dilatation (d_x et d_y), un pour la rotation (angle θ) et deux pour les translations (t_x et t_y). Les atomes $\psi_{\mathbf{p}}(\cdot)$ déduits d'une ondelette mère $\psi(\cdot)$ sont ainsi donnés par l'expression suivante :

$$\psi_{\mathbf{p}}(x, y) = \psi \left(\begin{pmatrix} d_x & 0 \\ 0 & d_y \end{pmatrix} \cdot \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \cdot (x - t_x \quad y - t_y)^T \right) , \quad (2.19)$$

\mathbf{p} désignant le vecteur colonne $(d_x \ d_y \ \theta \ t_x \ t_y)^T$.

2.4 Modélisation itérative de la fonction image

Dans la section précédente, nous avons décrit les principales méthodes de modélisation paramétrique d'une fonction par transformation, lesquelles permettent d'obtenir par un calcul direct le vecteur signature caractérisant une image, notamment d'un visage. Les données image étant considérées comme les valeurs d'une fonction définie sur un support compact et borné de \mathbb{Z}^2 , leur signature est constituée par les paramètres de cette modélisation par transformation. La définition des atomes, qui constituent une base des fonctions sur le support considéré, est le fondement de cette approche.

Cependant, quand les atomes considérés ne constituent pas une base orthornomale des fonctions, la transformation directe des données n'est plus possible et il faut revenir au problème initial de modélisation paramétrique exposé dans le paragraphe 2.2. La modélisation optimale est alors obtenue, connaissant l'ensemble des atomes, par recherche itérative des paramètres de la combinaison linéaire (2.2) qui minimise le résidu défini par l'équation (2.3).

Dans cette section, nous présentons en premier lieu un formalisme neuronal qui a été retenu par de nombreux auteurs afin de décrire ce mode de paramétrisation d'une combinaison linéaire d'atomes. La structure générale des RNA pour l'approximation est

présentée, puis le cas particulier des réseaux d'ondelettes est décrite en détail. Parmi les ondelettes dont les propriétés sont particulièrement intéressantes pour la modélisation paramétrique, nous décrivons l'ondelette de Gabor et le réseau d'ondelettes de Gabor (ROG) qui a servi de base à nos travaux.

2.4.1 RNA pour la modélisation d'une image

Comme nous l'avons montré dans le chapitre 1, les RNA peuvent jouer plusieurs rôles dans un processus de reconnaissance de visages. De façon générale, il a été montré qu'un RNA multicouches (MLP) dont les fonctions d'activation sont correctement choisies est un approximateur universel, c'est à dire une structure qui permet de résoudre le problème de la modélisation paramétrique de n'importe quelle fonction. Nous décrivons ici un cas particulier de RNA dédié à la modélisation d'une fonction image.

L'approche consiste à reformuler, avec un formalisme neuronal, la question de la modélisation paramétrique d'une image $I(x, y)$. Pour ce faire, on procède de la façon suivante. Considérons un ensemble de neurones dont les entrées sont les coordonnées entières (x, y) d'un point de la grille d'échantillonnage et la sortie un unique scalaire. La couche d'entrée est constituée de L neurones et la couche de sortie d'un seul neurone sommateur, comme représenté sur la figure 2.4.

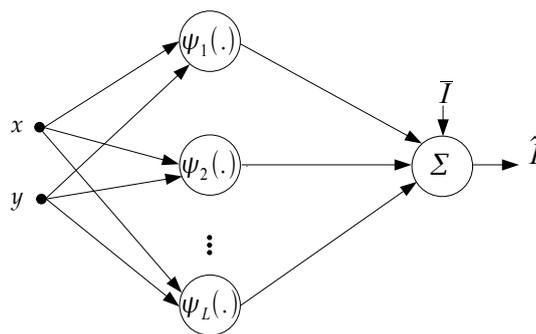


FIGURE 2.4 : Architecture du RNA pour l'approximation d'une image

En sélectionnant correctement les pondérations des connexions reliant les entrées à la première couche de neurones et les fonctions d'activation de ces derniers, la fonction de transfert du réseau peut correspondre exactement à une approximation paramétrique de

la forme :

$$\hat{I}_{\mathbf{p}}(x, y) = \sum_{l=1}^L \omega_l \psi_{\mathbf{p}_l}(x, y), \quad (2.20)$$

dans laquelle uniquement L atomes sont utilisés, les vecteurs \mathbf{p}_l codant les paramètres du réseau de connexions alimentant le neurone d'indice l . Cette expression est un cas particulier de celle décrite par l'équation (2.2) intégrant un nombre fini d'atomes.

L'erreur d'approximation est quantifiée par un résidu, similaire à celui de l'équation (2.3) :

$$\rho(\mathbf{p}) = \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} \left(I(x, y) - \hat{I}_{\mathbf{p}}(x, y) \right)^2, \quad (2.21)$$

qui correspond à la différence quadratique entre les données images $I(x, y)$ et leur approximation $\hat{I}(x, y)$ calculée par le réseau.

L'ajustement des paramètres \mathbf{p} est opéré en minimisant $\rho(\mathbf{p})$ de façon itérative. De part le formalisme neuronal, cette étape d'ajustement des paramètres est implantée via un algorithme apprentissage, dont il existe de nombreuses variantes. Pour plus de détails concernant les RNA pour l'approximation de fonctions et les algorithmes d'apprentissage associés, le lecteur peut se référer au deuxième chapitre de la thèse de Sylvain Tertois [Ter03] ou à différents travaux décrivant les propriétés de ces réseaux [DH96, WLLW97, LLO97].

Dans le domaine de la reconnaissance de visages, ces RNA ont été largement utilisés pour l'extraction des signatures permettant une bonne performance [Row99]. [LGTB97] présente une étude complète des différentes utilisations intégrant une comparaison des performances de plusieurs RNA. Parmi les fonctions d'activation les plus utilisées pour la modélisation de visages, on trouve la fonction gaussienne (réseaux RBF : Radial Basis Function). Les réseaux RBF ont été utilisés directement [EWLT02, Wan08] ou en association avec d'autres techniques [DWZ06, CrH08].

La performance de l'architecture RBF a motivé les scientifiques à développer d'autres types de RNA pour la modélisation paramétrique, notamment le réseau d'ondelettes. Ce dernier possède la même structure que le RNA présenté auparavant, mais les fonctions d'activation des neurones de la couche d'entrée sont des ondelettes. Le réseau d'ondelettes allie la performance des RNA pour la modélisation paramétrique au pouvoir discriminant des ondelettes [PSKK05].

2.4.2 Réseau d'ondelettes

Le réseau d'ondelettes (RO) a été introduit par Zhang et Benveniste en 1992 [ZB92] comme une alternative aux réseaux de neurones feed-forward pour l'approximation arbitraire des fonctions non linéaires. Depuis, il a été utilisé dans plusieurs applications et dans différents domaines, par exemple : le contrôle des systèmes [Lek97], l'estimation non paramétrique [Zha97], la modélisation statique et dynamique de processus [Ous98], etc. Le traitement des signaux et des images reste malgré tout le domaine phare d'utilisation des RO. Pour une étude complète des propriétés des RO, le lecteur peut se référer à la thèse de Richard Barron dans laquelle il a présenté une contribution très complète à l'étude des RO [Bar97].

Dans le domaine du traitement de signaux, plusieurs travaux ont exploité les RO, par exemple pour la détection des vibrations anormales dans un système mécanique, l'annulation d'écho dans un signal audio, la segmentation de la parole et la reconnaissance vocale [Thu00]. Quant au traitement d'images, ils ont été utilisés dans deux applications principales : la compression d'images [SAA03, JA05, AJ05] et la reconnaissance de visages [ZAA05, ZAA03, Kru00, KS02a, FCK01].

Zhang et Benveniste ont décrit une structure de RO permettant la modélisation paramétrique de fonctions multidimensionnelles $g(\mathbf{x})$, selon l'expression :

$$\hat{g}_{\mathbf{p}}(\mathbf{x}) = \sum_{l=1}^L \omega_l \psi(D_l \cdot R_l(\mathbf{x} - T_l)) + \bar{g} \quad (2.22)$$

dans laquelle L est le nombre de neurones sur la couche d'entrée, D_l , R_l et T_l désignent les transformations affines appliquées au vecteur \mathbf{x} avant traitement par l'ondelette mère dans le neurone d'indice l , et \bar{g} est la composante continue (valeur moyenne) de l'approximation. Le vecteur \mathbf{p} regroupe tous les paramètres du RO, à savoir les L poids ω_l , les $L \times K(K+3)/2$ coefficients des transformations affines et \bar{g} .

Le paramètre \mathbf{p} est ajusté par un algorithme itératif d'optimisation, qui vise à minimiser le résidu :

$$\rho(\mathbf{p}) = \sum_{\mathbf{x}} (g(\mathbf{x}) - \hat{g}_{\mathbf{p}}(\mathbf{x}))^2, \quad (2.23)$$

pour toutes les valeurs disponibles de l'entrée \mathbf{x} .

Comme indiqué dans la section 2.3.3, le choix de l'ondelette mère $\psi(\cdot)$ dépend principalement de l'application. Les ondelettes orthogonales ont été employées pour l'identi-

cation et surveillance de systèmes dynamiques [Zha99], pour la définition de modèles de régression [Zha93] et pour la classification des signaux [PKG98]. Pour leur part, Krüger et Feris ont employé des ondelettes non orthogonales pour la représentation d'images de visages [Kru00, KS02a, FCK01]. Dans un autre domaine, [Ous98][ORPD98] ont employé des ondelettes à structure oblique pour la modélisation de processus. L'étude de Marc Thuillard présente un aperçu général des différentes utilisations des RO [Thu00]. Une comparaison des performances du RO et du RBF pour la modélisation de fonctions a été présentée dans [BAA06].

L'ondelette bêta et l'ondelette de Gabor restent cependant les ondelettes les plus utilisées comme fonction d'activation dans la couche d'entrée d'un RO. La première a été utilisée pour modéliser des fonctions 1D et 2D [BAA05], pour la reconnaissance de visages [ZAA05, ZAA03] et pour la compression d'images [JA05, AJ05]. Les RO exploitant l'ondelette de Gabor, utilisés pour la modélisation paramétrique de l'image d'un visage exclusivement par Krüger d'après notre recherche bibliographique [KS02a], sont détaillés dans la section suivante. Contrairement aux autres techniques d'extraction de signatures, le RO a rarement été associé à d'autres techniques dans les approches relatées dans la littérature. Cela dit, il sert de base à quelques travaux récents, notamment visant à la modélisation 3D des visages [SAZ⁺08].

2.4.3 Ondelettes de Gabor et réseau d'ondelettes de Gabor

En 1946, Dennis Gabor a introduit la fonction qui porte maintenant son nom dans le but de pallier l'un des défauts des atomes de Fourier, à savoir l'impossibilité de conserver l'information spatiale lors de la transformation d'un signal. John G. Daugman a proposé une extension bidimensionnelle de la fonction de Gabor en 1985 [Dau85]. Dans les deux cas, la fonction de Gabor étant utilisée pour définir des atomes permettant la décomposition paramétrique d'une fonction, il y est souvent fait référence en tant qu'ondelette [FGTK01]. Pour plus de détails concernant les propriétés de cette ondelette, se référer aux publications de Daubechies [Dau90, Dau92].

2.4.3.1 Ondelettes de Gabor

L'ondelette de Gabor est une fonction à moyenne nulle, oscillante et définie sur un support compact. Dans le cas 1D, elle est définie par les expressions :

$$\psi^g(x) = \exp\left(\frac{-x^2}{2}\right) \exp(j\omega_0 x) \quad \text{ou} \quad \psi^g(x) = \exp\left(\frac{-x^2}{2}\right) \sin(\omega_0 x), \quad (2.24)$$

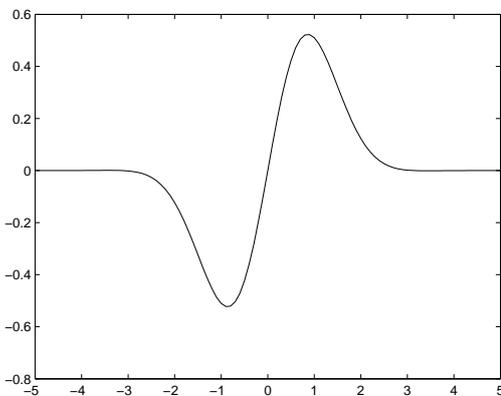
sous ses formes respectivement complexe et réelle. L'ondelette de Gabor ne peut pas être utilisée simplement pour établir une transformation, du fait qu'elle ne permet pas de constituer des atomes orthogonaux et que les ondelettes duales ne disposent pas d'une expression analytique. De ce fait, elle n'a été utilisée que pour la modélisation paramétrique avec une recherche itérative de la solution. En revanche, elle réalise un excellent compromis entre les descriptions spatiale et fréquentielle de la fonction modélisée.

L'expression de l'ondelette de Gabor 1D peut être généralisée à un espace de dimension K . Dans ce cas, on considère comme l'a fait Daugman que la partie oscillante est une fonction d'une seule composante du vecteur \mathbf{x} alors que la partie gaussienne est calculée en fonction de son module :

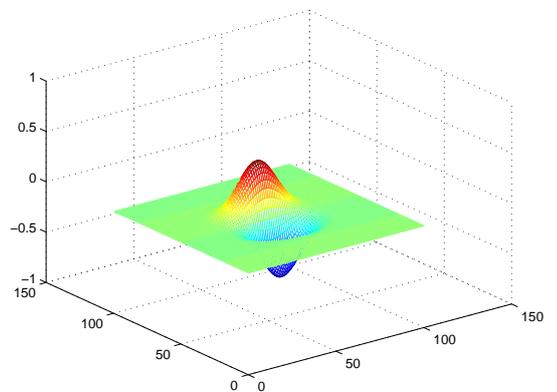
$$\psi^g(x) = \exp\left(\frac{-\mathbf{x}^T \mathbf{x}}{2}\right) \sin(\omega_0 x_1). \quad (2.25)$$

dans laquelle x_1 est la première coordonnée du vecteur \mathbf{x} .

La figure 2.5 présente les graphes d'ondelettes mère de Gabor en une et deux dimensions, pour un paramètre ω_0 égal à 1.



(a) ondelette de Gabor 1D



(b) ondelette de Gabor 2D

FIGURE 2.5 : Graphes de l'ondelette de Gabor

L'ondelette de Gabor possède des propriétés remarquables qui justifient sa popularité dans le domaine de la modélisation paramétrique des signaux et des images. L'une d'elles est le fait qu'elle reproduit parfaitement la réponse de certains systèmes biologiques, notamment des neurones constituant les premières couches du système visuel animal. Par exemple, Daugman a comparé les graphes de certaines fonctions de Gabor avec les champs réceptifs de neurones du cortex visuel du chat (cf. figure 2.6) et a montré qu'ils étaient très similaires [Dau85, JP87].

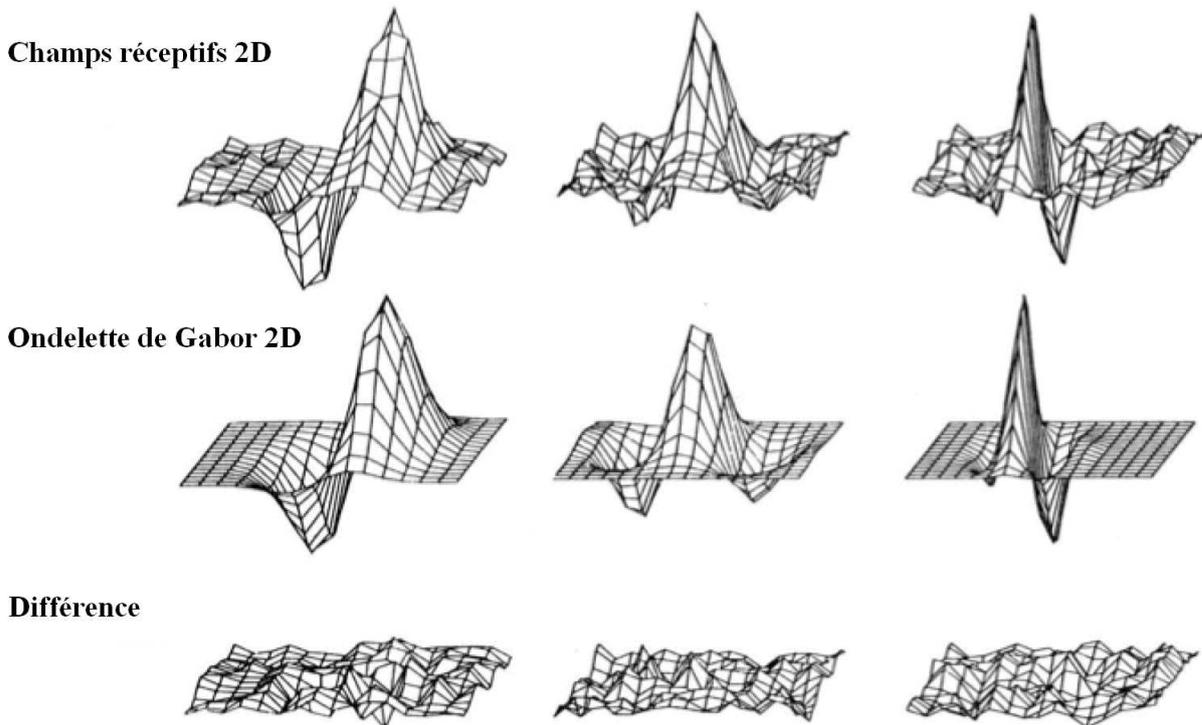


FIGURE 2.6 : Allure des champs réceptifs 2D expérimentaux et modélisés (extraite de [Dau85])

Les fonctions de la figure 2.6 sont obtenues par paramétrisation de l'ondelette de Gabor 2D selon l'expression suivante :

$$\begin{aligned} \psi_{\mathbf{p}}^g(x, y) = & \exp \left[-\frac{\omega_0}{2} \left([(d_x \cos \theta(x - t_x) - d_x \sin \theta(y - t_y)]^2 \right. \right. \\ & \left. \left. + [d_y \sin \theta(x - t_x) + d_y \cos \theta(y - t_y)]^2 \right) \right] \\ & \times \sin \left(d_x \cos \theta(x - t_x) - d_x \sin \theta(y - t_y) \right), \end{aligned} \quad (2.26)$$

qui correspond à appliquer une translation de paramètres t_x et t_y , puis une rotation d'angle

θ et enfin une dilatation de facteurs d_x et d_y au vecteur \mathbf{x} avant d'utiliser l'ondelette mère de l'équation (2.25) dans le cas 2D.

Dans la modélisation paramétrique d'une image numérique, les atomes de Gabor permettent de coder très efficacement les transitions du niveau de gris au niveau des contours. Dans ce cas, l'ondelette doit avoir une partie oscillante limitée à une seule période (comme c'est le cas sur les ondelettes de la figure 2.6). Cela correspond à fixer le paramètre ω_0 à la valeur 1. Daugman, puis Krüger, ont justifié ce choix par des considérations sur la bande passante du filtre correspondant à la convolution du signal ou de l'image avec les atomes de Gabor [Kru00, Dau85].

La première utilisation des ondelettes de Gabor conjointement à un RNA a été décrite dans [Dau88]. Il s'agissait d'associer une modélisation paramétrique par des ondelettes de Gabor 2D à un RNA assurant la compression de la signature. Dans le même contexte, [Lee96] a décrit une technique de représentation d'une image numérique par un ensemble d'ondelettes de Gabor 2D. Sur la base de ces premiers travaux prometteurs, notamment ceux de Quighua Zhang [ZB92], le réseau d'ondelettes de Gabor a été introduit en 2000 par Volker Krüger.

2.4.3.2 Réseau d'ondelettes de Gabor 2D

Afin d'obtenir une modélisation paramétrique d'une image numérique, des ondelettes de Gabor 2D peuvent être utilisées comme fonctions d'activation dans un RO. On obtient ainsi un réseau d'ondelettes de Gabor (ROG) dont la structure est présentée sur la figure 2.7. Les atomes de Gabor constituant les fonctions d'activation des neurones de la couche d'entrée sont paramétrés par les translations T_l , rotations R_l et dilatations D_l . Les sorties de ces neurones sont pondérées par les coefficients ω_l et un biais \bar{I} permet de tenir compte de la composante continue de l'image.

La fonction de transfert de ce réseau, c'est à dire l'expression de la sortie $\hat{I}_{\mathbf{p}}(x, y)$ en fonction de l'entrée (x, y) , est donnée par :

$$\hat{I}_{\mathbf{p}}(x, y) = \sum_{l=1}^L w_l \psi_{\mathbf{p}_l}^g(x, y) + \bar{I}, \quad (2.27)$$

dans laquelle \mathbf{p}_l désigne le vecteur paramètre $(d_x \ d_y \ \theta \ t_x \ t_y)^T$ associé à chaque neurone et \mathbf{p} regroupe à la fois les \mathbf{p}_l et les poids ω_l . La composante continue \bar{I} de l'image, qui peut

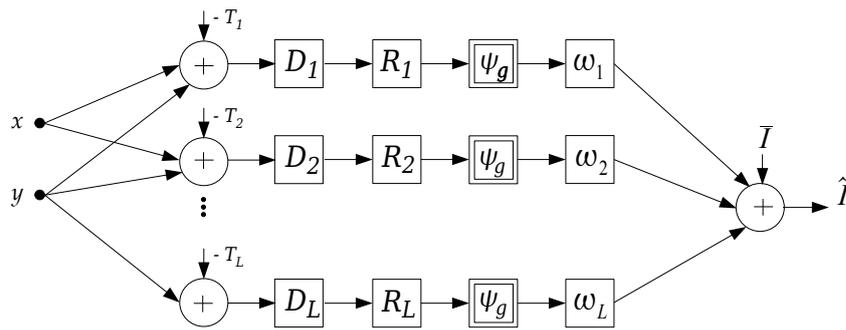


FIGURE 2.7 : Architecture du réseau d'ondelettes de Gabor 2D

être annulée sans modifier le contenu informationnel de cette dernière, sera omise par la suite pour simplifier les notations.

L'image $I(x, y)$ étant fixée, l'ajustement des paramètres du ROG est assuré par une technique itérative d'optimisation. Pour ce faire, on calcule le gradient de l'erreur quadratique $(I(x, y) - \hat{I}_{\mathbf{p}}(x, y))^2$ pour l'ensemble des entrées (x, y) et on ajuste les paramètres du réseau par rétro-propagation de ce gradient. Cette procédure élémentaire est répétée pendant de nombreuses itérations jusqu'à convergence.

Pour améliorer la convergence vers un ensemble de paramètres pour lesquels le résidu est faible, Krüger a proposé d'utiliser l'algorithme de Levenberg-Marquardt comme technique d'optimisation [HM94]. D'autres auteurs, pour apprendre les paramètres de réseaux de structure similaire, ont proposé d'utiliser des techniques stochastiques, notamment les algorithmes génétiques [CL07].

Depuis son apparition en 2000, le réseau d'ondelettes de Gabor n'a cessé d'être utilisé et adapté à diverses applications. On le trouve utilisé pour la localisation des attributs discriminants de visage [FGTK01], le suivi du visage dans une séquence vidéo [FCK01, PL08, FJ00], l'estimation de pose [KS02b], l'alignement d'images de visage [FCK01] et la reconnaissance de visages [ZM05b, KS02a, ZM05a].

Pour l'identification de visages, Krüger utilise les paramètres des ondelettes de Gabor et les poids du réseau comme signature associée à chaque image [KS02a]. Nous reviendrons en détail sur la méthode de Krüger dans la suite de ce manuscrit.

2.4.4 Paramétrisation par déformation de graphe

Les techniques de modélisation paramétrique décrites auparavant utilisent des itérations visant à minimiser le résidu. Au cours des itérations, les paramètres du RNA évoluent, la convergence se produisant vers un minimum local. Aucune méthode d'optimisation ne peut garantir la convergence vers un minimum global, si tant est qu'il existe pour ce problème particulier. En pratique, aucune contrainte n'étant imposée sur les valeurs des paramètres des atomes intervenant dans la modélisation, le résultat obtenu après convergence n'est pas toujours satisfaisant.

Pour pallier ce problème, Krüger a proposé une méthode de sélection des paramètres initiaux du ROG permettant d'aboutir presque certainement à une solution acceptable. Nous reviendrons sur cette méthode dans le chapitre suivant. D'autres auteurs ont décrit des méthodes de modélisation par un graphe plutôt que par un réseau, approche qui consiste en fait à ajouter explicitement un terme de régularisation à la fonction résidu qui est minimisée de façon itérative.

Un graphe associe une topologie au RNA assurant la modélisation paramétrique, en considérant que certains neurones sont connectés les uns aux autres (par une liaison topologique et non une liaison pondérée qui sert à transférer une information). Les neurones sont ainsi considérés en tant que noeuds d'un graphe, les liaisons entre noeuds étant qualifiées d'arêtes. Une énergie est associée au graphe, laquelle dépend de sa géométrie [Jol01].

Dans cette section, nous présentons les deux approches de modélisation par un graphe déformable qui ont été exploitées en reconnaissance de visages, à savoir l'EGM (Elastic Graph Matching) et l'EBGM (Elastic Bunch Graph Matching) [Vis05].

2.4.4.1 EGM : Elastic Graph Matching

Introduite par Lades en 1993, l'EGM permet de modéliser une image par une fonction codée via un graphe formé d'un ensemble de noeuds réunis par une topologie de type grille. Chaque noeud a donc quatre voisins, sauf ceux situés sur les bords ou les coins de la grille [SKC07]. A ce graphe est associée une énergie, calculée en fonction de la différence entre les positions courantes des noeuds et leurs positions initiales, mais également en fonction du contenu de l'image à la position courante des noeuds.

La fonction d'énergie permet donc de coder l'adéquation du modèle à l'image, par

l'intermédiaire d'un terme similaire au résidu, mais également la régularité du modèle en terme géométrique. Cette formulation correspond parfaitement aux techniques de régularisation proposées par Poggio en 1985 pour résoudre les problèmes mal posés [Pog85], largement utilisée par la suite, notamment par Kass *et coll.* pour définir le modèle de contour actif [KWT87].

La figure 2.8 présente un exemple de modélisation d'une image de visage par EGM. En premier lieu, un graphe de référence est calculé sur l'image du visage stockée dans la base, en ajustant un modèle paramétrique sans tenir compte du terme de régularisation. Dans cette image de référence, les noeuds sont donc positionnés sur une grille régulière, à maille carrée dans l'exemple de la figure 2.8. Afin de modéliser une autre image de visage, on initialise tout d'abord un graphe à la même position que celui présent dans l'image de référence. Ensuite, les paramètres du nouveau graphe sont ajustés, par minimisation de son énergie, de telle sorte qu'il décrive le nouveau visage tout en restant régulier. Ici, la régularité est mesurée par la similarité des deux grilles, la régulière correspondant à la référence et celle déformée au cours des itérations.

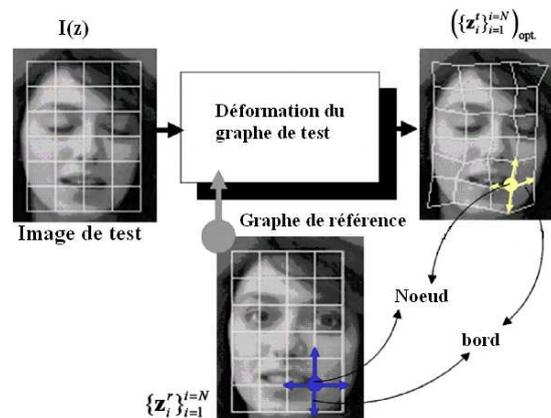


FIGURE 2.8 : Graphe d'EGM rectangulaire sur une image de visage comparée à une image de référence (extraite de [SKC07])

L'EGM n'extrait pas à proprement parler une signature du visage, mais permet de comparer deux visages au moyen de la distance implicite que constitue l'énergie du graphe après minimisation de son énergie. En fonction de cette distance, éventuellement post-traitée par une technique relevant de la classification de données, on détermine la classe d'appartenance de l'image cible. Par exemple, dans [TKP00], l'auteur a associé l'EGM aux

séparateurs à vaste marge, aboutissant à un taux d'erreur de reconnaissance de seulement 5.2% sur tous les visages de la base XM2VTS⁷.

Des méthodes hybrides ont été développées visant à associer l'EGM à d'autres techniques d'extraction de signatures. Dans [SW02], les auteurs ont choisi de détecter les yeux, le nez et la bouche par Eigenface, ensuite de positionner les noeuds du graphe de référence sur ces positions particulières de l'image. La même approche mixte a été appliquée dans [LLC01] et [ZTP05], suivie d'une étape de réduction de dimension par FisherFaces.

2.4.4.2 EBGM : Elastic Bunch Graph Matching

Inspirée de l'EGM, l'EBGM a été introduite par Wiskott en 1997 afin de réduire le nombre de graphes calculés sur la base d'apprentissage [WF99, WFKvdM97]. L'idée consiste à utiliser un seul graphe pour toutes les images d'un même individu de la base, ce qui est cohérent du fait que tous les visages possèdent des caractéristiques morphologiques très similaires. Les noeuds sont positionnés sur l'image du visage pour coder ses éléments caractéristiques, à savoir les yeux, le nez et la bouche, ainsi que des éléments de saillance comme les contours. L'unique graphe de référence est appelé FGB (Face Bunch Graph) et sa construction est obtenue de façon supervisée, c'est à dire avec l'intervention d'un opérateur humain. Pour tester un visage cible, on utilise une méthode de minimisation d'énergie similaire à celle de l'EGM, la distance n'étant calculée qu'avec le FBG de chaque individu.

Depuis son introduction, l'EBGM a fait l'objet de nombreuses études et améliorations, soit sur des images brutes du visage [ZM05a, ZM05b], soit sur des images pré-traitées pour éliminer l'effets des variations d'éclairage et d'orientation du visage [KRG06].

Une variante importante consiste à utiliser d'autres descripteurs du contenu de l'image que les atomes de Gabor, notamment un histogramme des orientations de gradients (HOG) dans des fenêtres centrées sur les points du graphe [MAS08, AMM⁺08]. Avec ces descripteurs, qui sont plus robustes vis à vis des changements de rotation et d'illumination, le taux de reconnaissance passe 85.5% à 97% sur la base de données YaleFaces.

Récemment, Cheung *et coll.* ont proposé une amélioration de l'EBGM [CCM08] qui permet de tenir compte de d'une variation de l'orientation du visage dans l'image. La

7. <http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/>

méthode a été testée sur des images avec une orientation qui peut atteindre un angle de 30 degrés, tout en aboutissant un taux de reconnaissance de 97%.

En dépit de ses très bonnes performances, l'EBGM reste une technique dont l'implémentation est complexe, notamment sur des architectures spécifiques de traitement comme les DSP, et qui nécessite des temps de calcul prohibitifs sur des architectures standard de traitement comme les PC. Les auteurs relatent tous des temps de traitement de l'ordre de la seconde ou même de la dizaine de secondes rien que pour la mise en correspondance de deux graphes. En pratique, cela rend cette technique inutilisable dans la plupart des applications de biométrie.

2.5 Conclusion

L'importance du module d'extraction de signatures est cruciale dans le processus de reconnaissance de visage. De ce fait, nous avons consacré ce chapitre à la présentation des principales méthodes de modélisation paramétrique décrites dans la littérature scientifique. Nous avons présenté au début de ce chapitre un formalisme général permettant de bien situer les méthodes les unes par rapport aux autres. Ce formalisme permet de distinguer les méthodes en fonction des atomes utilisés comme base de la modélisation.

Toutes les méthodes présentées possèdent des avantages et des inconvénients, ce qui explique qu'aucune n'occupe le premier rang devant toutes les autres. En termes de taux de classification, la modélisation par un graphe (EGM ou EBGM) est très performante. Elle bénéficie en fait des avantages des autres modélisations paramétriques (par RNA ou ROG) tout en introduisant une contrainte qui évite la divergence du processus de modélisation. Malheureusement elle n'est pas utilisable en l'état actuel, du fait d'un temps de calcul prohibitif.

Parmi les autres méthodes de modélisation paramétrique très performantes, nous avons retenu celle qui exploite un réseau d'ondelettes de Gabor, proposée par Krüger. Dans la version décrite par cet auteur, soit il faut associer un ROG à chaque image de la base, soit on associe un seul ROG à un individu en ne tenant compte que d'une seule expression faciale. Les deux options ne présentent pas les mêmes performances, du fait qu'il faut réaliser un compromis entre le temps de calcul nécessaire à la reconnaissance

(avec plusieurs ROG) et le taux de reconnaissance (*a priori* moins bon avec un seul ROG).

Dans le chapitre suivant, nous présentons la méthode que nous avons développée durant cette thèse, qui permet de modéliser une série d'images d'un même individu avec un seul ROG. Cette approche permet d'atteindre des meilleurs taux de reconnaissance que celle de Krüger, tout en conservant un temps de calcul en-ligne similaire.

Chapitre 3

Reconnaissance de visages par réseaux d'ondelettes de Gabor

3.1 Introduction

Après avoir présenté le contexte scientifique et l'état de l'art, nous décrivons dans ce chapitre les aspects théorique et pratique d'une nouvelle méthode de reconnaissance de visages. En nous basant sur les travaux de Krüger [Kru00], nous proposons une amélioration de son approche utilisant des réseaux d'ondelettes de Gabor (ROG). Krüger propose de modéliser une image de visage par un ROG et de sauvegarder les paramètres de ce dernier comme signature de l'individu.

Après avoir mis en évidence des limitations de cette approche en termes de non généralisation et de dépendance entre la signature et l'expression faciale, nous proposons une nouvelle méthode qui comble ces lacunes. Nous décrivons également les modules composant le système de reconnaissance de visages exploitant notre nouvelle méthode d'extraction de signatures. Nous détaillons les traitements réalisés dans chaque module, suivant l'architecture standard présentée dans le chapitre 1 et rappelée par le synoptique de la figure 3.1.

Ce chapitre comporte quatre parties principales. Dans la première, nous présentons les aspects théoriques qui nous ont motivés à explorer cette piste. Nous exposons ensuite un aperçu général du système que nous avons conçu pour la reconnaissance de visages. La deuxième partie explique le pré-traitement effectué pour préparer les imagerielettes alimentant

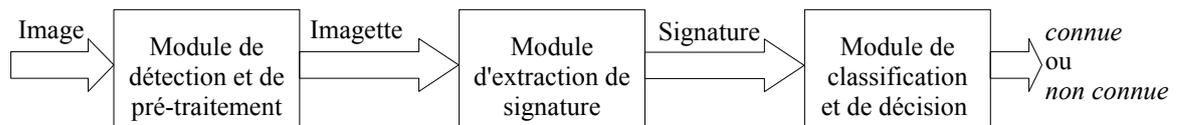


FIGURE 3.1 : Synoptique général d'un système de reconnaissance de visage

le module d'extraction de signatures. Nous décrivons la méthode proposée pour la détection et le pré-traitement du visage. La troisième et la quatrième partie sont consacrées respectivement à l'explication des phases d'enrôlement et d'identification du fonctionnement de notre système.

3.2 Principe de l'approche proposée

Le réseau d'ondelettes de Gabor a prouvé son efficacité pour la modélisation paramétrique d'images. Ceci nous a motivés à l'exploiter en tentant de l'adapter au mieux à l'application de reconnaissance de visages. Pour montrer l'efficacité de la modélisation opérée par un ROG, nous présentons dans la figure 3.2 plusieurs images originales (dans la ligne du haut) et leurs reconstructions par ROG (dans la ligne du bas). Nous constatons que les détails sont bien codés, du fait que l'expression faciale est conservée dans les images reconstruites.

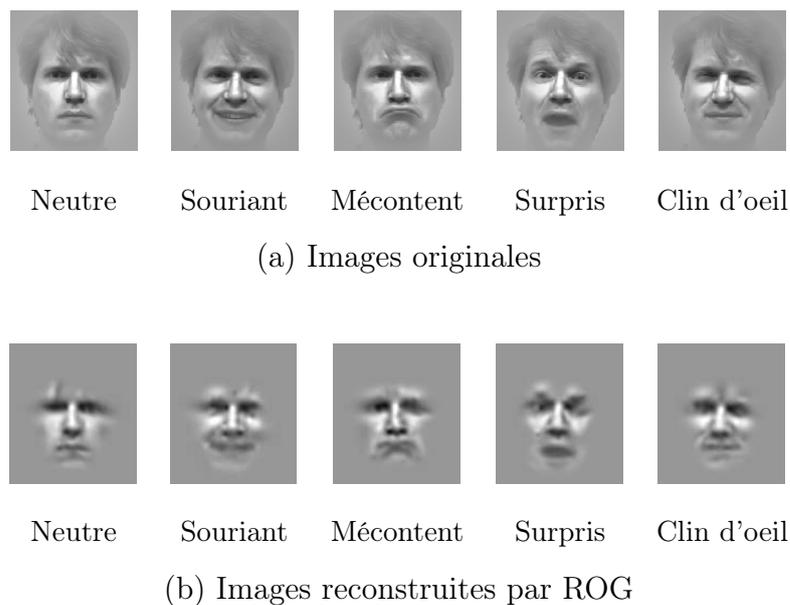


FIGURE 3.2 : Images originales et images reconstruites par un ROG

La principale qualité du ROG, à savoir la précision du modèle obtenu, devient un défaut dans le cadre de l'extraction de signatures. En effet, si le réseau parvient à reconstruire une image qui intègre trop de détails, il fournit une signature qui risque d'être non discriminante pour un individu, car dépendante de l'expression faciale.

Dans l'approche proposée par Krüger, la signature est extraite à partir d'une image unique. Elle ne modélise donc un individu qu'à partir d'une seule expression faciale, ce qui est une limitation majeure de cette méthode. Pour y remédier, notre technique d'apprentissage tient compte de plusieurs expressions faciales de la même personne. Cela permet d'obtenir une modélisation robuste vis-à-vis des différentes conditions d'acquisition de l'image.

En parallèle avec cet apprentissage sur plusieurs images, notre méthode détermine des coefficients mesurant l'importance qu'on peut accorder à chaque ondelette dans le modèle. Ces coefficients permettent de calculer un masque, représenté par une image codant l'importance de chaque pixel de l'image pour la reconnaissance. Nous présentons dans le schéma synoptique de la figure 3.3 l'entrée et la sortie de notre module d'extraction de signatures.

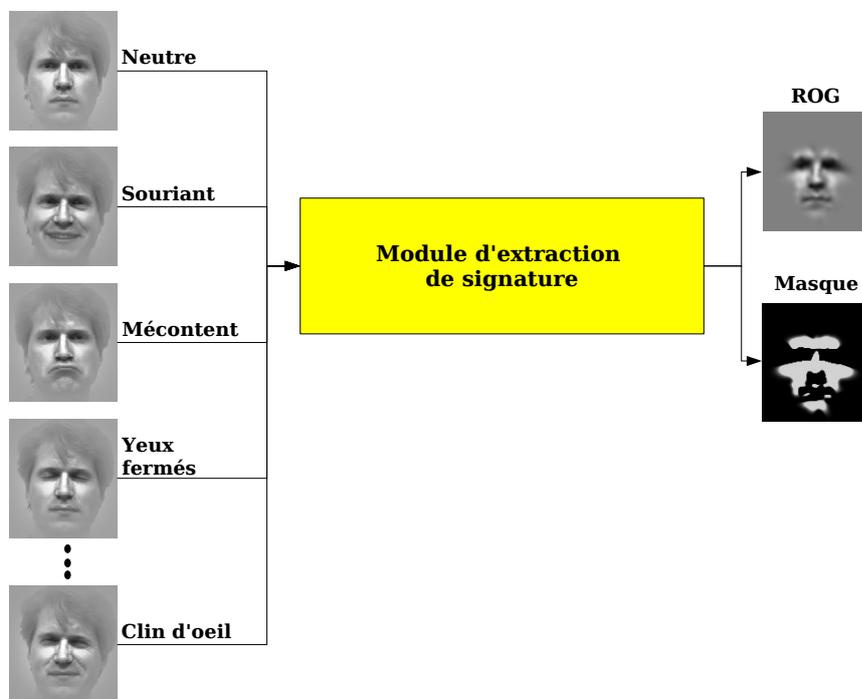


FIGURE 3.3 : Schéma synoptique d'extraction de la signature d'un individu

En définitive, notre méthode fournit pour chaque individu une signature constituée des

paramètres du ROG et d'un masque qui reflète l'importance de chaque ondelette dans le modèle.

3.3 Pré-traitement

Comme nous l'avons indiqué dans le chapitre 1, l'étape de pré-traitement est très importante du fait qu'elle influence directement les performances d'un système de reconnaissance de visages. Dans ce fait, nous avons accordé une attention particulière à cette étape durant notre travail. La figure 3.4 présente le résultat de l'étape de pré-traitement sur une image particulière de visage (extraite de la base YaleFaces).

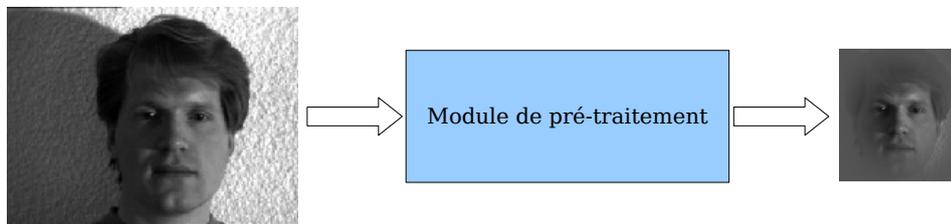


FIGURE 3.4 : Synoptique du module de pré-traitement

Notre module de pré-traitement intègre 2 étapes, la première assurant le recadrage du visage dans l'image et la deuxième un filtrage visant à éliminer les informations non pertinentes présentes sur les bords. Nous détaillons ces deux opérations dans les sections suivantes.

3.3.1 Recadrage des images

Cette étape consiste à délimiter, pour l'instant manuellement, le visage dans les images originales servant à l'apprentissage. De nombreuses techniques automatiques de recadrage ont été décrites dans la littérature, ce qui devrait permettre par la suite d'éliminer l'intervention d'un opérateur. Comme notre objectif premier est de tester l'amélioration de l'étape d'extraction de signatures par ROG, il nous a semblé judicieux de travailler sur des images parfaitement recadrées.

Actuellement, nous sélectionnons par pointage à la souris une partie précise dans l'image, puis élargissons la fenêtre pour qu'elle visualise tous les organes (yeux, nez et bouche) à des positions quasiment identiques dans toutes les images d'une personne (cf.

figure 3.5). Nous avons également envisagé de déformer les images, par une homothétie et/ou une inclinaison, de telle sorte que les organes apparaissent toujours à la même position dans l'imagette. Sur les images utilisées pour l'évaluation, cette déformation ne s'est pas avérée nécessaire.

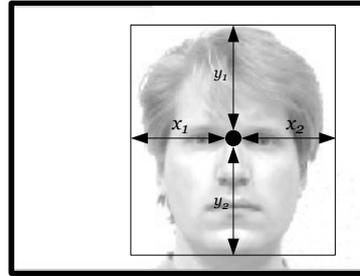


FIGURE 3.5 : Recadrage d'une image

La figure 3.6 présente le recadrage des images correspondant à un même individu de la base. On constate que les organes principaux du visage sont correctement recadrés sans qu'il soit nécessaire de redimensionner les images.

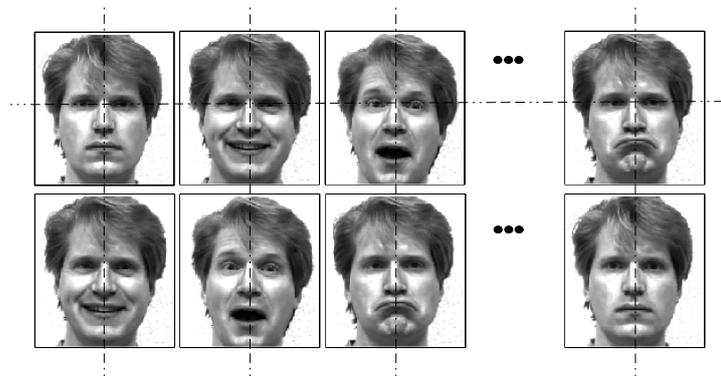


FIGURE 3.6 : Recadrage de toutes les images d'un même individu

3.3.2 Filtrage

L'étape de filtrage permet d'atténuer les effets de bord qui pourraient perturber le processus de modélisation par le ROG. En effet, si on optimise les paramètres du ROG sur une imagette brute, certaines ondelettes vont se positionner à proximité du contour extérieur du visage (séparant le visage du fond), afin de modéliser ce changement notable du niveau de gris.

Nous considérons que ces contours extérieurs ne sont pas significatifs pour l'extraction d'une signature, en tout cas qu'ils le sont moins que les caractéristiques des organes comme les yeux, le nez ou la bouche. Dès lors, le filtrage consiste à diminuer l'amplitude de la fonction niveau de gris sur les bords de l'image, en la ramenant à une valeur constante correspondant à un gris moyen.

En pratique, cela consiste à combiner l'image et une fonction de pondération selon l'équation suivante :

$$I_f(x, y) = \bar{I}_o + F(x, y) \cdot (I_o(x, y) - \bar{I}_o) , \quad (3.1)$$

dans laquelle $I_o(x, y)$ désigne l'image originale, \bar{I}_o sa valeur moyenne, $I_f(x, y)$ l'image filtrée et $F(x, y)$ la fonction de pondération. Nous avons utilisé une fonction de pondération de Hamming, laquelle prend une valeur maximale égale à 1 au centre de la fenêtre et s'annule sur les bords.

La figure 3.8 présente une image originale, le graphe de la fonction de pondération de Hamming et l'image résultante.

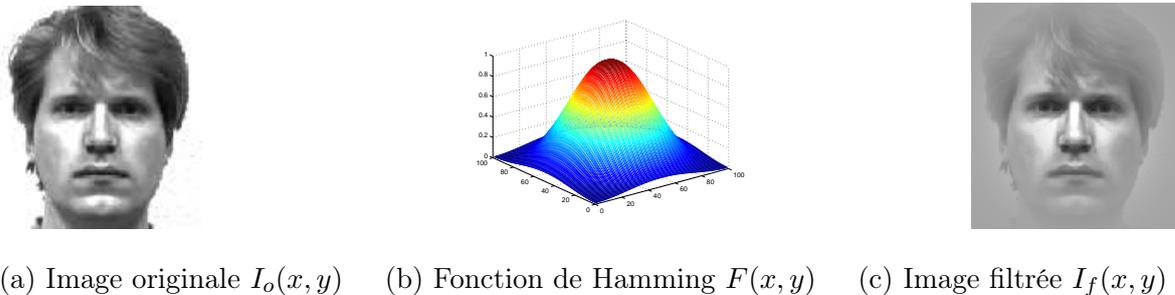


FIGURE 3.7 : Filtrage des images

Pour montrer l'intérêt de cette étape de filtrage, nous présentons sur la figure 3.8 deux résultats de modélisation de la même image d'un visage par un ROG, avec et sans filtrage. Dans l'image (a), on constate que quelques ondelettes se sont déplacées vers le bord du visage pour coder des éléments du fond ou les cheveux, alors que dans l'image (b) toutes les ondelettes modélisent des variations du niveau de gris correspondant à des organes caractéristiques.



(a) Modélisation par ROG sans filtrage (b) Modélisation par ROG avec filtrage

FIGURE 3.8 : Intérêt du filtrage des images

3.4 Phase d'entraînement

La phase d'entraînement utilise une base de données où sont stockées plusieurs images faciales pour chaque individu. Durant cette phase, le système est chargé d'extraire une signature pour chaque individu et de la stocker dans une base de données dédiée.

Le traitement d'entraînement est schématisé par la figure 3.9. Le module d'apprentissage traite plusieurs images de chaque individu, applique l'algorithme d'apprentissage et fournit la signature. Ce module contient 3 étapes principales, à savoir l'initialisation des réseaux, l'optimisation des paramètres des réseaux avec N_a images au lieu d'une seule et le calcul des masques. Nous détaillons les traitements de ces étapes dans les sections suivantes.

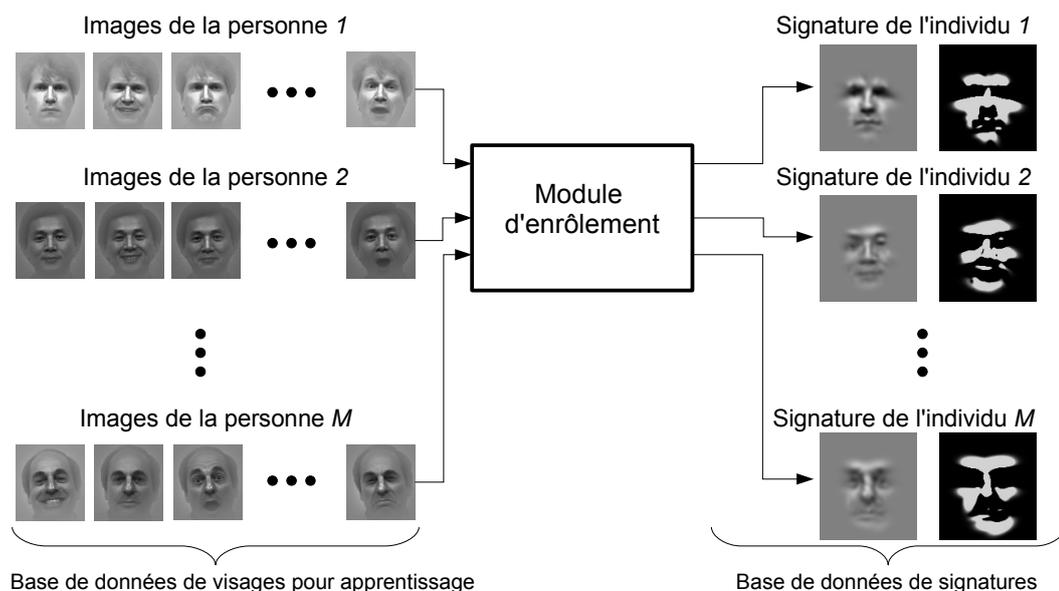


FIGURE 3.9 : Synoptique de la phase d'entraînement

3.4.1 Initialisation

L'étape d'initialisation du ROG est une phase importante, qui demande une attention particulière pour garantir la bonne convergence de l'algorithme de minimisation du résidu.

Avant de procéder à l'initialisation des paramètres du ROG, à savoir les translations t_x et t_y , les dilatations d_x et d_y , l'orientation θ et le poids ω de chaque ondelette, nous devons choisir le nombre d'ondelettes intervenant dans le réseau. La figure 3.10 montre l'influence du nombre d'ondelettes sur la qualité de l'image reconstruite. Nous constatons qu'en toute logique plus le nombre d'ondelettes est élevé plus l'image reconstruite présente de détails.

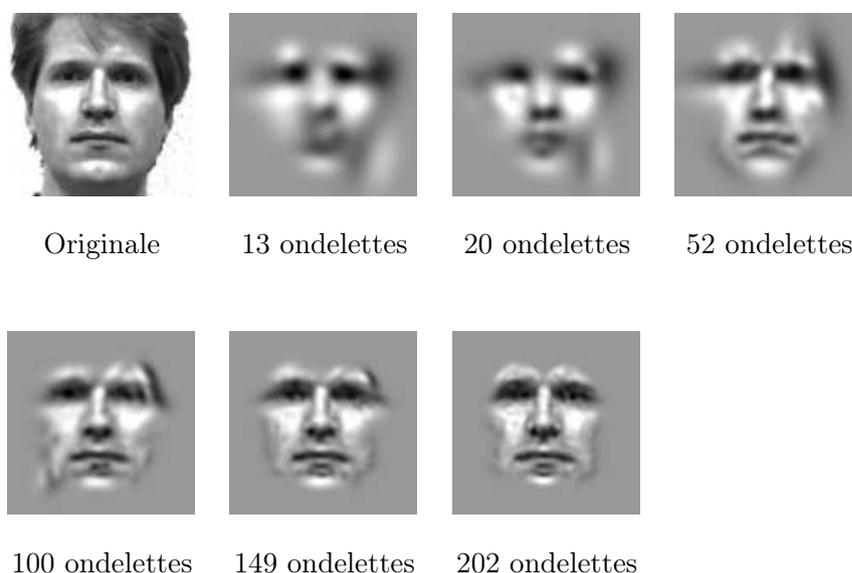


FIGURE 3.10 : Précision du modèle en fonction du nombre d'ondelettes

Dans les réseaux de neurones, l'initialisation des paramètres avant apprentissage se fait en général d'une manière aléatoire. Les paramètres des fonctions et les poids sont choisis aléatoirement dans des intervalles bornés. Cette initialisation aléatoire suppose que la technique de minimisation du résidu permet d'atteindre un minimum local qui décrive correctement les données. Malheureusement, les techniques d'optimisation dont nous disposons, même les plus sophistiquées, ne parviennent pas à ce résultat quand il s'agit de modéliser une image.

L'ondelette de Gabor est une fonction à décroissance rapide définie sur un support compact. L'essentiel de sa contribution à la modélisation d'une fonction se limite donc à

un intervalle restreint de cette dernière. Si les paramètres de translation sont initialisés en dehors d'un domaine d'intérêt (à l'extérieur de l'image dans notre cas), ou si les facteurs de dilatation initiaux sont inadaptés, la contribution d'une ondelette dans le modèle s'avère pratiquement nulle. Ceci entraîne une convergence difficile, sinon impossible de l'algorithme de minimisation, quel que soit le nombre d'itérations.

De ce fait, nous proposons comme l'a fait Krüger d'utiliser une procédure d'initialisation adaptée au contenu de l'image. Nous initialisons les paramètres de translation de manière à placer initialement toutes les ondelettes du ROG sur une grille positionnée au centre du visage. De cette manière, nous nous assurons du fait qu'elles modéliseront la partie intéressante de l'image, à savoir les organes distinctifs.

La figure 3.11 indique sur un exemple les positions initiales et finales des ondelettes avant et après l'apprentissage. Chaque point noir représente le centre d'une ondelette, c'est à dire les composantes de translation t_x et t_y de sa paramétrisation à partir de l'ondelette mère. Après apprentissage, on constate que ces paramètres ont été adaptés afin de modéliser correctement le visage alors que les ondelettes sont restées centrées sur la partie significative de la fonction image. Le minimum du résidu est donc local, mais significatif en terme de modélisation de la fonction image.



(a) Positions initiales des ondelettes sur l'image originale (b) Positions finales des ondelettes sur l'image reconstruite

FIGURE 3.11 : Positions des ondelettes sur l'image avant et après apprentissage

L'initialisation des autres paramètres est réalisée de manière classique. Les dilatations sont choisies afin que les ondelettes puissent initialement décrire de façon correcte le voisinage du point de la grille sur lequel elles sont positionnées. Les orientations et les poids sont choisis aléatoirement à partir d'un ensemble borné ($[-\pi, \pi]$ pour θ et $[0, 100]$ pour ω).

3.4.2 Apprentissage avec plusieurs images

Dans notre approche, nous proposons d'optimiser les paramètres du ROG en utilisant plusieurs images du visage d'un même individu, lesquelles correspondent à différentes expressions faciales. Pour ce faire, l'image cible est changée périodiquement durant l'apprentissage, sans remise à zéro des paramètres du réseau (cf. algorithme 1). En définitive, les paramètres du modèle sont ajustés afin qu'il représente aux mieux toutes ces images.

L'algorithme que nous proposons traite automatiquement tous les individus de la base de données. Avant de lancer l'exécution, l'opérateur doit définir les valeurs de quelques méta-paramètres, à savoir :

- les N_a images de chaque individu, de préférence avec différentes expressions faciales, qui seront utilisées pour l'ajustement du modèle ;
- le nombre L d'ondelettes de Gabor intervenant dans le modèle paramétrique ;
- le nombre Q_i d'itérations séparant deux changements d'image cible durant l'apprentissage ;
- le nombre Q_t de changements d'image cible durant l'apprentissage.

Au début des itérations d'optimisation visant à déterminer la signature d'un individu m , l'algorithme charge ses N_a images caractéristiques dans un buffer dédié. Ensuite, il crée un nouveau ROG et initialise ses paramètres selon la procédure décrite dans la section 3.4.1. Une fois les paramètres initialisés, l'algorithme sélectionne aléatoirement une image cible sur laquelle va porter la minimisation du résidu. Après Q_i itérations, une autre image cible est sélectionnée aléatoirement parmi les N_a images disponibles. L'apprentissage pour un individu nécessite donc $Q_t \times Q_i$ itérations de minimisation du résidu. Le processus d'apprentissage pour tous les individus est schématisé par l'algorithme 1.

La méthode de Levenberg-Marquardt [HM94] est utilisée pour minimiser le résidu. Elle réalise un bon compromis entre la descente du gradient (simple mais nécessitant de nombreuses itérations) et les méthodes qui impliquent de calculer la matrice Hessienne à chaque itération. Dans la méthode de Levenberg-Marquardt, la matrice Hessienne est évaluée progressivement au cours des itérations, ce qui réduit le temps de calcul tout en

conservant les propriétés de bonne convergence des méthodes quasi-newtoniennes.

```

for  $m \leftarrow 1$  to  $M$  do
  | for  $n \leftarrow 1$  to  $N_a$  do
  | |  $I[n] \leftarrow \text{Charger-Image}(I_{m,n}(x, y));$ 
  | end
  |  $(\mathbf{p}[m], \omega[m]) = \text{Initialiser-ROG}();$ 
  | for  $k \leftarrow 1$  to  $Q_t$  do
  | |  $f = \text{random}(N_a);$ 
  | | for  $l \leftarrow 1$  to  $Q_i$  do
  | | |  $(\mathbf{p}[m], \omega[m]) = \text{Optimiser-ROG}(\mathbf{p}[m], \omega[m], I[f]);$ 
  | | |  $\mathbf{P}[m][Q_i \cdot (k - 1) + l] = \mathbf{p}[m];$ 
  | | |  $\Omega[m][Q_i \cdot (k - 1) + l] = \omega[m];$ 
  | | end
  | end
end

```

Algorithm 1: Algorithme d'apprentissage

Durant l'apprentissage, tous les paramètres du réseau ainsi que le numéro d'image cible sont sauvegardés à chaque itération. Cela permet en premier lieu d'analyser *a posteriori* si l'apprentissage a fourni des signatures significatives. En outre, une analyse statistique de l'évolution des paramètres au cours des itérations nous permet d'évaluer la confiance qu'on peut accorder à chaque ondelette du modèle paramétrique.

3.4.3 Analyse des données d'apprentissage

Une étude détaillée de l'évolution au cours de l'apprentissage des paramètres caractérisant les ondelettes a été menée en premier lieu manuellement. Nous avons constaté que certaines ondelettes voient leurs paramètres évoluer de façon significative au cours des itérations, notamment après un changement de l'image cible. En revanche, pour certaines ondelettes, les paramètres convergent rapidement vers des valeurs stables, qui ne dépendent que très peu de l'image cible.

En nous référant à la position des ondelettes dans l'image du visage, nous avons constaté (ce qui était prévisible) que les ondelettes dont les paramètres sont instables sont celles qui modélisent des parties du visage dont l'apparence est très dépendante

de l'expression faciale. Les zones correspondantes du visage sont qualifiées de *variables*. En revanche, les ondelettes qui modélisent une zone *stable* du visage, conservent des paramètres assez similaires quelle que soit l'expression faciale visible dans l'image.

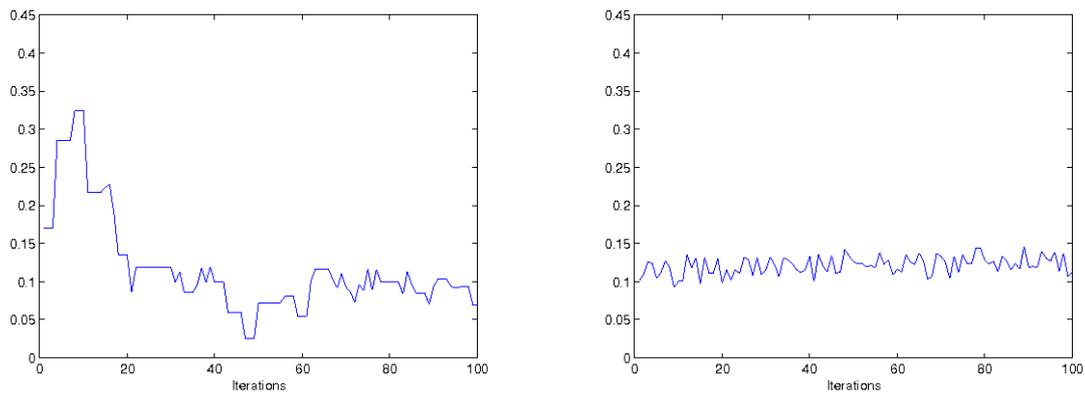
La variabilité des paramètres s'explique aisément en considérant l'influence de chaque ondelette dans la valeur du résidu. Si une région est très variable, la différence entre le modèle et l'image est importante dans cette région, notamment à chaque changement de l'image cible. Le processus de minimisation tend alors à modifier les paramètres des ondelettes modélisant les régions variables, car leur contribution dans le résidu est importante [CCSM08].

Par exemple, l'aspect de la région située autour de la bouche est très variable, alors que celui de la région située à proximité du nez est relativement stable. On vérifie aisément cette propriété en comparant visuellement les régions présentées sur la figure 3.12, qui correspondent à des fenêtres rectangulaires extraites des images de différentes expressions faciales d'un même individu. Toutes les images du nez (ligne du haut) apparaissent identiques, alors que les images de la bouche (ligne du bas) apparaissent très différentes.



FIGURE 3.12 : Zoom sur les zones stables et variables du visage pour différentes expressions faciales

Pour visualiser plus clairement les variations des paramètres en fonction de la région que modélise une ondelette, nous avons représenté sur la figure 3.13 l'évolution d'un paramètre particulier, à savoir le coefficient de dilatation d_x , pour deux ondelettes situées initialement à proximité du nez (région stable) et de la bouche (région instable). On vérifie aisément que d_x évolue constamment pour l'ondelette modélisant une région proche de la bouche (figure 3.13(a)), alors qu'il reste presque constant pour l'ondelette modélisant une région proche du nez (figure 3.13(b)).



(a) Ondelette à proximité de la bouche

(b) Ondelette à proximité du nez

FIGURE 3.13 : Variation de d_x en fonction du numéro d'itération

On peut également visualiser la stabilité des paramètres de translation t_x et t_y en marquant dans l'image du visage les positions successives du centre de l'ondelette au cours des itérations. La figure 3.14 montre clairement que le centre d'une ondelette modélisant une région de l'image proche de la bouche varie fortement durant les itérations alors que le centre d'une ondelette modélisant une région proche du nez reste presque fixe¹.



(a) Ondelette placée sur la bouche (b) Ondelette placée sur le nez

FIGURE 3.14 : Variations des positions des centres d'ondelettes particulières durant l'apprentissage

En conclusion, nous constatons que la variabilité des paramètres d'une ondelette est directement liée à celle de la région du visage qu'elle modélise. Cette propriété nous a incités à associer à chaque ondelette un indice de confiance, dont la valeur dépend de la variabilité des paramètres de cette ondelette durant l'apprentissage.

¹. Comme dans la section 3.4.1, les positions des ondelettes sont repérées par des points noirs dans les images.

3.4.4 Indice de confiance associé à une ondelette

Afin de constituer la signature d'un individu, nous avons précédemment indiqué qu'il fallait sélectionner des éléments qui permettent de le différencier d'une autre personne, tout en restant le plus indépendant possible des facteurs de variation de l'image, notamment l'expression faciale. De ce fait, les régions de l'image que nous avons qualifiées de stables peuvent être considérées comme caractéristiques de l'individu, contrairement aux régions instables.

Pour tenir compte de la confiance qu'on peut accorder à une région du visage en terme de reconnaissance, nous proposons de tenir compte de la variabilité de cette région selon l'expression faciale. Nous avons vu que la variabilité d'une région peut être quantifiée en analysant celle, durant la phase d'apprentissage, des paramètres des ondelettes qui la modélisent. De ce fait, il suffit d'associer à chaque ondelette, plutôt qu'à chaque région, un indice de confiance qui est d'autant plus faible que les paramètres sont variables. Ces indices de confiance, associés aux paramètres du ROG, constitueront la signature de l'individu.

Pour mesurer la variabilité d'une ondelette, nous évaluons en premier lieu les variances de toutes les valeurs prises par ses cinq paramètres au cours des itérations d'apprentissage. Une variance élevée indique que le paramètre varie beaucoup et donc que la confiance qu'on peut lui accorder est faible. On peut trouver de nombreuses expressions combinant les cinq valeurs de variance afin de déterminer un unique coefficient de confiance. Notre objectif étant simplement de valider le principe d'association d'une confiance à chaque ondelette, nous avons utilisé une approche triviale.

Elle consiste à normaliser les variances, puis à calculer leur somme. Une transformation non linéaire, par une fonction sigmoïde, permet de ramener le résultat dans l'intervalle $[0 \dots 1]$. Le résultat est une valeur proche de 0 quand les variances sont élevées, et proche de 1 quand les variances sont faibles [CCSM09]. Plus précisément, les coefficients de confiance sont calculés de la façon suivante :

1. on calcule les variances des cinq paramètres t_x , t_y , d_x , d_y et θ pour tous les apprentissages réalisés sur tous les individus de la base. Ces variances globales vont permettre la normalisation des variances de ces mêmes paramètres pour un individu particulier. Pour le calcul de normalisation, on détermine les cinq valeurs moyennes

$[\bar{V}_{t_x}, \bar{V}_{t_y}, \bar{V}_\theta, \bar{V}_{d_x}, \bar{V}_{d_y}]$ et les cinq écarts-types $[\sigma_{t_x}, \sigma_{t_y}, \sigma_\theta, \sigma_{d_x}, \sigma_{d_y}]$ de ces différentes variances.

- pour chaque ondelette l du ROG modélisant le visage de l'individu m , on normalise les variances des séries de valeurs prises par les cinq paramètres. Par exemple, on normalise la variance du paramètre t_x de l'ondelette l du ROG de l'individu m en utilisant l'expression :

$$\Delta_{m,l}^{t_x} = \frac{\text{Var}(\{t_{x_q} \mid q \in [1 \cdots Q_t \times Q_i]\}) - \bar{V}_{t_x}}{\sigma_{t_x}},$$

dans laquelle le premier terme du membre de droite désigne la variance de la série constituée de toutes les valeurs prises par le paramètre t_x lors de l'apprentissage pour l'individu m .

- on transforme la somme des variances normalisées via une fonction sigmoïde pour recadrer l'indice entre 0 et 1 :

$$C_{m,l} = \frac{1}{1 + \exp(\tau(\Delta_{m,l}^{t_x} + \Delta_{m,l}^{t_y} + \Delta_{m,l}^\theta + \Delta_{m,l}^{d_x} + \Delta_{m,l}^{d_y}))}, \quad (3.2)$$

dans laquelle τ est un méta-paramètre permettant d'ajuster l'étendue de la sigmoïde.

Le méta-paramètre τ permet de régler la variabilité de l'indice de confiance. La figure 3.15 représente les courbes représentatives de la fonction sigmoïde pour deux valeurs différentes de τ . Plus on augmente τ , plus l'indice de confiance associé à une ondelette varie rapidement quand les variances de ses paramètres se rapprochent de leurs valeurs moyennes.

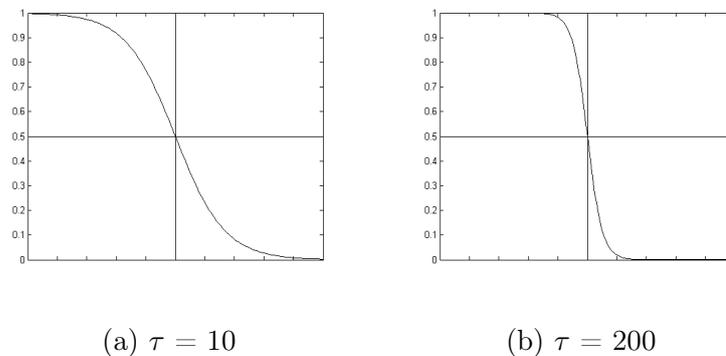


FIGURE 3.15 : Fonction sigmoïde pour deux valeurs du méta-paramètre τ

3.4.5 Masque des confiances

Afin de tenir compte des coefficients de confiance dans le processus de reconnaissance, nous avons choisi de faire intervenir un masque de pondération déterminé à partir de ces coefficients. Le masque permet de transformer les confiances associées à chaque ondelette en confiances associées aux pixels de l'image. Une valeur proche de zéro dans le masque indique que le pixel n'est pas significatif, c'est à dire qu'il est situé dans une région variable, alors qu'une valeur élevée indique qu'il appartient à une région stable.

Le masque des confiances est construit sous la forme d'une image en niveau de gris $I_m^c(x, y)$ (ayant la dimension de l'image optimisée), ce qui permet de le visualiser simplement. Dans cette image, les zones claires correspondent à des régions considérées comme stables et les zones sombres à des régions instables.

Le masque $I_m^c(x, y)$ est déterminé par une combinaison linéaire de fonctions de Gabor modifiées $\psi_{\mathbf{p}_l}^{g'}(x, y)$, selon l'expression :

$$I_m^c(x, y) = \sum_{l=1}^L C_{m,l} \psi_{\mathbf{p}_l}^{g'}(x, y), \quad (3.3)$$

dans laquelle \mathbf{p}_l désigne le vecteur paramètre $(d_x \ d_y \ \theta \ t_x \ t_y)^T$ associé à l'ondelette d'indice l dans le ROG de l'individu m après apprentissage.

Dans les fonctions de Gabor modifiées $\psi_{\mathbf{p}_l}^{g'}(x, y)$, le terme oscillant a été supprimé de telle sorte que la confiance associée à une ondelette soit prise en compte de façon maximale dans l'image au niveau du centre de l'ondelette. $\psi_{\mathbf{p}_l}^{g'}(x, y)$ est donnée par l'expression suivante :

$$\begin{aligned} \psi_{\mathbf{p}_l}^{g'}(x, y) = & \exp \left[-\frac{1}{2} \left(d_{x_l} \cos \theta (x - t_{x_l}) \right. \right. \\ & - \left. d_{x_l} \sin \theta (y - t_{y_l}) \right)^2 - \frac{1}{2} \left(d_{y_l} \sin \theta (x - t_{x_l}) \right. \\ & \left. \left. + d_{y_l} \cos \theta (y - t_{y_l}) \right)^2 \right]. \end{aligned} \quad (3.4)$$

La figure 3.16 montre l'allure de cette fonction modifiée. Son support est identique à celui de l'ondelette initiale, c'est à dire que la confiance porte sur une zone ayant la même étendue et la même orientation que l'ondelette de Gabor qui intervient dans le modèle.

La figure 3.17 illustre l'intérêt du masque des confiances associé à un modèle du visage obtenu par un ROG. Dans cette figure, l'image (a) représente les centres des ondelettes

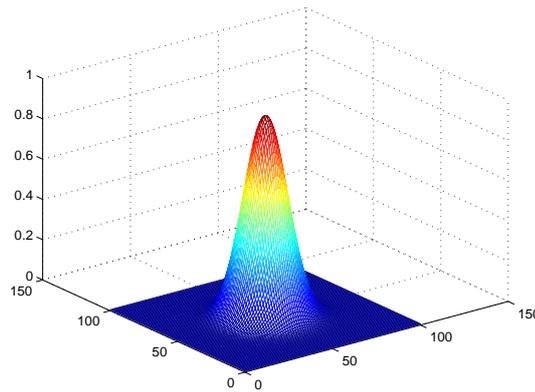
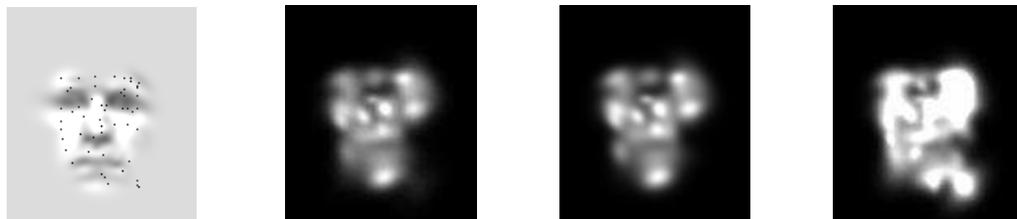


FIGURE 3.16 : Allure d'une fonction $\psi_{\mathbf{p}}^g(x, y)$ avec des valeurs initiales de \mathbf{p} et w

superposés au modèle paramétrique codé par le ROG après apprentissage pour un individu. L'image (b) de cette même figure représente le masque associé, obtenu avec des coefficients de confiance avec une valeur du méta-paramètre τ égale à 10. Une zone claire dans ce masque correspond à une région qui peut être considérée comme stable dans l'image. On constate par exemple que la région située à proximité du nez est relativement stable. Les images (c) et (d) représentent des masques obtenus avec des valeurs plus élevées du méta-paramètre τ . On constate que les images sont plus contrastées, ce qui indique des transitions plus franches entre les régions stables et variables. L'augmentation de contraste est due au fait que les coefficients de confiance associés à chaque ondelette vérifient également de façon plus rapide quand les paramètres de ces dernières s'approchent des variances moyennes.



(a) Positions finales des ondelettes sur \hat{I} (b) Masque obtenu pour $\tau = 10$ (c) Masque obtenu pour $\tau = 50$ (d) Masque obtenu pour $\tau = 200$

FIGURE 3.17 : Modèle du visage et masque associé

En conclusion, l'approche proposée nous permet d'obtenir un codage pertinent des attributs les plus discriminants d'un visage à partir de différentes expressions faciales. A

l'issue de la phase d'enrôlement, la base de données de signatures est constituée des ROG optimisés pour chaque individu à partir de plusieurs images, chacun étant associé à un masque indiquant les régions stables du visage. La signature \mathbf{s}_m d'un individu m est ainsi un vecteur \mathbf{p}_m contenant les paramètres $\mathbf{p}_{m,l}$ (5 scalaires), les poids $\omega_{m,l}$ et les confiances $C_{m,l}$ des L ondelettes du Gabor intervenant dans le ROG.

Nous avons vu qu'à partir de la signature \mathbf{s}_m de l'individu, on peut construire les deux images $\hat{I}_m(x, y)$ et $I_m^c(x, y)$, qui désignent respectivement le modèle de la fonction image et le masque des confiances. Dans la suite, nous utilisons indifféremment ces deux représentations de la signature d'un individu.

3.5 Phase d'identification

Durant la phase d'identification, étant donnée une image de visage $I(x, y)$ qualifiée d'image cible, il s'agit de déterminer à quel individu m de la base elle correspond. Pour ce faire, nous utilisons uniquement les signatures \mathbf{s}_m , $1 \leq m \leq M$, extraites pour chaque individu durant la phase d'enrôlement.

Une approche triviale consisterait à calculer les distances entre l'image $I(x, y)$ et les modèles $\hat{I}_m(x, y)$ pour tous les individus, puis à sélectionner la distance minimale comme critère d'association. Krüger a montré qu'il est plus intéressant de vérifier si les ondelettes constituant chaque signature peuvent servir de base pour obtenir une modélisation paramétrique précise de l'image cible.

Pour chaque individu m , il s'agit donc de déterminer les L poids $\omega'_{m,l}$ permettant la modélisation optimale (avec le plus faible résidu) de l'image $I(x, y)$ en utilisant comme atomes les L ondelettes du modèle paramétrique du ROG codé par la signature \mathbf{s}_m . Ensuite, plutôt que de déterminer la distance entre l'image cible et le modèle, on détermine la distance entre la modélisation paramétrique — sur la base des ondelettes du ROG — de cette image et le modèle.

3.5.1 Calcul direct du vecteur poids

Pour déterminer la modélisation paramétrique de l'image cible sur la base des ondelettes du ROG de l'individu m , on peut utiliser le fait que les ondelettes de Gabor

possèdent la propriété de bi-orthogonalité. De ce fait, on peut déterminer une ondelette duale pour chaque ondelette du ROG. Les ondelettes duales permettent de calculer directement le vecteur des poids $\omega'_{m,l}$ correspondant à la modélisation paramétrique recherchée.

Étant donné l'ensemble $\Psi_m = \{\psi_{m,l} \mid 1 \leq l \leq L\}$ des L ondelettes du ROG modélisant l'individu m , on peut déterminer l'ensemble $\tilde{\Psi}_m = \{\tilde{\psi}_{m,l} \mid 1 \leq l \leq L\}$ des L ondelettes duales en utilisant les techniques standard de transformation inverse décrites dans la littérature [Kru00]. On utilise le fait que les ondelettes originales et les ondelettes duales vérifient la propriété suivante :

$$\langle \psi_{\mathbf{p}_i}, \tilde{\psi}_{\mathbf{p}_j} \rangle = \delta_{i,j}, \quad (3.5)$$

dans laquelle $\langle \cdot, \cdot \rangle$ désigne le produit scalaire de deux fonctions.

En notation matricielle, cette propriété s'exprime selon l'équation :

$$\Psi_m \tilde{\Psi}_m = (\langle \psi_{p_i}, \tilde{\psi}_{p_j} \rangle)_{i,j} = \mathbb{I}^L \quad (3.6)$$

dans laquelle \mathbb{I}^L désigne la matrice identité de dimension $L \times L$. En inversant cette équation matricielle, on obtient l'expression qui permet de calculer les ondelettes duales à partir des ondelettes originales :

$$\tilde{\Psi}_m = (\Psi_m^T \Psi_m)^{-1} \Psi_m^T. \quad (3.7)$$

Au moyen des ondelettes duales, on peut déterminer directement les poids qu'il faut affecter aux ondelettes Ψ_m pour obtenir la modélisation paramétrique recherchée. Le poids de l'ondelette d'indice l est donné par le produit scalaire de la fonction image $I(x, y)$ et de l'ondelette duale $\tilde{\psi}_{m,l}$:

$$\omega'_{m,l} = \langle I, \tilde{\psi}_{m,l} \rangle. \quad (3.8)$$

3.5.2 Projection / reconstruction

Étant donnée l'image cible $I(x, y)$, nous avons vu qu'il est possible de déterminer la modélisation paramétrique optimale de cette image sur la base des ondelettes du ROG caractérisant l'individu m . Le résultat de la modélisation est une image, déterminée à partir de la combinaison linéaire des L ondelettes du ROG de l'individu m par les poids $\omega'_{m,l}$ déterminés précédemment :

$$\text{PR}_m(I)(x, y) = \sum_{l=1}^L \omega'_{m,l} \psi_{\mathbf{p}_{m,l}}^g(x, y), \quad (3.9)$$

dans laquelle $PR_m(\cdot)$ désigne la transformation globale permettant de passer de l'image originale au modèle paramétrique sur la base des ondelettes du ROG de l'individu m . Nous avons choisi la notation $PR_m(\cdot)$, car la transformation peut être considérée comme une projection de l'image sur la base des ondelettes, suivie d'une reconstruction.

La figure 3.18 montre l'intérêt de cette opération de projection/reconstruction. Sur la première ligne, on trouve deux images (a) et (b) d'un même individu ainsi qu'une image d'un autre individu. L'image (a) a été utilisée afin de déterminer la signature de l'individu 1 via un ROG obtenu selon la méthode de Krüger. La ligne du bas présente les images obtenues par la transformation $PR_1(\cdot)$ sur les trois images de la première ligne. On constate que les deux images (d) et (e), qui correspondent à l'individu 1, restent globalement reconnaissables après cette transformation. En revanche, la projection de l'image du visage de l'individu 2 sur le ROG codant l'individu 1, suivie d'une reconstruction, fournit un résultat très différent (image (f)), du fait que les ondelettes du ROG ne sont pas adaptées à l'individu 2.

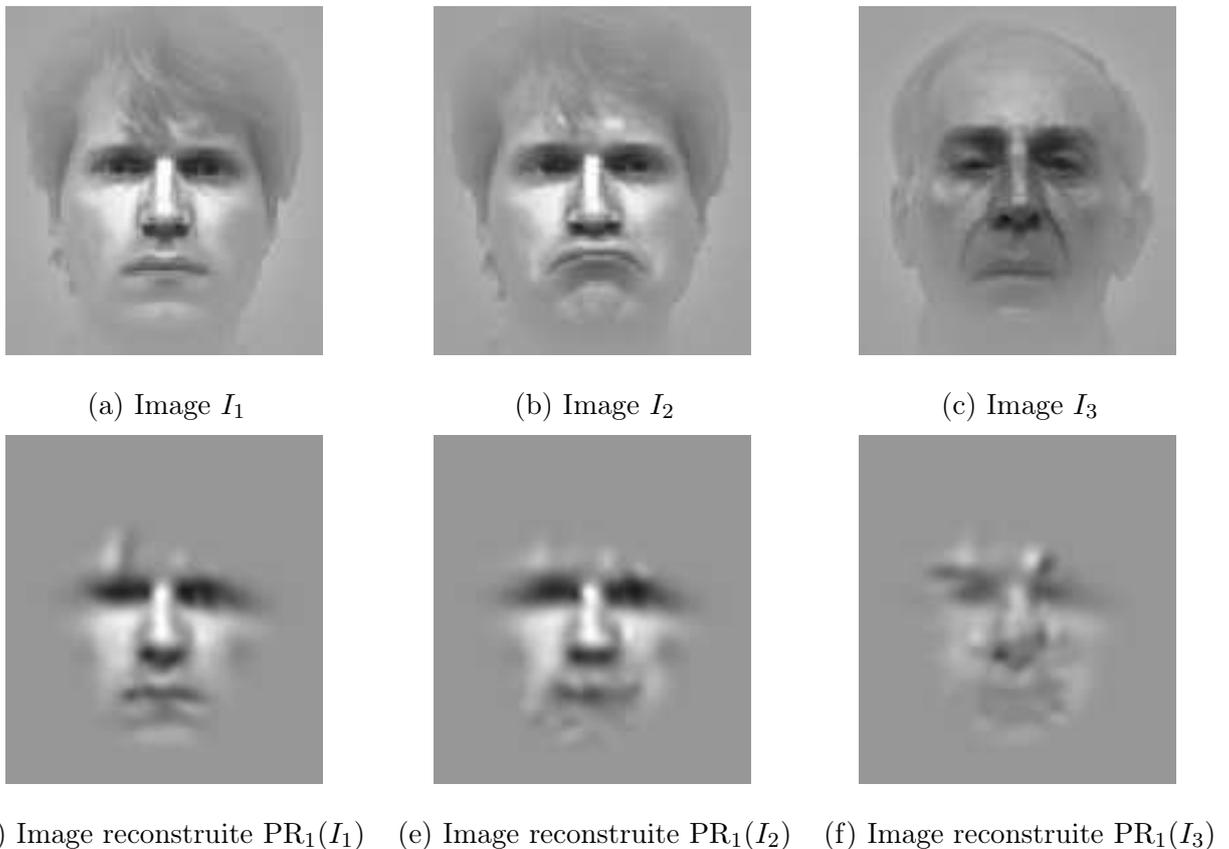


FIGURE 3.18 : Images initiales et après projection / reconstruction

3.5.3 Calcul des distances

Pour déterminer un indice de dissimilarité entre une image cible et le modèle de la base caractérisant l'individu m on peut maintenant calculer une distance entre le ROG de ce dernier et l'image obtenue par projection/reconstruction de l'image cible. Parmi toutes les expressions possibles définissant une distance, nous avons retenu la distance quadratique (comme base de référence) et la distance de corrélation.

La distance quadratique entre deux images $I_1(x, y)$ et $I_2(x, y)$ de taille $X \times Y$, qui correspond au carré de la distance euclidienne, est obtenue selon l'expression suivante :

$$d^2(I_1, I_2) = \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} (I_1(x, y) - I_2(x, y))^2, \quad (3.10)$$

et la distance de corrélation par l'expression :

$$d^c(I_1, I_2) = 1 - \frac{\sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} (I_1(x, y) - \bar{I}_1)(I_2(x, y) - \bar{I}_2)}{\sqrt{\left(\sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} (I_1(x, y) - \bar{I}_1)\right) \cdot \left(\sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} (I_2(x, y) - \bar{I}_2)\right)}} \quad (3.11)$$

dans laquelle \bar{I}_1 et \bar{I}_2 sont respectivement les niveaux de gris moyens des images $I_1(x, y)$ et $I_2(x, y)$. La fraction intervenant dans cette expression est le coefficient de corrélation calculé entre les deux séries de valeurs constituées des niveaux de gris des deux images.

Pour calculer l'indice de dissimilarité entre l'image cible $I(x, y)$ et le modèle $\hat{I}_m(x, y)$ de l'individu m , on peut calculer les distances d^2 et d^c entre les images $\text{PR}_m(I)(x, y)$ et $\hat{I}_m(x, y)$, ou entre ces mêmes images mais après pondération par le masque de confiance contenu dans la signature de cet individu. Dans ce deuxième cas, les pixels appartenant à des régions variables de l'image interviennent avec un poids plus faible dans l'expression de la distance.

On détermine ainsi, pour chaque individu m de la base, quatre indices de dissimilarité associés à l'image $I(x, y)$ calculés de la façon suivante :

$$\begin{aligned} S_m^2(I) &= d^2(\text{PR}_m(I), \hat{I}_m), \\ S_m^c(I) &= d^c(\text{PR}_m(I), \hat{I}_m), \\ S_m^{2p}(I) &= d^2(I_m^c \cdot \text{PR}_m(I), I_m^c \cdot \hat{I}_m), \\ S_m^{cp}(I) &= d^c(I_m^c \cdot \text{PR}_m(I), I_m^c \cdot \hat{I}_m). \end{aligned} \quad (3.12)$$

Dans les expressions de $S_m^{2p}(I)$ et de $S_m^{cp}(I)$, l'opérateur \cdot désigne le produit pixel à pixel des deux images.

3.5.4 Classification et décision

Pour obtenir une déclinaison particulière de la méthode proposée, on sélectionne en premier lieu l'un des quatre indices de dissimilarité définis en (3.12). Ce dernier permet de comparer l'image cible à chacune des M signatures stockées dans la base de données résultant de l'apprentissage.

Une fois que tous les indices de dissimilarité entre l'image cible et les M individus de la base ont été calculés, la décision amenant à une reconnaissance de l'individu est obtenue en sélectionnant la dissimilarité minimale, comme nous l'avons décrit dans la section 1.6.1 du premier chapitre. Cela correspond à une technique de classification par la méthode du plus proche voisin, à ceci près qu'on ne dispose que des centres des classes (signatures \mathbf{s}_m des individus) dans l'espace des paramètres.

On peut introduire un seuil permettant de considérer l'image cible comme appartenant à un individu inconnu si tous les indices de dissimilarité ont des valeurs trop élevées.

3.6 Conclusion

Nous avons détaillé dans ce chapitre les aspects théorique et pratique des nouvelles méthodes que nous avons développées dans le cadre de cette thèse. Ces méthodes améliorent le concept initial d'utilisation de réseaux d'ondelettes de Gabor pour la reconnaissance de visages.

La limitation de la méthode proposée par Krüger réside dans le fait que l'extraction de signatures est réalisée sur une seule image. Nous avons donc proposé une nouvelle méthode d'apprentissage permettant de remédier à cette limitation en utilisant plusieurs images du même individu pour ajuster les paramètres du ROG. D'autre part, nous avons proposé d'inclure dans le modèle un ensemble de coefficients définissant la confiance qu'on peut apporter à chaque ondelette du ROG.

Les traitements nécessaires à l'implantation de cette méthode, tant dans la phase hors-ligne d'enrôlement que dans la phase en-ligne d'identification, ont également été détaillés afin que notre approche soit facilement reproductible.

Dans le chapitre suivant, nous présentons une évaluation expérimentale des performances de la méthode proposée.

Chapitre 4

Résultats, évaluations et discussions

4.1 Introduction

Pour vérifier l'efficacité des méthodes proposées dans le chapitre 3, nous avons opté pour une validation sur une base de données standard, à savoir YaleFaces. En effet, comme toute problématique de reconnaissance automatique, les techniques doivent être validées sur un ensemble consistant de données. Toutes les images de la base sont étiquetées, ce qui permet de déterminer *a posteriori* les performances des méthodes, par comparaison du résultat de classification et de la vérité terrain.

Nous présentons dans ce chapitre les résultats obtenus par notre méthode en la comparant avec d'autres techniques, notamment celle proposée par Krüger. Rappelons que la méthode de Krüger utilise également des réseaux d'ondelettes de Gabor pour extraire les signatures (paramètres du ROG) à partir d'une image unique, puis reconnaît ensuite l'image de test en la projetant sur les ROG de tous les individus. Notre approche généralise la phase d'apprentissage afin de tenir compte de plusieurs expressions faciales pour déterminer la signature d'un individu.

Au début du chapitre, nous présentons la base de visages sur laquelle nous avons réalisé les comparaisons, c'est à dire YaleFaces. Nous décrivons le contenu et les caractéristiques de cette base. Ensuite, nous présentons les différentes expérimentations que nous avons réalisées selon plusieurs protocoles, chacun correspondant à un paramétrage donné des méthodes et utilisant un sous-ensemble particulier de la base YaleFaces.

4.2 Base de données YaleFaces

Nous avons présenté brièvement dans le chapitre 1 quelques bases de données de visages qui sont souvent référencées dans la littérature scientifique. Le tableau ci-dessous présente plus précisément les caractéristiques des principales bases de données de visages, notamment en terme de diversité des paramètres intrinsèques et extrinsèques de modification des visages.

Base	Nb d'individus	Pose	Occultation	Illumination	Expression
FERET	1199	1-9	-	2	2
AT&T	40	1-2	1-2	-	-
AR	120	1	2	4	4
PF01	107	8	-	4	5
YaleFaces	15	1	1	3	6

TABLE 4.1 : Caractéristiques des principales bases de données de visages

Comme notre méthode utilise principalement la variabilité de l'image induite par les différentes expressions faciales, YaleFaces¹ est la base la plus adaptée du fait qu'elle contient six expressions différentes pour chaque individu. D'autre part, elle présente l'avantage de contenir à la fois des images des individus pour ces différentes expressions faciales (variations intrinsèques), ainsi que pour différentes conditions d'acquisition (facteurs extrinsèques, comme le changement d'illumination ou le port de lunettes). Enfin, cette base a été largement utilisée dans la littérature, ce qui autorise la comparaison des performances avec un nombre significatif de méthodes.

YaleFaces a été constituée par le centre CVC² (Center for Computational Vision and Control) du département informatique de l'université de Yale³. Elle contient 165 images en niveaux de gris de 15 individus. Chaque individu est représenté par 11 images faciales prises dans des conditions différentes. On y trouve des images standard du visage (expression neutre, lumière centrée et sans lunettes), des images du visage dont l'aspect a été modifié par des facteurs intrinsèques (expressions faciales : souriant, triste, yeux

1. <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>

2. <http://cvc.yale.edu/>

3. <http://www.yale.edu/>

fermés, surpris et clin d'oeil) et des images du visage dont l'aspect a été modifié par des facteurs extrinsèques (lunettes, éclairage à gauche, éclairage à droite).

Nous présentons dans la figure 4.1 les 11 images du premier individu de la base YaleFaces, chaque image étant accompagnée par la description soit de l'expression faciale, soit de la condition particulière d'acquisition qui la caractérise.

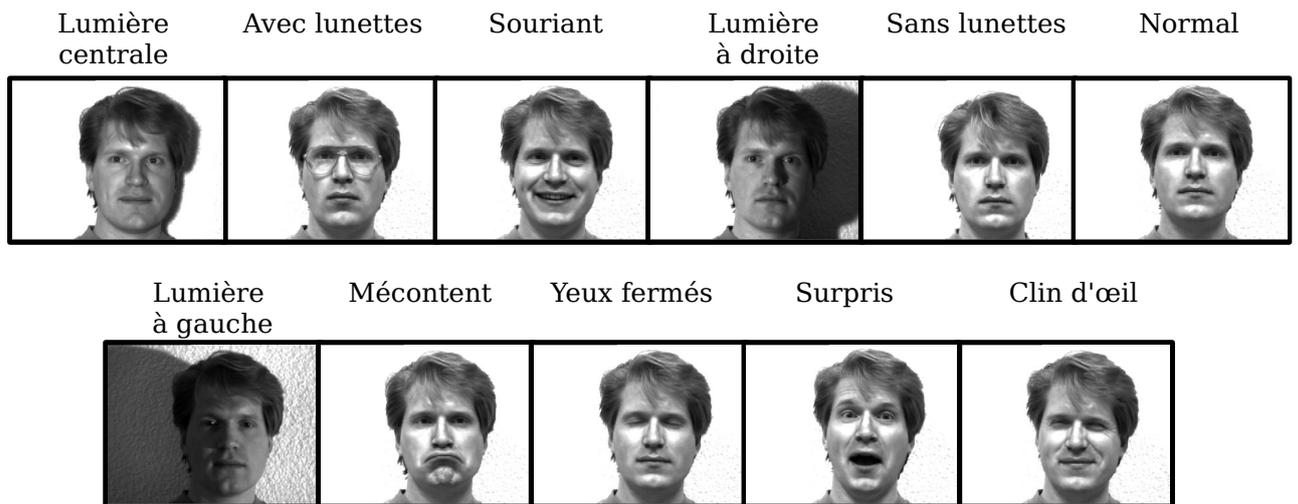


FIGURE 4.1 : Images de l'individu 1 de la base YaleFaces

Ces différentes configurations sont disponibles pour les 15 individus, dont les images correspondant à l'expression neutre sont présentées dans la figure 4.2.

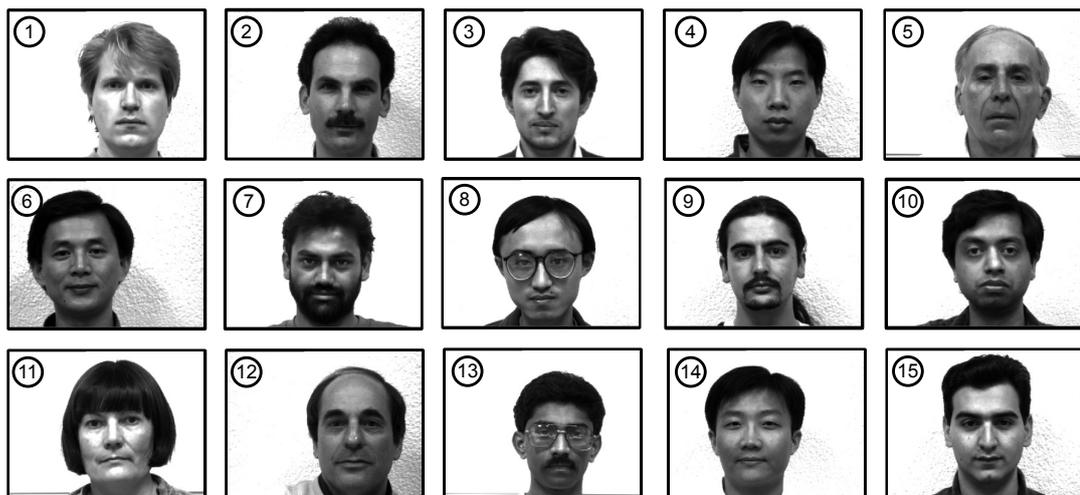


FIGURE 4.2 : Image 1 de chaque individu de la base YaleFaces

Selon les notations utilisées dans ce manuscrit, une base de données β est composée

de N images de M individus : $\beta = \{I_{1,1}, I_{1,2}, \dots, I_{m,n}, \dots, I_{M,N}\}$. Pour YaleFaces, N et M valent respectivement 11 et 15. Pour tester les méthodes, nous avons choisi de partitionner la base en deux sous-ensembles, l'un pour l'apprentissage, l'autre pour le test. Cette approche nous semble plus rigoureuse, quand on dispose d'assez de données, que celles qui consistent à apprendre sur toutes les données sauf une, puis de tester sur cette donnée laissée de côté (approche *leave one out*).

Avant de procéder à l'apprentissage (mode enrôlement), nous sélectionnons pour chaque individu un ensemble de N_a images afin de constituer la base d'apprentissage β_a . N_t images (avec $N_t \leq N - N_a$) de chaque individu constituent la base de test β_t . Cette base est employée pour tester la performance du système en termes de reconnaissance (mode identification). À l'issue de l'étape d'enrôlement, les M individus de la base sont caractérisés par M signatures, également regroupées dans une base $S = \{s_1, s_2, \dots, s_M\}$. Le contenu exact d'une signature change selon la méthode utilisée, par exemple dans notre approche elle est composée des paramètres du ROG et des coefficients de confiance associés aux ondelettes. Dans le mode identification, le système reçoit une image de visage, y applique un pré-traitement puis la classifie. Nous présentons dans la figure 4.3 un schéma synoptique illustrant le fonctionnement de ce mode de validation.

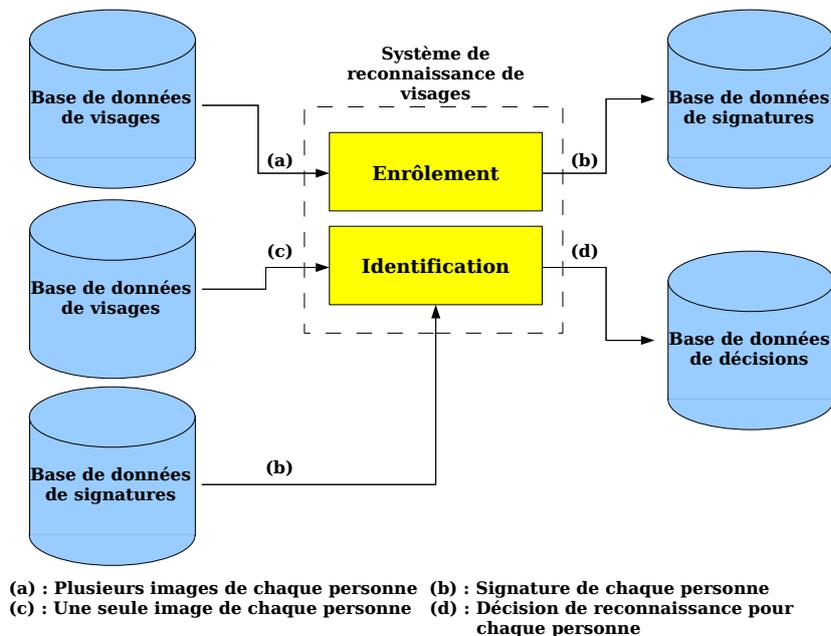


FIGURE 4.3 : Synoptique du système de validation de notre méthode

4.3 Protocole expérimental 1

Le premier protocole expérimental reproduit aussi fidèlement que possible celui qui a été utilisé par Krüger dans sa thèse et dans plusieurs des articles qu’il a publiés. Dans ce protocole, nous utiliserons $L = 52$ ondelettes dans le ROG, une seule image ($N_a = 1$) pour l’apprentissage des paramètres du ROG, et $N_t = 5$ images pour calculer le taux de reconnaissance.

La figure 4.4 présente un ensemble d’images de la base YaleFaces correspondant à un unique individu. Elles indiquent les expressions faciales sélectionnées par Krüger pour évaluer sa méthode. L’image correspondant à l’expression neutre est utilisée en enrôlement pour extraire les signatures, et les cinq autres expressions faciales sont utilisées dans le mode identification pour évaluer les taux de reconnaissance.



(a) Normale (b) Souriant (c) Mécontent (d) Yeux fermés (e) Surpris (f) Clin d’oeil

FIGURE 4.4 : Images utilisées dans le protocole expérimental 1

Dans un souci de reproduction exacte du protocole expérimental de Krüger, nous n’avons pas utilisé de filtrage ni sur les images de test ni sur l’image d’apprentissage. Nous présentons dans le tableau 4.2 les taux de reconnaissance obtenus en fondant la décision soit sur une distance quadratique d^2 , soit sur une distance de corrélation d^c . Dans les deux cas, l’image cible est projetée sur la base d’ondelettes de chacun des ROG, puis reconstruite avant de calculer la distance.

Si on cumule tous les taux, comme l’a fait Krüger, on aboutit à un taux global de 86.66%, soit un résultat qui est différent de ce que cet auteur annonçait dans sa thèse [Kru00]. Nous avons constaté que pour aboutir au taux de 96% que Krüger affiche dans ses différentes publications, il fallait retirer l’individu 13 de la base YaleFaces. Dans ce deuxième cas, on obtient les taux de reconnaissance présentés dans le tableau 4.3.

Image de test	Taux de reconnaissance avec d^c	Taux de reconnaissance avec d^2
Souriant	93.33%	80.00%
Mécontent	80.00%	93.33%
Yeux fermés	93.33%	93.33%
Surpris	86.66%	73.33%
Clin d'oeil	80.00%	73.33%
Taux global	86.66%	82.66%

TABLE 4.2 : Protocole 1 : taux de reconnaissance selon l'image de test (individu 13 inclus)

Image de test	Taux de reconnaissance avec d^c	Taux de reconnaissance avec d^2
Souriant	100.00%	93.33%
Mécontent	93.33%	93.33%
Yeux fermés	93.33%	100.00%
Surpris	93.33%	93.33%
Clin d'oeil	100.00%	93.33%
Taux global	96.66%	94.66%

TABLE 4.3 : Protocole 1 : taux de reconnaissance selon l'image de test (individu 13 exclu)

Discussion

Après une étude approfondie des résultats, nous avons donc constaté que l'individu 13 n'est pas bien reconnu et qu'il entraîne une diminution notable du taux de reconnaissance. C'est probablement la raison pour laquelle Krüger a exclu cet individu de la base avant de mener ses expérimentations. Nous avons vérifié cette hypothèse *a posteriori*, en regardant en détail les figures présentes dans les publications de Krüger. Par exemple, la figure 4.3, extraite de [Kru00], présente l'histogrammes des distances de corrélation calculées entre l'image reconstruite à partir de la signature de l'individu 1 (subject01.expression dans la figure) et les images (pour l'expression faciale neutre) des autres individus (subject*i*.normal dans la figure). Sur cet histogramme, on constate que Krüger n'a pas fourni de résultat pour l'individu 13.

Sur les images du visage de l'individu 13, on remarque que contrairement aux autres

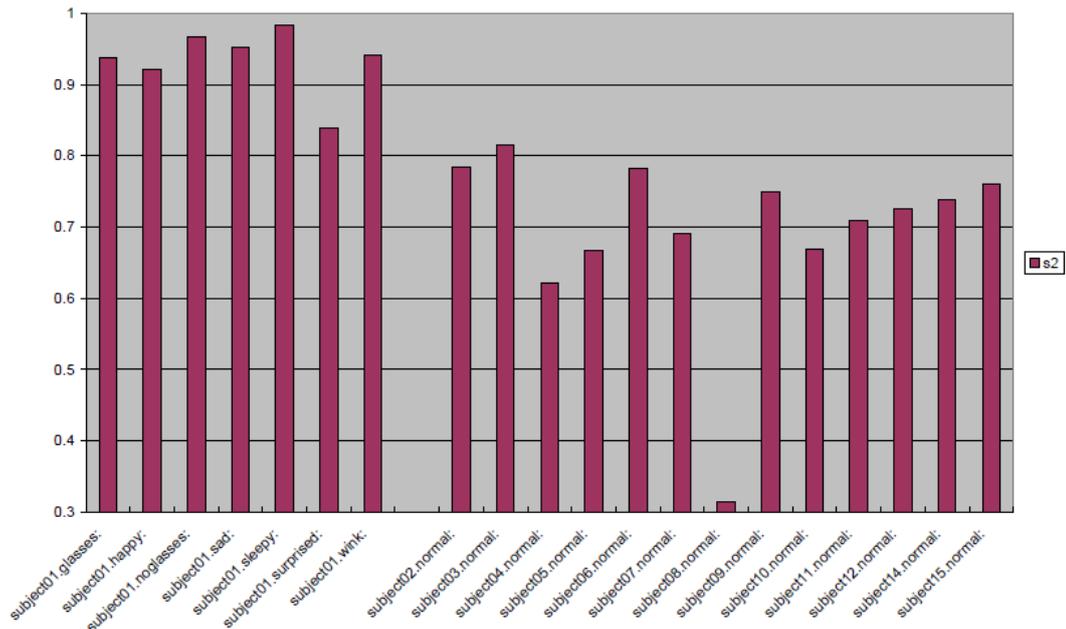


FIGURE 4.5 : Distances de corrélation entre différentes images (extraite de [Kru00])

personnes de la base, il porte des lunettes (cf. figure 4.6). Ces lunettes sont présentes pour toutes les expressions faciales et toutes les conditions d’acquisition, à l’exception d’une seule qui correspond en fait à celle dans laquelle les autres individus portent des lunettes. D’autre part, le visage de l’individu 13 est légèrement plus petit que celui des autres personnes, probablement parce que le facteur de zoom était légèrement différent lors de l’acquisition de la série d’images.

Un taux de reconnaissance légèrement plus faible est également obtenu pour l’individu numéro 8, qui porte également des lunettes. En revanche, pour cet individu la taille du visage dans l’image est la même que pour les autres personnes.



(a) Individu 8 (b) Individu 13

FIGURE 4.6 : Présentation des images des individus 8 et 13

Ces constatations semblent indiquer d’une part que la présence de facteurs de modification intrinsèque comme les lunettes perturbent le processus de reconnaissance, et d’autre part que la méthode de Krüger est très sensible à la variation des dimensions du

visage. Bien sur, cette dernière hypothèse est formulée sur un nombre très limité d'images, mais notre objectif n'est pas ici de revalider en totalité l'approche de Krüger.

Les taux de reconnaissance sont très similaires pour les deux distances, avec une légère supériorité de la distance de corrélation. Cela correspond à ce que Krüger indique dans ses publications. De ce fait, pour les comparaisons présentées dans la suite de ce chapitre, nous n'avons indiqué que les résultats obtenus avec la distance de corrélation, bien que toutes nos simulations aient été réalisées avec les deux distances.

4.4 Protocole expérimental 2

Ce protocole est une version plus générale du précédent, reposant toujours sur la méthode de Krüger. Notre objectif est ici de déterminer les facteurs de modification (intrinsèque ou extrinsèque) qui mettent cette méthode en défaut d'une part, et de prouver l'intérêt de l'étape de filtrage des images avant leur modélisation par un ROG.

Les paramètres du ROG sont toujours ajustés en utilisant uniquement l'image qui correspond à une expression faciale neutre. Par contre, nous utilisons les $N_t = 10$ autres images disponibles pour chaque individu afin d'évaluer les taux de reconnaissance. Tous les individus de la base sont utilisés pour l'évaluation, y compris ceux qui posent problème (individus 8 et 13). Nous présentons les résultats obtenus, d'une part en utilisant les images brutes disponibles dans la base, d'autre part pour des images recadrées et filtrées par pondération avec une fenêtre de Hamming (cf. tableau 4.4).

Discussion

Le taux de reconnaissance global pour la technique de Krüger sans filtrage préalable des images, indiqué dans la dernière ligne du tableau 4.4, atteint 80%. Quand on réalise un recadrage et un filtrage des images avant d'appliquer la méthode de Krüger, le taux global de reconnaissance atteint 82.66%. Cela indique l'intérêt de réaliser un pré-traitement efficace des images, notamment pour éliminer les effets de bord par pondération avec une fonction de fenêtrage.

L'intérêt du pré-traitement apparaît de façon plus marquée sur certaines images, notamment pour celles acquises avec une modification de la source d'éclairage. Par exemple,

Image de test	Résultats sans filtrage	Résultats avec filtrage
Lumière centrale	73.33%	73.33%
Avec lunettes	80.00%	80.00%
Souriant	93.33%	93.33%
Lumière à droite	66.66%	66.66%
Sans lunettes	100.00%	100.00%
Lumière à gauche	46.66%	60.00%
Mécontent	80.00%	80.00%
Yeux fermés	93.33%	93.33%
Surpris	86.66%	93.33%
Clin d'oeil	80.00%	86.66%
Taux global	80.00%	82.66%

TABLE 4.4 : Protocole 2 : taux de reconnaissance selon l'image de test

pour la situation d'éclairage à gauche, le taux passe de 46.66% à 60.00% en moyenne sur tous les individus quand on introduit l'étape de filtrage. Cela s'explique par le fait que les effets de bord sont prépondérants sur ce type d'image, du fait que les niveaux de gris sur les côtés des images sont très variables.

Cependant, sur les images pour lesquelles la source lumineuse a changé de position, le taux de reconnaissance global reste relativement faible en dépit de l'amélioration apportée par le filtrage. Nous supposons que cela est causé par la chute de performance de l'étape de projection/reconstruction quand l'éclairage n'est pas uniforme. En effet, les poids calculés par le produit scalaire entre l'image et les ondelettes duales sont très perturbés par cette variation significative du niveau de gris dans des zones de surface importante.

4.5 Protocole expérimental 3

L'objectif de ce protocole expérimental est de comparer trois techniques de reconnaissance basées sur des réseaux d'ondelettes de Gabor :

- la technique de Krüger exploitant une seule image pour l'apprentissage. La signature d'un individu est dans ce cas un unique ROG ;

- la technique de Krüger, mais avec plusieurs ROG, chacun modélisant une expression faciale différente. Cela correspond en fait à considérer qu'un même individu est représenté par plusieurs signatures différentes, donc par plusieurs ROG ;
- la technique que nous avons décrite dans le chapitre précédent, qui utilise directement plusieurs images afin d'extraire une signature unique pour chaque individu.

Les réseaux utilisés dans les trois méthodes intègrent tous $L = 52$ ondelettes, comme dans la méthode initiale de Krüger. En revanche, nous intégrons systématiquement l'étape de pré-traitement par recadrage et filtrage dans les trois cas, afin de considérer exactement les mêmes images lors de la comparaison des performances.

Nous utilisons une base d'apprentissage constituée de $N_a = 4$ images de chaque individu et une base de test constituée des $N_t = 7$ images restantes (également pour chaque individu). La figure 4.7 présente les images qui ont été sélectionnées pour constituer la base d'apprentissage, lesquelles correspondent aux expressions neutre, souriant, mécontent et surpris. La figure 4.8 présente les images intégrées dans la base de test, à savoir : lumière centrale, avec lunettes, lumière à droite, sans lunettes, lumière à gauche, yeux fermés et clin d'oeil.



FIGURE 4.7 : Images utilisées en mode enrôlement dans le protocole 3

Pour la méthode que nous proposons, les méta-paramètres sont fixés aux valeurs suivantes : le nombre d'itérations Q_i séparant deux modifications de l'image de référence durant l'apprentissage est fixé à 10 ; les masques de confiances associés aux images sont obtenus avec un coefficient τ d'étalement de la sigmoïde fixé à la valeur 200 ; enfin, le nombre total d'itérations d'apprentissage est fixé à 500 (comme pour les deux autres méthodes), ce qui correspond à $Q_t = 50$ tirages au sort d'une image de référence pour l'apprentissage.

Le tableau 4.5 présente les résultats obtenus sur les différentes expressions faciales par les trois méthodes basées sur un ROG.

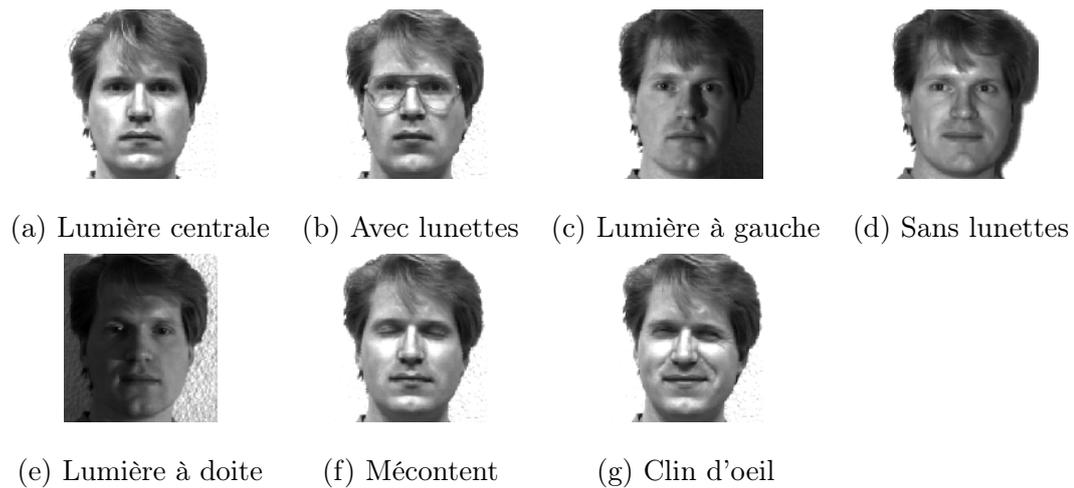


FIGURE 4.8 : Images utilisées en mode identification dans le protocole 3

Image de test	Méthode de Krüger	Méthode de Krüger sur 4 images	Notre méthode
Lumière centrale	73.33%	73.33%	100.00%
Avec lunettes	80.00%	80.00%	93.33%
Lumière à droite	66.66%	60.00%	86.66%
Sans lunettes	93.33%	93.33%	100.00%
Lumière à gauche	60.00%	60.00%	86.66%
Yeux fermés	93.33%	93.33%	93.33%
Clin d'oeil	86.66%	86.66%	93.33%
Taux global	79.04%	78.09%	93.33%

TABLE 4.5 : Protocole 3 : taux de reconnaissance selon l'image de test pour les trois méthodes

Discussion

En analysant les résultats présentés dans le tableau 4.5, nous constatons que les taux de reconnaissance caractérisant les deux variantes de la méthode de Krüger sont très similaires. Contrairement à ce qu'on pourrait attendre, il y a une légère dégradation avec la méthode utilisant 4 signatures, causée par la non reconnaissance d'une image avec un éclairage à droite. On peut en déduire que la simple augmentation globale du nombre de signatures n'améliore pas les performances du système, tant que l'on ne tient pas compte explicitement de la variabilité de la signature en fonction de l'expression faciale.

En revanche, les taux obtenus par la méthode que nous proposons sont bien supérieurs. Quelle que soit l'expression faciale ou les conditions d'éclairage, nous obtenons un meilleur taux de classification qu'avec les deux déclinaisons de la méthode de Krüger. Cela montre l'intérêt d'utiliser plusieurs expressions faciales durant l'apprentissage. D'autre part, l'amélioration du taux de reconnaissance est dûe principalement à l'utilisation du masque des confiances dans le calcul des distances.

L'intérêt d'intégrer un masque des confiances dans le traitement apparaît clairement sur les images acquises avec des conditions d'éclairage variables. La première ligne de la figure 4.9 présente les deux images de la base, pour l'individu 1, dans lesquelles la source de lumière est placée soit à gauche (a), soit à droite (b) durant l'acquisition. Sur la deuxième ligne, nous avons représenté deux fois le masque des confiances associé à la signature de cet individu. Enfin, la dernière ligne présente les deux images de cet individu après pondération par le masque. Sur ces images pondérées, on voit clairement que le masquage élimine une majeure partie des zones de l'image dans lesquelles le changement d'éclairage a une influence. Naturellement, le changement d'éclairage modifie partiellement les images pondérées, mais son influence est moindre sur le taux de reconnaissance.

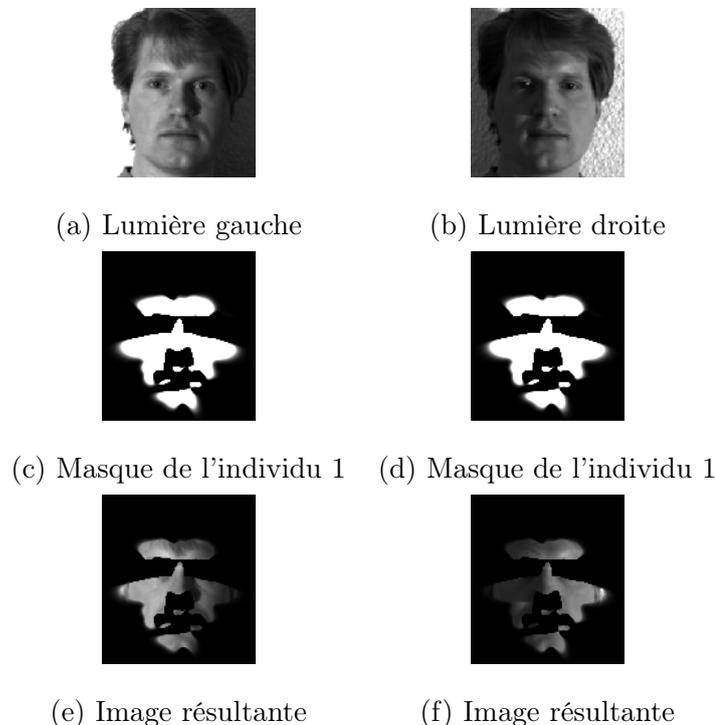


FIGURE 4.9 : Pondération des images par le masque des confiances

Outre cet aspect de minimisation de l'effet d'une variation de l'éclairage, le masque

délimite également les zones du visage significatives pour la reconnaissance. Le masque des confiances permet en fait de ne conserver dans une image cible que les zones qui sont significatives dans le modèle de l'individu auquel on la compare. La figure 4.10 permet de vérifier visuellement cette propriété de l'opération de pondération par un masque. La première colonne présente les images, pour l'expression faciale neutre, des individus 1 (a) et 8 (d) de la base YaleFaces. La deuxième colonne présente les masques des confiances déterminés pour ces deux individus durant la phase d'apprentissage, mais ces derniers sont placés sur la ligne correspondant à l'image de l'autre individu. La dernière colonne présente les images après pondération, c'est à dire telles qu'elles seront utilisées pour le calcul de distance. Ainsi, l'image (c) est utilisée pour calculer la distance entre l'image de l'individu 1 (a) et le modèle de l'individu 8. Sur les images de la dernière colonne, on vérifie aisément que le masque ne sélectionne pas les zones pertinentes des images de la première colonne quand ce dernier ne correspond pas au même individu.

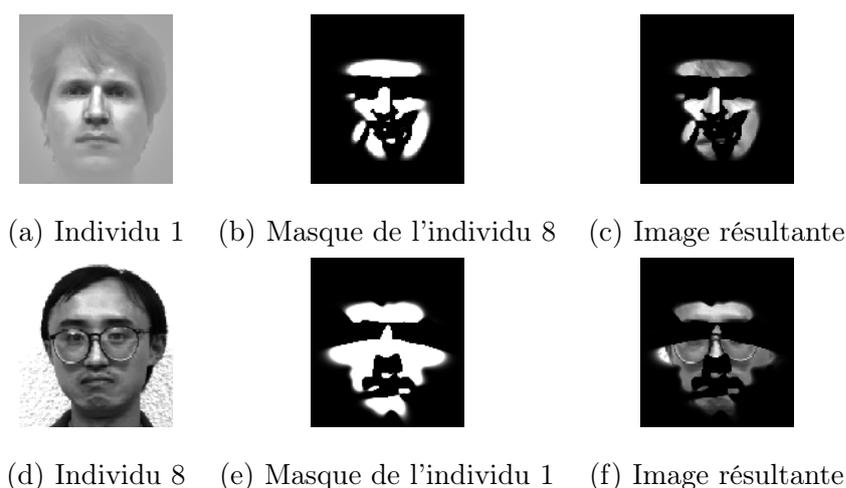


FIGURE 4.10 : Application du masque sur les images

4.6 Performances en terme de temps de calcul

Pour évaluer les performances en termes applicatifs de notre méthode, nous avons déterminé les temps de calcul nécessaires durant les phases d'enrôlement de d'identification. A titre de comparaison, nous avons également déterminé les temps de calcul pour les deux déclinaisons de la méthode de Krüger. Les simulations ont été réalisées sur un ordinateur équipé d'un processeur Intel® Quad cadencé à 2.4MHz. Le code n'a pas été optimisé et

les images sont chargées à partir du disque dur (pas de stockage initial en mémoire de toutes les images). Toutes les méthodes utilisent des ROG intégrant $L = 52$ ondelettes et traitent des images de taille 119×132 .

Le tableau 4.6 indique les temps de calcul qui sont nécessaires pour le calcul de la signature (méthodes de Krüger et notre méthode) ou des signatures (méthode de Krüger avec 4 signatures) pour un individu de la base. On constate que notre méthode nécessite environ 60% de temps de calcul supplémentaire en comparaison avec la méthode de Krüger utilisant 4 signatures. La phase d'apprentissage de la méthode de Krüger initiale est quant à elle nettement plus rapide. Il faut rappeler que l'apprentissage est réalisé hors-ligne, donc que le temps de calcul nécessaire dans cette phase est rarement critique dans les applications de reconnaissance de visages.

Méthode	Temps d'exécution
Krüger	3 min 15 sec
Krüger (4 images)	13 min
Notre méthode	20 min et 31 sec

TABLE 4.6 : Temps nécessaire au calcul d'une signature (ou des signatures) pour un individu

Le tableau 4.7 présente les temps de calcul correspondant à la phase de reconnaissance d'une image de visage. On constate, qu'en toute logique, il faut 4 fois plus de temps avec la méthode de Krüger qui exploite 4 signatures qu'avec la méthode originale de cet auteur. En revanche, avec notre méthode, le temps nécessaire à la reconnaissance d'un visage est le même qu'avec la méthode de Krüger. En fait, les calculs nécessaires à l'évaluation d'une distance dans les deux méthodes ne diffèrent qu'au niveau de la multiplication des images par le masque des confiances, opération élémentaire peu coûteuse en temps de calcul.

Méthode	Temps d'exécution
Krüger	13 sec
Krüger (4 images)	51 sec
Notre méthode	13 sec

TABLE 4.7 : Temps nécessaire à la reconnaissance d'un individu de la base parmi les 15

4.7 Comparaison avec d'autres méthodes de reconnaissance

Dans les sections précédentes, la méthode que nous proposons a été comparée avec celle dont elle dérive, à savoir la méthode de Krüger. Dans cette partie, nous comparons maintenant notre méthode avec la plupart de celles décrites dans l'état de l'art ayant été validées sur la base YaleFaces. Cette comparaison nous permettra de bien positionner notre approche, notamment au niveau de l'efficacité de reconnaissance.

Nous retrouvons ici les principales références décrites dans le chapitre 2. Nous rappelons brièvement le principe de chaque technique et réalisons une comparaison en nous plaçant autant que possible dans les mêmes conditions de fonctionnement dans les phases d'enrôlement et d'identification. Nous présentons cette section selon un ordre chronologique, les références les plus récentes étant présentées en premier.

La CSFR (Compressed Sensing for Face Recognition) est une technique qui a été récemment proposée par Vo *et coll.* [VVCM09]. Elle a été validée sur la base YaleFaces. Les auteurs ont choisi de faire varier le nombre d'images consacrées à l'apprentissage, puis d'évaluer le taux de reconnaissance global obtenu sur les images restantes. Dans le cas où 4 images par individu ont été utilisées pour l'apprentissage, cette méthode atteint un taux de reconnaissance global égal à 83.54%. Ce taux est inférieur à celui obtenu par notre méthode, laquelle atteint 93.33% sur le même jeu de données.

Les réseaux de neurones de type RBF ont été également utilisés dans l'application de reconnaissance de visages. Chang *et coll.* utilisent un classifieur RBF particulier dont les centres des fonctions radiales sont fixés par une technique d'analyse en composantes principales [CrH08]. Ce classifieur atteint un taux de reconnaissance de 97.33% sur la base YaleFaces. Les auteurs ont mentionné qu'ils utilisent 5 images pour le test mais ils n'ont pas précisé lesquelles. Pour ne pas biaiser la comparaison (en sélectionnant les images qui nous favoriseraient), nous avons déterminé le taux de reconnaissance obtenu par notre méthode en tirant 5 images au hasard de chaque individu afin de constituer la base de test. Dans ces conditions, nous atteignons un taux de reconnaissance de 94.66%, légèrement inférieur à celui de la méthode de Chang *et coll.* .

Nous avons utilisé le même principe pour comparer notre méthode à celle de Mazloom

et Ayat [MA08]. Ces auteurs ont décrit une méthode de reconnaissance de visages basée sur l'association de la transformée en ondelettes (TO), l'analyse en composantes principales (ACP) et les réseaux de neurones artificiels (RNA). Cette combinaison donne un taux de reconnaissance variable selon le nombre des composantes principales sélectionnées. Dans le cas de l'utilisation de 15 composantes principales, le système atteint un taux de reconnaissance de 86.56%. Ce taux passe à 90.35% si le système utilise 25 composantes principales.

Avec 10 images utilisées pour le test, le taux de reconnaissance obtenu avec notre méthode est de 96.36%. Ce taux est légèrement inférieur à celui obtenu avec la méthode combinant HOG et EBGM [AMM⁺08]. Cette dernière utilise les HOG (Histogrames of Oriented Gradients) plutôt que les ondelettes de Gabor dans l'EBGM. Elle atteint un taux de reconnaissance variable selon la dimension des HOG : 98.2% et 97.6% correspondent respectivement à des HOG de dimensions 12×12 et 16×16 . Si nous testons notre système sur 10 images (dans ce cas 4 images sont communes aux deux bases d'apprentissage et de test), nous atteignons un taux de 95.14%.

Zhang *et coll.* utilisent 5 images pour l'apprentissage et 6 images pour le test [ZGD07] afin de valider et comparer 4 méthodes statistiques de reconnaissance de visages : l'ACI par blocs, l'ACP classique, l'ACP2D et l'ACI classique. Ces quatre méthodes atteignent respectivement des taux de reconnaissance de 75.66%, 67.78%, 72.22% et 65.56%. Avec les mêmes bases d'apprentissage et de test, notre méthode atteint un taux de reconnaissance de 94.44%.

Les réseaux d'ondelettes bêta ont également été utilisés pour la reconnaissance de visage [ZAA05]. La technique de validation proposée par Zaied *et coll.* est similaire à celle que nous avons présenté dans le protocole 2 — une image pour l'extraction de la signature et 10 images pour le test —. Dans ce protocole, notre méthode a atteint un taux de reconnaissance de 82.66%, contre 77% pour les réseaux d'ondelettes bêta.

Les résultats de comparaison décrits précédemment sont récapitulés dans le tableau 4.8. On constate que notre méthode se positionne correctement par rapport aux autres techniques.

Méthode	Taux de reconnaissance	Notre méthode
CSFR [VVCM09]	83.54%	93.33%
ACP+RBF [CrH08]	97.33%	94.66%
TO+ACP+RNA [MA08]	90.35%	94.66%
HOG+EBGM [AMM ⁺ 08]	98.2%	95.14%
ACI en block [ZGD07]	75.66%	94.44%
Réseau d'ondelettes bêta [ZAA05]	77.00%	82.66%

TABLE 4.8 : Comparaison des taux de reconnaissance obtenus par notre méthode et d'autres méthodes de la littérature dans le cadre de protocoles expérimentaux similaires

4.8 Conclusion

Nous avons étudié dans ce chapitre les performances obtenues par notre méthode sur la base de données YaleFaces en les comparant à celles obtenues par d'autres méthodes. Un premier protocole nous a permis de spécifier le cadre expérimental utilisé par Krüger dans ses différentes publications. Un deuxième protocole a servi à vérifier l'intérêt des pré-traitements réalisés sur les images avant leur modélisation. Enfin, dans un troisième protocole expérimental, nous avons comparé notre méthode à deux déclinaisons de celle de Krüger. Dans tous les cas, les taux de reconnaissance que nous obtenons sont supérieurs à ceux relatés par Krüger dans ses articles.

Nous avons également comparé les temps de calcul nécessaires pour extraire les signatures et pour reconnaître un visage. Notre approche nécessite un temps de calcul bien supérieur à celui de la méthode de Krüger pour l'apprentissage des ROG. Malgré tout, cela n'est pas trop pénalisant car ce traitement est réalisé en hors-ligne. En revanche, les temps de calcul pour la reconnaissance d'un visage durant la phase d'identification sont équivalents, bien que notre méthode utilise une signature plus discriminante.

Enfin, nous avons comparé notre méthode à un ensemble d'autres méthodes qui avaient également été validées par leurs auteurs sur la base YaleFaces. Pour chaque comparaison, nous avons reproduit le protocole expérimental décrit par les auteurs, notamment en ce qui concerne les bases d'apprentissage et de test. Dans la plupart des cas (sauf pour l'EBGM), notre méthode s'avère supérieure en terme de taux de reconnaissance des visages.

Conclusion générale et perspectives

Synthèse

Les travaux menés durant la préparation de cette thèse concernent le développement d'une nouvelle technique de reconnaissance de visages basée sur les réseaux d'ondelettes de Gabor. La reconnaissance de visages à partir d'une image est classée parmi les principales techniques de biométrie actuellement explorées. La figure I.4, présentée dans l'introduction, montre clairement l'intérêt qu'elle suscite auprès de la communauté scientifique.

Nous avons organisé ce manuscrit en quatre chapitres. Le premier a présenté une synthèse bibliographique se rapportant aux techniques de reconnaissance de visages. D'abord, la capacité du système visuel humain à reconnaître les visages a été mise en évidence. Ensuite nous avons présenté les systèmes automatiques selon une architecture modulaire, à savoir : un module de détection et pré-traitement, un module d'extraction de signatures et un module de classification et décision. Les aspects théoriques et pratiques des principales techniques ont été explicités dans les descriptions de chacun des modules. Les meilleurs résultats relatés par la communauté scientifique ont également été présentés.

Le chapitre 2 a décrit les méthodes paramétriques utilisées dans le module d'extraction de signatures. Ce module a une importance cruciale dans le processus de reconnaissance de visages. Une introduction des méthodes de paramétrisation a été présentée dans ce chapitre. Nous avons passé en revue les techniques les plus utilisées pour l'extraction de signatures. Une présentation détaillée a ensuite été proposée pour chaque méthode, dans laquelle les aspects théorique et pratique ont été mis en évidence. L'impact de ces techniques sur les performances globales du processus de reconnaissance a également été examiné. Une grande partie de ces travaux repose sur l'approximation de la fonction image, notamment au moyen de réseaux de neurones et de réseaux d'ondelettes.

Dans le chapitre 3, une nouvelle technique de reconnaissance de visages basée sur les réseaux d'ondelettes de Gabor a été présentée. Les éléments originaux ont été détaillés aussi bien lors de la phase d'enrôlement que lors de la phase d'identification. La première phase introduit l'utilisation de plusieurs images de différentes expressions faciales pour l'apprentissage du réseau et la création d'un masque des confiances. Ce masque indique quelles sont les zones de l'image qui sont significatives en terme de reconnaissance. La deuxième phase exploite ce masque lors de la reconnaissance en-ligne d'un visage.

Des résultats expérimentaux évalués sur la base de données YaleFaces ont été décrits dans le chapitre 4. L'évaluation des nouvelles techniques proposées a été réalisée selon différents protocoles expérimentaux. Ceci a permis de discuter l'utilisation des réseaux d'ondelettes de Gabor pour la reconnaissance faciale, et de prouver l'efficacité de l'approche proposée.

Contributions

Les contributions de cette thèse portent essentiellement sur la représentation optimale du visage par les réseaux d'ondelettes de Gabor. Lors de l'apprentissage, un calcul statistique mené sur l'évolution des paramètres des ondelettes permet d'associer à chacune d'entre elles un coefficient de confiance. Nous avons proposé un nouvel algorithme d'apprentissage de manière à pouvoir changer l'image cible au cours des itérations. Ce changement amène les ondelettes à s'adapter à chaque fois à une nouvelle image pendant un nombre fixé d'itérations. Le modèle résultant, codé dans les paramètres du réseau, représente ainsi toutes les images de l'ensemble d'apprentissage, et non une seule.

Durant l'apprentissage, l'évolution des paramètres est suivie et analysée. Nous avons constaté que les ondelettes placées dans des zones peu sensibles à l'expression faciale (le nez, la joue, etc.) sont stables pendant l'optimisation, alors que les ondelettes placées sur des zones très sensibles à cette expression (la bouche, les yeux, etc.) varient fortement. Forts de cette constatation, nous avons proposé un deuxième algorithme qui vise à associer un coefficient élevé — proche de 1 — aux ondelettes "stables" et un coefficient faible — proche de 0 — aux ondelettes "instables". Ces coefficients de confiance servent à construire un masque indiquant les zones importantes du visage, afin qu'elles contribuent d'avantage

à la précision de la reconnaissance.

Les techniques habituelles utilisées dans la phase d'identification ont été modifiées afin de prendre en compte les masques des confiances calculés durant la phase d'enrôlement. Ces masques sont ainsi intégrés dans les signatures des individus.

Perspectives

De point de vue théorique, cette thèse ouvre plusieurs pistes pour de futurs travaux de recherche. Certaines améliorations de la méthode proposée peuvent être envisagées à court-terme. Leur intégration dans le processus de reconnaissance devrait permettre d'améliorer les performances du système. Nous suggérons particulièrement deux opérations liées au traitement des images durant la phase d'apprentissage réalisée en hors-ligne.

La première consiste à affiner et à rendre plus adaptative l'étape d'initialisation des ROG. Dans notre système, nous avons réalisé cette étape d'une manière très supervisée (en définissant la position initiale des ondelettes sur une grille) et identique pour toutes les images. Or, pour obtenir après apprentissage un ROG qui code parfaitement toutes les particularités d'un visage, il est nécessaire de s'adapter au mieux aux caractéristiques discriminantes de ce dernier. L'initialisation des paramètres du ROG, notamment des translations (t_x, t_y) , devrait donc être adaptée à chaque personne. Elle pourrait par exemple être réalisée en tenant compte explicitement de la taille du visage dans l'image. Pour l'initialisation, plusieurs informations fournies par le module de pré-traitement seraient alors très utiles.

Dans le même ordre d'idée, nous proposons une deuxième amélioration qui permettrait d'automatiser l'étape de construction du masque des confiances. Ce dernier a été construit en réglant le méta-paramètre τ d'étendue de la sigmoïde d'une manière empirique. Nous proposons d'explorer quelques pistes qui pourraient aboutir à un réglage automatique de ce méta-paramètre, par exemple en reportant le réglage à celui du poids total du masque (c'est à dire à la somme de tous les coefficients de confiance associés aux pixels).

D'autres améliorations de notre approche pourraient être envisagées à plus long terme, du fait qu'elles impliquent des validations plus complexes à la fois sur les aspects théorique et pratique.

Nous pensons particulièrement à une amélioration du modèle de ROG utilisé pour modéliser la fonction image. Dans le réseau actuel, chaque ondelette évolue de façon indépendante au cours des itérations de minimisation du résidu. Il peut arriver que certaines ondelettes se rapprochent les unes des autres, tendant de ce fait à modéliser le même élément caractéristique de l'image. Ce comportement non souhaité pourrait être évité en introduisant des contraintes entre ondelettes, c'est à dire en ajoutant un terme de régularisation au résidu. Un terme de régularisation est utilisé notamment dans les approches EBGM, quoique sous une forme assez différente.

La classification qui intervient durant la phase en ligne repose pour l'instant sur la recherche du plus proche voisin dans l'espace des paramètres. Nous pensons qu'il serait possible d'utiliser les calculs statistiques réalisés sur l'évolution des paramètres en cours d'apprentissage afin de définir des hypersurfaces de séparation entre classes plus adaptées (actuellement, la technique du plus proche voisin consiste en fait à définir des hyperplans de séparation des classes définissant une partition de Voronoï de l'espace des paramètres).

Sur le long terme, nous pensons qu'il faudrait explorer plus précisément toutes les possibilités offertes par la modélisation paramétrique de fonctions multidimensionnelles au moyen d'ondelettes de Gabor nD. Au vu des performances particulièrement intéressantes obtenues dans les cas des images numériques, il ne fait aucun doute que cette technique devrait également être très efficace sur des données multidimensionnelles.

Bibliographie

- [AJ05] C. B. Amar et O. Jemai. Wavelet networks approach for image compression. *International Journal on Graphics, Vision and Image Processing*, SI1 :37–45, mai 2005.
- [AMM⁺08] A. Albiol, D. Monzo, A. Martin, J. Sastre, et A. Albiol. Face recognition using HOG-EBGM. *Pattern Recognition Letters*, 29(10) :1537–1543, juillet 2008.
- [ANR74] N. Ahmed, T. Natarajan, et K. Rao. Discrete cosine transform. *IEEE Transactions on Computers*, C-23(1) :90–93, janvier 1974.
- [ANRS07] A. F. Abate, M. Nappi, D. Riccio, et G. Sabatino. 2D and 3D face recognition : A survey. *Pattern Recognition Letters*, 28(1885–1906), juillet 2007.
- [AVE97] H. Abdi, D. Valentin, et B. G. Edelman. Eigenfeatures as intermediate level representations : the case for PCA models. *Brain and Behavioural Sciences*, pages 175–177, septembre 1997.
- [AYS07] P. Abouzar, S. Yousefi, et S. Setarehdan. Hybrid WT based-DCT based face recognition. Dans *IEEE International Conference on Signal Processing and Communications*, pages 963–966, Dubai, United Arab Emirates, novembre 2007.
- [BAA05] W. Bellil, C. B. Amar, et A. M. Alimi. *Beta wavelet networks for function approximation*, chapter I, pages 18–21. décembre 2005.
- [BAA06] W. Bellil, C. B. Amar, et M. Alimi. Comparison between beta wavelets neural networks, RBF neural networks and polynomial approximation for 1D, 2D function approximation. Dans *C. Transactions on Engineering et*

- Technology, editors, *ICSP'06*, volume 13, pages 102–107, Budapest, Hungary, mai 2006.
- [Bar97] R. Barron. *Contribution à l'étude des réseaux d'ondelettes*. Thèse, École Normale Supérieure de Lyon, février 1997.
- [BJF09] E. Barbeau, S. Joubert, et O. Felician. *Traitement et reconnaissance des visages : du percept à la personne*. SOLAL, janvier 2009.
- [BJL03] S. Bres, J.-M. Jolion, et F. Lebourgeois. *Traitement et analyse des images numériques*. 2-7462-0741-9. Paris, Hermes Science Pub, Lavoisier, septembre 2003.
- [BMS02] M. Bartlett, J. Movellan, et T. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13(6) :1450 – 1464, novembre 2002.
- [Bor04] H. L. Borgne. *Analyse de Scènes Naturelles par Composantes Indépendantes*. Thèse, Institut National Polytechnique de Grenoble, janvier 2004.
- [BP93] R. Brunelli et T. Poggio. Face recognition : features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10) :1042–1052, octobre 1993.
- [BPT92] R. Brunelli, T. Poggio, et I. P. Trento. Face recognition through geometrical features. Dans *European Conference on Computer Vision*, pages 792–800, Santa Margherita Ligure, Italy, mai 1992.
- [Car02] J.-F. Cardoso. Analyse en composantes indépendantes. Dans *Journées de Statistique, JSBL 2002*, Bruxelles-Belgique, mai 2002.
- [CCM08] K.-W. Cheung, J. Chen, et Y.-S. Moon. Pose-tolerant non-frontal face recognition using EBGM. Dans *2nd IEEE International Conference on Biometrics : Theory, Applications and Systems*, pages 1–6, Arlington, Virginia, Sept.-Oct. 2008.
- [CCSM08] A. Chaari, F. Cabestaing, et D. Sellami-Masmoudi. A new approach to face image coding using gabor wavelet networks. Dans *Proceedings of the International Workshops on Image Processing Theory, Tools and Applications (IPTA'08)*, Sousse, Tunisie, novembre 2008.

- [CCSM09] A. Chaari, F. Cabestaing, et D. Sellami-Masmoudi. Modélisation de visages par un réseau d'ondelettes de gabor. Dans *Actes du 22eme Colloque sur le Traitement du Signal et des Images (GRETSI'09)*, Dijon, France, septembre 2009.
- [CGY96] I. Cox, J. Ghosn, et P. Yianilos. Feature-based face recognition using mixture-distance. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR'96*, pages 209 – 216, San Francisco, CA, USA, juin 1996.
- [Cha01] S. Chaptal. La biométrie s'infiltré dans le quotidien. *01 Informatique*, (1641), août 2001.
- [Chu92] C. K. Chui. *An introduction to wavelets*. Academic Press, Boston, septembre 1992.
- [CL07] B. Chen et A.-Q. Liu. A novel face coding scheme based on gabor wavelet networks and genetic algorithm. Dans *Proceedings of International Conference on Wavelet Analysis and Pattern Recognition*, Beijing, China, novembre 2007.
- [Com94] P. Comon. Independent Component Analysis, a new concept. *Signal Processing, Elsevier*, 3(287-314), avril 1994. Special issue on Higher-Order Statistics.
- [CrH08] C.-Y. Chang et H. rung Hsu. Apply an adaptive center selection algorithm to radial basis function neural network for face recognition. Dans *3rd International Conference on Innovative Computing Information and Control*, Dalian, China, juin 2008.
- [CSB07] M. S. Charifa, A. Suliman, et M. Bikdash. Face recognition using a hybrid general backpropagation neural network. Dans *IEEE International Conference on Granular Computing*, pages 510 – 516, Silicon Valley, USA, novembre 2007.
- [Dau85] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2(7) :1160–1169, juillet 1985.

- [Dau88] J. G. Daugman. Complete discrete 2D Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(7) :1169–1179, July 1988.
- [Dau90] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE transactions on information theory*, 36(5) :961–1005, septembre 1990.
- [Dau92] I. Daubechies. *Ten lectures on wavelets*. Regional conference series in applied mathematics, septembre 1992.
- [DH96] T. Draelos et D. Hush. A constructive neural network algorithm for function approximation. Dans *IEEE International Conference on Neural Networks*, volume 1, pages 50 – 55, Washington, DC, USA, juin 1996.
- [DHHR07] A. De-Heering, S. Houthuys, et B. Rossion. Holistic face processing is mature at 4 years of age : Evidence from the composite face effect. *Journal of Experimental Child Psychology*, 96 :57–70, novembre 2007.
- [DR07] V. J. Dattatray et S. H. Raghunath. Feature extraction and dimensionality reduction using Radon and Fourier transforms with application to face recognition. Dans *International Conference on Computational Intelligence and Multimedia Applications*, volume 2, pages 254–260, Sivakasi, Tamil Nadu, Inde, décembre 2007.
- [DWZ06] J. Dang, Y. Wang, et S. Zhao. Face recognition based on radial basis function neural networks using subtractive clustering algorithm. Dans *World Congress on Intelligent Control and Automation*, Dalian, China, juin 2006.
- [EAEK00] H. Elbakry, M. Abo-Elsoud, et M. Kamel. Integrating fourier descriptors and PCA with neural networks for face recognition. Dans *17th National Radio Science Conference*, C22/1-C22/8, Minufiya, Egypt, février 2000.
- [EC97] K. Etemad et R. Chellappa. Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America A*, 14(8) :1724–1733, octobre 1997.
- [ECW05] M. J. Er, W. Chen, et S. Wu. High-speed face recognition based on discrete

- cosine transform and RBF neural networks. *IEEE Transactions on Neural Networks*, 16(3) :676–691, mai 2005.
- [EWLT02] M. J. Er, S. Wu, J. Lu, et H. L. Toh. Face recognition with radial basis function (RBF) neural networks. *IEEE Transactions on Neural Networks*, 13(3) :697–710, mai 2002.
- [FCK01] R. S. Feris, R. Cesar, et V. Krüger. Efficient real-time face tracking in wavelet subspace. Dans *Proceedings of the 8th International Conference on Computer Vision (ICCV'01)*, pages 113–118, Vancouver, BC, Canada, juillet 2001.
- [FGTK01] R. S. Feris, J. Gemmell, K. Toyama, et V. Krüger. Hierarchical wavelet networks for facial feature localization. Dans *Proceedings of the 8th International Conference on Computer Vision (ICCV'01)*, Vancouver, BC, Canada, juillet 2001.
- [FJ00] R. Feris et R. C. Jr. Tracking facial features using Gabor wavelet networks. Dans *Proceedings XIII Brazilian Symposium on Computer Graphics and Image Processing, 2000.*, pages 22–27, Gramado, Brazil, octobre 2000.
- [GA04] R. Gottumukkal et V. K. Asari. An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letters*, 25 :429 – 436, mars 2004.
- [GD04] C. Garcia et M. Delakis. Convolutional face finder : A neural architecture for fast and robust face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11), novembre 2004.
- [Geo01] N. George. MEG-EEG et reconnaissance des visages. Cours séminaire, Centre MEG-EEG, juin 2001.
- [GLC00] G. Guo, S. Z. Li, et K. Chan. Face recognition by support vector machines. Dans *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 10.1109, pages 196 – 201, mars 2000.
- [GW04] C. Gasquet et P. Witomski. *Analyse de Fourier et applications*. 2-10-007684-1. Dunod, Paris, septembre 2004.
- [HHP01] B. Heisele, P. Ho, et T. Poggio. Face recognition with support vector

- machines : Global versus component-based approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 688–694, avril 2001.
- [Hiz09] W. Hizem. *Capteur intelligent pour la reconnaissance de visage*. these, Institut National des Télécommunications, Paris, 2009.
- [HL01] Z. Hafed et M. D. Levine. Face recognition using the discrete cosine transform. *International Journal of Computer Vision*, 43(3) :167–188, juillet 2001.
- [HM94] M. T. Hagan et M. B. Menhaj. Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6) :989–993, novembre 1994.
- [HSW98] J. Huang, X. Shao, et H. Wechsler. Face pose discrimination using support vector machines (SVM). Dans *Proceedings of the 14th International Conference on Pattern Recognition, ICPR'98*, volume 1, pages 154 – 156, Brisbane, Australia, août 1998.
- [JA05] O. Jemai et C. B. Amar. A new schema of image compression using wavelet networks. Dans *Third International Conference on Systems, Signals & Devices*, volume 3, Sousse, Tunisie, mars 2005.
- [JK08a] J.-S. Jang et J.-H. Kim. Fast and robust face detection using evolutionary pruning. *IEEE Transactions on Evolutionary Computation*, 12(5) :562–572, octobre 2008.
- [JK08b] Y. B. Jemaa et S. Khanfir. Automatic Gabor features extraction for face recognition using neural networks. Dans *Proceedings of the International Workshops on Image Processing Theory, Tools and Applications (IPTA'08)*, Sousse, Tunisie, novembre 2008.
- [Jol01] J. Jolion. Graph matching : what are we really talking about? Dans *IAPR workshop on Graph-based Representations in Pattern Recognition*, mai 2001.
- [Jou06] R. Jourani. Reconnaissance de visages. Rapport de Master, Faculté des Sciences, université Mohamed V, Rabat, Maroc., novembre 2006.

- [JP87] J. Jones et L. Palmer. The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6) :1187–1211, décembre 1987.
- [KCA02] N. Kwak, C.-H. Choi, et N. Ahuja. Face recognition using feature extraction based on independent component analysis. Dans *Proceedings of the IEEE International Conference on Image Processing (ICIP'02)*, volume 2, pages 337–340, Rochester, NY, USA, septembre 2002.
- [KP97] C. Kotropoulos et I. Pitas. Rule based face detection in frontal views. Dans *Acoustics, Speech, and Signal Processing*, volume 4, pages 2537–2540, avril 1997.
- [KP07] K.-C. Kwak et W. Pedrycz. Face recognition using an enhanced independent component analysis approach. *IEEE Transactions on Neural Networks*, 18(2) :530 – 541, mars 2007.
- [KRG06] N. Kela, A. Rattani, et P. Gupta. Illumination invariant elastic bunch graph matching for efficient face recognition. Dans *Conference on Computer Vision and Pattern Recognition Workshop*, pages 42–48, New York, USA, juin 2006.
- [Kru00] V. Kruger. *Gabor Wavelet Networks for Object Representation*. Thèse, Christian-Albrechts-Universität zu Kiel, Allemagne, septembre 2000.
- [KS02a] V. Krüger et G. Sommer. Wavelet networks for face processing. *Journal of the Optical Society of America A*, 19(6) :1112–1119, June 2002.
- [KS02b] V. Krüger et G. Sommer. Gabor wavelet networks for efficient head pose estimation. *Image and Vision Computing*, 20(9-10) :665–672, août 2002.
- [KSC06] V. Kumar, M. Savvides, et X. Chnyan. Correlation pattern recognition for face recognition. *Proceedings of the IEEE*, 94(11) :1963–1976, novembre 2006.
- [KWT87] M. Kass, A. Witkin, et D. Terzopoulos. Snakes : Active contour models. *International Journal of Computer Vision*, 1(4) :321–331, 1987.
- [Lee96] T. S. Lee. Image representation using 2D gabor wavelets. *IEEE Tran-*

- sactions on Pattern Analysis and Machine Intelligence*, 18(10) :959–971, octobre 1996.
- [Lek97] G. Lekutai. *Adaptive Self-Tuning Neuro Wavelet Network Controllers*. Phd thesis, Polytechnic Institute, Faculty of the Virginia, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, mars 1997.
- [LGTB97] S. Lawrence, C. L. Giles, A. C. Tsoi, et A. D. Back. Face recognition : a convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 8(1) :98–113, janvier 1997.
- [LHZZ08] B. Luo, Y.-J. Hao, W.-H. Zhang, et Z.-S. Liu. Comparison of PCA and ICA in face recognition. Dans *International Conference on Apperceiving Computing and Intelligence Analysis*, pages 241 – 243, Chengdu, China, décembre 2008.
- [Liu04] C. Liu. Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5) :572–581, mai 2004.
- [LJ04] S. Z. Li et A. K. Jain. *Handbook of Face Recognition*. Springer, NewYork, USA, septembre 2004.
- [LK00] R. Lotlikar et R. Kothari. Fractional-step dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(6) :623–628, juin 2000.
- [LLA06] C. Lu, W. Liu, et S. An. Face recognition with only one training sample. Dans *Chinese Control Conference*, pages 2215–2219, Zhangjiajie, Hunan, China, août 2006.
- [LLC01] H.-J. Lee, W.-S. Lee, et J.-H. Chung. Face recognition using Fisherface algorithm and elastic graph matching. Dans *Proceedings of the 8th International Conference on Image Processing, ICIP'2001*, volume 1, pages 998–1001, Thessaloniki, Greece, octobre 2001.
- [LLO97] J. Lampinen, J. Laaksonen, et E. Oja. Neural network systems, techniques and applications in pattern recognition. Technical report, Helsinki University of Technology, février 1997.

- [LPV03] J. Lu, K. Plataniotis, et A. Venetsanopoulos. Face recognition using LDA-based algorithms. *IEEE Transactions on Neural Networks*, 14(1) :195–200, janvier 2003.
- [LW99] C. Liu et H. Wechsler. Comparative assessment of independent component analysis (ICA) for face recognition. Dans *Second International Conference on Audio- and Video-based Biometric Person Authentication*, pages 211–216, Washington D.C., USA, mars 1999.
- [LWL07] M. Li, F. Wu, et X. Liu. Face recognition based on WT, fastICA and RBF neural network. Dans *Third International Conference on Natural Computation*, volume 2 of 10.1109/ICNC.2007.371, pages 3–7, août 2007.
- [LYYL08] H.-J. Lin, S.-H. Yen, J.-P. Yeh, et M.-J. Lin. Face detection based on skin color segmentation and SVM classification. Dans *Second International Conference on Secure System Integration and Reliability Improvement*, pages 230 – 231, juillet 2008.
- [MA08] M. Mazloom et S. Ayat. Combinational method for face recognition : Wavelet, PCA and ANN. Dans *Digital Image Computing : Techniques and Applications, 2008. DICTA '08.*, décembre 2008.
- [MAHE07] N. Morizet, F. Amiel, I. D. Hamed, et T. Ea. A comparative implementation of PCA face recognition algorithm. Dans *14th IEEE International Conference on Electronics, Circuits and Systems*, volume 2, pages 865 – 868, décembre 2007.
- [Mal98] S. Mallat. *A wavelet tour of signal processing*. 0124666051. Academic Press, San Diego, London, Boston, septembre 1998.
- [MAS08] D. Monzo, A. Albiol, et J. Sastre. HOG-EBGM vs. Gabor-EBGM. Dans *15th IEEE International Conference on Image Processing*, pages 1636–1639, San Diego, California, octobre 2008.
- [Mc08] Z. Mu-chun. Face recognition based on fastICA and RBF neural networks. Dans *International Symposium on Information Science and Engineering*, volume 1, pages 588 – 592, Dec. 2008.

- [MK01] A. Martinez et A. Kak. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2) :228–233, septembre 2001.
- [MP01] H. Moon et P. Phillips. Computational and performance aspects of PCA-based face recognition algorithms. *Perception*, 30 :303–321, septembre 2001.
- [MSP05] M. Meade, S. C. Sivakumar, et W. J. Phillips. Comparative performance of principal component analysis, Gabor wavelets and discrete wavelet transforms for face recognition. *Canadian Journal of Electrical and Computer Engineering*, 30(2) :93 – 102, septembre 2005.
- [MWJI08] A. Mohamed, Y. Weng, J. Jiang, et S. Ipson. Face detection based neural networks using robust skin color segmentation. Dans *5th International Multi-Conference on Systems, Signals and Devices*, Amman, Jordan, juillet 2008.
- [Ngu05] T. P. Nguyen. Détection de visage dans images de couleur en utilisant une caractéristique invariante. Dans *Sciences of Electronic, Technologies of Information and Telecommunications*, Sousse, Tunisie, mars 2005.
- [OFG97] E. Osuna, R. Freund, et F. Girosit. Training support vector machines : an application to face detection. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR'97*, pages 130 – 136, San Juan, PR, USA, juin 1997.
- [ORPD98] Y. Oussar, I. Rivals, L. Personnaz, et G. Dreyfus. Training wavelet networks for nonlinear dynamic input-output modeling. *Neurocomputing*, 20(1-3) :173–188, août 1998. Elsevier Science.
- [Ous98] Y. Oussar. *Réseaux d'ondelettes et réseaux de neurones pour la modélisation statique et dynamique de processus*. Thèse, Université Pierre et Marie Curie, juillet 1998.
- [Par02] F. Parrain. *Capteur intégré d'empreintes digitales à microstructures piezo-résistives*. Thèse, Institut National Polytechnique de Grenoble, décembre 2002.
- [Pie07] J. Pierson. *La biométrie : L'identification par le corps*. 2-7462-1831-3. Lavoisier, 1 édition, septembre 2007.

- [PKB05] S. Pang, D. Kim, et S. Y. Bang. Face membership authentication using SVM classification tree generated by membership-based LLE data partition. *IEEE Transactions on Neural Networks*, 16(2) :436–447, mars 2005.
- [PKG98] S. Pittner, S. V. Kamarthi, et Q. Gao. Wavelet networks for sensor signal classification in flank wear assessment. *Journal of Intelligent Manufacturing*, 9 :315–322, septembre 1998.
- [PL08] M. Park et S. Lee. Face modeling and tracking with Gabor filter response prior. Dans *19th International Conference on Pattern Recognition*, Tampa, Florida, USA, décembre 2008.
- [Pog85] T. Poggio. Early vision : From computational structure to algorithms and parallel hardware. *Computer Vision, Graphics, and Image Processing*, 31(2) :139–155, 1985.
- [PSKK05] I. Paliy, A. Sachenko, V. Koval, et Y. Kurylyak. Approach to face recognition using neural networks. Dans *IEEE Intelligent Data Acquisition and Advanced Computing Systems : Technology and Applications*, pages 112–115, Sofia, Bulgarie, septembre 2005.
- [RB76] C. Rader et N. Brenner. A new principle for fast Fourier transformation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(3) :264–266, septembre 1976.
- [RCM07] A. Rekik, A. Chaari, et D. S. Masmoudi. Face recognition by applying wavelet and bidimensional PCA. Dans *Fourth International Conference on Systems, Signals and Devices*, Sousse, Tunisie, mars 2007.
- [Row99] H. A. Rowley. *Neural Network-Based Face Detection*. Thèse, School of Computer Science-Carnegie, Mellon University, mai 1999.
- [RS08] R. Rioux et T. Simon. Redondance de classes pour la reconnaissance de visages par SVM. Dans *14ème Colloque National de la Recherche en IUT*, Campus de la Doua , Lyon-Villeurbanne, France, mai 2008.
- [RY90] K. R. Rao et P. Yip. *Discrete Cosine Transform : Algorithms, Advantages, Applications*. Academic Press, Boston, 1990.
- [SAA03] A. Samet, C. B. Amar, et M. A. Alimi. Ondelettes, paquets d’ondelettes et

- réseaux d'ondelettes pour la compression d'images. Dans *3ème journée des Jeunes Chercheurs en Génie Electrique et Informatique*, volume 2, pages 108–112, Mahdia, Tunisie, mars 2003.
- [SAI05] A. Samra, S. E. T. G. Allah, et R. Ibrahim. Face recognition using wavelet transform, fast Fourier transform and discrete cosine transform. Dans *IEEE International Symposium on Micro-NanoMechatronics and Human Science*, volume 1, pages 272–275, décembre 2005.
- [Sam07] C. Samir. *Analyse des déformations des visages 3D utilisant les chemins géodésiques dans l'espace des surfaces faciales*. Thèse, Université de Lille 1 et Institut National des Télécommunications, Lille, octobre 2007.
- [SAZ⁺08] S. Said, B. B. Amor, M. Zaied, C. B. Amor, et M. Daoudi. Vers une reconnaissance faciale 3D rapide utilisant les réseaux d'ondelettes. Dans *Traitement et analyse de l'information : Méthodes et Applications*, Hammamet, Tunisie, mai 2008.
- [SG08] H. Sahoolizadeh et Y. A. Ghassabeh. Face recognition using eigen-faces and Fisherfaces and neural networks. Dans *7th IEEE International Conference on Cybernetic Intelligent Systems*, pages 1–6, septembre 2008.
- [Sid07] D. D. Sidibe. *Une technique de relaxation pour la mise en correspondance d'images : Application à la reconnaissance d'objets et au suivi du visage Jury*. Thèse, Université Montpellier 2, décembre 2007.
- [SKC07] H. Shin, S.-D. Kim, et H.-C. Choi. Generalized elastic graph matching for face recognition. *Pattern Recognition Letters*, 28(9) :1077–1082, janvier 2007. Elsevier Science Inc.
- [SL07] Y. Sabrina et M. T. Laskri. Un modèle basé templates matching/réseau de neurones pour la reconnaissance des visages humains. Dans *JED'2007*, Annaba, Algérie, mai 2007.
- [SP96] K. Sobottka et I. Pitas. Face localization and facial feature extraction based on shape and color information. Dans *Proceedings of the 3rd International Conference on Image Processing, ICIP'96*, volume 3, pages 483–486, Lausanne, Switzerland, septembre 1996.

- [SRH03] A. Schwaninger, S. Ryf, et F. Hofer. Configural information is processed differently in perception and recognition of faces. *Vision Research*, 43 :1501–1505, mars 2003.
- [SW02] D.-R. Sun et L.-N. Wu. A local-to-holistic face recognition approach using elastic graph matching. Dans *International Conference on Machine Learning and Cybernetics*, volume 1, pages 240–242, Beijing, China, novembre 2002.
- [Ter03] S. Tertois. *Réduction des effets des non-linéarités dans une modulation multiporteuse à l'aide de réseaux de neurones*. Thèse, Université de Rennes 1, décembre 2003.
- [Thu00] M. Thuillard. A review of wavelet networks, wavenets, fuzzy wavenets and their applications. Dans *European Symposium on Intelligent Techniques, ESIT'2000*, pages 5–16, Aachen, Germany, septembre 2000.
- [TKP00] A. Tefa, C. Kotropoulos, et J. Pitas. Face authentication by using elastic graph matching and support vector machines. Dans *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2409–2412, Istanbul, Turkey, juin 2000.
- [TP91] M. Turk et A. Pentland. Face recognition using eigenfaces. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'91)*, 10.1109, pages 586 – 591, Lahaina, Maui, HI, USA, juin 1991.
- [V08] J.-M. Vézien. *Traitement des Images et Vision par Machine*. Cours Master Professionnel Informatique, Université Paris XI, École Supélec., novembre 2008.
- [Vap95] V. N. Vapnik. *The Nature of Statistical Learning Theory*. 0-387-98.780.0. Library of Congress Cataloging in Publication Data, septembre 1995.
- [VGJ04] M. Visani, C. Garcia, et J.-M. Jolion. Two-dimensional-oriented linear discriminant analysis for face recognition. Dans K. Wojciechowski, B. Smolka, H. Palus, R. Kozera, W. Skarbek, et L. Noakes, editors, *International Conference on Computer Vision and Graphics, ICCVG'2004*. Published in special issue of *Computational Imaging and Vision*, Vol. 32, Dec. 2005, pages 1008–1017. Springer, septembre 2004.

- [VGJ05] M. Visani, C. Garcia, et J.-M. Jolion. Une nouvelle méthode de représentation des visages pour leur reconnaissance : l'analyse discriminante bilinéaire. Dans *Actes du Colloque Compression et Représentation des Signaux Audiovisuels (CORESA'05)*, pages 103–108, Rennes, France, novembre 2005.
- [Vis05] M. Visani. *Vers de nouvelles approches discriminantes pour la reconnaissance automatique de visages*. Thèse, Institut National des Sciences Appliquées de Lyon, novembre 2005.
- [VVCM09] N. Vo, D. Vo, S. Challa, et B. Moran. Compressed sensing for face recognition. Dans *IEEE Symposium on Computational Intelligence for Image Processing (CIIPBet'09)*, Nashville, TN, USA, mars 2009.
- [Wan08] W. Wang. Face recognition based on radial basis function neural networks. Dans *International Seminar on Future Information Technology and Management Engineering*, pages 41–44, Leicestershire, UK, novembre 2008.
- [WF99] L. Wiskott et J.-M. Fellous. *Face Recognition by Elastic Bunch Graph Matching*, chapter 11, pages 355–396. Press International Series on Computational Intelligence. janvier 1999.
- [WFKvdM97] L. Wiskott, J.-M. Fellous, N. Krüger, et C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) :775–779, juillet 1997.
- [WLLW97] W.-Y. Wang, T.-T. Lee, C.-L. Liu, et C.-H. Wang. Function approximation using fuzzy neural networks with robust learning algorithm. *IEEE transactions on Systems, Man and Cybernetics*, 27(4), août 1997.
- [XSK05] C. Xie, M. Savvides, et K. V. Kumar. Quaternion correlation filters for face recognition in wavelet domain. Dans *Procs. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'05)*, volume 2, pages 85–88, Philadelphia, PA, USA, mars 2005.
- [Yam00] W. S. Yambor. Analysis of PCA-based and Fisher discriminant-based image recognition algorithms. Technical report, Computer Science Department Colorado State University Fort Collins, septembre 2000.

- [YFQ07] H. Yin, P. Fu, et J. Qiao. Face recognition based on DCT and 2D LDA. Dans *Second International Conference on Innovative Computing, Information and Control*, pages 581–581, Kumamoto, Japan, septembre 2007.
- [YH94] G. Yang et T. S. Huang. Human face detection in a complex background. *Pattern Recognition*, 27(1) :53–63, septembre 1994.
- [YKA02] M.-H. Yang, D. J. Kriegman, et N. Ahuja. Detecting faces in images : A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1) :34–59, janvier 2002.
- [YZFyY04] J. Yang, D. Zhang, A. F. Frangi, et J. yu Yang. Two-dimensional PCA : A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1) :131–138, janvier 2004.
- [ZAA03] M. Zaeid, C. B. Amar, et A. Alimi. Award a new wavelet based beta function. Dans *Second International Conference on Signals Systems Decision and Information Technology*, mars 2003.
- [ZAA05] M. Zaied, C. B. Amar, et A. Alimi. Beta wavelet networks for face recognition. *Journal of Decision Systems*, 14 :109–122, septembre 2005.
- [Zai08] M. Zaied. *Étude des réseaux d'ondelettes Bêta : Application à la reconnaissance de visages*. Thèse, École Nationale d'Ingénieurs de Sfax, Sfax, Tunisie, février 2008.
- [ZB92] Q. Zhang et A. Benveniste. Wavelet networks. *IEEE Transactions on Neural Networks*, 3(6) :889 – 898, novembre 1992.
- [ZCK98] W. Zhao, R. Chellappa, et A. Krishnaswamy. Discriminant analysis of principal components for face recognition. Dans *3rd IEEE International Conference on Face and Gesture Recognition*, pages 36–341, avril 1998.
- [ZCMC06] Y. Zana, R. Cesar, et J. Mena-Chalco. Human and machine recognition of fourier-bessel filtered face images. Dans *7th International Conference on Automatic Face and Gesture Recognition*, pages 299–304, Southampton, United Kingdom, avril 2006.
- [ZCRP03] W. Y. Zhao, R. Chellapa, A. Rosenfeld, et P. Phillips. Face recognition : A

- literature survey. Technical Report 4, National Institute of Standards and Technology, décembre 2003.
- [ZGD07] L. Zhang, Q. Gao, et Z. D. Block independent component analysis for face recognition. Dans *14th International Conference on Image Analysis and Processing*, 10.1109/ICIAP.2007.4362782, pages 217 – 222, septembre 2007.
- [Zha93] Q. Zhang. Regressor selection and wavelet network construction. Dans *Proceedings of the 32nd Conference on Decision and Control*, pages 3668–3693, December 1993.
- [Zha97] Q. Zhang. Using wavelet network in nonparametric estimation. *IEEE Transactions on Neural Networks*, 8(2) :227–236, mars 1997.
- [Zha99] Q. Zhang. *Identification et Surveillance de Systèmes Dynamiques*. HDR, Université de Rennes 1, janvier 1999.
- [Zha07] S. Zhao. Apprentissage et recherche par le contenu visuel de catégories sémantiques d’objets vidéo. Rapport de Master, Université Paris Descartes, juillet 2007.
- [ZKC⁺98] W. Zhao, A. Krishnaswamy, R. Chellappa, D. Swets, et J. Weng. *Face Recognition : From Theory to Applications*, volume 163 of *Computer and Systems Sciences*, chapter Discriminant Analysis of Principal Components for Face Recognition, pages 73–85. juin 1998.
- [ZM05a] Y. Zhi et G. Ming. Adaptive matching wavelet networks for face recognition. Dans *Proceedings of the 12th International Conference on Image Processing (ICIP’2005)*, volume 1, pages 345–8, Genova, Italia, septembre 2005.
- [ZM05b] Y. Zhi et G. Ming. A SOM-wavelet networks for face identification. Dans *IEEE International Conference on Multimedia and Expo*, pages 4–8, Amsterdam, Netherlands, juillet 2005.
- [ZTP05] S. Zafeiriou, A. Tefas, et I. Pitas. Exploiting discriminant information in elastic graph matching. Dans *Proceedings of the 12th International Confe-*

rence on Image Processing (ICIP'2005), 3, pages 768–771, Genova, Italia, septembre 2005.

Résumé

Les travaux de recherche menés dans le cadre de cette thèse portent sur la reconnaissance de visages par réseaux d'ondelettes de Gabor. Ce réseau est très utilisé pour la représentation d'image. Après une étape d'apprentissage, il reconstruit l'image en se basant sur un nombre défini d'ondelettes dont les paramètres ont été ajustés.

L'originalité du travail se situe au niveau de l'utilisation de plusieurs images pour l'extraction de la signature d'un individu. Un algorithme d'apprentissage dédié a été développé de façon à prendre en compte plusieurs images et à en extraire une signature discriminante associée à un masque des confiances accordées aux pixels. Ce masque est construit en associant un coefficient d'importance à chaque ondelette. Ce coefficient est calculé en suivant et analysant l'évolution des paramètres des ondelettes au cours de l'apprentissage.

Une étude comparative sur la base de données YaleFaces a montré qu'avec cette méthode la discrimination des individus est améliorée par rapport aux méthodes conventionnelles.

Mots clés : Reconnaissance de visages, ondelettes, réseau d'ondelettes de Gabor, apprentissage automatique, identification biométrique.

Abstract

This research work is oriented toward face recognition by Gabor Wavelet networks. This network has proved to be very efficient for image representation. It reconstructs, after a learning step, the image with a set of adapted wavelets.

The originality of this work lies in the use of multiple images to extract the signature of an individual. A special learning algorithm was developed to use several images and extract a distinct signature, associated to a mask of significant pixels. This mask is built from coefficients of significance associated to each wavelet. The coefficients are calculated by analyzing the evolution of wavelet parameters during the learning step.

A comparative study, on YaleFaces database, has shown that with this method face recognition is improved compared to conventional methods.

Key-words : Face recognition, wavelets, Gabor wavelet network, machine learning, biometric identification.