

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



Mémoire de magister

**Présenté à l'Université de Guelma
Faculté des Sciences et de l'Ingénierie**

Département de Génie Electrique
Spécialité : Signaux et Images

Présenté par Mr. **Khelil Jalel**

Reconnaissance Automatique de Visages par Approche Locale

Sous la direction du **Pr. Hicham Tebbikh**

Jury

Dr. S. KECHIDA	Université de Guelma	Présidente
Pr. H. TEBBIKH	Université de Guelma	Rapporteur
Dr. A. BOUKROUCHE	Université de Guelma	Examineur
Dr. M. ROUAINIA	Université de Skikda	Examinatrice
M^{elle} C. ROUABHIA	Université de Guelma	Invitée

2009

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude à mon directeur de mémoire le professeur **Tebbikh Hicham** directeur du Laboratoire d'Automatique et Informatique de Guelma –LAIG- pour m'avoir encadré et guidé avec clairvoyance, pour ses précieux conseils, sa confiance et son soutien tout au long de la réalisation de ce travail de recherche.

Je remercie vivement M^{elle} **Kechida Sihem** docteur à l'université de Guelma pour avoir accepté de présider mon jury de mémoire.

Mes sincères remerciements vont aussi aux membres de jury M^{er} **Boukrouche** docteur à l'université de Guelma, Madame **Rouainia Mounira** docteur à l'université de Skikda.

Je suis également très reconnaissant envers M^{elle} **Rouabhia Chahrazed** pour son aide précieuse qu'elle m'a apporté, sa disponibilité ainsi que ses nombreux conseils.

Je remercie également mon frère **Khelil Khaled** docteur à l'université de Souk Ahras pour son aide considérable et son soutien constant.

Un grand merci à mes chers parents, à toute ma famille et à ceux qui m'ont soutenu de près ou de loin.

Enfin, rien n'aurait été possible sans le soutien profond de ma femme qui m'a supporté patiemment, et m'a encouragé aux moments difficiles durant tout le chemin de ce mémoire.

Jalel Khelil

Résumé

Le travail présenté dans ce mémoire s'inscrit dans le contexte de l'identification automatique de visages à partir de séquences d'images vidéo. Le système de reconnaissance automatique de visages mis au point se décompose en trois étapes fondamentales:

Une première étape consacrée à la localisation des différentes caractéristiques du visage telles que les yeux, le nez et la bouche. Ensuite, nous avons procédé à la localisation des différents points fiduciels (les points d'intérêts).

Quant à la deuxième étape, elle consiste à la représentation du visage. Pour cela, nous avons appliqué à chacun des points fiduciels localisés un banc de filtres de Gabor de différentes fréquences et différentes tailles (40 filtres) en effectuant la convolution de l'image du visage autour de ces points avec plusieurs ondelettes complexes 2D de Gabor paramétrées. Chaque point caractéristique du visage sera ainsi représenté par un vecteur de coefficients complexes de Gabor appelé jet.

Enfin, l'étape de reconnaissance basée sur la maximisation d'une mesure de similarité entre les différents jets de Gabor d'un visage test et tous les jets de la base d'apprentissage.

Mots clés: Biométrie, EBGM, Jet de Gabor, Points fiduciels, Reconnaissance automatique de visages, Ondelettes.

Abstract

In this memory, we propose an automatic face recognition system from video sequences. This system contains three main stages:

In the first one, the facial features positions (eyes, nose and mouth) are located and then the fiducial points extracted.

Second, face recognition is carried out using Gabor filters. For this purpose, a family of Gabor wavelet responses are computed (using convolution) at each of the fiducial points, where the Gabor wavelets can be tuned in frequency, orientation and spatial range of action in the image. So, each characteristic point of the face will be represented by a complex vector called jet.

Finally, face recognition is implemented by computing a similarity measure between each jets vector of an unknown face and the ones of all the images in the training database.

Key words: Biometric, EBGM, Gabor jet, Fiducial points, Localization, Automatic face recognition, Gabor wavelets.

المخلص

العمل المقترح في هذه المذكرة يندرج في إطار المعرفة الآلية للوجوه باستعمال سلاسل صور فيديو، من أجل ذلك النظام الآلي المقترح يتضمن ثلاثة مراحل أساسية:

- تعيين مختلف العناصر المميزة للوجه و هي العينان، الأنف و الفم ثم تحديد عدة نقاط مميزة،
- تمثيل الوجه باستعمال مجموعة مرشحات قابور ذات تواترات و حجوم مختلفة (40مرشحا). ينتج عن هذه المرحلة تمثيل كل وجه بشعاع عناصره هي المعاملات المركبة لمرشحات قابور،
- معرفة الوجوه باستعمال القيمة الأعظمية لقياس التشابه بين الشعاع الممثل للوجه المراد معرفته وبقية الأشعة الممثلة لكل وجوه قاعدة التدريب .

Sommaire

Introduction Générale	1
------------------------------	----------

Chapitre I : Biométrie

I.1 Introduction	3
I.2 Biométrie	3
I.3 Classification des techniques biométriques	3
I.4 Domaines d'application de la biométrie	3
I.5 Quelques caractéristiques biométriques	4
I.5.1 Empreintes digitales	4
I.5.2 Géométrie de la main	5
I.5.3 Iris	5
I.5.4 Rétine	6
I.5.5 Voix	6
I.5.6 Signature	7
I.5.7 Visage	7
I.6 Principe de fonctionnement d'un système biométrique	8
I.7 Evaluation des systèmes biométriques	10
I.8 Performances des systèmes biométriques	12
I.9 Conclusion	12

Chapitre II : Reconnaissance de Visages

II.1 Introduction	13
II.2 Reconnaissance de visages	13
II.3 Processus de reconnaissance des visages	14
II.3.1 Images Fixes	14
II.3.2 Images Vidéo	14
II.3.3 Images 3D	15
II.4 Différentes étapes d'un processus de reconnaissance de visages	15
II.5 Extraction des caractéristiques du visage	17
II.5.1 Méthodes génériques	18
II.5.2 Méthodes basées sur les modèles de caractéristiques	18
II.5.3 Méthodes basées sur la similitude structurelle	18
II.6 Approches existantes de reconnaissance de visages	18
II.6.1 Approches globales	19
II.6.2 Approches locales	24
II.6.2.1 Approches basées sur les éléments caractéristiques du visage	24
a. Approches géométriques	24
b. Approches basées sur les graphes	24
II.6.2.2 Approches basées sur les régions faciales	25
II.6.3 Approches Hybrides	26
II.6.3.1 Approches modulaires	26
II.6.3.2 Analyse des caractéristiques locales	26

II.7 Quelques difficultés inhérentes à la reconnaissance de visages	26
II.7.1 Variations de pose	26
II.7.2 Changements d'illumination	27
II.7.3 Variations d'expressions faciales	28
II.7.4 La taille de la base de visages	28
II.8 Evaluation des performances des systèmes de reconnaissance de visages	29
II.9 Conclusion	30

Chapitre III : Approche EBGM

III.1 Introduction	31
III.2 Approche EBGM	31
III.3. Généralités sur les ondelettes de Gabor	32
III.3.1 Transformée de Fourier	33
III.3.2 Transformée de Gabor	36
III.3.3 Ondelettes	38
III.3.4 Analyse multi résolution et analyse temps échelle	38
III.3.5 Ondelettes de Gabor	39
III.3.6 Ondelettes de Gabor 2D	40
III.4. Etapes de l'algorithme EBGM	43
III.5. Analyse de sensibilité des différents points fiduciels	45
III.6 Caractérisation du visage	47
III.6.1 Extraction de caractéristiques de la texture	47
III.6.2 Jet de Gabor	47
III.6.3 Représentation des visages	48
III.6.4 Calcul de similarité	49
III.7 Applications de l'algorithme EBGM	50
III.8 Conclusion	51

Chapitre IV : Simulations & Résultats

IV.1 Introduction	52
IV. 2. Base de visages utilisés	52
IV.3 Système de reconnaissance de visages	54
IV.3.1 Redimensionnement des images de visages	56
IV.3.2 Localisation des éléments caractéristiques du visage	56
1. Localisation des iris	56
2. Localisation du nez et de la bouche	58
IV.3.3. Localisations de sept points fiduciels	60
IV.3.4. Représentation des visages	62
IV.3.5 Mesure de similarité	62
IV.6. Résultats de la reconnaissance	64
IV.7. Amélioration du taux de reconnaissance	68
IV.8 Conclusion	73
Conclusion Générale	74
Bibliographie	75

Introduction Générale

Reconnaître un visage c'est lui affecter une identité parmi celles d'un ensemble de visages connus. Les humains sont dotés d'une excellente aptitude à identifier leurs semblables. Les études biologiques tendent à prouver que la reconnaissance humaine des visages constitue un processus spécifique de reconnaissance d'objets qui serait mené dans une région particulière du cerveau. Ainsi, la reconnaissance d'une personne passe par l'utilisation de nombreuses informations visuelles (visage, posture, coupe de cheveux, etc.), ou non visuelles (voix, parfum, etc.).

Beaucoup de psycho-physiciens et neuroscientifiques ont cherché à déterminer si le cerveau humain se base plutôt sur l'étude de caractéristiques globales ou locales du visage pour sa reconnaissance. De nos jours, la plupart des chercheurs s'accordent à dire que les deux types d'informations sont utilisés. Si l'analyse de l'information des basses fréquences de l'image (reflétant sa globalité) sont suffisantes à la reconnaissance d'un visage très familier, les cheveux, la forme du visage, les yeux et la bouche sont reconnues comme étant des caractéristiques primordiales dans la perception et la mémorisation des visages.

Les premiers travaux portant sur la reconnaissance de visages humains ont été menés dans le domaine de la neuropsychologie dès le début des années 1950. Depuis, notamment grâce aux avancées dans le domaine de l'imagerie médicales, des études sont venues compléter ces recherches. C'est avec un grand intérêt que les scientifiques cherchant à automatiser le processus suivent ces travaux, dans l'espoir de concevoir des algorithmes capables de copier les prodigieuses facultés de reconnaissance du cerveau humain. En effet, les premiers travaux concernant la reconnaissance automatique de visages remontent au début des années 1970. Les techniques introduites à l'époque utilisaient pour la plupart des mesures estimées autour des éléments faciaux des visages tels que les yeux, le nez et la bouche. Mais ce n'est qu'au début des années 1990 que le volume de recherches concernant la reconnaissance de visages a réellement commencé à croître. Par la suite, l'engouement pour cette problématique n'a fait qu'augmenter si bien qu'aujourd'hui la reconnaissance de visages constitue l'un des domaines les plus explorés de la reconnaissance de formes et de l'analyse d'images. Néanmoins, on ne peut pas considérer jusqu'à ce jour que la reconnaissance automatique de visages soit un problème résolu, comme l'a montré la récente évaluation des principales techniques proposées dans ce contexte menée par le National Institute of Standards and Technology (NIST). En effet, malgré la maturité des meilleures techniques évaluées, celles-ci peuvent être insuffisantes dans le contexte d'applications réelles, où les sources de variabilité sont multiples (changements d'éclairage, de poses et des expressions faciales). Etant donné que les visages sont des objets structurellement très proches, ces changements peuvent engendrer des différences plus importantes entre deux vues d'un même visage qu'entre deux vues de visages différents. Cela fait de la reconnaissance automatique de visages un problème de classification particulièrement difficile.

Dans le cadre de la préparation de notre mémoire de magister, nous nous intéressons aux méthodes à base de caractéristiques locales notamment à la méthode dite **Elastic Bunch Graph Matching: EBGM**. Cet algorithme trouve ses fondements dans les neurosciences en imitant le fonctionnement de certaines cellules spécialisées localisées dans le cortex visuel humain. C'est un algorithme local (il ne traite pas directement les valeurs de niveaux de gris des pixels d'une image de visage), ce qui lui confère une plus grande robustesse aux changements d'éclairément, de pose et d'expressions faciales. Basé sur des points caractéristiques du visage, cet algorithme effectue une transformation en ondelettes 2D de Gabor afin de représenter chacun de ces points caractéristiques par un jeu de coefficients complexes appelé jet. Pour effectuer une reconnaissance d'une image test, on effectue une mesure de similarité entre les différents Jets.

Afin de détailler ces apports, nous avons choisi d'articuler notre mémoire autour de quatre chapitres principaux.

Dans le **premier chapitre**, nous avons présenté des généralités sur la biométrie et le principe de fonctionnement des systèmes biométriques. Un panorama de techniques biométriques les plus utilisées qu'il s'agisse de l'iris, de la géométrie de la main, des empreintes digitales, de la voix, de la signature manuscrite, ou du visage humain a été détaillé. Leurs avantages et inconvénients ont été également discutés.

Dans le **deuxième chapitre**, nous avons présenté les différentes étapes d'un processus de reconnaissance de visages et nous avons passé en revue les principales techniques de reconnaissance proposées à ce jour. Ces techniques se décomposent en trois grandes familles: i) les approches globales pour lesquelles les caractéristiques du visage sont extraites directement depuis l'ensemble des valeurs de pixels des images, ii) les approches locales basées sur l'extraction de caractéristiques des régions bien définies du visage comme les yeux, le nez, la bouche, etc. Parmi les approches locales utilisées, on compte les approches géométriques et les approches basées sur les graphes dites Elastic Bunch Graph Matching, iii) les techniques dites hybrides qui combinent les méthodes locales et globales. En outre, nous avons présenté un aperçu sur les principales difficultés inhérentes à la reconnaissance automatique de visages. Enfin, nous avons discuté des différents protocoles d'évaluation des performances des systèmes de reconnaissance de visages.

Le **troisième chapitre** a été consacré à présenter une méthode de reconnaissance de visages qui tient ses origines de la biologie c'est l'Elastic Bunch Graph Matching (EBGM). Elle imite le fonctionnement du système visuel humain en traitant l'information fréquentielle autour des points caractéristiques du visage en se basant sur les ondelettes 2D de Gabor entièrement paramétrées. Nous avons détaillé ses origines en allant de la transformée de Fourier jusqu'au filtre de Gabor 2D et expliqué son principe avant d'arriver à son application dans le domaine de la reconnaissance de visages.

Le **quatrième chapitre** a été consacré à l'application de l'algorithme EBGM sur des séquences vidéo. Pour cela, nous avons détaillé les différentes étapes de la mise en œuvre du système de reconnaissance faciale, reporté et commenté les différents résultats de simulations menés sur la base de visages B-LAIG acquises au laboratoire d'Automatique et Informatique de Guelma.

Une conclusion terminera ce mémoire avec les principales perspectives des futures recherches.

Chapitre I
BIOMETRIE

I.1 Introduction

Reconnaître ou identifier une personne revient à déterminer son identité en utilisant, par exemple, une pièce d'identité, une signature manuscrite, une clé magnétique ou un mot de passe. Cependant, ces moyens peuvent être volés, perdus ou oubliés, c'est pourquoi la biométrie est une alternative à ces moyens classiques. Ce chapitre vise donc à présenter, outre la biométrie et le principe des systèmes biométriques, un panorama de techniques biométriques les plus utilisées qu'il s'agisse des empreintes digitales, de l'iris, de la géométrie de la main, de la voix, de la signature manuscrite, ou du visage humain. Leurs avantages et inconvénients sont également présentés.

I.2 Biométrie

La biométrie est la partie de la biologie qui analyse statistiquement les mesures faites sur l'homme [1]. Les systèmes biométriques électroniques sont relativement récents, mais la biométrie elle-même a toujours été utilisée par les êtres humains. Lorsque vous reconnaissez quelqu'un en voyant son visage- en personne ou sur une photo- c'est essentiellement la **biométrie** qui vous permet d'identifier cette personne.

Les techniques appliquant la biométrie visent à établir **l'identité d'une personne** en mesurant une de ses caractéristiques physiques, comportementales ou biologiques. Ces techniques biométriques permettent donc la mesure et la reconnaissance de **ce que l'on est**, à la différence d'autres techniques de mêmes finalités, mais permettant de mesurer ou vérifier **ce que l'on possède** (carte, badge, document,..) ou **ce que l'on sait** (mot de passe, code pin,..)[2][3]

I.3 Classification des techniques biométriques

On peut classer les techniques biométriques en trois catégories principales :

- Celles basées sur l'analyse des caractéristiques biologiques (odeur corporelle, salive, sang, ADN, ...), et
- Celles basées sur l'analyse des caractéristiques comportementales (dynamique du tracé de signature, frappe sur un clavier d'ordinateur,...), et
- Celles basées sur l'analyse des caractéristiques morphologiques (empreintes digitales, forme de la main, **traits du visage**, réseau veineux de la rétine, iris de l'œil,...).

Ces caractéristiques sont variables, les unes plus fiables que d'autres, mais toutes doivent être uniques, discriminatives, permanentes, collectables, mesurables et infalsifiables pour pouvoir être représentatives d'un seul et même individu [3].

I.4 Domaines d'application de la biométrie

L'industrie développe de plus en plus d'applications basées sur l'utilisation de la biométrie dans de nombreux domaines et à finalités diverses. L'intérêt de ces applications se résume généralement en trois objectifs :

- Faciliter le mode de vie,
- Eviter la fraude,
- Contrôler la population.

En effet, la liste des domaines utilisant la biométrie comme moyen de contrôle d'accès (physique ou logique) est très longue et elle n'est limitée que par l'imagination de chacun dans son domaine d'activité. Parmi ces domaines on trouve:

1. Contrôle d'accès aux zones surveillées telles que les aéroports et les sites sensibles (service de recherche, sites nucléaires,...),
2. Systèmes d'information: commerce électronique, transactions financière, et les logiciels utilisant des mots de passe,
3. Etat/Administration (fichier judiciaire, pièces d'identité (carte d'identité, passeport, permis de conduire),
4. Machines et équipements divers tels que les distributeurs automatiques de billets et les voitures (anti-démarrage).

I.5 Quelques caractéristiques biométriques

Il existe plusieurs caractéristiques physiques ou comportementales uniques pour identifier un individu [2] [4] [5]. En effet, on va présenter quelques unes les plus répandues ainsi que leurs domaines d'applications:

I.5.1 Empreintes digitales

L'empreinte digitale est la technique la plus ancienne et la plus répandue et qui a été adoptée par Scotland Yard dès 1901 et par la préfecture de police de Paris (1903) pour rechercher et identifier les criminels [1].



Figure I.1. Scanner d'empreintes digitales.

Crochets, îlots, interruptions, bifurcations, embranchements, chaque détail de votre empreinte digitale est analysé pour vous identifier. Ces particularités nommées minuties sont repérées et enregistrées par un logiciel dans une carte d'empreintes. Le nombre de minuties peut atteindre 80 pour des systèmes de haute sécurité. La formation des empreintes dépend des conditions initiales du développement embryogénique, ce qui les rend uniques à chaque personne et même à chaque doigt. L'image d'empreinte est prise soit selon le mode traditionnel qui est un scanning du doigt couvert d'encre soit en utilisant un dispositif d'acquisition d'image spécifique (live-scan fingerprint scanner, voir figure I.1).

L'attractivité majeure de la technique des empreintes digitales est son acceptance, son confort d'utilisation ainsi que sa fiabilité relativement élevée. Parmi les applications des systèmes biométriques utilisant les empreintes digitales, on cite :

- Sécurité réseaux (défense militaire, service de police),
- Contrôle d'accès physique et logique.

I.5.2 Géométrie de la main

Ce type de mesure biométrique est l'un des plus répandus, Cela consiste à mesurer plusieurs caractéristiques de la main (jusqu'à 90 caractéristiques) telles que sa forme tridimensionnelle, longueur et largeur des doigts, formes des articulations, longueurs inter articulations, etc. La technologie associée à cela est principalement de l'imagerie infrarouge.

D'une façon générale, cette technique est simple à mettre en œuvre, peu intrusive, et appréciée par les utilisateurs. Néanmoins, le système basé sur la géométrie de la main peut être facilement trompé par les personnes de la même famille ou bien encore des jumeaux. Elle est également limitée par les modifications de la forme de la main liées au vieillissement [4].

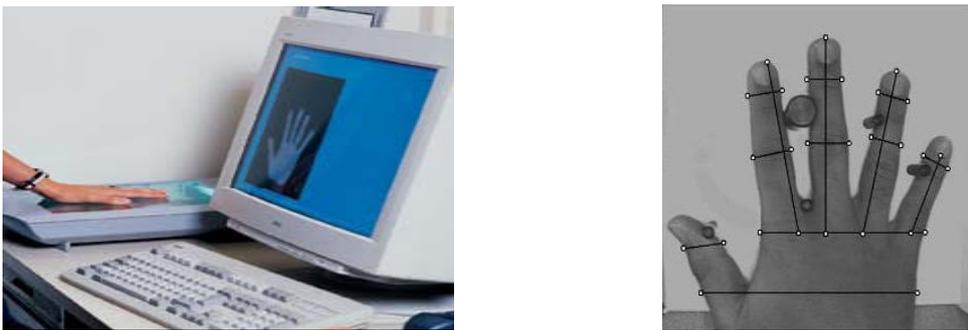


Figure I.2. Scanner de la main.

Cette technique biométrique est particulièrement appréciée dans les cantines scolaires, pointeuses dans les entreprises, Coca-Cola, Pfizer ou TWA, par exemple, l'utilisent déjà [6]. Au Musée de Louvre, l'accès à certaines salles est réservé pour le nettoyage. L'ouverture se fait par indentation palmaire.

I.5.3 Iris

L'iris est unique, très complexe et indépendant du code génétique. Une même personne a deux iris différents, deux vrais jumeaux en auront quatre différents. L'iris est la partie de l'œil entourant la pupille (Figure I.3 (a)). Un scanner d'iris (Figure I.3 (b)) crée tout d'abord une image très détaillée, avant que cette dernière ne subisse une analyse approfondie.

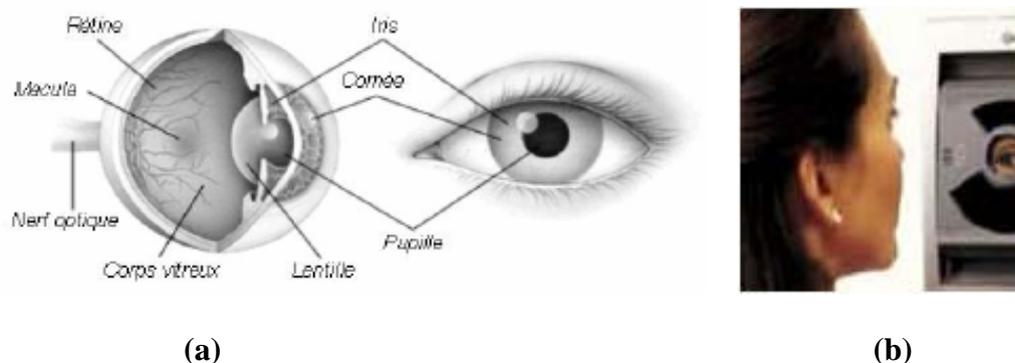


Figure I.3. (a) Anatomie de l'œil, (b) Scanner d'iris.

De multiples points de comparaison peuvent être relevés dans un iris comme par exemple les ligaments arches, les « furrows », les « contours », les « crypts », les « rings » de la corona, les « freckles », et les collerettes zigzag, ce qui en fait l'un des meilleurs systèmes d'identification. Les logiciels les plus performants peuvent même démasquer de faux yeux (image) ou une lentille.

Les applications de vérification de l'iris les plus répandues, sont implémentées dans les domaines de la sécurité aéronautique, l'accès automatique des frontières, l'accès à des bases de données à restrictions extrêmes, programmes gouvernementaux à profil de sécurité d'état et l'accès à des lieux restreint dans le monde des finances.

I.5.4 Rétine

La rétine est la paroi interne et opposée de l'œil sur laquelle se projettent les images que nous voyons. Les éléments qui permettent de distinguer deux rétines sont les veines qui les tapissent (Figure I.4). La disposition de ces veines est stable et unique d'un individu à l'autre (d'un œil à l'autre).

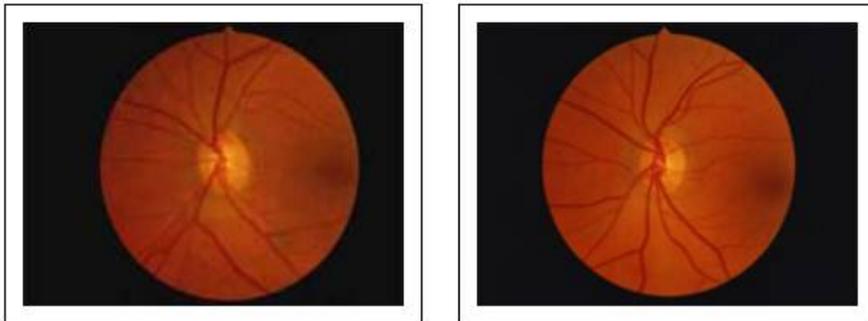


Figure I.4. Photographie des deux rétines d'un individu.

Afin d'obtenir une image d'une rétine, il est nécessaire d'éclairer le fond de l'œil à l'aide d'un faisceau lumineux à travers la pupille et le corps vitreux, pour cette raison, la biométrie par la rétine requiert une collaboration étroite de la part de l'utilisateur car il doit placer son œil extrêmement près de la caméra (quelques centimètres). La mesure peut ainsi fournir jusqu'à 400 points caractéristiques (rappelons qu'une empreinte digitale n'en compte que 30 à 40). La mesure rétinienne est la plus difficile à utiliser mais également la plus dure à contrefaire, ce qui réserve cette technique aux applications de haute sécurité (sites militaires, salles de coffres forts, etc.).

I. 5.5 Voix

L'identification par la voix est considérée par les utilisateurs comme une des formes les plus normales de la technologie biométrique, car elle n'est pas intrusive et n'exige aucun contact physique avec le lecteur du système. Les systèmes d'identification de la voix se concentrent sur les seules caractéristiques de voix qui sont uniques à la configuration de la parole d'un individu. Ces configurations de la parole sont constituées par une combinaison des facteurs comportementaux (vitesse, rythme, etc.), et physiologiques (tonalité, âge, sexe, fréquence, accent, harmoniques) [6].

On distingue les systèmes à texte prédéterminé (En anglais: dependent text), où l'utilisateur doit répéter un texte (prédéfini) qu'il ne choisit pas, et les systèmes où la personne peut parler librement (En anglais: independent text). Parmi les applications des systèmes biométriques utilisant l'analyse de la voix (aussi appelée analyse du locuteur), on trouve :

- Les centres d'appels téléphoniques,
- Les transactions bancaires,
- Accès à des comptes sur PC.

I.5.6 Signature

La vérification de la signature est la technologie de biométrie la plus confortable dans l'utilisation. La signature est une biométrie comportementale liée à la dynamique de l'individu, par conséquent beaucoup plus difficile à reproduire. Un dispositif comprenant une palette graphique et un stylo (Figure I.5) analyse plusieurs caractéristiques pendant l'apposition de la signature de la personne:

- La vitesse de déplacement du stylo,
- L'ordre des pressions exercées et des accélérations,
- L'inclinaison et les contours d'images.



Figure I.5. Palette graphique.

I.5.7 Visage

L'identification et/ou la vérification par le visage est la technique biométrique la plus simple et la moins contraignante puisqu'elle correspond à ce que les humains utilisent dans l'interaction visuelle. Cette technique présente l'avantage comparé à d'autres caractéristiques biométriques d'être non intrusive et acceptable par les utilisateurs.

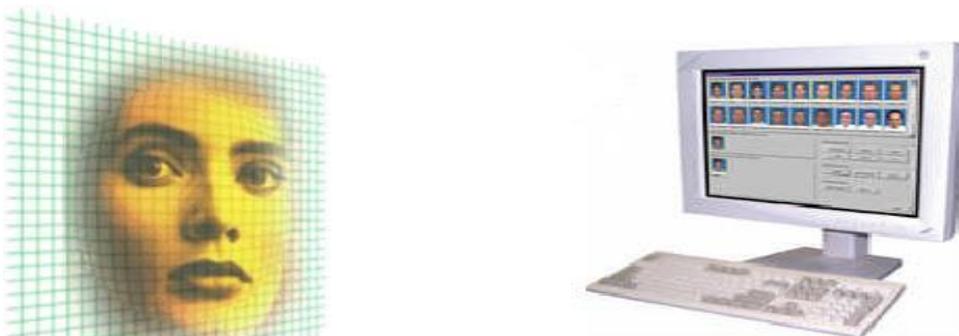


Figure I.6. Exemple d'identification de visage

Au début des années 1970, la reconnaissance par le visage était principalement basée sur des attributs faciaux mesurables comme l'écartement des yeux, des sourcils, des lèvres, la forme du menton, etc. Depuis les années 1990, les différentes technologies utilisées exploitent toutes les découvertes effectuées dans le domaine du traitement d'image [6]. En position fixe et à éclairage stable, les taux de reconnaissance sont effectivement très élevés, mais en vidéo, la technique n'est pas encore au point.

La technologie de la reconnaissance faciale est aujourd'hui en pratique dans plusieurs domaines:

- Contrôle d'accès hautement sécurisé,
- Sécurité dans les entreprises,
- Accès restreint,
- Sécurité des réseaux.

Chacune de ces techniques biométriques a ses propres avantages ainsi que ses inconvénients en termes de performance, coût, acceptation de la part des utilisateurs etc.

En particulier, le visage reste trop changeant et dépend fortement des conditions d'éclairage, de la position de la tête, d'expressions faciales, du temps et d'apparence comme le port de lunettes et le maquillage [3].

I.6 Principe de fonctionnement d'un système biométrique

Un système biométrique comporte généralement quatre modules comme le schématise la figure I.7. Il peut être appliqué à des fins d'identification (Figure I.8) ou d'authentification (vérification) (Figure I.9). Dans ce paragraphe, nous allons détailler les différentes étapes d'un système biométrique ainsi que les deux modes de fonctionnement.

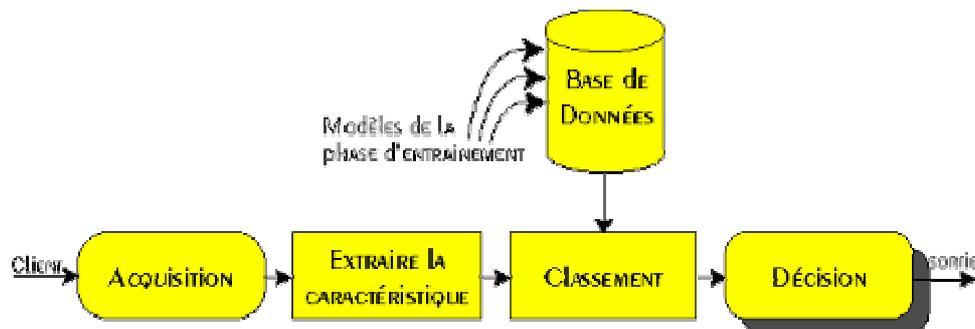


Figure I.7. Structure générale d'un système d'identification.

1. Acquisition

Un système d'acquisition équipé d'un capteur est utilisé pour acquérir la caractéristique biométrique de l'utilisateur, par exemple une caméra dans le cas d'une reconnaissance de visage ou un microphone dans le cas de la voix.

2. Extraction

Ayant des données biométriques en entrée, cette étape permet d'extraire les caractéristiques dont le processus de reconnaissance aura besoin. Ces caractéristiques forment les modèles des utilisateurs stockés dans une base de référence.

3. Classification

En examinant les modèles stockés dans la base de données, le système collecte un certain nombre de modèles qui ressemblent le plus à celui de la personne à identifier, et constitue une liste limitée de candidats. Cette classification intervient uniquement dans le cas d'identification car l'authentification ne retient qu'un seul modèle (i.e. celui de la personne proclamée).

4. Décision

Pour l'étape de décision, on distingue entre le mode d'identification et celui d'authentification.

Le processus d'identification (Figure I.8) consiste à comparer le modèle de l'utilisateur avec les modèles de tous les utilisateurs, déjà stockés dans une base de données. L'utilisateur inconnu est identifié comme l'utilisateur ayant le modèle qui ressemble le plus, selon un critère donné, au modèle d'entrée.

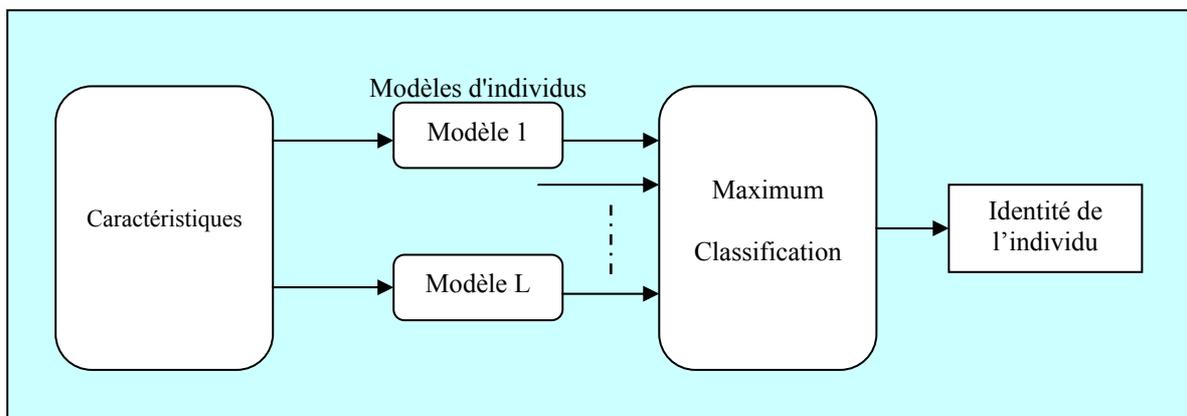


Figure I.8. Système d'identification de personnes.

Contrairement à l'identification, l'authentification (Figure I.9) consiste à comparer le modèle d'entrée avec seulement celui de l'identité proclamée. Ici, il s'agit de classer l'utilisateur comme un vrai utilisateur ou un imposteur.

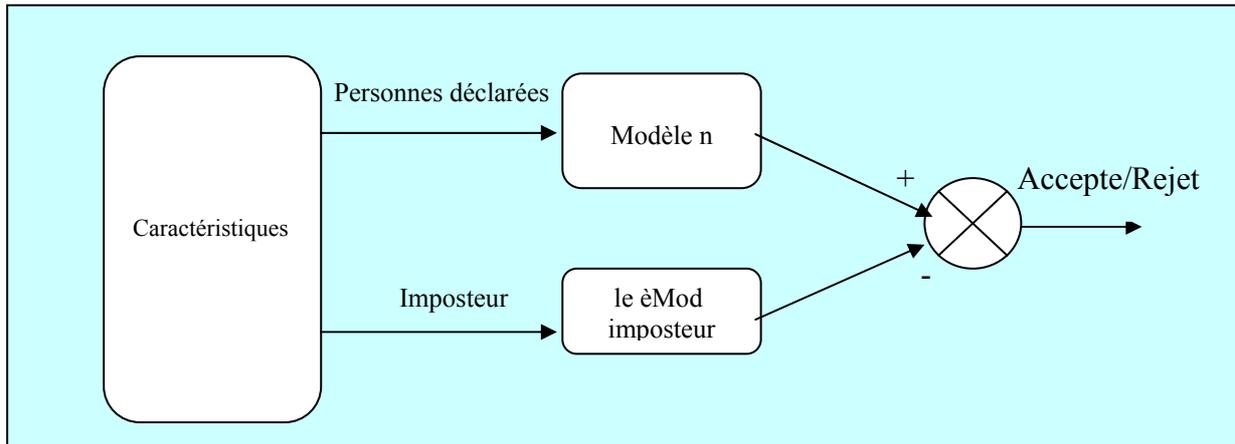


Figure I.9. Système d'authentification de personnes.

I.7 Evaluation des systèmes biométriques

Dans cette diversité de techniques biométriques, l'analyse comparative des performances des différentes techniques biométriques s'avère importante avant la prise de décision d'implantation d'un système biométrique qui sera le mieux adapté à la situation et aux finalités recherchées. Plusieurs études de ce type ont été menées afin de trouver les critères d'évaluation et de comparaison. Ainsi, **l'International Biometric Group**, une société américaine d'intégration et de conseil a établi un classement de huit techniques biométriques en se basant sur quatre critères d'évaluation (Figure I.10) [40]:

1. **l'effort** requis par l'utilisateur lors de la prise de mesures biométriques et qui doit être réduit le plus possible,
2. **l'intrusivité** doit être minimale, c'est le cas de la reconnaissance faciale qui est relativement bien acceptée par les utilisateurs par comparaison avec d'autres techniques biométriques comme les empreintes digitales où un contact direct est nécessaire entre le doigt de l'utilisateur et le capteur, cette obligation se retrouve également lorsqu'on utilise l'iris ou la rétine,
3. **le coût** de la technologie (lecteurs, capteurs, etc.) doit être modéré,
4. **la fiabilité** ou la performance qui caractérise l'efficacité de la technique biométrique à identifier les personnes.

Cette étude a mené aux résultats suivants :

- Des techniques les moins intrusives aux plus intrusives: la voix, la frappe sur le clavier, la signature, la main, le visage, l'empreinte digitale, l'iris, et enfin la rétine,
- Des techniques les plus fiables aux moins fiables: l'iris, la rétine, l'empreinte digitale, le visage, la main, la voix, et enfin à un niveau équivalent la frappe sur le clavier et la signature,
- Des techniques les moins chères aux plus chères: la frappe sur le clavier, la voix, la signature, l'empreinte digitale, le visage, la main, la rétine, et enfin l'iris,

- Des techniques les plus faciles point de vue utilisation aux plus difficiles: le visage, la signature, l'iris, la frappe sur le clavier, la voix, l'empreinte digitale, la main, et enfin la rétine.

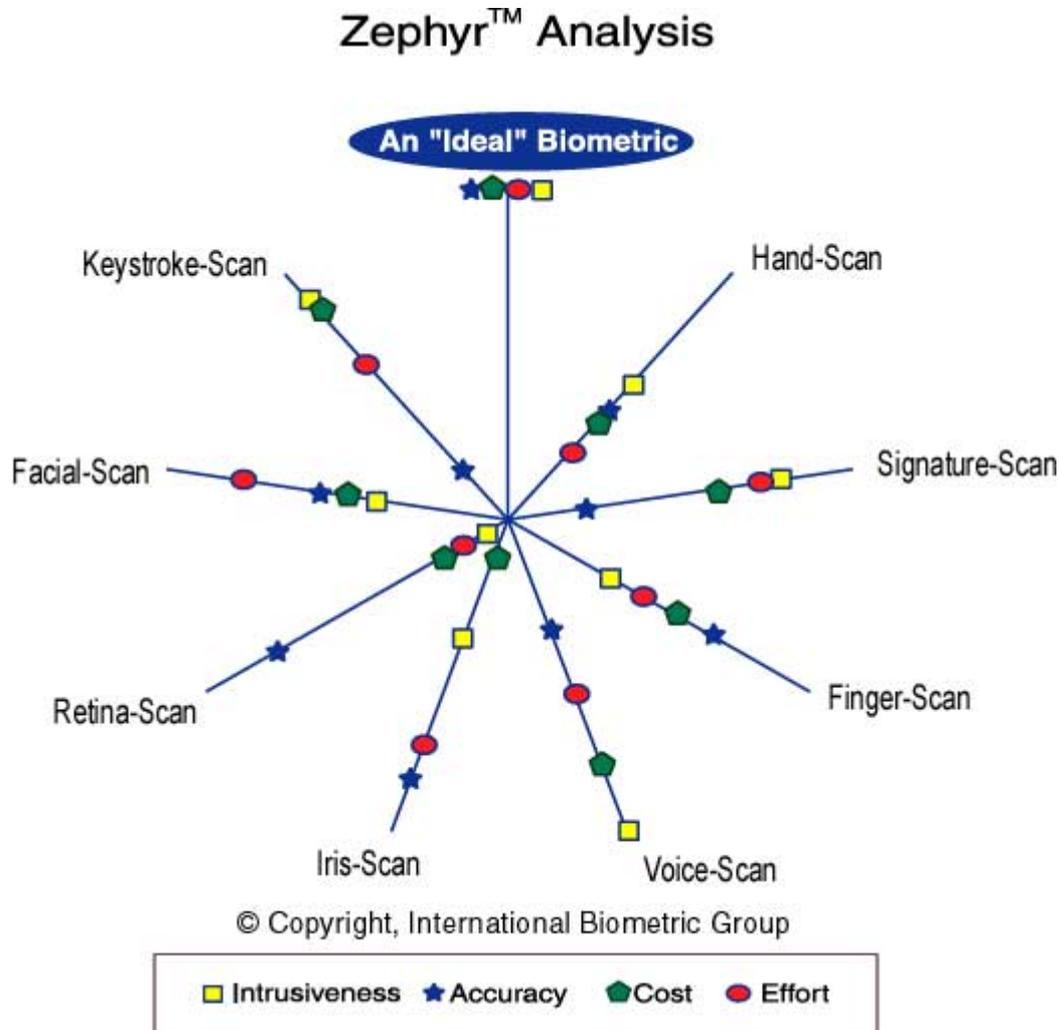


Figure I.10. Analyse Zephyr (source : International Biometric Group / IBG)

I.8 Performances des systèmes biométriques

Deux notions sont importantes pour évaluer les performances d'un système biométrique:

- **Le taux de fausses acceptations** (False Acceptance Rate: **FAR**) donné par l'équation (I.1) est le rapport entre le nombre de personnes qui ont été acceptées alors qu'elles n'auraient pas du l'être et le nombre total de personnes non autorisées qui ont tenter de se faire accepter,

- **Le taux de faux rejets** (False Reject Rate: **FRR**) donné par l'équation (I.2) est le rapport entre le nombre de personnes autorisées dont l'accès a été refusé et le nombre total de personnes autorisées s'étant présentées.

$$FAR = \frac{\text{Nombre de FA}}{\text{Nombre Im posteurs}} \dots\dots\dots \text{(I.1)}$$

$$FRR = \frac{\text{Nombre de FR}}{\text{Nombre Clients}} \dots\dots\dots \text{(I.2)}$$

$$TER = \frac{FR + FA}{\text{Nombre Total d' acces}} \dots\dots\dots \text{(I.3)}$$

Les deux types de taux sont reliés de façon inverse, l'équilibre entre ces deux taux qui permet d'obtenir les meilleurs résultats doit être adapté en fonction des besoins (taux de fausses acceptations très diminué pour un système nécessitant une haute sécurité avec conséquence inévitable un taux de faux rejets élevé). La figure I.11 illustre cet équilibre [40].

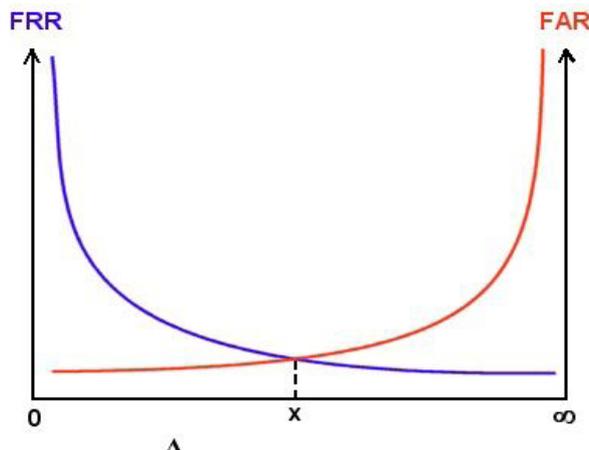


Figure I.11. Le point x est le compromis entre les taux FRR et FAR.

I.9 Conclusion

Dans un futur proche, les systèmes biométriques vont peu à peu remplacer l'utilisation de mots de passe, de badges, voire de clés qui sont utilisés actuellement pour les ordinateurs, les voitures et les accès contrôlés à des bâtiments ou à Internet. Les systèmes qui rencontreront le plus de succès seront ceux qui offriront l'interface la plus simple et la moins contraignante à l'utilisateur, tout en garantissant un bon niveau de sécurité. Néanmoins, la finalité des utilisations de la biométrie doit être définie avec précision afin d'éviter la collecte et le traitement de plus de données biométriques qu'il n'en faut. Il y a des règles à respecter, des méthodes à définir et des contrôles à effectuer régulièrement. Parmi les techniques biométriques existantes, nous nous intéressons à la reconnaissance de visages qui sera discutée dans le chapitre suivant.

Chapitre II

RECONNAISSANCE DE VISAGES

II.1 Introduction

Si pour un être humain, reconnaître un visage (identifier un individu) relève d'une action réflexe, aussi instantanée que "naturelle", il en va tout autrement pour un système automatique. Pour une machine, une telle opération suppose au contraire, une chaîne de traitements complexes reposant sur des principes mathématiques avancés.

Les premiers travaux concernant la reconnaissance de visages remontent au début des années 1970. Les techniques introduites à l'époque utilisaient pour la plupart des mesures estimées autour des éléments faciaux des visages [13] [14]. Mais ce n'est qu'au début des années 1990 que le volume de recherche concernant la reconnaissance de visages a réellement commencé à croître. Par la suite, l'engouement pour cette problématique n'a fait qu'augmenter, si bien qu'aujourd'hui la reconnaissance de visages constitue l'une des applications les plus explorées de la reconnaissance de formes et de l'analyse d'images. Cet intérêt croissant s'est soldé par l'apparition de conférences internationales sur le sujet, telles que l'International Conference on Automatic Face and Gesture Recognition (AFGR) en 1995 et l'International Conference on Audio and Video Based Authentication (AVBPA) en 1997 [22].

Ainsi, et vu les enjeux croissants notamment dans les domaines d'indexation et de la sécurité, la technique de la reconnaissance de visages continue à attirer de nombreux chercheurs issus de différentes disciplines telles que la neurophysiologie, la neuropsychologie, le traitement d'images, la reconnaissance de formes et la vision par ordinateur [37].

II.2 Reconnaissance de visages

Les visages constituent une catégorie de stimulus unique par la richesse des informations qu'ils véhiculent. Ils sont à la fois les vecteurs visuels principaux de l'identité individuelle, et des vecteurs essentiels de communication (verbale et non verbale) d'intentions et d'émotions entre individus, via, en particulier, la direction du regard et les expressions faciales. Sur le plan perceptif, la reconnaissance des visages constitue une des capacités les plus étonnantes, et les plus performantes du système visuel humain. En effet, l'être humain est capable de discriminer et de reconnaître un très grand nombre de visages, alors même que les visages constituent une catégorie de stimulus homogènes, partageant un ensemble de traits similaires et une structure commune, et qu'ils peuvent subir d'importants changements avec l'âge par exemple, ou encore selon l'expression faciale [7]

Le but de la reconnaissance des visages est de concevoir des systèmes informatiques capables de copier les prodigieuses facultés de reconnaissance du cerveau humain. Cependant, et malgré l'avancement de la recherche réalisé durant ces trois dernières décennies, la reconnaissance robuste de visages reste très difficile comme le montre la récente évaluation menée par le National Institut of Standards and Technology (NIST) [8]. Les difficultés proviennent essentiellement de la forte variabilité des données à classer (dues à des changements dans les expressions faciales, dans la pose de la tête, dans les conditions d'illumination et à des occultations partielles), et dans le fait que les images de visages constituent des données de très grandes dimensions. Néanmoins, et malgré toutes ces difficultés, le domaine de la reconnaissance faciale demeure important pour de nombreux chercheurs vu la multitude des champs d'applications. Cette variabilité des applications a donné lieu à la disponibilité de plusieurs produits commerciaux permettant:

- L'identification, par comparaison d'un visage à ceux mémorisés dans une base de données,
- La vérification, par comparaison des identités déclarées avec les identités associées aux visages mémorisés,
- La supervision, qui permet de suivre l'image d'une personne dans une séquence vidéo, [5]
- La surveillance qui permet de retrouver, en temps réel, une personne dans une séquence vidéo à partir d'une liste de visages.

II.3 Processus de reconnaissance des visages

II.3.1 Images Fixes

La plus part des algorithmes de reconnaissance automatique de visages portent sur la classification d'images fixes 2D notamment en niveau de gris. Les algorithmes prenant en compte la couleur ne sont, pour la plus part, que des généralisations à trois canaux des techniques utilisées en niveaux de gris [22].

On dispose d'une base de connaissance contenant un ensemble d'images de visages de personnes connues, c'est-à-dire les personnes que le système est censé reconnaître. Cette base de connaissance peut contenir plusieurs images d'une même personne sous des conditions de prises de vue différentes. Chaque image est étiquetée par son identité associée. Lorsqu'un visage requête (image d'un visage à reconnaître) est présenté au système, celui-ci va chercher à lui assigner une identité parmi celles contenues dans la base de connaissance. Le système de reconnaissance est basé sur un modèle généralement construit à partir d'une base d'apprentissage contenant un ensemble d'images de visages par le biais d'un algorithme d'apprentissage.

Dans certains cas, la base d'apprentissage et la base de connaissance sont confondues. Le modèle est alors spécifiquement conçu pour les visages qu'il vise à reconnaître. Dans d'autres cas, au contraire, les deux bases sont distinctes soit à cause de l'application à réaliser soit à cause des spécificités de la technique utilisée [22].

II.3.2 Images Vidéo

Ces dernières années, la reconnaissance de visages depuis des images vidéo commence à être explorée [17], mais utilisant toujours les mêmes techniques conçues dans le contexte d'images fixes. Trois méthodes de classification sont utilisées à savoir :

1. Méthode qui consiste à sélectionner une image clé représentative de l'ensemble des vues du visage requête (Pour la vidéo, on dispose d'un nombre important de vues) [22],
2. Méthode qui utilise un algorithme de vote sur les résultats de classification obtenus pour chacune de ces vues [50],
3. Méthode qui construit un modèle spécifique pour chaque personne de la base de connaissance ainsi que pour le visage requête à partir de l'ensemble des vues de la personne considérée dont on dispose. Les caractéristiques du modèle du visage requête sont alors comparées à celles des modèles de la base de connaissance afin de lui assigner l'identité du modèle le plus proche [71].

II.3.3 Images 3D

Très récemment, les avancées dans le domaine de l'acquisition de données tridimensionnelles (acquisition par scanners laser) ont permis l'émergence de technologies de reconnaissance 3D. Deux grandes familles de méthodes sont apparues ces dernières années [22]:

1. Celles basées sur la correspondance d'objets 3D directement (on dispose d'images 3D aussi bien dans la base d'apprentissage que pour les visages requêtes) [42], et
2. Celles se ramenant pour la classification à un problème de reconnaissance 2D (seules les vues 3D de la base d'apprentissage sont requises) [72].

L'inconvénient de ces deux méthodes c'est qu'elles sont intrusives en utilisant un capteur 3D.

II. 4 Différentes étapes d'un processus de reconnaissance de visages

Après avoir donné un aperçu sur les différents types d'images de visages utilisées dans le domaine de la reconnaissance automatique de visages, ainsi que les algorithmes développés, nous allons nous limiter, dans notre travail, aux images vidéo en niveaux de gris. Comme toute tâche de reconnaissance de formes, le processus de reconnaissance automatique de visages se décompose en deux étapes principales [22] :

1. l'extraction de caractéristiques (appelés signatures), et
2. la classification de ces caractéristiques.

Pour cela, le système doit prendre en compte plusieurs facteurs qui contribuent à la complexité de sa tâche, car le visage est une entité dynamique qui change constamment sous l'influence de plusieurs facteurs [9] [10]. La figure II.1 illustre la démarche générale adoptée pour réaliser de tels systèmes. Les paragraphes suivants détaillent les différentes étapes de ce système:

Dans le monde physique, il y'a trois paramètres à considérer: l'éclairage, la variation de posture, et l'échelle. La variation de l'un de ces trois paramètres peut conduire à une différence entre deux images du même individu, dépassant celle séparant deux images de deux individus différents.

II.4.1 Le codage

L'acquisition et le codage des images s'effectuent en utilisant des dispositifs d'acquisition et de numérisation tels que :

- Une caméra,
- Un appareil photos numérique,
- Un fichier de sauvegarde dans une mémoire de masse,
- Un scanner.

Cependant, cette étape retourne des images en niveaux de gris bruitées.

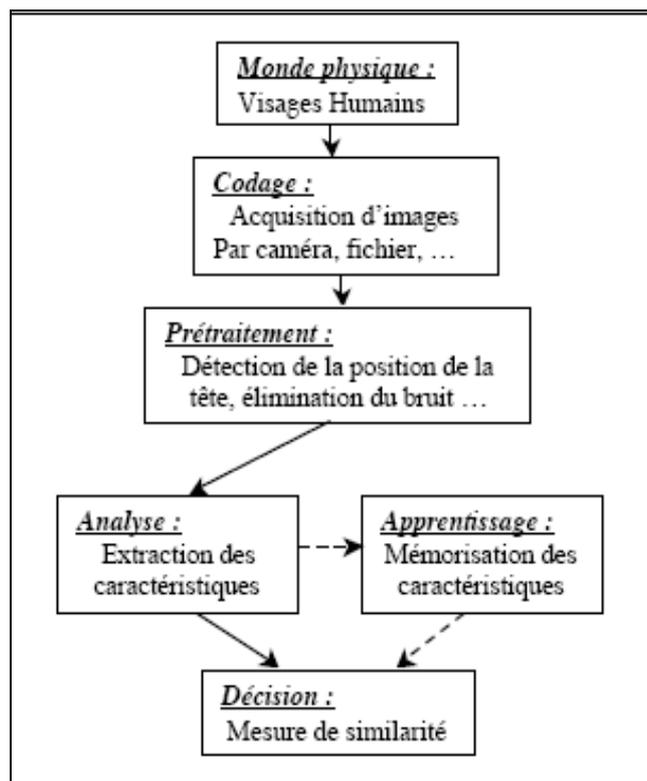


Figure II.1. Extrait de [10] Schéma général d'un système de reconnaissance de visages.

II.4.2 Le prétraitement

Cette étape consiste à éliminer les bruits résiduels issus de la digitalisation par des techniques de traitement et de restauration d'images. La restauration d'images ou l'élimination du bruit consiste à compenser les dégradations connues ou estimées et rétablir la qualité initiale de l'image [11] [12]. Le prétraitement peut concerner également la localisation de la tête dans l'image.

II.4.3 L'analyse

Dans une image, une quantité gigantesque d'informations peut être extraite, mais ces informations ne sont pas toutes pertinentes, et il est important de n'utiliser que les informations représentatives, résumant ainsi le contenu de l'image. L'opération qui retourne ce type d'informations est dite « **extraction de caractéristiques** ». L'analyse (appelée aussi indexation, représentation, modélisation ou extraction de caractéristiques) revient donc à établir un modèle pour le visage qui sera sauvegardé en mémoire pour être utilisé plus tard dans la phase « décision ».

II.4.4 L'apprentissage

L'apprentissage consiste à mémoriser les modèles de personnes obtenus dans la phase « Analyse » pour les individus connus.

II.4.5 La décision

Pour estimer la différence entre deux images de visages, il faut introduire une mesure de similarité telle que la distance euclidienne.

II.5 Extraction des caractéristiques du visage

L'extraction de caractéristiques faciales est une étape primordiale dans plusieurs applications telles que la reconnaissance de visages, et la reconnaissance d'expressions faciales. Cette étape peut se faire manuellement [25] ou automatiquement à partir des images de vue de face [19], (figure II.2 (a)), ou de vue orientée [26] [27] (figure II.2 (b)).



Figure II.2. Exemples de localisation de caractéristiques faciales et de distances géométriques: (a) vue de face et (b) vue de profile.

Deux types de caractéristiques faciales peuvent être extraites: caractéristiques géométriques et caractéristiques d'apparence [41]. Les caractéristiques géométriques présentent les formes et les positions des différents éléments du visage (y compris les yeux, le nez, la bouche, etc.) (Voir figure II.2). Quant aux caractéristiques d'apparence (ou d'aspect), elles présentent les changements d'apparences (texture de la peau) du visage tels que les rides et les sillons. Ces caractéristiques peuvent être extraites à partir du visage entier ou des régions spécifiques du visage.

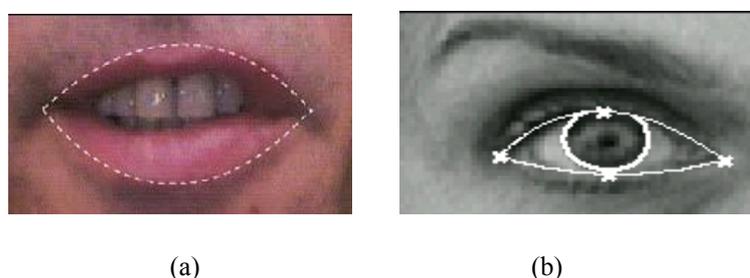


Figure II.3. Exemples d'extraction de contours à base de modèles: (a) à deux paraboles, (b) à une courbe de Bézier pour le contour supérieur et une parabole pour le contour inférieur.

Dans tous les cas, un algorithme approprié d'extraction ou de localisation d'éléments faciaux basé sur l'information contenue dans les pixels (intensité, couleur, fréquence) [22] renvoie des points caractéristiques qui seront par la suite utilisés dans les différentes applications (reconnaissance de visages, détection et suivi des visages ou reconnaissance des expressions faciales). On peut distinguer trois types de méthodes d'extraction d'éléments faciaux [17]:

II.5.1 Méthodes génériques (Generic Method)

Elles sont basées sur les lignes (curves) et les contours (edges), comme celle proposée par Craw et al. [31] et sur les contours actifs [19] [32] dont l'inconvénient majeur est l'initialisation des courbes, et le réglage de nombreux paramètres.

II.5.2 Méthodes basées sur les modèles de caractéristiques (Feature Template-based Methods)

Consistent à assigner des modèles prédéfinis ou déformables [32] aux différents attributs faciaux. Hallinan [17] et Funabiki et al. [33] ont proposé des approches qui détectent les yeux, Yuille et al. [32] ont utilisé deux modèles pour la détection des yeux et la bouche respectivement. Ces approches trouvent des difficultés lorsque les apparences des caractéristiques changent de manière significative (yeux fermés, yeux avec lunettes, bouche ouverte) [62].

II.5.3 Méthodes basées sur la similitude structurelle (Structural Matching)

Afin de mieux détecter les différentes caractéristiques du visage, les approches récentes ont utilisé les méthodes basées sur la similitude structurelle. Parmi ces méthodes, on trouve le modèle actif de forme (Active Shape Model) de Cootes et al. [17] [35] qui offre plus de flexibilité et de robustesse. Afin de prendre en considération les variations de texture, le modèle actif de forme (ASM) s'est développé au modèle statistique d'apparence donnant lieu au modèle flexible d'apparence proposé par Lanitis et al., et le modèle actif d'apparence (Active Appearance Model AAM) de Cootes et al. [17] [36].

II.6 Approches existantes de reconnaissance de visages

La reconnaissance automatique de visage humain basée sur le traitement des images 2D s'est bien développée dans ces dernières années, et plusieurs approches ont été proposées. Ces approches sont regroupées en trois catégories principales à savoir [15] [22] :

- **Les approches globales** appelées approches pixels au sens où les caractéristiques sont directement extraites depuis la totalité des pixels de l'image du visage. Ces méthodes traitent le visage comme un tout,
- **Les approches locales** basées sur l'étude de caractéristiques extraites localement de différentes régions du visage (yeux, bouche, nez, etc.). On les appelle aussi les méthodes à traits, à caractéristiques locales ou géométriques,
- **Les approches hybrides**: Cette catégorie utilise conjointement les deux catégories citées auparavant (globale et locale). Les méthodes hybrides proches du fonctionnement du système visuel humain sont généralement très performantes et plus robustes à des changements d'apparence du visage dus par exemple à des variations dans l'expression faciale

que les techniques globales. Néanmoins, elles sont plus coûteuses en temps de calcul, pour la phase d'apprentissage comme pour la classification [22].

II.6.1 Approches globales

Les approches globales ont la particularité d'utiliser la région entière du visage comme entrée au système de reconnaissance. En introduisant les « Eigen pictures », Sirovitch et Kirby (1987) ont donné une nouvelle impulsion à la reconnaissance automatique de visages, en démontrant que l'Analyse en Composantes Principales (ACP) constitue un outil efficace pour la représentation des visages [17] [22].

Ces approches, qui s'appuient sur des propriétés statistiques et sur l'algèbre linéaire sont relativement rapides à mettre en œuvre, cependant leur inconvénient majeur est qu'elles sont sensibles aux problèmes d'éclairage, de pose, et d'expressions faciales [16].

L'algorithme le plus populaire utilisant une approche globale est sans aucun doute celui des « eigenfaces ». Cet algorithme, basé sur l'analyse en composantes principales, est né des travaux de Turk et Pentland au MIT Media Lab. [16]. L'eigenfaces est à la base de nombreux algorithmes actuels qui ont, depuis apporté des améliorations ou quelques variantes [16] [17], c'est pourquoi, il sera détaillé dans les paragraphes qui suivent.

Un autre algorithme aussi important que eigenfaces est « Fisherfaces » à base de l'Analyse Discriminante Linéaire (ADL). Il est né des travaux de Belhumeur et al [16][17]. Contrairement à l'ACP, l'ADL effectue une véritable séparation de classes, selon un critère mathématique qui minimise les variations entre les images d'un même individu (variations intra-classe) tout en maximisant les variations entre les images d'individus différents (variations inter-classes).

De nombreux autres algorithmes appartenant à cette catégorie ont été développés dont on peut citer [17] :

- Méthodes d'Analyse en Composantes Indépendantes (**ICA**) qui peuvent fournir des taux de reconnaissance plus élevés que le l'ACP,
- Feature Line Based Methods,
- Linear/Fisher Discriminant Analysis,
- Méthodes Bayésienne qui utilisent les distances métriques probabilistes.

Eigenfaces (Visages Propres)

Eigenfaces vise à capturer la variation dans une collection d'images de visages et d'utiliser cette information pour coder et comparer les visages [16]. En termes mathématiques, cela revient à trouver les vecteurs propres de la matrice de covariance formée par les différentes images de notre base d'apprentissage.

- ❖ Soit une image de visage $I_i(m \times n)$, elle est traitée comme un vecteur $\Gamma_{i(m \times n, 1)}$ dans un espace vectoriel de grande dimension ($N = mxn$), par concaténation des colonnes (ou des lignes).

$$I_1 = \begin{pmatrix} a_{1,1} & \dots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,m} \end{pmatrix} \Rightarrow \Gamma_1 = \begin{pmatrix} a_{1,1} \\ \vdots \\ a_{n,1} \\ \vdots \\ a_{1,m} \\ \vdots \\ a_{n,m} \end{pmatrix} \dots \dots \dots \text{(II.1)}$$

Les coefficients $a_{i,j}$ représentent les valeurs des pixels en niveau de gris codés de 0 à 255.

- ❖ Après avoir rassemblé les M images dans une unique matrice, nous obtenons une matrice d'images Γ , où chaque colonne représente une image Γ_i :

$$\Gamma = \begin{pmatrix} a_{1,1} & b_{1,1} & \dots & z_{1,1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n,1} & b_{n,1} & \dots & z_{n,1} \\ \vdots & \vdots & \dots & \vdots \\ a_{1,m} & b_{1,m} & \dots & z_{1,m} \\ \vdots & \vdots & \dots & \vdots \\ a_{n,m} & b_{n,m} & \dots & z_{n,m} \end{pmatrix} \dots \dots \dots \text{(II.2)}$$

- ❖ On calcule ensuite l'image moyenne Ψ de toutes les images collectées. Cette image peut être vue comme le centre de gravité du jeu d'images.

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \dots \dots \dots \text{(II.3)}$$

- ❖ On ajuste ensuite les données par rapport à la moyenne. L'image moyenne est alors soustraite de chaque image avec la formule suivante :

$$\Phi_i = \Gamma_i - \Psi, \quad i = 1 \dots M \dots \dots \dots \text{(II.4)}$$

- ❖ On calcule ensuite la matrice de covariance du jeu de données. Cette matrice peut être vue comme une matrice de moment d'ordre 2 :

$$C = \sum_{i=1}^M \Phi_i \Phi_i^T = AA^T, \quad A = [\Phi_1 \quad \Phi_2 \quad \dots \quad \Phi_M] \dots \dots \dots \text{(II.5)}$$

- ❖ Cette étape consiste à calculer les vecteurs propres et les valeurs propres de la matrice de covariance C de taille $(N \times N)$, c'est-à-dire de l'ordre de la résolution de l'image. Il est à remarquer que si $N > M$ (si la résolution est supérieure au nombre d'images) il y'aura seulement $(M-1)$ vecteurs propres qui contiendront de l'information (les vecteurs propres restants auront des valeurs propres associées nulles).

❖ Soient les vecteurs propres e_i de $C = AA^T$ associées aux valeurs propres λ_i on a :

$$Ce_i = \lambda_i e_i \quad \dots\dots\dots \text{(II.6)}$$

Les valeurs propres v_i de $L = A^T A$, associées aux valeurs propres μ_i sont tels que :

$$Lv_i = \mu_i v_i \quad \dots\dots\dots \text{(II.7)}$$

Soit :

$$A^T Av_i = A\mu_i v_i \quad \dots\dots\dots \text{(II.8)}$$

En multipliant à gauche par A des deux côtés de l'égalité, nous obtenons :

$$A A^T Av_i = A\mu_i v_i \quad \dots\dots\dots \text{(II.9)}$$

Puisque $C = AA^T$, nous pouvons simplifier :

$$C(Av_i) = \mu_i (Av_i) \quad \dots\dots\dots \text{(II.10)}$$

de (II.6) et (II.10) nous voyons que Av_i et μ_i sont respectivement les vecteurs propres et les valeurs propres de C .

$$\begin{cases} e_i = Av_i \\ \lambda_i = \mu_i \end{cases} \quad \dots\dots\dots \text{(II.11)}$$

Nous pouvons donc trouver les valeurs propres de cette énorme matrice C en trouvant les valeurs propres d'une matrice L beaucoup plus petite.
 Pour trouver les vecteurs propres de C , il suffit de multiplier les vecteurs propres de L par la matrice A .

- ❖ Les vecteurs propres trouvés sont ensuite ordonnés selon leurs valeurs propres correspondantes, de manière décroissante (plus une valeur propre est grande plus la variance capturée par le vecteur propre est importante. Cela implique que la majeure partie des informations est contenue dans les premiers vecteurs propres),
- ❖ Retenir k vecteurs propres correspondants à k valeurs propres les plus grandes,
- ❖ On définit ainsi un espace vectoriel engendré par ces k vecteurs propres, que l'on appelle **l'espace des visages** E_v

Les images originales peuvent alors être reconstituées par combinaison linéaire de ces vecteurs propres. Les représentations graphiques de ces vecteurs propres nous donnent les **visages propres (eigenfaces)** schématisés dans la figure II.4.

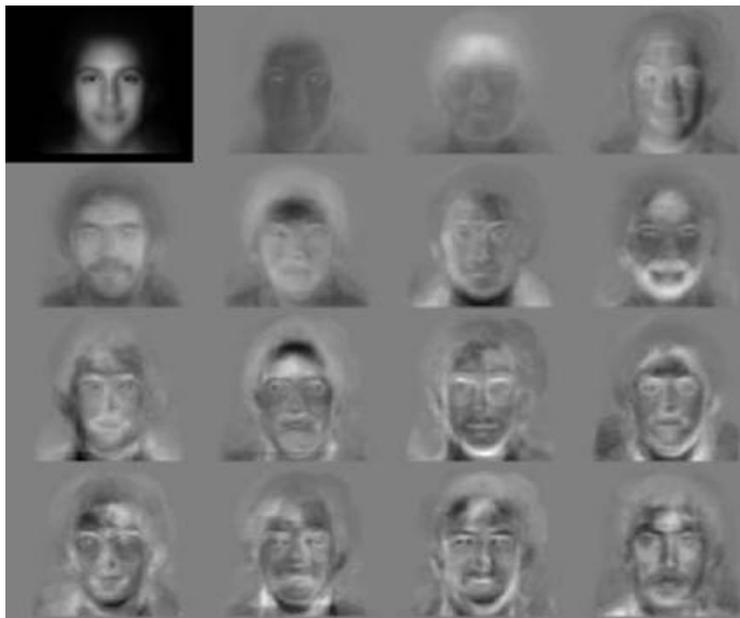


Figure II.4. Extrait de [16]. Image moyenne et les 15 premiers Eigenfaces.

- ❖ Nous allons maintenant projeter nos images de départ sur l'espace de visages E_v . Une image Γ_i est alors transformée en ses composantes Eigenfaces par une simple opération de projection vectorielle :

$$\omega_k = e_k^T (\Gamma_i - \Psi), \quad k = 1, \dots, M' \quad \dots \dots \dots \quad \text{(II.12)}$$

Les vecteurs ω_k sont appelés poids et forment une matrice $\Omega^T = [\omega_1, \omega_2, \dots, \omega_M]$ qui décrit la contribution de chaque eigenface dans la représentation de l'image d'entrée.

Sélection des vecteurs propres à retenir

Un point important et qui reste une voie de recherche est le choix du paramètre g . celui-ci détermine la dimensionnalité intrinsèque de l'espace des visages (nombre de vecteurs propres associés aux plus grandes valeurs propres [22] [62]).

Plusieurs solutions ont été proposées afin d'optimiser ce paramètre g :

1. Etudier le spectre des valeurs propres λ_i par le biais d'un graphe de l'éboullis des valeurs propres, et chercher la valeur charnière à partir de laquelle les valeurs propres (normalisées) sont très petites. (figure II.5).
2. Turk et Pentland [43] ont déterminé g d'une manière heuristique à partir de l'étude des valeurs propres. Pour leurs tests sur une base de 16 individus $k = 7$ valeurs propres,
3. Moon et Phillips [44] préconisent d'éliminer 40% des derniers vecteurs propres.
4. Zhao et al. [62] ont retenu 300 vecteurs propres pour une base de 1038 images après avoir observé que pour un nombre très élevé, les eigenfaces ne représentent pas des visages,

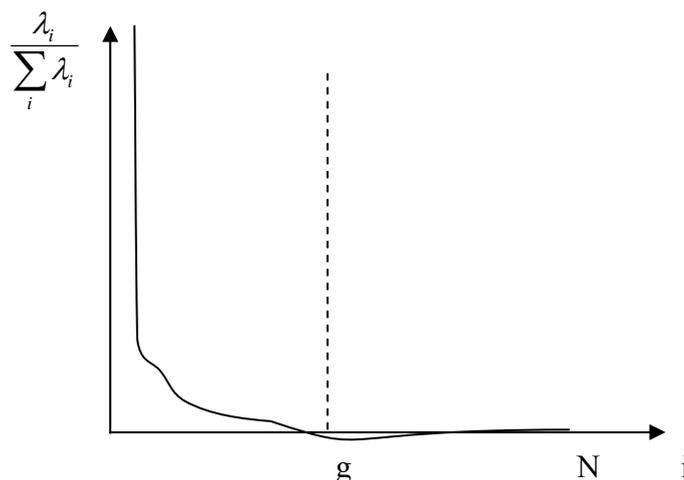


Figure II.5. Allure typique du graphe d'éboulis des valeurs propres associées à des Eigenfaces. Les valeurs propres λ_i sont rangées par ordre décroissant.

5. Swets et Weng [45] préconisent l'utilisation d'un taux $\tau = 5 \%$. Quant à Kirby [46] il a choisi $\tau = 10 \%$,
6. Kirby [46] a introduit un critère d'étirement défini comme le ratio $S_i = \frac{\lambda_i}{\lambda_1}$ entre la valeur propre de w_i et la plus grande valeur propre λ_1 . Seuls les vecteurs propres dont les S_i sont supérieurs à un certain seuil τ sont retenus (souvent $\tau = 1 \%$).

Choix de la mesure de dissimilarité

La classification des signatures fournies par la technique des eigenfaces est généralement menée à l'aide d'une distance au plus proche voisin: le visage requête est affecté à la classe d'appartenance dont la signature est la plus proche. Parmi les distances les plus utilisées on trouve la distance Manhattan, Euclidienne, Mahalanobis et la similarité cosinus [16] [22].

En conclusion, eigenfaces présente les avantages suivants [62]:

- Facile à mettre en œuvre,
- Ne nécessite aucune connaissance a priori sur l'image d'entrée,
- C'est un algorithme de réduction de dimensions où il peut alors être utilisé en amont d'autres algorithmes.

Cet algorithme souffre néanmoins de quelques problèmes :

- Les trois premières composantes principales des eigenfaces retiennent les variations dues à l'éclairage, et aux expressions faciales [47],
- Eigenfaces est plus adapté à la représentation de visages qu'à leur classification,
- Eigenfaces dépend du nombre de vecteurs propres retenus.

II.6.2 Approches locales

On les appelle aussi les méthodes à traits, à caractéristiques locales, ou analytiques. L'analyse du visage humain est donnée par la description individuelle de ses parties comme les yeux, le nez, la bouche, etc., et de leurs relations (longueur, largeur, distances, rapports de distances, angles,..), et ceci à partir des vues de faces [19], de profil [20], ou les deux [21].

Les méthodes locales peuvent être réparties en deux catégories principales [15] [22] :

II.6.2.1 Approches basées sur les éléments caractéristiques du visage (Local Feature-Based Method)

a. Approches géométriques

Ces approches font partie des plus anciennes techniques utilisées dans le cadre de la reconnaissance des visages. Elles consistent à extraire entre autres paramètres, les positions relatives des caractéristiques faciales telles que les yeux, le nez, la bouche, etc.

Parmi les travaux réalisés, on trouve **Kelly** [17][48] qui a utilisé comme éléments faciaux, parmi d'autres, la largeur de la tête, la distance entre les yeux et les distances entre les yeux et la bouche. Le système de **Kanade** (1973) qui portait sur l'utilisation des angles et distances entre les coins des yeux, les commissures de la bouche, les narines et l'extrémité du menton. **Brunneli et Poggio** [15] [22], ont utilisé un ensemble de trente cinq (35) éléments géométriques extraits automatiquement pour la représentation du visage. A l'aide du classificateur de Bayes, ils arrivaient à réaliser un taux de reconnaissance de 90 % sur une base contenant 188 images de 47 personnes [15]. **Cox et al.** (1996) ont proposé une approche basée sur un mélange de distances (mixture-distance). Ils avaient extrait manuellement 30 distances de 685 individus et arrivaient à un taux de reconnaissance de 95 %. Récemment, dans [49], une approche géométrique utilisant 8 distances comme entrée d'un perceptron multicouche a été proposée et testée sur une base de 40 personnes. Le taux de reconnaissance était 94.50%.

b. Approches basées sur les graphes

Ces techniques sont basées sur la méthodologie d'Architecture de Lien Dynamique (Dynamic Link Architecture: **DLA**). A partir d'une image de visage, on localise des points caractéristiques (coins des yeux, centre de la bouche, nez, etc.). Cette localisation peut se faire manuellement ou automatiquement à l'aide d'un algorithme.

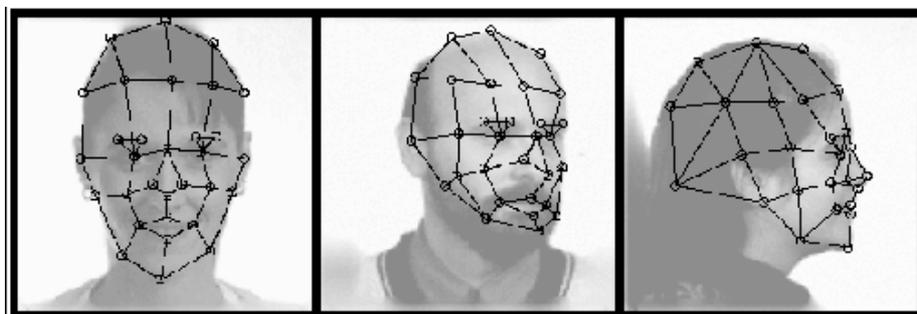


Figure II.6. Création du treillis.

Un treillis élastique virtuel est ensuite appliqué sur l'image de visage à partir de ces points (figure II.6). Chaque point caractéristique représente un nœud labellisé auquel on associe un jeu de coefficients d'ondelettes complexes de Gabor appelés **Jet** [16]. Pour effectuer une reconnaissance avec une image test, on fait une mesure de similarité entre les différents Jets.

Parmi les systèmes les plus réussis dans cette catégorie, on trouve le système **Elastic Graph Matching** (EGM) [16] [17] [22], où à chaque image de la base d'apprentissage est associé un graphe qui lui est propre (Figure II.7 (a)).

Pour pallier au problème de variation de pose, Okada et al. et Wiskott et al. ont proposé un système **Elastic Bunch Graph Matching** (EBGM). La différence avec Elastic Graph Matching (EGM) est que EBGM utilise un même graphe pour la modélisation de tous les visages d'apprentissage. Au lieu de construire un modèle pour chaque image, on construit donc un modèle général de représentation appelé **Face Bunch Graph** (FBG) depuis l'intégralité de la base d'apprentissage (Figure II.7 (b)). Tous les vecteurs correspondants à un même nœud sont regroupés de manière à représenter l'ensemble des états possibles de ce nœud (point caractéristique).

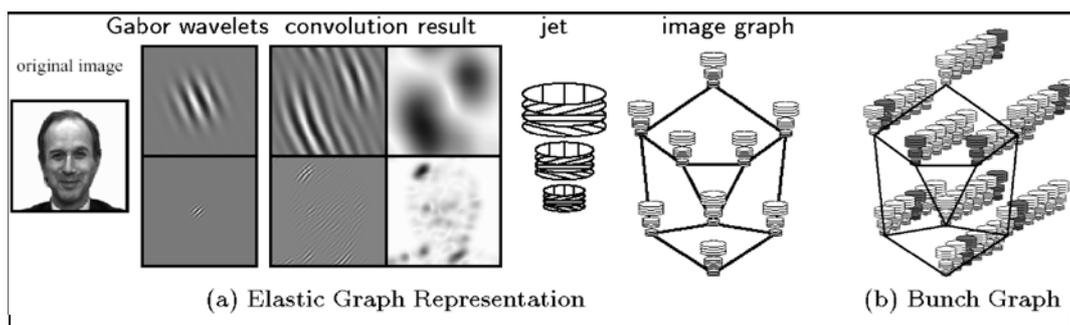


Figure II.7. The Bunch Graph Representation of faces used in Elastic Graph Matching.

L'approche EBGM est plus robuste aux changements d'éclairage, de pose et d'expressions faciales. Cependant elle est plus difficile à implémenter que les méthodes globales comme Eigenfaces et Fisherfaces. L'approche EBGM sera traitée en détail dans le prochain chapitre

II.6.2.2 Approches basées sur les régions faciales

On compte, dans cette catégorie, l'approche Eigenfeatures de Pentland et al.[17] [26] où l'ACP a été appliquée sur les yeux et le nez et celle de Visani [22] pour laquelle l'ADL bilinéaire a été appliquée pour l'extraction de caractéristiques à partir d'une région contenant les sourcils, les yeux et le nez et une autre qui ne contient que les sourcils et les yeux. Dans [62], trois approches bidimensionnelles (ACP2D, ADL2D et ACPDL2D) ont été testées pour comparer la discrimination des yeux et du nez sur une base de séquences vidéo.

II.6.3 Approches Hybrides

Les approches hybrides utilisent conjointement les deux types de modélisation cités auparavant (globale et locale) pour la reconnaissance du visage. Parmi les approches les plus importantes on trouve:

II.6.3.1 Approches modulaires

Ces techniques reposent sur des approches globales, mais appliquées de manière modulaire à différentes régions faciales, et combinées de manière à obtenir un modèle global alliant plusieurs modèles locaux [17] [22]. L'idée est que les différentes régions faciales ne sont pas affectées de la même manière par les différentes sources de variabilité.

Brunnelli et Poggio [24] ont appliqué la technique de corrélation sur plusieurs régions faciales (les yeux, le nez, la bouche, et la région faciale dans sa globalité du haut des sourcils jusqu'au menton). Pentland et al. [26] ont introduit l'approche des Eigenspaces Modular. Les régions faciales retenues englobent la totalité du visage, les yeux et le nez. Une ACP est appliquée sur chacune de ces régions et les résultats de classification sont agrégés.

En 2003, Heisel et al. [17][22] ont introduit une technique modulaire utilisant les Machines à Vecteurs de Support (SVM). Après la détection de dix caractéristiques faciales, des blocs de pixels englobant ces régions faciales sont extraits. Chacun de ces blocs est transformé en un vecteur par concaténation des lignes (ou colonnes) de pixels. Les vecteurs correspondants aux dix caractéristiques sont concaténés pour obtenir un vecteur par observation, ces vecteurs constituant les signaux d'entrée.

II.6.3.2 Analyse des caractéristiques locales

En 1996, Penev et Atick ont proposé une technique hybride de représentation de visages basée sur l'utilisation conjointe de l'Analyse des Caractéristiques Locales (ACL) et de l'Analyse en Composantes Principales (ACP) [15] [22]. Cette technique a été utilisée par le groupe **Visionic** dans le logiciel commercial de reconnaissance de visages **FaceIt**.

II.7 Quelques difficultés inhérentes à la reconnaissance de visages

Les systèmes automatiques de reconnaissance de visages doivent rester invariants à tout facteur indépendant de l'identité du visage, même si ce facteur engendre des changements d'apparence du visage. Or de nombreux facteurs extérieurs au visage ou en lien avec sa nature intrinsèque peuvent influencer sur celle-ci. Les variations entre des images d'un même visage dues à l'illumination et à l'angle de vision sont presque toujours plus importantes que les variations entre images dues à un changement de l'identité du visage [28]. Dans cette section, nous allons donner un aperçu sur quelques facteurs parmi tant d'autres qui peuvent contribuer à la complexité de la tâche de tout système de reconnaissance de visages.

II.7.1 Variations de pose

Un changement de l'angle d'inclinaison du visage engendre de nombreux changements d'apparence dans l'image collectée. En effet, si la pose du visage requête diffère significativement de celle des visages enregistrés, les performances des systèmes de reconnaissance de visages sont affectées et les taux de reconnaissance baissent sensiblement.

Selon le FRVT 2000, la rotation de la tête n'entraîne pas de baisses des taux de reconnaissance significatives jusqu'à $\pm 25^\circ$, alors qu'à partir de $\pm 40^\circ$ on constate une chute des performances.

Gross et al. ont mis en évidence le fait qu'un modèle construit à partir de poses frontales présente une meilleure capacité de généralisation à d'autres poses qu'un modèle construit à partir de poses non frontales. Parmi les solutions proposées pour ce genre de problème, il est nécessaire de construire plusieurs modèles de reconnaissance (un par type de pose). En phase de reconnaissance, il faudra alors utiliser un **classifieur de pose** en amont de la reconnaissance de manière à ce que tout visage requête ne soit comparé qu'au modèle de sa pose [22][29].



Figure II.8. Exemple de variations de la pose de la tête [22].

II.7.2 Changements d'illumination

L'intensité et la direction d'illumination, lors de la prise de vue, influent toutes deux énormément sur l'apparence du visage dans l'image. Dans la plus part des applications réelles, des changements dans les conditions d'illumination sont néanmoins inévitables, notamment lorsque les vues sont collectées à des dates différentes en intérieur ou en extérieur [17] [22].

Etant donné qu'un visage humain est un objet intrinsèquement 3D, des changements d'illumination peuvent faire apparaître sur le visage des ombres accentuant ou au contraire masquant certaines caractéristiques faciales (occultation d'une partie du visage) comme la montre la figure II.9.

L'évaluation du protocole FRVT [29] conclut que des changements importants dans les conditions d'illumination peuvent mener à des baisses considérables dans les taux de reconnaissance. Enfin, le couplage de variations d'illumination et de variations de pose constitue une difficulté importante.



Figure II.9. Effets de variations dans les directions d'illumination sur l'apparence d'un visage [22].

II.7.3 Variations d'expressions faciales

Les visages sont des objets non rigides. L'expression faciale de l'émotion surtout combinée avec la parole peut produire des changements d'apparence importants des visages. Comme les expressions faciales affectent la forme géométrique et les positions des caractéristiques faciales, il semble logique que les techniques globales et hybrides soient plus robustes aux expressions faciales que les techniques géométriques.

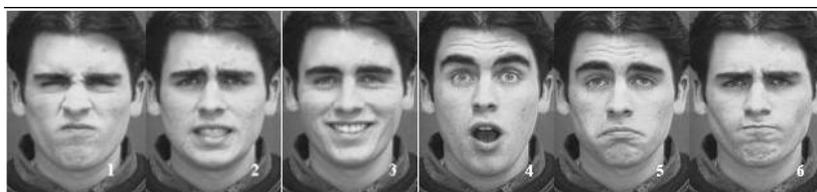


Figure II.10. Exemples de variations d'expressions faciales

Parmi les études réalisées sur l'impact des changements d'expressions faciales sur la reconnaissance, on trouve celle de Gross et al. [37] qui ont utilisé la base des expressions faciales de Kohn-Kanade [38] et ont démontré que généralement les algorithmes de reconnaissance sont relativement robustes aux changements d'expressions faciales à l'exception des cas extrêmes engendrant d'importantes déformations de la bouche (tel que le cri) et le rétrécissement ou la fermeture complète des yeux.

II.7.4 La taille de la base de visages

C'est lors du protocole d'évaluation FRVT 2002 que l'impact de la taille de la base sur les performances du système de reconnaissance a été étudié pour la première fois, grâce à la très grande taille de la base HC Int utilisée [39]. Les meilleurs systèmes fournissent des taux de reconnaissance de l'ordre de 85% pour 800 personnes, 83% pour 1600 personnes, et 73% pour 37437 personnes. En conclusion, les performances décroîtraient de manière Log linéaire en fonction de la taille de la base.

II.8 Evaluation des performances des systèmes de reconnaissance de visages

Etant donné le nombre et la variété des algorithmes de reconnaissance de visages proposés ces dernières années, il est très important de définir des protocoles expérimentaux permettant leur évaluation. A cet effet, les premiers travaux réalisés sur l'évaluation ont fourni les principaux facteurs que doit prendre en compte tout protocole d'évaluation pour qu'il soit efficace. Parmi ces facteurs, le nombre d'images test utilisé doit être important et les images des visages données doivent être semblables à celles rencontrées en situation réelle. De plus, la mesure des performances doit se faire selon une méthodologie prenant en compte le coût des erreurs en fonction du niveau de sécurité des applications [22].

La première évaluation a été menée en Août 1994 par le National Institut of Standards and Technology (NIST) avec le protocole **FERET** [40], il visait le test des algorithmes destinés à localiser, normaliser, et identifier automatiquement les visages. Depuis, deux autres évaluations utilisant le protocole FERET ont eu lieu, l'une en mars 1995, l'autre entre septembre 1996 et mars 1997. Ces évaluations menées sur la base de visages FERET (14126 images) ont permis la comparaison de neuf algorithmes proposés par des institutions et entreprises dans des contextes d'authentification et d'identification [17] [22]. Les résultats obtenus ont montré la supériorité de trois algorithmes à savoir: Probabilistic eigenfaces de MIT, Subspace LDA [17] [68] [69] de UMD et Elastic Graph Matching [70] de USC. Ces évaluations ont également permis de mettre en évidence les principaux facteurs pouvant influencer sur les performances des algorithmes, tels que les changements de pose ou d'illumination.



Figure II.11. Extrait d'images d'une personne de la base FERET.

A ces évaluations, ont succédé celles du Face Recognition Vendor Test (**FRVT**) [29] [39]. En 2000 (FRVT 2000) cinq compagnies ont participé à cette évaluation. Un protocole expérimental similaire à celui de FERET (session septembre 1996) a été utilisé, à la différence près que la base utilisée en 2000 comptait beaucoup plus d'images. En 2002, dix participants (industriels) ont été évalués entre juillet et août 2002 sur la base de visages High Computational Intensity (HCInt) qui contient 121589 images de 37437 personnes différentes.

Un autre protocole d'évaluation **XM2VTS** a été introduit essentiellement pour l'évaluation d'algorithmes de reconnaissance multimodaux (voix et visages) et 3D, et ceci pour l'authentification uniquement [17] [22].

II.9 Conclusion

Durant ces trois dernières décennies, la reconnaissance automatique des visages est devenue l'une des applications les plus réussies en traitement d'images et en reconnaissance de formes. Néanmoins, et malgré la multitude des techniques proposées et les progrès réalisés, la reconnaissance robuste de visages reste très difficile.

Dans ce chapitre, nous avons présenté les différentes étapes d'un processus de reconnaissance de visages et nous avons passé en revue les principales techniques de reconnaissance proposées à ce jour. Ces techniques se décomposent en trois grandes familles: i) les approches globales pour les quelles les caractéristiques du visage sont extraites directement depuis l'ensemble des valeurs de pixels des images, ii) les approches locales basées sur l'extraction de caractéristiques des régions bien définies du visage comme les yeux, le nez, la bouche, etc. Parmi les approches locales utilisées, on compte l'approche géométrique et l'approche basée sur les graphes dite **Elastic Bunch Graph Matching** qui fera l'objet du chapitre suivant et iii) les techniques dites hybrides qui combinent les méthodes locales et globales.

En outre, nous avons présenté un aperçu sur les principales difficultés inhérentes à la reconnaissance automatique de visages. Enfin, nous avons discuté les différents protocoles d'évaluation des performances de nombreux systèmes de reconnaissance de visages.

Chapitre III

APPROCHE EBG

III.1. Introduction

Durant les vingt dernières années, des efforts importants sont fournis dans le domaine de la reconnaissance automatique de visages, et plusieurs techniques ont été proposées. Ces techniques, comme nous les avons détaillé dans le chapitre précédent, sont regroupées selon quelques critères en deux catégories principales [50]: les techniques dites globales et les techniques basées sur des caractéristiques locales. Parmi les approches les plus prometteuses dans les techniques globales on trouve Eigenfaces et Fisherfaces.

Quant à la deuxième catégorie qui englobe les techniques locales, on peut distinguer deux types d'approches. Celles qui exigent une localisation approximative seulement des différentes caractéristiques locales du visage (les yeux, le nez, la bouche,...etc.) et celles qui exigent une localisation précise des points caractéristiques.

Dans ce chapitre, nous allons détailler une approche dite **Elastic Bunch Graph Matching (EBGM)** que nous allons l'utiliser dans notre travail. Cette approche est considérée comme l'une des méthodes locales les plus importantes qui exigent une localisation précise des points caractéristiques.

Elle consiste à localiser des points caractéristiques (appelés points fiduciels) du visage tels que les coins des yeux, la pointe du nez, les commissures de la bouche, etc. Ensuite, un treillis élastique virtuel est appliqué sur l'image de visage à partir de ces points fiduciels. Chaque point représente un noeud labellisé auquel on associe un vecteur de coefficients d'ondelettes complexes de Gabor, appelés **Jet**. Pour effectuer une reconnaissance avec une image test, on fait une mesure de similarité entre les différents Jets.

III.2 Approche EBG

L'algorithme EBG est né des travaux de Wiskott et al. de la Southern California University (USC - USA) et de la Ruhr University (Allemagne), en 1997 [70][16]. Cet algorithme trouve ses fondements dans les neurosciences en imitant le fonctionnement de certaines cellules spécialisées localisées dans le cortex visuel humain. Ces cellules sont spécialisées dans la reconnaissance des contours et sont activées lors de l'apparition ou de la disparition d'un stimulus visuel. Un modèle simple de réponses de ces cellules spécialisées est appelé champ réceptif linéaire (**RF**: Receptive Field).

Pollen et Ronner [57] ont montré que la plupart de ces cellules simples peuvent être combinées par paire. La première avec une symétrie paire que l'on modélise mathématiquement par une fonction cosinus, la seconde possédant une symétrie impaire que l'on modélise par une fonction sinus. En 1987, Jones et Palmer [73] suggèrent de modéliser la forme des champs réceptifs linéaires par des filtres de Gabor 2D (Figure III.1)

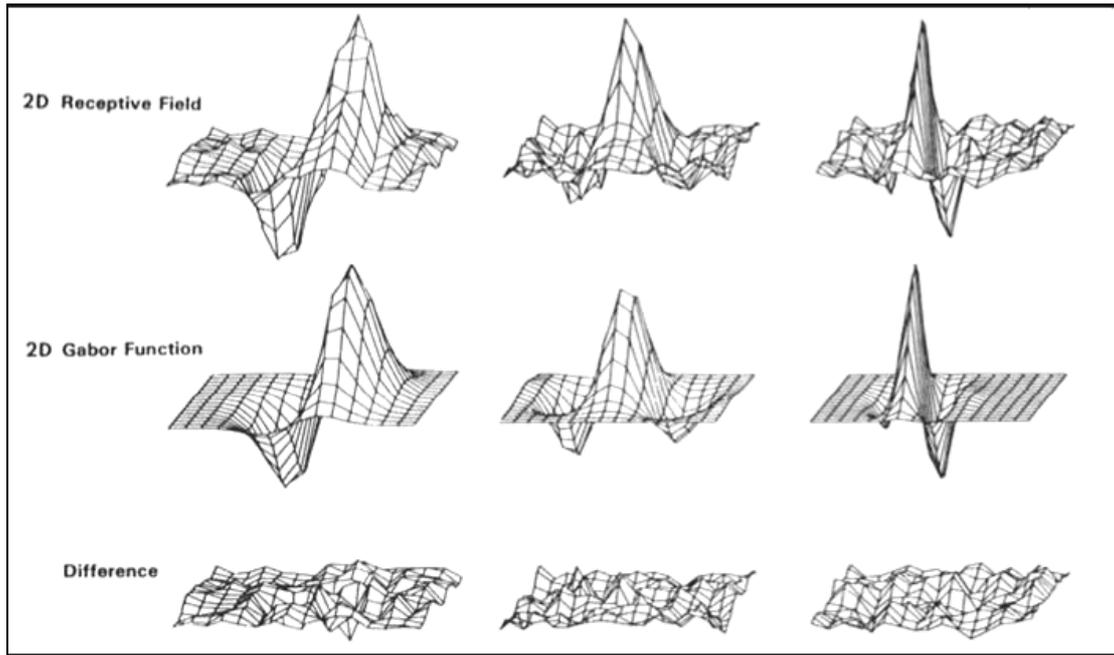


Figure III.1. Structure spectrale 2D de RF simples dans un cortex strié de chat. On peut apprécier la grande ressemblance entre le modèle mathématique des ondelettes 2D de Gabor et les réponses en fréquence des cellules simples de V1.

Soit $a_j(x_0)$ la réponse d'une telle cellule, elle est exprimée par la corrélation de la donnée en entrée (i.e. une image $I(x)$) et le modèle simple **RF** $\Psi_j(x - x_0)$:

$$a_j(x_0) = \int I(x) \Psi_j(x - x_0) dx \quad \dots\dots\dots \text{III.1}$$

Par conséquent, et comme nous pouvons modéliser le modèle simple de réponses RF par des ondelettes 2D de Gabor, le Jet dans l'algorithme EBG correspond bien à la réponse électrophysiologique $a_j(x_0)$ [16].

Ainsi, un Jet est basé sur une transformée en ondelettes [58], défini comme la convolution d'une image avec une famille de noyaux de Gabor. Ces noyaux de Gabor peuvent être assimilés à des ondes localisées dans le temps, modulées par une Gaussienne.

III.3. Généralités sur les ondelettes de Gabor

Pour comprendre comment les ondelettes de Gabor parviennent à jouer un rôle fondamental dans l'algorithme EBG, il est nécessaire de donner un aperçu sur l'analyse de Fourier et par ondelettes.

III.3.1 Transformée de Fourier

L'analyse de Fourier nous enseigne qu'un signal quelconque peut être décomposé en une somme de sinusoïdes de fréquences et d'amplitudes variables [60]. Un signal est entièrement caractérisé par l'ensemble des amplitudes des sinusoïdes, qui forme ce que l'on appelle sa transformée de Fourier. Cette transformée de Fourier est porteuse de précieuses informations sur le signal analysé [59].

La transformée de Fourier monodimensionnelle de la fonction x est donnée par :

$$f(x(t))(\omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t} dt \quad \dots\dots\dots \text{III.2}$$

qui peut être aussi écrite sous la forme :

$$f(x(t))(\omega) = \int_{-\infty}^{+\infty} x(t)\cos(\omega t)dt - i \int_{-\infty}^{+\infty} x(t)\sin(\omega t)dt \quad \dots\dots\dots \text{III.3}$$

Cette équation montre que la transformée de Fourier décompose la fonction originale en fonctions sinus et cosinus.

Evaluer $f(x(t))(\omega)$ à une fréquence ω bien déterminée donnera un nombre complexe dont les parties réelle et imaginaire sont respectivement les amplitudes des termes cosinus et sinus de la fonction originale à cette fréquence.

Si on somme tous les signaux sinus et cosinus pour toutes les valeurs de fréquence ω on pourra reconstruire le signal original $x(t)$ donné par :

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x(t))(\omega)e^{j\omega t} d\omega \quad \dots\dots\dots \text{III.4}$$

Pour chaque valeur de fréquence ω le signal original $x(t)$ possède une partie réelle ($a_{\omega,r}$) et une partie imaginaire ($a_{\omega,i}$):

$$x_{\omega}(t) = a_{\omega,r} \cos(t) + a_{\omega,i} \sin(t) \quad \dots\dots\dots \text{III.5}$$

$$x_{\omega}(t) = a \cos(t + \Phi) \quad \dots\dots\dots \text{III.6}$$

Où a et Φ représentent respectivement le module et la phase de la sinusoïde en la fréquence ω .

Limites de la transformée de Fourier

L'analyse de Fourier est un outil performant pour l'analyse des signaux périodiques et suffisamment réguliers notamment grâce à son algorithme rapide (Transformée de Fourier rapide dit **FFT**: **F**ast **F**ourier **T**ransform).

Cependant, si le signal est irrégulier, la quantité de paramètres à conserver devient trop importante. Par conséquent, même avec la transformée de Fourier rapide, l'analyse de Fourier présente des inconvénients majeurs pour une analyse satisfaisante du signal. Parmi les principaux inconvénients, on peut citer [60]:

1. Dans le spectre de $f(x(t))(\omega)$ tous les aspects temporels du signal $x(t)$ disparaissent, la transformée de Fourier est ainsi incapable de localiser les portions du signal dans lesquelles les variations sont rapides, ni celles où elles sont lentes. Un exemple est donné sur la figure III.2, où deux signaux différents présentent des transformées de Fourier similaires.

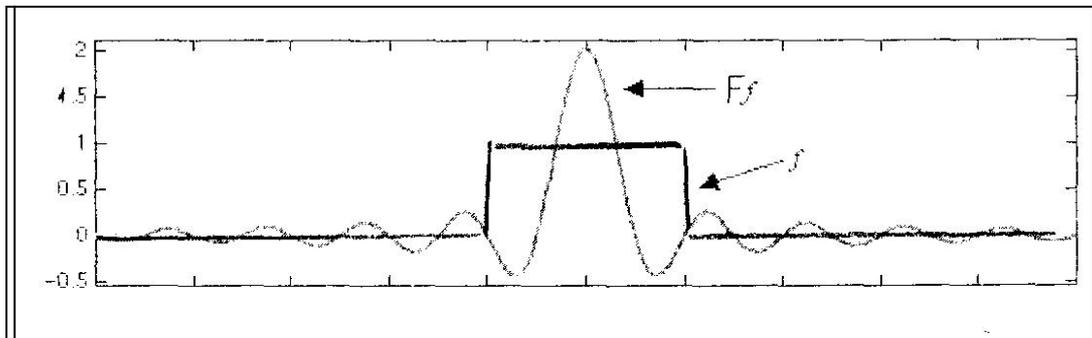


Figure III.2.a : f est un signal créneau et Ff : sa transformée de Fourier.

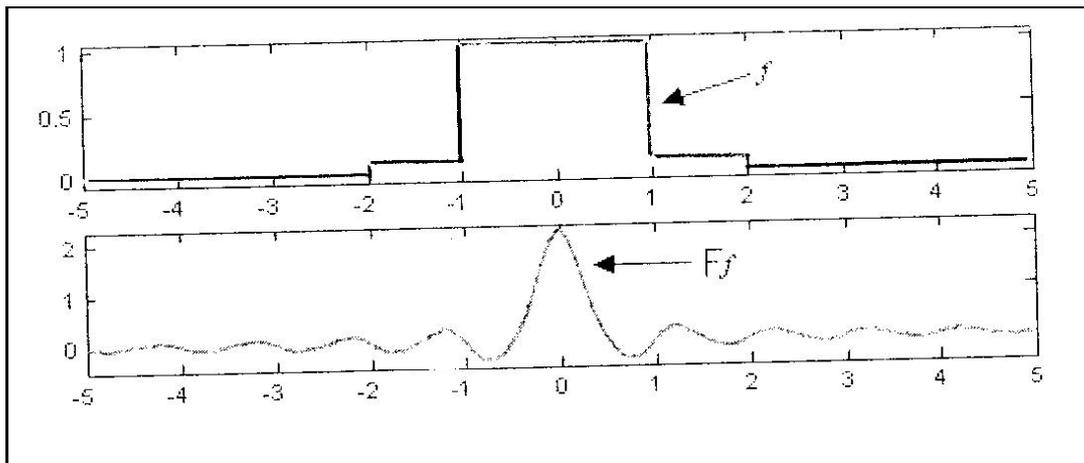


Figure III.2.b: $f = 0.9 1[-1, 1] + 0.1 1[-2, 2]$ et Ff : sa transformée de Fourier.

2. La non causalité de la transformée de Fourier. Ainsi, le calcul de $f(x(t))(\omega)$ nécessite la connaissance de $x(t)$ sur \mathbb{R} tout entier. Une analyse en temps réel est donc impossible, et on ne peut pas connaître le spectre de $f(x(t))(\omega)$ si on ignore la fonction. L'exemple de la figure III.3 montre que deux fonctions identiques sur un même intervalle peuvent avoir des transformées de Fourier très différentes.

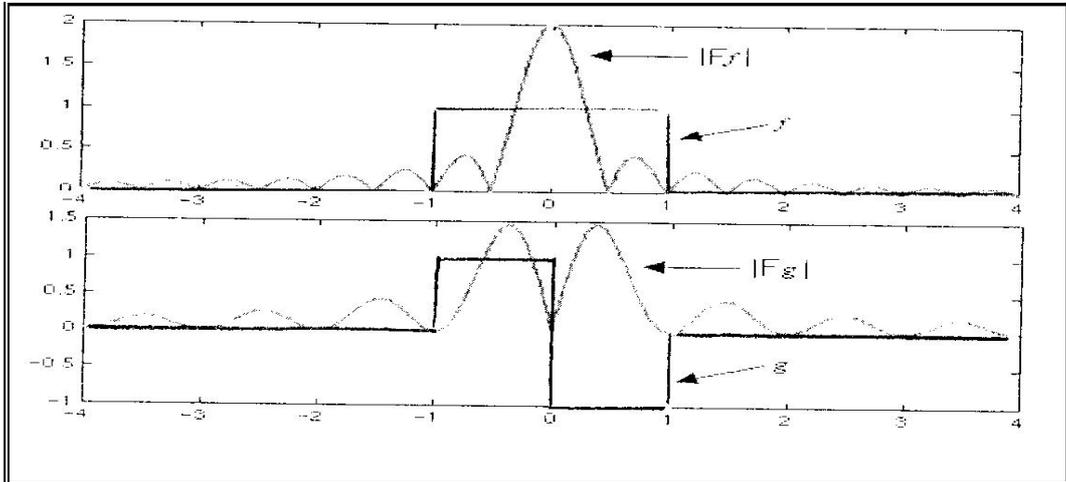


Figure III.3. Exemple de transformée de Fourier de deux fonctions identiques sur un même intervalle de temps $[-1 0]$

3. Le troisième inconvénient est dû au principe d'incertitude de Heisenberg :

$$\Delta t * \Delta \lambda \geq \frac{1}{4\pi}$$

Plus le support de $x(t)$ est petit plus le support de $f(x(t))(\omega)$ est grand et réciproquement. La fonction de Dirac (δ) est un exemple de ce cas.

$$f(\delta)(\omega) = 1, \text{ et } f(1) = \delta$$

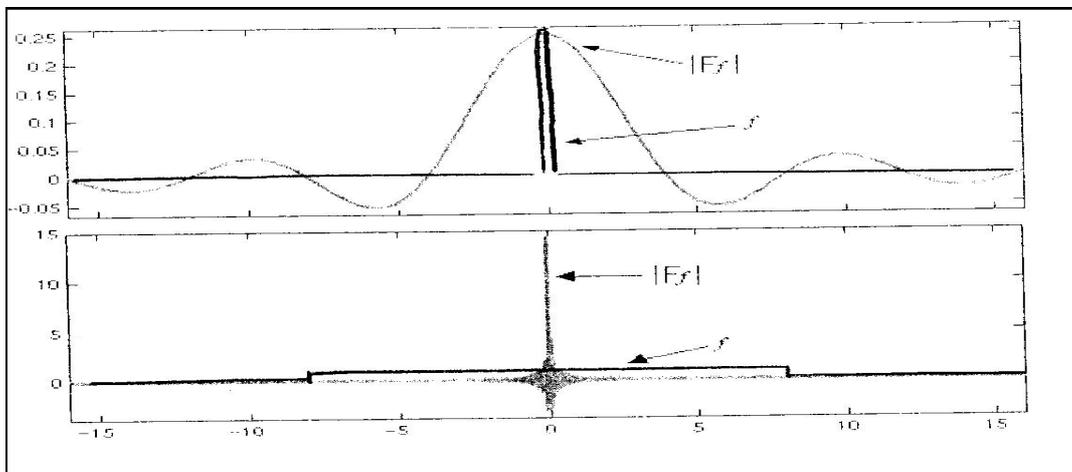


Figure III.4. Exemple de transformée de Fourier de la fonction Dirac

III.3.2 Transformée de Gabor (transformée de Fourier à fenêtre)

Afin de donner une représentation des signaux permettant de faire apparaître simultanément des informations temporelles (localisation dans le temps, durée) et fréquentielles, facilitant par là, l'identification des caractéristiques physiques de la source du signal, un prototype d'analyse par ondelettes avait été proposé au milieu des années 1940 par le physicien Denis Gabor (qui reçut par la suite le prix Nobel pour ses travaux sur l'holographie).

Gabor suggérait de rendre locale l'analyse de Fourier, en s'aidant de fenêtres. Une fenêtre est une fonction régulière, lentement variable, et bien localisée (ce qui signifie qu'elle est nulle en dehors d'une certaine zone, son support). En multipliant la fonction étudiée par une fenêtre, on en obtient une version «**locale**», dont on peut déterminer le contenu fréquentiel par analyse de Fourier classique. On renouvelle alors l'opération en déplaçant la fenêtre d'analyse. L'ensemble de ces transformées de Fourier ainsi localisées forme la transformée de Gabor du signal, et fournit donc une analyse fréquentielle locale.

On obtient alors une famille de coefficients $W_x(\omega, b)$ remplaçant les valeurs de $f(x(t))(\omega)$ où ω représente la fréquence, b localise l'analyse et g est la fenêtre :

$$W_x(\omega, b) = \int_{-\infty}^{+\infty} x(t) \overline{g(t-b)} e^{-2i\pi\omega t} dt \quad \dots\dots\dots \text{III.7}$$

L'application $f \longrightarrow W_x$ s'appelle la transformée de Fourier à fenêtre glissante ou encore la transformée de Gabor. On note alors :

$$g_{\omega, b}(t) = g(t-b) e^{-2i\pi\omega t} \quad \dots\dots\dots \text{III.8}$$

Les figures suivantes illustrent, respectivement, les réponses temporelles et fréquentielles des principales fenêtres.

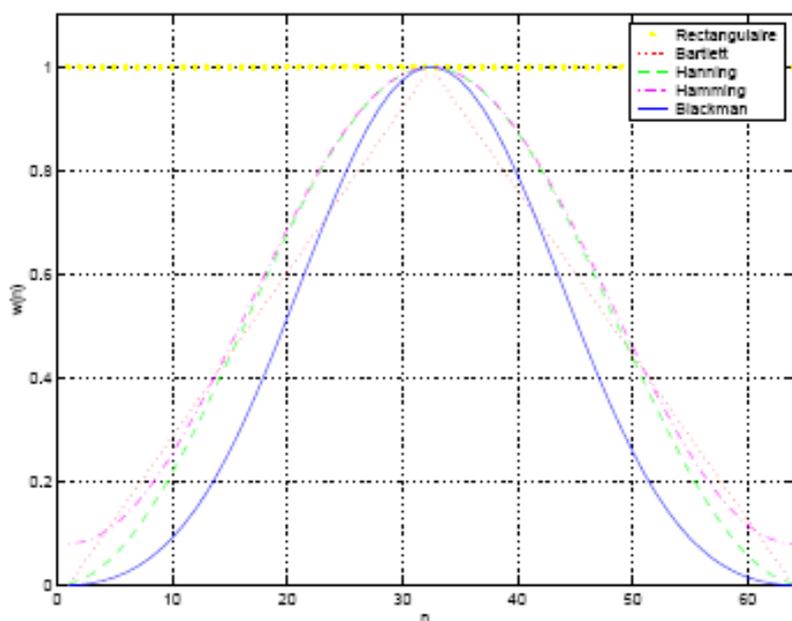


Figure III.5. Réponses temporelles des principales fenêtres.

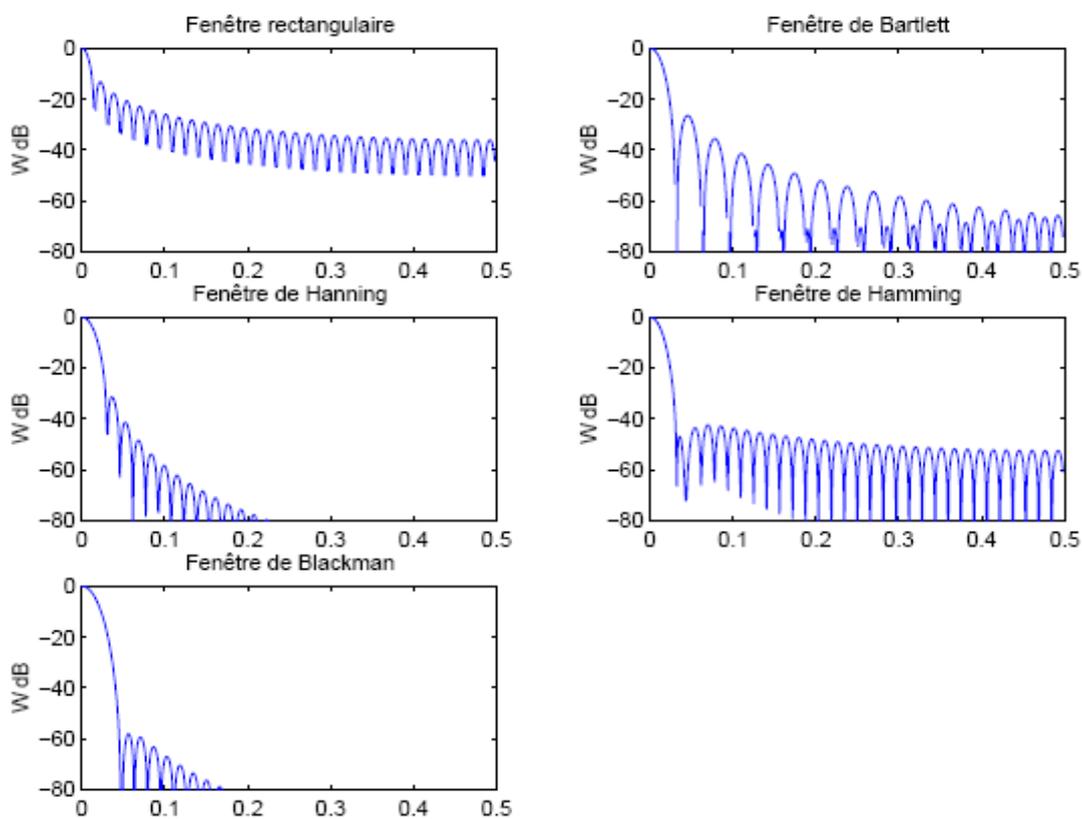


Figure III.5. Réponses fréquentielles des principales fenêtres.

III.3.3 Ondelettes

Nous avons vu précédemment que la transformée de Fourier est une transformation globale, la transformée de Gabor est locale, mais toutes les deux sont de résolution temporelle fixe. Cet inconvénient peut être résolu en utilisant la transformée en ondelettes. En effet, la transformée en ondelettes est non seulement une analyse locale, mais de plus sa résolution temporelle est variable. Elle permet donc de décrire le comportement local des signaux (fonctions) à différentes échelles de temps.

La transformation en ondelettes consiste à créer à partir d'une ondelette mère Ψ (qui ressemble à une petite onde) une famille d'ondelettes.

$$\Psi_{ab}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \dots\dots\dots \text{III.9}$$

Où $a \in \mathbb{R}^+$ est l'échelle et $b \in \mathbb{R}$ est la position (a sert à dilater et b à translater). On dit également que tout couple $(a,b) \in \mathbb{R}^+ \times \mathbb{R}$ définit un atome de la transformée.

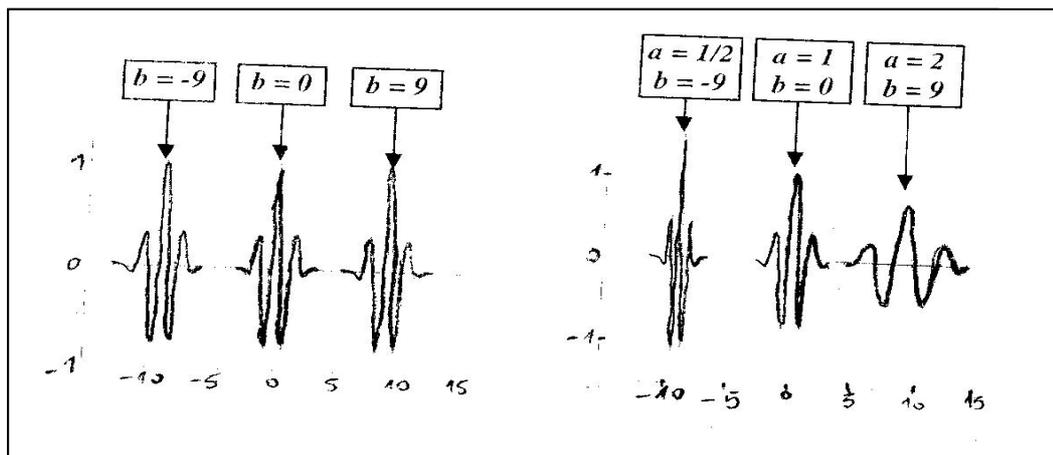


Figure III.6. Exemple de translation et dilatation d'une ondelette.

III.3.4 Analyse multi résolution et analyse temps échelle

Les ondelettes étant une extension de l'analyse de Fourier, le principe est le même : on reconstruira un signal à partir de sa transformée en ondelettes en "additionnant" des ondelettes de différentes tailles de même qu'on reconstruit un signal à partir de sa transformée de Fourier.

Toutefois, en comprimant ou en étirant les ondelettes (varier la valeur de a) fait varier leur fréquence permettant aux ondelettes de s'adapter automatiquement aux différentes composantes du signal: elles utilisent une fenêtre étroite pour regarder les composantes transitoires de hautes fréquences (révéler les détails de plus en plus fins), et une fenêtre large pour regarder les composantes de longue durée, de basse fréquence. C'est cette procédure qu'on appelle **multi résolution**.

Le terme échelle désigne la corrélation entre la taille de l'ondelette et la taille des composantes que l'on peut voir. On parle alors d'analyse **temps-échelle** pour les ondelettes comme on parlait d'analyse **temps-fréquence** pour la transformée de Gabor.

III.3.5 Ondelettes de Gabor

Les caractéristiques de l'ondelette mère sont nettement différentes de celles d'une fenêtre ordinaire. En effet, une fenêtre avait plutôt l'allure d'un créneau alors que la définition de Ψ lui imposera d'être d'intégrale nulle. Par ailleurs, on imposera à Ψ une bonne localisation autour de 0 et pour $\hat{\Psi}$ une localisation autour d'une certaine fréquence $\omega > 0$. On obtient ainsi des fonctions à décroissance rapide : elles oscillent puis s'amortissent ; ce qui leur donne leur nom d'ondelette.

Les ondelettes de Gabor sont fondamentalement des ondes localisées dans le temps, modulées par une Gaussienne (Figure III.7). Ainsi, la convolution d'une fonction avec l'ondelette de Gabor nous fournit l'information fréquentielle autour du centre de la gaussienne par contre l'information fréquentielle lointaine du centre sera négligeable.

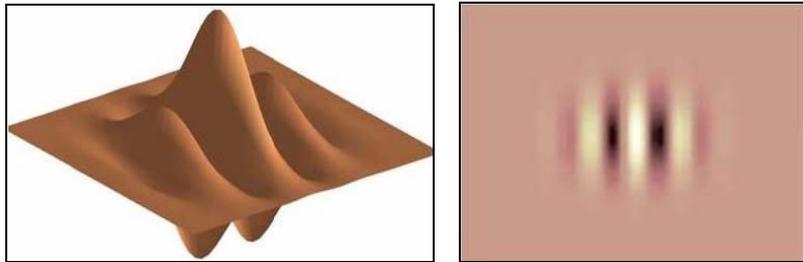


Figure III.7. Représentation 3D (à gauche) et 2D (à droite) de la partie réelle d'une ondelette de Gabor.

L'ondelette de Gabor monodimensionnelle est donnée par l'équation :

$$W(t, t_0, \omega) = e^{-\sigma(t-t_0)^2} e^{-i\omega(t-t_0)} \dots\dots\dots \text{III.11}$$

La convolution de la fonction $x(t)$ avec l'ondelette de Gabor est donnée par :

$$C(x(t))(t_0, \omega) = \int_{-\infty}^{+\infty} x(t) W(t, t_0, \omega) dt \dots\dots\dots \text{III.12}$$

Par extension de la fonction ondelette, la convolution ressemble à la transformée de Fourier :

$$C(x(t))(t_0, \omega) = \int_{-\infty}^{+\infty} x(t) e^{-\sigma(t-t_0)^2} e^{i\omega(t-t_0)} dt \quad \dots\dots\dots \text{III.13}$$

$$C(x(t))(t_0, \omega) = \int_{-\infty}^{+\infty} x(t) e^{-\sigma(t-t_0)^2} \cos(\omega(t-t_0)) + i \int_{-\infty}^{+\infty} x(t) e^{-\sigma(t-t_0)^2} \sin(\omega(t-t_0)) \quad \dots\dots\dots \text{III.14}$$

Ce qui donne un coefficient complexe $C(x(t))(t_0, \omega)$ décrivant le contenu fréquentiel local de la fonction $x(t)$ aux valeurs spécifiques de fréquence ω et de temps t_0

III.3.6 Ondelettes de Gabor 2D

L'algorithme EBGM utilise les ondelettes 2D de Gabor pour le traitement des images. Elle consiste en une onde sinusoïdale plane multipliée par une Gaussienne bidimensionnelle. L'onde sinusoïdale est ainsi activée par l'information fréquentielle de l'image, quant à la Gaussienne, elle assure que la convolution soit dominée par la région de l'image près du centre de l'ondelette [16] [61]. Les ondelettes 2D de Gabor sont des fonctions de type $f(\theta, \lambda, \phi, \sigma, \gamma)$. Voici une explication des cinq (05) paramètres qui contrôlent l'ondelette.

1. Orientation de l'ondelette (θ)

Ce paramètre fait pivoter l'ondelette autour de son centre. L'orientation de l'ondelette détermine l'angle des contours ou des lignes de l'image auxquelles l'ondelette va être sensible. Dans la plupart des cas θ prend ses valeurs entre 0 et π . Les valeurs de π à 2π sont redondants vu la symétrie de l'ondelette.

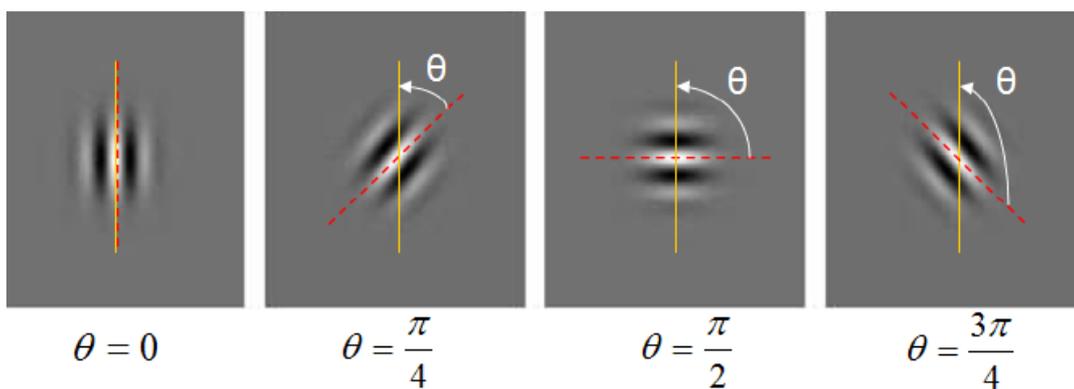


Figure III.8. Orientation de l'ondelette.

2. Fréquence centrale de l'ondelette (λ)

Ce paramètre spécifie la longueur d'onde du cosinus ou inversement la fréquence centrale de l'ondelette. Les ondelettes avec une grande longueur d'onde seront sensibles à des changements progressifs d'intensité dans une image. Les ondelettes avec une petite longueur d'onde seront sensibles à des contours et des lignes abruptes.

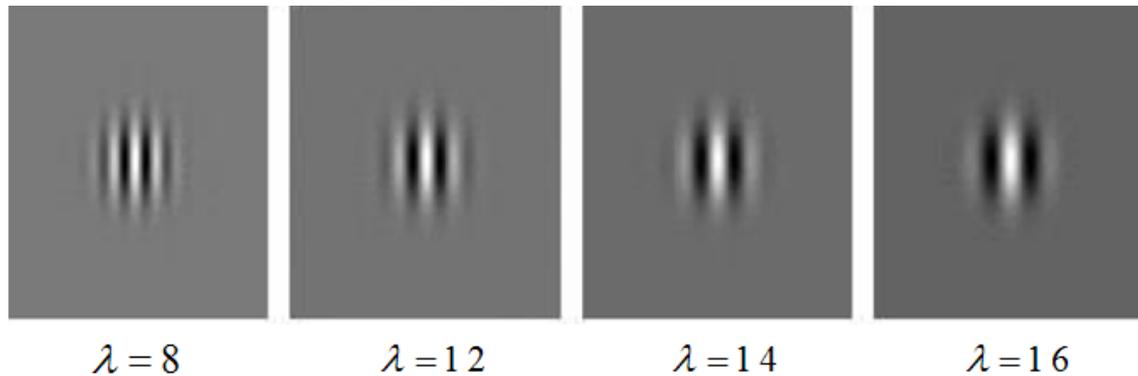


Figure III.9. Fréquence centrale de l'ondelette.

3. Phase de la sinusoïde (ϕ)

On utilise à la fois la partie réelle et la partie imaginaire de l'ondelette complexe de Gabor. Ce qui nous donne deux ondelettes : une ondelette paire et une ondelette impaire. La convolution (incluant alors deux phases) donne un coefficient complexe basé sur deux ondelettes qui sont déphasées de $\frac{\pi}{2}$.

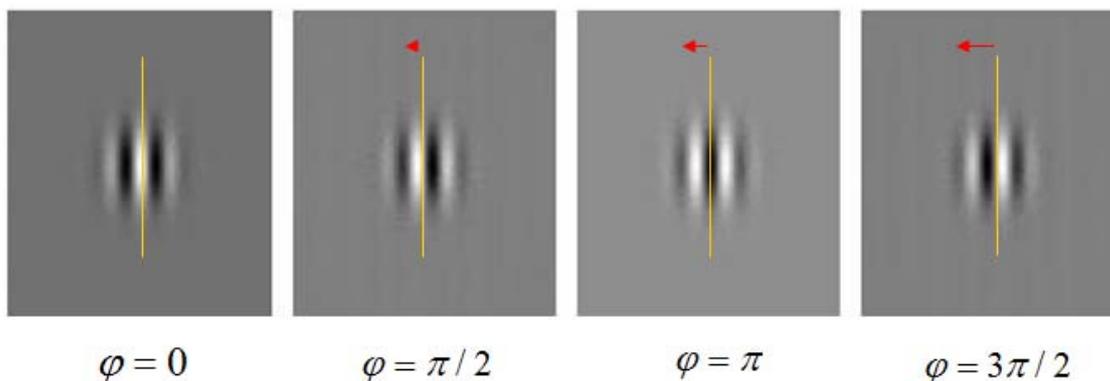


Figure III.10. Phase de l'ondelette.

4. Support temporel de l'ondelette (σ)

Ce paramètre spécifie le rayon de la Gaussienne. La taille de la Gaussienne détermine la quantité de pixels de l'image qui vont être pris en compte dans la convolution.

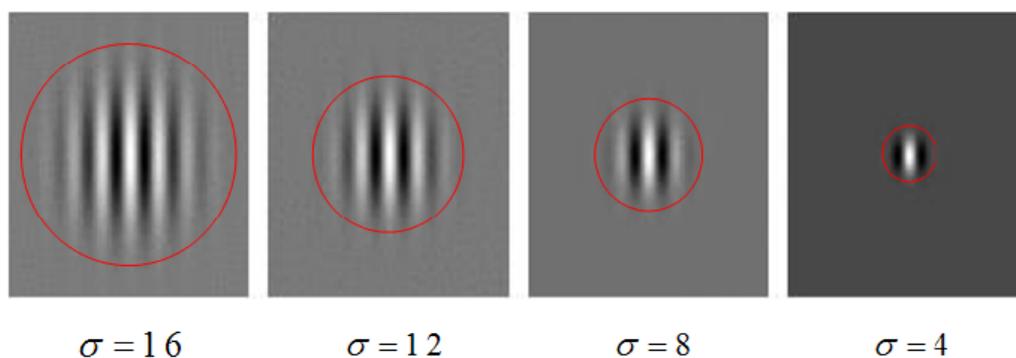


Figure III.11. Support temporel de l'ondelette.

5. Enveloppe de la Gaussienne (γ)

Ce paramètre agit sur la forme de l'enveloppe Gaussienne, en l'étirant spatialement. Ce paramètre a été inclus de manière à ce que les ondelettes puissent approximer certains modèles biologiques.

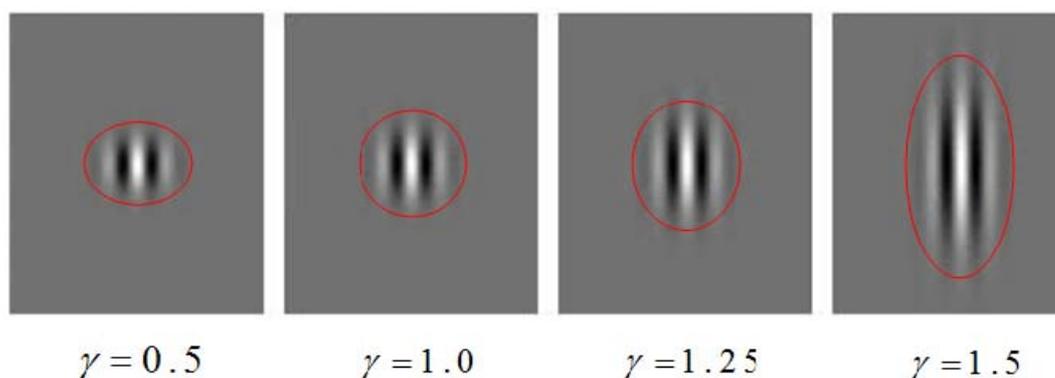


Figure III.12. Forme de l'enveloppe Gaussienne

En utilisant 5 fréquences différentes, 8 orientations différentes et 2 phases différentes. On obtient un total de 80 masques (filtres) d'ondelettes de Gabor différents (Figure III.13).

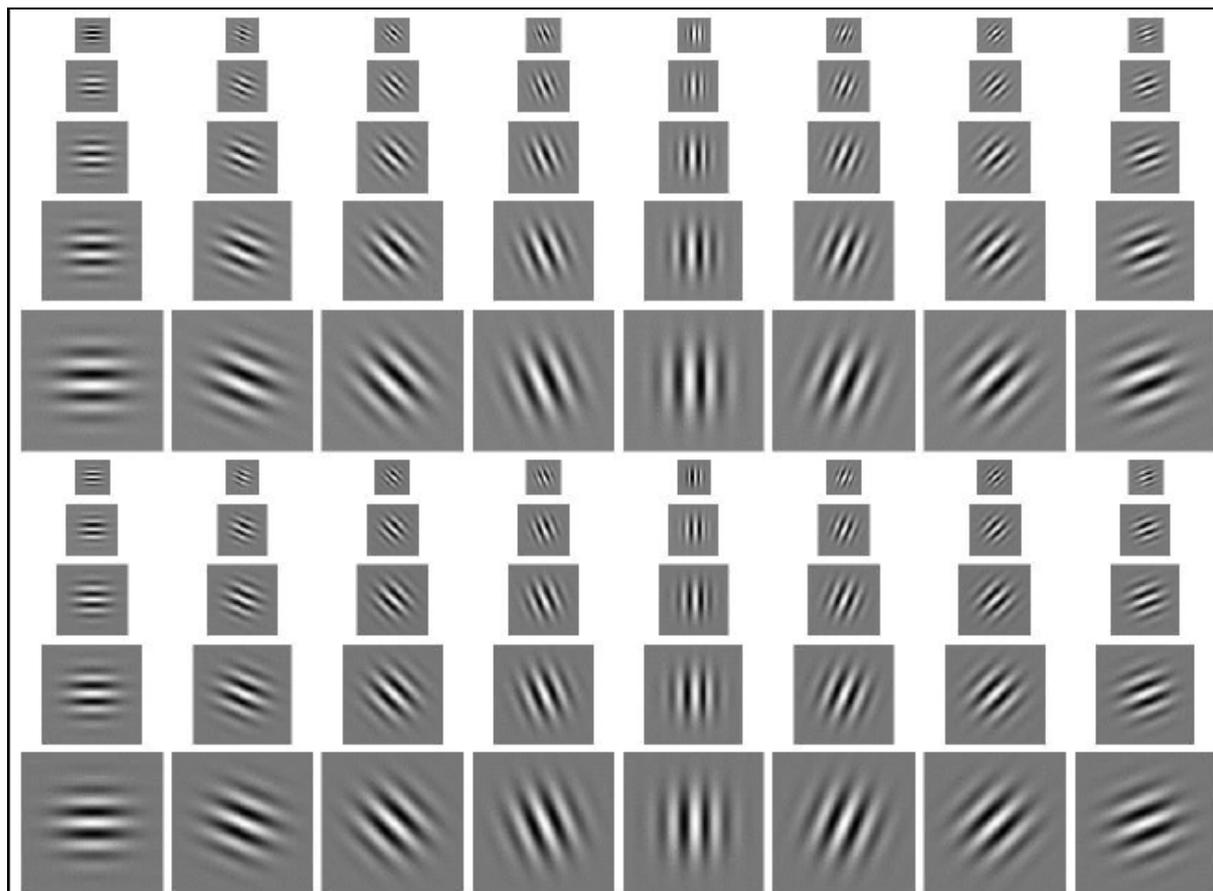


Figure III.13. Les 80 masques d'ondelettes de Gabor.

III.4. Etapes de l'algorithme EBG

L'algorithme EBG est principalement décomposé en trois étapes principales :

1. La première étape consiste à localiser les points de repère (Landmark locations) tels que les iris des yeux, la pointe du nez, les coins des yeux, les commissures de la bouche, le bas du menton, etc. Pour cela, plusieurs méthodes ont été proposées comme nous l'avons discuté précédemment (chapitre II).



Figure III.14. Sélection des points caractéristiques (points de repères).

2. Ensuite un treillis virtuel est construit en reliant les différents points repère (caractéristiques) précédemment trouvés. A chaque nœud du treillis correspond un point caractéristique et contient un jeu de coefficients complexes d'ondelettes de Gabor: le **Jet**. Les différents jets sont calculés en convoluant l'image de visage autour des points caractéristique avec plusieurs ondelettes de Gabor paramétrées (banc de filtres de Gabor).



Figure III.15. Création du treillis.

3. Calcul de similarité de deux images: Une fois la structure du FBG (Face Bunch Graph) trouvée, l'algorithme va procéder au calcul de similarité entre une image test et l'ensemble des images d'apprentissage. Pour cela, les points caractéristiques de l'image test sont trouvés le treillis est mis en place et les nouveaux jets calculés. La similarité de deux images est alors une fonction de correspondance des treillis (basée sur les différents jets de Gabor). La classification (ou la reconnaissance) finale se fait en maximisant cette fonction. Les mesures de similarité seront détaillées plus loin.

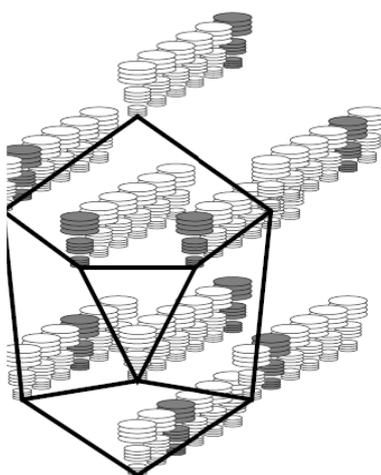


Figure III.16. Le “Face Bunch Graph” sert de représentation générale pour les visages.
(Chaque empilement de disques représente un Jet)

III.5 Analyse de sensibilité des différents points fiduciels :

Dans cette section, nous allons étudier l'importance de chaque point fiduciel du visage humain. Pour cela, on a employé 34 points fiduciels ainsi que leurs coefficients d'ondelettes de Gabor. Il est évident que chaque point fiduciel ne présente pas le même degré d'importance, par exemple les points sur la joue ne devraient pas être très instructifs.

Cette étude s'avère importante pour différentes raisons :

- Afin d'éliminer les points fiduciels les moins importants, et ainsi réduire la dimension de l'espace d'entrée.
- Il est difficile de localiser tous les points fiduciels avec une grande précision (par exemple les points sur la joue sont difficiles à localiser avec précision).
- Améliorer les performances des systèmes de reconnaissance [63] [75] [76].

Afin de déterminer l'importance des différents points fiduciels du visage (surtout utilisé pour la reconnaissance d'expressions faciales (FER)) [66], on utilise une méthode d'analyse de sensibilité. Si une variable d'entrée (point fiduciel) est importante dans la FER, enlever cette entrée causera un changement crucial du rendement i.e. le gradient de la fonction erreur en ce point devrait être élevée. Dans notre cas, nous avons un vecteur de coefficients d'ondelettes de Gabor comme entrée pour chaque point fiduciel.

Soit $X^{(P)}$ le vecteur caractéristique du point fiduciel P et $x_i^{(P)}$ son $i^{\text{ème}}$ élément

Nous calculons le gradient de la fonction erreur E , pour le vecteur $X^{(P)}$ comme suit:

$$G^{(P)} = \left\| \frac{\partial E}{\partial X^{(P)}} \right\| = \sqrt{\sum_i \left(\frac{\partial E}{\partial x_i^{(P)}} \right)^2} \quad \dots\dots\dots\text{III. 21}$$

Les points fiduciels dont le gradient est faible diffusent peu d'information utile pour la FER, et peuvent être moins importants que ceux qui disposent d'un gradient plus grand [66].

Les résultats obtenus sont résumés dans le tableau ci-après, nous fournissons également l'ordre des points fiduciels selon les valeurs de gradient: un point avec l'ordre 0 a la plus grande valeur de gradient; un point avec l'ordre 33 a la moindre valeur de gradient.

Point fiduciel	Gradient	Ordre	Point fiduciel	Gradient	Ordre
0	7.16990	6	17	0.31891	33
1	10.1341	2	18	0.32449	32
2	5.67114	12	19	0.42730	30
3	1.59910	23	20	0.49084	29
4	10.6303	1	21	0.49871	28
5	3.17509	19	22	3.96846	16
6	6.07363	9	23	0.54490	26
7	12.3932	0	24	0.82790	25
8	6.44077	8	25	3.34998	18
9	4.34609	15	26	6.67880	7
10	9.02400	3	27	2.47474	20
11	7.26315	5	28	5.60849	13
12	5.82878	10	29	1.96330	22
13	0.52836	27	30	0.38853	31
14	4.84604	14	31	2.47427	21
15	5.69166	11	32	8.16229	4
16	0.89122	24	33	3.84180	17

Tableau III.1. Gradient de la fonction Erreur des différents points fiduciels

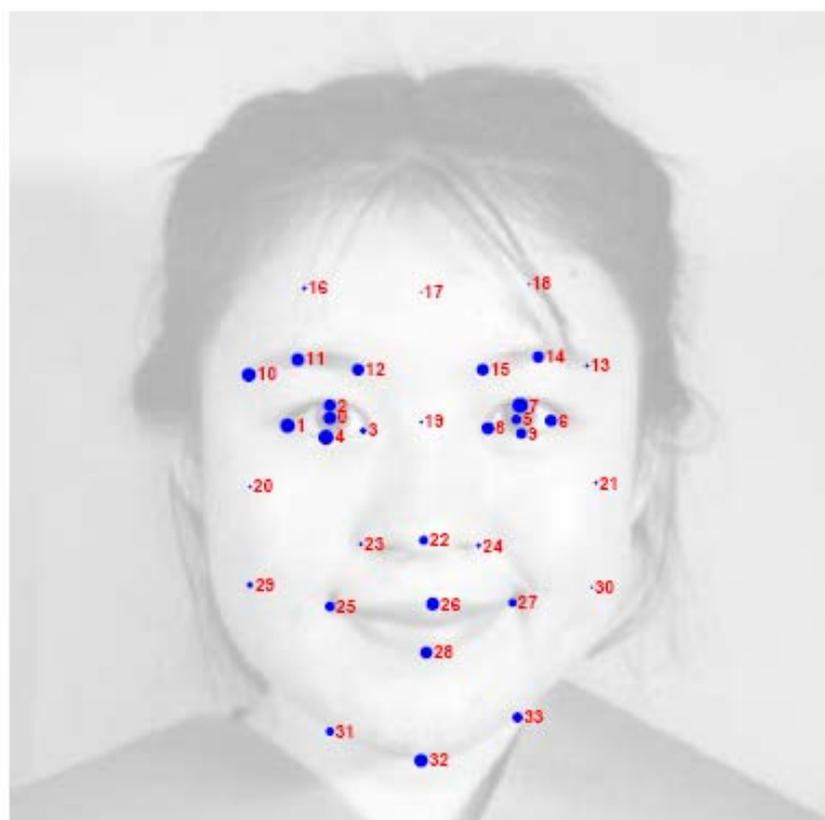


Figure III.21. Importance de chaque point fiduciel selon l'analyse de sensibilité. L'importance est illustrée par la taille du point.

III.6. Caractérisation du visage

III.6.1 Extraction de caractéristiques de la texture

La texture est une caractéristique importante de l'apparence des objets dans des scènes réelles et la comprendre est une partie essentielle de la compréhension de la vision humaine [64]. Une texture représente, à une échelle donnée, le même aspect quelle que soit la zone observée. Dans ces conditions, on considère l'image comme la réalisation d'un processus stochastique local et stationnaire. C'est-à-dire que chaque pixel est caractérisé par un petit voisinage et que cette caractérisation est la même pour tous les pixels de l'image.

De nombreuses approches de l'analyse de textures ou d'objets par attributs fréquentiels sont proposées : matrice de cooccurrence, dimension fractale, Transformation de Fourier, transformée en ondelettes, la pyramide laplacienne et la pyramide gaussienne. L'algorithme EBGGM que nous avons adopté dans notre travail utilise les ondelettes de Gabor pour l'extraction de caractéristiques de texture car il imite le fonctionnement de certaines cellules spécialisées, localisées dans le cortex visuel primaire humain [16].

III.6.2 Jet de Gabor

Le Jet est un ensemble de coefficients complexes d'ondelettes de Gabor. Il est calculé en convoluant l'image de visage autour d'un point repère (x, y) avec une collection de filtres de Gabor (Figure III.13). Ces coefficients sont ainsi générés en utilisant des ondelettes à différentes fréquences, différentes orientations, et différentes tailles.

En conséquence, les jets de Gabor présentent une bonne description de l'information fréquentielle au voisinage du point repère (x, y) .

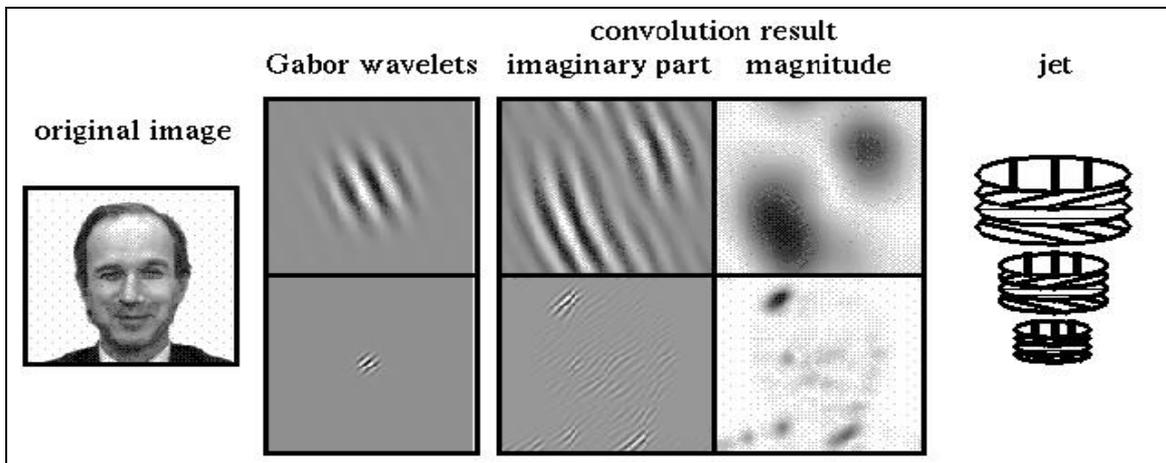


Figure III.19. Exemple de convolution d'une image de visage avec ondelette de Gabor.

Le jet est ainsi représenté par :

$$G_j = a_j \exp(i\Phi) \quad \dots\dots\dots \text{III.15}$$

Avec une amplitude a_j qui varie peu avec la position et une phase Φ_j dont la variation en rotation est plus importante.

III.6.3 Représentation des visages

Etant donné les points repères sont extraits, l'algorithme EBGGM procède à caractériser chacun de ces points en termes d'une portion en niveau de gris voisinant ce point de repère. Pour une configuration standard [61] [65], les jets de Gabor sont à base de 40 ondelettes complexes où chaque ondelette comprend une partie réelle et une partie imaginaire. On utilise pour ces jets huit (08) orientations (θ), cinq (05) fréquences (λ) et deux (02) phases (ϕ).

Les ondelettes de Gabor comme nous les avons défini sont des ondes sinusoïdales avec une fréquence et une orientation particulière modulée par une enveloppe gaussienne. [64] :

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{\left(-\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}\right)} \left[e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}} \right] \dots\dots\dots\text{III.16}$$

Où

$$\vec{k}_{u,v} = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos \theta_u \\ k_v \sin \theta_u \end{pmatrix}, \quad k_v = \frac{f_{\max}}{2} \frac{v}{2}$$

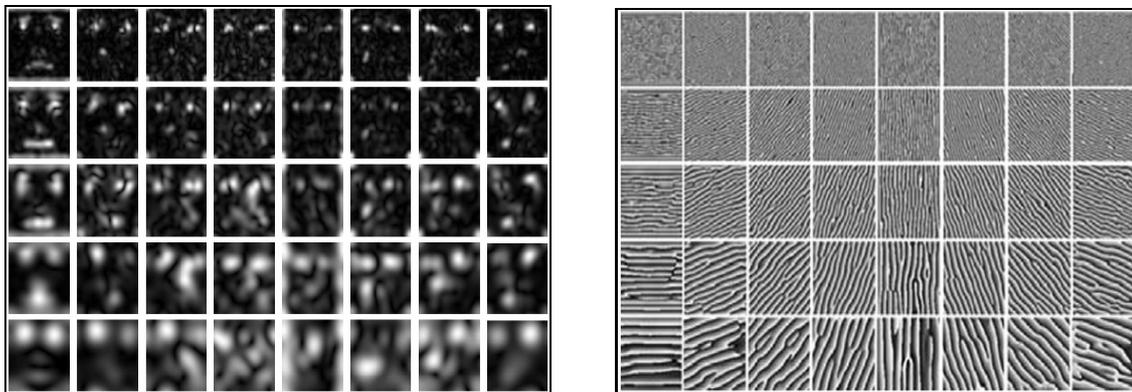
$$\Phi_u = u \left(\frac{\pi}{8} \right), \quad \sigma = 2\pi, \quad f_{\max} = \pi, \quad z = (x, y)$$

Ici v est la fréquence (échelle) entre 0 et 4, u est l'orientation entre 0 et 7

La transformation de Gabor est la convolution complexe d'une image par un filtre de Gabor :

$$G_{u,v}(z) = I(z) * \Psi_{u,v}(z) \dots\dots\dots\text{III.17}$$

Les résultats de cette transformation sont illustrés sur la figure III. 20.



(a) (b)
Figure III.20. Transformation de Gabor: (a) Module et (b) Phase.

III.6.4 Calcul de similarité

Etant donnée une image de visage test à reconnaître, l'algorithme EBG va effectuer une mesure de similarité entre cette image test et les différentes images de la base d'apprentissage [16][65].

Calculer la similarité entre l'image test et une image de la base d'apprentissage revient à déterminer la ressemblance entre leurs points repères correspondants. La similarité de jets de Gabor est un bon indicateur de similarité de points repères [61]. Trois mesures de similarité sont possibles que nous allons les présenter ci-après.

1. Similarité de phase (S_ϕ)

$$S_\phi(j, j') = \frac{\sum_{j=1}^N a_j a'_j \cos(\phi_j - \phi'_j)}{\sqrt{\sum_{j=1}^N a_j^2 \sum_{j=1}^N a'_j{}^2}} \dots\dots\dots \text{III.18}$$

S_ϕ : Appelée similarité de phase, elle est similaire à la corrélation,
 N : Nombre de coefficients d'ondelettes dans le jet.

2. Similarité d'amplitude (S_a)

Cette fonction mesure la similarité de l'énergie des différentes réponses fréquentielles. Le déphasage ϕ n'est pas pris en compte. Le résultat est une mesure de similarité fondée sur la covariance des amplitudes.

$$S_a(j, j') = \frac{\sum_{j=1}^N a_j a'_j}{\sqrt{\sum_{j=1}^N a_j^2 \sum_{j=1}^N a'_j{}^2}} \dots\dots\dots \text{III.19}$$

Cette méthode de similarité présente des tolérances aux petits déplacements des points repères, et elle est invariable aux changements de phases. Cette mesure de similarité qu'on va adopter dans notre travail.

3. Similarité conjointe

Cette mesure de similarité se base simultanément sur l'amplitude et la phase des coefficients, elle présente aussi la possibilité de compenser les petites différences de phases.

$$S_D(j, j', \vec{d}) = \frac{\sum_{j=0}^N a_j a'_j \cos(\phi_j - (\phi'_j + \vec{d} \cdot \vec{k}_j))}{\sqrt{\sum_{j=0}^N a_j^2 \sum_{j=0}^N a'_j{}^2}} \dots\dots\dots \text{III.20}$$

\vec{d} : Vecteur déplacement, différentes méthodes proposées pour le déterminer.

III.7. Applications de l'algorithme EBGM

Motivés principalement par le fait que le filtrage par ondelettes de Gabor est un processus présent dans le système de vision humain, Lyons, Akamatsu, Kamachi et Gyoba [74] ont démontré qu'il est possible de construire un système de reconnaissance des expressions uniquement basé sur des images filtrées par des ondelettes de Gabor. Ils tentaient de classer les expressions du visage parmi les six classes universelles.

Zhang [66] compare alors l'approche par filtres de Gabor aux approches classiques basées sur des mesures de composantes. Le but est de mesurer la qualité de l'information véhiculée par ces deux approches pour la reconnaissance des expressions universelles. Les deux informations sont données en entrée à un perceptron. Les résultats de la comparaison indiquent qu'une image filtrée par ondelettes de Gabor est beaucoup plus porteuse d'informations, pour la reconnaissance des expressions, que des mesures géométriques.

Dans [67] à des fins de reconnaissance de visages, une variante d'EBGM a été fusionnée avec un perceptron multicouche entraîné par l'algorithme de rétro propagation de l'erreur. Les tests ont été réalisés sur deux bases de visages: CMU et PIE en comparant l'approche proposée (84.50%) avec l'algorithme EBGM classique (75.29%) et la méthode dite «Log-Polar Gabor» (77.38%).

Les coefficients du filtre de Gabor ont été également fusionnés avec l'algorithme Ada-Boost dans [79]. Les tests ont été menés sur des images de visages en niveau de gris issues des deux bases XM2VTS et ORL. Les taux de reconnaissance atteints sont respectivement **92,2 %** pour la base XM2VTS et **97 %** pour la base ORL.

Paola Campadelli, et Raffaella Lanzarotti [80] ont appliqué l'approche EBGM sur un échantillon de 200 personnes extraites de la base FERET, et ont montré que cette approche atteignant un taux de reconnaissance de **96 %**, présente une grande robustesse aux variations de pose et aux expressions faciales.

Les filtres de Gabor ont été également appliqués pour extraire des caractéristiques faciales (les contours). Les résultats ont été comparés avec ceux obtenus avec Canny [81].

III.8. Conclusion

Ce chapitre a été consacré à présenter une méthode de reconnaissance de visages qui tient ses origines de la biologie c'est l'Elastic Bunch Graph Matching (EBGM). Elle imite le fonctionnement du système visuel humain en traitant l'information fréquentielle autour des points caractéristiques du visage en se basant sur les ondelettes 2D de Gabor entièrement paramétrées (elle ne traite pas directement les valeurs de niveaux de gris des pixels d'une image de visage), ce qui lui confère une grande robustesse aux changements d'éclairage, de pose et d'expressions faciales. Nous avons détaillé ses origines en allant de la transformée de Fourier jusqu'au filtre de Gabor 2D et expliqué son principe avant d'arriver à son application dans le domaine de la reconnaissance de visages. En outre, nous avons étudié l'importance des différents points fiduciels et leur impact sur l'amélioration des performances du système de reconnaissance. Notre contribution porte donc sur l'application de l'algorithme EBGM à des fins de reconnaissance de visages. Les résultats de simulations feront l'objet du prochain chapitre.

Chapitre IV

SIMULATIONS & RESULTATS

IV.1. Introduction

Notre contribution consiste à mettre en œuvre un système d'identification de visages basé sur une approche locale dite EBGM (Elastic Bunch Graph Matching) au sens où elle utilise des caractéristiques locales. Pour cela, le système passe par les trois étapes suivantes :

- 1) localisation des différentes caractéristiques du visage (yeux, nez, la bouche...), à partir de ces dernières, nous déterminons les différents points fiduciels dit également points repères comme les iris, les coins des yeux, la pointe du nez, les commissures de la bouche, etc.
- 2) application des ondelettes 2D de Gabor sur les différents points fiduciels localisés précédemment pour obtenir ainsi un vecteur de coefficients d'ondelettes complexes de Gabor appelé Jet, et
- 3) la reconnaissance d'un visage requête qui s'effectue en deux étapes: i) extraction de la signature à l'aide de la même technique que celle appliquée sur la base d'apprentissage, et ii) mesure de similarité entre les différents jets de l'image requête (obtenus aux différents points fiduciels) et ceux de la base d'apprentissage. La reconnaissance est obtenue en maximisant cette fonction de similarité.

IV. 2. Base de visages utilisés

La base de données utilisée pour évaluer notre système de reconnaissance automatique de visages est la base B-LAIG élaborée entre décembre 2003 et mars 2004 au sein du laboratoire d'Automatique et Informatique de l'université de Guelma [66]. Cette base de données de 100 personnes (50 hommes et 50 femmes) est de taille totale de 40,08 GB. Nous avons travaillé sur un échantillon de 36 personnes (18 hommes et 18 femmes) comprenant chacune une séquence vidéo de 10 images faciales avec meilleur éclairage (voir Figure IV.1).

Stockage et nomenclature

Les images sont de taille 174 x 144 pixels à 24 bits, stockées et référencées comme suit :

- « H-xx-yz-w.jpg » pour les hommes,
- « F-xx-yz-w.jpg » pour les femmes

Où

'xx' représente le numéro de la personne (de «00 » à «20 » pour les deux sexes),

'y' présente la session (0 pour la session de décembre et 1 pour la session de mars),

'z' présente l'état de la personne (z = 1 : 8 selon les différentes positions et différents états d'illumination avec et sans sourire),

'w' le numéro de l'image (w = 0 : 9 car 10 images sont extraites de chaque état).



Figure IV.1. La base de visage B-LAIG utilisée.

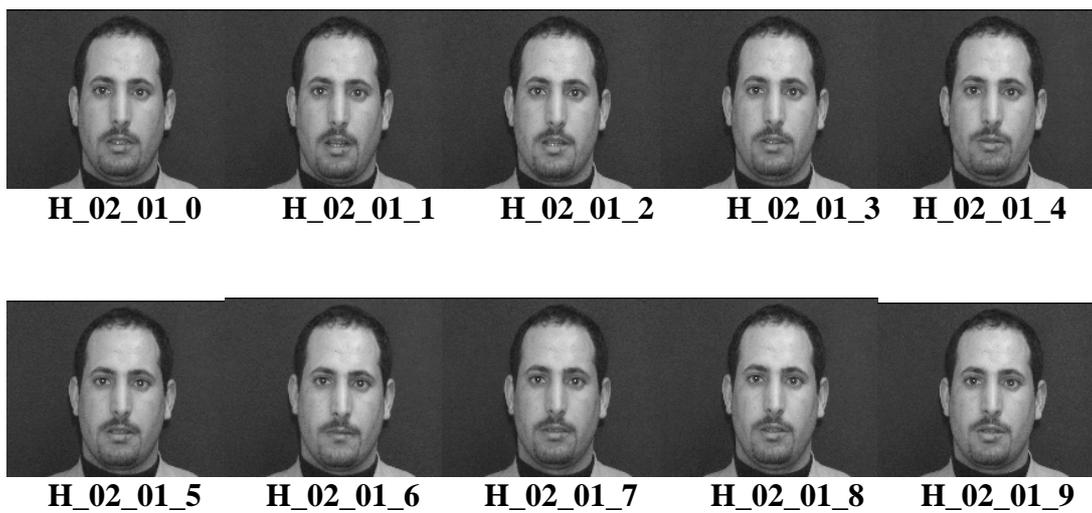


Figure IV.2. Exemple de séquence vidéo de la personne N°02.

IV.3 Système de reconnaissance de visages

Les organigrammes des figures IV.3 et IV.4 représentent, respectivement, la phase d'apprentissage et celle de test de notre système de reconnaissance faciale et dont les différentes étapes seront détaillées dans les paragraphes qui suivent.

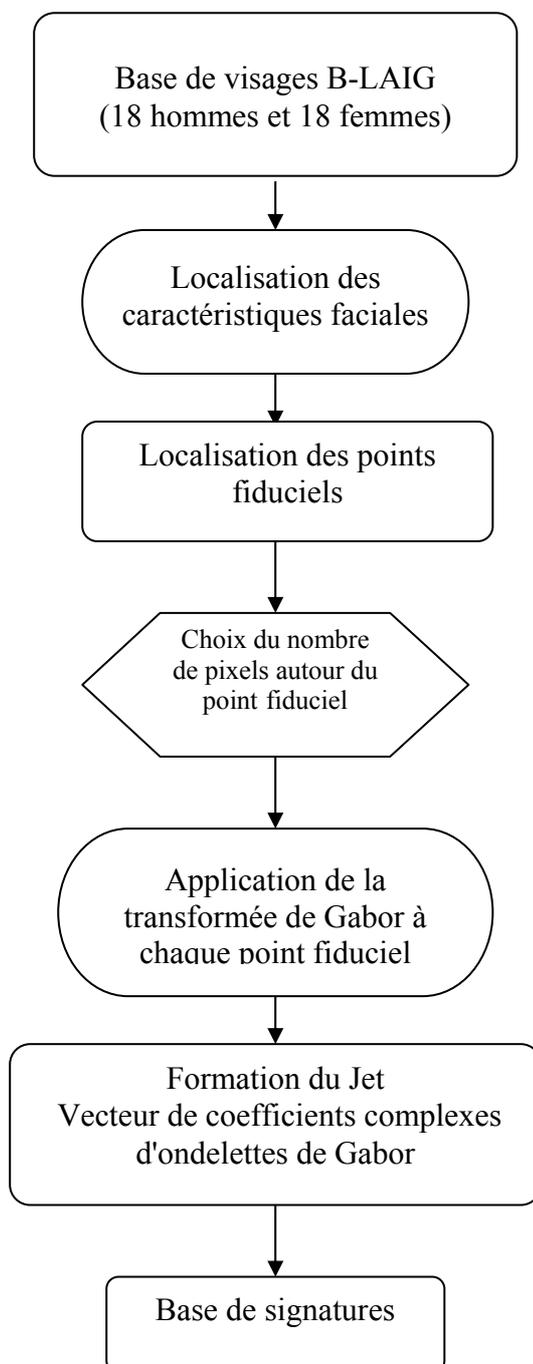


Figure IV.3. Organigramme de la phase d'apprentissage.

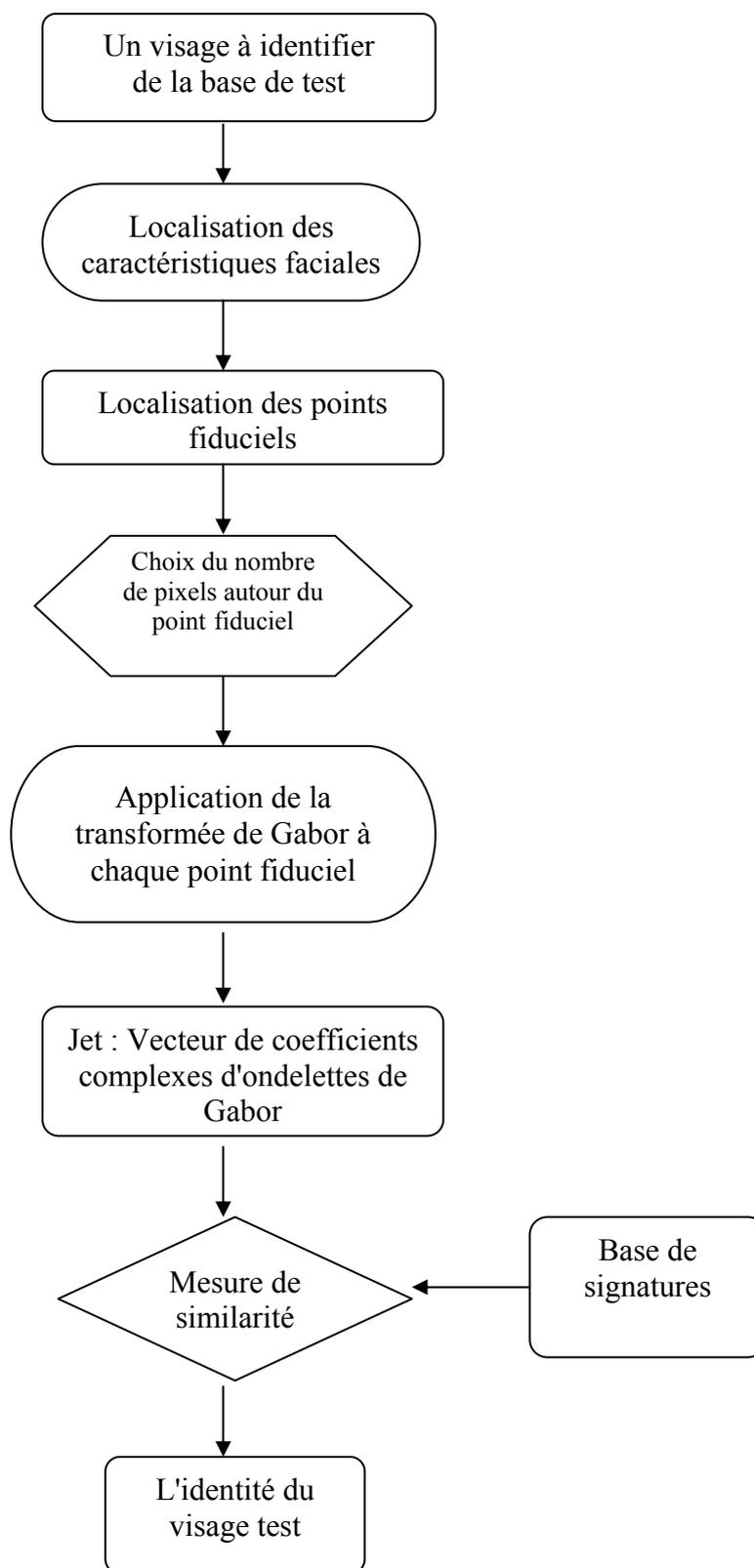


Figure IV.4. Organigramme de la phase de test.

IV.3.1 Redimensionnement des images de visages

Puisque nous adoptons une approche basée sur les caractéristiques locales et nous disposons d'une base de visages dont l'arrière plan est noir, nous avons pensé à redimensionner toutes les images pour réduire le temps de calcul ainsi que la capacité mémoire utilisée. Pour cela, nous avons procédé à un réajustement des images de visages en réduisant la taille de l'image de (174 x 144) pixels à une taille de (126x131). La figure IV.5 donne des exemples d'images réajustées.

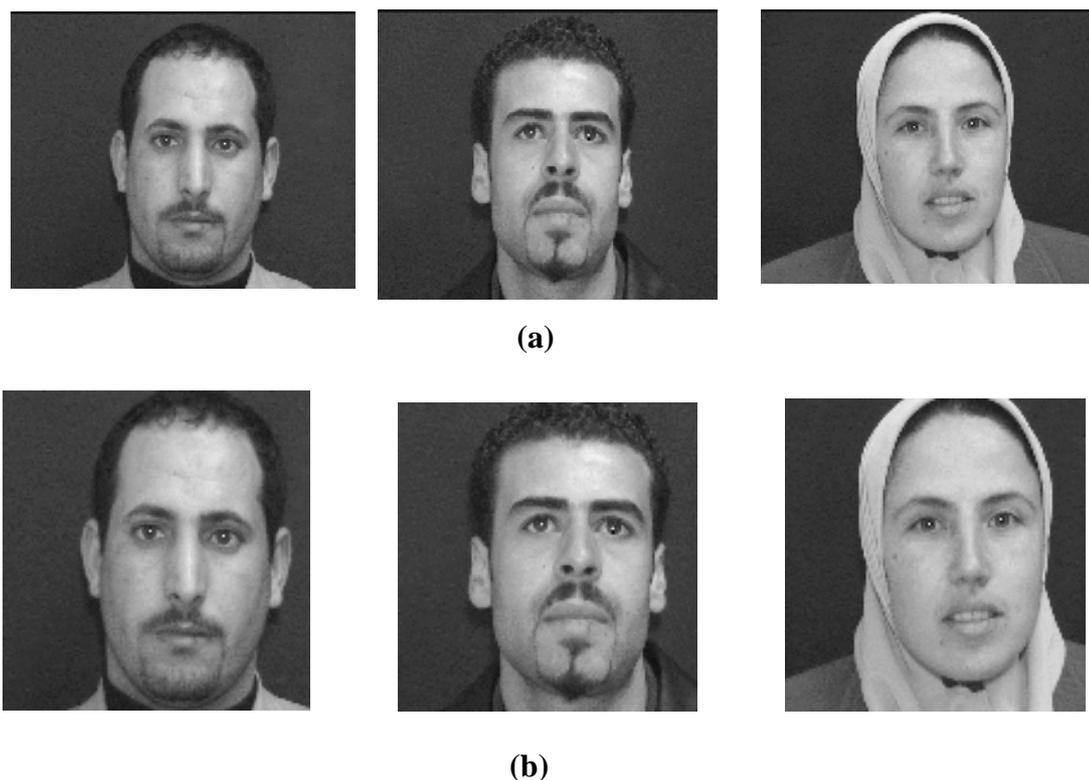


Figure IV.5. Exemples d'images de la base B-LAIG
(a) images originales de tailles (174 x 144) pixels et (b) images réajustées de tailles (126 x 131) pixels.

IV.3.2 Localisation des éléments caractéristiques du visage

1. Localisation des iris

Pour localiser les iris, nous avons adopté la méthode de [19]. L'existence d'une tache de réflexion spéculaire, due aux sources lumineuses frontales sur l'iris nous permet de localiser les yeux en effectuant un balayage de toute l'image du visage de haut en bas en vue de rechercher les maxima de la fonction d'intensité de niveau de gris (selon l'organigramme présenté dans la figure IV.6).

L'application de cette méthode sur la base de visages nous a fourni des résultats satisfaisants et nous a permis de localiser avec précision les iris des yeux comme l'illustre la figure IV.7.

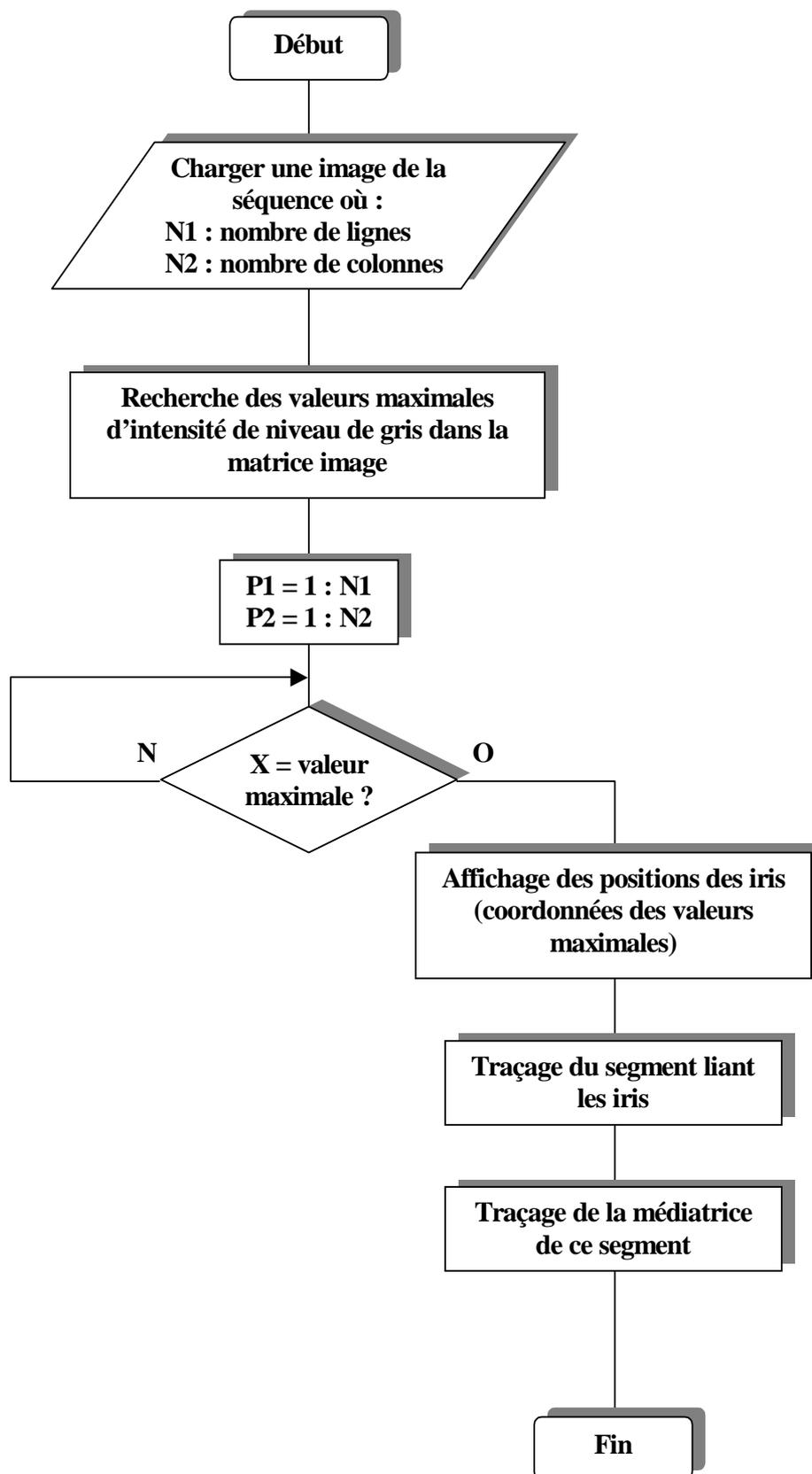


Figure IV.6. Organigramme de localisation des iris des yeux [49][78].

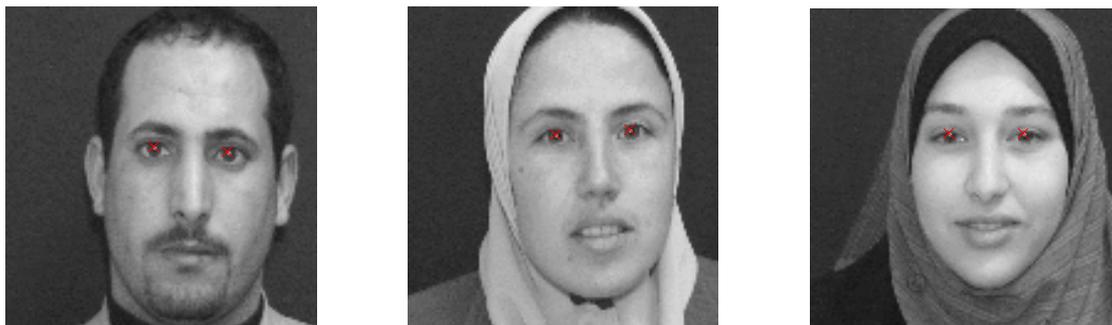


Figure IV.7. Exemples de localisation des iris.

2 Localisation du nez et de la bouche

Pour localiser le nez et la bouche, nous utilisons l'information spatio-temporelle de façon à caractériser les points ayant enregistré la plus forte variation d'intensité au cours de la séquence [19]. L'image I_{som} ainsi obtenue vérifie :

$$I_{som} = \sum_{i=0}^{i=n} \sqrt{\left(\left(\frac{\partial I(x, y, t)}{\partial x} \right)^2 + \left(\frac{\partial I(x, y, t)}{\partial y} \right)^2 + \mu \left(\frac{\partial I(x, y, t)}{\partial t} \right)^2 \right)}$$

$I(x, y, t)$: L'intensité du niveau de gris.

x, y : Les composantes spatiales.

t : La composante temporelle

μ : Un terme de normalisation dépendant de l'échantillonnage temporel de la séquence et du mouvement du visage.

n : Le nombre des images de la séquence.

La localisation du nez et de la bouche consiste à chercher le long de la médiatrice passant par les iris, les deux régions ayant les plus fortes valeurs sur l'image I_{som} . Leurs positions correspondent à la partie inférieure du nez et au centre de la bouche.

Les étapes de cette méthode de localisation sont détaillées sous forme d'un organigramme donné à la figure IV.8.

Quant aux résultats de localisation du nez et de la bouche, la figure IV.9 illustre quelques exemples.

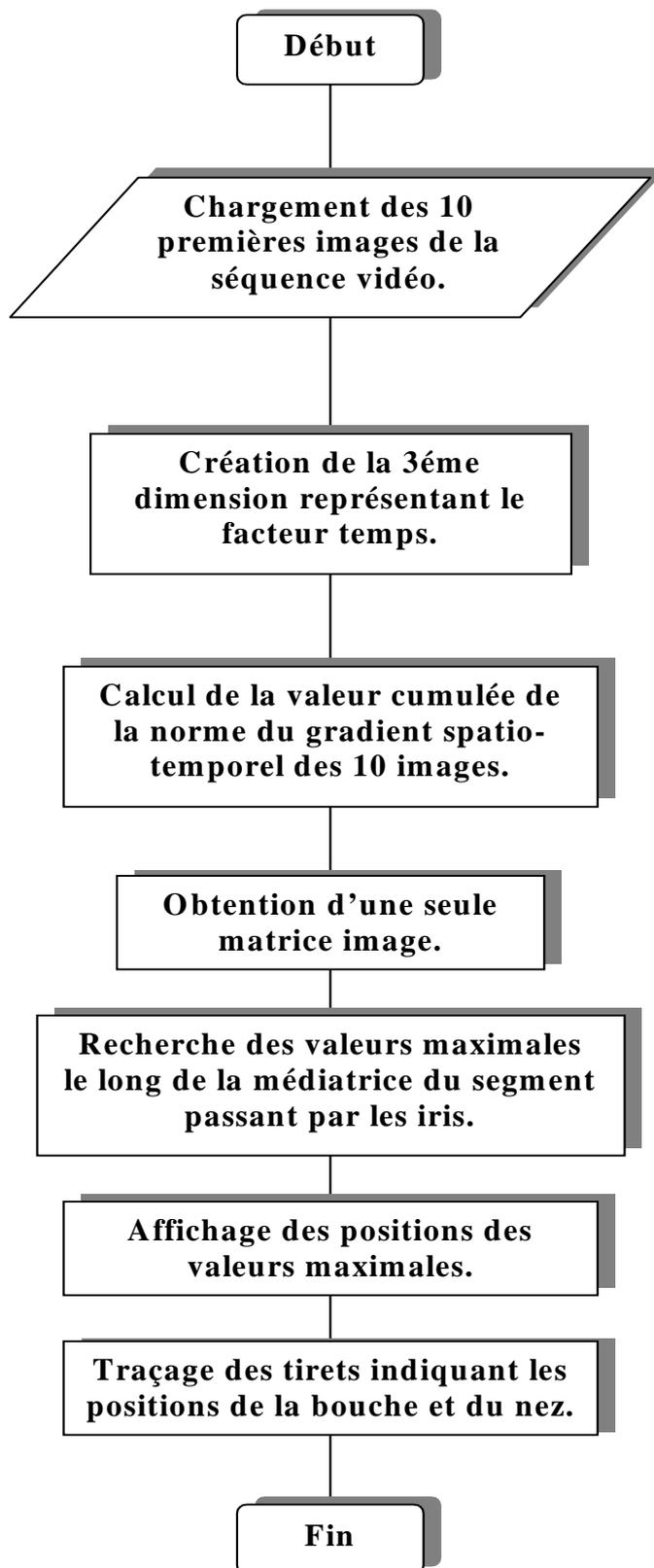


Figure IV.8. Organigramme de localisation du nez et du centre de la bouche [49][78].



Figure IV.9. Exemples de localisation du nez et du centre de la bouche.

IV.3.3. Localisation de sept points fiduciels

Une fois que les caractéristiques du visage (les yeux, le nez et la bouche) sont localisées, nous allons procéder à la localisation des différents points fiduciels (les deux iris, le point milieu entre les deux iris, la pointe du nez, les deux extrémités latérales du nez, le point milieu de la bouche) [51] [77].

- Les deux iris: I_1 et I_2
- Le point M médiatrice entre les deux iris: ce point est déterminé en prenant comme coordonnées (x_m, y_m) :

$$x_m = \frac{x_1 + x_2}{2}; \quad y_m = \frac{y_1 + y_2}{2}$$

- Le point B milieu de la bouche,
- La pointe du nez N en prenant le milieu entre le point M et N,
- Les deux extrémités latérales du nez : L_1 et L_2 :

L'extrémité latérale gauche du nez est déterminée en faisant l'intersection de la droite horizontale passant par la pointe du nez (point N) et la droite reliant l'iris gauche (point I_1) avec le point milieu de la bouche (point B).

L'extrémité latérale droite du nez est déterminée en faisant l'intersection de la droite horizontale passant par la pointe du nez (point N) et la droite reliant l'iris droit (point I_2) avec le point milieu de la bouche (point B).

Dans notre travail, nous aurons un total de sept (07) points fiduciels sur les quels nous allons appliquer les ondelettes 2D de Gabor. Quelques exemples de résultats de localisation des différents points fiduciels obtenus sont illustrés dans les figures IV.10 et IV.11.



Figure IV.10. Exemples de localisations correctes de points fiduciels.

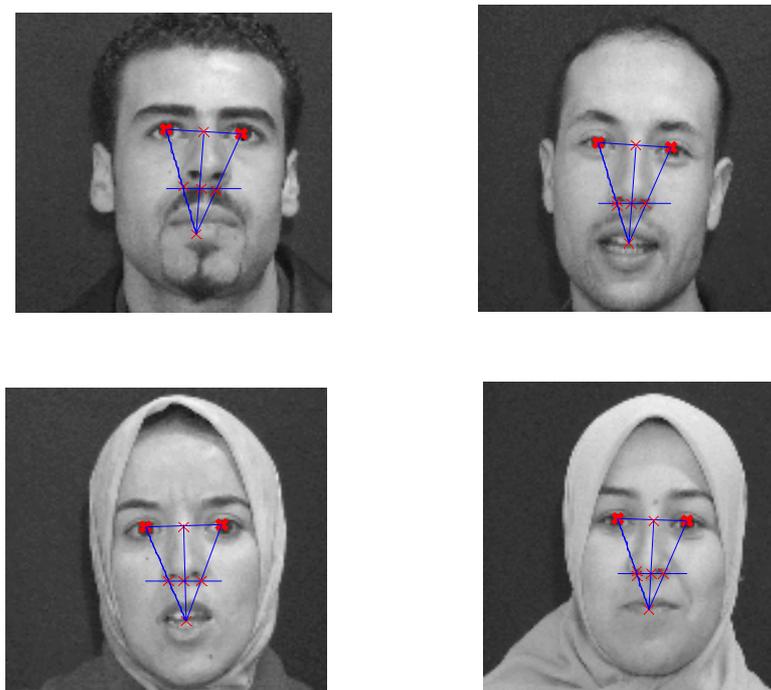


Figure IV.11. Exemples de mauvaises localisations des points fiduciels.

IV.3.4. Représentation des visages

Une fois les points repères (points fiduciels) sont localisés, et afin de représenter les visages on procède à convoluer l'image du visage autour de ces points avec plusieurs ondelettes complexes 2D de Gabor paramétrées (banc de filtres).

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{\left(-\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2} \right)} \left[e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}} \right]$$

Avec

$$\begin{aligned} \vec{k}_{u,v} &= \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos \theta_u \\ k_v \sin \theta_u \end{pmatrix} \\ k_v &= \frac{f_{\max}}{2 \frac{v}{2}} \quad \Phi_u = u \left(\frac{\pi}{8} \right) \\ \sigma &= 2\pi \quad f_{\max} = \pi \quad z = (x, y) \end{aligned}$$

En utilisant 5 fréquences différentes et 8 orientations indexées pour $v = 0 \dots 4$, et $u = 0 \dots 7$ respectivement, on aura un banc de filtres de Gabor composé de 40 filtres (figure IV.12). Nous obtenons comme résultat de la convolution de l'image de visage $I(z)$ autour d'un point fiduciel avec l'ensemble des filtres de Gabor 40 coefficients complexes formant ainsi un jet de Gabor et dont on a pris le module de ces coefficients.

$$G_{u,v}(z) = I(z) * \Psi_{u,v}(z)$$

Nous avons appliqué cette méthode pour l'ensemble des points fiduciels du visage, nous obtenons ainsi un vecteur caractéristique du visage composé de jets ($40 \times P$) où P est le nombre total des points fiduciels (dans notre cas $P=7$).

IV.3.5 Mesure de similarité

Pour identifier une image de visage test I , nous mesurons la similarité entre son vecteur de jets (40×7) et l'ensemble des vecteurs G_i de jets des différentes images de la base d'apprentissage. L'image test I sera affectée au G_i qui maximise cette fonction de similarité. On définit la mesure de similarité entre deux vecteurs de jets comme étant la moyenne des similarités entre les paires de jets correspondants.

$$S_v(V^1, V^2) = \frac{1}{N} \sum_{n=1}^N S(J_n^1, J_n^2)$$

Où N représente le nombre de Jets (Dans notre cas $N=7$)

$$S(J_n^1, J_n^2) = \frac{\sum_i J_i^1 J_i^2}{\sqrt{\left(\sum_i (J_i^1)^2 (J_i^2)^2\right)}} \quad i = (0 \dots, 39)$$

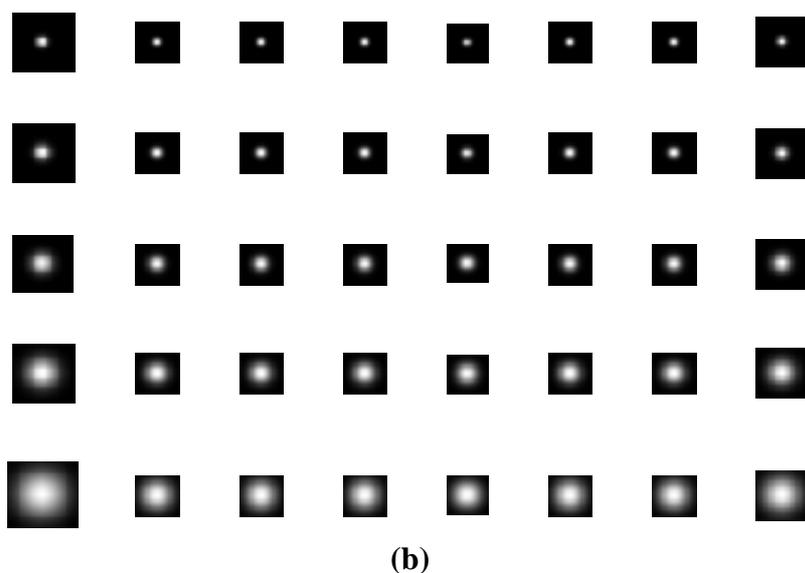
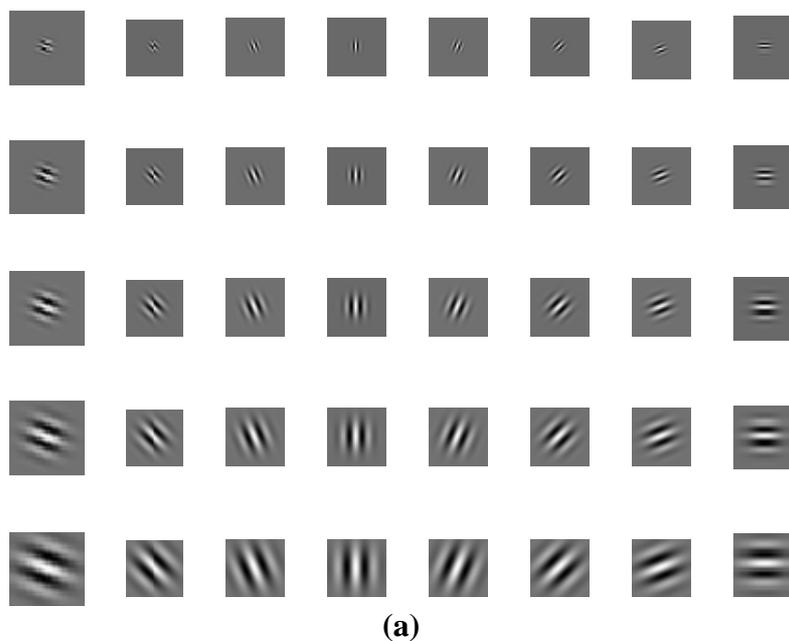


Figure IV.12. Banc de filtres généré par 5 différentes fréquences et 8 différentes orientations: (a) partie réelle et (b) le module.

IV.6. Résultats de la reconnaissance

La base de données que nous avons utilisé contient 36 personnes (18 hommes et 18 femmes) ayant chacune 10 images soit un total de 360 images de visages. Pour chaque visage on prend 07 images pour la base d'apprentissage et 03 images pour la base de test, soit au total une base d'apprentissage contenant **252** images (7x 36 visages), et une base de test contenant **108** images (3x36 visages).

Afin de mettre en valeur les performances du système proposé, nous avons divisé notre travail de simulation en trois phases :

- Réaliser des tests de reconnaissance sur la base des visages des hommes seuls.
- Réaliser des tests de reconnaissance sur la base des visages des femmes seules.
- Enfin, réaliser des tests de reconnaissance sur la totalité de la base de visages du LAIG " B-LAIG" utilisée.

En outre, pour évaluer l'influence et l'importance de la taille du voisinage autour des points fiduciels, nous avons effectué différents tests en fonction du nombre de pixels autour des sept points que nous avons localisé. Nous avons appliqué la convolution de chaque point fiduciel avec l'ensemble des 40 filtres 2D de Gabor:

1. Le point fiduciel seul sans voisinage.
2. Un voisinage d'un pixel soit une imagerie 3x3 pixels
3. Un voisinage de deux pixels, soit une imagerie 5x5 pixels
4. Un voisinage de trois pixels, soit une imagerie 7x7 pixels
5. Un voisinage de quatre pixels, soit une imagerie 9x9 pixels
6. Un voisinage de cinq pixels, soit une imagerie 11x11 pixels
7. Un voisinage de six pixels, soit une imagerie 13x13 pixels

Le tableau ci-dessous récapitule les différents taux de reconnaissance obtenus en fonction du nombre de pixels voisins pour la base des hommes et celle des femmes séparément.

Nombre de pixels voisins (N)	Taille de l'imagerie	Taux de reconnaissance (%)	
		Hommes	Femmes
0	1x1	47,5	42,65
1	3x3	53,75	39,71
2	5x5	71,25	51,47
3	7x7	70,00	48,53
4	9x9	73,75	42,65
5	11x11	73,75	55,88
6	13x13	76,25	57,35
7	15x15	72,40	52,94

Tableau IV.1. Taux de reconnaissance en fonction du nombre de pixels voisins.

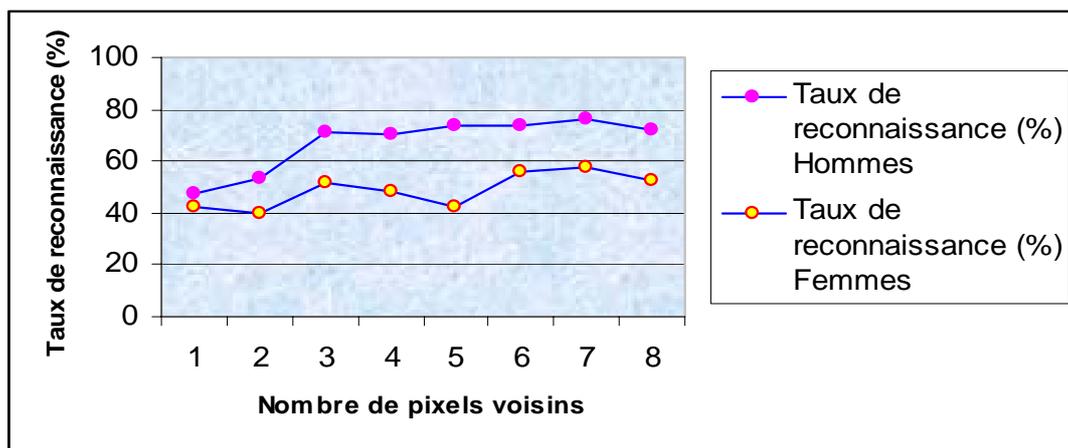


Figure IV.13. Evolution du taux de reconnaissance en fonction du nombre de pixels voisins pour 7 points fiduciels.

Nous constatons que pour un nombre de 07 points fiduciels nous avons obtenu un taux de reconnaissance **de 76,25 %** pour les hommes seuls, et un taux de **57,35 %** pour les femmes seules. Ces valeurs coïncident avec un nombre de pixels voisins $N=6$ soit une taille de l'imagette de (13 x 13) pixels. On remarque aussi que le taux de reconnaissance obtenu pour les hommes est nettement supérieur à celui obtenu pour les femmes.

Les figures IV.14 et IV.15 illustrent quelques résultats de reconnaissance réalisés par notre système proposé.

Image test : visage N° 01



Image similaire: visage N° 1



Image test : visage N° 02



Image similaire: visage N° 2



Image test: visage N° 09



Image similaire: visage N° 9



Image test: visage N° 03

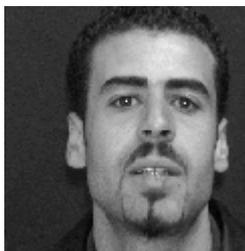


Image similaire: visage N° 3

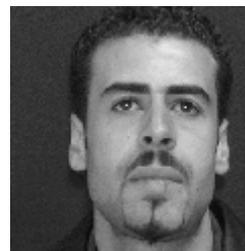


Figure IV.14. Exemples de visages test identifiés.

Image test : Image (base de test) de la personne à identifier ou à reconnaître.

Image similaire : Image (base d'apprentissage) de la personne dont la valeur de similarité est maximale.

Image test : visage N° 04



Image similaire: visage N° 11



Image test : visage N° 05



Image similaire: visage N° 16



Image test : visage N° 07



Image similaire: visage N° 20



Image test : visage N° 05



Image similaire: visage N° 11



Figure IV.15. Exemples de visages test non identifiés.

Image test : Image (base de test) de la personne à identifier ou à reconnaître.

Image similaire : Image (base d'apprentissage) de la personne dont la valeur de similarité est maximale.

IV.7. Améliorations du taux de reconnaissance

Pour améliorer les performances de notre système, nous avons augmenté le nombre de points fiduciels à 13 points au lieu de 07 points. Pour cela, nous avons rajouté les quatre coins des yeux et les deux commissures de la bouche puis nous avons procédé à une **localisation manuelle** de ces treize points pour l'ensemble des visages de la base B-LAIG utilisée en utilisant la fonction **Imview** de Matlab. La figure VI.16 illustre quelques exemples de cette localisation.



Figure IV.16. Exemples de localisation de 13 points fiduciels.

1. Applications sur la base des hommes

Nous récapitulons, dans le tableau ci-dessous, les résultats de simulations menées sur la base des hommes en fonction du nombre de pixels voisins.

Nombre de pixels voisins (N)	Taille de l'imagette	Nombre total de visages test	Nombre visages reconnus	Taux de reconnaissance %
0	1x1	54	52	96,30
1	3x3	54	47	87,04
2	5x5	54	50	92,59
3	7x7	54	48	88,89
4	9x9	54	48	88,89
5	11x11	54	49	90,74
6	13x13	54	47	87,04
7	15x15	54	47	87,04
8	17x17	54	48	88,89

Tableau IV.2. Taux de reconnaissance en fonction du nombre de pixels voisins.

Nous constatons une nette amélioration du taux de reconnaissance de **76,25 %** à **96,30 %** pour 7 et 13 points respectivement.

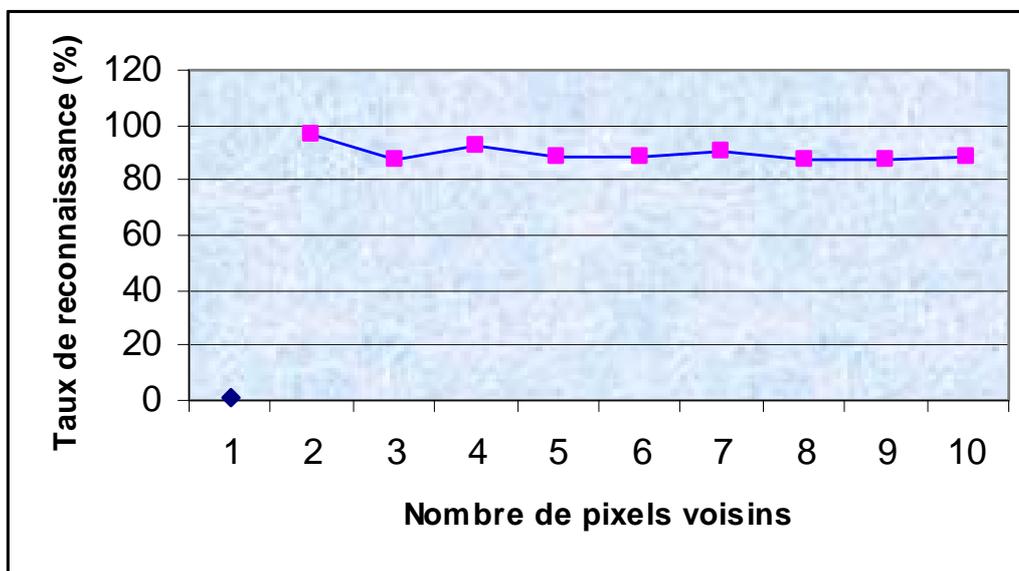


Figure IV.17. Evolution du taux de reconnaissance de la base hommes en fonction du nombre de pixels voisins pour 13 points fiduciels.

Un exemple de simulation d'une personne identifiée pour un voisinage de trois pixels (N=3) est illustré dans la figure ci-dessous.

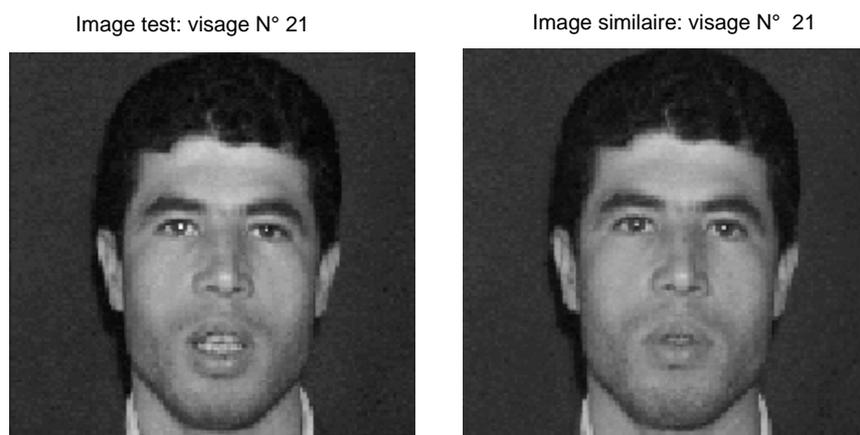


Figure IV.18. Exemple d'une personne identifiée avec un voisinage N=3.

2. Applications sur la base des femmes

Nous récapitulons, dans le tableau ci-dessous, les résultats de simulations menées sur la base des femmes en fonction du nombre de pixels voisins.

Nombre pixels voisins (N)	Taille de l'imagette	Nombre total visages test	Nombre de visages reconnus	Taux de reconnaissance %
0	1x1	54	43	79,63
1	3x3	54	30	55,56
2	5x5	54	25	46,30
3	7x7	54	22	40,74
4	9x9	54	18	33,33
5	11x11	54	24	44,44
6	13x13	54	21	38,89
7	15x15	54	29	53,70
8	17x17	54	35	64,81
9	19x19	54	36	66,67
10	21x21	54	31	57,41
12	25x25	54	38	70,37
13	27x27	54	38	70,37
14	29x29	54	34	62,96
15	31x31	54	35	64,81

Tableau IV.3. Taux de reconnaissance en fonction du nombre de pixels voisins.

Pour les images des femmes seules, le taux de reconnaissance qui était de **57,35 %** s'est amélioré à **79,63 %**, néanmoins ce taux reste insuffisant ceci est du principalement au nombre de points fiduciels qui reste toujours assez réduit. On remarque aussi que le taux maximum (79,63%) est obtenu pour les points fiduciels sans voisinage (N=0) vu que la localisation des différents points s'est réalisée d'une façon manuelle et que l'erreur de localisation est assez faible (un à deux pixels).

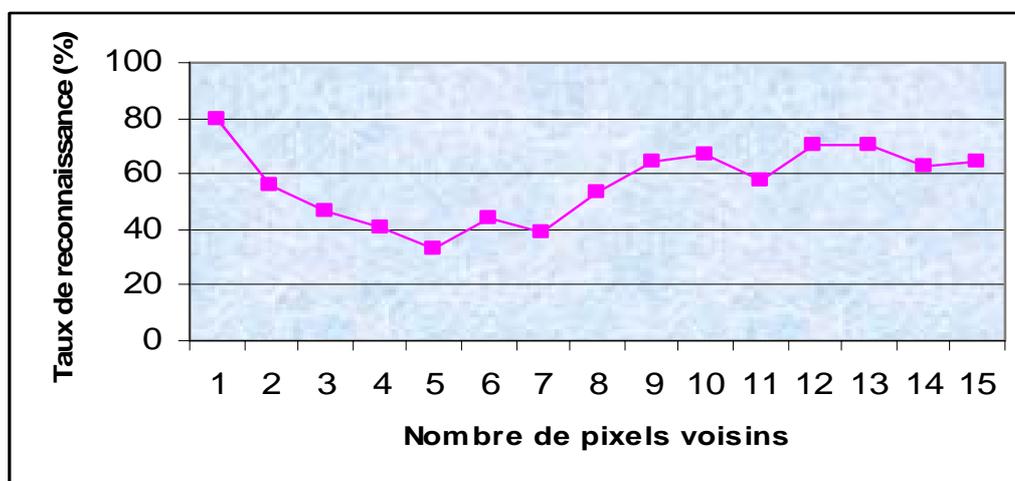


Figure IV.19. Evolution du taux de reconnaissance de la base femmes en fonction du nombre de pixels voisins pour 13 points fiduciels.

Un exemple de simulation d'une personne identifiée pour (N=0) sans pixels voisins autour des points fiduciels est illustré dans la figure ci-dessous.

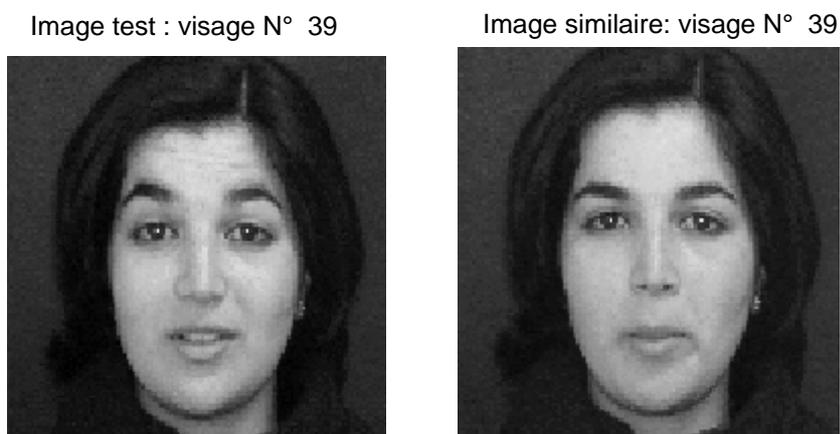


Figure IV.20. Exemple de personne identifiée pour un voisinage **N=0**

3. Applications sur la base de test complète (Hommes & Femmes)

Le tableau ci-dessous résume les résultats de simulations menées sur la base complète en fonction du nombre de pixels voisins.

Nombre pixels voisins (N)	Taille de l'imagette	Nombre total des visages test	Nombre visages reconnus	Ecart	Taux de reconnaissance (%)
0	1x1	108	87	21	80,56
1	3x3	108	62	46	57,41
2	5x5	108	68	40	62,96
3	7x7	108	67	41	62,04
4	9x9	108	63	45	58,33
5	11x11	108	70	38	64,81
6	13x13	108	67	41	62,04
7	15x15	108	79	29	73,15
8	17x17	108	83	25	76,96
9	19x19	108	85	23	78,70
10	21x21	108	82	26	75,93

Tableau IV.4. Taux de reconnaissance en fonction du nombre de pixels voisins.

Pour la totalité de la base des images de visages (108 images), nous avons atteint un taux de reconnaissance de **80,56 %** qui reste à notre avis insuffisant et que nous pouvons

toujours l'améliorer en augmentant d'abord notre base de visages, en augmentant le nombre de points fiduciels (au minimum 20 points) et enfin en améliorant la qualité des images.

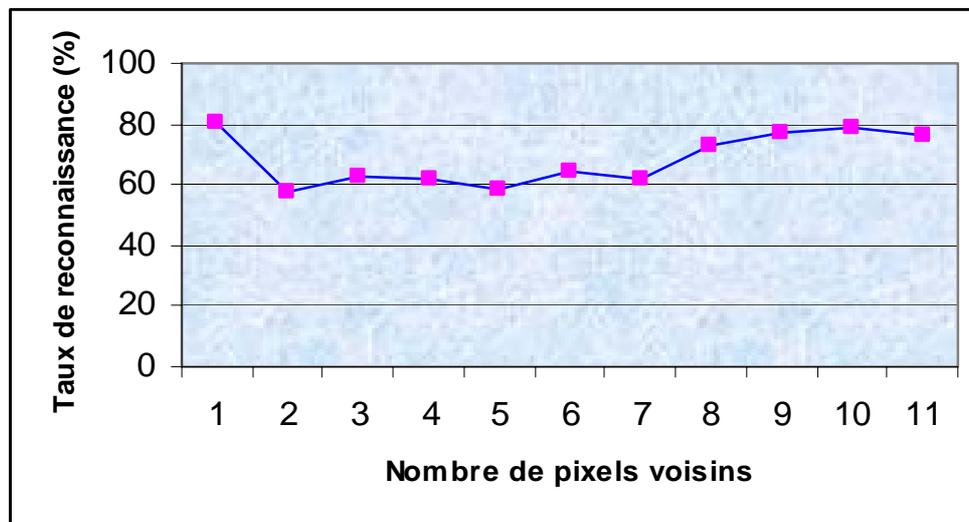


Figure IV.21. Evolution du taux de reconnaissance de la base totale en fonction du nombre de pixels voisins pour 13 points fiduciels.

Des exemples de visages test non identifiés pour différents voisinages sont illustrés dans la figure ci-dessous.

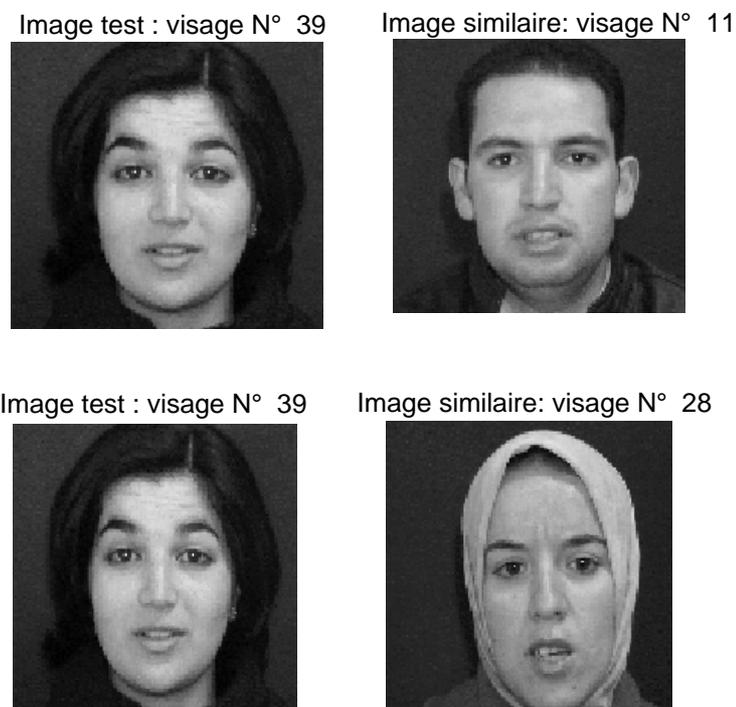


Figure IV.22. Exemples d'une personne non identifiée avec un voisinage N=2 et N=3.

IV.8 Conclusion

Nous avons proposé un système de reconnaissance automatique de visages en utilisant l'algorithme EBG. Les expérimentations sont menées sur la base de données B-LAIG où nous avons pris des échantillons de 36 personnes (18 hommes et 18 femmes). Chaque personne est représentée par dix images soit un total de 360 images. Les images initialement de taille (174 x 144) sont normalisées de manière à être centrées sur les visages et redimensionnées à une taille suffisante (126 x 131).

La base étant divisée en deux sous bases l'une servant pour l'apprentissage et l'autre pour le test. Notre algorithme se décompose en trois parties principales: localisation des caractéristiques du visage (yeux, nez et la bouche), la deuxième étape consiste à déterminer les points fiduciels sur les quels nous avons appliqué la transformée en ondelettes 2D de Gabor pour ainsi extraire les caractéristiques. La troisième étape est la classification basée sur une mesure de similarité entre l'image test et les différentes images de la base d'apprentissage.

Dans une première étape, nous avons procédé à une localisation automatique des caractéristiques du visage (yeux, nez et la bouche), puis nous avons déterminé sept points fiduciels sur les quels nous avons appliqué la transformée en ondelettes de Gabor afin d'extraire les caractéristiques du visage. Les résultats de simulations ainsi obtenus ont donné des taux de reconnaissance moyens (**57,35 %** pour les femmes et **76,25 %** pour les hommes). Ceci est dû essentiellement au nombre réduit des points fiduciels utilisé (07 points) et à leurs localisations automatiques qui n'est pas très précise.

Dans une deuxième étape, et afin d'améliorer les performances de notre système de reconnaissance, nous avons augmenté le nombre de points fiduciels à 13 points. Ces derniers ont été localisés manuellement afin d'assurer la précision. Les résultats de simulation ainsi obtenus révèlent une nette amélioration des différents taux de reconnaissance que ce soit pour les hommes (**96,30 %**) ou pour les femmes (**79,63 %**). Quant à la base complète, le meilleur taux est **80.56 %**.

Conclusion Générale

Le travail présenté dans ce mémoire s'inscrit dans le contexte de l'identification automatique de visages à partir de séquences d'images vidéo. Pour cela, la base de visages utilisée contient 360 images de 36 personnes de sexes différents. Elle a été acquise au LAIG en 2003/2004 dans des conditions d'illumination contrôlées avec des vues frontales et presque frontales et des expressions faciales dues à la prononciation de la phrase « reconnaissance automatique de visages ».

Notre travail est décomposé en deux grandes parties principales:

Une première partie consacrée à la localisation automatique des différentes caractéristiques du visage (les yeux, le nez et la bouche). La localisation des yeux se fait grâce à l'existence de la tache de réflexion spéculaire due aux sources lumineuses frontales sur l'iris où nous avons balayé la matrice image de haut en bas en vue de la recherche des deux régions possédant l'intensité maximale de niveau de gris. Pour la localisation de la bouche et du nez, nous avons utilisé l'information spatio-temporelle qui permet de caractériser les points ayant enregistré les plus fortes variations d'intensités au cours de la séquence. Une fois que les caractéristiques du visage sont localisées, nous avons procédé en premier lieu à une localisation automatique de sept (07) points fiduciels (les deux iris, le point médian entre les deux iris, la pointe du nez, les deux extrémités latérales du nez, le point milieu de la bouche), puis nous avons augmenté le nombre de ces points fiduciels à 13 points en ajoutant les quatre coins des yeux, et les deux commissures de la bouche. Ces points ont été localisés manuellement afin d'améliorer la précision de leurs localisation et par conséquent les performances de notre système de reconnaissance.

La deuxième partie de notre travail consiste à la représentation du visage, pour cela, nous avons appliqué à chacun de ces points fiduciels un banc de filtres de Gabor de différentes fréquences et différentes tailles (40 filtres) en effectuant la convolution de l'image du visage autour de ces points avec plusieurs ondelettes complexes 2D de Gabor paramétrées. Chaque point caractéristique du visage sera ainsi représenté par un vecteur de coefficients complexes de Gabor appelé jet. En fin la phase de reconnaissance se base sur une mesure de similarité entre les différents jets de Gabor d'un visage test et tous les jets d'apprentissage.

En conclusion, l'algorithme EBGM est un algorithme local qui ne traite pas directement les niveaux de gris des pixels d'une image de visage mais il applique des transformations en ondelettes de Gabor afin d'extraire l'information fréquentielle des différents points caractéristiques du visage. Cet aspect lui confère une grande robustesse aux changements d'éclaircements, de pose et d'expressions faciales [79]. Cependant, l'algorithme EBGM présente quelques inconvénients à savoir: i) il est difficile à implémenter, ii) nécessite une localisation précise des différents points fiduciels et iii) les vecteurs caractéristiques des coefficients de Gabor sont de tailles importantes ce qui rend la méthode assez coûteuse en temps de calcul.

En perspective, nous suggérons l'application de cette approche à des images en couleur, sur des bases de visages internationales, d'augmenter le nombre de points fiduciels et d'associer d'autres types de classificateurs.

Bibliographie

- [1] K.M. Aban. Le rôle de l'anthropologie dans les techniques émergentes de reconnaissance faciales. Université Claude Bernard –Lyon 1, France, Septembre 2005
- [2] J.D. Woodward, Jr.Christopher Horn, J. Gatune, and A. Thomas. Biometrics: A look at Facial Recognition, Published 2003 by Rand.
- [3] Technique de contrôle d'accès par biométrie, commission technique de sécurité physique, dossier technique publié par CLUSIF, Juin 2003
- [4] D. Mahmoudi, Biométrie et Authentification, Septembre 2000
- [5] M. Marzouki, Enjeux des techniques de biométrie. Troisième conférence internationale des commissaires à la protection des données, vie privée, droits de l'homme, septembre 2001. Paris, France
- [6] Internaute- Biométrie : L'identité par le corps, Juillet 2006
- [7] MEG/EEG et reconnaissance des visages
Nathalie George, CNRS UPR 640 - LENA, Hôpital de la Salpêtrière, Paris
- [8] P.J Phillips, P.Grother, R.J.Micheals, D.M. Black-Burn, E Tabassi et J.M. Bone. Face recognition vendor test 2002. Evaluation Report. Technical Report 6965, NIST, Mars 2003.
- [9] M.Visani, C.Garcia, J-M.Jolion. Une nouvelle méthode de représentation des visages pour leur reconnaissance: l'analyse discriminante bilinéaire
- [10] M.T.Laskri, D.Chefrour. Who Is. Système d'identification des visages humains, ARIMA, volume 1, pp.39-61,2002.
- [11] K.ACHOUR, N.ZENATI, H.LAGA. Contribution of restoration of degraded images. Actes du 4^{ème} Colloque Africain sur la Recherche en Informatique CARI'98 (INRIA). Dakar, Sénégal, 12-15 octobre 1998.
- [12] P. Charbonnier, B.FERAU L, G.AUBERT, M.BARLAU. Deterministic edge preserving regularisation in computer imaging. IEEE Transaction on Image Processing, Vol.6, Number 2, pp. 298, February 1997.
- [13] W.W. Bledsoe. The model method in facial recognition. Technical Report PRI: 15 Panoramic Research. Inc, Palo Alto, CA, 1966.
- [14] M.D.KELLY. Visual identification of people by computer. Technical Report AI-130, Stanford AI Project, Stanford CA, 1970.
- [15] Face Recognition from a Single Image per Person: A Survey Xiaoyang Tan 1,2 Songcan Chen 1,3,* Zhi-Hua Zhou 2 Fuyan Zhang 2 Department of Computer Science and Engineering Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China
- [16] N. Morizet, Thomas EA, F.Rossant, F.Amara." PCA, LDA et EBGm utilisés en reconnaissance 2D du visage pour la biométrie". Institut Supérieur d'Electronique de Paris (ISEP), Département d'Electronique. Revue des algorithmes PCA.
- [17] W.Zhao, R.Chellappa, A.Rosenfeld and P.J.Phillips. Face Recognition: A literature survey. ACM Computing Surveys, Vol.35, No.4, pp.399-458, December 2003.

- [18] J.R. BEVERIDGE, K.SHE, B.A. DRAPER, et G.H. GIVENS. A nonparametric statistical comparison of principal component and linear discriminant subspaces for face recognition. In proceeding of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Volume 1, pp.535-542, 2001.
- [19] B. Leroy, A. Chouakria, I.L.Herlin et E. Diday. Approche Géométrique et Classification pour la Reconnaissance de Visages, INRIA, France, 1996.
- [20] L.D.Harmon, M.K. Khan, R.Lasch and P.F. Raming. Machine identification of human faces. Pattern Recognition, 13(2): 97-110, 1981.
- [21] A.J.Goldstein, L.D.Harmon and A.B.Lesk. Identification of human faces. Proc.of the IEEE, 59(5): 748-760, 1971.
- [22] M.Visani, Vers de nouvelles approches discriminantes pour la reconnaissance de visages, thèse de doctorat, France, Novembre 2005
- [23] A. Danalet. Modèles de choix discrets pour la reconnaissance des expressions faciales statiques, Lausanne, mars 2007.
- [24] R.Brunnelly and T.Poggio "Face Recognition: Features versus Templates", IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(10): 1042-1052, 1993.
- [25] I.J. Cox, J. Ghosn and P.N. Yianilos. Feature-based recognition using mixture distance. In Proceedings, IEEE Conference on Computer Vision and pattern Recognition, pp.209-216, 1996.
- [26] A.Pentland, B.Moghaddam and T.Starner. View based and modular eigen spaces for face recognition. IEEE Conference on Computer Vision and Pattern Recognition, 1994.
- [27] X.Lu and A.K.Jain. Automatic feature extraction for multiview 3D face recognition. To appear in Proc. International Conference on Automatic Face and Gesture Recognition, Southampton, UK, April 2006.
- [28] Y.Moses, Y.Adini et S.Ullman. Face recognition: the problem of compensating for changes in Illumination direction. Proceedings of the European Conference on Computer Vision (ECCV), Volume 1, pp. 286-296, 1994.
- [29] D.M. Blackburn, M. Bone, et P.J. Philips " Facial Recognition Vendor Test 2000: Evaluation Report". Technical Report A269514, National Institute of Standards and Technology, 2001. 70 pages
- [30] O'Toole, A.J., Abdi.H. Low-dimensional representation of faces in higher dimensions of the face space. Optical Society of America, 10(3), 405-411, 1993.
- [31] I. Craw, H.Ellis and J.Lishman. Automatic Extraction of Face Features. Pattern Recognition Lettery, Vol.5, pp. 183-187, 1987.
- [32] N. Eveno, P. Delmas et P-Y. Coulon. Vers l'Extraction Automatique des lèvres d'un Visage Parlant. Colloque GRETSI sur le traitement du signal et des images (GRETSI 01) Toulouse, France, Septembre 2001.
- [33] N.Funabiki, M.Isogai, T.Higashino and M.Oda. An Eye Contour Extraction Algorithm From Face Image Using Deformable Template Matching. Memoirs of the Faculty of Engineering Okayama university, Vol.40, pp.78-82, January, 2006.
- [34] Yuille, D.S Cohen and P.W. Hallinan. Feature extraction from faces using deformable templates. Int.J. Comput. Vis.8, pp.99-112, 1992.
- [35] T. Cootes, C. Taylor, D. Cooper and J. Graham. Active Shape Models-their Training and Application. Comput. Vis. Image Understand. 61, 18-23, 1995.
- [36] T. F. Cootes, G.J. Edwards and C. J. Taylor. Active Appearance Models. Proc. European Conference on Computer Vision. Vol.2, pp. 484 - 498, Springer, 98
- [37] R.Gross, J.Shi, et J.F Cohn. Quo vadis face recognition? The current State of the

- Art in Face Recognition. Technical Report CMU-RI-TR-01-17, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 2001.
- [38] T.Kanade, J.F Cohn et Y.L Tian. Comprehensive database for facial expression analysis. Proceeding of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR) pp. 46-53, 2000.
 - [39] P.J. Philips, P. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi et J.M.Bone. Facial recognition vendor test 2002: Evaluation report. Technical Report 6965, National Institute of Standards and Technology, 2003.
 - [40] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET Database and Evaluation Procedure for Face Recognition Algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.
 - [41] Handbook of face recognition / editors, Stan Z. Li & Anil K. Jain. Includes bibliographical references and index. TA 1650. H36. 2004
 - [42] K.I.Chang, K.W Bowyer et P.J. Flynn " An Evaluation of Multi-Modal 2D+3D Face Biometrics", *IEEE Transactions on pattern Analysis and Machine Intelligence*, 27 (4): 619-624, Avril 2005
 - [43] M.A. Turk et A.D. Pentland " Eigenfaces For Recognition" *Journal of Cognitive Neuroscience*, 3:71-86, 1991
 - [44] H.Moon et J.Phillips "Analysis of PCA- Based Face Recognition Algorithms" Dans K. Boyer et P.J Phillips, éditeurs *Empirical Evaluation Techniques in Computer Vision*, pages 835-855. IEEE Computer Society Press,1998
 - [45] D.L. Swets et J.J Weng "Using Discriminant Eigenfeatures for Image Retrieval ". *IEEE transactions on Pattern Analysis and Machine Intelligence*, 18(8): 831-836, 1996.
 - [46] M. Kirby *Dimensionality Reduction and Pattern Analysis: An Empirical Approach*. Wiley, New York, 2000
 - [47] P.N Belhumeur, J.P. Hespanha and D.J. Kriegman. Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pages 711-720, July 1997.
 - [48] M.D. Kelly "Visual Identification of People by Computer", Technical Report AI-130, Stanford AI Project, Stanford, CA, 1970. 247 pages
 - [49] A.H BOUALLEG " La reconnaissance Automatique des Visages" Mémoire de Magister, Université de Guelma. 2004.
 - [50] S.GONG, S.J. Mc KENNA et A.PSARROU. *Dynamic Vision: From Images to Recognition*. Imperial College Press, Londres, Angleterre, 2000, 364 pages
 - [51] S.Arca, P.Campadelli, R.Lanzarotti. A Face Recognition System based on automatically determined facial fiducial points. 2005.
 - [52] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042-1062, 1993.
 - [53] B. Heisele, P. Ho, J. Wu, and T. Poggio. Face recognition: component-based versus global approaches. *Computer Vision and Image Understanding*, 91:6-21, 2003.
 - [54] A.Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. *IEEE international conference on Computer Vision and pattern Recognition (CVPR)*, pages 84-91, 1994.
 - [55] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696-710, 1997.
 - [56] P.S. Penev and J.J. Atick. Local feature analysis: A general statistical theory for object representation network. *Computation in Neural Systems*, 7(3):477-500, 1996.

- [57] D.A.Pollen, S.F.Ronner. Phase relationship between adjacent simple cells in the visual cortex. *Science*, 212:1409-1411, 1981.
- [58] H. Burke B., "Ondes et ondelettes, La Saga d'un Outil mathématique" Belin pour la science, 1995.
- [59] A. Grossmann et B.Torrésani, Les ondelettes, Centre de Physique Théorique, CNRS-Luminy, Case 907, 13288 Marseille Cedex 09 et laboratoire d'analyse, topologie et probabilités, CMI, Université de Provence, Marseille, France.
- [60] S. Hamnane, De l'analyse de Fourier traditionnelle aux ondelettes : une approche par le signal musical. Université Paris VI - Pierre et Marie Curie, année 2005.
- [61] David S. Bolme. Elastic Bunch Graph Matching. Thesis Colorado State University Fort Collins, Colorado, 2003.
- [62] C. ROUABHIA, Approches Bidimensionnelles Hybride Neuro-ACPD2D & Locales pour la reconnaissance Automatique de Visages. Mémoire de Magister, Université de Guelma. 2007.
- [63] G. McLachlan. Discriminant Analysis and Statistical Pattern Recognition. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 1992
- [64] S. ZHAO. Apprentissage et Recherche par le Contenu Visuel de Catégories Sémantiques d'Objets Vidéo. Laboratoire des Equipes Traitement des Images et du Signal, Université Paris Descartes, Juillet 2007.
- [65] S. Arca, P. Campadelli, E. Casiraghi, R. Lanzarotti : An Automatic Feature Based Face Authentication System
- [66] Z. Zhang. Feature-Based Facial Expression Recognition: Experiments With a Multi-Layer Perceptron. Technical Report 3354, Institut National de Recherche en Informatique et en Automatique, Février 1998.
- [67] A. Bhuiyan and C.H.Liu. On face recognition using Gabor filters. Proceedings of world academy of science engineering and technology. Volume 22, ISSN 1307-6884, 2007.
- [68] W.Zhao, R.Chellappa and A. Krishnaswamy. Discriminant Analysis of Principal Components for Face Recognition. In Proceedings, International Conference on Automatic Face and Gesture Recognition. 336-341, 1998.
- [69] W.Zhao, R.Chellappa, and P.J. Phillips. Subspace Linear Discriminant and Analysis for Face Recognition. Tech. Rep. CAR-TR-914, Center for Automation Research, University of Maryland College Park, MD, 1990.
- [70] L.Wiscott, J-M. Fellous and V. DerMalsburg. Face Recognition by Elastic Bunch Graph Matching. *IEEE Trans. Pattern and Analysis. Mach. Intelligence*, 19, 775-779, 1997
- [71] O.YAMAGUCHI, K. FUKUI et K.I. MAEDA "Face Recognition Using Temporal Image Sequence" In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR), pages 318-323, 1998.
- [72] S. ROMDHANI, V. BLANZ, C. BASSO et T. VETTER " Morphable Models of Faces" Dans S.Z. LI et A.K. JAIN, éditeurs, Handbook of Face recognition, Chapitre 10, Springer - Verlag Reidel, Dordrecht, 2004. 33 pages
- [73] Jones J. P. Palmer L. A., "The twodimensional spectral structure of simple receptive fields in cat striate cortex", *Journal of Neurophysiology*, Vol.58, No. 6, pp. 1187-1211, December 1987.
- [74] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, Apr. 1998.
- [75] C. Bishop. Neural Networks for Pattern Recognition. Clarendon Press, Oxford,

- 1995.
- [76] R. Gnanadesikan. Methods for Statistical Data Analysis of Multivariate Observations. Wiley Series in Probability and Statistics. Wiley, New York, 2nd edition, 1997.
 - [77] R. LANZAROTTI "Facial Feature Detection and Description" these de Doctorat. Année Académique 2002-2003
 - [78] C. BENCHERIET, Approches Géométriques pour la reconnaissance de Visages. Mémoire de Magister, Université de Guelma. 2000.
 - [79] B. RUF "Face Recognition using Boosting" EPFL / Computer Science Department June 29, 2007.
 - [80] P. Campadelli, R. Lanzarotti" A face recognition system based on local feature characterization" DSI, Université de Milano, ITALIE, 2003
 - [81] J. Wei , Z. Jian , S. Ting-zhi, W. Xiao-hua" A novel facial features extraction algorithm using Gabor wavelets", Department of Electronic Engineering, School of Information Science and Technology, Beijing Institute of Technology, Beijing 100081,China. Congress on Image and Signal Processing 2008