

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Université Batna 2 – Batna
Faculté des Mathématiques et d'Informatique
Département d'Informatique



THESE

Présentée par :

Mme. Leïla BOUSSAAD

En vue de l'obtention du diplôme de

Doctorat en Sciences

Spécialité : Informatique

THEME :

RECHERCHE PAR INDEXATION D'IMAGES PAR LE CONTENU

Soutenue le: 29/06/2017

Devant le jury composé de :

Président	Azeddine Bilami	Prof.,	Université de Batna2
Rapporteur	Mohamed Benmohammed	Prof.,	Université de Constantine2
Co-rapporteur	Rédha Benzid	Prof.,	Université de Batna2
Examineurs	Samir Zidat	Dr.,	Université de Batna2
	Salim Chikhi	Prof.,	Université de Constantine2
	Med Chaouki Babahnini	Prof.,	Université de Biskra

Remerciements

Tous mes remerciements s'adressent tous d'abord à tout puissant ALLAH, d'avoir guidé mes pas vers le chemin du savoir.

Je tiens à remercier Monsieur M. Benmohammed, Professeur à l'université de Constantine2, et Monsieur R. Benzid, Professeur à l'université de Batna2 pour avoir accepté de diriger et d'encadrer cette thèse et aussi pour leur disponibilité permanente et leurs précieux conseils qui m'ont permis d'aboutir dans ce travail.

Je tiens à remercier également Monsieur A. Bilami, Professeur à l'université de Batna2 pour l'honneur qu'il m'a fait en acceptant la présidence du jury de ma soutenance.

Je remercie vivement Monsieur S. Zidat, Maître de conférence à l'université de Batna2, Monsieur S. Chikhi, Professeur à l'université de Constantine2, et Monsieur M. C. Babahnini, professeur à l'université de Biskra, pour avoir accepté la lourde charge d'évaluer cette thèse et d'en être examinateurs.

Je tiens aussi à remercier mon époux Mohamed pour sa patience, son soutien moral très précieux et inestimable qui, à aucun moment n'a cessé de m'encourager pour l'achèvement de ma thèse.

Enfin, que toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail trouvent ici l'expression de mes sincères remerciements

RESUME

Cette thèse s'intéresse à **l'indexation et la recherche d'images fixes par le contenu**, où les images sont décrites au niveau visuel qui peut être très général tel le cas de la recherche d'images sur le Web, ou spécifique à un domaine particulier tel que l'identification biométrique des individus par leurs empreintes digitales, palmaires ou **la reconnaissance faciale**.

L'objectif principal de ce travail a été de développer une technique de description d'images fixes, qui est destinée particulièrement à être utilisée pour la reconnaissance faciale à travers la progression d'âge. La technique combine trois outils populaires utilisés dans le domaine de la reconnaissance de forme qui sont les modèles actifs d'apparence (MAA), la transformée en cosinus discrète (TCD), et l'analyse discriminante non linéaire. Pour atteindre ce but, d'abord, une représentation de l'image basée sur les MAA est générée, ensuite, une TCD est effectuée pour extraire le descripteur qui est représenté par un vecteur dans un espace de grande dimension, enfin, une analyse discriminante non linéaire est effectuée pour réduire la dimension et la classification est faite en utilisant la méthode des k plus proches voisins basée sur la distance euclidienne.

Pour tester l'efficacité du descripteur proposé, une étude expérimentale a été réalisée en utilisant la base de données FG-NET, et en suivant le protocole (Leave-one-person-out) qui sont utilisés dans la majorité des études relatives à l'âge. L'étude a concerné, à la fois, l'effet de la croissance et du vieillissement du visage ainsi que la différence d'âge entre l'image requête et celle de référence sur les performances de la technique proposée.

Les résultats fournis par cette étude encouragent l'utilisation de cette combinaison comme descripteur pour l'identification et la vérification faciale à travers la progression d'âge, surtout pour les individus qui sont âgés de plus de 40 ans et lorsque la différence d'âge est inférieure ou égale à vingt ans.

Mots clés

Recherche d'images par le contenu, Reconnaissance faciale, Invariance à la progression d'âge, croissance faciale, vieillissement faciale, modèles actifs d'apparence, Analyse discriminante non linéaire, transformée en cosinus discrète.

ABSTRACT

This thesis Deals with content-based image indexing and retrieval, where the still images are described at the visual level which can be very general like the case of image search on the Web or specific to a particular field such as biometric identification of individuals by their fingerprints, palmprints or face recognition.

The aim of this work was to develop an approach to extract features from images that are designed specially to be used for face recognition across age progression. the approach combines three popular tools used in pattern recognition, which are the Active Appearance Model (AAM), the two-dimensional discrete cosine transform (2D-DCT), and Kernel Fisher Analysis (KFA). For this purpose, we first used AAM to generate an AAM-based face representation; then, we applied 2D-DCT to get the descriptor of the image; and finally, we used a multiclass KFA for dimension reduction. Classification was made through a K-nearest neighbor classifier, based on Euclidean distance.

To test the effectiveness of the proposed feature extraction approach, an experimental study was performed using the FG-NET face database and following the LOPO (Leave-one-person-out) protocol that are used in the most studies relating to age progression. The study included both, the effect of facial growth, facial aging as well as the age difference between the query image and the reference one on the performance of the proposed technique.

Results from this study encourages the use of this combination as features extraction method for both face identification and verification across age progression, especially for groups of individuals who are older than 40 years and when the age difference is less or equal than twenty years.

Keywords:

Content-based image retrieval (CBIR), face recognition, age invariance, facial growth, facial aging, Active Appearance Model, Kernel Fisher Analysis, 2D-Discrete Cosine Transform.

ملخص

ترتكز هذه الأطروحة على فهرسة و استرجاع الصور الثابتة عن طريق المحتوى، حيث توصف الصور باستعمال الخصائص البصرية التي قد تكون عامة كما هو الحال في البحث عن الصور على الويب ، أو محصورة في مجال معين مثل التعرف على هوية الأفراد من خلال بصمات راحة الكف، بصمات الأصابع أو التعرف على الوجه.

الهدف الرئيسي من هذا البحث هو إعداد تقنية وصف للصور موجهة خصيصا لاستعمالها في التعرف على الوجه عبر تقدم السن. التقنية تجمع ثلاث أدوات مشهورة تستخدم كثيرا في مجال التعرف على الأنماط وهي نماذج المظهر الفعالة (MAA)، تحويل جيب التمام المتقطع (TCD) و التحليل التمييزي غير الخطي.

لتحقيق هذا الهدف، أولا يتم إنشاء تمثيل لصورة الوجه بالاعتماد على ال MAA، ثم باستعمال ال TCD يتم استخلاص واصف للصورة على شكل شعاع من فضاء ذو بعد كبير، بعدها يطبق التحليل التمييزي غير الخطي لتخفيض البعد. أما التصنيف فيتم باستخدام خوارزمية ال "k" أقرب جوار مقاسة بالمسافة الإقليدية.

لاختبار فعالية الواصف المقترح، تم إجراء دراسة تجريبية باستخدام قاعدة المعطيات FG-NET و بروتوكول (Leave- one- person- out) اللذان يتم استخدامهما في معظم الدراسات المتعلقة بالعمر. الدراسة شملت كلا من تأثير مرحلة نمو ومرحلة شيخوخة الوجه وكذلك فارق السن بين صورة الوجه المطلوبة والصورة الموجودة في القاعدة على فعالية التقنية المقترحة. نتائج هذه الدراسة تدعم استخدام هذا الواصف لمعرفة أو تحديد هوية الأفراد والتحقق منها وخاصة بالنسبة للأفراد التي تتجاوز أعمارهم 40 عاما، وكذلك عندما يكون فارق السن أقل أو يساوي 20 سنة.

الكلمات المفتاحية

البحث عن الصور من خلال المحتوى، التعرف على الوجه عبر تقدم السن، نمو الوجه، شيخوخة الوجه، نماذج المظهر الفعالة، تحليل التمييز غير الخطي، تحويل جيب التمام المتقطع.

Liste des figures

Figure 1.1-	Exemples d'images de visage prises à différents âges	007
Figure 2.1-	Architecture d'un système typique de recherche d'images par le contenu...	012
Figure 2.2-	Différents propriétés de texture.....	013
Figure 2.3-	Deux images différentes avec histogrammes de couleur similaires.....	016
Figure 2.4-	Représentation macroscopique et microscopique d'une texture	017
Figure 2.5-	Illustration de la direction et la distance utilisées dans la GLCM.....	018
Figure 2.6-	Séparation de deux classes de données par SVM.....	031
Figure 2.7-	Exemple d'un SVM-BDT.....	034
Figure 2.8-	Courbe de rappel et précision du système Ikona.....	039
Figure 3.1-	Les étapes principale d'un système de reconnaissance faciale.....	044
Figure 3.2-	Exemple d'un visage d'une même personne subissant des variations de pose.....	047
Figure 3.3-	Effets de variations dans les directions d'illumination sur l'apparence d'un visage.....	048
Figure 3.4-	Exemple d'images extraites à partir de la base de données d'expressions faciales JAFEE.....	049
Figure 3.5-	Exemples d'occultations partielles (Base de données AR).....	049
Figure 3.6-	Exemples d'images d'un individu, les images sont de la base de données de visages ORL.....	050
Figure 3.7-	Le concept de l'ACP.....	052
Figure 3.8-	L'image moyenne à gauche, suivie par les sept premières Eigenfaces.....	054
Figure 3.9-	Un exemple de projection de deux classes sur un axe ACP et sur un axe ADL.	056
Figure 3.10-	Images de base obtenues avec la méthodes ACI pour l'architecture 1 en haut et pour l'architecture 2 en bas.....	057
Figure 3.11-	Processus de reconnaissance de visage basé sur les modèles actifs d'apparence.....	059
Figure 3.12-	Caractéristiques géométriques utilisées dans	060
Figure 3.13-	Mesures géométriques utilisées dans	061
Figure 3.14-	Exemple d'un visage divisé en six régions locales sous forme d'ellipse.....	062
Figure 3.15-	Exemple d'une image de visage originale, sa projection (SOM-Face) et l'image reconstruite.....	062
Figure 3.16-	Technique de division du visage.....	063
Figure 3.17-	Le graphe Elastic Graph Matching sous forme d'une grille rectangulaire sur une image de visage comparée à une image de référence	064
Figure 3.18-	Le Face Bunch Graph (FBG) représenté d'un point de vue artistique et scientifique.....	065
Figure 3.19-	Réseau de neurones discriminant pour la reconnaissance faciale.....	069

Figure 3.20-	Illustration d'exemples typiques de graphiques d'un système biométrique...	070
Figure 3.21-	Exemple d'une courbe CMC.....	071
Figure 4.1-	Simulation de la progression d'âge de l'âge X à l'âge Y.....	075
Figure 4.2-	Anthropométrie du visage.....	077
Figure 4.3-	Résultats de transformation d'âge sur des individus différents.....	077
Figure 4.4-	Configuration des différents types de muscles faciaux.....	078
Figure 4.5-	Les 48 points de repère qui sont utilisés ainsi que leur correspondance avec les muscles faciaux de la figure 3.4	078
Figure 4.6-	Un aperçu du modèle de vieillissement facial proposé dans	079
Figure 4.7-	Exemples d'images faciales obtenues par simulation du processus de vieillissement d'un individu âgé de 2 ans.....	080
Figure 4.8-	Pyramide d'orientation du gradient.....	081
Figure 4.9-	Calcul de la pyramide des LBP à partir d'une image.....	082
Figure 4.10-	Un exemple de conversion des coordonnées cartésiennes aux coordonnées polaires.....	083
Figure 4.11-	Les dérivés (drifts) des traits faciaux de quelques visages à différent âges...	083
Figure 4.12-	Illustration des différentes étapes de l'approche TED-GM pour la reconnaissance de visage invariante à l'âge.....	085
Figure 4.13-	Représentation des caractéristiques locales d'une image d'un visage.....	087
Figure 4.14-	Un exemple d'un individu à différents âges.....	088
Figure 5.1-	Différentes étapes pour le calcul du descripteur.....	091
Figure 5.2-	Classes de points d'intérêts.....	093
Figure 5.3-	Exemple d'un visage de la base de donnée FG-NET annoté par 68 points de repère et d'une main annotée par 56 points de repère.....	093
Figure 5.4-	Un exemple d'un ensemble de formes non alignées.....	094
Figure 5.5-	Un exemple d'un ensemble de formes alignées.....	095
Figure 5.6-	Image originale, sa forme et sa texture à l'intérieur de la forme.....	096
Figure 5.7-	Propriétés de Delaunay	097
Figure 5.8-	Exemple de déformation de texture.....	098
Figure 5.9-	images Normalisées d'une personne à différents âges.....	099
Figure 5.10-	Formation du vecteur des caractéristiques dans le domaine de la DCT	100
Figure 5.11-	Images normalisées d'une personne à des âges différents et leur images reconstruites.....	102
Figure 6.1-	Exemples d'images d'individus avec des différences d'âge	109
Figure 6.2-	Alignement manuel.....	111
Figure 6.3-	Courbe CMC pour les expérimentations faites sur la base de données FG-NET.....	114
Figure 6.4 -	Taux de reconnaissance versus nombre de coefficients de basse fréquence mis à zéro.....	115
Figure 6.5-	Image normalisée d'une personne et les images reconstruites en mettant à zéro un pourcentage de coefficients de basse fréquence.....	115

Liste des tableaux

Tableau 6.1-	Taux de reconnaissance au rang-1 (identification) et taux d'erreurs égales (vérification).....	113
Tableau 6.2-	Comparaison de performances pour la reconnaissance de visage invariante à la progression d'âge.....	116
Tableau 6.3-	Taux d'identification versus différence d'âge.....	119
Tableau 6.4-	Détails des groupes d'âge.....	120
Tableau 6.5-	Taux d'identification versus groupes d'âge.....	120

Liste des acronymes

A	
ACI:	Analyse en Composantes Indépendantes
ACP (PCA):	Analyse en Composantes Principale (Principal Component Analysis)
ADL (LDA):	Analyse Discriminante Linéaire (Linear discriminant analysis)
ANMRR:	Average Normalized Modified Retrieval Rank
C	
CBIR:	Content Based Images Retrieval
CMC-C:	Cumulative Match Characteristic Curve
CSLBP:	Center-Symmetric Local Binary Patterns
CSS:	Curvature Scale Space
D	
DLTP:	Differential Local Ternary Patterns
E	
EBGM:	Elastic Bunch Graph Matchnig
EGM:	Elastic Graph Matching
F	
FCM:	Fuzzy c-means
FERET:	Facial Evaluation REcognition Test
FG-NET:	Face and Gesture recognition Research Network
FRVT:	Face Recognition Vendor Test
G	
GLCM:	Gray Level Co-occurrence Matrix
GMM:	Gaussian Mixture Models
GOP:	Gradient Orientations Pyramid
I	
ILBP:	Improved Local Binary Patterns
INFACE:	Illumination Invariant Face Recognition Toolbox
K	
KFA:	Kernel Fisher Analysis
K-ppv:	K-Plus Proches Voisins
L	
LBP:	Local Binary Patterns
LFA:	Local Features Analysis
LGPBP:	Local Gabor Phase Binary Pattern

LGXP:	Local Gabor XOR Pattern
LOPO:	Leave One Person Out
LXP:	Local XOR Pattern
M	
MAA (AAM):	Modèle Actif d'Apparence (Actif Appearance Model)
MBLBP:	Multi-Scale Block Local Binary Pattern
MDL:	Multiview Discriminative Learning
MFDA:	Multi-Feature Discriminant Analysis
MMC:	Modèles de Markov Cachés
MPEG:	Moving Picture Experts Group
N	
NMRR:	Normalized Modified Retrieval Rank
P	
Phd:	Pretty Helpful Development
PSO:	Particle Swarm Optimization
R	
ROC- C:	Receiver Operating Characteristic Curve
RVB (RGB):	Rouge Vert Bleu (Red Green Blue)
S	
SGBD:	Système de Gestion de Base de Données
SIFT:	Scale Invariant Feature Transform
SOM:	Self Organizing Maps
SVM :	Séparateurs à Vastes Marges (Support Vector Machines)
T	
TBIR:	Text Based Images Retrieval
TCD (DCT):	Transformée en Cosinus Discrète (Discret Cosine Transform)
TED-GM:	Texture Embedded Discriminative Graph Matching
TEE (EER):	Taux d'Erreurs Egales (Equal Error Rate)
TFA (FAR):	Taux de Fausses Acceptations (False Accept Rate)
TFR (FRR):	Taux de Faux Rejets (False Reject Rate)
TPLBP:	Three Patch Local Binary Pattern
TSV(HSV):	Teinte Saturation et Valeur (Hue Saturation Value)
U	
UDP	Unsupervised Discriminant Projection
W	
WLD	Weber Local Descriptor

Table des matières

Remerciements	i
Résumé	ii
Abstract	iii
ملخص	iv
Liste des figures	v
Liste des tableaux	vii
Liste des acronymes	viii
Table des matières	x
CHAPITRE1 Introduction générale	001
1.1 Introduction	002
1.2 Motivation et contribution	005
1.3 Structure de la thèse.....	008
 Partie I: Introduction sur le domaine de recherche	 009
 CHAPITRE 2 Principes fondamentaux de la recherche d'images par le contenu	 010
2.1 Introduction	011
2.2 Description d'images	012
2.2.1 Descripteurs de couleur	013
2.2.2 Descripteurs de texture	016
a) Les méthodes statistiques	018
b) Les méthodes basées sur la modélisation	019
c) Les méthodes spatio-fréquentielles	019
2.2.3 Descripteurs de forme	020
2.2.3.1 Descripteurs de forme basés sur les contours	020
2.2.3.2 Descripteurs de forme basés sur les régions	021
a) Les moments géométriques	021
b) Les moments orthogonaux	023
2.2.4 Combinaison de descripteurs	025
2.3 Mesures de similarité	026
2.4 Structure d'indexation	029
2.5 Les techniques d'apprentissage	030
2.5.1 Les techniques d'apprentissage supervisé	030
2.5.2 Les techniques d'apprentissage non supervisé	034
2.6 Spécification de la requête	036
2.7 Retour de pertinence	037
2.8 Evaluation des performances	038

2.9 Conclusion	040
CHAPITRE 3 Reconnaissance automatique de visage	042
3.1 Introduction	043
3.2 Applications de la reconnaissance faciale	045
3.3 Difficultés rencontrées par les systèmes de reconnaissance automatique du visage	046
3.4 Techniques de description de visage	051
3.4.1 Les approches globales	051
3.4.1.1 L'analyse en composantes principales	051
3.4.1.2 L'analyse discriminante linéaire, ADL	054
3.4.1.3 L'analyse en composantes indépendantes ACI	056
3.4.1.4 Les méthodes de projection non linéaires	057
3.4.1.5 Les modèles actifs d'apparence	057
3.4.1.6 Les réseaux de neurones	059
3.4.2 Les approches locales	060
3.4.2.1 Les approches géométriques	060
3.4.2.2 Les techniques modulaires	061
3.4.2.3 Les Modèles de Markov Cachés (MMC)	063
3.4.2.4 Les approches basées sur les graphes	063
a) L'Elastic Graph Matching (EGM)	064
b) L'Elastic Bunch Graph Matching (EBGM)	065
3.4.3 Les approches hybrides	065
3.5 les approches basées sur la transformée en cosinus discrète	066
3.6 L'identification (la reconnaissance)	068
3.6.1 Classification par k-plus proches voisins	068
3.6.2 Classification par réseaux de neurones	068
3.6.3 Classification par SVM	069
3.7 Mesures d'évaluation	069
3.7.1 Vérification	069
3.7.2 Identification	071
3.8 Conclusion	071
CHAPITRE 4 Reconnaissance de visage invariante à la progression d'âge	073
4.1 Introduction	074
4.2 Les techniques de reconnaissances de visages invariantes à la progression d'âge	074
4.2.1 Les approches génératives	075
4.2.2 Limites des approches génératives	080
4.2.3 Les approches non génératives	081
4.4 Conclusion	088
Partie II: Contribution	089

CHAPITRE 5 Méthode de description de visage invariante à la progression d'âge -----	090
5.1 Introduction -----	091
5.2 La normalisation d'un visage dans une image -----	092
5.2.1 L'alignement des formes -----	094
5.2.2 La déformation de la texture de l'image -----	096
5.2.3 Processus de normalisation -----	098
5.3 Extraction de caractéristiques -----	099
5.4 La transformée en cosinus discrète -----	099
5.4.1 Principe -----	099
5.4.2 Choix de la DCT -----	101
5.5 Réduction de la dimension du descripteur -----	102
5.5.1 Analyse de Fisher à noyau -----	103
5.6 Reconnaissance (identification) -----	105
5.7 Conclusion -----	106
CHAPITRE 6 Expérimentations, évaluations et discussions -----	107
6.1 Introduction -----	108
6.2 Choix de la base de données -----	108
6.3 Description de l'expérimentation -----	110
6.4 Résultats expérimentaux -----	110
6.4.1 Importance de l'étape de normalisation -----	111
6.4.2 Importance d'utiliser la DCT -----	112
6.4.3 Importance d'utiliser la DCT et fixer un petit nombre des premiers coefficients à zéro -----	112
6.4.4 Choix du pourcentage de coefficients de basses fréquences à être ignorés -----	114
6.4.5 Intérêt d'utiliser la KFA -----	115
6.4.6 Résultats obtenus avec certaines approches proposées dans le domaine de la reconnaissance faciale invariante à l'âge -----	116
6.5 Influence de la différence d'âge entre l'image requête et celle de référence sur les performances de la reconnaissance -----	118
6.6 Effets de croissance et de vieillissement sur les performances de la reconnaissance de visage -----	119
6.7 Conclusion -----	120
Conclusion Générale -----	122
ANNEXE A Coordonnées barycentriques -----	126
ANNEXE B Interpolation bilinéaire -----	128
Références bibliographiques -----	130

CHAPITRE 1:

Introduction générale

1.1 Introduction	002
1.2 Motivation et contribution	005
1.3 Structure de la thèse.....	008

CHAPITRE 1:

Introduction générale

1.1 Introduction

Avec le développement des nouvelles technologies de capture et de stockage d'images numériques, ainsi que l'avènement de l'internet, les collections d'images numériques produites par les applications scientifiques, éducatives, médicales et industrielles ont augmenté rapidement. Au cours des deux dernières décennies, un certain nombre de systèmes de recherche d'images ont été conçus pour répondre à cette demande croissante de la recherche, de la navigation et de la reconnaissance d'images dans de grandes collections d'images. Les systèmes de recherche d'images peuvent être classés en deux classes: les systèmes de recherche d'images basée sur le texte (en anglais text based images retrieval: TBIR) [1], et les systèmes de recherche d'images basée sur le contenu (en anglais: content based images retrieval: CBIR) [2]. Dans la première classe, les images sont annotées par les caractéristiques textuelles telles que les noms des fichiers, les légendes et les mots clés, décrivant leur contenu, et par la suite, un système de gestion de base de données traditionnel (SGBD) est utilisé pour effectuer la recherche d'images. Une étude exhaustive sur les méthodes de recherche d'images basées sur l'annotation textuelle peut être trouvée dans [3, 4]. L'inconvénient majeur de cette technique, c'est qu'elle nécessite l'intervention humaine pour l'annotation manuelle qui est une tâche lourde et longue, surtout avec le volume des bases de données images qui devient aujourd'hui très important, ainsi que l'annotation inexacte en raison de la subjectivité de la perception humaine, des limites de la langue et d'expertise du domaine [5, 6]. Cet inconvénient a constitué la cause principale pour laquelle la recherche d'images par le contenu a été introduite [2].

Dans un système de recherche d'images par le contenu, les images sont indexées et recherchées par leur contenu visuel tel que la couleur, la texture et la forme, ainsi les résultats fournis par ces systèmes sont plus pertinents que ceux fournis par les systèmes

basés sur l'annotation textuelle. En outre, la recherche d'images par le contenu est aussi efficace en termes de temps de calcul. En effet la couleur, la forme et la texture sont des caractéristiques de bas niveau qui peuvent être extraites en utilisant des algorithmes simples et rapides [7]. C'est pour cette raison, la recherche d'images par le contenu, depuis son origine, demeure un thème de recherche passionnant pour de nombreuses applications de navigation et de recherche d'images numériques, et des progrès remarquables ont été réalisés tant en recherches théoriques qu'en développement de systèmes de recherche. Ainsi, un bon nombre de systèmes de recherche d'images par le contenu académiques et commerciaux ont été développés par les universités, les organismes gouvernementaux et les entreprises tels que: Photobook [8], Virage [9], VisualSEEK [10], Netra [11], SIMPLIcity [12] pour n'en citer que quelques uns.

Selon l'objectif pour lequel est construit le système de recherche d'images par le contenu, certains systèmes sont conçus pour reconnaître un type d'objets en particulier, c'est le cas d'une base d'images spécialisée, qui contient des images de même type (image médicale, géographique, visage,...), d'autres sont conçus pour des bases d'images généralistes qui contiennent des images de nature variée (c'est le cas du web). Pour la première catégorie, on connaît à priori le type d'images que l'on peut y rencontrer, par exemple : rechercher un individu parmi d'autres, tous les traits du visage ont un arrangement spatial similaire : deux yeux, un nez et une bouche dans un contour rond, dont la localisation varie très peu. C'est le domaine de recherche le plus récent et le plus difficile. Pour la seconde catégorie, par contre, on ne sait pas ce que contiennent les images, et cela dû à la diversité du contenu des images du monde réel, car on ne peut pas établir un lien entre les caractéristiques de bas niveau et les concepts de haut niveau c'est à dire le système ne sait pas sémantiquement ce que recherche l'utilisateur. Diverses techniques ont été étudiées pour améliorer les performances des systèmes de recherche en réduisant les écarts entre les caractéristiques de bas niveau et les concepts de haut niveau.

Egalement, les techniques conçues pour usage général peuvent ne pas produire de bon résultats pour des domaines spécifiques, par exemple, un système de recherche général peut distinguer entre l'image d'un arbre et celle d'un bâtiment, mais, il est peut être incapable de reconnaître deux jumeaux identiques. Par conséquent, le choix de

caractéristiques de bas niveau qui sont effectivement appropriées à un domaine spécifique fait l'objet de diverses recherches intéressantes. Plusieurs chercheurs ont développé des techniques en se basant sur les caractéristiques de bas niveau pour diverses applications spécifiques à des domaines tels que la télédétection, la médecine, l'architecture, les systèmes biométriques, etc. [13]. Cependant, les chercheurs travaillant sur des systèmes de recherche spécialisés sont toujours confrontés par plus de défis. Les caractéristiques d'un domaine peuvent ne pas convenir à un autre domaine. C'est pourquoi, les principaux défis restent toujours l'identification et la sélection des méthodes appropriées d'extraction de caractéristiques pour le domaine cible.

D'autre part, une forte demande des systèmes de sécurité biométriques en raison des activités de terrorisme continues et les menaces de sécurité partout dans le monde [14]. Des rapports sur la sécurité des systèmes basés sur les mots de passe traditionnels soulignent la facilité de déchiffrer la majorité des mots de passe plus compliqués [15, 16, 17], en outre, les mots de passe ou les jetons de sécurité peuvent être soit oubliés, volés ou perdus [18]. L'utilisation des traits biométriques pour l'authentification peut facilement surpasser ces inconvénients des systèmes de sécurité basés sur les mots de passe. Un système biométrique est une technique de reconnaissance automatique des personnes à partir de leurs traits physiologiques tels que le visage, les empreintes digitales ou les traits comportementaux tels que la signature et la voix [19]. Parmi tous les traits biométriques, la reconnaissance faciale a connu un gain d'intérêt exponentiel dans les milieux universitaires et industriels. Car selon un rapport récent [20], d'ici 2018, la reconnaissance faciale sera la deuxième caractéristique biométrique la plus adoptée après les empreintes digitales, et ceci grâce à sa précision accrue et sa nature peu intrusif par rapport aux autres modalités biométriques, car elle ne nécessite pas d'avoir un contact direct entre le capteur (la caméra) et le sujet, elle est bien acceptée par les utilisateurs à l'inverse d'autres techniques comme l'iris où un contact direct est nécessaire entre le capteur et l'œil.

De nombreux chercheurs ont mentionné que les techniques de la recherche d'images par le contenu peuvent être très efficaces pour la reconnaissance des traits biométriques tels que le visage et les empreintes digitales [21, 22, 23]. Cela nous a encouragé à explorer les défis et les obstacles de la reconnaissance des traits biométriques en utilisant des

techniques de recherche d'images par le contenu et **de développer une technique pour la recherche dans des bases de données contenant des images de visage.**

1.2 Motivation et contribution

Dans cette thèse, plus précisément, nous nous intéressons à la reconnaissance faciale. Le visage humain constitue une classe particulière d'objet 3D qui est, depuis longtemps, un sujet d'intérêt pour les chercheurs, et plusieurs études ont largement contribué à mieux comprendre la perception des visages humains et à développer des modèles numériques qui aident à caractériser l'apparence du visage et de nombreuses méthodes qui ont été proposées pour la reconnaissance faciale à partir d'images 2D, 3D ou de la vidéo. Zhao et al. [24] ont fourni une analyse qualitative des différents algorithmes de reconnaissance proposés. Des décennies de recherches intensives dédiées complétées par la standardisation des algorithmes de reconnaissance faciale grâce aux méthodologies d'évaluation tels que FERET [25,26] et FRVT [27] ont, ainsi considérablement, aidé cette technologie à atteindre un certain niveau de maturité, ce succès est souvent limité à un certain nombre de paramètres contrôlés. Car il a été noté que les performances de nombreuses méthodes de reconnaissance faciale de référence se détériorent significativement dans les environnements réels qui sont incontrôlés [28]. Récemment, les recherches en reconnaissance faciale sont principalement axées sur des scénarios impliquant des images capturées dans des conditions non contrôlées. Dans le cadre des efforts réalisés pour faire face aux difficultés rencontrées par les systèmes de reconnaissance faciale, un nouveau défi s'est ajouté : L'âge comme un attribut lié au visage humain fait l'objet d'études pour la reconnaissance faciale à travers la progression d'âge, surtout pour des applications où la considération d'âge est nécessaire comme l'identification des enfants disparus, la surveillance, et la détection d'inscriptions multiples pour les documents délivrés par les gouvernements tels que les permis de conduire et les passeports. Ces trois scénarios ont deux caractéristiques communes :

- (i) une importante différence d'âge entre les images à connaître et celles de la base de données (images obtenues lors des étapes d'inscription et de vérification) ; par exemple dans des applications de surveillance, la variation d'âge est une source majeure de difficulté pour identifier les suspects dans une liste de surveillance. Les criminels commettent des crimes au cours de différentes

périodes de leur vie, en commençant souvent, dès leur jeune âge et en continuant tout au long de leur vie. Il n'est pas rare de rencontrer un laps de temps de dix à vingt ans entre la première (inscription) et les arrestations (vérification) qui suivent.

- (ii) l'incapacité d'obtenir l'image du visage d'un sujet pour mettre à jour la base de données.

Ling et al. [29] ont étudié la façon dont la différence d'âge affecte les performances de reconnaissance des visages dans une tâche réelle de vérification des photos des passeports. Leurs résultats montrent que le processus de vieillissement augmente la difficulté de la reconnaissance.

L'effet de la progression d'âge sur les performances de la reconnaissance faciale n'a pas été substantiellement étudié. Cela est dû principalement à:

- (i) la simulation des variations de forme et de texture causés par le vieillissement du visage est une tâche difficile, car les changements que peut subir le visage humain avec l'âge sont inévitables. Les effets du vieillissement du visage portent sur des variations importantes qui touchent à la fois la forme et la texture du visage. Le degré de variation est influencé par d'autres facteurs tels que le genre, l'ethnie (la race), l'exposition au soleil, les effets dû au tabagisme et à l'alcoolisme, la prise ou la perte du poids, etc. D'autres facteurs peuvent annuler ou au moins masquer les effets du vieillissement sur le visage tels que les traitements cosmétiques, dermatologiques et les interventions chirurgicales [30].
- (ii) Il n'y a pas de base de données du domaine public pour l'étude du vieillissement du visage jusqu'à récemment. Et même les bases de données disponibles pour étudier le vieillissement du visage ne sont pas seulement réduites mais aussi elles contiennent des variations internes et externes incontrôlées qui dégradent les performances de la reconnaissance de visage. (par exemple, la pose, l'éclairage, et l'expression). La figure 1.1 montre des exemples d'images prises à des âges différents, les images appartiennent à la base de données FG-NET.

Ces raisons expliquent bien l'absence d'études approfondies sur les effets du vieillissement par rapport aux autres facteurs de variations intra-classe dans l'apparence du visage.

Depuis peu, l'accent a été mis sur ce problème de vieillissement du visage dû à l'avancement d'âge qui peut considérablement dégrader les performances des algorithmes de la reconnaissance faciale.

C'est dans ce contexte de variabilité d'âge qu'intervient cette thèse. Un nouvel algorithme est proposé pour améliorer les résultats de la reconnaissance faciale à travers la progression d'âge.

L'objectif majeur de ce travail est donc, en se basant sur les caractéristiques de bas niveau, de proposer un descripteur qui soit discriminant, invariant et compact pour un système de reconnaissance faciale à travers la progression d'âge.

Le terme discriminant veut dire que le descripteur doit permettre de faire la différence entre des individus différents. Invariant veut dire qu'il soit moins affecté par des changements tels que : les changements d'éclairage, la position de l'individu dans l'image, son expression et surtout les changements dus aux changements temporels qui sont l'objectif principal de notre travail. Et enfin compact veut dire que la dimension du descripteur doit être réduite.

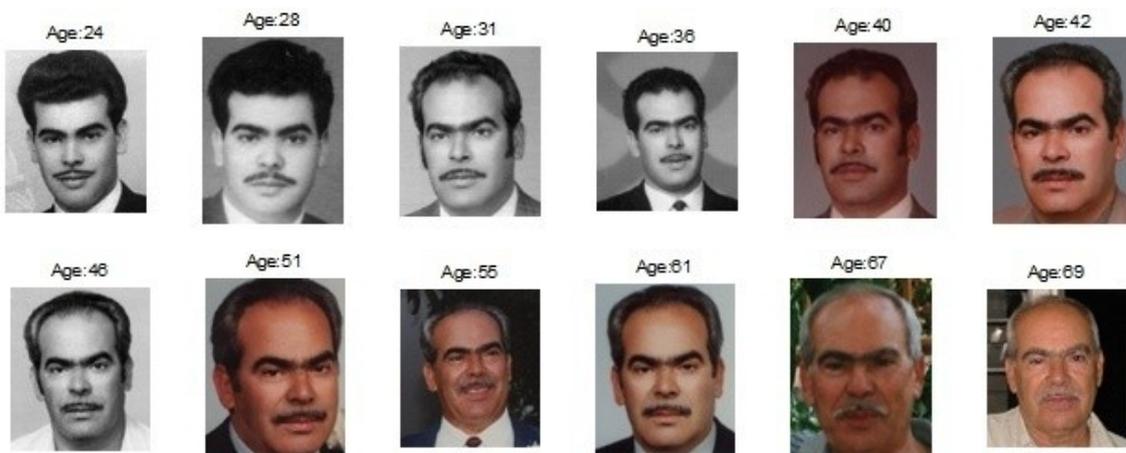


Figure 1.1- Exemples d'images de visage prises à différents âges.

1.3 Structure de la thèse

Ce manuscrit de thèse est organisé en deux parties et une conclusion, chaque partie présente plus précisément une part de nos travaux, la première partie consiste à présenter le domaine de recherche, tandis que la deuxième est consacrée à notre contribution.

La première partie de la thèse est structurée en trois chapitres:

Le chapitre 2 introduit une relecture des principes fondamentaux de la recherche d'image par le contenu.

Le chapitre 3 présente une revue des systèmes de reconnaissance de visages, Nous décrivons d'abord les trois modules constituant la chaîne de traitements intégrés dans l'approche standard. Puis, nous présentons les principales techniques qui ont été développées dans ce contexte.

Dans **le chapitre 4**, nous nous sommes intéressés au domaine d'étude ciblé, nous exposons un état de l'art des systèmes de reconnaissance de visages, à travers la progression d'âge, en présentant les principales techniques qui ont été développées dans ce contexte et les résultats décrits dans la littérature.

La deuxième partie de la thèse est aussi divisée en deux chapitres, Notre méthode originale de reconnaissance de visage à travers la progression d'âge est décrite dans **le chapitre 5**. Le choix du descripteur durant la phase d'enrôlement et la reconnaissance réalisée durant la phase d'identification sont détaillées, ces descriptions étant accompagnées de justifications de l'approche proposée.

Des résultats expérimentaux calculés sur les images de la base de données FG-NET sont présentés dans **le chapitre 6**. Nous commentons les performances de notre nouvelle technique et tirons des conclusions sur l'utilité du descripteur proposé pour la reconnaissance faciale à travers la progression d'âge.

La conclusion conclut la thèse en résumant nos principaux apports et présentant les différentes perspectives de nos travaux.

Partie I :

Introduction sur le domaine de recherche

CHAPITRE 2:

Principes fondamentaux de la recherche d'images par le contenu

2.1 Introduction	011
2.2 Description d'images	012
2.2.1 Descripteurs de couleur	013
2.2.2 Descripteurs de texture	016
a) Les méthodes statistiques	018
b) Les méthodes basées sur la modélisation	019
c) Les méthodes spatio-fréquentielles	019
2.2.3 Descripteurs de forme	020
2.2.3.1 Descripteurs de forme basés sur les contours	020
2.2.3.2 Descripteurs de forme basés sur les régions	021
a) Les moments géométriques	021
b) Les moments orthogonaux	023
2.2.4 Combinaison de descripteurs	025
2.3 Mesures de similarité	026
2.4 Structure d'indexation	029
2.5 Les techniques d'apprentissage	030
2.5.1 Les techniques d'apprentissage supervisé	030
2.5.2 Les techniques d'apprentissage non supervisé	034
2.6 Spécification de la requête	036
2.7 Retour de pertinence	037
2.8 Evaluation des performances	038
2.9 Conclusion	040

CHAPITRE 2

Principes fondamentaux de la recherche d'images par le contenu

L'indexation et la recherche d'images par le contenu visuel consiste à représenter et à rechercher des images en utilisant que les caractéristiques visuelles de l'image elle-même, sans aucune autre information. Ce chapitre présente les principes fondamentaux de l'indexation et de la recherche d'images basées sur le contenu, il détaille certaines méthodes largement utilisées pour la description du contenu visuel et précise les principales mesures de similarité utilisées permettant une recherche efficace par le contenu. En outre, il décrit les différents schémas d'indexation, les méthodes de formulation des requêtes, le retour de pertinence et les principaux critères d'évaluation des performances des systèmes de recherche d'images.

2.1 Introduction

La recherche d'images par le contenu, comme son nom l'indique, consiste à indexer et à rechercher des images en utilisant que les caractéristiques physiques telles que la couleur, la forme, la texture ou toute autre information qui peut être extraite de l'image elle-même. L'indexation et la recherche sont deux aspects inséparables, le premier correspond à la description et la représentation informatique de l'image et le second concerne l'utilisation de cette représentation pour rechercher des images. Un schéma typique d'un système de recherche d'images par le contenu peut être représenté par la figure 2.1. Deux phases principales doivent exister: une phase "hors-ligne", elle correspond à l'indexation, ou le calcul des descripteurs et une phase "en ligne", elle correspond à la recherche. Les contenus visuels des images sont extraits et décrits par des vecteurs multidimensionnels de caractéristiques, ces vecteurs forment la base de données de descripteurs (ou de signatures). Pour rechercher une image, les utilisateurs fournissent au système de recherche des exemples d'images ou des croquis. Ensuite, le système change ces exemples en représentations internes sous forme de vecteurs de descripteurs, la similarité entre le vecteur descripteur de la requête contenant l'exemple

ou le croquis, et ceux des images de la base de données est calculée, enfin, la récupération est effectuée à l'aide d'un schéma d'indexation. Le schéma d'indexation fournit un moyen efficace pour rechercher des images dans une base de données. Les systèmes récents de la recherche d'images par le contenu ont permis l'interaction de l'utilisateur avec le système, et ceci en intégrant une phase de bouclage de pertinence (Relevance feedback), dans le but de générer des résultats perceptivement et sémantiquement plus significatifs. Les descripteurs calculés doivent comprendre des caractéristiques discriminantes qui permettent une bonne description du contenu visuel de l'image et qui doivent être associé à des mesures de similarité efficaces.

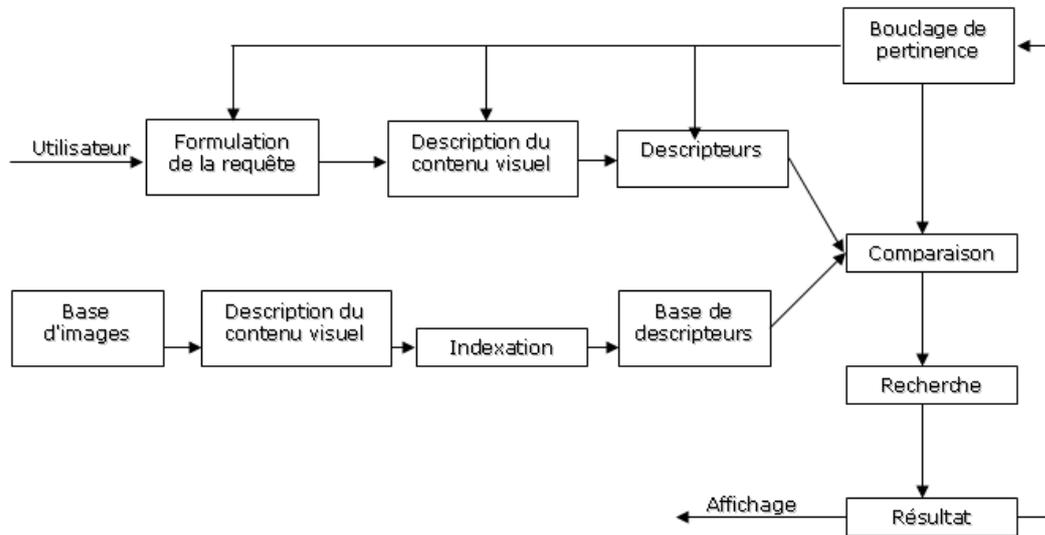


Figure 2.1- Architecture d'un système typique de recherche d'images par le contenu.

2.2 Description d'images

La description d'images constitue un moyen d'extraction d'une représentation plus compacte mais sémantiquement expressive qui résume le contenu visuel plus important de l'image, cette représentation est appelée descripteur ou signature. Des images similaires doivent avoir des signatures similaires. La description d'images doit considérer les caractéristiques les plus utiles pour représenter le contenu et choisir les approches qui codent effectivement ces caractéristiques. Par exemple, si nous examinons les images (a) à (d) de la figure 2.2, la couleur noire et la texture des surfaces sont des propriétés caractéristiques. En effet, les performances des systèmes de recherche d'images par le contenu dépendent directement de la sélection et du choix de ces caractéristiques et des méthodes utilisées pour les calculer.

D'une manière générale, un bon descripteur du contenu visuel doit vérifier les propriétés d'invariance, particulièrement, en translation, en rotation et en changement d'échelle.

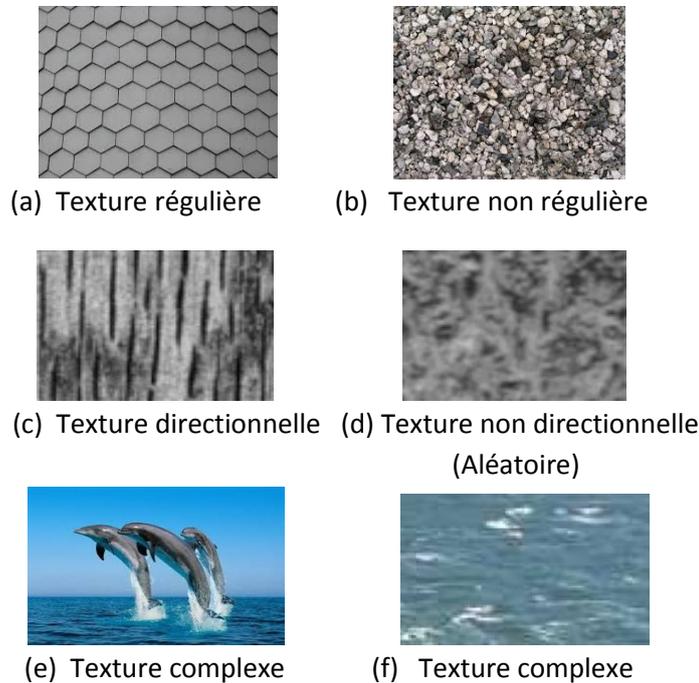


Figure 2.2- Différentes propriétés de texture.

Les images peuvent être décrites soit par des descripteurs globaux; soit par des descripteurs locaux. Un descripteur global utilise les attributs visuels de l'image entière, tandis qu'un descripteur local utilise les attributs visuels de régions ou d'objets dans l'image. Pour obtenir des descripteurs locaux, l'image doit être divisée en parties, la façon la plus simple pour effectuer cette division et de partitionner l'image en parties de tailles et de formes égales, cependant cette partition ne génère pas des régions perceptivement significatives mais elle constitue une façon de représenter les attributs globaux à une résolution plus fine. Une meilleure méthode consiste à segmenter l'image en objets sémantiquement significatifs en fonction de certains critères. Dans ce qui suit, nous allons présenter, brièvement, quelques techniques utilisées pour décrire le contenu visuel des images, à savoir, pour l'extraction de la couleur, la texture et la forme [5].

2.2.1 Descripteurs de couleur

La couleur constitue le critère le plus important dans la vision d'une image, c'est l'attribut largement utilisé dans la plupart des systèmes de recherche d'image par le

contenu, et le plus puissant qui simplifie l'identification des objets. Les couleurs sont définies sur des espaces colorimétriques appropriés où chaque pixel de l'image est représenté par un point dans un espace 3D de couleurs. Divers espaces de couleurs sont développés, ils servent souvent pour différentes applications. Les espaces de couleur couramment utilisés pour la recherche d'image comprennent: l'espace RVB, TSV, CIE $L^*a^*b^*$, CIE $L^*u^*v^*$. Bien qu'il n'y a pas d'accord sur quel espace de couleur est le meilleur pour la recherche d'image par le contenu, l'une des caractéristiques nécessaire pour le choix de l'espace de couleur approprié est l'uniformité. L'uniformité signifie que deux paires de couleur qui sont égales en termes de distance dans un espace colorimétrique sont considérées comme perceptivement égales. En d'autres termes la proximité mesurée entre les couleurs doit être directement liée à la similarité mentale entre celles-ci.

L'espace RVB (RGB) est l'espace de couleur largement utilisé pour l'affichage des images. Il est défini par les trois composantes de couleur rouge, vert et bleu, ces trois composantes sont appelées primaires additives, car une couleur dans l'espace RVB est produite en additionnant les trois couleurs primaires. En revanche, l'espace CMY est un espace de couleur principalement utilisé pour l'impression, il est composé des trois couleurs cyan, magenta et jaune. Ces trois composantes sont appelées primaires soustractives, car une couleur dans cet espace est produite par absorption de lumière. Les deux espaces RVB et CMY ne peuvent pas être utilisés pour la recherche d'images par le contenu, car ils ne sont pas uniformes et ils dépendent du dispositif de capture.

Les espaces CIE $L^*a^*b^*$ et CIE $L^*u^*v^*$ appartiennent à la famille des systèmes chromatiques uniformes et ils sont indépendants des dispositifs de capture. Ils sont constitués d'une composante de luminance (L) et de deux composantes chromatiques (de chrominance) a et b (ou u et v). CIE $L^*a^*b^*$ est conçu pour traiter les mélanges de colorants additifs. La formule de transformation de l'espace RVB en espace CIE $L^*u^*v^*$ ou CIE $L^*a^*b^*$ peut être trouvée dans [31].

L'espace TSV pour teinte saturation et valeur (en anglais HSV pour hue saturation value ou HSB pour hue saturation brightness) est particulièrement utilisé dans les applications de l'infographie. Il constitue le moyen le plus intuitif pour décrire une couleur. La composante de teinte est invariante aux changements d'illumination et à la direction de la caméra. Par conséquent, il est le plus adapté à la recherche d'objets. En outre, les

coordonnées RVB peuvent être facilement converties en coordonnées HSV par une simple formule [31].

De nombreux descripteurs de couleur ont été proposés, l'histogramme de couleur [32] constitue l'outil fondamental le plus utilisé pour représenter l'attribut couleur d'une image. Il peut être défini comme une fonction discrète qui associe à une couleur particulière de l'espace de couleur le nombre de pixels qui ont cette couleur. Les histogrammes sont en général normalisés, en divisant les valeurs de chaque couleur par le nombre total de pixels de l'image. La valeur d'une couleur varie alors entre 0 et 1, et peut s'interpréter comme la probabilité d'occurrence de la couleur dans l'image. Qu'il soit global ou local, l'histogramme de couleur est l'ensemble efficace permettant de représenter la distribution des intensités (ou des couleurs) de l'image entière ou d'une région de l'image. Un histogramme de couleur global ne considère pas la distribution des couleurs de régions d'images, ce qui est le contraire pour l'histogramme local qui divise une image en blocs ou régions fixes, et prend l'histogramme de couleur de chacun de ces blocs. L'histogramme local contient plus d'informations sur l'image, mais, il est coûteux en calcul lorsqu'on compare des images. En outre, Les histogrammes sont robustes à un certain nombre de transformations sur l'image. Car ils sont invariants en translation et en rotation suivant l'axe de vue [32]. Les histogrammes sont en revanche sensibles aux changements d'illumination et aux conditions d'éclairage.

Dans un histogramme de couleur, la quantification est un processus où le nombre de couleurs est réduit en prenant des couleurs qui sont semblables, et en les regroupant dans la même classe. La quantification réduit l'espace nécessaire pour le stockage de l'histogramme, et le temps de comparaison de deux histogrammes. C'est un compromis entre espace de stockage, temps de comparaison et précision des résultats.

L'histogramme de couleur global est connu comme un procédé traditionnel pour la recherche des images couleur. Comme il ne comprend pas la distribution de la couleur des régions, lorsque deux histogrammes globaux sont comparés, on ne peut pas toujours obtenir un bon résultat lorsqu'il est vu en termes de similarité d'images [33]. La figure 2.3 illustre bien cette limite où deux images différentes ont des histogrammes de couleur similaires.

Pour prendre en compte l'information locale, plusieurs modifications de l'histogramme de couleur sont proposées dans la littérature, telles que les vecteurs de cohérence de couleurs [34], les corrélogrammes de couleurs [35] et les histogrammes nommés Histogramme de blobs [36]

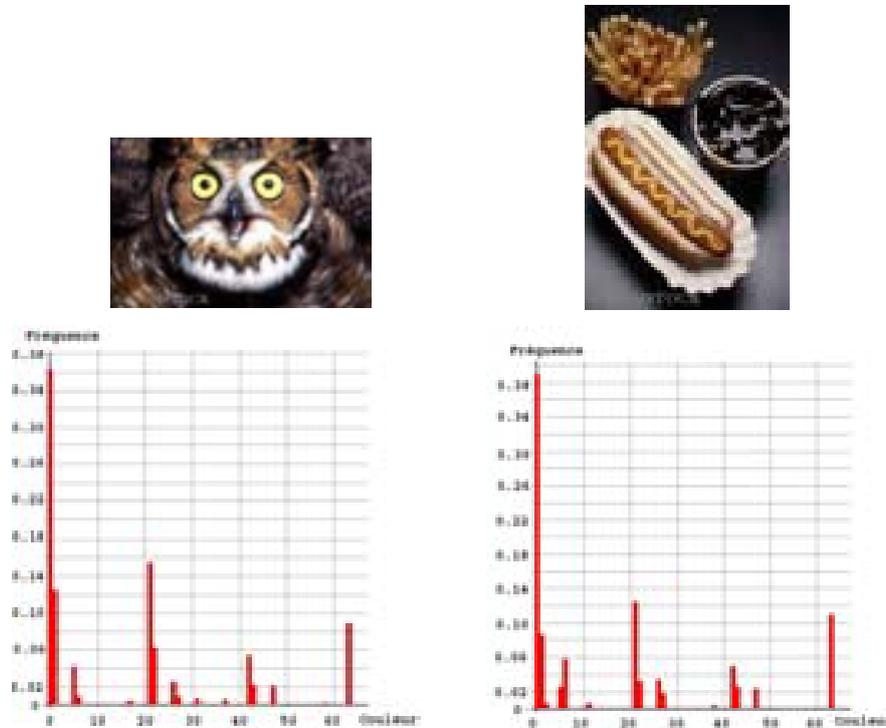


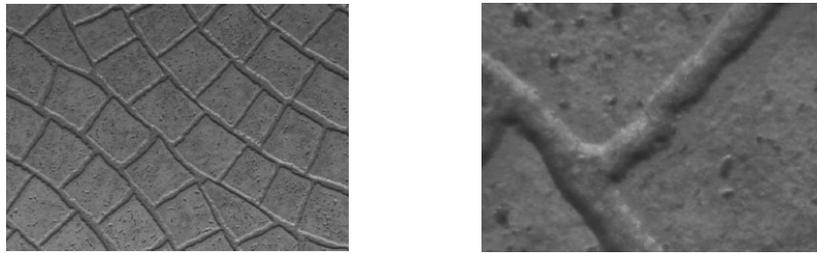
Figure 2.3- Deux images différentes avec histogrammes de couleur similaires [33].

2.2.2 Descripteurs de texture

La texture est une autre propriété fondamentale de l'image. Diverses représentations de texture ont été étudiées dans le domaine de reconnaissance de forme et de vision par ordinateur. Quoiqu'il soit facile de décrire, visuellement, des textures présentes dans une image, il est difficile de définir une texture précisément dans le domaine de l'analyse d'image, et cela dû à la diversité des propriétés des textures existantes. Une texture peut être considérée comme un agencement plus au moins régulier de motifs répétés ou de pixels sur un domaine spatial qui forme un ensemble. Les différentes propriétés de la texture telles qu'elles sont perçues par l'œil humain peuvent être exprimés par des adjectifs qualificatifs tels que: régulières, directionnelles, contrastées, grossières, fines, lisses, marbrées, granuleuses, etc. (voir figure 2.2 (a) à (d)). En véritables scènes, la perception de la texture peut être beaucoup plus compliquée. Les différentes intensités

de luminosité donnent lieu à un mélange de perception humaine différente de la texture (voir figure 2.2 (e) et (f)).

Dans la pratique, on peut distinguer deux approches pour définir une texture [37], la première correspond à une vision macroscopique de la texture (pour une analyse plus grossière), où elle est vue comme une répétition spatiale d'un motif élémentaire, appelé TEXTON. L'image de la figure 2.4 (a) représente un pavage où le motif est un pavé qui se répète d'une manière structurée. La deuxième, quant elle, correspond à une vision microscopique de la texture (pour des détails fins) où la représentation de la texture est considérée au niveau d'un voisinage du pixel lui-même. L'image de la figure 2.4(b) représente un pavé du pavage de la figure (a).



(a) Représentation macroscopique (b) Représentation microscopique

Figure 2.4 - Représentation macroscopique et microscopique d'une texture (les images sont de taille 746×538, elles sont extraites de la base de texture OuTex) [38].

Les méthodes d'analyse de texture sont exploitées dans plusieurs applications de traitement de l'image et de la vision par ordinateur notamment dans des problématiques de segmentation ou de classification de texture. En effet, la classification d'images de texture a été utilisée dans plusieurs applications de contrôle qualité industrielle, d'analyse de documents, de visage ou d'images satellites [38]. De nombreuses méthodes pour calculer l'attribut de texture ont été proposées, néanmoins, il n'existe pas de méthode unique qui fonctionne mieux avec tous les types de texture.

Selon Manjunah et Ma [42], les méthodes couramment utilisées pour la description de texture peuvent être regroupées en trois catégories:

- Les méthodes statistiques
- Les méthodes basées sur la modélisation

- Les méthodes spatio-fréquentielles

Ces catégories sont brièvement expliquées ci-dessous:

a) Les méthodes statistiques

Les méthodes statistiques analysent la distribution spatiale des valeurs de gris en calculant des caractéristiques locales à chaque pixel de l'image et en dérivant un ensemble de statistiques à partir de la distribution de ces caractéristiques locales. La matrice de cooccurrence des niveaux de gris (Gray level co-occurrence matrix: GLCM) [43] est la méthode statistique la plus utilisée pour extraire l'attribut texture. En effet, c'est la méthode qui décrit effectivement la notion de texture. C'est une matrice bidimensionnelle de probabilités conjointes entre des paires de pixels, séparés par une distance d , dans une direction donnée, r . Les directions utilisées sont 0, 45, 90, 135 degrés (voir figure 2. 5). Car la description de la texture est basée sur la configuration de la répétition de l'occurrence de certains niveaux de gris.

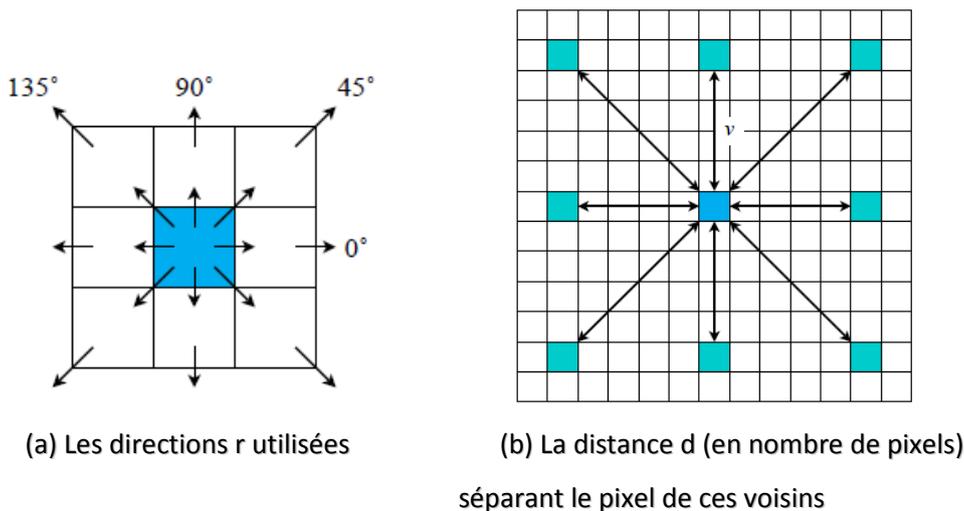


Figure 2.5 - Illustration de la direction et la distance utilisées dans la GLCM [38].

Si chaque image de la base est codée sur 256 niveaux de gris, alors la taille de la matrice de cooccurrence est de 256×256 , et pour une grande base d'images, on remarque qu'une énorme quantité d'informations qui doit être indexée. Au lieu d'utiliser la matrice complète, Haralick et al. [43] ont défini 14 caractéristiques statistiques de texture à partir de cette matrice, les plus utilisés sont:

- L'énergie qui interprète l'homogénéité globale de l'image [39],
- l'entropie qui fournit une information sur le désordre que peut présenter la texture [39],

- le contraste qui mesure l'uniformité de la texture [40]
- le moment différentiel inverse qui exprime l'homogénéité locale au sein de la texture [41].

b) Les méthodes basées sur la modélisation

Les méthodes basées sur la modélisation tentent de capturer le processus qui a généré la texture. En se servant des descripteurs qui sont basés sur la modélisation, certaines parties de l'image sont prises comme modèle et un algorithme d'estimation est utilisé pour définir les paramètres du modèle pour obtenir le meilleur ajustement [44]. Une image de texture peut être vue comme la réalisation d'un champ aléatoire, on suppose que l'image est modélisée par une fonction $f(r, w)$, où r est un vecteur de position représentant l'emplacement du pixel dans un espace 2D et w un paramètre aléatoire. Une fois qu'une texture spécifique w est choisie, $f(r, w)$ est une image, qui est une fonction définie sur une grille 2D indexée par r . La fonction $f(r, w)$ est appelée champ aléatoire. Les trois approches principales qui sont basées sur la modélisation de la texture sont les champs de Markov aléatoires proposés par Dubes et Jain [45], les fractales proposés par Pentland [46] et les caractéristiques d'auto-régression multi-résolution introduites par Mao et Jain [47].

c) Les méthodes spatio-fréquentielles

Les méthodes de description de la textures utilisant l'attribut spatio-fréquentiel se basent généralement sur la transformée de Fourier discrète, la transformée en cosinus discrète, la transformée en ondelettes discrète et les transformations de Gabor. La transformée de Fourier et la transformée en cosinus discrète permettent de représenter la texture uniquement dans le domaine fréquentiel. Ces transformées sont bien adaptées pour des textures grossières, où la totalité de l'information de la texture peut être représentée par un nombre réduit de coefficients qui correspondent aux basses fréquences. D'autre part, la transformation de Gabor et la transformée en ondelettes associent les deux représentations spatiale et fréquentielle. Parmi ces techniques, Manjunath et Ma [42] ont montré que la recherche d'images en utilisant les filtres de Gabor est plus performante que l'utilisation d'autres transformées. Aussi, les auteurs dans [48] ont montré que les moments calculés à partir des coefficients d'ondelettes de différentes bandes de fréquences peuvent être efficaces pour la représentation de la texture.

2.2.3 Descripteurs de forme

La forme est une autre caractéristique visuelle dans le domaine de la reconnaissance d'objets et de recherche d'images. Contrairement à la texture, la forme est un concept assez clairement défini qui représente des caractéristiques calculées à partir des contours d'un objet ou d'une région particulière de l'image. L'attribut forme d'objets ou de régions a été utilisé dans de nombreux systèmes de recherche d'images par le contenu [49, 50, 51, 52]. La reconnaissance de la forme est l'une des modalités à travers laquelle la perception humaine de l'environnement est effectuée, c'est un outil important dans les systèmes de recherche d'images par le contenu car elle correspond aux régions d'intérêts dans les images. Les représentations de la forme sont classées selon les techniques utilisées. On distingue deux techniques de modélisation de la forme: celle basée sur les contours et celle basée sur les régions [53]. Dans les techniques basées sur les régions, tous les pixels contenus dans une forme sont pris en compte pour obtenir la représentation de la forme. Ces méthodes utilisent généralement des moments pour décrire la forme [54], ces moments interprètent une fonction image de niveaux de gris normalisés comme une densité de probabilité d'une variable aléatoire 2D. Les sept premiers moments invariants dérivés des moments centraux normalisés du 2^{ème} et 3^{ème} ordre sont donnés par Hu [55]. Par comparaison aux descripteurs basés sur les régions, les descripteurs basés sur les contours sont plus populaires, ces derniers exploitent uniquement l'information des contours de la forme. Les plus simples incluent la surface, le périmètre, la compacité et l'orientation. Les plus compliqués incluent les descripteurs de Fourier, Grid Shape Descriptor et CHAIN CODES [53]. Les deux techniques de modélisation de la forme sont décrites rapidement ci-dessous:

2.2.3.1 Descripteurs de forme basés sur les contours

Les descripteurs de formes basés sur les contours capturent les caractéristiques d'une forme en se basant sur son contour, un exemple classique est le descripteur Curvature Scale-Space (CSS) [56] qui a été initialement proposé par Mokhtarian et al. [51], et qui est retenu dans la norme MPEG-7 pour décrire les formes [57]. Le descripteur se base sur la description des concavités d'une courbe et son filtrage successif pour suivre l'évolution de ses concavités. Il représente essentiellement les points d'inflexion d'une courbe pour caractériser les contours de la forme. Ce descripteur possède des propriétés importantes d'invariance en translation, rotation

et changement d'échelle, d'efficacité et de stabilité[58], ce qui permet une recherche par similarité très efficace.

2.2.3.2 Descripteurs de forme basés sur les régions

Les descripteurs de forme basés sur les régions intègrent tous les pixels situés dans la région occupée par la forme. Ils font principalement référence aux moments invariants. En plus de leur capacité à décrire diverses formes efficacement, ils sont aussi robustes aux transformations géométriques. Les processus d'extraction de caractéristiques et d'appariement sont simples, car ils ont une faible complexité de calcul [53].

Les moments d'image et leurs fonctions ont été utilisés dans de nombreuses applications de traitement d'image, à savoir, la reconnaissance des formes, la classification des images, l'identification des cibles, et l'analyse des formes. Ils comprennent les moments invariants, les moments de Zernike, les moments de Legendre et les moments exacts de Legendre. Les descripteurs basés sur les moments de Zernike s'avèrent très prometteurs [59, 60]; même ils ont été adoptés par la norme MPEG-7 comme descripteurs de forme basés sur les régions [57].

a) Les moments géométriques

Les moments géométriques m_{pq} d'ordre $(p+q)$ d'une fonction $f(x, y)$ peuvent être considérés comme une projection d'une fonction sur une base de polynômes $x^p y^q$. Ils sont définis par la formule:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y); p, q = 0, 1, 2, \dots \quad (2.1)$$

L'utilisation de ces moments pour l'analyse et la reconnaissance d'images a été proposée par Hu [55], où il a proposé sept moments qui sont invariants en translation, en rotation et en changement d'échelle [59, 60].

Pour une image f , le moment d'ordre $(p+q)$ est donné par la formule :

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^p y^q f(x, y); p, q = 0, 1, 2, \dots \quad (2.2)$$

Où M , N sont respectivement le nombre de lignes et de colonnes de l'image f et $f(x, y)$ est l'intensité du pixel (x, y) dans l'image.

Le moment m_{00} d'ordre 0, représente la surface de l'objet et le rapport des moments d'ordre 1 (m_{01}, m_{10}) et m_{00} représente le centre de gravité de la surface de l'objet, qui est donné par :

$$\bar{x} = \frac{m_{10}}{m_{00}} ; \bar{y} = \frac{m_{01}}{m_{00}} \quad (2.3)$$

De même, les moments géométriques centraux qui sont invariants en translation sont définis par:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y); p, q = 0, 1, 2, \dots \quad (2.4)$$

Pour obtenir une invariance aux changements d'échelle, on normalise les moments centrés en utilisant la formule suivante:

$$\eta_{pq} = \frac{\mu_{pq}}{m_{00}^{\frac{1+p+q}{2}}} , p + q \geq 2 \quad (2.5)$$

En utilisant ces moments centraux jusqu'à l'ordre trois, les sept fonctions des moments invariants de Hu sont données par les équations suivantes:

$$I_1 = \eta_{20} + \eta_{02} \quad (2.6)$$

$$I_2 = (\eta_{20} + \eta_{02})^2 + (2\eta_{11})^2 \quad (2.7)$$

$$I_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (2.8)$$

$$I_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (2.9)$$

$$I_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} - 3\eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (2.10)$$

$$I_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (2.11)$$

$$I_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (2.12)$$

b) Les moments orthogonaux

Les moments orthogonaux permettent de représenter ou de reconstruire une forme, avec précision et avec une quantité minimale de redondance de l'information [61, 62]. Les plus intéressants sont les moments orthogonaux de Zernike, de Legendre et les moments exacts de Legendre. Ils sont largement utilisés dans les systèmes de recherche d'images par le contenu comme descripteurs de forme [63, 64, 68, 69].

• Les moments de Zernike

Les moments de Zernike sont définis à partir des polynômes complexes de Zernike qui forment un ensemble orthogonal complet sur le disque unité

Les moments de Zernike d'ordre p et de répétition q sont définis sur $D = \{(p, q) | 0 \leq p \leq \infty, |p - q| \text{ est pair}\}$:

$$Z_{pq} = \frac{p+1}{\pi} \iint_{x^2+y^2 \leq 1} V_{pq}^*(x, y) f(x, y) dx dy \quad (2.13)$$

Où V_{pq}^* est le conjugué complexe de V_{pq} , et V_{pq} est défini par:

$$V_{pq}(\rho, \theta) = R_{pq}(\rho) \cdot e^{iq\theta} \quad (2.14)$$

Et

$$R_{pq}(\rho) = \sum_{\substack{k=|q| \\ |p-k| \text{ pair}}}^p \frac{(-1)^{\frac{p-k}{2}} \frac{p+q}{2}}{\frac{p-q}{2} \frac{k-q}{2} \frac{k+q}{2}} \rho^k \quad (2.15)$$

Les moments de Zernike sont invariants en rotation. En outre, ils sont robustes au bruit et aux variations mineures de forme. Une image peut être mieux décrite par un petit ensemble de ses moments de Zernike que tous les autres types de moments tels que les moments géométriques [65, 66]. Un ensemble relativement petit des moments de Zernike peut caractériser la forme globale du motif, les moments d'ordre inférieur représentent la forme globale du motif et les moments d'ordre

supérieur représentent le détail. En raison de leurs caractéristiques, les moments de Zernike ont été utilisés comme descripteurs dans de nombreuses applications telles que la reconnaissance de forme [62, 67] et la recherche d'images par le contenu [63].

- **Les moments de Legendre**

Les moments de Legendre utilisent les polynômes de Legendre comme fonction de noyau. Les moments à deux dimensions de Legendre d'ordre $(p + q)$ d'une image $f(x, y)$ sont définis par l'équation:

$$L_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-1}^1 \int_{-1}^1 P_p(x) \cdot P_q(y) f(x, y) dx dy \quad x, y \in [-1, 1] \quad (2.16)$$

Où le polynôme de Legendre $P_p(x)$ d'ordre 'P' est donné par l'équation:

$$P_p(x) = \frac{1}{2^p p!} \frac{d^p}{dx^p} (x^2 - 1)^p \quad x \in [-1, 1] \quad (2.17)$$

Les moments de Legendre ont été utilisés dans plusieurs applications de reconnaissance de forme [70]. Ils permettent de représenter les caractéristiques indépendantes de l'image avec une quantité de redondance d'information qui est presque nulle [71]. Les fonctions invariantes en translation et en changement d'échelle sont obtenues en utilisant les invariants correspondants des moments géométriques, ou en normalisant les images reçues après translation et changement d'échelle en utilisant les moments complexes ou les moments géométriques [72].

- **Les moments exacts de Legendre**

Les moments exacts de Legendre sont des moments continus et orthogonaux, ils peuvent être utilisés pour représenter une image avec une quantité minimum de redondance de l'information. De nombreux algorithmes sont développés pour le calcul des moments exacts de Legendre [73, 74, 75], mais ces méthodes se concentrent principalement sur les moments géométriques en 2-D. Quand elles sont appliquées à une image, une approximation numérique est nécessaire. L'erreur due à cette approximation augmente lorsque l'ordre du moment augmente. Une méthode précise pour calculer les moments exacts de Legendre est proposée par Hosney [76].

2.2.4 Combinaison de descripteurs

La description du contenu visuel d'une image peut être effectuée par un attribut unique résumant son contenu de couleur, de texture ou de forme, comme elle peut être effectuée par une combinaison de plusieurs attributs. La limite principale d'utiliser un seul attribut comme descripteur c'est que : un seul attribut n'est plus suffisant pour décrire le contenu de l'image, par exemple, des images ayant des textures ou des formes très différentes mais des informations de couleur similaires peuvent être retournées par le système en tant qu'images correspondantes, si seule la caractéristique de couleur est considérée. De même, une technique basée sur la caractéristique de forme peut récupérer une image très différente à une image requête si leurs formes sont similaires, la figure 2.3 illustre bien cette limite où deux images qui sont sémantiquement totalement différentes, mais elles possèdent deux histogrammes de couleurs qui sont similaires. Les performances des applications basées sur des caractéristiques uniques se dégradent parfois de manière significative surtout dans la présence d'une grande variété d'images. En effet, les travaux de recherche actuels tendent à combiner plusieurs caractéristiques pour réduire l'écart sémantique et augmenter les performances des systèmes de recherche.

La combinaison des caractéristiques telles que la couleur et la texture, la couleur et la forme, ou la forme et la texture a été largement utilisée pour distinguer des images, elle constitue une bonne solution pour une recherche plus efficace et plus discriminante. Dans [77], les auteurs ont combiné des ondelettes de Gabor avec l'histogramme de corrélation de couleurs pour calculer un descripteur, nommé Wavelet Correlogram, pour la recherche d'images basée sur le contenu. Egalement, les auteurs dans [78] ont proposé de classifier des images de texture en utilisant une transformée en ondelettes à plusieurs échelles, ils ont utilisé la moyenne et l'entropie dans le domaine fréquentiel pour former le vecteur descripteur. Par ailleurs, dans [79], les auteurs ont proposé une technique de description basée sur une combinaison pondérée de plusieurs descripteurs MPEG-7, où un algorithme par essaim de particules est utilisé pour ajuster les poids des différents descripteurs. Dans le même contexte, Les auteurs dans [48] ont présenté une combinaison d'ondelettes et des moments pour la recherche d'images par le contenu, où les vecteurs descripteurs sont formés en utilisant des moments géométriques calculés à

partir des coefficients normalisés d'une transformée en ondelettes discrète appliquée sur des images en niveaux de gris.

2.3 Mesures de similarité

La recherche d'images peut s'interpréter comme le calcul d'une similarité entre le descripteur de l'image requête et ceux des images de la base de données. Généralement, la similarité entre deux images peut être modélisée par une mesure de distance dans un espace métrique multidimensionnel. Le choix de la mesure de distance a un impact direct sur les performances des systèmes de recherche d'images par le contenu, et le type de vecteurs de caractéristiques choisis détermine le type de mesure de distance qui sera utilisée pour comparer leur similarité [80]. En outre, la mesure de distance utilisée doit exprimer la perception humaine, c.-à-d. deux images perceptivement similaires doivent avoir une distance faible, réciproquement deux images perceptivement différentes doivent avoir des distances importantes. En plus, puisque la recherche d'images s'effectue en temps réel, il faut tenir en compte la complexité de calcul de la distance, car elle doit être raisonnable.

En mathématiques, dans un espace métrique ε , on appelle une distance d une fonction de ε à valeur dans R^+ (ensemble des réels positifs) qui vérifie les propriétés suivantes:

$$\begin{aligned} \forall a, b \in \varepsilon, d(a, b) &\geq 0 \\ \forall a, b \in \varepsilon, d(a, b) &= d(b, a) \\ \forall a, b \in \varepsilon, d(a, b) = 0 &\Leftrightarrow a = b \\ \forall a, b \in \varepsilon, d(a, b) + d(b, c) &\geq d(a, c) \end{aligned}$$

En particulier, si les caractéristiques extraites à partir des images sont présentées sous forme de vecteurs multidimensionnels, on peut exprimer différentes distances entre deux vecteurs ainsi:

Soit q une image requête, et t une image de la base de données,

- **La distance de Minkowski**

C'est la distance la plus simple utilisée pour calculer la similarité entre deux images, elle est définie par la formule:

$$D_r = (\sum_{i=1}^n |q_i - t_i|^r)^{\frac{1}{r}} \quad (2.18)$$

Où $r \geq 1$, est le facteur de Minkowski et n est la dimension de l'espace de descripteurs.

- Pour $r=1$ on obtient **la distance de Manhattan**

Cette distance est connue sous le nom de City-block. Le standard MPEG-7 [57] recommande l'utilisation de cette distance pour comparer deux descripteurs ART (angular Radial Transform). La distance de Manhattan entre q et t est donnée par l'équation:

$$D_{Man}(q, t) = \sum_{i=1}^n |q_i - t_i| \quad (2.19)$$

- Pour $r=2$, on obtient **la distance euclidienne**

C'est la mesure de distance utilisée fréquemment pour mesurer la distance entre deux points dans un espace multidimensionnel [81]. Pour d'autres types de caractéristiques tels que les histogrammes de couleur, cette distance n'est pas optimale pour comparer ce type de descripteurs. L'intersection d'histogrammes [32] et l'entropie ont été parmi plusieurs mesures pour calculer la similarité entre deux histogrammes.

La distance euclidienne entre q et t est donnée par l'équation:

$$D_{Ece}(q, t) = \sqrt{\sum_{i=1}^n (q_i - t_i)^2} \quad (2.20)$$

Une petite valeur pour la distance euclidienne indique une similarité entre l'image q et l'image t.

- Pour $r = \infty$, on obtient **la distance de Chebychev**

Cette distance est souvent employée pour des données de grandes dimensions, surtout pour des applications où le temps de réponse est important. La distance de Chebychev entre q et t est donnée par l'équation:

$$D_{Che}(q, t) = \max_{1 \leq i \leq n} |q_i - t_i| \quad (2.21)$$

D'autres types de mesures de similarité sont utilisées par les applications de la recherche d'images par le contenu telles que: l'intersection d'histogrammes, la distance de Bhattacharya et la distance de Mahalanobis.

- **Intersection d'histogrammes**

C'est une mesure de similarité utilisée pour comparer deux histogrammes de couleurs. Elle permet de calculer les parties communes de deux histogrammes en négligeant les parties propres de ces deux derniers. L'intersection de deux histogrammes H , et H' , telle qu'elle est définie par Swain et Ballard [32], est donnée par l'équation:

$$D_{Swain}(H, H') = \frac{\sum_{m=1}^M \min(H_m, H'_m)}{\sum_{m=1}^M H_m} \quad (2.22)$$

Où M est le nombre de valeurs de chaque histogramme. Cette mesure n'est pas une distance puisque elle ne vérifie pas la propriété de symétrie.

Cette mesure a été étendue par Smith et Chang [82] en:

$$D_{hi}(H, H') = \frac{\sum_{m=1}^M \min(H_m, H'_m)}{\min(\sum_{m=1}^M H_m, \sum_{m=1}^M H'_m)} \quad (2.23)$$

Une version de mesure de similarité qui vérifie la propriété de symétrie et de l'inégalité triangulaire est celle définie dans [83]: où

$$D_{hi}(H, H') = \sum_{m=1}^M \left(H_m \log \frac{2H_m}{H_m + H'_m} + H'_m \log \frac{2H'_m}{H_m + H'_m} \right) \quad (2.24)$$

- **Distance de Bhattacharyya**

Cette distance mesure la similarité entre deux distributions de probabilité discrètes ou continues, c'est la distance populaire utilisée pour calculer la similarité entre deux distributions gaussiennes. La distance de Bhattacharyya entre une image requête q et une image t de la base de données est donnée par l'équation:

$$D_{Bhat}(q, t) = \frac{1}{8} (\mu_q - \mu_t)^T \left[\frac{\Sigma_q + \Sigma_t}{2} \right]^{-1} (\mu_q - \mu_t) + \frac{1}{2} \ln \frac{|(\Sigma_q + \Sigma_t)/2|}{\sqrt{|\Sigma_q| |\Sigma_t|}} \quad (2.25)$$

Où μ_q et μ_t sont respectivement les vecteurs moyens et Σ_q et Σ_t les matrices de covariance de q et de t.

- **La distance de Mahalanobis**

Cette distance est basée sur les corrélations entre les variables. Contrairement aux autres distances, cette dernière tient compte de la distribution statistique des données. La distance de Mahalanobis entre une image requête q et une image t de la base de données est définie par l'équation:

$$D_{Mahala}(q, t) = \sqrt{(\mu_q - \mu_t)^T \Sigma^{-1} (\mu_q - \mu_t)} \quad (2.26)$$

Où Σ est la matrice des covariances entre les descripteurs d'images.

2.4 Structure d'indexation

Un autre problème important en recherche d'images par le contenu est l'indexation efficace et la recherche rapide des images qui se base sur des caractéristiques visuelles. Vu que les vecteurs de caractéristiques des images peuvent avoir une grande dimension, ils ne sont pas bien adaptés aux structures d'indexation traditionnelles; la réduction de la dimension est généralement utilisée avant la création d'un système d'indexation efficace. L'une des techniques couramment utilisées pour la réduction de dimension est l'analyse en composantes principales (ACP). C'est une technique optimale qui fait correspondre de façon linéaire les données d'entrée dans un espace de coordonnées de telle sorte que les axes soient alignés pour tenir compte des variations maximales dans les données. Outre l'analyse en composantes principales, les réseaux de neurones ont également été démontrés comme un outil utile pour la réduction de la dimension des vecteurs de caractéristiques [85].

Après réduction de la dimension, les données multidimensionnelles sont indexées. De nombreuses approches ont été proposées à cet effet, y compris les R-tree (en particulier, R * -tree [86]), les K-d-B tree [87] et les grid files [88]. La plupart de ces méthodes d'indexation multidimensionnelles ont des performances raisonnables pour les petites dimensions (jusqu'à 20), mais malheureusement avec l'augmentation de la dimension, leurs performances se réduisent à celles d'une recherche séquentielle. En outre, ces schémas d'indexation supposent que la mesure de comparaison se base sur la distance

euclidienne, ce qui n'est pas nécessairement vrai pour de nombreuses applications de recherche d'images. Egalement les cartes auto-organisation (Self organizing maps: SOM) sont parmi les structures d'indexation utilisées dans les systèmes de recherche d'images par le contenu [89, 90]. Outre de bénéficier de l'indexation, SOM fournit aux utilisateurs un moyen utile pour parcourir les images représentatives de chaque type.

2.5 Les techniques d'apprentissage

Comme les attributs de bas niveau ne peuvent pas grouper des images dans des classes sémantiques, il faut utiliser des outils formels tels que les techniques d'apprentissage automatique supervisé ou non supervisé [91, 92, 93, 94]. L'objectif de l'apprentissage supervisé est de prédire la valeur d'un résultat (par exemple la classe sémantique) en se basant sur un ensemble de mesures en entrée (caractéristiques de bas niveau). Dans l'apprentissage non supervisé, il n'y a pas de mesures de résultats, ici l'objectif est de décrire comment les données d'entrée sont organisées ou classées [95]. Ces techniques peuvent être considérées comme une étape pour accélérer la recherche d'images dans de grandes bases de données et d'améliorer la précision de la recherche.

2.5.1 Les techniques d'apprentissage supervisé

Les techniques d'apprentissage supervisé telles que les séparateurs à vastes marges (SVM) [96, 97, 98], les classificateurs bayésiens [99] sont souvent utilisés pour apprendre les concepts de haut niveau à partir des caractéristiques de bas niveau d'image. Ces techniques réalisent le processus de classification à l'aide des données d'apprentissage déjà classées. Pour les données d'apprentissage, l'entrée (les caractéristiques de bas niveau de l'image) et la sortie désirée (la classe) sont déjà connues. Etant donné une image requête, Ses caractéristiques de bas niveau sont extraites et elles sont passées en tant qu'entrée à l'un des algorithmes d'apprentissage automatique, qui sont déjà entraînés avec les données d'apprentissage. L'algorithme d'apprentissage automatique prédit la classe de l'image requête qui est le concept sémantique de cette dernière. Donc, au lieu de trouver la similarité entre l'image requête et toutes les images dans la base de données, il trouve la similarité entre l'image requête et seulement les images appartenant à la classe d'image requête. De même, lorsque l'ensemble de la base de données est parcouru, le résultat de recherche contient des images de différentes classes. Mais, lorsque les techniques d'apprentissage automatique

sont utilisées, sauf les images de la classe de la requête (concept sémantique) sont prédites, les résultats de la recherche contiennent les images appartenant à cette unique classe.

Vu leurs bases théoriques solides disponibles, les SVM ont été utilisés pour la reconnaissance d'objets, la classification des textes, et ils sont considérés comme des supports puissants pour l'apprentissage dans les systèmes de recherche d'images [100, 101]. Les SVM ont été initialement conçus pour la classification binaire. Supposons que nous avons un ensemble de données d'apprentissage comme des vecteurs dans un espace appartenant à deux classes avec leurs étiquettes. L'objectif de la méthode est de trouver un hyper-plan optimal (Optimal Separating Hyperplane) pour séparer les données en maximisant les marges entre les deux classes (voir figure 2.6). L'hyperplan est déterminé de façon que la distance entre ce dernier et les points les plus proches dans chaque classe soit maximale. Une fois déterminé, les points les plus proches sont appelés vecteurs support et la distance entre les vecteurs supports et l'hyperplan est appelée marge. Pour apprendre plusieurs concepts pour la recherche d'images, un SVM doit être entraîné pour chaque classe. Par exemple, dans [97], un SVM est employé pour l'annotation des images. Lors de l'apprentissage, un SVM binaire est entraîné pour chacune des 23 classes choisies, et lors de l'étape de test, les images non étiquetées sont introduites dans tous les modèles où la classe dont le résultat positif est plus élevé est associée à l'image.

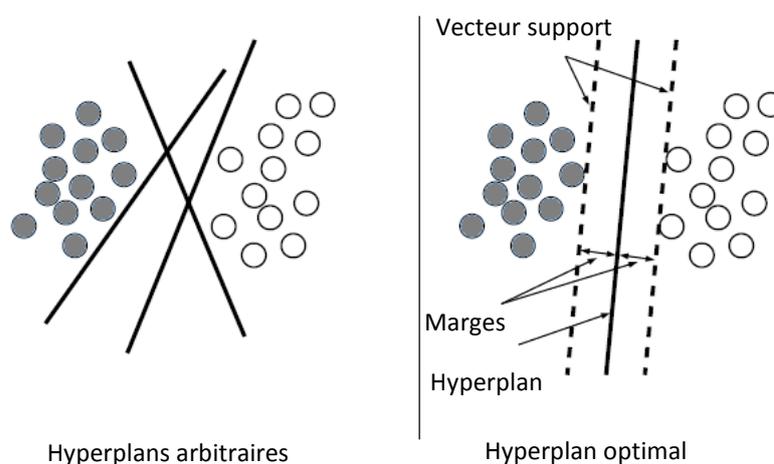


Figure 2.6- Séparation de deux classes de données par SVM.

Une autre méthode d'apprentissage largement utilisée est la classification bayésienne [102]. Dans [93] en utilisant un classificateur bayésien binaire, les concepts de haut niveau qui représentent les classes des scènes naturelles sont obtenues à partir des caractéristiques de bas niveau des images, où les images de la base sont automatiquement classées en types généraux comme intérieur ou extérieur et les images d'extérieur ensuite sont classées en deux classes ville ou paysage, etc. Aussi, dans [94], un réseau bayésien est utilisé pour le classement d'images.

Autres techniques d'apprentissage comme les réseaux neuronaux ont été également utilisées pour l'apprentissage des concepts de haut niveau. Dans [92], après avoir choisi 11 classes: brique, nuage, fourrure, herbe, glace, route, roche, sable, peau, arbre et eau, une grande quantité de données d'apprentissage (des caractéristiques de bas niveau) est introduite dans le réseau neuronal pour établir le lien entre les caractéristiques de bas niveau d'une image et sa sémantique de haut niveau (étiquette de la classe). L'inconvénient de ces algorithmes est qu'ils nécessitent une grande quantité de données d'apprentissage, ce qui implique des calculs très intensifs [103]. Les auteurs dans [96] ont indiqué que les algorithmes d'apprentissage classiques souffrent de deux problèmes: une grande quantité d'échantillons d'apprentissage étiquetés avec leurs classes sémantiques est nécessaire, car réellement, il est très difficile de fournir de telles données; l'ensemble d'apprentissage est fixé pendant la phase d'apprentissage pour une application bien définie, si le domaine d'application change, de nouveaux échantillons étiquetés doivent être fournis pour assurer l'efficacité du classificateur. Pour résoudre ce problème, une approche est représentée dans [96], en se basant sur un algorithme de co-apprentissage, un petit ensemble d'apprentissage est utilisé pour entraîner deux classificateurs qui sont statistiquement indépendants et co-annoter les échantillons non étiquetés. En comparant leur méthode, en utilisant seulement 20 images étiquetées pour l'apprentissage avec un SVM en utilisant 400 images étiquetées, leurs résultats montrent qu'une amélioration de 10 % dans la précision de la recherche est obtenue par rapport à un SVM utilisé tout seul.

Outre ces algorithmes, les arbres de décision ont été également utilisés pour prédire les caractéristiques sémantiques (les classes des images). Les méthodes d'induction basées sur les arbres de décision telles que ID3, C4.5 (une version améliorée de ID3) et CART permettent de construire une structure arborescente en divisant récursivement l'espace

d'attributs en un ensemble d'espaces non chevauchés [95]. Un ensemble de règles de décision peut être obtenu en traversant l'arbre de la racine vers les feuilles. Par exemple dans [91], en se servant de l'algorithme basé sur l'arbre de décision CART, des règles de décision ont été dérivées pour faire correspondre la distribution globale des couleurs dans une image donnée à sa description textuelle.

Comparée avec les autres méthodes d'apprentissage supervisé, la méthode d'apprentissage par arbre de décision est conceptuellement simple et robuste par rapport aux données d'entrée incomplètes, en outre les arbres de décision peuvent être facilement traduites en un ensemble de règles qui peuvent être intégrées dans un système expert pour la prise de décision automatique [95, 104]. Pareillement, les SVM et les arbres de décision ont été combinés dans une technique dite: SVM-BDT (SVM-Binary Decision tree) [105, 106], la technique SVM-BDT combine l'avantage du calcul efficace de l'architecture des arbres binaires, et la classification précise des SVM. Cette technique s'effectue en deux étapes principales, la première consiste à construire la SVM-BDT en regroupant les différentes classes de la base de données. Dans la deuxième étape un SVM binaire est associé à chaque nœud de l'arbre construite à l'étape précédente. Pour un problème d'apprentissage à N classes, le modèle sémantique de chaque classe (centre de classe) est calculé, les distances entre les modèles sémantiques de toutes les N classes sont calculées, ainsi une matrice de distance de dimension $N \times N$ est obtenue. Les deux classes qui ont la plus grande distance entre elles sont choisies pour former les deux premiers clusters de l'arbre, leurs modèles sémantiques sont considérés comme centre de cluster. Ensuite, chaque paire de classes est affectée au cluster qui est le plus proche et le centre du cluster est mis à jour, toutes les classes sont affectées à l'un des deux clusters possibles. Et le processus se poursuit, jusqu'à ce qu'il y a une seule classe par groupe. Un SVM binaire est utilisé pour entraîner les échantillons dans chaque nœud non feuille de l'arbre de décision [106]. La figure 2.7 représente un exemple d'un arbre SVM-BDT construit à partir de 10 classes.

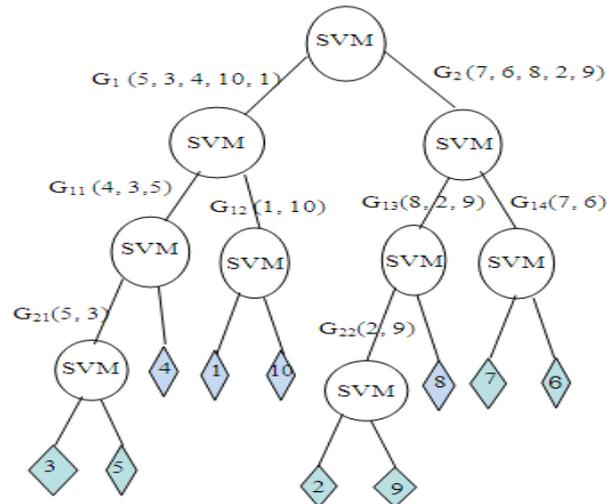


Figure 2.7 - Exemple d'un SVM-BDT [105].

Lors de la phase de test, le vecteur de caractéristique de l'image requête est donné comme entrée du SVM-BDT, le SVM binaire à chaque nœud non feuille est utilisé pour faire traverser l'arbre et ainsi étiqueter la requête avec sa classe. Par conséquent, les distances de similarité entre l'image requête et les images de la classe prédite sont calculées et les images avec une faible distance sont récupérées.

2.5.2 Les techniques d'apprentissage non supervisé

Contrairement à l'apprentissage supervisé dans lequel la présence de la variable résultat guide le processus d'apprentissage, l'apprentissage non supervisé n'a aucune mesure de résultat, la tâche est plutôt de découvrir comment les caractéristiques d'entrée sont organisées ou regroupées. Le regroupement d'images (Clustering en anglais) est la technique typique d'apprentissage non supervisé utilisée pour la recherche. Il permet de regrouper un ensemble de données d'images de manière à maximiser la similarité à l'intérieur des clusters et de minimiser la similarité entre les différents clusters. Chaque cluster résultant est associé à une étiquette d'une classe et les images dans le même cluster sont censées être similaires.

L'algorithme de regroupement en K-moyennes (ou K-means en anglais) traditionnel et ses variantes sont souvent utilisés pour le regroupement d'images. L'algorithme vise à diviser n observations données en k groupes (ou clusters). La moyenne de chaque cluster est calculée et l'image est placée dans un cluster, dont la distance euclidienne entre la moyenne de celui-ci et le vecteur caractéristique de l'image est minimale.

Par exemple dans [107], l'algorithme des k-moyennes est appliqué aux caractéristiques de bas niveau de couleur d'un ensemble d'images d'apprentissage. Ensuite, des statistiques mesurant la variation avec chaque cluster sont utilisées pour dériver un ensemble de correspondances entre les caractéristiques de bas niveau et la caractérisation textuelle optimale (mots clés) du cluster correspondant. Les règles de correspondance dérivées ont été utilisées pour indexer de nouvelles images non étiquetées ajoutées à la base de données.

En raison de la distribution complexe des données d'image, les méthodes traditionnelles telles que l'algorithme des k-moyennes ne peuvent souvent pas séparer des images avec des concepts différents [108]. Pour traiter ce problème, la méthode de regroupement N-Cut [109] a été proposée et utilisée avec succès dans plusieurs applications telles que la segmentation d'images, le regroupement d'images. Une extension de la méthode N-Cut peut être consultée dans [110].

Dans [111], une méthode appelée 'CLUE' est présentée pour réduire le 'fossé sémantique' dans les CBIR. Contrairement aux autres systèmes CBIR qui affichent comme résultat, les images qui sont considérées similaires à la requête au sens de la distance utilisée, ce système tente de récupérer des clusters d'image sémantiquement cohérents. Étant donné une image requête, une collection d'images similaires à la requête est sélectionnée comme voisin de la requête. Basé sur l'hypothèse que les images de la même sémantique ont tendance à être regroupées; N-Cut clustering est utilisé pour regrouper ces images cibles dans différentes classes sémantiques. Ensuite, le système affiche les clusters d'images et ajuste le modèle de mesure de similarité en fonction des commentaires des utilisateurs. Bien que réussie dans le regroupement de données multiples, N-Cut ne peut pas produire une fonction de correspondance explicite. Pour traiter les nouveaux points de données, les similarités entre les nouveaux points et toutes les données d'apprentissage doivent être mesurées. Le calcul des similarités pourrait être très compliqué en raison de la grande taille de l'ensemble d'apprentissage [108]. Pour remédier à ces problèmes, dans [108], une méthode dite locality preserving clustering (LPC) est proposée pour le regroupement d'images, cette méthode peut fournir une fonction de correspondance explicite. Les résultats expérimentaux montrent que la LPC donne une précision de recherche comparable à

celle de l'algorithme N-Cut. En outre, le résultat de recherche de LPC est prouvé être plus précis que celui des k-moyennes.

Egalement, les classificateurs probabilistes basés sur la théorie de Bayes sont parmi les outils de regroupement les plus puissants. Le classificateur maximum-a-posteriori (MAP), tout comme la méthode du maximum de vraisemblance (ML) sont de plus en plus prometteurs pour la recherche d'images par le contenu [112,113]. Cependant, Il est difficile d'appliquer ces classificateurs en raison de la complexité de la fonction de similarité MAP, car dans [113], l'auteur a montré que la fonction de similarité peut être calculée efficacement lorsque des quantificateurs vectoriels et des mélanges gaussiens sont utilisés comme modèles pour des fonctions de densité de probabilité des caractéristiques de l'image.

Aussi, les modèles de regroupement flou fournissent une solution prometteuse au problème de regroupement. L'algorithme de regroupement en K-moyennes flou (Fuzzy c-means en anglais ou FCM) est l'algorithme de classification floue le plus fréquemment utilisé. Cet algorithme est basé sur une optimisation itérative d'une fonction objective floue. Le degré d'appartenance d'un élément de données à un cluster est compris entre 0 et 1. Pour obtenir une image requête donnée, la sortie du FCM est la valeur d'appartenance de l'image avec chacun des K classes. L'image requête appartient à la classe pour laquelle la valeur d'appartenance est élevée [114, 105, 84].

2.6 Spécification de la requête

Spécifier quel type d'images l'utilisateur souhaite récupérer à partir de la base de données peut être fait de plusieurs façons. Les formulations de requête couramment utilisés sont: la navigation par catégorie, la recherche par concept, la requête par croquis, et la recherche par l'exemple. La navigation par catégorie consiste à parcourir la base de données en fonction des catégories d'images. A cet effet, les images de la base de données doivent être classées en différentes catégories en fonction de leur contenu sémantique ou visuel [115, 116]. La requête par concept consiste à récupérer des images conformément à la description conceptuelle associée à chaque image dans la base de données. La requête par croquis et la requête par l'exemple [117] sont faites par fournir un croquis ou un exemple d'image à partir de laquelle les images avec des

caractéristiques visuelles similaires seront extraites de la base de données. Les deux premiers types de requêtes sont liés à la description sémantique des images.

La requête par croquis permet à l'utilisateur de dessiner une esquisse d'une image avec un outil d'édition graphique fourni soit par le système de recherche ou par un autre logiciel. Les requêtes peuvent être formées en tirant plusieurs objets avec certaines propriétés comme la couleur, la texture, la forme, la taille et les emplacements. Dans la plupart des cas, une esquisse grossière est suffisante, et la requête peut être affinée en fonction des résultats de recherche.

La recherche par l'exemple permet à l'utilisateur de formuler une requête en fournissant une image par exemple. Le système convertit cette image en une représentation interne de caractéristiques. Les images stockées dans la base de données avec des caractéristiques similaires sont ensuite récupérées.

La requête par groupe d'exemples permet aux utilisateurs de sélectionner plusieurs images. Le système va alors trouver les images qui correspondent le mieux aux caractéristiques communes du groupe d'exemples. De cette façon, une cible peut être définie plus précisément en spécifiant les variations de caractéristiques pertinentes et en éliminant les variations non pertinentes dans la requête. En outre, les propriétés du groupe peuvent être affinées en ajoutant des exemples négatifs. De nombreux systèmes développés fournissent à la fois la requête par des exemples positifs et négatifs.

2.7 Retour de pertinence

La perception humaine de la similarité des images est subjective, sémantique, et dépend de la tâche à réaliser. Bien que les méthodes basées sur le contenu fournissent des orientations prometteuses pour la recherche d'images, généralement, les résultats de la recherche basée sur la similarité entre les caractéristiques visuelles pures ne sont pas nécessairement perceptivement et sémantiquement significatifs. En outre, chaque type de caractéristique visuelle tend à capturer un seul aspect de la propriété de l'image, et il est généralement difficile pour un utilisateur de spécifier clairement comment les différents aspects sont combinés. Pour résoudre ces problèmes, les systèmes récents intègrent l'utilisateur, dans une phase de bouclage de pertinence (ou relevance feedback en anglais), pour guider et affiner les résultats de la recherche. Le retour de pertinence

interactif a été déjà utilisé par les systèmes de recherche d'informations basés sur le texte traditionnels [118]. Il permet d'établir un lien entre les concepts de haut niveau et les caractéristiques de bas niveau et d'améliorer les performances des systèmes de recherche d'images par le contenu [119, 120]. Son principe est le suivant [121]: pour une requête donnée par l'utilisateur, le système récupère d'abord une liste d'images classées selon une métrique de similarité prédéfinie. Ensuite, l'utilisateur marque les images extraites comme (exemples positifs) pertinents à la requête ou (exemples négatifs) ne sont pas pertinents. En appliquant des techniques d'apprentissage automatique pour apprendre le feedback de l'utilisateur, le système va affiner les résultats de récupération et présenter une nouvelle liste d'images à l'utilisateur, ce processus se poursuivra jusqu'à la satisfaction de l'utilisateur. Une manière souvent utilisée pour apprendre les feedbacks de l'utilisateur est d'ajuster des poids aux différentes caractéristiques de bas niveau et /ou ajuster la mesure de similarité [122, 119]. Une étude détaillée sur les systèmes de recherche d'images utilisant le retour de pertinence peut être consultée dans [123]

2.8 Evaluation des performances

Pour évaluer un système de recherche d'images par le contenu et comparer les performances de différentes approches, une manière générale, suivie par la communauté scientifique est d'utiliser des bases de données avec vérité terrain (Ground truth). Autrement dit, les images de la base de données doivent être classées en classes d'images disjointes qui couvrent l'ensemble de la base de données. Pour une image requête Q appartenant à une classe C , toute image appartenant à C est considérée comme pertinente à la requête.

Deux mesures empruntées de la recherche d'information traditionnelle sont largement utilisées, à savoir, le rappel et la précision [124]. La précision de la recherche est définie comme le pourcentage d'images pertinentes retrouvées au sein de l'ensemble des images retournées par le système ; le rappel donne le pourcentage d'images pertinentes retournées par rapport au nombre total d'images pertinentes dans la base, la précision et le rappel sont définis par :

$$Rappel = \frac{|images\ pertinentes\ retrouvées|}{|images\ pertinentes|} \quad (2.27)$$

$$Précision = \frac{|images\ pertinentes\ retrouvées|}{|images\ retrouvées|} \quad (2.28)$$

Où $|E|$ désigne la cardinalité de l'ensemble E.

En général, la sélection d'un ensemble de données pertinentes est beaucoup moins stable en raison de diverses interprétations des images. En outre, lorsque le nombre d'images pertinentes est supérieur au nombre des images récupérées, le rappel n'a pas de sens. Par conséquent, la précision et le rappel sont des descriptions seulement approximatives de la performance du système de récupération.

Une représentation graphique des deux mesures peut être illustrée par la figure 2.8. Cette courbe permet de tracer la précision en fonction du rappel. Elle correspond à la performance du système pour une requête donnée en fonction du pourcentage d'images pertinentes retournées. Un système parfait aura une courbe constante de valeur 1. Plus les courbes tendent vers cette dernière plus les méthodes sont plus performantes.

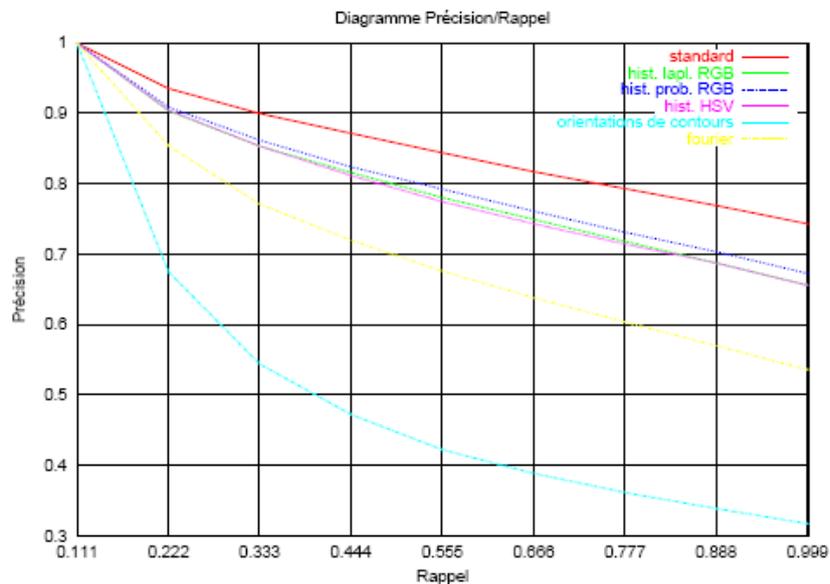


Figure 2.8 - Courbe de rappel et précision du système Ikona.

Une autre mesure d'évaluation, qui est recommandée par le standard MPEG-7, est la norme: ANMRR (Average Normalized Modified Retrieval Rank) [125], elle tient à la

fois compte du nombre d'images pertinentes retournées par le système et de l'ordre de ces images dans la liste des résultats. Elle est définie comme la moyenne des NMRR (Normalized Modified Retrieval Rank) sur le nombre de requêtes NQ et qui est donnée par l'équation :

$$ANMRR = \frac{1}{NQ} \sum_{q=1}^{NQ} NMRR(q) \quad (2.29)$$

La NMRR est définie par l'équation suivante:

$$NMRR = \frac{\left(\sum_{k=1}^{NG(I_q)} \frac{Rank(I_k)}{NG(I_q)} \right) - 0.5 - \frac{NG(I_q)}{2}}{K(I_q) + 0.5 - 0.5 * NG(I_q)} \quad (2.30)$$

Où $NG(I_q)$ est le nombre d'éléments dans la vérité terrain associée à l'image requête I_q , $Rank(I_k)$ est égale à l'ordre selon lequel est retrouvée la $K^{ième}$ image de la vérité terrain, $K(I_q)$ la valeur du classement valide associé à l'image requête I_q , cette valeur dépend de la taille de la vérité terrain.

NMRR varie de 0 à 1. Plus sa valeur est petite, meilleure est l'efficacité de la recherche. Il est évident que cette mesure n'aura d'intérêt pratique qu'au cas où les images de la vérité terrain sont ordonnées, ce qui n'est en général pas le cas.

2.9 Conclusion

La recherche d'image par le contenu a été un domaine de recherche actif depuis plusieurs décennies et on lui a porté de plus en plus d'attention au cours des dernières années en raison de l'augmentation spectaculaire et rapide dans le volume d'images numériques. De nombreuses techniques nouvelles ont été proposées pour relever les défis. Toutefois, tous les systèmes actuels de recherche d'images par le contenu souffrent de la capacité de généralisation insuffisante et la précision car, ils ne sont pas en mesure d'établir un lien solide entre les caractéristiques de bas niveau de l'image et les concepts de haut niveau.

Dans ce chapitre, nous avons présenté un bref aperçu sur les travaux relatifs aux domaines passionnants de la recherche d'image par le contenu, et nous avons abordé également les différentes méthodes utilisées pour extraire les caractéristiques de bas niveau, et les diverses mesures de distance pour calculer la similarité entre les images

dans le but de réduire l'écart sémantique entre les caractéristiques de bas niveau et les concepts sémantiques de haut niveau. La recherche d'images basée sur le contenu fournit une solution intelligente et automatique pour une recherche efficace des images et la plupart des techniques actuelles sont basées sur des caractéristiques de bas niveau.

En général, chacune des caractéristiques de bas niveau a tendance à capturer un seul aspect d'une d'image. Ni une seule caractéristique, ni une combinaison de plusieurs caractéristiques a une signification sémantique explicite. En outre, les mesures de similarité entre les caractéristiques visuelles ne correspondent pas nécessairement à la perception humaine. Les utilisateurs sont intéressés à récupérer des images qui sont sémantiquement et perceptivement similaires, les résultats de recherche des approches basées sur les caractéristiques de bas niveau sont généralement insatisfaisantes et souvent imprédictibles. Bien que le retour de pertinence fournisse un moyen de combler le fossé entre la recherche sémantique et le traitement des données au niveau bas, ce problème reste soulevé et plus de recherches s'avèrent nécessaires.

Comme nous avons déjà mentionné, les techniques destinées pour usage général ne sont pas convenable pour utilisation dans des domaines spécifiques, aussi le choix de caractéristiques de bas niveau qui sont effectivement appropriées reste comme un défi de recherche persistant.

Dans le chapitre qui suit, nous allons détailler les différentes techniques de description du contenu proposées pour un domaine spécifique qui est très intéressant, à savoir la reconnaissance de visage.

CHAPITRE 3:

Reconnaissance automatique de visage

3.1 Introduction	043
3.2 Applications de la reconnaissance faciale	045
3.3 Difficultés rencontrées par les systèmes de reconnaissance automatique du visage	046
3.4 Techniques de description de visage	051
3.4.1 Les approches globales	051
3.4.1.1 L'analyse en composantes principales	051
3.4.1.2 L'analyse discriminante linéaire, ADL	054
3.4.1.3 L'analyse en composantes indépendantes ACI	056
3.4.1.4 Les méthodes de projection non linéaires	057
3.4.1.5 Les modèles actifs d'apparence	057
3.4.1.6 Les réseaux de neurones	059
3.4.2 Les approches locales	060
3.4.2.1 Les approches géométriques	060
3.4.2.2 Les techniques modulaires	061
3.4.2.3 Les Modèles de Markov Cachés (MMC)	063
3.4.2.4 Les approches basées sur les graphes	063
a) L'Elastic Graph Matching (EGM)	064
b) L'Elastic Bunch Graph Matching (EBGM)	065
3.4.3 Les approches hybrides	065
3.5 les approches basées sur la transformée en cosinus discrète	066
3.6 L'identification (la reconnaissance)	068
2.6.1 Classification par k-plus proches voisins	068
2.6.2 Classification par réseaux de neurones	068
2.6.3 Classification par SVM	069
3.7 Mesures d'évaluation	069
2.7.1 Vérification	069
2.7.2 Identification	071
3.8 Conclusion	071

CHAPITRE 3

Reconnaissance automatique de visage

L'une des applications les plus importantes de la recherche d'images par le contenu est la reconnaissance faciale qui est considérée parmi les principales techniques biométriques. Cette technologie qui ne nécessite aucun contact direct ou aucune coopération de l'utilisateur représente aujourd'hui un axe de recherche viable pour la communauté des chercheurs académiques et des entreprises privées dans la biométrie.

Dans ce chapitre, nous introduisons quelques notions de base liées à cette technologie. Nous donnerons le principe de fonctionnement d'un système de reconnaissance faciale, de même que les différentes difficultés rencontrées par cette technologie. Egalement, nous donnons une idée de l'état de l'art sur les différentes techniques développées dans ce domaine, ainsi que les outils d'évaluation pour mesurer leurs performances. Vu la diversité des techniques proposées, la liste des méthodes exposées dans ce chapitre n'est pas exhaustive.

3.1 Introduction

Comme nous avons déjà mentionné, la reconnaissance automatique du visage a pris une part de plus en plus considérable dans le domaine de la recherche. La reconnaissance faciale consiste à affecter une identité à une personne à partir d'une image de son visage. On peut considérer que la reconnaissance faciale est différente de la reconnaissance d'objet. En effet reconnaître un objet particulier, c'est classer cet objet dans sa classe d'appartenance, dans le cas de visage, cela correspond à la détection de visages. En revanche, dans le cadre de la reconnaissance faciale, nous connaissons l'objet qui est le visage, mais nous voulons établir une correspondance exacte entre le visage d'un individu présenté dans une image requête et les visages de sa classe qui sont enregistrés dans la base de données.

Généralement, un système de reconnaissance automatique de visage est constitué des trois phases principales qui sont illustrées dans la figure 3.1, à savoir, (1) la détection et la normalisation du visage, (2) l'extraction des caractéristiques du visage (ou le calcul de

descripteur) et (3) la classification ou la reconnaissance. Les méthodes employées dans chaque phase sont très critiques pour les applications biométriques, et doivent, ainsi, être améliorées pour obtenir des résultats plus performants.

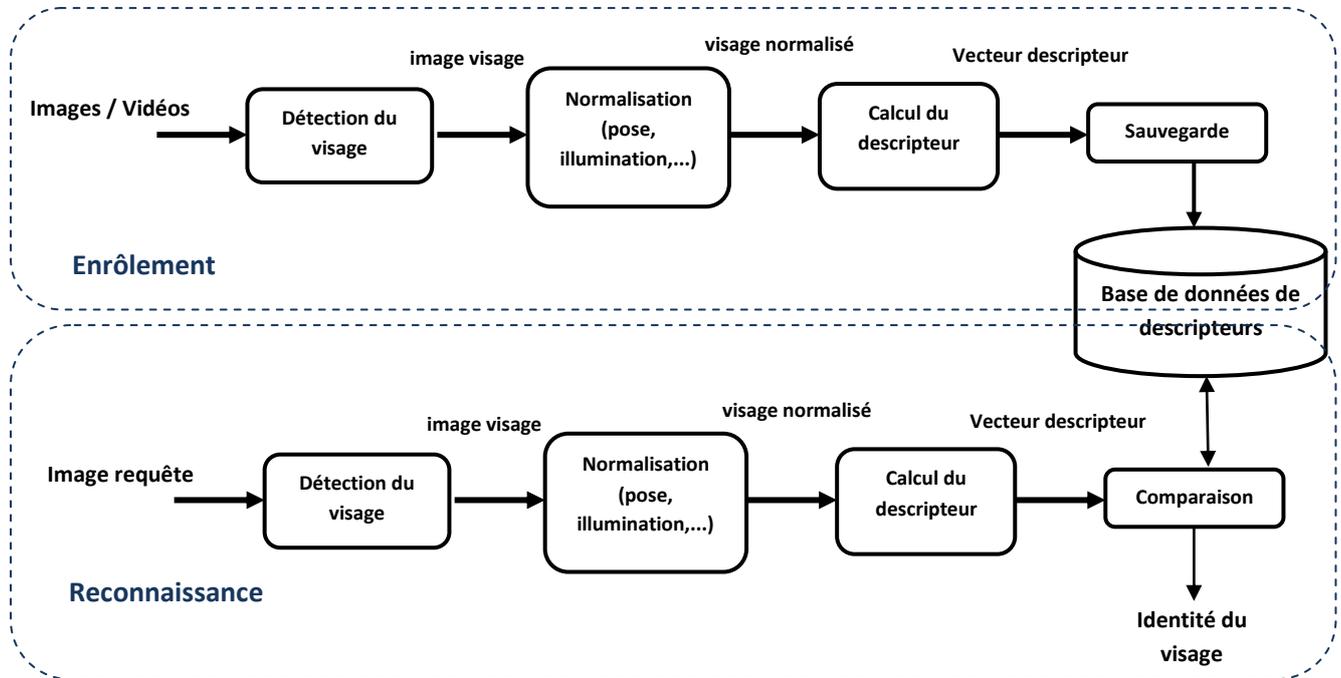


Figure 3.1- Les étapes principales d'un système de reconnaissance faciale.

a) La détection de visage :

La détection de visage se réfère à la localisation d'un visage dans une image. Plusieurs techniques ont été développées allant d'une simple localisation d'un visage complet dans une image à la localisation précise des différentes parties du visage telles que le nez, la bouche, les yeux, etc. Cependant, ces techniques proposées souffrent de la variabilité d'apparence des visages dans des conditions non contrôlées, notamment dues à :

- La variabilité des caractéristiques des visages humains (couleur, taille, forme)
- La présence ou l'absence de caractéristiques particulières (cheveux, moustache, barbe, lunettes...)
- Les expressions faciales qui peuvent modifier complètement la géométrie du visage
- L'occultation par d'autres objets ou d'autres visages
- L'orientation et la pose (de face, de profil)
- Les conditions d'illumination et la qualité de l'image

Cette étape n'est pas détaillée dans cette thèse, car elle est hors de la portée de la thèse, en effet, notre objectif principal est la reconnaissance de visage invariante à la progression d'âge, c'est-à-dire reconnaître l'identité d'une personne à partir de sa photo de visage et ceci quelque soit l'âge de l'image requête et celui de l'image de la base de données.

b) L'extraction de caractéristiques

Cette phase constitue l'étape la plus importante dans un système de reconnaissance faciale, elle consiste à représenter une image par un descripteur, qui doit être invariant pour des images d'une même personne et différent d'une personne à une autre. Cette étape sera détaillée en section 3.4.

c) La classification

Cette étape utilise les caractéristiques du visage ainsi extraites à l'étape précédente, elle dépend du mode de comparaison utilisé: vérification ou identification. Elle sera détaillée en section 3.6

3.2 Applications de la reconnaissance faciale

Dans un système de reconnaissance faciale typique, un certain nombre d'images de visage sont inscrites dans le système en tant que base de données et la mise en correspondance entre l'image requête et celles de la base d'images est effectuée suivant deux modes : un-à-un (one-to-one) ou un-à-plusieurs (one-to-many) [126]. Le premier mode correspond à la vérification, il vise à vérifier que l'identité prétendue est bien la bonne tandis que le dernier représente l'identification, il consiste à déterminer l'identité d'un individu parmi un ensemble d'entités connues, présentes dans la base de données. La reconnaissance automatique du visage est devenue de plus en plus très importante en raison de sa large variété d'applications. De nombreux domaines d'application dans lesquels la reconnaissance faciale peut être exploitée dans les deux modes d'identification ou de vérification, dont quelques uns sont cités ci dessous:

- L'identification ou la recherche des criminels connus qui ont été filmé par des caméras de vidéosurveillance.
- La recherche des personnes disparues ou de membre de famille dont on a juste une photo.

- La vérification générale d'identification personnelle (utilisation des cartes à puce), telle que les cartes d'identité nationales, les passeports, badge d'accès aux bâtiments d'une entreprise, etc.
- Le contrôle d'accès à des sites sensibles
- La sécurité des réseaux informatiques
- Les points de contrôle aux frontières, où elle permet d'alléger les procédures de vérification d'identité des voyageurs.
- Les applications commerciales où elle permet de sécuriser les transactions en ligne.

3.3 Difficultés rencontrées par les systèmes de reconnaissance automatique du visage

Bien qu'aujourd'hui les systèmes de reconnaissance faciale ont atteint des performances de plus en plus élevées dans des environnements contrôlés, leurs performances se dégradent de façon significative dans des environnements non contrôlés [127]. Cela dû principalement à la variabilité intra-classes des données, causée par des facteurs tels que les conditions d'éclairage du visage, son orientation, les expressions faciales, les occlusions du visage, par exemple, le port d'accessoires comme des lunettes, un cache-col, etc. Ces variations changent considérablement l'apparence de la personne et les caractéristiques utilisées pour la reconnaissance sont, par conséquent, influencées et les performances globales du système diminuent. Les auteurs dans [128] ont démontré que les différences entre les images de visage d'une même personne qui sont dû à des changements d'illumination et de l'angle de vision sont presque toujours plus importantes que la différence entre des images de visage de deux personnes différentes prises dans les mêmes conditions. Gross et al. [129] ont réalisé une étude sur l'effet des différents facteurs sur les performances de la reconnaissance faciale, les facteurs étudiés sont : la pose de la tête, le changement de l'illumination, l'expression faciale, les occultations et l'intervalle de temps entre deux prises de vues.

Les évaluations des travaux de littérature sur la reconnaissance faciale effectuées au fil des années telles que les évaluations FERET [25, 130], FRVT2000 [131], FRVT2002 [27] et FAT2004 [132] ont montré que les variations d'illumination, de position et d'âge constituent les trois problèmes majeurs des systèmes de reconnaissance faciale [133].

En se basant sur les travaux de la littérature, nous résumons les principales difficultés rencontrées par les systèmes de reconnaissance faciales dans les conditions réelles dans les points suivants:

- **Les variations de la pose**

Les variations de la pose sont parmi les principales sources de dégradation des performances de la reconnaissance faciale. Le visage est un objet 3D qui apparaît de différentes façons selon la direction dans laquelle l'image du visage est prise. Ainsi, il est possible que des images prises à deux points de vue différents de la même personne peuvent apparaître plus différentes que deux images prises à partir du même point de vue pour deux personnes différentes. Car un changement de l'angle d'inclinaison du visage implique des changements d'apparence entre les images. Comme il est montré dans [131], si la position de la tête du visage de la requête diffère significativement de celles enregistrées dans la base de données, les taux de reconnaissance baissent sensiblement. Car selon la FRVT2000 [131], la rotation de la tête jusqu'à plus ou moins 25° , ne cause pas de dégradation de taux de reconnaissance, mais, à partir de plus ou moins 40° , on constate une diminution des taux de reconnaissance. La figure 3.2 présente un exemple d'un visage subissant une rotation.



Figure 3.2- Exemple d'un visage d'une même personne subissant des variations de pose [134].

- **Les variations d'illumination**

Les variations de l'intensité et de la direction d'illumination entraînent des variations considérables dans l'apparence du visage dans l'image. Combinées avec la structure 3D du visage, ces changements peuvent faire apparaître sur les visages des ombres accentuant, ou même masquent certaines caractéristiques faciales. La figure 3.3 présente

un exemple de visage avec effets de variations dans les directions d'illumination sur l'apparence d'un visage.

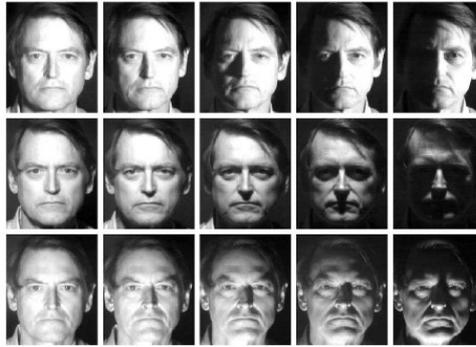


Figure 3.3- Effets des variations dans les directions d'illumination sur l'apparence d'un visage [128].

L'évaluation FRVT2000 [131] conclut que les variations d'illumination peuvent dégrader significativement les performances des systèmes de reconnaissance faciales. En effet, on note des baisses de taux de reconnaissance de l'ordre de 40% avec les meilleurs systèmes testés par la FRVT surtout dans des conditions d'illumination en extérieur. Il a également été révélé par la FRVT2006 que les variations d'éclairage sont parmi les problèmes majeurs que rencontrent les systèmes de reconnaissance faciale dans la pratique [135]. De nombreuses techniques ont été proposées pour gérer ces problèmes d'éclairage. Certaines approches se concentrent sur la modélisation implicite des variations d'illumination [136, 137], d'autres se concentrent sur l'extraction de caractéristiques invariantes aux changements de l'illumination [138, 139]. Notons enfin, qu'un certain nombre d'approches tentent de corriger les effets dus aux variations de l'illumination avant la reconnaissance [140, 141].

- Les expressions faciales

Les expressions faciales peuvent produire des changements d'apparence importants du visage (voir figure 3.4). Dans leurs études sur l'impact de changement d'expressions faciales sur la reconnaissance, Gross et al. [129] montrent que 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 (telles que le cri) et le rétrécissement ou la fermeture complète des yeux. Certaines approches [142] ont classé les émotions en sept expressions basiques : neutre, colère,

dégout, peur, joie, tristesse et surprise. D'autres considèrent que ce genre d'expression n'est que rarement observé dans des applications réelles. Et ils ont pris en considération d'autres expressions plus compliquées [143]. Il semble logique que les systèmes fondés sur les techniques géométriques ou les techniques basées sur les points d'intérêts sont plus sensibles aux changements d'expressions que les techniques globales.



Figure 3.4- Exemple d'images extraites à partir de la base de données d'expressions faciales JAFEE [144].

- Les occultations partielles

Les occultations partielles apparaissent fréquemment dans les applications réelles, elles peuvent être intentionnelles ou non, et elles peuvent être causées, par exemple, par le port de lunettes de vue ou de soleil, ou par tout autre objet (foulard, cache-nez), ainsi qu'une main ou un téléphone sur le visage, etc. (voir figure 3.5). Généralement, tout objet qui se situe entre la surface frontale du visage et la caméra peut provoquer le problème d'occultation. Gross et al. [129] ont étudié l'effet d'occulter la partie inférieure du visage en portant un cache-nez ou le port des lunettes de soleil sur les performances de différents algorithmes de reconnaissance, leurs résultats montrent que, dans ces conditions, les performances des algorithmes testés sont en général faibles. De plus, les différents algorithmes présentent un comportement différent vis-à-vis des occultations.



Figure 3.5- Exemples d'occultations partielles (Base de données AR) [145].

Plusieurs travaux ont révélé que la connaissance préalable des occlusions peut améliorer considérablement la précision des techniques de reconnaissance faciale basées sur les caractéristiques locales [146, 147].

- **Les variations dues à la progression d'âge**

Un autre facteur très important est l'âge de la prise de l'image, c'est à dire le moment auquel l'image est capturée. En effet, l'apparence d'un visage peut changer au cours du temps (surtout lors de l'adolescence) et un écart de temps important entre deux images peut engendrer des difficultés de reconnaissance.



Figure 3.6- Exemple d'images d'un individu, les images sont de la base de données de visages ORL.

Dans le cadre de l'évaluation FRVT2002 [27], les taux de reconnaissance des meilleurs algorithmes testés sont diminués de 5% par année d'écart entre l'image de référence et celle à reconnaître.

Egalement, dans le même contexte de l'effet de la progression d'âge sur les performances de la reconnaissance faciale, si nous prenons l'exemple de l'approche Eigenfaces qui, depuis sa proposition par Turk et Pentland [154] en 1991 pour la reconnaissance faciale, a été largement étudiée et en devenant l'une des approches les plus réussies et les plus populaires dans le domaine de la reconnaissance faciale. Cette méthode, combinée avec la distance euclidienne, a fourni un taux de reconnaissance de 97.5% [213] sur la base d'images ORL (The ORL Database of Faces), malgré que la base de visages inclut des variations d'illumination, et d'expression (voir figure 3.6). Cette même approche, combinée avec la même distance a fourni un taux de 17.39% [214] sur la base d'images FG-NET. Ce qui met en évidence l'effet de la progression d'âge sur les performances de la reconnaissance faciale.

Comme l'objectif principal de cette thèse est la reconnaissance faciale invariante à la progression d'âge, ce point sera présenté avec davantage de détails dans le chapitre 4.

3.4 Techniques de description de visage

De nombreuses techniques ont été proposées pour décrire une image. Certaines méthodes se basent sur la représentation 2D du visage, par contre d'autres utilisent l'information 3D de celui-ci. De même, il existe des méthodes appelées 2.5D qui effectuent la reconnaissance en se basant sur l'information de profondeur de l'image. Dans cette section, nous nous concentrons sur les techniques en 2D, nous résumons les différentes techniques les plus connues de la reconnaissance de visage, qui sont détaillées davantage dans [149]. Ces techniques de description en deux dimensions peuvent être classées en trois classes : Les méthodes holistiques ou globales, les méthodes locales, ainsi que les méthodes hybrides. Plusieurs auteurs ont réalisé des états de l'art et ont évalué les techniques de reconnaissances de visage sous différents aspects [24, 150, 151], une excellente synthèse sur ces méthodes peut être trouvée dans [152]. Un résumé des principales méthodes déjà existantes est brièvement présenté ci-dessous:

3.4.1 Les approches globales

Ces techniques utilisent le visage complet comme entrée pour le système de reconnaissance. En effet, les descripteurs sont directement calculés à partir des valeurs de niveau de gris de la totalité des pixels de l'image du visage. Cela permet de mieux représenter l'apparence globale du visage que les techniques basées sur les caractéristiques locales [153]. Parmi les techniques globales les plus connues de cette catégorie, nous pouvons citer la méthode Eigenfaces [154, 156], Fisherfaces [155] ou encore l'analyse en composantes indépendante (ACI) [157]. Ces techniques peuvent être résumées comme suit [149]:

3.4.1.1 L'analyse en composantes principales (ou Principal Component Analysis PCA en anglais)

L'analyse en composantes principales ACP [158] est une technique de réduction de la dimensionnalité d'un espace qui cherche un sous espace qui est le plus représentatif des données dans l'espace initial, elle fait ressortir l'information nécessaire et élimine celle qui est considérée comme du bruit. Sous l'hypothèse de linéarité et d'orthogonalité, la méthode ACP consiste à trouver une nouvelle base de vecteurs qui est la combinaison linéaire des vecteurs de la base initiale dont les vecteurs sont tous orthogonaux entre eux. Le sous espace construit par l'ACP est appelé sous espace principal. La première

composante principale correspond à la direction de variance maximale des données initiales. La seconde qui est orthogonale à la première correspond à la deuxième plus grande variance des données initiales et la n ème composante principale est la combinaison linéaire avec la n ème plus grande variance, elle doit être orthogonale aux $n-1$ premières composantes principales. L'idée de l'ACP est représentée dans la figure 3.7 :

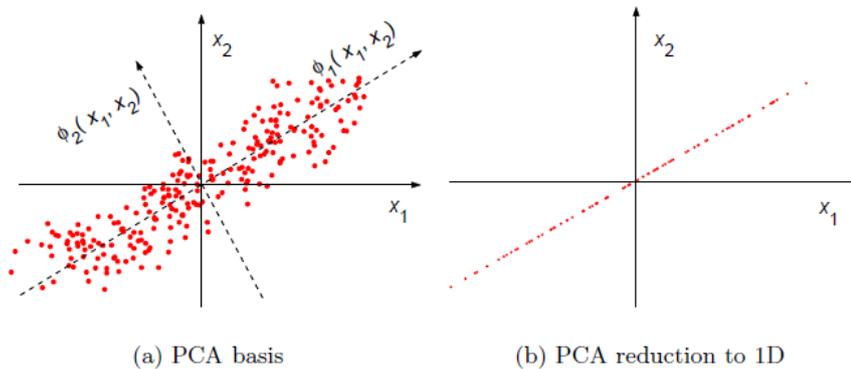


Figure 3.7- Le concept de l'ACP. (a) x_1, x_2 : la base originale. ϕ_1, ϕ_2 : la nouvelle base. (b) la projection des données en utilisant seulement la première composante principale. [159]

L'axe ϕ_1 correspond à la direction de variance maximale, il est choisi comme première composante principale. Dans le cas de données à deux dimensions, la deuxième composante principale est alors déterminée par contrainte d'orthogonalité. Dans un espace de dimension supérieure, le processus de sélection se poursuivra guidé par les variances des projections.

La méthode Eigenfaces

Depuis les travaux de M. A. Turk et A. P. Petland [154], L'ACP a été utilisée pour la reconnaissance faciale sous le nom de Eigenfaces (visage propres), c'est une méthode très utilisée dans le domaine de la reconnaissance faciale, et cela est due à sa simplicité, ses fondements mathématiques forts et ses résultats bons. De plus, plusieurs extensions de l'ACP ont aussi été proposées et utilisées pour la reconnaissance faciale. Son principe est le suivant:

Etant donnée une base de données contenant M images de même taille, les images sont normalisées en représentant chaque image i de taille $n * m$ par un vecteur Γ_i de dimension $N = n * m$, cela revient à représenter un visage comme un point dans un

espace de dimension N . La matrice Γ contenant la base d'apprentissage est l'ensemble des vecteurs images, tels que:

$$\Gamma = [\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_M]$$

La méthode EigenFaces est constituée de deux phases, une d'apprentissage et une de classification.

* Phase d'apprentissage

1. Calculer l'image moyenne Ψ de la base d'apprentissage, tel que:

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (3.1)$$

2. Calculer ϕ_i qui correspond à l'image i dont on a soustrait l'image moyenne tel que :

$$\phi_i = \Gamma_i - \Psi, \quad i=1,2,\dots, M \quad (3.2)$$

3. Si on considère la matrice des vecteurs de différences $A = [\phi_1, \phi_2, \dots, \phi_M]$ qui ne conserve que les caractéristiques distinctives des images de visage et élimine les caractéristiques communes. Les visages propres sont obtenus en calculant la matrice de covariances C des vecteurs visages de la base d'apprentissage, où C est donnée par:

$$C = A.A^T \quad (3.3)$$

4. Dû à la grande dimension de C , on considère la matrice L de taille $M_t \times M_t$ inférieure à C qui donne le même résultat avec une dimension réduite. Les vecteurs propres de la matrice C (matrice U) peuvent être obtenus en calculant les vecteurs propres de la matrice L (matrice V), tel que:

$$U_i = AV_i \quad (3.4)$$

Les valeurs propres sont : $[U_1, U_2, \dots, U_M]$. La figure 3.8 montre l'image moyenne, suivie des sept premières eigenfaces associés aux plus grandes valeurs propres.

5. Au lieu d'utiliser M eigenfaces, on réorganise les valeurs propres dans l'ordre croissant pour n'en conserver que les m' ($m' < M$) plus grandes valeurs. les eigenfaces associées aux m' valeurs sont choisies comme espace propre.

Ensuite, le poids de chaque vecteur propre ω_i représentant une image i dans l'espace propre est donné par:

$$\omega_i = U_i^T (\Gamma - \Psi), i = 1, 2, \dots, m' \quad (3.5)$$

La matrice des poids : $\Omega = [\omega_1, \omega_2, \dots, \omega_{m'}]^T$.

La projection de la classe moyenne:

$$\Omega_{\Psi} = \frac{1}{X_i} \sum_{i=1}^{X_i} \Omega_i \quad (3.6)$$



Figure 3.8- L'image moyenne à gauche, suivie par les sept premières Eigenfaces [154].

* Phase de classification

La classification ou l'identification consiste à trouver le visage de la base de données d'apprentissage qui minimise la distance δ_i entre le vecteur de projection Ω et celui de l'image test Ω_t . Où:

$$\delta_i = \|\Omega - \Omega_t\| \quad (3.7)$$

3.4.1.2 L'analyse discriminante linéaire, ADL (ou Linear Discriminant Analysis LDA, en anglais)

L'idée principale de cette technique est de trouver les directions de projection les plus discriminantes dans l'espace de visage. Elle a été proposée par Belhumeur et al. [155] sous le nom de FisherFaces. Contrairement à l'ACP qui permet de maximiser les variations inter-classe, l'ADL permet de maximiser les variations inter-classe et minimiser les variations intra-classe, car elle permet d'effectuer une vraie séparation de classe [156]. Un exemple graphique de réduction de dimension de données en 2D en données monodimensionnelles est illustré dans la figure 3.9, où l'axe principal de l'ACP (axe en bleu) ne permet pas de séparer en projection les deux classes, tandis que l'axe ADL (axe en rouge) sépare au mieux les données des deux classes après projection [149]. L'approche FisherFaces est une méthode d'apprentissage supervisé, cela nécessite que chaque personne de la base d'images doit avoir au moins deux images, et toute

image doit être manuellement associée à une classe. Pareillement, comme dans l'approche ACP, l'algorithme FisherFaces consiste à trouver un espace de projection approprié. L'identification est effectuée en comparant la projection de l'image test avec chacune des projections des images de la base d'apprentissage. Comme dans l'approche ACP, chaque image i de taille $n * m$ de la base d'apprentissage est représentée par un vecteur Γ_i de dimension $N = n * m$

L'image moyenne de chaque classe c_i est définie par:

$$\Psi_{c_i} = \frac{1}{N_c} \sum_{i=1}^{N_c} \Gamma_i \quad (3.8)$$

Où N_c est le nombre d'images de la classe c_i .

L'image moyenne de toutes les classes c_i est définie par:

$$\Psi = \frac{1}{c} \sum_{i=1}^c N_c \Psi_{c_i} \quad (3.9)$$

ϕ_i correspond à l'image i dont on a soustrait l'image moyenne de la classe c_i tel que :

$$\phi_i = \Gamma_i - \Psi_{c_i},$$

Les variations inter-classe S_b et intra-classe S_w sont définies comme suit:

$$S_b = \sum_{i=1}^{N_c} N_c (\Psi_{c_i} - \Psi)(\Psi_{c_i} - \Psi)^T \quad (3.10)$$

$$S_w = \sum_{i=1}^c \sum_{\Gamma_k \in c} N_c (\Gamma_k - \Psi_{c_i})(\Gamma_k - \Psi_{c_i})^T \quad (3.11)$$

Le sous espace est engendré par la matrice W qui maximise le critère de Fisher [33]:

$$R = \frac{|W^T S_b W|}{|W^T S_w W|} \quad (3.12)$$

Ce qui est équivalent à chercher les vecteurs propres de la matrice supposée symétrique $S_w^{-1} S_b$ satisfaisant l'équation aux valeurs propres suivante:

$$S_w^{-1} S_b \lambda = w \cdot \lambda \quad (3.13)$$

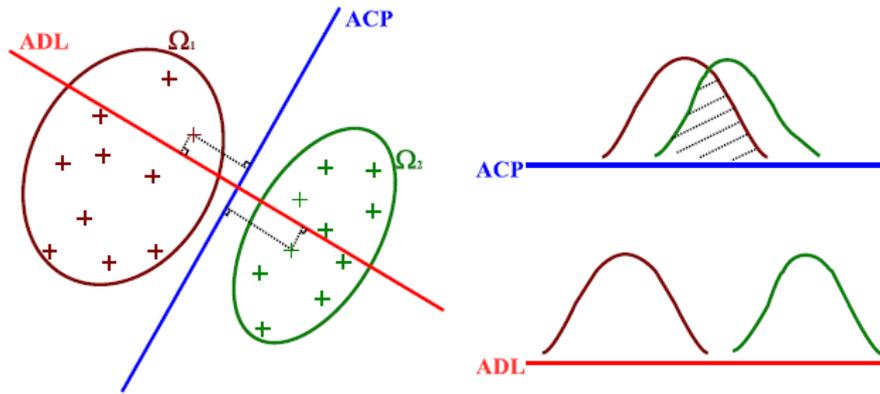


Figure 3.9- Un exemple de projection de deux classes sur un axe ACP et sur un axe ADL [149].

3.4.1.3 L'analyse en composantes indépendantes ACI (ou Independent Component Analysis : ICA en anglais)

C'est une technique initialement formulée par [160] pour la séparation de sources, mais aujourd'hui elle est appliquée à divers problèmes. Dans un contexte de reconnaissance de visage, elle peut être considérée comme une généralisation de la méthode ACP, car elle permet non seulement de dé-corréler les données, mais, également de les rendre statistiquement indépendantes [149]. Il existe de nombreux algorithmes d'implémentation de l'ACI, dont les plus connus sont: INFOMAX [161], JADE [162] et FastICA [163]. Dans le même contexte de la reconnaissance faciale, Bartlett et al. [157] ont proposé une approche dont le principe est le suivant:

Soit S un vecteur de sources inconnues, et X une matrice d'observations (les images du visage), le modèle de mélange est décrit comme suit:

$$X^t = AS^t \quad (3.14)$$

Où A est la matrice de mélange. Si on suppose que les sources sont indépendantes et la matrice A est inversible, on peut également définir une matrice de séparation W qui permet à partir de X d'estimer S tel que:

$$\mu^t = WX^t = WAS^t \quad (3.15)$$

Où μ est l'estimation de la source S .

Le but de l'ACI est d'estimer la matrice A (ou W) et les sources S en réduisant au minimum la dépendance de ses composantes. Cela revient à trouver une fonction pour

mesurer cette dépendance et de la minimiser. Dans [157], en se basant sur l'algorithme INFOMAX, les auteurs ont proposé deux architectures différentes pour exprimer la matrice des observations X , où la première architecture considère que les images sont des variables aléatoires et la seconde considère les pixels comme variables aléatoires. Comme illustré sur la figure 3.10, il est intéressant de noter que la première architecture permet d'extraire des structures de base telles que les yeux, le nez ou la bouche, caractéristiques principales d'un visage, alors que la deuxième architecture fournit plus d'information sur la structure globale du visage.

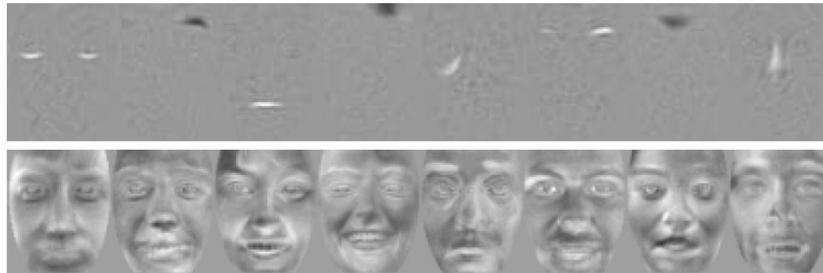


Figure 3.10- Images de base obtenues avec la méthode ACI pour l'architecture 1 en haut et pour l'architecture 2 en bas [164].

3.4.1.4 Les méthodes de projection non linéaires

Les techniques précédemment présentées sont basées sur l'hypothèse que les visages sont linéairement séparables dans l'espace de représentation. Ce n'est pas toujours vrai, surtout, dans la présence des changements drastiques d'illumination [149], c'est pourquoi de nombreux auteurs, ont proposé des versions non linéaires pour les méthodes citées ci-dessus, particulièrement l'ACP, et l'ADL, en utilisant des fonctions noyaux (kernel). L'idée principale de ces méthodes est d'exprimer les données dans un nouvel espace non linéaire et d'appliquer ensuite la méthode dans le nouvel espace. De bons résultats ont été achevés avec l'algorithme ACP à noyau [165] et l'algorithme LDA à noyau [166].

3.4.1.5 Les modèles actifs d'apparence

Les Modèles Actifs d'Apparence (MAA) [167] sont des modèles déformables permettant de faire simultanément l'analyse et la synthèse d'images. Ils constituent un outil performant pour l'extraction de caractéristiques de forme et de texture des visages.

Un modèle actif d'apparence est constitué d'une forme (un ensemble de points d'intérêt reliés entre eux) et d'une texture (ensemble des pixels à l'intérieur d'une forme). Les modèles sont créés à partir d'une base d'apprentissage d'images sur lesquelles des points d'intérêts sont manuellement déterminés et avec précision, et les variations entre les points constituant la forme et le contenu (la texture) sont apprises à partir d'une ACP. A partir d'un ensemble de vecteurs de forme alignés à la forme moyenne, une ACP est appliquée pour obtenir le modèle suivant:

$$x = \bar{x} + P_f b_f \quad (3.16)$$

Où \bar{x} est la forme moyenne de la base d'apprentissage, P_f est une matrice orthonormée contenant les principaux modes de variation de la forme et b_f est la projection de x sur P_f , c'est un vecteur contrôlant la déformation de la forme synthétisée.

Pour construire un modèle statistique de texture, les textures des visages sont déformées pour que leur forme correspond à la forme moyenne, on obtient ainsi une représentation des visages dites Shape-free, les processus d'alignement et de déformation de texture sont détaillés davantage dans le chapitre 5. Pour chaque visage, on obtient un vecteur de texture g , qui est indépendant de la forme. En appliquant une ACP sur ces vecteurs, le modèle statistique de texture est donné par:

$$g = \bar{g} + P_g b_g \quad (3.17)$$

Où \bar{g} est la texture moyenne, b_g est le vecteur de paramètres de texture associé au visage, et P_g est une matrice contenant les modes principaux de variation de la texture.

La forme et la texture associées à chaque visage peuvent être représentées par ses vecteurs de forme et de texture b_f et b_g qui sont corrélés. Pour modéliser cette corrélation, on concatène les vecteurs obtenus, on obtient alors le modèle combiné suivant:

$$x = \bar{x} + Q_f c \quad (3.18)$$

$$g = \bar{g} + Q_g c \quad (3.19)$$

Où c est le vecteur d'apparence contrôlant à la fois la forme et la texture du modèle, et Q_f et Q_g sont respectivement les matrices de projection de c dans les espaces de variations de forme et de texture.

c est considéré comme descripteur de l'image du visage, et la classification peut se faire par le calcul d'une distance entre les vecteurs c [169]. Le processus de classification est illustré dans la figure 3.11

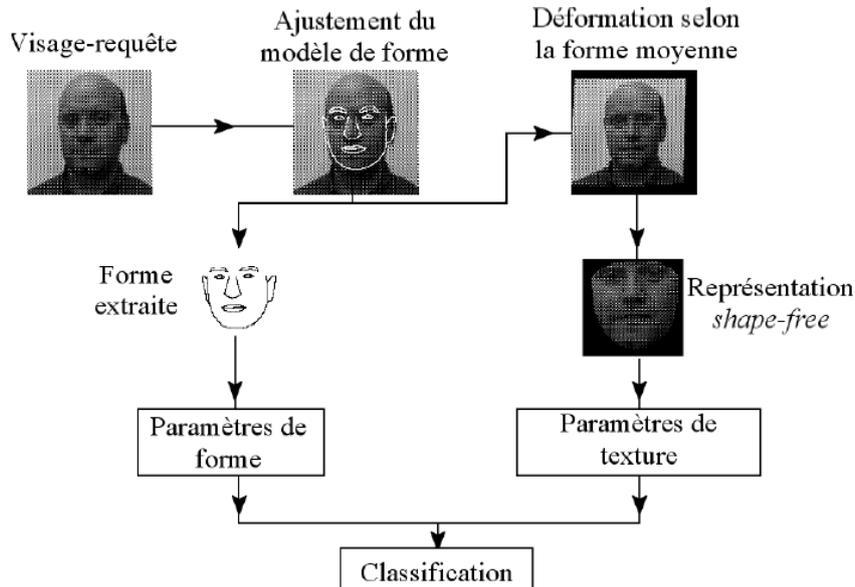


Figure 3.11- Processus de reconnaissance de visage basé sur les modèles actifs d'apparence [168].

3.4.1.6 Les réseaux de neurones

Les réseaux de neurones ont été employés dans toutes les phases constitutives d'un système de reconnaissance faciale. Ils sont utilisés pour la détection du visage [170, 171], pour l'extraction de caractéristiques [172, 173, 174] et pour la classification de celles-ci [175, 176]. Ils sont considérés parmi les techniques non linéaires pour l'extraction de caractéristiques. On note en particulier, les cartes auto-organisatrices ou cartes de Kohonen introduites par Kohonen [177] qui sont utilisées pour l'extraction de caractéristiques [172]. Celles-ci sont constituées d'une grille, dans chaque nœud de la grille se trouve un neurone et chaque neurone est lié à un vecteur référent, responsable d'une zone de l'espace d'entrée. Les vecteurs référents sont positionnés de telle sorte qu'ils conservent la forme topologique des classes, c.-à-d. la similarité entre les entrées est conservée en sortie.

3.4.2 Les approches locales

Ces méthodes permettent d'extraire des caractéristiques locales du visage, elles peuvent être classées en deux classes, celles qui extraient les descripteurs des traits du visage (local feature-based method) tels que le nez, la bouche, et les yeux, aussi elles utilisent les relations pouvant exister entre ces traits (distances, angles,...). Et celles qui extraient les descripteurs à partir de régions prédéfinies du visage, ces régions peuvent être obtenues en divisant l'image en parties de dimensions égales et de formes similaires. Dans ce qui suit, nous présenterons quelques techniques locales utilisées pour la description de visage :

3.4.2.1 Les approches géométriques

Elles sont considérées parmi les techniques les plus anciennes utilisées pour la reconnaissance faciale. Elles sont basées généralement sur les mesures et les positions relatives des éléments distinctifs du visage telles que la position et la largeur du nez et de la bouche, l'épaisseur des sourcils, etc. Dans [178], les auteurs proposent d'extraire 35 mesures du visage qui représentent, par exemple, la position verticale du centre de l'œil, l'épaisseur des sourcils, la position verticale et la largeur du nez et de la bouche ainsi que les segments de droite caractérisant la forme du menton, voir figure 3.12.

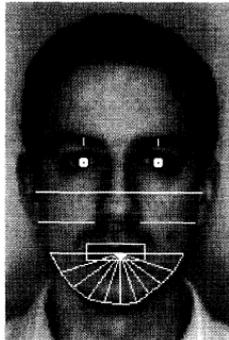


Figure 3.12- Caractéristiques géométriques utilisées dans [178].

Toutefois, ces méthodes nécessitent une très grande précision dans la détection des éléments faciaux. D'autres méthodes de détection ont été proposées, telles que le détecteur de Viola et Jones [179]. De plus ces méthodes restent très sensibles à des changements d'expression faciale ou de pose de la tête. Dans leur comparaison entre les méthodes géométriques et les méthodes modulaires, Brunelli et Poggio [178] ont montré clairement ces limitations. Pareillement, les auteurs dans [180] ont montré que

la méthode géométrique en utilisant les mesures présentées dans la figure 3.13 combinée avec un réseau de neurones donne un taux de reconnaissance de 84.2% alors qu'une combinaison d'une transformation en ondelette de Gabor avec la même architecture de réseau de neurones atteint un taux de reconnaissance de 99.4%.

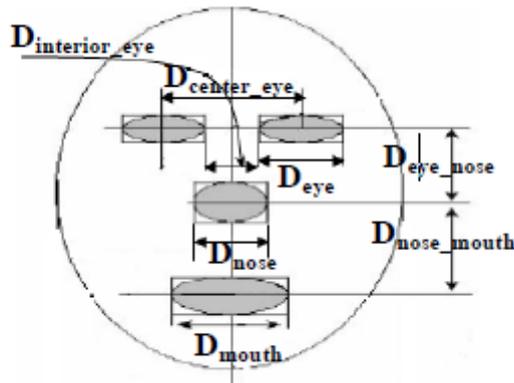


Figure 3.13- Mesures géométriques utilisées dans [180].

3.4.2.2 Les techniques modulaires

Ces techniques se basent sur les méthodes globales telles que la méthode ACP et ADL, appliquées de manière modulaire à différentes régions du visage. Par exemple, dans [181], Pentland et al. ont proposé la technique des "Modular Eigenspaces" (ou espace propres modulaires) où une ACP est appliquée sur la totalité du visage, sur les yeux et sur le nez, et les résultats de classification obtenus de chacune des régions sont agrégés. Les auteurs ont montré que cette approche est plus efficace que les techniques globales ou strictement locales c.-à-d. la même technique sans utiliser la totalité du visage. Egalement, Price et Gee [182] ont proposé une technique modulaire basée sur une variante de l'ADL où les régions considérées sont: la totalité du visage, la partie couvrant le front jusqu'au dessous du nez, et la partie contenant les yeux. Leurs résultats montrent que cette technique est plus stable par rapport aux techniques Eigenfaces et Fisherfaces, en présence des changements d'illumination, d'expressions faciales et d'occultations partielles.

Martinez et al. [183] ont présenté une technique probabiliste pour la reconnaissance de visages en présence des occultations partielles et des variations d'expression. Dans leur approche, chaque visage est divisé en six régions locales sous forme d'ellipses (voir figure 3.14). Les régions ayant les mêmes positions pour chaque visage sont classées

séparément dans un sous-espace (6 espaces au total) et chaque sous-espace de visage est à nouveau transformé dans un *eigenspace*, où la distribution est évaluée par un modèle de mélange de Gaussiennes en utilisant l'algorithme EM. Dans l'étape d'identification, les images de test sont aussi divisées en six régions locales et sont projetées respectivement dans l'*eigenspace* calculé et la similarité est mesurée par une approche probabiliste.

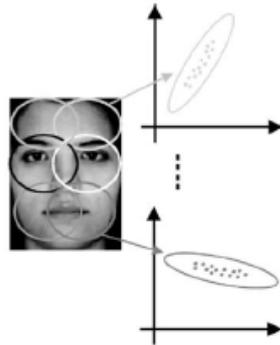


Figure 3.14- Exemple d'un visage divisé en six régions locales sous forme d'ellipse [183].

Tan et al. [184] ont développé cette approche en représentant le sous espace de visage par des cartes SOM (Self organizing maps) [185] où chaque image I de visage est divisée en M sous blocs locaux R_i ($i = 1..M$), puis un réseau SOM est entraîné sur les sous blocs obtenus de toutes les images d'apprentissage disponibles quelque soit leur classes. Ensuite, chaque sous bloc R_i de la même image I du visage est mappé avec les meilleures unités qui lui correspondent (Best Matching Units) en suivant une stratégie du plus proche voisin. La position de chaque bloc dans l'espace topologique SOM est un vecteur de position $l_i = \{x_i, y_i\}$. Tous les vecteurs de position du même visage peuvent être groupés en un ensemble $I = \{l_i, i = 1..M\}$. Cet ensemble est appelé: la représentation "SOM-face" du visage (voir figure 3.15 (b)).

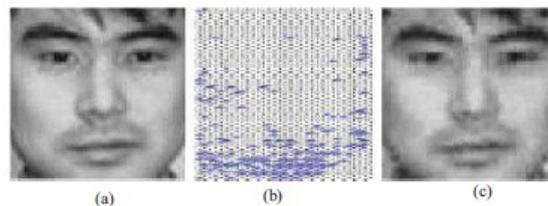


Figure 3.15- Exemple : (a) une image du visage originale, (b) sa projection (SOM-Face) et (c) image reconstruite [184].

Les éléments d'une SOM-face encodent les différentes parties locales du visage dans l'espace d'entrée, ils sont donnés par des vecteurs de poids stockés dans les nœuds représentants de la carte SOM.

3.4.2.3 Les Modèles de Markov Cachés (MMC)

Les modèles de Markov Cachés ont été utilisés pour la reconnaissance de visage par Samaria [186]. Dans leur approche, un modèle MMC est construit en divisant l'image d'un visage en cinq régions qui sont considérées comme des états cachés, ces régions couvrent toute la largeur du visage et se chevauchent, elles incluent le front, les yeux, le nez, la bouche et le menton (voir figure 3.16), et les observations émises par chacun de ces états sont les vecteurs obtenus par concaténation des lignes de pixels de ces états. Un modèle de visage est alors considéré comme une séquence d'observations constituée de cinq états, dont chacun peut être modélisé par une distribution gaussienne multi-variables, et les transitions probabilistes entre les états peuvent être apprises à partir des frontières entre les régions. Après une étape d'apprentissage, un visage est reconnu par le calcul de la probabilité de sortie de son ordre d'observation.

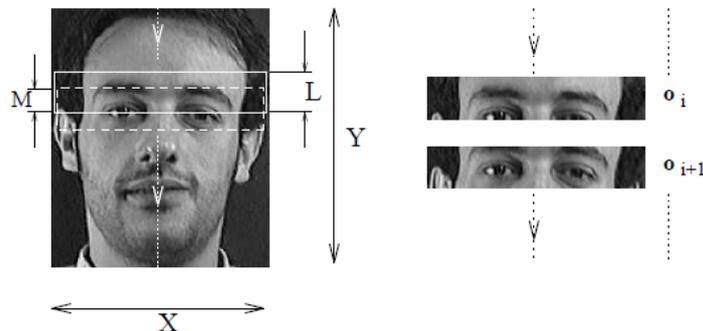


Figure 3.16- Technique de division du visage [186].

3.4.2.4 Les approches basées sur les graphes

Certains auteurs ont choisi de représenter les caractéristiques du visage par déformation de graphe [187, 188]. Ces techniques permettent de représenter une image d'un visage à l'aide d'un graphe formé par un ensemble de nœuds reliés entre eux par des arêtes. Dans [187], les éléments locaux d'un visage sont localement détectés par des ondelettes de Gabor. Les points ainsi détectés constituent les nœuds du graphe qui sont caractérisés par leur position dans l'image et par un vecteur de caractéristiques calculé au voisinage

de ce point. La topologie du graphe est obtenue en calculant la distance euclidienne $d_{i,j}$ entre deux nœuds voisins i et j . Et la reconnaissance s'effectue en comparant le graphe de l'image requête avec ceux contenus dans la galerie. Le visage associé au graphe qui minimise une fonction de coût correspond au visage à identifier. Cette approche est sensible aux variations qui peuvent modifier l'apparence du visage telle que la position.

a) L' Elastic Graph Matching (EGM)

Pour faire face aux problèmes rencontrés dans l'approche [187], Une méthode basée sur une Architecture de Liens Dynamiques (Dynamic Link Architecture ou DLA) a été proposée dans [189] où l'image cette fois-ci est représentée par un graphe avec une topologie de type grille déformable comme le montre la figure 3.17.

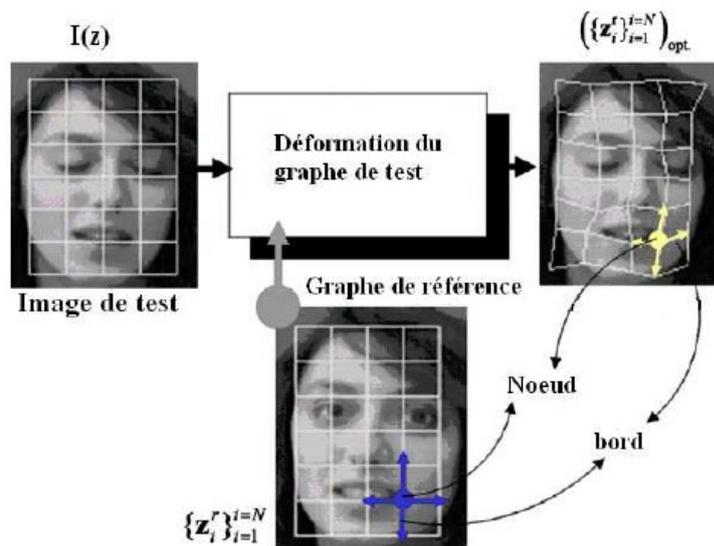


Figure 3.17- Le graphe Elastic Graph Matching sous forme d'une grille rectangulaire sur une image de visage comparée à une image de référence [190].

La technique est connue sous le nom de "Elastic Graph Matching" (EGM), Pour cela une grille régulière déformable est placée sur les images de visages. Les nœuds de la grille contiennent des ensembles de coefficients obtenus par une transformation de Gabor, et ils sont liés d'une manière élastique, de telle sorte qu'elle permet au système de reconnaissance de s'adapter aux déformations des parties du visage et des changements de point de vue.

b) L' Elastic Bunch Graph Matching (EBGM)

La technique d'Elastic Bunch Graph Matching (EBGM) [191] imite celle de l'Elastic Graph Matching, en se distinguant par l'utilisation d'un seul graphe pour représenter les différentes variations d'apparence du visage. Son principe est le suivant: A partir d'une image d'un visage, on détermine les points caractéristiques tels que les coins des yeux, de la bouche et du nez. Cette détection peut se faire manuellement ou automatiquement [193]. Ensuite, un graphe avec une topologie de type grille déformable est construit en reliant ces points, la représentation obtenue est appelée: "Face Bunch Graph" (voir figure 3.18). Chaque nœud contient un ensemble de coefficients complexes d'Ondelette de Gabor, appelés *jets*. Pour comparer deux images, l'algorithme calcule la similarité entre les différents Jets et les longueurs des arêtes des deux images. Zhang et al. [192] ont souligné deux inconvénients de la méthode EBGM, car elle nécessite un temps de calcul important ce qui la rend plus difficile à implémenter dans la pratique, et seules les informations sur les positions caractéristiques de l'image telles que les yeux, le nez et la bouche sont utilisées pour l'identification. Bien que ce soit un attribut important qui contribue à la robustesse de la méthode, les performances de la méthode diminuent lors des situations où les parties caractéristiques sont occultées.

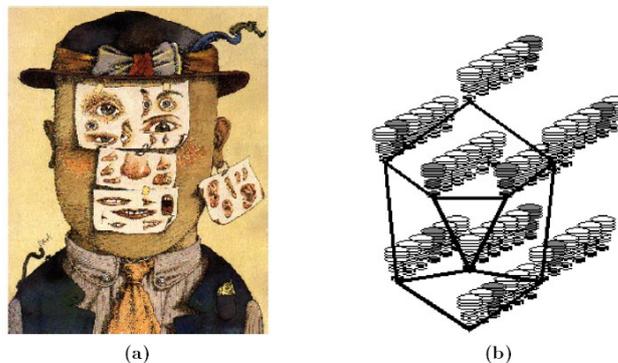


Figure 3.18- Le Face Bunch Graph (FBG) représenté d'un point de vue (a) artistique et (b) scientifique [149].

3.4. 3 Les approches hybrides

Les méthodes hybrides sont des approches qui combinent les caractéristiques holistiques et locales afin d'améliorer les performances de la reconnaissance de visages.

En effet, les caractéristiques locales et les caractéristiques globales ont des propriétés tout à fait différentes.

3.5 Les approches basées sur la transformée en cosinus discrète

La transformée en cosinus discrète DCT est une transformée mathématique puissante utilisée pour de nombreuses applications de traitement d'image telles que le codage d'images; le principe de cette transformée sera détaillé dans le chapitre 5.

Depuis longtemps, la transformée en cosinus discrète a été utilisée comme outil d'extraction de caractéristiques dans plusieurs études de reconnaissance de visage :

Par exemple, dans [194], une DCT par blocs est appliquée pour extraire les caractéristiques, où le sous ensemble des coefficients obtenus de la DCT est quantifié pour former le vecteur descripteur, et la classification est réalisée par une approche des K-plus proches voisins. Egalement, dans [195], la DCT a été combinée par un modèle de Markov Caché (Hidden Markov Model HMM), dans cette étude, la DCT a été appliquée sur des blocs d'image du visage chevauchés, et les coefficients obtenus sont disposés d'une manière de balayage en zigzag et seulement les premiers coefficients sont choisis comme descripteur de chaque bloc. Les blocs de l'image sont traités séquentiellement de la partie supérieure gauche de l'image du visage jusqu'à la partie en bas à droite et la séquence d'observations extraite est introduite dans un HMM unidimensionnel. Aussi, dans [196], les coefficients de la DCT appliquée sur l'image entière ou sur ses blocs sont extraits, et seulement un sous ensemble d'entre eux est sélectionné par un balayage diagonal de la partie supérieure gauche. Ces coefficients sont ensuite, introduits comme entrée à un perceptron multicouche.

Dans le même contexte, Dans [197], les coefficients de la DCT par bloc sont utilisés comme vecteurs d'observations d'un modèle intégré de Markov caché. Dans cette approche, chaque région du visage (le front, les yeux, le nez, la bouche et le menton) est modélisé par un (super state) et chaque (super state) contient des états intégrés (embedded states). Une autre méthode qui a utilisé les coefficients de la DCT comme vecteur d'observations d'un modèle de Markov caché est décrite dans [198].

Egalement, dans [199], La DCT est effectuée sur l'image entière et un sous ensemble de coefficients carré de la partie supérieure gauche est utilisée comme vecteur descripteur,

la classification est effectuée par la méthode des K plus proches voisins en utilisant la mesure L2 comme distance.

De même dans [200], les trois approches (fisherfaces [155], bayesian face recognition [201] et la DCT) sont appliquées sur l'image entière, puis, sur des parties de l'image du visage en utilisant trois schémas de partitionnement, le premier divise l'image du visage horizontalement en quatre parties égales, le deuxième et le troisième segmentent l'image verticalement et horizontalement; le deuxième donne quatre parties égales et le dernier douze parties égales. Les sorties de classification de chaque schéma sont fusionnées avec celles de classification de l'image entière par une règle de somme [202]. Pour la DCT, le sous ensemble carré de coefficients en haut à gauche est utilisé pour la classification.

En outre, dans [203], l'ACP et l'ADL sont effectuées dans le domaine de la DCT, la DCT est appliquée par blocs et les coefficients de la DCT sont quantifiés et disposés suivant un balayage en zigzag, et seul un nombre de coefficients est utilisé. Les coefficients obtenus de chaque bloc sont concaténés. PCA et ADL sont appliquées sur le vecteur obtenu.

Pareillement, dans [204], La DCT est appliquée sur l'image entière du visage et les coefficients des basses fréquences sont éliminés afin de fournir une robustesse du descripteur en présence des variations d'illumination.

Egalement, dans [205], la DCT est appliquée sur des images logarithmiques afin de normaliser les images en présence des variations d'illumination. La DCT est effectuée sur l'image entière, les valeurs des coefficients sont normalisées en fonction de la valeur du coefficient DC. Et les coefficients de basses fréquences, qui sont sensibles aux variations d'illumination, sont éliminés du descripteur.

Dans la même perspective, dans [206], une DCT par bloc est appliquée sur l'image du visage, et les caractéristiques optimales sont choisies par un algorithme d'optimisation par essaim de particule (Particle Swarm Optimization PSO). La technique a permis aussi bien de réduire le nombre de caractéristiques obtenues ainsi de fournir une robustesse vis-à-vis des variations d'illumination, d'expression et de pose de l'image du visage.

3.6 L'identification (la reconnaissance)

Après avoir extrait les caractéristiques discriminantes d'un visage, cette étape permet de le classer et d'identifier son identité. Les approches utilisées dans cette étape appartiennent au domaine de la classification automatique, elles dépendent principalement sur des attributs utilisés comme descripteur lors de l'étape de l'extraction de caractéristiques. Certaines de ces méthodes s'emploient lors de l'étape d'identification telles que la méthode des k-plus proches voisins qui se base principalement sur le calcul d'une distance entre les descripteurs de la base de données et celui du visage à reconnaître. D'autres méthodes, qui se base sur l'apprentissage supervisé, s'emploient en deux phases, une d'apprentissage qui s'effectue lors de l'étape d'enrôlement pour déterminer les frontières entre les différentes classes d'individus et une phase, pendant l'identification, qui permet de classer le visage dans sa classe cible à partir des connaissances apprises lors de l'apprentissage. Les réseaux de neurones et les séparateurs à vastes marges (SVM) sont les méthodes les plus utilisées pour la reconnaissance de visage.

3.6.1 Classification par k-plus proches voisins

Ces méthodes permettent de calculer le degré de ressemblance entre les visages de la base de données et celui de test, en calculant des distances. Les distances qui peuvent être utilisées sont celles décrites au chapitre précédent. La distance euclidienne et la distance de Mahalanobis sont parmi les distances les plus utilisées pour la reconnaissance de visage.

Avec cette technique, le visage de la base de données le plus proche au sens de la distance utilisée est reconnu comme le visage cible, ce qui peut aboutir à des erreurs de classification. Pour faire face à ce problème, un seuil de rejet est défini, si la distance minimale est inférieure à ce seuil, le visage a été reconnu, si non le visage ne correspond pas à un visage de la base de données.

3.6.2 Classification par réseaux de neurones

Les réseaux de neurones discriminants sont largement utilisés dans le domaine de la reconnaissance faciale pour la classification ou l'identification du visage. Cette méthode utilise un réseau de neurones unique qui donne en sortie la classe d'appartenance (identité d'un individu) d'un visage inconnu en entrée. Le réseau est composé d'une

couche cachée et de N sorties, où N correspond au nombre d'individus dans la base de données (voir figure 3.19). Ce type de réseau a été utilisé pour classifier différents descripteurs qui sont calculés, par exemple, à partir d'une ACP [207], une transformée en cosinus [204], ou encore une transformée en ondelette [208].

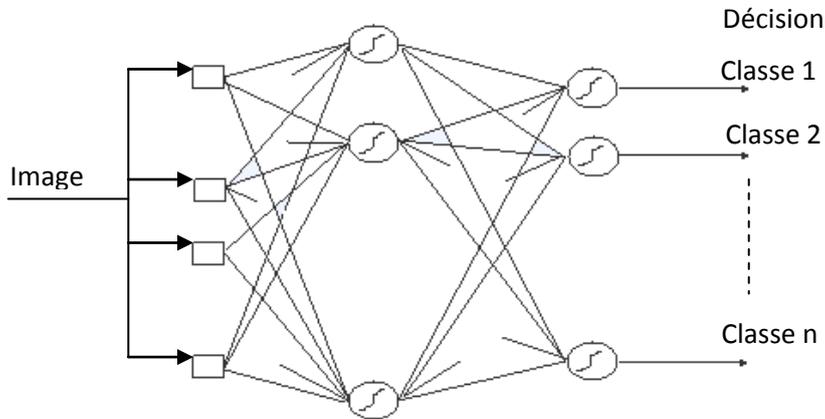


Figure 3.19- Réseau de neurones discriminant pour la reconnaissance faciale.

3.6.3 Classification par SVM

Les séparateurs à vastes marges sont des méthodes de classification supervisée, un bref aperçu sur ces méthodes est donné au chapitre 2. De nombreuses études ont démontré l'efficacité de la discrimination par les SVM, par exemple Huang [209] a atteint un taux de reconnaissance de 100% sur la base de données FERET. Des SVM à noyau ont également été évalués pour la classification de visage [210] où un noyau gaussien est utilisé.

3.7 Mesures d'évaluation

Étant donné que les systèmes biométriques standards peuvent être mis en œuvre dans deux modes différents [126] (à savoir l'identification ou la vérification), les algorithmes de reconnaissance peuvent être évalués selon les deux modes : identification ou vérification. Dans cette section, nous présentons brièvement les 2 modes d'évaluation des systèmes de reconnaissance de visage.

3.7.1 Vérification

Dans un contexte de vérification (authentification), les mesures suivantes sont utilisées :

Taux de faux rejets (TFR) [false reject rate (FRR)], ce taux représente le pourcentage de personnes qui devrons être acceptées mais qui sont rejetées par le système.

Taux de fausse acceptation (TFA) [false accept rate (FAR)], ce taux représente le pourcentage de personnes qui devront être rejetées mais qui sont acceptées par le système.

L'acceptation ou le rejet d'un individu est pris par un score de mise en correspondance en fonction d'un paramètre de seuil, si le score est supérieur au seuil, alors l'utilisateur est accepté, si non il est rejeté. Plus le seuil est grand plus le taux de fausses acceptations est important. Par contre, un seuil trop faible engendre un taux de faux rejets important, donc le seuil doit être choisi de manière à ce que les deux erreurs soient égales (TFA=TFR).

Taux d'erreurs égales (TEE) [Equal error rate (EER)], est la valeur de l'une des erreurs précédente qui correspond à l'endroit où TFR = TFA, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations.

Courbe ROC (Receiver Operating Characteristic): L'évaluation peut être faite également en utilisant la courbe ROC (Receiver Operating Characteristic), cette courbe trace le TFR en fonction du TFA pour un seuil variable (voir figure 3.20(b)).

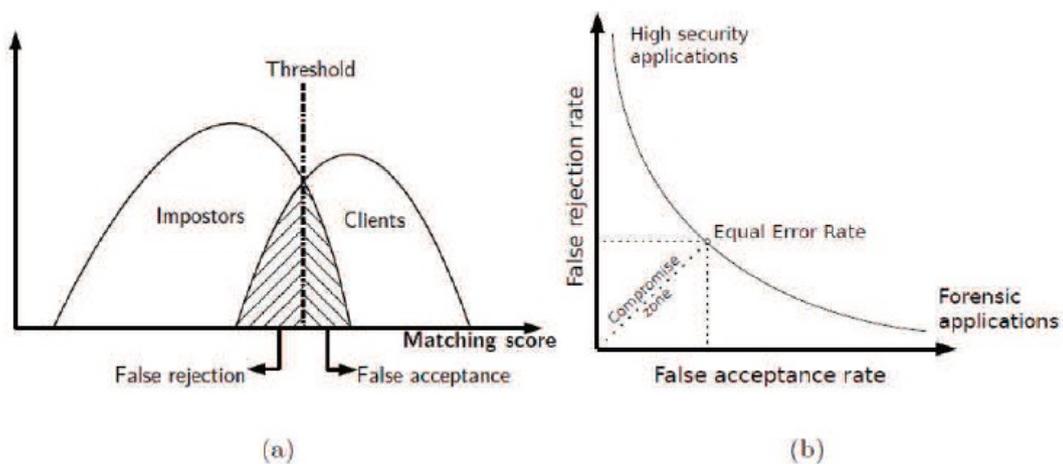


Figure 3.20- Illustration d'exemples typiques de graphiques d'un système biométrique, (a) les deux distributions représentent les scores client / imposteur; en faisant varier les valeurs du seuil différentes FAR et TFR peut être calculée. (b) Une courbe ROC est utilisé pour résumer les points opérationnels d'un système biométrique, pour chaque application différente des performances différentes sont nécessaires pour le système [211].

3.7.2 Identification

Dans un contexte d'identification (reconnaissance), les mesures suivantes sont utilisées :

Taux de reconnaissance (Rank-I Recognition Rate): Il mesure le pourcentage des personnes dont la classe est correctement identifiée.

Cumulative Match Characteristic (CMC): La courbe CMC donne le pourcentage de personnes reconnues en fonction d'une variable que l'on appelle le rang (voir figure 3.21). On dit qu'un système reconnaît au rang r , lorsqu'il choisit, parmi r images, celle qui correspond le mieux à l'image d'entrée. On peut dire que plus le rang augmente, plus le taux de reconnaissance correspondant est lié à un niveau de sécurité faible [212].

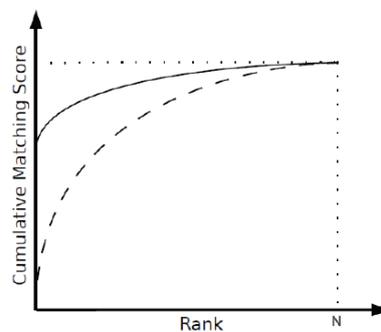


Figure 3.21- Exemple d'une courbe CMC, La ligne continue représente un système avec meilleures performances. N est le nombre d'image dans la base de données [211].

3.8 Conclusion

Dans ce chapitre, nous avons donné une idée générale sur la reconnaissance automatique du visage. Nous avons présenté l'architecture générale d'un système de reconnaissance faciale, en détaillant les deux principales phases, celle de description et celle d'identification, ainsi qu'une revue des principales méthodes de base existantes. La description de l'image a une influence remarquable sur les performances globales du système. Cela est dû, d'une part, à la complexité des objets dont on manipule, (car un visage regroupe un ensemble d'éléments qui sont à la fois semblables et très discriminants) et les variations que peut subir le visage de l'autre part.

Les variations d'âge constituent l'une des principales difficultés rencontrées par les systèmes de reconnaissance faciale, dans le chapitre qui suit, nous allons présenter une

revue des différentes techniques proposées dans la littérature pour faire face au facteur de la progression d'âge sur les systèmes de reconnaissance faciale.

CHAPITRE 4:

Reconnaissance de visage invariante à la progression d'âge

4.1 Introduction	074
4.2 Les techniques de reconnaissances de visages invariantes à la progression d'âge	074
4.2.1 Les approches génératives	075
4.2.2 Limites des approches génératives	080
4.2.3 Les approches non génératives	081
4.4 Conclusion	088

CHAPITRE 4

Reconnaissance de visage invariante à la progression d'âge

Ce chapitre se focalise sur le domaine d'étude de cette thèse qui est la reconnaissance de visage invariante à la progression d'âge, où nous présenterons une relecture de certains travaux antérieurs qui ont été réalisés pour faire face à l'impact de la progression d'âge sur les systèmes de reconnaissance automatique de visage, en décrivant les principales techniques qui ont été développées dans ce contexte et les résultats donnés dans la littérature.

4.1 Introduction

Le visage humain constitue une caractéristique importante pour les systèmes d'identification et de vérification surtout pour les applications biométriques, car il fournit l'information nécessaire pour identifier un individu, cependant, c'est la partie la plus affectée par le processus de croissance et de vieillissement dû à la progression d'âge. D'après les résultats de plusieurs études [24, 25, 26, 27, 28], la progression d'âge dégrade significativement les performances des algorithmes de reconnaissance faciale standards, d'où une nécessité de développer des techniques qui sont dédiées à la reconnaissance faciale invariante à la progression d'âge.

La plupart des études menées relatives à la progression d'âge dans le domaine de la reconnaissance de forme et de la vision par ordinateur se sont concentrées sur l'estimation de l'âge à partir du visage [215, 216, 217, 218, 219, 30] et la modélisation ou la simulation du processus de vieillissement du visage [220, 221, 222, 223]. En outre, certaines études se sont concentrées sur l'effet de la progression d'âge sur les apparences du visage [224, 225, 226]. Par contre, il existe peu d'algorithmes qui sont proposés pour construire des systèmes de reconnaissance de visage qui soient invariants à la progression d'âge.

4.2 Les techniques de reconnaissances de visages invariantes à la progression d'âge

Les méthodes de reconnaissance de visage invariantes à la progression d'âge peuvent être classées en deux parties [227]. Les approches de la première catégorie sont appelées "génératives" car elles suivent une approche de modélisation ou de simulation, dans laquelle la reconnaissance est effectuée en synthétisant l'image de visage à l'âge de celles de la base de données. L'autre catégorie comprend des approches "non-génératives" ou "discriminatives", elles se concentrent sur la définition de caractéristiques discriminantes et des méthodes de projection qui sont peu affectées par les variations d'âge pour permettre l'identification précise de l'identité. Dans ce qui suit, nous présentons quelques méthodes réalisées selon les deux approches génératives ou non génératives :

4.2.1 Les approches génératives

Les approches de cette catégorie sont principalement liées à la modélisation ou la simulation du processus de changement du visage dû à la progression d'âge, par exemple, deviner le visage d'un individu à l'âge de 60 ans à partir de l'image de son visage à l'âge de 20 ans. La figure 4.1 représente un exemple de simulation du processus de la progression d'âge proposé dans [228].

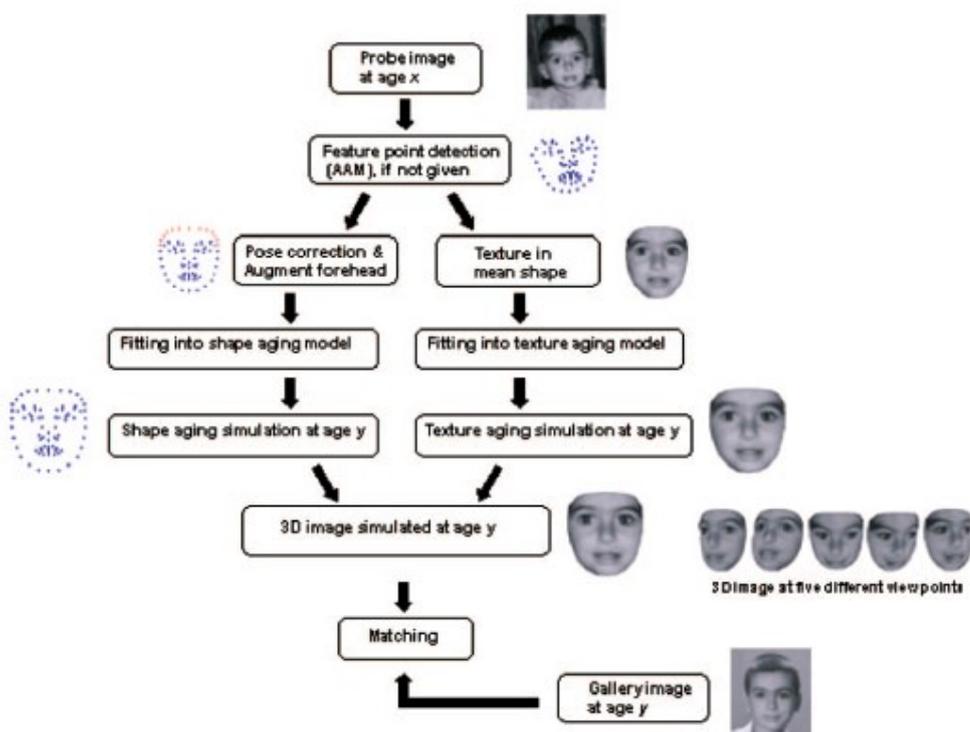


Figure 4.1- Simulation de la progression d'âge de l'âge X à l'âge Y [228].

Il existe plusieurs outils pour construire un modèle simulant ce processus, comme exemple, nous pouvons mentionner les travaux réalisés par Ramanathan et Chellappa [225] [226]. Le premier se concentre sur la modélisation du processus de croissance du visage chez l'enfant, tandis que le deuxième simule les changements du visage pendant l'âge adulte. Selon les auteurs, les changements faciaux dus à la progression d'âge chez l'enfant sont totalement différents que ceux chez l'adulte.

Dans leur approche [225], en se basant sur le modèle de croissance crâno-faciale [229], et autres études anthropométriques sur le visage [230], les auteurs ont proposé de modéliser les variations de la forme du visage qui sont dues à la croissance chez l'enfant. Le point clé dans leur technique est de modéliser la croissance du visage au moyen des paramètres de croissance définis sur des points de repères du visage. Ces points de repère sont aussi ceux les plus couramment utilisés dans les études anthropométriques. Le modèle proposé satisfait les critères des invariants géométriques qui caractérisent le modèle de croissance crâno-faciale, qui peuvent être définis comme suit [231]:

- les coordonnées angulaires de chaque point d'un objet dans un système de coordonnées polaires étant préservées;
- la symétrie bilatérale autour de l'axe vertical étant maintenue;
- la continuité des contours des objets étant préservée.

Les études anthropométriques sur le visage indiquent que différentes régions du visage changent de manières différentes avec le temps, et certaines régions sont plus stables que d'autres. Pour la reconnaissance faciale, certaines caractéristiques du visage sont extraites et des distances et des angles sont calculés entre elles. Ces mesures sont considérées comme invariantes, les auteurs dans [225] ont utilisé les mêmes mesures fournies dans [230, 232] pour construire le modèle de croissance. La figure 4.2 illustre les 24 points de repère et les différentes mesures utilisés.

L'approche présentée dans [225] ne traite que la croissance géométrique du visage et ne considère pas d'autres attributs comme la texture par exemple. La figure 4.3 présente un exemple d'images résultats de la transformation d'âge sur des individus différents ainsi que les images originales de l'individu à l'âge cible.

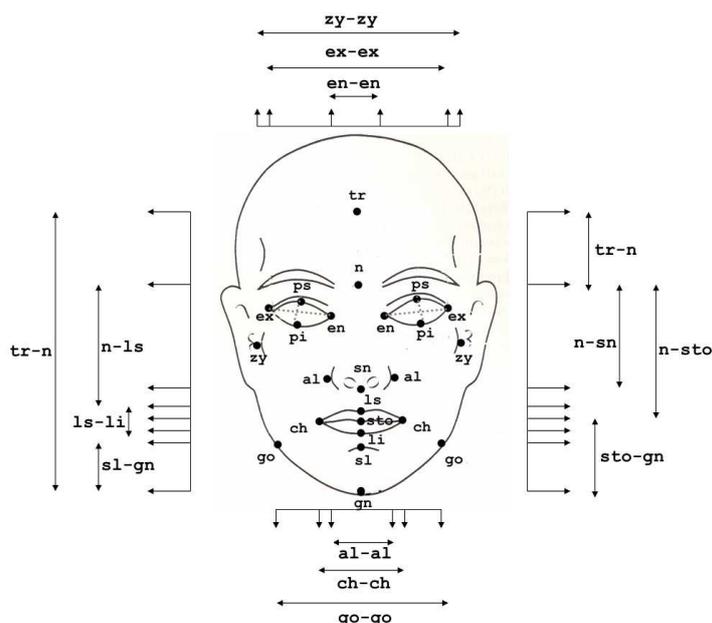


Figure 4.2 - Anthropométrie du visage: Point de repère et mesures utilisée pour la construction du modèle proposé dans [225].



Figure 4.3- Résultats de transformation d'âge sur des individus différents [225] (Les images originales sont de la base d'images FG-NET).

Dans leur seconde approche [226], les auteurs ont présenté une technique pour modéliser le processus de vieillissement chez l'adulte, la technique considère les variations de forme et celles de texture, où les transformations de forme ont été formulées par un modèle paramétrique qui capture les déformations que peuvent subir les traits du visage avec l'âge. Le modèle développé est basé sur l'hypothèse suivante: les degrés de liberté associés aux déformations des traits faciaux sont directement liés aux propriétés physiques et des orientations géométriques des muscles faciaux sous-jacents.

En s'inspirant du modèle musculaire de Waters [233], les auteurs ont défini trois types de muscles, à savoir les muscles linéaires, les muscles en feuilles et les Sphincters, ainsi qu'un modèle de transformation pour chaque type de muscles. La figure 4.4 illustre les 18 muscles faciaux qui ont été utilisés pour développer le modèle.

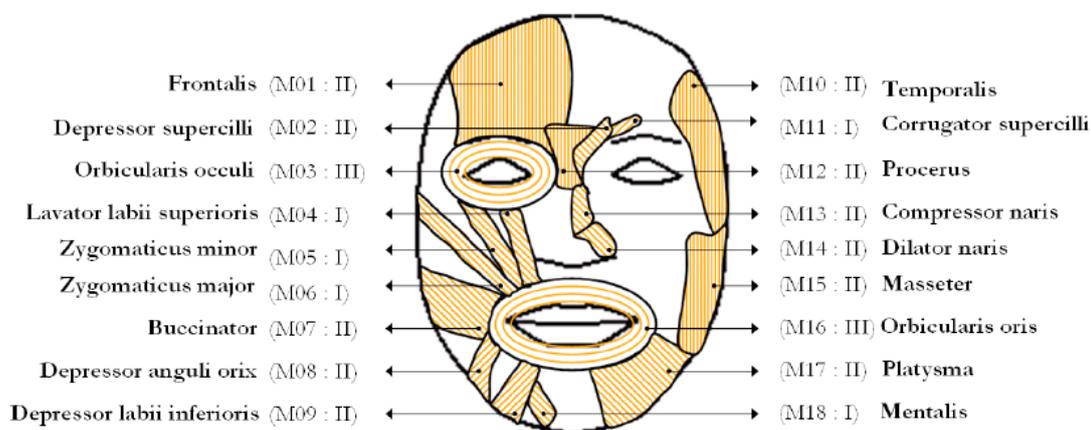


Figure 4.4 - Configuration des différents types de muscles faciaux [226].

Par ailleurs, quarante-huit points de repère ont été sélectionnés sur lesquels quarante-quatre mesures projectives ont été extraites pour construire le modèle, la figure 4.5 présente ces 48 points et leur correspondance avec les muscles faciaux utilisés. De plus des facteurs tels que la perte ou le gain du poids à travers l'âge peuvent influencer la déformation des parties du visage. Le modèle de forme proposé a été formulé de telle sorte que ces facteurs externes sont pris en compte.

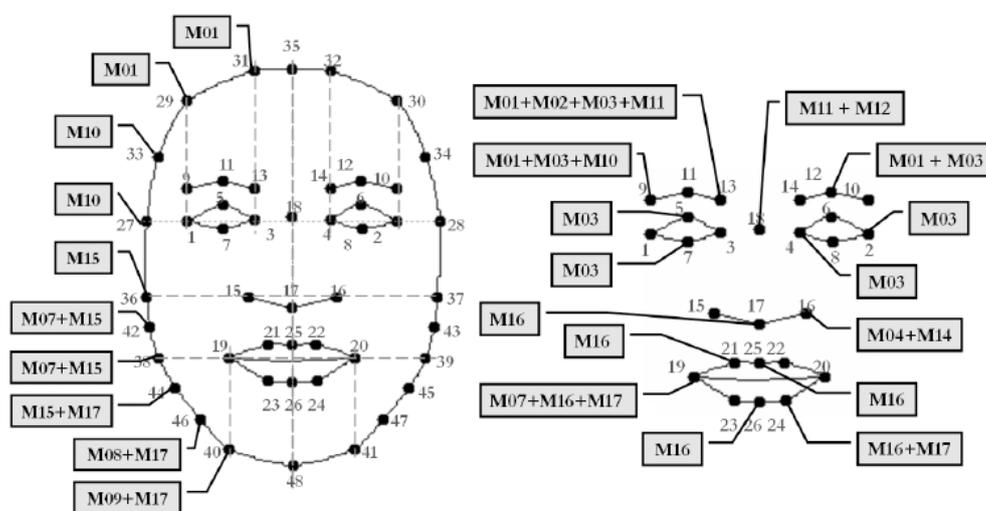


Figure 4.5- Les 48 points de repère qui sont utilisés ainsi que leur correspondance avec les muscles faciaux de la figure 4.4 [226].

Aussi, les transformations de texture, qui sont caractérisées par les rides ou d'autres déformations de la peau (taches brunes) à travers le temps, ont été formulées en utilisant les fonctions de transformation du gradient de l'image. La figure 4.6 présente un aperçu du modèle de vieillissement facial proposé dans [226].

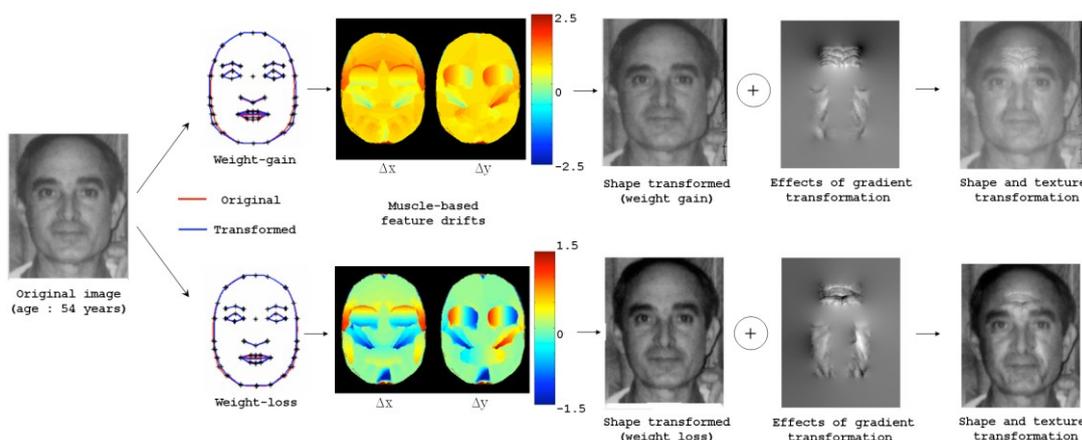


Figure 4.6- Un aperçu du modèle de vieillissement facial proposé dans [226], Les changements de la forme du visage dans le cas d'une perte ou d'une prise de poids sont illustrés. En outre les effets des transformations du gradient qui engendrent les variations de texture sont bien illustrés.

Nous pouvons également citer le travail de Park et al. [228], où ils ont proposé un modèle simulant les variations de forme et de texture du visage, qui sont causés par le vieillissement. Le modèle est utilisé pour la reconnaissance faciale invariante à l'âge. Lors d'une interrogation, l'image requête est transformée au même âge que les images de la base de données en utilisant le modèle de vieillissement 3D déjà formé. Dans leur approche, les auteurs considèrent que le vrai modèle de vieillissement crâno-facial [234] ne peut être formulé correctement que dans le domaine 3D et l'utilisation d'un modèle 3D offre une caractérisation plus puissante qu'un modèle 2D puisque les changements de l'apparence du visage humain se produisent en 3D. En outre, l'extension du modèle de forme du domaine 2D au domaine 3D peut compenser encore le problème de changement de pose et potentiellement celui d'éclairage. Les auteurs ont montré que leur approche peut gérer à la fois les effets de croissance chez l'enfant (développement de l'enfant) et le vieillissement du visage chez l'adulte. Un exemple d'images faciales obtenues par simulation du processus de vieillissement d'un individu âgé de 2 ans est présenté dans la figure 4.7.

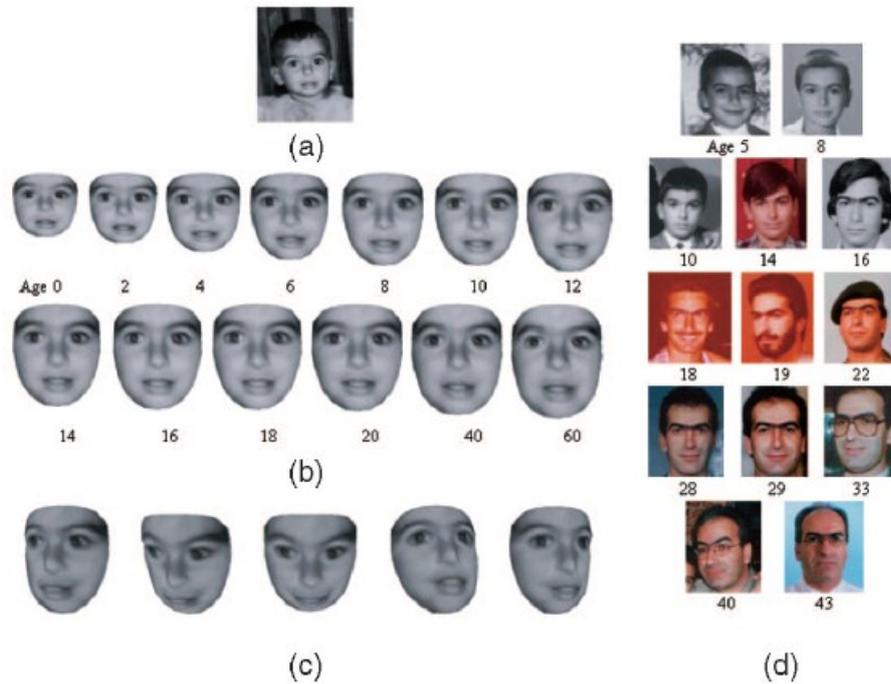


Figure 4.7- Exemple d'images faciales obtenues par simulation du processus de vieillissement d'un individu âgé de 2 ans [228]
 (a) individu âgé de 2 ans, (b) visages obtenus de la simulation (de 0 à 60 ans), (c) visage obtenu à l'âge de 20 ans en 5 positions différentes et (d) la vérité terrain.

Dans le même contexte, Lanitis et al. [220] ont proposé une approche permettant de caractériser les variations de la forme et de l'intensité du visage à travers la progression d'âge par un modèle statistique basé sur une analyse en composantes principales qui fournit un codage compact et réversible de l'image. A partir d'un certain nombre d'images d'apprentissage, la relation entre la représentation du visage et l'âge réel de l'individu dans les images correspondantes est apprise. Une fois cette relation est établie, il est possible de prédire l'âge d'un nouvel individu. La méthodologie proposée a été utilisée pour concevoir un système de reconnaissance de visage, robuste aux variations d'âge.

4.2.2 Limites des approches génératives

L'inconvénient majeur des approches génératives, c'est que l'image produite (synthétisée) à l'âge estimé peut ne pas être forcément semblable à l'image réelle de l'individu à cet âge, car prédire avec précision comment les individus peuvent apparaître à un âge progressé est une tâche très difficile, d'autres facteurs comme le mode de vie et l'environnement contribuent également au changement du visage en plus des facteurs

biologiques. Et également, transformer une image requête à un âge spécifique, lui même fait appel à un autre axe de recherche qui est l'estimation de l'âge d'un individu à partir de sa photo.

4.2.3 Les approches non génératives

Parmi les premières études appartenant à cette catégorie, nous pouvons citer les travaux de Ling et al. [29, 238]. En s'appuyant sur des travaux antérieurs, qui ont justifié l'application des orientations du gradient pour la reconnaissance faciale invariante aux changements d'illumination [235], et en se basant sur d'autres études précédentes du modèle de croissance cranô-faciale [231] et celui de la couleur de la peau [236, 237], les auteurs ont constaté qu'une représentation pyramidale des directions du gradient (GOP: Gradient orientation pyramide) peut être utilisée pour construire des descripteurs qui sont insensibles aux changements causés par le vieillissement du visage. Partant de cette idée, les auteurs ont proposé de combiner les différences entre les paires d'images calculées à partir du cosinus entre les orientations du gradient à toutes les échelles avec une machine à vecteurs de support (SVM) pour une tâche de vérification du visage à travers la progression d'âge. La figure 4.8 représente une image de visage en pyramide.

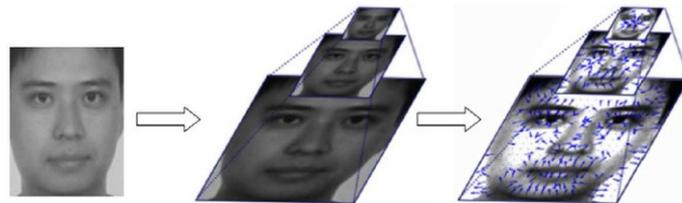


Figure 4.8- Pyramide d'orientation du gradient [29].

L'approche proposée a été testée en utilisant deux bases de données de passeports privées pour étudier comment l'augmentation des écarts d'âge affectent les performances de vérification. Ils ont également étudié les effets liés aux vieillissements tels que la qualité d'image, la présence de lunettes, et la présence d'une barbe sur le visage. Aussi, les auteurs ont démontré qu'en éliminant la magnitude du gradient et en utilisant certaines techniques hiérarchiques, le nouveau descripteur fournit une représentation plus robuste et plus discriminante.

Dans le même contexte, Mahalingam et Kambhamettu [239] ont développé une approche simple, mais efficace pour la vérification faciale à travers la progression d'âge dans laquelle ils ont défini un opérateur de visage en se basant sur les histogrammes des motifs binaires locaux (Local Binary Pattern LBP). La figure 4.9 représente un exemple de calcul de descripteur dans [239]. L'information spatiale est incorporée par la construction d'une pyramide gaussienne pour chaque image du visage et le calcul de l'histogramme des LBP à chaque niveau. Les caractéristiques les plus discriminantes du visage sont extraites de l'histogramme des LBP en utilisant l'algorithme d'apprentissage AdaBoost. Leurs résultats indiquent que les meilleurs taux de reconnaissance sont obtenus avec des différences d'âge plus petites entre les paires d'images et ils se stabilisent lorsque la différence augmente. En plus, des facteurs comme la barbe, le port de lunettes, etc. fournissent des indices discriminant pour le système de vérification de visage.

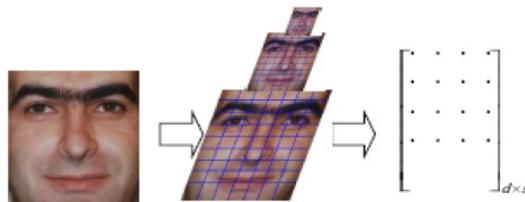


Figure 4.9- Calcul de la pyramide des LBP à partir d'une image [239].

Egalement, Singh et al. [240], ont présenté un algorithme de transformation pour minimiser les variations des traits du visage causée par la progression d'âge, cet algorithme enregistre les images de la galerie et l'image requête dans le système de coordonnées polaires. Les coordonnées des yeux et de la bouche sont utilisées pour former un triangle et le centre de ce triangle est utilisé comme centre pour la conversion aux coordonnées polaires. Egalement cette conversion permet d'éliminer les variations de la pose. La figure 4.10 représente une image de visage en coordonnées cartésiennes et la même image transformée en coordonnées polaires. L'efficacité de la méthode de transformation proposée est validée en utilisant un algorithme de reconnaissance faciale basé sur la transformée de Gabor en log 2D (2D log polar Gabor transform) pour la comparaison des images de visages. L'approche proposée fournit les meilleures performances pour les groupes d'individus au-delà de 19 ans.

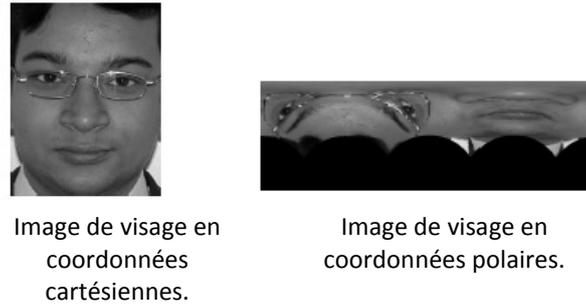


Figure 4.10 Un exemple de conversion des coordonnées cartésiennes aux coordonnées polaires [240].

Une autre équipe qui a adopté une approche similaire est Biswas et al. [241], les auteurs ont donné un algorithme de vérification qui analyse la cohérence des dérivées dans divers traits du visage pour vérifier si deux images prises à différents âges sont pour la même personne ou non. Selon les auteurs, les cohérences de certains traits sélectionnés du visage sur deux images de la même personne avec différent âges sont très élevées (voir figure 4.11). Dans leur étude, sauf quelques traits faciaux frontaux qui sont choisis parce que les autres points sur les limites extérieures du visage se changent rapidement avec les mouvements de la tête et les expressions (le rire ou même le sourire).



Figure 4.11- Les dérivées (drifts) des traits faciaux de quelques visages à différent âges. Les dérivées des images du même individu apparaissent être cohérentes (les deux premières lignes) alors qu'elles sont assez incohérentes (troisième ligne) lorsque les images appartiennent à des d'individus différents [241].

Dans la même optique, Mahalingam et al. [242] ont proposé une méthode de reconnaissance de visage invariante à l'âge en représentant le visage par un graphe qui contient la géométrie et l'apparence des points caractéristiques du visage. Un modèle probabiliste de vieillissement est construit pour chaque individu en utilisant des modèles de mélange gaussien (Gaussian mixture models - GMMs), le modèle intègre à la fois les informations de la forme et de texture. Un graphe est construit pour chaque individu en utilisant un algorithme simple, les points qui correspondent aux traits caractéristiques du visage constituent les sommets du graphe, chaque sommet est étiqueté par le descripteur associé, qui est calculé en utilisant les motifs binaires locaux (LBP). Les traits caractéristiques sont extraits en utilisant une version modifiée de l'analyse statistique locale des caractéristiques (Local features analysis, LFA) [243], dans cette technique, des noyaux sont construits, et ils sont utilisés, par la suite, comme vecteurs de base pour l'extraction des traits caractéristiques. La géométrie du graphe est définie en imposant des contraintes sur la longueur des arêtes entre un sommet et ses voisins. La mise en correspondance de deux graphes est effectuée en deux étapes: Dans la première étape, une probabilité de maximum à postériori (a maximum a posteriori probability) est calculée en utilisant le modèle de vieillissement des individus pour réduire l'espace de recherche et identifier les individus potentiels pour la deuxième étape. Dans la seconde étape, un algorithme déterministe d'appariement de graphe qui exploite la similarité spatiale entre les graphes est proposé. D'après les auteurs, la combinaison d'un modèle de vieillissement avec une représentation graphique est plus performante pour un système de reconnaissance de visage à travers la progression d'âge. Aussi, une représentation efficace de la relation spatiale entre les points caractéristiques du visage peut améliorer les performances d'un système de reconnaissance faciale à travers la progression d'âge.

Par ailleurs, Yang et al. [244] ont proposé une méthode nommée Texture embedded discriminative graph matching (TED-GM) dans laquelle le problème de la reconnaissance faciale invariante à la progression d'âge est formulé comme un problème d'appariement de graphe. L'approche proposée étend la méthode Elastic Bunch Graph Matching EBGM [191] suivant deux aspects principaux à savoir la représentation de la texture et le schéma d'appariement. Plus précisément, l'image du visage est convertie en un graphe en utilisant une triangulation de Delaunay. Les nœuds du graphe représentent

la texture de régions chevauchées et les arêtes correspondent à la topologie géométrique de ces régions. La texture est extraite en appliquant la méthode Local Gabor Binary Partterns Histogram Sequence (LBPHS) proposées dans [245], puis une PCA et une LDA sont introduites pour réduire la dimension. La similarité entre l'image requête et celles de la base d'images est calculée par une fonction objective qui est conçue pour l'appariement de graphes dans une phase d'identification. La figure 4.12 représente les différentes étapes d'extraction de caractéristiques et d'identification proposées dans [244].

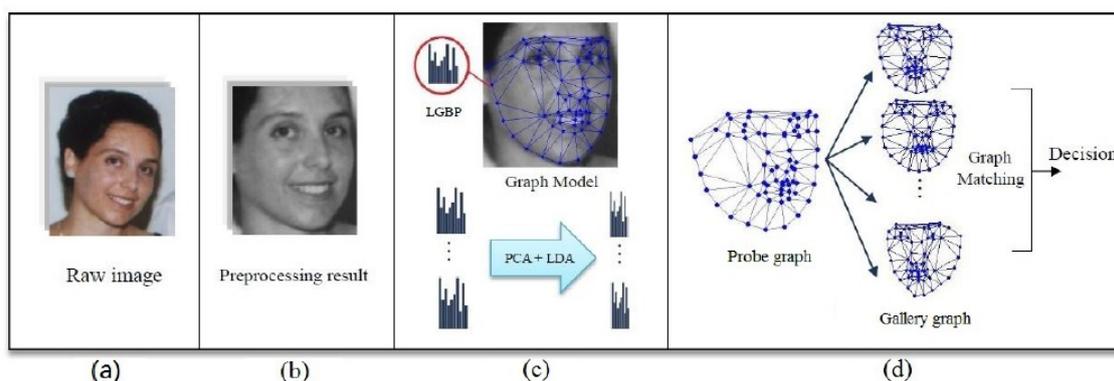


Figure 4.12- Illustration des différentes étapes de l'approche TED-GM pour la reconnaissance de visage invariante à l'âge [244].

Ainsi, Guo et al. [246] ont étudié l'influence de la progression d'âge sur les systèmes de reconnaissance faciale. Ils ont évalué les performances de l'utilisation de l'analyse en composante principale (PCA) et la méthode EBGM (Elastic Bunch Graph Matching) [191] dans le contexte de la reconnaissance faciale à travers la progression d'âge. Les deux méthodes ont été testées séparément, combinées et complétées par des signes de biométrie douce (Soft Biometrics) tels que la race, la taille et le poids, etc. Les auteurs ont démontré que le taux de reconnaissance ne se dégrade pas linéairement par rapport aux intervalles d'âge, et lorsque les différences d'âge sont supérieures à 15 ans, le taux de reconnaissance devient considérablement inférieur à ceux rapportés dans des intervalles inférieurs à 15 ans. Aussi, les auteurs ont montré que l'utilisation de plusieurs signes de biométrie douce combinés avec les caractéristiques principales peut significativement améliorer les performances de la reconnaissance faciale à travers la progression d'âge.

Dans le même contexte d'évaluation, Meng et al. [247] ont réalisé une étude comparative entre les pyramides d'orientations du gradient (GOP), les motifs binaires

locaux et les ondelettes de Gabor dans la présence des variations d'âge; les trois représentations sont suivies par une analyse en composantes principales. Selon les auteurs, les ondelettes de Gabor en utilisant 5 échelles et 8 orientations ont fourni les meilleurs taux de reconnaissance parmi les trois méthodes testées. Aussi, les LBP fournissent de bonnes performances dans les intervalles d'âge 7 à 9 ans et 10 à 12 ans.

Egalement, Bereta et al. [248] ont évalué dix descripteurs locaux largement utilisés dans la reconnaissance du visage, qui sont: les LBP (local binary pattern), les ILBP (improved LBP), les CSLBP (center-symmetric LBP), les DLTP (differential local ternary pattern), les TPLBP (three patch LBP), les MBLBP (multi-scale block LBP), les WLD (simplified form of Weber local descriptor), les LXP (local XOR pattern), les LGPBP (local Gabor phase binary pattern), et les LGXP (local Gabor XORpattern), dans le contexte de la progression de l'âge. Les descripteurs choisis pour l'étude sont ceux qui ont fourni des taux de reconnaissance prometteurs lors de l'utilisation d'images d'individus sans aucune différence d'âge. Dans leur étude la classification est effectuée par le calcul d'une distance entre des vecteurs descripteurs contenant la texture locale en appliquant des mesures de similarité différentes telles que la distance euclidienne, cosinus, Manhattan et corrélation, etc. Leurs résultats montrent que des taux de reconnaissance élevés sont obtenus en combinant les différents descripteurs locaux avec des magnitudes du filtre de Gabor et plus particulièrement le descripteur MBLBP qui apparaît le plus stable dans la présence des variations d'âge.

Dans le même but de la reconnaissance faciale à travers la progression d'âge, Li et al. [249] ont proposé un modèle discriminant où chaque visage est représenté par un schéma de description qu'ils l'ont appelé: " A densely sampled local feature description scheme" dans lequel chaque image de visage est divisée en un ensemble de patches imbriqués et chaque patch est représenté par les deux descripteurs locaux, à savoir la transformation de caractéristiques invariante à l'échelle (Scale invariant feature transform SIFT) et les motifs binaires locaux multi-échelles (MLBP) (voir figure 4.13). Les deux descripteurs sont suivis par une analyse discriminante multi-descripteurs (Multi-feature discriminant analysis : MFDA) pour la réduction de la dimension. La MFDA est une extension et une amélioration de l'analyse discriminante linéaire (Linear discriminant analysis LDA) en utilisant plusieurs descripteurs combinés avec deux méthodes différentes d'échantillonnage aléatoire dans l'espace de descripteurs et

d'échantillons. Par un échantillonnage de l'ensemble d'apprentissage ainsi que l'espace de descripteurs, plusieurs classifieurs basés sur la LDA sont construits et combinés pour générer une décision robuste via une règle de fusion.

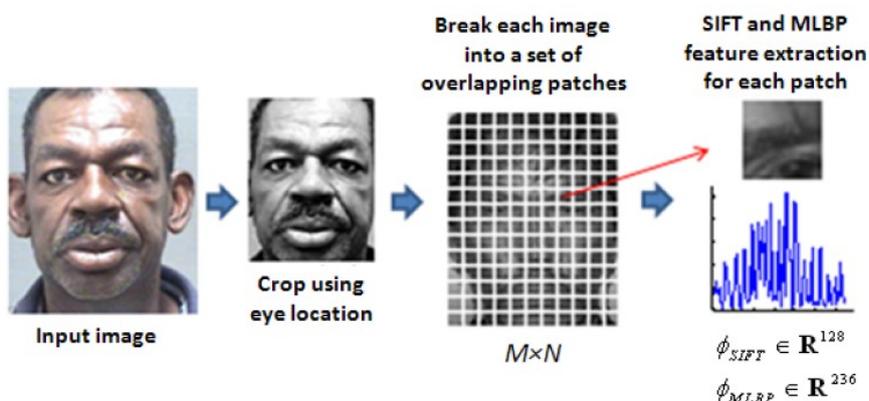


Figure 4.13- Représentation des caractéristiques locales d'une image d'un visage [249].

Par ailleurs, Sungatullina et al. [250] ont présenté une méthode d'apprentissage nommée (multiview discriminative learning MDL method) pour la reconnaissance faciale invariante à l'âge. De la même façon que dans [249], chaque image de visage est divisée en un ensemble de patches chevauchés et chaque patch est représenté par trois descripteurs locaux: la transformation de caractéristiques invariante à l'échelle SIFT, les motifs binaires locaux LBP et la pyramide d'orientation du gradient GOP. Les trois descripteurs sont projetés dans un sous espace latent de faible dimension en utilisant la méthode d'apprentissage MDL proposée, où pour chaque descripteur, la variation entre les éléments de la même classe est minimisée et la variation entre les éléments de classes différentes est maximisée.

En outre, J-Xu et al. [251] ont considéré que le visage humain avec sa structure complexe se change facilement avec le temps en terme de couleur, de texture et de structure, c'est pourquoi sa modélisation à travers l'âge est très difficile, cependant, la région péri-oculaire est plus stable à travers l'âge, car la forme et l'emplacement des yeux restent pratiquement inchangés (voir figure 4.14) tandis que la bouche, le nez, le menton et les joues sont plus sensibles à la progression d'âge, surtout la peau. En plus, la région péri-oculaire possède les caractéristiques discriminantes les plus complexes dans le visage telles que: les contours, les paupières et les sourcils, etc. qui pourraient tous varier en forme, en couleur et en taille, c'est pourquoi cette région peut être la zone du

visage la plus importante pour distinguer les personnes. Partant de cette hypothèse, les auteurs ont proposé une approche d'extraction de caractéristiques pour la reconnaissance faciale invariante à l'âge en appliquant la combinaison de la transformée de Walsh-Hadamard avec les motifs binaires locaux sur des régions péri-oculaires prétraitées, suivie d'une projection discriminante non supervisée (Unsupervised Discriminant Projection UDP) pour la modélisation des sous-espaces.

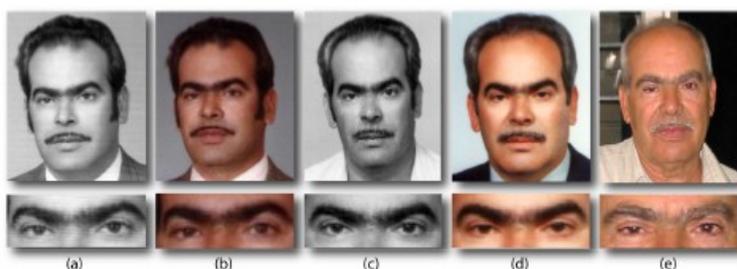


Figure 4.14- Un exemple d'un individu à différents âges. (a) 31 ans, (b) 40 ans, (c) 46 ans, (d) 61 ans et (e) 69 ans. La région péri-oculaire est perceptivement plus stable à travers l'âge que le visage complet [251].

Egalement, Nayak et al. [252] ont développé une méthode nommée Self-PCA pour la reconnaissance faciale invariante à l'âge, où ils ont appliqué une analyse en composantes principales pour chaque individu car les auteurs ont considéré que les effets de la progression d'âge sont spécifiques à chaque individu. Lors d'une étape de test, l'image requête est appariée en la projetant sur l'espace propre de chaque individu de manière indépendante et en fonction de la distance euclidienne. En se basant sur l'étude [251], les auteurs ont choisi seulement la région péri-oculaire pour construire les espaces propres.

4.4 Conclusion

La reconnaissance faciale a déjà son sac plein de challenges et, récemment l'accent est mis sur un autre problème, qui est en effet, plus difficile: c'est celui de l'effet de la progression d'âge sur les systèmes de reconnaissance de visage. C'était l'objectif principal de ce chapitre, où nous avons présenté un état de l'art des techniques proposées dans ce domaine. Certainement, chaque approche a ses avantages et ses inconvénients. Par exemple, nous pouvons conclure que l'inconvénient majeur des approches génératives, c'est que ces dernières se basent sur la modélisation d'un processus qui est en réalité incontrôlable, pour cela nous nous sommes orientés vers le deuxième type qui

sont les approches non génératives où une approche de description de visage est proposée, les détails de cette dernière font l'objet du chapitre 5.

Partie II :

Contribution

CHAPITRE 5:

Méthode de description de visage invariante à la progression d'âge

5.1 Introduction	091
5.2 La normalisation d'un visage dans une image	092
5.2.1 L'alignement des formes	094
5.2.2 La déformation de la texture de l'image	096
5.2.3 Processus de normalisation	098
5.3 Extraction de caractéristiques	099
5.4 La transformée en cosinus discrète	099
5.4.1 Principe	099
5.4.2 Choix de la DCT	101
5.5 Réduction de la dimension du descripteur	102
5.5.1 Analyse de Fisher à noyau	103
5.6 Reconnaissance (identification)	105
5.7 Conclusion	106

CHAPITRE 5

Méthode de description de visage invariante à la progression d'âge

Dans ce chapitre, nous allons détailler les différentes étapes qui constituent notre méthode de calcul de descripteur invariant à la progression d'âge ainsi que le processus d'identification qui sont proposés dans [253]. En justifiant notre choix pour chaque méthode utilisée dans toutes les étapes.

5.1 Introduction

Récemment des résultats prometteurs ont été fournis dans le domaine de la reconnaissance faciale, cependant, l'invariance à la progression d'âge reste toujours compliquée, Dans ce chapitre nous allons présenter une approche simple discriminative mais efficace pour décrire et identifier des visages à travers la progression d'âge.

Le processus général de calcul de descripteur est illustré dans la figure 5.1, les différentes étapes qui le composent sont détaillées dans les paragraphes qui suivent:

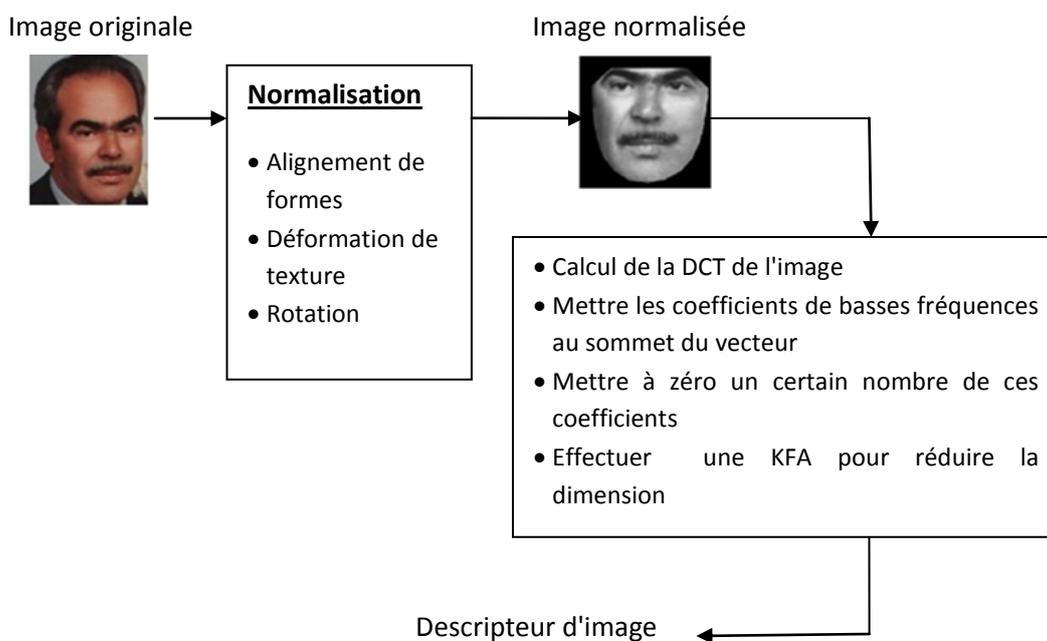


Figure 5.1- Différentes étapes pour le calcul du descripteur.

5.2 La normalisation d'un visage dans une image

L'étape de normalisation des images originales présente une partie essentielle dans un système de reconnaissance faciale, car cette étape a un grand impact sur les performances et la robustesse de la procédure de reconnaissance. Par exemple Tan et Triggs [254] ont démontré que la robustesse de plusieurs méthodes de sous espace linéaires et des motifs binaires locaux peut être considérablement améliorée en ajoutant une simple étape de prétraitement basée sur une correction Gamma ou une normalisation de variance. La boîte à outils réalisée en MATLAB : INFACE (illumination invariant face recognition toolbox) [255] inclue un ensemble de fonctions pour effectuer des normalisations d'illumination, par conséquent, relever l'un des plus grand défis de la reconnaissance faciale. Aussi, la variation de la pose constitue un facteur critique qui peut limiter les performances des systèmes de reconnaissance faciale. Son impact sur les algorithmes de reconnaissance peut être minimisé en utilisant des techniques qui transforment des visages dans une position non frontale en une position frontale, ce type de traitement peut être effectué par des méthodes 2D ou 3D. Parmi les méthodes 2D, les modèles d'apparence actifs [167] sont considérés comme des outils très puissants qui permettent de présenter un visage indépendamment de sa position.

Dans notre approche proposée [253], nous nous sommes intéressés à la correction de la pose où nous avons choisi d'utiliser la même approche que dans les modèles d'apparence actifs (AAM) [256], qui est décrite en détail dans les paragraphes qui suivent:

Avant de présenter les différentes étapes constituant cette phase de normalisation, nous allons définir ce qu'est une forme? et comment la décrire?

Selon I.-L. Dryden et al. [257]: une forme est toute information géométrique qui reste invariante lorsque la position, l'échelle, et les effets de rotation sont ignorés. Elle peut être représentée en localisant un nombre fini de points sur son contour, ces points sont appelés points de repère (ou landmark en anglais), ils doivent être appropriés pour représenter la forme générale.

Selon Cootes et al. [256], ces points de repère peuvent être anatomiques, ce sont des points définis et affectés par des experts, des points mathématiques, qui sont déterminés

en fonction des propriétés mathématiques ou géométriques telles que les coins et les jonctions ou bien des points intermédiaires, la figure 5.2 représente les différents types de points d'intérêts.

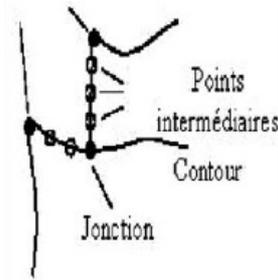
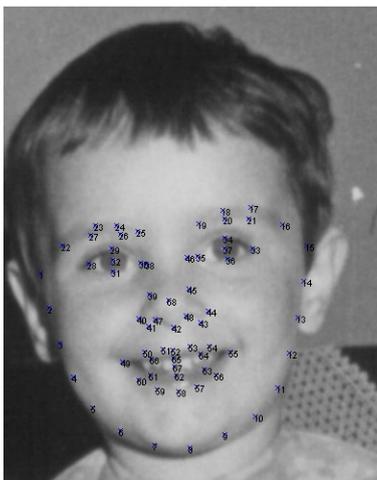


Figure 5.2- Classes des points de repère (ou d'intérêts) [256].

La figure 5.3 montre un Exemple d'un visage de la base de données FG-NET annoté par 68 points de repère (a), et un autre exemple d'une main annotée par 56 points de repère [258] (b).



(a)



(b)

Figure 5.3- (a) Exemple d'un visage de la base de donnée FG-NET annoté par 68 points de repère.

(b) Exemple d'une main annotée par 56 points de repère [258]

Mathématiquement, une forme représentée par n points dans un espace de K dimensions est décrite par un vecteur de dimension K_n qui concatène chaque dimension. Une forme en 2D, X , composée de n points de repère est représentée par :

$$X = (x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)^T$$

Donc, après l'annotation de la base d'apprentissage, on obtient un ensemble de vecteurs $\{X_i\}$, où X_i représente la forme associée à l'image i , $i = \overline{1, n}$.

La figure 5.4 représente un exemple d'un ensemble de formes qui sont associées aux images du visage pris de la base FG-NET.

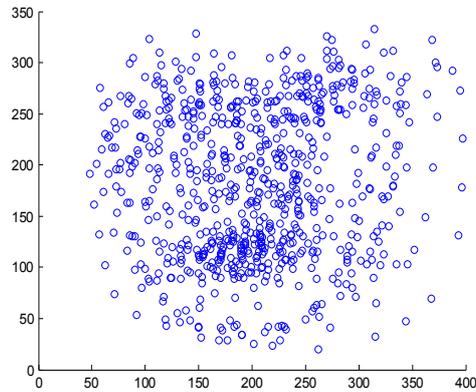


Figure 5.4- Un exemple d'un ensemble de formes.

5.2.1 L'alignement des formes:

Cette étape consiste à aligner tous les vecteurs de la base d'apprentissage, autrement dit: il s'agit d'enlever les différences de position, d'orientation et d'échelle entre les différents vecteurs d'apprentissage [259] pour les représenter dans un même repère de référence. Plusieurs solutions qui existent pour aligner deux formes dans un repère commun, la solution classique est l'analyse procrustéenne [256], son principe est le suivant:

Etant données deux formes S_1 et S_2 , Aligner la forme S_2 sur la forme S_1 consiste à trouver les paramètres d'une similarité T composée d'une translation (t_x, t_y) , d'un angle de rotation θ et d'un facteur d'échelle s qui, lorsqu'elle est appliquée à la forme S_2 , la forme S_2' obtenue par rotation par l'angle θ , multiplication par le facteur d'échelle s et translation de la forme S_2 , soit le plus proche de S_1 au sens des moindres carrés. C.-à-d., après alignement la distance de Procruste entre S_1 et S_2 doit être minimale.

$$D_{procrustes}(s_1, s_2) = \sqrt{\sum_{i=1}^n (x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (5.1)$$

Avec respect de s , θ et (t_x, t_y) .

Cootes [256] a proposé une simple méthode pour estimer s et θ , où:

$$a = (s1^T s2)/|s1|^2 \quad (5.2)$$

$$b = \sum_{i=1}^{68} (x_{i,1}y_{i,2} - y_{i,1}x_{i,2})/|s1|^2 \quad (5.3)$$

Et l'angle de rotation θ et le coefficient de mise à l'échelle s sont donnés par:

$$\theta = \arctan(b/a) \quad (5.4)$$

$$s = \sqrt{a^2 + b^2} \quad (5.5)$$

Dans notre cas, cette procédure d'alignement doit être appliquée à toutes les formes de la base d'apprentissage par une approche simple qui est décrite dans l'algorithme suivant [256]:

1. Faire une translation de chaque forme de telle sorte qu'elle soit centrée à l'origine
2. calculer la forme moyenne de toutes les formes
3. Choisir la forme moyenne comme forme de référence x_0 et changer son échelle de telle sorte que $|\bar{x}_0| = 1$.
4. Aligner toutes les formes sur la forme moyenne
5. Recalculer la moyenne des formes alignées
6. Aligner la forme moyenne sur la forme de référence et la normaliser ($|\bar{x}| = 1$)
7. Aller à l'étape 4 jusqu'à convergence (la forme moyenne des formes alignées ne varie pas significativement par rapport à l'itération précédente).

La figure 5.5 représente le même ensemble présenté dans la figure 5.4 après alignement.

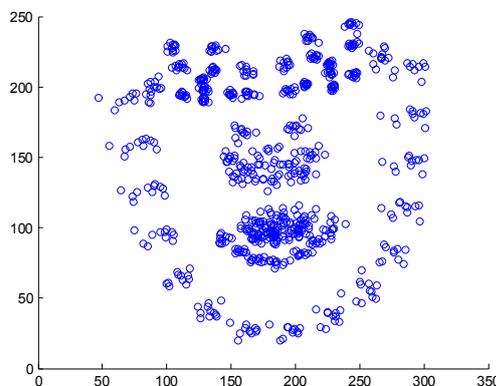


Figure 5.5- Un exemple d'un ensemble de formes alignées.

5.2.2 La déformation de la texture de l'image

La texture d'une forme constitue l'ensemble des échantillons de pixels de niveaux de gris situés à l'intérieur du polygone constitué par les points de frontière de cette forme. La figure 5.6 montre un exemple d'une image d'un visage annoté par 68 points d'intérêt qui constituent sa forme et sa texture à l'intérieur de cette forme. Sa représentation mathématique est donnée par un vecteur g_{im} composé de m niveaux de gris où:

$$g_{im} = (g_1, g_2, \dots, g_m)^T$$

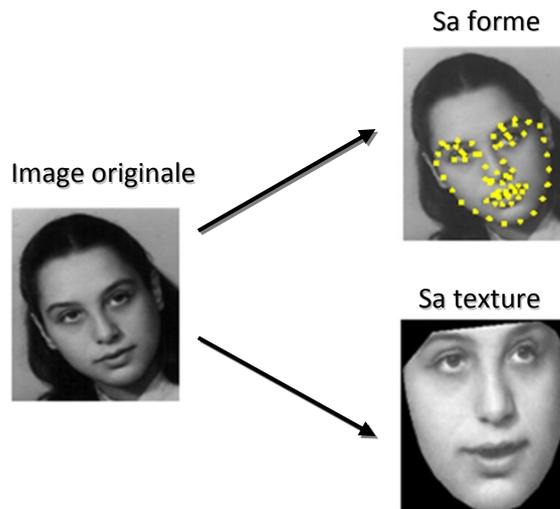


Figure 5.6- Image originale, sa forme et sa texture à l'intérieur de la forme [253].

La déformation de la texture, ou le *Warping* en anglais, permet de rendre la représentation de la texture indépendante des variations de sa forme. Son principe est le suivant: pour une forme donnée dans une image, on fait extraire sa texture et la déformer de telle sorte que ses points d'intérêt correspondent à ceux de la forme moyenne qui est calculée à l'étape d'alignement. La triangulation de Delaunay est utilisée pour établir des triangles qui seront utilisés, par la suite, pour déformer la texture. Cette triangulation stipule que le cercle circonscrit d'un triangle n'ait en son intérieur aucun autre point d'intérêt, la figure 5.7 illustre les propriétés de Delaunay.

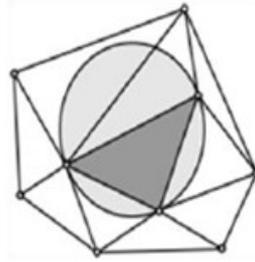


Figure 5.7- Propriétés de Delaunay.

La déformation de la texture peut être vue comme une fonction de transfert qui associe à chaque pixel de la forme moyenne le pixel de la forme à déformer.

Une transformation affine pour chaque triangle est utilisée, où chaque pixel appartenant à un triangle est transféré dans le triangle correspondant dans la forme moyenne en utilisant ses coordonnées barycentriques [256]. Donc, la déformation de la texture consiste à trouver une fonction de correspondance f qui permet d'associer l'ensemble des points $\{X_i\}$ représentant la première forme X à l'ensemble des points $\{X'_i\}$ représentant la forme X' , telle que :

$$f(X_i) = X'_i \quad \forall i = 1..n$$

Une fois cette fonction est déterminée, nous pouvons projeter chaque pixel de l'image I dans une nouvelle image I' . Dans la pratique, pour éviter les trous et les problèmes d'interpolation, il vaut mieux de trouver f' qui prend x' dans x . Autrement dit: pour chaque pixel de l'image déformée I' , nous déterminons d'où il provient de I et le mettre. En général $f' \neq f^{-1}$, mais c'est une bonne approximation.

Une méthode simple, définie dans [256], est de supposer que la fonction f est localement linéaire. Après avoir effectué une triangulation de Delaunay afin de diviser l'enveloppe convexe des points de repère de la forme en un ensemble de triangles. Aux points de chaque triangle I , nous appliquons une transformation affine qui fait correspondre les sommets de ce triangle à leur nouvelles position dans I' .

Si x_1, x_2, x_3 sont les trois sommets du triangle, tout point x qui se situe à son intérieur peut s'écrire comme suit:

$$x = x_1 + \beta(x_2 - x_1) + \gamma(x_3 - x_1) = \alpha x_1 + \beta x_2 + \gamma x_3 \quad (5.6)$$

Où $\alpha = 1 - (\beta + \gamma)$ et $\alpha + \beta + \gamma = 1$. Pour que x appartienne au triangle, $0 \leq \alpha, \beta, \gamma \leq 1$.

Sous une transformation affine, ce point correspond à:

$$x' = f(x) = \alpha x'_1 + \beta x'_2 + \gamma x'_3 \quad (5.7)$$

Pour déformer une image, pour chaque pixel x' dans I' , son triangle englobant est déterminé, puis ses coordonnées barycentriques α, β et γ , qui donnent sa position relative dans le triangle, sont calculées et utilisées pour trouver le point équivalent dans l'image originale I , après une interpolation bilinéaire, la valeur de ce point est copier dans le pixel x' de I' . La figure 5.8 montre un exemple de déformation de texture en utilisant une triangulation de Delaunay.

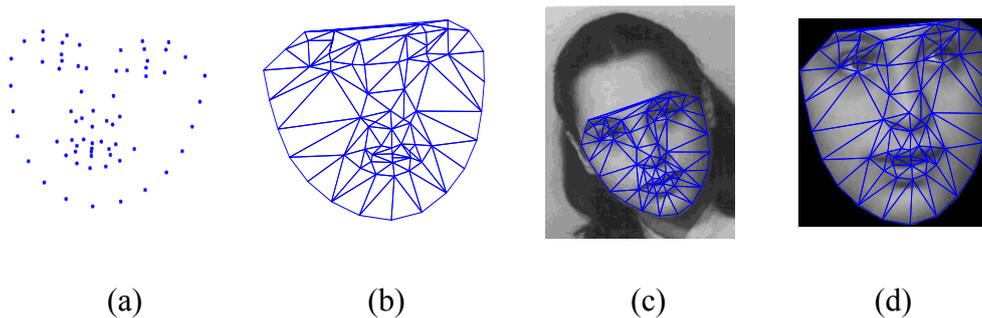


Figure 5.8- Exemple de déformation de texture. (A) La forme moyenne, (b) la triangulation de Delaunay de la forme moyenne, (c) l'image actuelle et (d) la texture déformée [253].

5.2.3 Processus de normalisation

La normalisation des visages est constituée des différentes étapes suivantes:

1. Convertir toutes les images en niveaux de gris.
2. Aligner toutes les formes associées aux images de la base d'apprentissage et transférer les textures associées à chaque forme dans la forme moyenne en utilisant la méthode décrite ci-dessus.
3. Pivoter toutes les images pour obtenir les yeux à des points fixes dans l'image de telle sorte que le segment intraoculaire soit horizontal, en se basant sur les fichiers de coordonnées des yeux fournis avec la base de données FG-NET originale.

4. Redimensionner toutes les images à la taille de 128x128. La figure 5.9 montre les images des visages de la figure 1.1 après normalisation



Figure 5.9- images Normalisées d'une personne à différents âges [253].

5.3 Extraction de caractéristiques

Afin d'obtenir le vecteur descripteur d'une image prétraitée en utilisant l'approche décrite dans la section 5.2, nous appliquons une transformée DCT en 2D, et nous conservons un sous-ensemble des coefficients. Le nombre de coefficient sélectionné est choisi de telle sorte qu'il peut représenter un visage, notre but n'est plus la réduction de la dimension des données, mais d'écarter les coefficients qui représentent trop de détails de texture. Les caractéristiques sélectionnées qui construisent le vecteur de caractéristique sont représentées comme des points dans un espace de grande dimension; une analyse de Fisher à noyau (KFA Kernel Fisher Analysis) est effectuée pour réduire la dimension. La description de la KFA sera donnée en section 5.5. Les détails de la transformée DCT seront présentés dans la section suivante:

5.4 La transformée en cosinus discrète

5.4.1 Principe

La transformée en cosinus discrète (DCT) transforme une image du domaine spatial dans le domaine fréquentiel. Divers classes de DCT ont été proposées [261]. En particulier, la DCT a été classée [261] en quatre classes nommées DCT-I, DCT -II, DCT-III et DCT -IV. La DCT-II est la variante la plus courante de DCT appliquée dans le codage de signal, en particulier dans la compression, car elle était l'idée principale de

la compression jpeg [262]. La DCT bidimensionnelle pour une image d'entrée A de dimension MxN est définie comme suit:

$$B(p, q) = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{m, n} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N}, \quad 0 \leq p \leq M-1, \quad 0 \leq q \leq N-1 \quad (5.8)$$

Et la transformée inverse est définie comme

$$A(m, n) = \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} \alpha_p \alpha_q B_{p, q} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N}, \quad (5.9)$$

Où

$$\alpha_p = \begin{cases} \frac{1}{\sqrt{M}}, & p = 0 \\ \sqrt{\frac{2}{M}}, & 1 \leq p \leq M-1 \end{cases} \quad (5.10)$$

Et

$$\alpha_q = \begin{cases} \frac{1}{\sqrt{N}}, & q = 0 \\ \sqrt{\frac{2}{N}}, & 1 \leq q \leq N-1 \end{cases} \quad (5.11)$$

M et N représentent le nombre de lignes et de colonnes de A.

Contrairement à la norme de compression JPEG, où l'image est d'abord divisée en blocs non chevauchés (blocs 8x8) et la DCT est effectuée indépendamment sur des blocs d'images [262], dans notre approche, la DCT est calculée sur la totalité de l'image de visage. Pour construire le vecteur descripteur, les coefficients de la DCT sont disposés d'une manière de balayage en zigzag, de telle sorte de placer les coefficients des basses fréquences au sommet du vecteur (voir Figure 5.10).

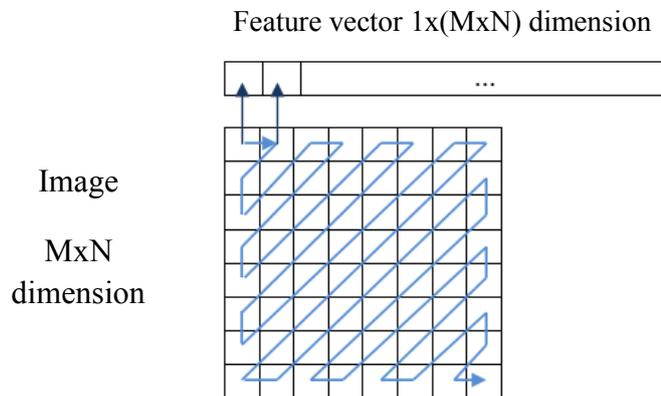


Figure 5.10- Formation du vecteur des caractéristiques dans le domaine de la DCT.

5.4.2 Choix de la DCT

La DCT possède une caractéristique très importante, elle permet de séparer l'image en plusieurs parties d'importance inégale. Une fois l'image originale a été transformée, les coefficients de la DCT reflètent l'importance des fréquences qui sont présents dedans. Le DC-coefficient (le premier coefficient) montre l'illumination totale de l'image, les coefficients des basses fréquences de la DCT sont fortement liés à la variation d'éclairage, et les coefficients des fréquences élevés montrent des détails et des informations bien éventuellement avoir été causé par le bruit. Les coefficients situés entre le premier et les derniers coefficients présentent différents niveaux d'information de l'image originale [263].

Notre but est d'obtenir des caractéristiques qui sont invariantes à la progression d'âge, nous savons que:

- En général, les processus de croissance et de vieillissement du visage se composent des variations de la forme du visage et de la texture de la peau.
- L'apparition d'une surface de texture dépend de l'éclairage ainsi qu'elle change dans différentes conditions d'éclairage.
- Les variations d'éclairage se trouvent dans la bande des basses fréquences.
- En règle générale, pour reconnaître une personne, nous recherchons les principales caractéristiques du visage comme les formes et les relations géométriques des principales composantes du visage y compris les yeux, le nez et la bouche, et nous ignorons presque certains détails de texture de la peau.
- A titre d'illustration, la première rangée (a) sur la Figure 5.11 affiche quatre images du visage de la même personne à différents âges (24, 31, 42 et 61 ans) après normalisation, la deuxième rangée (b) montre les images reconstruites de la même personne après l'application de la transformée DCT, en ne conservant qu'un petit ensemble de coefficients de basses fréquences et application de la DCT inverse. La troisième rangée (c) montre des images reconstruites de la même personne après l'application de la transformation DCT, la mise à zéro d'un petit ensemble de coefficients de basse fréquence et de l'application de la DCT inverse. De la deuxième et la troisième ligne, nous pouvons voir que les images reconstruites en rejetant un ensemble de coefficients de basse fréquence présentent une diminution

dans certains détails, principalement de la texture, mais des caractéristiques importantes qui identifient le visage comme les yeux et le nez sont conservées. A partir de ces observations, on peut conclure que l'utilisation de la transformation DCT en ignorant certains coefficients de basses fréquences est cruciale pour l'extraction de caractéristiques avec succès pour la reconnaissance faciale à travers la progression d'âge.

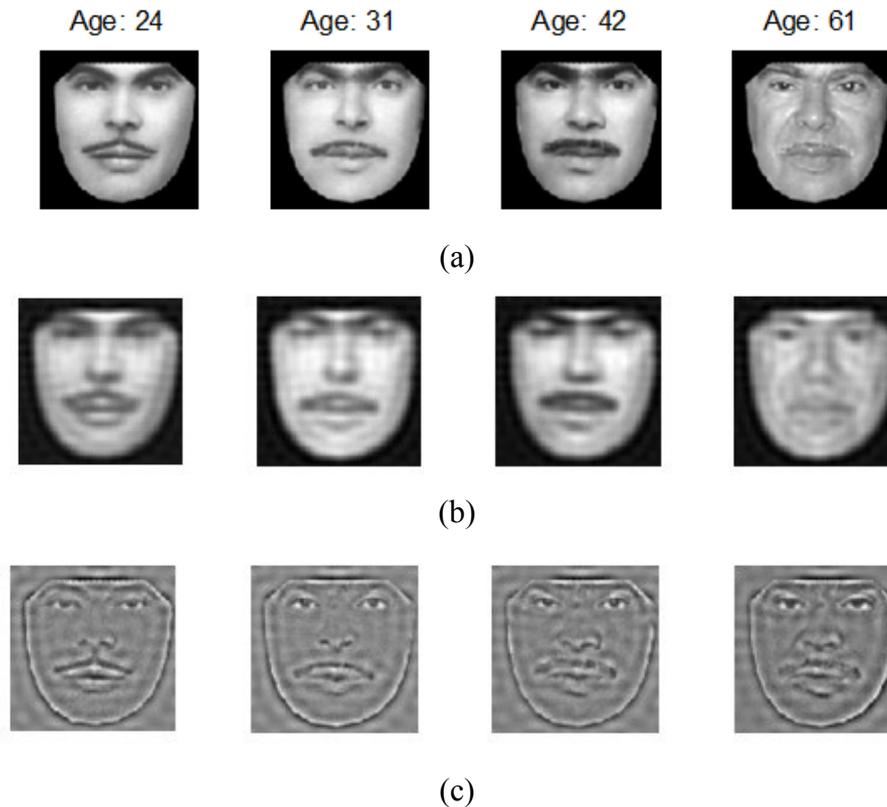


Figure 5.11- Images normalisées d'une personne à des âges différents et leurs images reconstruites [253].

5.5 Réduction de la dimension du descripteur

Cette étape est considérée comme une phase importante dans le processus d'extraction des descripteurs, car elle permet de réduire l'effet de la malédiction de la dimension et d'éviter certaines propriétés indésirables des espaces de grande dimension [267]. Plusieurs méthodes qui ont été utilisées, l'ACP et l'ADL sont parmi les méthodes les plus utilisées dans la reconnaissance de visage. D'après les résultats expérimentaux fournis par [267], les méthodes ACP et ADL appliquées sur des coefficients d'une DCT ont donné des taux de reconnaissance remarquables pour la reconnaissance faciale. Pour

traiter des données complexes qui sont non linéaires, des versions non linéaires ont été proposées comme alternative à ces méthodes linéaire. En se basant sur l'étude [166] ou une comparaison entre les méthodes ACP, ADL, ACP et ADL à noyaux, ICA pour la vérification de visage. Leurs résultats expérimentaux montrent que la méthode ADL à noyau est capable d'extraire des caractéristiques non linéaires et d'obtenir des taux d'erreur plus faibles, nous avons choisi cette méthode pour la réduction de dimension.

5.5.1 Analyse de Fisher à noyau

L'analyse discriminante de Fisher à noyau, ou la méthode KFA, est une version d'analyse discriminante linéaire avec noyau. Elle a été appliquée avec succès à la reconnaissance biométrique telle que les empreintes palmaires [264], l'identification et la vérification du visage [166, 265, 266].

Dans cette approche, les données d'entrée sont projetées à partir de l'espace d'entrée, R^n , dans un espace implicite de grande dimension, R^f , appelé espace caractéristique par une fonction non linéaire (à noyau): $\Phi: R^n \rightarrow R^f$, $f > n$, puis, une analyse discriminante linéaire est adopté à cette espace de caractéristiques. Les noyaux fréquemment utilisés sont le noyau polynomial de degré K et le noyau gaussien.

Nous utilisons la même terminologie et le même algorithme décrit dans [166] pour effectuer l'analyse KFA. En supposant que les échantillons projetés $\Phi(X_i)$ sont centrés dans R^f , avec toute l'information de tous les échantillons et leurs classes, la matrice de dispersion entre classe S_B^Φ et la matrice de dispersion au sein de la classe S_W^Φ sont définis comme suit :

$$S_B^\Phi = \sum_{i=1}^c l_i \mu_i^\Phi (\mu_i^\Phi)^T \quad (5.12)$$

$$S_W^\Phi = \sum_{i=1}^c \sum_{j=1}^{l_i} \Phi(X_{ij})\Phi(X_{ij})^T \quad (5.13)$$

Où l_i est le nombre d'échantillons de la classe i , c est le nombre de classes, μ_i^Φ est la moyenne d'une classe i dans R^f .

Pour appliquer une LDA dans un espace à noyau, nous avons besoin des valeurs propres λ et des vecteurs propres W^Φ de:

$$\lambda S_W^\Phi W^\Phi = S_B^\Phi W^\Phi \quad (5.14)$$

Qui peuvent être obtenus en cherchant un ensemble de vecteurs W_{OPT}^Φ qui est obtenu par:

$$W_{OPT}^\Phi = \arg \max_{W^\Phi} \left| \frac{(W^\Phi)^T . S_B^\Phi . W^\Phi}{(W^\Phi)^T . S_W^\Phi . W^\Phi} \right| = [W_1^\Phi \dots W_m^\Phi] \quad (5.15)$$

Où $\{W_i^\Phi, i = 1, 2, \dots, m\}$ sont les vecteurs propres qui correspondent aux m plus grandes valeurs propres $\{\lambda_i, i = 1, 2, \dots, m\}$.

Considérons c est le nombre de classes, le $r^{i\text{ème}}$ échantillon d'une classe t et le $s^{i\text{ème}}$ échantillon d'une classe u sont respectivement X_{tr} et X_{us} , et la classe t a l_t échantillon et la classe u a l_u échantillon, la fonction du noyau est définie comme suit :

$$(k_{rs})_{tu} = k(X_{tr}, X_{us}) = \Phi(X_{tr}) . \Phi(X_{us}) \quad (5.16)$$

Supposons que K est une matrice symétrique de dimension $m \times m$ définie par les éléments: $(k_{tu})_{u=1, \dots, c}^{t=1, \dots, c}$, où k_{tu} est une matrice composée du produit point $l_t \times l_u$ dans l'espace de caractéristiques \mathcal{R}^f , i.e.,

$$K = (k_{tu})_{u=1, \dots, c}^{t=1, \dots, c}, \text{ où } k_{tu} = (k_{rs})_{s=1, \dots, l_u}^{r=1, \dots, l_t} \quad (5.17)$$

Nous pouvons aussi définir Z , telle que Z est une matrice diagonale à bloc de dimension $m \times m$, chaque bloc (Z_t) est une matrice $l_t \times l_t$ avec des éléments tous égaux à $\frac{1}{l_t}$:

$$Z = (Z_t)_{t=1, \dots, c} \quad (5.18)$$

Toute solution $W^\Phi \in \mathcal{R}^f$ doit se situer dans la portée de tous les échantillons d'apprentissage \mathcal{R}^f , i.e.

$$W^\Phi = \sum_{p=1}^c \sum_{q=1}^{l_p} \alpha_{pq} \Phi(X_{pq}) \quad (5.19)$$

Il en résulte que la solution de (5.19) peut être obtenue en résolvant l'équation:

$$\lambda KK\alpha = KZK\alpha \quad (5.20)$$

Par conséquent, l'équation (5.15) peut être écrite comme

$$W_{OPT}^{\Phi} = \arg \max_{W^{\Phi}} \left| \frac{(W^{\Phi})^T \cdot S_B^{\Phi} \cdot W^{\Phi}}{(W^{\Phi})^T \cdot S_W^{\Phi} \cdot W^{\Phi}} \right| = \arg \max_{W^{\Phi}} \left| \frac{\alpha K Z K \alpha}{\alpha K K \alpha} \right| = [W_1^{\Phi} \dots W_m^{\Phi}] \quad (5.21)$$

Par la suite la projection des données dans le nouvel espace se fait suivant l'équation:

$$W^{\Phi} \cdot \Phi(X) = \sum_{i=1}^m \alpha_i (\Phi(X_i) \cdot \Phi(X)) = \sum_{i=1}^m \alpha_i k(X_i, X) \quad (5.22)$$

Pour éviter des singularités dans le calcul de W_{OPT}^{Φ} , nous utilisons la même technique dans Fisherfaces [155].

5.6 Reconnaissance (identification)

Comme nous avons déjà mentionné, au chapitre 3, cette étape permet de classer une image de visage et d'identifier son identité. Les approches utilisées sont issues du domaine de la classification automatique, et elles dépendent principalement des descripteurs choisis à l'étape précédente.

Vu que le descripteur choisi dans cette étude est basé sur une KFA appliquée sur des coefficients d'une DCT. Et en prenant en considération l'étude [268] faite par Jonsson et al., qui a démontré que, dans un contexte de reconnaissance de visage, les SVM sont capables d'extraire des informations discriminantes à partir des données d'apprentissage. Cependant, lorsqu'une ADL est déjà utilisée à l'étape précédente, le SVM n'a apporté aucun résultat car l'ADL a séparé correctement les différentes classes dans l'espace de représentation et l'étape de maximisation de la marge devenant ainsi inutile [268].

La méthode de classification du plus proche voisin a été choisie pour la reconnaissance de visage. C'est une méthode simple et efficace, aussi appelée: algorithme de classification paresseux, où il n'y a pas de traitement à faire à l'étape d'apprentissage, et tous le travail est effectué au cours de l'étape de test. La classification est réalisée en comparant le descripteur de l'image requête avec tous les images de la base d'apprentissage et en cherchant le visage de descripteur de la base de données qui a la distance la plus proche. Plusieurs mesures de distance peuvent être utilisées pour le classement du plus proche voisin, comme la norme L1, L2 et la corrélation normalisée.

Le choix de distance est très important et les performances de la classification peuvent varier considérablement en fonction de la distance utilisée. La distance utilisée dans l'approche proposée est la distance euclidienne.

5.7 Conclusion

Comme nous avons déjà mentionné, les méthodes qui sont destinées à être utilisées dans un contexte d'invariance à la progression d'âge sont classées en approches génératives ou approches discriminatives, notre choix s'est fixé sur les approches de la seconde classe, vu leur stabilité et leur simplicité par rapport aux autres approches. Dans ce chapitre nous avons exposé les différentes étapes qui constituent notre méthode de description et d'identification faciale proposée en justifiant notre choix. L'évaluation des performances de la méthode fait l'objectif du chapitre qui suit.

CHAPITRE 6:

Expérimentations, évaluations et discussions

6.1 Introduction	108
6.2 Choix de la base de données	108
6.3 Description de l'expérimentation	110
6.4 Résultats expérimentaux	110
6.4.1 Importance de l'étape de normalisation	111
6.4.2 Importance d'utiliser la DCT	112
6.4.3 Importance d'utiliser la DCT et fixer un petit nombre des premiers coefficients à zéro	112
6.4.4 Choix du pourcentage de coefficients de basses fréquences à être ignorés	114
6.4.5 Intérêt d'utiliser la KFA	115
6.4.6 Résultats obtenus avec certaines approches proposées dans le domaine de la reconnaissance faciale invariante à l'âge	116
6.5 Influence de la différence d'âge entre l'image requête et celle de référence sur les performances de la reconnaissance	118
6.6 Effets de croissance et de vieillissement sur les performances de la reconnaissance de visage	119
6.7 Conclusion	120

CHAPITRE 6

Expérimentations, évaluations et discussions

Nous décrivons dans ce chapitre les différentes expérimentations réalisées au cours de cette thèse pour évaluer les performances de la méthode de reconnaissance de visage invariante à la progression d'âge décrite dans le chapitre 5. Après avoir présenté la base de données des images utilisées et décrit les expérimentations réalisées dans cette étude, nous présenterons les résultats expérimentaux concernant l'évaluation de l'approche de description proposée, qui est basée, principalement sur les AAM, la DCT, et la KFA.

6.1 Introduction

L'évaluation d'un système de reconnaissance faciale, ou plus généralement d'un système de recherche d'images par le contenu est une étape cruciale pour sa validation et son utilisation dans des applications réelles. L'évaluation concerne à la fois sa capacité de reconnaissance (c.-à-d. Est-ce que le résultat fourni par le système satisfait les besoins de l'utilisateur?) et son efficacité (c.-à-d. sa rapidité). Dans cette étude, nous nous concentrons sur le pouvoir de reconnaissance, où nous allons évaluer la robustesse de la technique proposée en présence des variations qui sont dus à la croissance ou le vieillissement du visage.

6.2 Choix de la base de données

Les performances des algorithmes de reconnaissance faciales sont très dépendantes de la base de données utilisée pour leur évaluation. De nombreuses bases de données qui existent dans le domaine de la reconnaissance faciale, chacune ayant des caractéristiques différentes et chacune convient à étudier une des difficultés rencontrées par les systèmes de reconnaissance dans les situations réelles, telles que la pose, l'illumination, et les occultations, etc. Pour les études concernant les variations d'âge, à l'heure où nous avons rédigé cette thèse, il existe deux bases de données qui sont fournies gratuitement, à savoir, les bases de données FG-NET et Morph. Pour l'étude expérimentale effectuée dans cette thèse, nous avons utilisé la base de données FG-NET (Face and Gesture Recognition Research Network) [148], qui est bien connu dans le

domaine de la reconnaissance faciale invariante à la progression d'âge. La base de données est utilisée dans la majorité des études relatives à l'âge dans le domaine de la reconnaissance de forme et de la vision par ordinateur [269], que se soit de l'estimation de l'âge à partir du visage, de la modélisation du processus du vieillissement du visage ou de la reconnaissance de visage invariante à l'âge. C'est une collection d'images publiques contenant des images de visage montrant un certain nombre de personnes à différents âges. La base de données a été développée et fournie gratuitement comme une aide pour les chercheurs qui étudient les effets du vieillissement sur l'apparence du visage et leurs conséquences sur les performances des systèmes de reconnaissance faciale. La base de données contient 1002 images provenant de 82 personnes différentes avec des âges allant de la naissance à 69 ans. Des images typiques de la base de données sont représentées dans la figure 6.1. Les images de la base de données ont été recueillies par numérisation de photos d'individus trouvées dans des collections personnelles, par conséquent, la qualité des images de la base de données dépendent de l'état général des photos utilisées, aussi ces images présentent une grande variabilité dans la résolution, l'illumination, la position du visage et les occlusions comme le port de lunettes ou autre accessoire sont également présents dans un certain nombre d'images. Des fichiers de données contenant les emplacements de 68 points de repère du visage qui ont été identifiés manuellement, et l'âge de la personne dans chaque image sont, également, disponibles.

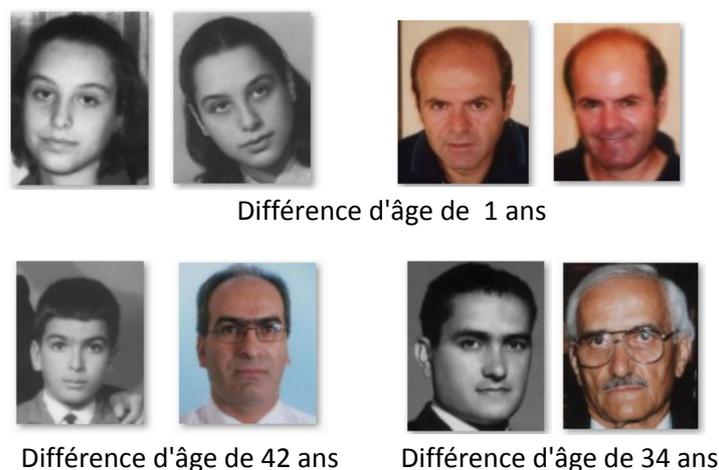


Figure 6.1- Exemples d'images d'individus avec des différences d'âge. Les images sont de la base d'images FG-NET [148].

6.3 Description de l'expérimentation

Tous les algorithmes et les étapes de normalisation ont été implémentés dans MATLAB (R 2012a). Pour l'évaluation de notre approche, nous adoptons le schéma LOPO (leave one person out) qui est le schéma d'évaluation souvent utilisé dans la plupart des travaux qui utilisent la base de données FG-NET pour l'évaluation [269] et cela dans le but de faciliter la comparaison avec d'autres résultats obtenus sur cette base de données. Cette stratégie sélectionne une image d'une personne pour le test, en utilisant toutes les images restantes comme un ensemble d'apprentissage. Cette opération est répétée pour toutes les personnes dans la base de données.

Pour chaque image de la base d'apprentissage, nous calculons son descripteur en suivant l'approche de description déjà présentée au chapitre 5, où le nombre de coefficients de basses fréquences à mettre à zéro est choisi empiriquement. Même le choix du noyau et de ces paramètres est fait empiriquement, dans cette étude, nous utilisons le noyau polynomial de degré deux. A l'étape de reconnaissance, nous utilisons une classification par K-plus proches voisins, où K est choisi égal à un ($K = 1$) pour calculer les taux de reconnaissance au Rank-1.

6.4 Résultats expérimentaux

Comme nous avons déjà mentionné au chapitre 3, un algorithme de reconnaissance peut être évalué selon deux modes: identification ou vérification. En mode identification, les résultats de nos expérimentations sont présentés de deux façons:

- Un tableau montrant les taux d'identification au rang un (Rank-one recognition rate)
- Une courbe de correspondance (CMC pour cumulative match characteristic curve) [25], montrant les taux de reconnaissance au rang un et supérieur.

En cas de vérification, les résultats sont présentés par un tableau contenant les taux d'erreurs égales (EER).

Les résultats présentés sont la moyenne de plusieurs résultats obtenus à partir d'expériences suivant le schéma LOPO.

Les mesures de performance utilisées sont définies au chapitre 3 et sont calculées en utilisant la boîte à outils PhD (Pretty Helpful Development) créée dans MATLAB pour

la reconnaissance faciale [270, 271] et qui peut être téléchargée du site web Mathworks¹.

Pour tester l'approche proposée et évaluer ces performances, nous effectuons trois expérimentations et cela dans le but de :

- Tester l'intérêt de la méthode de normalisation
- Tester l'utilité de la transformée en cosinus pour la reconnaissance faciale invariante à la progression d'âge en mettant à zéro un pourcentage des coefficients de basses fréquences
- Tester l'importance de la méthode de réduction non linéaire utilisée.

Les différentes expérimentations sont décrites ci-dessous:

6.4.1 Importance de l'étape de normalisation

Dans cette première expérimentation, nous étudions l'utilité d'utiliser la méthode de correction de la pose en se basant sur l'approche décrite au chapitre 5, nous comparons les résultats obtenus avec ceux obtenus en utilisant l'approche suivante:

Premièrement, les images sont tournées pour obtenir les yeux à des points fixes dans l'image de telle sorte que le segment interoculaire est horizontal, cela est réalisé en se basant sur les fichiers contenant les coordonnées des yeux fournis avec la base de données FG-NET, ensuite elles sont coupées (en utilisant toujours les coordonnées des yeux) pour éliminer l'arrière-plan et redimensionnée pour avoir la taille de 128x128, cette approche est appelée dans ce qui suit alignement manuel (AM) (voir figure 6.2).

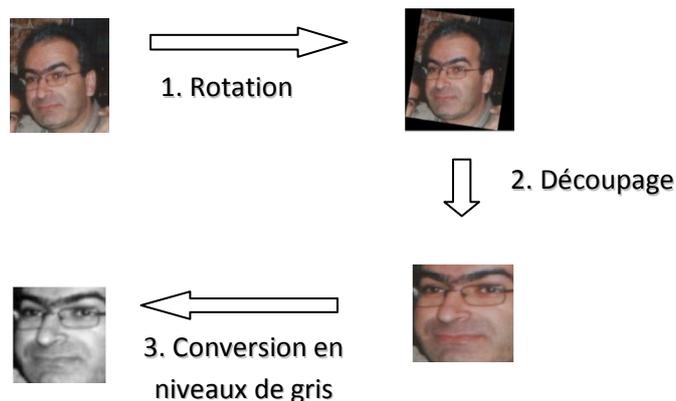


Figure 6.2- Alignement manuel (AM)

¹ <https://www.mathworks.com/matlabcentral/fileexchange/35106-the-phd-face-recognition-toolbox>

Les résultats obtenus sont présentés dans le tableau 6.1, une analyse de ces derniers montre clairement l'avantage d'utiliser l'approche proposée comme méthode de correction de pose, où il y a une amélioration de 19,15% pour le taux d'identification et une diminution dans le taux d'EER de 22,68 à 7,2% pour la vérification dans le cas où nous utilisons la DCT, et il y a une amélioration de 23,14% pour le taux d'identification, et une diminution dans le taux d'EER de 37,70% à 17,23%, sans utiliser la DCT.

6.4.2 Importance d'utiliser la DCT

Dans cette deuxième expérimentation, nous étudions l'importance d'utiliser la transformée DCT. Tout d'abord, nous comparons les résultats obtenus en appliquant la DCT sur les images alignées manuellement avec ceux sans appliquer la DCT. Deuxièmement, nous comparons les résultats combinant AAM et DCT avec ceux utilisant seulement AAM. D'après les résultats obtenus (voir tableau 6.1), nous pouvons observer que l'ajout de la DCT augmente le taux de reconnaissance par 14,22% et diminue le taux d'EER par 15,02% pour la vérification dans le premier cas, de même pour le second cas, le taux de reconnaissance est augmenté de 10,23% et le taux d'EER est diminué de 10,03% pour la vérification. Cela conduit à la conclusion que l'utilisation de la DCT fournit des informations utiles pour franchir la variance causée par les variations d'âge dans les systèmes de reconnaissance faciale.

6.4.3 Importance d'utiliser la DCT et fixer un petit nombre des premiers coefficients à zéro

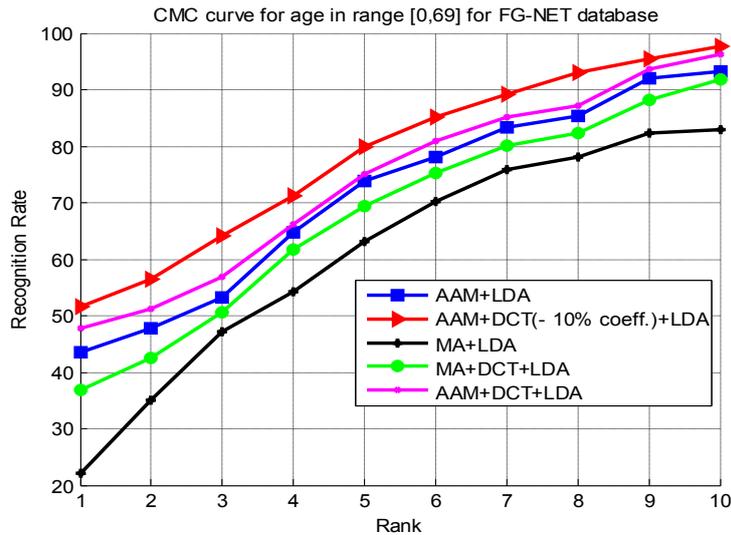
Dans la troisième expérimentation, nous évaluons les performances de la transformation en cosinus discrète, en présence des variations d'âge, nous comparons les résultats obtenus en effectuant une DCT sur les images prétraitées et en fixant un petit nombre des premiers coefficients à zéro avec ceux obtenus sans écarter des coefficients. A partir des résultats fournis dans le tableau 6.1, nous pouvons observer que l'application de la DCT en ignorant un certain nombre de coefficients a atteint un taux de 61,41% dans le cas de l'identification et le taux d'EER de 7,2% dans le cas de la vérification ce qui est équivalent à une amélioration de 7,85% pour l'identification et 6,03% pour la vérification, ce qui démontre explicitement l'impact des coefficients de basse fréquence sur les performances de la reconnaissance du visage en présence des variations d'âge.

Tableau 6.1. Taux de reconnaissance au rang-1 (identification) et taux d'erreurs égales (vérification) .

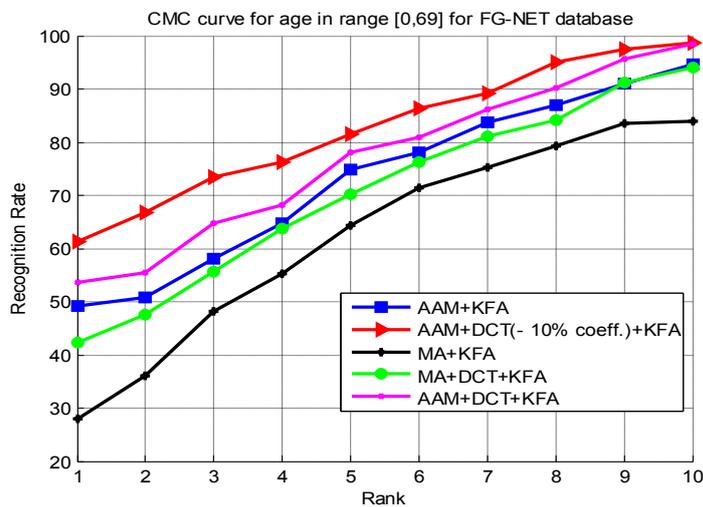
Méthode	Taux d'identification au rang-1		Taux d'erreurs égales	
	LDA	KFA	LDA	KFA
MA	22.16%	28.04%	43.29%	37.70%
MA+DCT (-10% de coeff.)	36.85%	42.26%	27.28%	22.68%
AAM	46.51%	51.18%	19.45%	17.23%
AAM+DCT	47.70%	53.56%	16.25%	13.23%
AAM+DCT (-10% de coeff.)	51.58%	61.41%	13.85%	7.2%

D'après les résultats présentés dans le tableau 6.1, nous pouvons, encore, constater que dans le cas de la présence des variances temporelles, la transformation DCT en ignorant certains coefficients de basse fréquence, lorsqu'elle est associée à la méthode de normalisation basée sur les AAM et celle de projection en utilisant la KFA produit la meilleure performance parmi tous les cas étudiés.

La Figure 6.3 démontre bien l'intérêt de l'approche proposée pour la reconnaissance de visage invariante à la progression d'âge, où l'approche proposée fournit le meilleur résultat au rang un, et le seul meilleur résultat reste toujours réservé à des rangs supérieurs. On peut voir également, l'importance de l'étape de prétraitement où les combinaisons dans lesquelles nous utilisons la méthode AAM (rouge, magenta et courbe bleue dans la figure 6.3) donnent toujours les meilleurs résultats.



(a) La courbe CMC en utilisant la LDA



(b) La courbe CMC en utilisant la KFA

Figure 6.3- Courbe CMC pour les expérimentations faites sur la base de données FG-NET.

6.4.4 Choix du pourcentage de coefficients de basses fréquences à être ignorés

Pour choisir le meilleur pourcentage de coefficients de basses fréquences à être ignorés, nous procédons à une succession d'expériences, dans chaque expérience, nous avons mis à zéro un pourcentage de coefficients de basses fréquences (0%, 5%, 10%, 15%, ..., 90%), le meilleur taux de reconnaissance est atteint entre 10% et 15% (voir Figure 6.4), et ce dernier est diminué lorsque nous avons choisi des pourcentages élevés, tel est le cas dans la dernière expérience où le taux de reconnaissance obtenu était de 0,6 %.

De toute évidence, il existe une perte importante d'information, nous pouvons voir cette perte d'information sur la Figure 6.5. La figure 6.5 (a) représente l'image originale, la Figure 6.5 (b, c, d, e) représentent respectivement les images reconstruites après avoir éliminé 5%, 10%, 30% et 85% des coefficients de basse fréquence. La dernière image (e) ne contient presque pas d'informations.

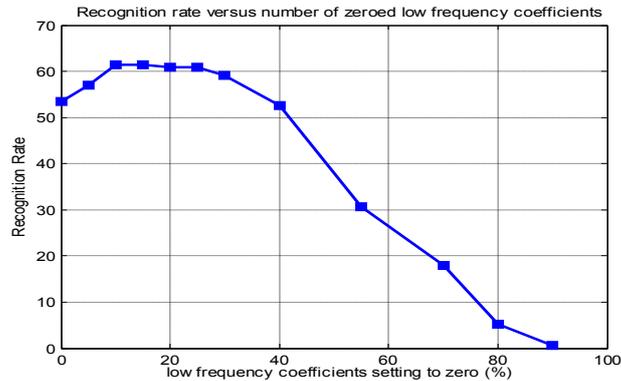


Figure 6.4 Taux de reconnaissance versus nombre de coefficients de basse fréquence mis à zéro.

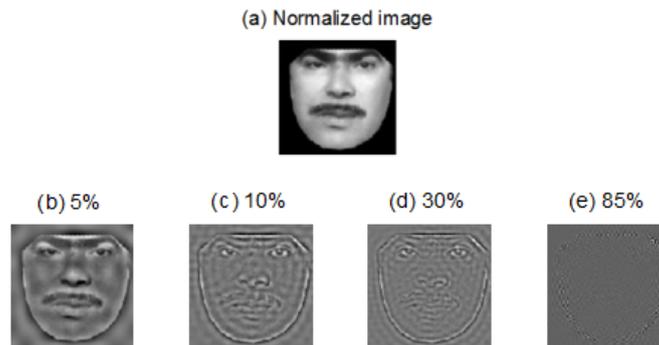


Figure 6.5- Image normalisée d'une personne et les images reconstruites en mettant à zéro un pourcentage de coefficients de basse fréquence.

6.4.5 Intérêt d'utiliser la KFA

En outre, nous démontrons l'importance de l'utilisation de la méthode à noyau pour la réduction de la dimensionnalité, à cet effet, nous répétons les trois expériences en utilisant l'analyse discriminante linéaire (LDA) au lieu d'utiliser l'analyse de Fisher à noyau. D'après les résultats (voir le tableau 6.1), les taux obtenus avec la KFA sont toujours plus élevés que ceux obtenus à l'aide de l'analyse discriminante linéaire (LDA).

6.4.6 Résultats obtenus avec certaines approches proposées dans le domaine de la reconnaissance faciale invariante à l'âge

Les résultats obtenus par notre approche sont comparés par ceux obtenus par certaines approches proposées dans le domaine de la reconnaissance faciale invariante à l'âge. Dans le tableau 6.2, nous résumons le principe sur lequel est basé chaque approche, la base de données utilisée pour l'évaluation ainsi que le taux de reconnaissance obtenu.

Table 6.2. Comparaison des performances pour la reconnaissance de visage invariante à la progression d'âge.

Auteur	Approche	Base de données	Nombre d'images	Nombre d'individus	Taux de reconnaissance au rang-1 (%)
Lanitis et al. (2002) [220]	Construction d'une fonction de vieillissement en termes de coefficients d'une ACP appliquée sur les vecteurs de forme et de texture	Base de données privée	85	12	68.5 %
Ramanathan et al. (2006) [225]	Modélisation de la croissance de la forme du visage jusqu'à l'âge de 18 ans	Base de données privée	109	109	15 %
Geng et al. (2007) [216]	Apprentissage d'un modèle de vieillissement en se basant sur les coefficients concaténés d'une ACP appliquée sur les vecteurs de forme et de texture	FG-NET	10	10	38.1 %
Park et al. (2008 - 2010) [273,228]	Apprentissage d'un modèle de vieillissement en se basant sur les coefficients d'une ACP appliquée dans l'espace de texture et de forme en 3D à partir d'une base d'images en 2D.	FG-NET	1002	82	37.4 %
		MORPH I	612	612	66.4 %
		MORPH II	20000	10000	79.8 %
		Browns	100	100	28.1 %

Juefei-Xu et al. (2011) [251]	Une combinaison de la transformée de Walsh-Hadamard et des motifs binaires locaux appliquée sur les régions péri-oculaire, suivie d'une projection discriminante non supervisée UDP, pour la modélisation des sous espaces.	FG-NET	1002	82	100%
Yang et al. (2014) [244]	Image de visage convertie en un graphe dont les noeuds représentent la texture des régions chevauchées et les arêtes la topologie de ces régions. La texture est représentée par un descripteur nommé LGBPMS et projeté dans le sous espace créée par l'ADL.	FG-NET	1002	82	64.47 %
Li et al. (2011) [249]	Les descripteurs SIFT et MLBP sont combinés à des classificateurs basés sur une ADL comme descripteurs locaux.	FG-NET	1002	82	47.5 %
		MORPH II	20000	10000	83.9 %
Bereta et al. (2013) [248]	Divers descripteurs locaux de texture, combinés par des classifieurs basés sur différentes distances	FG-NET	1002	82	< 45 %
L'approche proposée [253]	DCT appliquée à des images prétraitées en utilisant les AAM, et projetées dans les sous espaces créés par la KFA	FG-NET	1002	82	61.41%

La plupart des résultats inscrits dans le tableau 6.2 ne peuvent pas être comparés, car pour pouvoir comparer deux approches et dire qu'un résultat est meilleur que l'autre, on doit suivre les mêmes conditions pour réaliser les évaluations, c.-à-d. la même base de données et le même protocole de séparation entre les images d'apprentissage et celles de test.

D'après les résultats inscrits dans le tableau 6.2, le meilleur taux de reconnaissance est obtenu par Juefei-Xu et al [251], ce qui indique que l'utilisation de l'information locale et plus particulièrement la région péri-oculaire est bien désirable pour la reconnaissance faciale invariante à la progression d'âge, mais il faut tenir en compte que la reconnaissance faciale basée sur cette région pourrait échouer car dans des situations réelles, cette région utilisée seule peut être affectée par les problèmes d'expression tel que la fermeture totale des yeux ou la présence des occlusions tel que le port de lunette.

6.5 Influence de la différence d'âge entre l'image requête et celle de référence sur les performances de la reconnaissance

Dans cette partie, nous allons étudier l'impact de la différence d'âge entre l'image requête et celles de référence sur les performances de l'approche proposée [272]. Pour cela, en suivant le même schéma utilisé dans [248], nous choisissons une image d'un visage de chaque classe (individu) qui a à peu près 18 ans. La base d'images contient 82 individus, ce qui résulte en 82 images utilisées comme base de test et le reste 920 sont utilisées comme base d'images, ces dernières sont groupées en 9 groupes, de telle sorte que les différences d'âge entre les images de la base d'images et celles de la base de test appartiennent aux intervalles suivants:

0 à 1 ans, 2 à 3 ans, 4 à 5 ans, 6 à 10 ans, 11 à 15 ans, 16 à 20 ans, 21 à 30 ans, 30 à 40 ans et supérieur ou égale à 41 ans.

D'après les résultats mentionnés dans le tableau 6.3, nous observons que les meilleurs taux de reconnaissance sont obtenus lorsque la différence d'âge est de 0 à 3 ans, de même les taux de reconnaissance deviennent très faibles lorsque la différence est supérieure à 20 ans, cela peut être expliqué par le fait que deux images qui sont séparées par un laps de temps d'environ 3 à 6 mois peuvent être tout à fait similaires, mais deux images qui sont séparées par une durée de 30 à 40 ans sont significativement

différentes, en particulier, lorsque la première image est prise à l'âge de l'enfance (<18 ans) et la seconde à l'âge adulte (exemple 62 ans) .

Tableau 6.3. Taux d'identification versus différence d'âge.

Différence d'âges	0-1	2-3	4-5	6-10	11-15	16-20	20-30	30-40	>40
Taux d'identification au rang 1 (%)	89.70	89.23	87.56	85.20	62.36	61.42	52.17	47.30	30.15

On peut remarquer aussi, que dans l'intervalle 0 à 10 ans, les taux de reconnaissance ne sont pas très différents, même chose dans l'intervalle de 11 à 20 ans, il est donc évident que la différence d'âge entre les images à un impact significatif sur les taux de reconnaissance des algorithmes de reconnaissance. Egalement, nous pouvons conclure que le descripteur proposé basé sur la DCT peut obtenir des résultats satisfaisants lorsque la différence d'âge est inférieure à 20 ans.

6.6 Effets de croissance et de vieillissement sur les performances de la reconnaissance de visage

L'objectif principal de cette partie est d'étudier séparément les effets de la croissance et les effets du vieillissement sur les performances de la reconnaissance de visage, car les transformations que peut subir le visage pendant la croissance sont totalement différentes de ceux que peut subir pendant le vieillissement [272].

En suivant le même schéma de partitionnement utilisé dans [240], nous répartissons la base de visages en trois groupes, le premier contient des images d'individus, qui sont âgés de 0 à 18 ans, le second inclut les images d'individus de 19 à 40 ans et le dernier contient ceux qui ont un âge supérieur à 40 ans, le nombre d'images et d'individus dans chaque groupe sont présentés dans le tableau 6.4 , les individus qui ont une seule ou deux images sont supprimés, nous tenant en compte des individus qui ont plus de deux images, nous utilisons une image d'un individu comme image-test et les autres images comme base de données.

Tableau 6.4. Détails des groupes d'âge.

Groupes d'âge	0-18 ans	19-40 ans	> 40 ans
Nombre d'individus	78	52	16
Nombre d'images	708	213	62

D'après les résultats présentés dans le tableau 6.5, la première remarque que l'on peut noter, c'est que les taux obtenus dans les différents groupes d'âge ne sont pas les mêmes, et le meilleur taux de reconnaissance est obtenu dans le groupe d'individus qui sont âgés de plus de 40 ans. Ces résultats peuvent être interprétés comme suit:

Le développement du visage chez l'enfant est totalement différent du vieillissement de visage chez l'adulte, où dans l'enfance, la croissance faciale se caractérise par des changements beaucoup plus dans la forme du visage, alors que dans l'âge adulte, le vieillissement du visage se caractérise par des changements beaucoup plus dans la texture, et le descripteur étudié préserve des caractéristiques importantes qui constituent la forme du visage et ignore certains détails de texture. D'où les taux élevés sont atteints dans les groupes qui sont âgés de plus de 19 ans et les taux faibles dans les groupes de moins de 19 ans.

Tableau 6.5. Taux d'identification versus groupes d'âge.

Groupes d'âge	0-18 ans	19-40 ans	> 40 ans
Taux d'identification au rang 1 (%)	43.72%	69.15%	71.91%

6.7 Conclusion

Dans ce chapitre, nous avons présenté une évaluation expérimentale de l'approche de reconnaissance faciale invariante à l'âge décrite au chapitre 5.

Nous avons montré dans ce chapitre que la combinaison d'une méthode de correction de pose basée sur les AAM, de la DCT en ignorant un nombre de coefficients de basses fréquences et de l'analyse discriminante non linéaire est une solution prometteuse pour créer des descripteurs qui sont robustes à la progression d'âge.

Egalement, nous avons étudié l'influence de la durée qui sépare l'image requête et celles de la base de données, nous avons remarqué que l'approche fournit des résultats acceptables si cette durée est inférieure à 20 ans.

Conclusion générale

Conclusion générale

La recherche d'images par le contenu visuel, comme son nom l'indique, permet de représenter les images en fonction des caractéristiques de bas niveau de leur contenu, et la recherche se fait par la comparaison de ces caractéristiques.

Le contenu visuel peut être général, il inclut la couleur, la texture ou la forme des images, c'est le cas de la recherche d'images dans le Web, ou spécifique à un domaine particulier comme la reconnaissance des empreintes digitales ou de visage.

Dans le cadre de cette thèse, premièrement, nous avons présenté un état de l'art des différentes techniques proposées dans la littérature de la recherche d'images par le contenu surtout destinées aux bases de données génériques. Nous avons constaté que ces techniques souffrent toujours du manque de l'aspect sémantique de l'information, car les utilisateurs cherchent à récupérer des images sémantiquement et perceptivement similaires, mais cet aspect est absent dans les caractéristiques de bas niveau (couleur, texture ou forme) .

Ensuite, nous nous sommes orientés vers les bases d'images spécifiques, et plus particulièrement celles qui contiennent des visages. Nous avons exposé une revue des différentes techniques proposées dans le domaine ainsi que les difficultés rencontrées par les systèmes de reconnaissance faciale qui sont dues principalement à des variabilités causées par des facteurs tels que l'illumination, l'expression, la pose et les variations d'âge, qui peuvent changer complètement l'apparence du visage.

L'objectif de ce travail a été de développer une méthode efficace de recherche d'image par le contenu visuel qui est principalement destinée à être utilisée pour la reconnaissance faciale invariante à la progression d'âge. Après avoir exposé un état de l'art des techniques proposées pour faire face à ce problème de l'invariance d'âge, une approche de description a été proposée. L'approche combine trois outils populaires utilisés pour la reconnaissance de forme qui sont les modèles actifs d'apparence, la transformée en cosinus discrète et l'analyse discriminante non linéaire.

L'approche est testée en mode identification et en mode vérification du visage en utilisant la base de donnée FG-NET et en suivant le protocole LOPO qui sont utilisés dans la plupart des études relatives à l'âge, où elle a fourni un taux de reconnaissance de 61.41% dans le cas d'identification et un taux d'erreurs égales de 7.2% dans le cas de la vérification. Comparée par les résultats obtenus par d'autres approches de la littérature, notre approche encourage l'utilisation de cette combinaison comme descripteur pour la reconnaissance de visage invariante à l'âge.

Egalement, l'approche proposée a été évaluée en terme de la longueur de la durée séparant l'image requête et celle de référence ainsi que l'influence de la croissance et le vieillissement de visage sur les performances de la méthode proposée. D'après les résultats atteints, nous avons conclu que la technique proposée est plus performante dans les groupes d'individus qui sont âgés de plus de 40 ans et lorsque la différence d'âge entre l'individu de l'image requête et celle de référence est inférieure ou égale à 20 ans.

Perspectives

- Premièrement, il est envisageable de tester la méthode avec d'autres bases de données telle que la base de donnée MORPH, pour étudier le comportement de la méthode avec des bases de données plus grandes.
- Deuxièmement, il est intéressant de noter que de meilleurs résultats peuvent être obtenus, sans doute, si une approche par composant est utilisée; le même processus est appliqué pour chaque composante du visage (yeux, nez et bouche) séparément.
- Enfin, deux aspects majeurs de la recherche dans les bases d'images sont l'efficacité et la rapidité. L'efficacité d'un système de recherche est liée à la qualité du résultat retourné à l'utilisateur, elle mesure sa capacité à retrouver les images pertinentes tout en retenant les images non pertinentes. La rapidité correspond au temps de réponse du système de recherche. Il est intéressant d'évaluer la technique proposée en terme de temps de réponse surtout si une approche par composante est utilisée, et de proposer, si nécessaire une stratégie d'indexation et de recherche pour offrir des temps de réponse qui sont raisonnables.

ANNEXES

<u>ANNEXE A</u> Coordonnées barycentriques	126
<u>ANNEXE B</u> Interpolation bilinéaire	128

ANNEXE A

Coordonnées barycentriques

Si (X_1, X_2, X_3) représentent les coordonnées des sommets d'un triangle T, chaque point $X = (x, y)^T$ appartient à T peut être défini en fonction de ses sommets:

$$\begin{aligned}
 X &= f(X_1, X_2, X_3) = X_1 + \beta(X_2 - X_1) + \gamma(X_3 - X_1) \\
 &= (1 - (\beta + \gamma))X_1 + \beta X_2 + \gamma X_3 \\
 &= \alpha X_1 + \beta X_2 + \gamma X_3
 \end{aligned}$$

où $\alpha, \beta, et \gamma$ sont des réels et $\alpha + \beta + \gamma = 1$

Les coefficients $\alpha, \beta, et \gamma$ sont appelés les coordonnées barycentriques de X par rapport à X_1, X_2, X_3 .

Les équations précédentes peuvent être écrites sous forme d'un système d'équations de trois équations à trois inconnus $\alpha, \beta, et \gamma$:

$$\begin{cases}
 \alpha x_1 + \beta x_2 + \gamma x_3 = x \\
 \alpha y_1 + \beta y_2 + \gamma y_3 = y \\
 \alpha + \beta + \gamma = 1
 \end{cases}$$

Avec la solution:

$$\alpha = 1 - (\beta + \gamma) \tag{A.1}$$

$$\beta = \frac{yx_3 - x_1y - x_3y_1 - y_3x + x_1y_3 + x_1y}{-x_2y_3 + x_2y_1 + x_1y_3 + x_3y_2 - x_3y_1 - x_1y_2} \tag{A.2}$$

$$\gamma = \frac{xy_2 - xy_1 - x_1y_2 - x_2y + x_2y_1 + x_1y}{-x_2y_3 + x_2y_1 + x_1y_3 + x_3y_2 - x_3y_1 - x_1y_2} \tag{A.3}$$

Dans notre contexte, les coordonnées barycentriques sont utilisées pour la déformation de la texture, plus précisément pour connaître si un point appartient à un triangle, où chaque forme est représentée par un maillage de triangles, en utilisant une triangulation de Delaunay.

un point X appartient à un triangle T de sommets X_1, X_2 et X_3 si ses coordonnées barycentriques α, β , et γ satisfont la condition suivante:

$$0 \leq \alpha, \beta, \gamma \leq 1$$

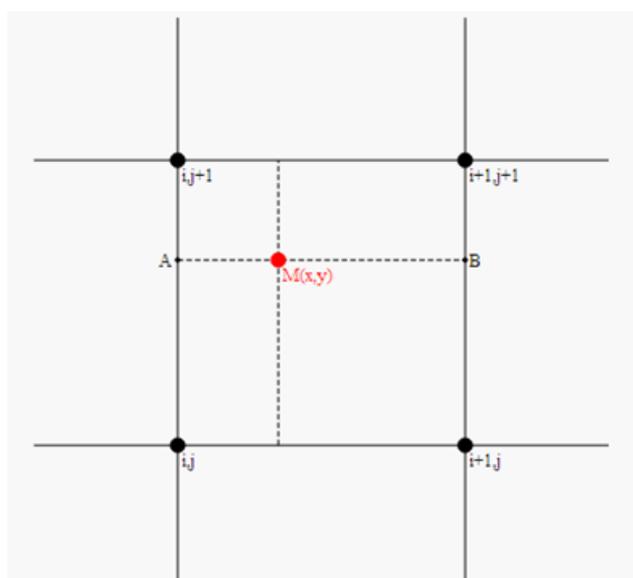
ANNEXE B

Interpolation bilinéaire

De nombreux algorithmes de traitement d'images après certains calculs, se trouvent souvent dans des situations qui nécessitent de calculer la valeur d'un niveau de gris sur un point de coordonnées qui ne sont pas des valeurs entières, c.-à-d. qui ne coïncide pas avec des pixels. Cette valeur peut être calculée par interpolation bilinéaire avec les 4 pixels voisins.

Interpolation bilinéaire

Si on a une image en niveaux de gris. Les coordonnées des pixels sont des indices, c.-à-d. des entiers. On note $V_{i,j}$ la valeur du pixel (i,j) . La figure suivante montre un point $M(x,y)$ quelconque, qui ne coïncide pas avec un pixel, avec les 4 pixels qui l'entourent.



On note:

$$t = x - i \tag{B.1}$$

$$u = y - j \tag{B.2}$$

La valeur au point A est obtenue par une interpolation linéaire sur le côté gauche :

$$V_A = (1 - u)V_{i,j} + uV_{i,j+1} \tag{B.3}$$

De même , la valeur au point B est obtenue par une interpolation linéaire sur le côté droit :

$$V_B = (1 - u)V_{i+1,j} + uV_{i+1,j+1} \quad (\text{B.4})$$

La valeur au point M est obtenue par une interpolation linéaire entre les points A et B :

$$\begin{aligned} V_M &= (1 - t)V_A + tV_B \\ &= (1 - t)(1 - u)V_{i,j} + t(1 - u)V_{i+1,j} + tuV_{i+1,j+1} + (1 - t)uV_{i,j+1} \end{aligned} \quad (\text{B.5})$$

Références bibliographiques

Références bibliographiques

- [1] N.-S. Chang and K.-S. Fu, "*A Relational Database System for Images*", in Pictorial Information Systems, Springer Berlin Heidelberg, pp. 288-321, 1980.
- [2] T. Kato, "*Database architecture for content-based image retrieval*", in SPIE/IS&T symposium on electronic imaging: science and technology, International Society for Optics and Photonics, pp. 112-123, 1992.
- [3] S.-K. Chang and A. Hsu, "*Image information systems: where do we go from here?*", IEEE Transaction on Knowledge and Data Engineering, vol.5, no.5, pp. 431-442, 1992.
- [4] H. Tamura and N. Yokoya, "*Image database systems: A survey*", Pattern Recognition, vol.17, no.1, pp. 29-43, 1984.
- [5] J. Eakins and M. Graham, "*Content-based image retrieval*", Technical Report, University of Northumbria at Newcastle, 1999.
- [6] I.-K. Sethi, I.-L. Coman and D. Stan, "*Mining association rules between low-level image features and high-level concepts*", in Aerospace/Defense Sensing, Simulation, and Controls, International Society for Optics and Photonics, pp. 279–290, 2001.
- [7] M. Sultana, and M.-L. Garilova, "*Face recognition using multiple content-based image features for biometric security applications*", International Journal of Biometrics, vol. 6, no. 4, pp.414–434, 2014.
- [8] A. Pentland, R.-W. Picard and S. Scaroff, "*Photobook: content-based manipulation for image databases*", International Journal of Computer Vision, vol. 18, no. 3, pp. 233–254, 1996.
- [9] A. Gupta and R. Jain, "*Visual information retrieval*", Communications of the ACM, vol. 40, no. 5 , pp. 70–79, 1997.
- [10] J.-R. Smith and S.-F. Chang, "*VisualSeek: a fully automated content based image query system*", in Proceedings of the 4th ACM International Conference on Multimedia, pp. 87–98, 1997.
- [11] W.-Y. Ma and B.-S. Manjunath, "*Netra: a toolbox for navigating large image databases*", in Proceedings of the IEEE International Conference on Image Processing, pp. 568–571, 1997.
- [12] J.-Z. Wang, J. Li and G. Wiederhold, "*SIMPLIcity: semantics-sensitive integrated matching for picture libraries*", IEEE Transactions on pattern analysis and machine intelligence, vol. 23, no. 9, pp. 947–963, 2001.
- [13] M. Sultana and M.-S. Uddin, "*Efficient Trademark Retrieval Using Weighted Image Features: A GPU Based Approach*", 1st edition, VDM Verlag, Germany, 2011.

- [14] M.-M. Monwar and M.-L. Gavrilova, "**FES: a system for combining face, ear and signature biometrics using rank level fusion**", in the 5th IEEE international conference on information technology: New Generations, ITNG 2008, pp.922–927, 2008.
- [15] K. Saeed, J. Pejas and R. Mosdorf, "**Biometrics, Computer Security Systems and Artificial Intelligence Applications**", Springer Science & Business Media, 2007.
- [16] M.-L. Gavrilova and M.-M. Monwar, "**Fusing multiple matcher's outputs for secure human identification**", International Journal of Biometrics, vol. 1, no. 3, pp.329–348, 2009.
- [17] R.-V. Yampolskiy and M.-L. Gavrilova, "**Artimetrics: biometrics for artificial entities**", IEEE Robotics & Automation Magazine, vol. 19, no. 4, pp.48–58, 2012.
- [18] M.-L. Gavrilova and M.-M. Monwar, "**Multimodal Biometrics and Intelligent Image Processing for Security Systems**", IGI Global, 2013.
- [19] A.-K. Jain, A. Ross and S. Prabhakar, "**An introduction to biometric recognition**", IEEE Transaction on Circuits and Systems for Video Technology, Special Issue on Image and Video-Based Biometrics, vol. 14, no. 1, pp.4–20, 2004.
- [20] "**Biometrics Face Off**" [online] <http://www.asmag.com/showpost/15098.aspx> (accessed JUNE 2015).
- [21] K. Iqbal, M.-O. Odetayo and A. James, "**Content-based image retrieval approach for biometric security using color, texture, and shape features controlled by fuzzy heuristics**", Journal of Computer and System Sciences, vol. 78, no. 4, pp.1258–1277, 2012.
- [22] H.-B. Kekre, D. Mishra and A. Kariwala, "**A Survey of CBIR techniques and semantics**", International Journal of Engineering Science and Technology (IJEST), vol. 3, no. 5, pp.4510–4517, 2011.
- [23] R.-S. Choras, "**Image feature extraction techniques and their applications for CBIR and biometrics systems**", International Journal of Biology and Biomedical Engineering, vol. 1, no. 1, pp.6–16, 2007.
- [24] W.-Y. Zhao and R. Chellappa, P.-J. Phillips, and A. Rosenfeld, "**Face recognition : A literature survey**", ACM Computing Surveys, vol. 35, no. 4, pp. 399–458, 2003.
- [25] P.-J. Phillips, H. Moon, S.-A. Rizvi and P.-J. Rauss, "**The FERET evaluation methodology for face recognition algorithms**", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp. 1090–1104, 2000.
- [26] S.-A. Rizvi, P.-J. Phillips and H. Moon, "**The verification protocol and statistical performance analysis for face recognition algorithm**", in IEEE Conference on Computer Vision and Pattern Recognition, pp. 833–838, 1998.
- [27] P.-J. Phillips, P. Grother, R. Micheals, D.-M. Blackburn, E. Tabassi and M. Bone, "**Face recognition vendor test 2002**", Evaluation Report, 2003.

- [28] K. Choudhary and N. Goel, "*A review on face recognition techniques*", in International Conference on Communication and Electronics System Design, International Society for Optics and Photonics, pp.87601E–87601E, 2013.
- [29] H. Ling, S. Soatto, N. Ramanathan and D.-W. Jacobs, "*A study of face recognition as people age*", in Proceedings of IEEE International Conference on Computer Vision, pp. 1-8, 2007.
- [30] N. Ramanathan and R. Chellappa. "*Face verification across age progression*", in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 462-469, 2005.
- [31] T. Acharya and A. K.-Ray, "*Image processing: principles and applications*", John Wiley & Sons, 2005.
- [32] M.-J. Swain, and D.-H. Ballard, "*Color indexing*", *International journal of computer vision*, vol. 7, no. 1, pp. 11-32, 1991.
- [33] N. Boujemaa, S. Boughorbel, C. Vertan, "*Description de la répartition spatiale de la couleur pour l'Indexation d'Images*", Reconnaissance des formes et Intelligence Artificielle (RFIA'2002). Vol. 2. pp. 407-414. Angers. France. Jan 2002.
- [34] G. Pass and R. Zabih. "*Histogram refinement for content based image retrieval*", in proceedings of 3rd IEEE Workshop on Application of Computer Vision, pp. 96 - 102, 1996.
- [35] J. Huang, S. R. Kumer, M. Mitra, W-J Zhu, and R. Zabih. "*Image indexing using color correlograms*", in Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 762- 768, 1997.
- [36] N.-R. Howe. "*Percentile blobs for image similarity*", in Proceedings of IEEE Workshop on Content- Based Access of Image and Video Libraries, pp. 78 - 83, 1998.
- [37] J. P. Cocquerez and S. Philipp-Foliguet, "*Analyse d'images: filtrage et segmentation*", 1995.
- [38] A. Porebski, N. Vandenbroucke, L. Macaire, "*Attributs de texture couleur*", dans Actes de l'Ecole d'Hiver sur l'Imagerie Numérique Couleur (EHINC'09), Toulon (France), pp. 95-108, 2009.
- [39] M. Sharma, "*Evaluation of image segmentation and texture extraction methods in scene analysis*", Ph.D. Thesis, University of Exeter, 2001.
- [40] M. Sharma and S. Singh, "*Evaluation of texture methods for image analysis*", in *Proceedings of the 7th Australian and New Zealand Intelligent Information Systems Conference*, pp. 117–121, 2001.
- [41] S. Paquis, "*Reconnaissance des surfaces de chaussée par analyse de texture image*", Thèse de doctorat, Université Jean Monnet, 2003.

- [42] B. S. Manjunath and W. Y. Ma, " *Texture features for Browsing and retrieval of image data*", IEEE transactions on pattern analysis and machine intelligence, vol. 18. No. 8, pp. 837-842, 1996.
- [43] R. Haralick, K. Shanmugam, " *Textured Features for Image Classification*" , IEEE Transactions on Systems, Man and Cybernetics, vol. 6, no. 3, pp. 610–621, 1973.
- [44] M. Chantler, and J. Wu, "*Rotation Invariant Classification of 3D Surface Textures using Photometric Stereo and Surface Magnitude Spectra*", Proceedings of British Machine Vision Conference, Vol.2, pp. 486-495, 2002.
- [45] R. Dubes, and A. Jain, "*Random field models in image analysis*", Journal of Applied Statistic, vol. 16, no. 2, pp.131-164, 1989.
- [46] A. Pentland, "*Fractal-Based Description of Natural Scenes*", IEEE Transaction on Pattern Analysis Machine Intelligence, vol. 6, no. 6, pp. 661-674, 1984.
- [47] J. Mao, and A. Jain, "*Texture Classification and Segmentation using Multi-Resolution Simultaneous Autoregressive Models*", Pattern Recognition, vol. 25, no. 2, pp. 173-188, 1992.
- [48] P. Srivastava, O. Prakash, & A. Khare, "*Content-Based Image Retrieval using moments of wavelet transform*", in *International Conference on Control, Automation and Information Sciences (ICCAIS)*, pp. 159-164, 2014.
- [49] J. E. Gary, and R. Mehrotra, "*Shape similarity-based retrieval in image database systems*", Image Storage and Retrieval Systems, vol. 1662, pp. 2-8, 1992.
- [50] W. I. Grosky, and R. Mehrotra, "*Index based object recognition in pictorial data management*", Computer Vision, Graphics, and image processing, vol. 52, no. 3, pp. 416-436, 1990.
- [51] H. V. Jagadish, "*A retrieval technique for similar shapes*", ACM SIGMOID Record, vol. 20, no. 2, pp. 208-217, 1991.
- [52] D. Tegolo, "*Shape analysis for image retrieval*", in International Symposium on Electronic Imaging: Science and Technology , pp. 59-69, 1994.
- [53] D. Zhang & G. Lu, "*Review of shape representation and description techniques*", Pattern recognition, vol. 37, no.1, pp. 1-19, 2004.
- [54] C. Teh and T. Roland, "*On image analysis by the methods of moments*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 4, pp. 496-513, 1988.
- [55] M. K. Hu, "*Visual Pattern Recognition by Moment Invariants*", IRE Transactions on Information Theory, vol. 8, no. 2, pp. 179-187, 1962.
- [56] F.Mokhtarian, S.Abbasi, and J.Kittler, "*Efficient and robust retrieval by shape content through curvature scale space*", *Image Databases and MultiMedia Search*, vol.8, pp 51-58, 1998.

- [57] M. Bober, "**MPEG-7 Visual Shape Descriptors**", IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6, pp. 716-719, 2001.
- [58] B. Tellez, and A. Baskurt, "**Une adaptation du descripteur CSS pour la recherche de parties communes entre objets**", in Proceedings of Colloque CORESA'03, Lyon 2003.
- [59] M. Yang, K. Kpalma, J. Ronsin, "**A survey of shape feature extraction techniques**", Pattern Recognition, IN-TECH, Peng-Yeng Yin , pp. 43-90, 2008.
- [60] A. Chaudhry, J. Ullah, M.-A. Jaffar, J.-Y. Kim, T.-A. Tuan, "**Human Activity Recognition System: Using Improved Crossbreed Features and Artificial Neural Network**", Life Science Journal, vol. 9, no. 4, pp. 5351-5356, 2012.
- [61] M. Zhenjiang, "**Zernike moment - based image shape analysis and its application**", Pattern Recognition Letters, Vol. 21, pp. 169-177, 2000.
- [62] A. Khotanzad, Y.H.Hong, "**Invariant image recognition by Zernike moments**", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 5, pp. 489-498, 1990.
- [63] Y.-S. Kim, W.-Y. Kim, "**Content-based trademark retrieval system using a visually salient feature**", Image and Vision Computing, vol. 16, no. 12, pp. 931-939, 1998.
- [64] W.-Y. Kim, Y.-S. Kim, "**A region-based shape descriptor using Zernike moments**", Signal Processing: Image Communication, vol. 16, no. 1, pp.95-102, 2000.
- [65] S. Nugroho and D. Utomo, "**Rotation Invariant Indexing For Image Using Zernike Moments and R-Tree**", TELKOMNIKA (Telecommunication Computing Electronics and Control), vol.9, no.2, pp. 335-340, 2013.
- [66] R.-S. Choras, "**Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems**", international journal of biology and biomedical engineering, vol.1, no.1, pp. 6-16, 2007.
- [67] T. Arif, Z. Shaaban, L. Krekor, and S. Baba, "**Object classification via geometrical, zernike and legendre moments**", Journal of Theoretical and Applied Information Technology, vol.7, no.1, pp.31-37, 2009.
- [68] B. Jyothi, Y.-M. Latha, P.-K. Mohan, and V.-S.-K. Reddy, , "**Medical Image Retrieval Using Moments**", International Journal of Application or Innovation in Engineering & Management, vol.2, no.1, pp.195-200, 2013.
- [69] C. Rao, S.-S. Kumar, and B.-C. Mohan, , "**Content based image retrieval using exact legendre moments and support vector machine**", International journal of multimedia and its applications (IJMA), vol. 2, no. 2, pp. 69-79, 2010.
- [70] C.-W. Chonga, P. Raveendranb and R. Mukundan, "**Translation and scale invariants of Legendre moments**", Pattern Recognition, vol. 37, no. 1, pp. 119 – 129, 2004.

- [71] C.H. Teh and R.T. Chin, "*On image analysis by the methods of moments*", IEEE Transactions on Pattern analysis and machine intelligence, vol. 10, no. 4 ,pp. 496–513, 1988.
- [72] R. Mukundan, K.-R. Ramakrishnan, "*Moment Functions in Image Analysis—Theory and Applications*", World Scientific, 1998.
- [73] J.-D. Zhou, H.-Z. Shu, L.-M. Luo and W.-X. Yu, "*Two new algorithms for efficient computation of Legendre moments*", Pattern Recognition, Vol.35, no. 5, pp. 1143-1152, 2002.
- [74] P.-T. Yap and R. Paramesaran, "*An Efficient Method for the Computation of Legendre Moments*", IEEE Transactions on Pattern Recognition and Machine Intelligence, Vol.27, No.12, pp.1996-2002, 2005.
- [75] S.-X. Liao and M. Pawlak, "*On image analysis by moments*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, no. 3, pp. 254-266, 1996.
- [76] K.-M. Hosney, "*Exact Legendre moments computation for gray level images*", Pattern Recognition, Vol. 40, no. 12, pp. 3597-3605, 2007.
- [77] H.-A. Moghaddam, T. Khajoie, A.-H. Rouhie, and M.-S. Tarzjan, "*Wavelet Correlogram: A New Approach for Image Indexing and Retrieval*", Pattern Recognition, vol. 38, no. 12, pp. 2506-2518, 2005.
- [78] T. Çelik, T. Tjahjadi, "*Multi-scale texture classification and retrieval based on magnitude and phase features of complex wavelet subbands*", Computers & Electrical Engineering, vol. 37, no 5, pp. 729–743, 2011.
- [79] L. Boussaad, M. Benmohammed, "*Pondération dynamiques de descripteurs par essaim de particules pour la recherche d'images par le contenu*" ,Conférence nationale sur la modélisation et la simulation informatique des systèmes industriels, MSISI' 2010, Skikda, Algérie, 2010.
- [80] F. Long, H. Zhang and D.-D. Feng, "*Fundamentals of content-based image retrieval*", In *Multimedia Information Retrieval and Management*, Springer Berlin Heidelberg, pp. 1-26, 2003.
- [81] G. Qian, S. Sural, Y. Gu, and S. Pramanik, "*Similarity between Euclidean and cosine angle distance for nearest neighbor queries*", in Proceedings of the ACM Symposium on Applied Computing, pp. 1232-1237, 2004.
- [82] J.-R. Smith and S.-F. Chang , "*Tools and techniques for color image retrieval*", in Electronic Imaging: Science & Technology, International Society for Optics and Photonics, pp. 426-437, 1996.
- [83] J. Puzicha, T. Hofman and J. Buhman, "*Non-parametric similarity measures for unsupervised texture segmentation and image retrieval*", in proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 267-272, 1997.

- [84] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "**Query by image and video content: The QBIC system**", IEEE Computer, Vol.28, No.9, pp. 23-32, 1995.
- [85] J.-A. Catalan, and J.-S. Jin, "**Dimension reduction of texture features for image retrieval using hybrid associative neural networks**", IEEE International Conference on Multimedia and Expo, Vol.2, pp. 1211 -1214, 2000.
- [86] N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger, "**The R*-tree: An efficient robust access method for points and rectangles**", in ACM SIGMOD record, vol. 19, no. 2, pp. 322-331, 1990.
- [87] J.-T. Robinson, "**The k-d-B-tree: a search structure for large multidimensional dynamic indexes**", In Proceedings of the ACM SIGMOD international conference on Management of data, pp. 10-18, 1981.
- [88] J. Nievergelt, H. Hinterberger, and K.-C. Sevcik, "**The grid file: an adaptable symmetric multikey file structure**", ACM Transactions on Database Systems, vol. 9, no. 1, pp. 38-71, 1984.
- [89] H.-J. Zhang, and D. Zhong, "**A Scheme for visual feature-based image indexing**", Proceedings of SPIE, Storage and Retrieval for Image and Video Databases III, pp. 36-46, 1995.
- [90] J. Laaksonen, M. Koskela, S. Laakso, and E. Oja, "**Picsom-content-based image retrieval with self-organizing maps**", Pattern Recognition Letters, vol. 21, no. 13, pp. 1199– 1207, 2000.
- [91] I.-K. Sethi, I.-L. Coman and D. Stan, "**Mining association rules between low-level image features and high-level concepts**", in SPIE Proceedings, Data Mining and Knowledge Discovery: Theory, tools and technology III, pp. 279–290, 2001.
- [92] C. Town and D. Sinclair, "**Content-based image retrieval using semantic visual categories**", TR2000-14, AT&T Labs Cambridge, 2001.
- [93] A. Vailaya, M.A. Figueiredo, A.-K. Jain, H.-J. Zhang, "**Image classification for content-based indexing**", IEEE transactions on image processing, vol. 10, no 1, pp. 117–130, 2001.
- [94] J. Luo, A. Savakis, "**Indoor vs outdoor classification of consumer photographs using low-level and semantic features**", International Conference on Image Processing (ICIP), vo.l 2, pp. 745–748, 2001.
- [95] T. Hastie, R. Tibshirani, J. Friedman, "**The Elements of Statistical Learning: Data Mining, Inference, and Prediction**", Springer series in statistics, 1st edition, 2001.
- [96] H. Feng, T.-S. Chua, "**A bootstrapping approach to annotating large image collection**", in Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, pp. 55–62, 2003.

- [97] R. Shi, H. Feng, T.-S. Chua and C.-H. Lee, "***An adaptive image content representation and segmentation approach to automatic image annotation***", International Conference on Image and Video Retrieval (CIVR), pp. 545–554, 2004.
- [98] V.-N. Vapnik and V. Vapnik, "***Statistical Learning Theory***", Wiley, 1998.
- [99] W. Jin, R. Shi, and T.-S. Chua, "***A semi-naïve bayesian method incorporating clustering with pairwise constraints for auto image annotation***", In Proceedings of the 12th annual ACM international conference on Multimedia, pp. 336-339, 2004.
- [100] S. Tong and E. Chang, "***Support vector machine active learning for image retrieval***", in Proceedings of the 9th ACM International Multimedia Conference, pp. 107–118, 2001.
- [101] L. Zhang, F. Lin, B. Zhang, "***Support vector machine learning for image retrieval***", International Conference on Image Processing, pp. 7–10, 2001.
- [102] N. Vasconcelos and A. Lippman, "***Library-based coding: a representation for efficient video compression and retrieval***", in Proceedings of the Data Compression Conference (DCC97), pp. 121–130, 1997.
- [103] Y. Liu, D. Zhang, G. Lu and W.-Y. Ma. "***A survey of content-based image retrieval with high level semantics***". Pattern Recognition, vol. 40, no. 1, pp. 262-282, 2007.
- [104] M. Pal, P.-M. Mather, "***Decision tree based classification of remotely sensed data***", in paper presented in the 22nd Asian Conference on Remote Sensing (ACRS), 2001.
- [105] I.-F. Rajam and S. Valli "***SRBIR: semantic region based image retrieval by extracting the dominant region and semantic learning***". Journal of Computer Science, Vol. 7, No. 3, pp.400–408, 2011.
- [106] I.-F. Rajam and S. Valli "***Region-based image retrieval using the semantic cluster matrix and adaptive learning***", International Journal of Computational Science and Engineering, Vol. 7, No. 3, pp.239–252, 2012.
- [107] D. Stan and I.-K. Sethi, "***Mapping low-level image features to semantic concepts***", In Photonics West 2001-Electronic Imaging, International Society for Optics and Photonics, pp. 172–179, 2001.
- [108] X. Zheng, D. Cai, X. He, W.-Y. Ma, X. Lin, "***Locality preserving clustering for image database***", In Proceedings of the 12th annual ACM international conference on Multimedia, pp. 885-891, 2004.
- [109] J. Shi, J. Malik, "***Normalized cuts and image segmentation***", IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 8, pp.888–905, 2000.
- [110] A.-Y. Ng, M.-I. Jordan, Y. Weiss, "***On spectral clustering: analysis and an algorithm***", Advances in Neural Information Processing Systems, vol. 14, no. 2, pp. 849-856, 2001.

- [111] Y. Chen, J.-Z. Wang, R. Krovetz, "***An unsupervised learning approach to content-based image retrieval***", IEEE Proceedings of the International Symposium on Signal Processing and its Applications, vol. 1, pp. 197–200, 2003.
- [112] N. Vasconcelos, "***The design of end-to-end optimal image retrieval systems***", in Proceedings of the International Conference on ANN, Istanbul, Turkey, 2003.
- [113] N. Vasconcelos, "***On the efficient evaluation of probabilistic similarity functions for image retrieval***", IEEE Transactions on information theory, vol. 50, no. 7, pp. 1482–1496, 2004.
- [114] M.-M. Rahman, P. Bhattacharya and B.C. Desai, "***A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback***", IEEE transactions on Information Technology in Biomedicine, vol. 11, no. 1, pp.58–69, 2007.
- [115] G. QIU, J. Morris and X. Fan, "***Visual guided navigation for image retrieval***", *Pattern Recognition*, vol. 40, no. 6, pp.1711–1721, 2007.
- [116] W.-Y. MA et B.-S. Manjunath, "***A texture thesaurus for browsing large aerial photographs***", *Journal of The American Society for Information Science*, vol. 49 no. 7, pp. 633–648, 1998.
- [117] J. Assfalg, A. Del Bimbo, and P. Pala, "***Using multiple examples for content-based image retrieval***", in IEEE International Conference on Multimedia and Expo, vol. 1, pp. 335-338, 2000.
- [118] G. Salton, "***Automatic text processing: The transformation, analysis, and retrieval of information by computer***", Addison-Wesley publishing, 1989.
- [119] Y. Rui, T.-S. Huang, M. Ortega, S. Mehrotra, "***Relevance feedback: a power tool for interactive content-based image retrieval***", IEEE Transactions on circuits and systems for video technology, vol.8, no. 5, pp. 644–655, 1998.
- [120] Y. Rui, T. Huang, "***Optimizing learning in image retrieval***", in Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 236–243, 2000.
- [121] G.-D. Guo, A.-K. Jain, W.-Y. Ma and H.-J. Zhang, "***Learning similarity measure for natural image retrieval with relevance feedback***", IEEE Transactions on Neural Networks, vol. 13, no. 4, pp.811–820, 2002.
- [122] F. Jing, M. Li, L. Zhang, H.-J. Zhang and B. Zhang, "***Learning in region based image retrieval***", in Proceedings of the International Conference on Image and Video Retrieval, pp. 206–215, 2003.
- [123] X.-S. Zhou and T.-S. Huang, "***Relevance feedback in image retrieval: a comprehensive review***", *Multimedia Systems*, vol. 8, no.6, pp. 536–544, 2003.

- [124] A.-W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "**Content-based image retrieval at the end of the early years**", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.12, pp. 1349-1380, 2000.
- [125] S. Krishnamachari and M. Abdel-Mottaleb, "**Compact color descriptor for fast image and video segment retrieval**", In Proceedings of IST/SPIE Conference on Storage and Retrieval of Media Databases, 2000.
- [126] S.-Z. Li. "**Encyclopedia of Biometrics: I-Z** ", Springer Science & Business Media, vol. 1, 2009.
- [127] J. Yang, X. Chen, and W. Kunz, "**A PDA-based face recognition system**", in Proceedings of sixth IEEE Workshop on Applications of Computer Vision, pp.19-23, 2002.
- [128] Y. Moses, Y. Adini and S. Ullman, "**Face recognition: The problem of compensating for changes in illumination direction**", In European conference on computer vision, Springer Berlin Heidelberg, pp. 286-296, 1994.
- [129] R. Gross, J. Shi, J. Cohn, "**Quo vadis Face Recognition? - The current state of the art in Face Recognition**", Technical Report, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA, 25 pages, 2001.
- [130] P. -J. Phillips, H. Wechsler, J.Huang, and P.-J. Rauss, "**The FERET database and evaluation procedure for face-recognition algorithm**", Image and Vision Computing, vol.16, no. 5, pp.295-306, 1998.
- [131] D.-M. Blackburn, M. Bone, and P.-J. Phillips, "**Face recognition vendor test 2000**", Technical report, Defense Advanced Research Projects Agency, Arlington, VA, 2001.
- [132] K. Messer, J. Kittler, M. Sadeghi, M. Hamouz, A. Kostin, F. Cardinaux, S. Marcel, S. Bengio, C. Sanderson, J. Czyz, L. Vandendorpe, C. McCool, S. Lowther, S. Sridharan, V. Chandran, R. P. Palacios, E. Vidal, L. Bai, L. Shen, Y. Wang, Y.-H. Chiang, H.-C. Liu, Y.-P. Hung, A. Heinrichs, M. Müller, A. Tewes, C.-v.-d. Malsburg, R. P. Würtz, Z. Wang, F. Xue, Y. Ma, Q. Yang, C. Fang, X. Ding, S. Lucey, R. Goss, H. Schneiderman, N. Poh, and Y. Rodriguez, "**Face Authentication Test on the BANCA Database**", in the 17th International Conference on Pattern Recognition, Vol.4, pp.523-532, 2004.
- [133] X. Ding and C. Fang, "**Discussions on some problems in face recognition**", in Advances In Biometric Person Authentication, Springer Berlin, Heidelberg, pp.47-56, 2004,
- [134] F. Tarrés, A. Rama, "**GTAV Face Database**", available at: <http://gps-tsc.upc.es/GTAV/ResearchAreas/UPCFaceDatabase/GTAVFaceDatabase.htm>
- [135] P.-J. Phillips, W.-T. Scruggs, A.-J. O'Toole, P.-J. Flynn, K.-W. Bowyer, C.-L. Schott, M. Sharpe "**FRVT 2006 and ICE 2006 large-scale experimental results**", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no.5, pp. 831-846, 2010.

- [136] H. Shim, J. Luo, T. Chen, "*A subspace model-based approach to face relighting under unknown lighting and poses*", IEEE Transactions on Image Processing, vol. 17, no. 8, pp. 1331–1341, 2008.
- [137] A.-S. Georghiades, P.-N. Belhumeur, D.-J. Kriegman, "*From few to many: Illumination cone models for face recognition under variable lighting and pose*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no.6, pp. 643–660, 2001.
- [138] T. Zhang, Y.-Y. Tang, B. Fang, Z. Shang, X. Liu, "*Face recognition under varying illumination using gradientfaces*", IEEE Transactions on Image Processing, vol. 18, no. 11, pp. 2599–2606, 2009.
- [139] J.-Y. Zhu, W.-S. Zheng, J.-H. Lai, "*Complete gradient face- a novel illumination invariant descriptor*", In Chinese Conference on Biometric Recognition, pp. 17–24, 2012.
- [140] S.-M. Pizer, E.-P. Amburn, J.-D. Austin, R. Cromartie, "*Adaptive histogram equalization and its variations*", Computer Vision, Graphics, and Image Processing, vol. 39, no. 3, pp. 355–368, 1987.
- [141] P.-H. Lee, S.-W. Wu and Y.-P. Hung, "*Illumination compensation using oriented local histogram equalization and its application to face recognition*", IEEE Transactions on Image Processing, vol. 21, no. 9, pp.4280–4289, 2012.
- [142] M.-S. Bartlett, G. Littlewort, I. Fasel, and J.-R. Movellan, "*Real Time Face detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction*", In Computer Vision and Pattern Recognition Workshop, CVPRW'03, vol. 5, pp. 53-53, 2003.
- [143] Y.-L. Tian, T. Kanade, and J.-F. Cohn, "*Recognizing action units for facial expression analysis*", series in machine perception and artificial intelligence, vol. 48, pp. 32-66, 2002.
- [144] M. Lyons, J. Budynek and S. Akmatsu, "*Automatic classification of single facial images*", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 21, no. 12, pp. 1357-1362, 1999.
- [145] A.-M. Martinez and R. Benavente, "*The AR face database*", CVC Technical Report, no. 24, 1998.
- [146] A. Rama, F. Tarres, L. Goldmann, and T. Sikora, "*More robust face recognition by considering occlusion information*", in 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG '08, pp. 1-6, 2008.
- [147] H.-J. Oh, K.-M. Lee, and S.-U. Lee, "*Occlusion invariant face recognition using selective local non-negative matrix factorization basis images*", Image and Vision Computing, vol. 26, no. 11, pp. 1515–1523, 2008.
- [148] "*FG-NET Aging Database*", Available: <http://www.fgnet.rsunit.com>.

- [149] M. Visani, "*Vers de nouvelles approches discriminantes pour la reconnaissance automatique de visages*", Thèse de doctorat, INSA de Lyon, 2005.
- [150] A. Samal, P.-A. Iyengar, "*Automatic recognition and analysis of human faces and facial expressions: a survey*", Pattern Recognition, vol. 25, no.1, pp. 65–77, 1992.
- [151] D. Valentin, H. Abdi, A. J. O'Toole, and G. W. Cottrell, "*Connectionist models of face processing: a survey*", Pattern Recognition, vol.27, no.9 , pp.1209–1230, 1994.
- [152] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "*Face recognition from a single image per person: A survey*", Pattern recognition, vol. 39, no.9, pp. 1725-1745, 2006.
- [153] A.-J. O'Toole, K. A. Deffenbacher, D. Valentin and H. Abdi, "*Low-dimensional representation of faces in higher dimensions of the face space*", JOSA A vol. 10, no.3, pp. 405–411,1993.
- [154] M.-A. Turk, and A.-P. Pentland, "*Face recognition using eigenfaces*", in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 586-591, 1991.
- [155] P.-N. Belhumeur, J.-P. Hespanha and D.-J. Kriegman, "*Eigenfaces vs. Fisherfaces : recognition using class specific linear projection*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.19, no.7, pp.711–720, 1997.
- [156] A.-M. Martinez et A.-C. Kak, "*PCA versus LDA*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 228–233, 2001.
- [157] M.-S. Bartlett, J.-R. Movellan et T.-J. Sejnowski, "*Face recognition by independent component analysis*", IEEE Transactions on Neural Networks, vol. 13, no. 6, pp. 1450–1464, 2002.
- [158] I.-T. Jolliffe, "*Principal component analysis*", Springer series in statistics, 1986.
- [159] G. Shakhnarovich, and B. Moghaddam, "*Face recognition in subspaces*", in Handbook of Face Recognition, Springer London, pp. 19-49, 2011.
- [160] C. Jutten, and J. Herault, "*Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture*", Signal processing, vol. 24, no. 1, pp. 1-10, 1991.
- [161] A.-J. Bell and T.-J. Sejnowski, "*An information-maximization approach to blind separation and blind deconvolution*", Neural computation, vol.7, no. 6, pp. 1129-1159, 1995.
- [162] J.-F. Cardoso, "*High-order contrasts for independent component analysis*", Neural computation, vol. 11, no.1, pp. 157-192, 1999.
- [163] A. Hyvärinen, and E. Oja, "*A fast fixed-point algorithm for independent component analysis*", Neural computation, vol. 9, no.7, pp. 1483-1492, 1997.

- [164] B.-A. Draper, K. Baek, M.-S. Bartlett, and J.-R. Beveridge, "**Recognizing faces with PCA and ICA**", *Computer vision and image understanding*, vol. 91, no. 1, pp. 115-137, 2003.
- [165] B. Scholkopf, A. Smola et K.-R. Muller: "**Nonlinear Component Analysis as a Kernel Eigenvalue Problem**", *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [166] M.-H. Yang: "**Kernel Eigenfaces vs. Kernel Fisherfaces : Face Recognition Using Kernel Methods**", in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 215-220, 2002.
- [167] T.-F. Cootes, G.-J. Edwards, and C.-J. Taylor. "**Active appearance models**", *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 681-685, 2001.
- [168] A. Lanitis, C.-J. Taylor, and T.-F. Cootes, "**Automatic face identification system using flexible appearance models**", *Image and vision computing*, vol.13, no.5, pp. 393-401, 1995.
- [169] N. Kumar, V. Abhishek, and G. Gautam, "**A novel approach for person authentication and content-based tracking in videos using kernel methods and active appearance models**", in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, pp. 1384-1389, 2003.
- [170] C. Garcia, and M. Delakis, "**Convolutional face finder: A neural architecture for fast and robust face detection**", *IEEE Transactions on pattern analysis and machine intelligence*, vol. 26, no. 11, pp.1408-1423, 2004.
- [171] A. Mohamed, Y. Weng, J. Jiang, and S. Ipson, "**Face detection based neural networks using robust skin color segmentation**", in *5th International Multi-Conference on Systems, Signals and Device*, pp. 1-5, 2008.
- [172] S. Lawrence, C.-L. Giles, A.-C. Tsoi, and A.-D Back,. "**Face recognition: A convolutional neural-network approach**", *IEEE transactions on neural networks*, vol. 8, no.1, pp. 98-113, 1997.
- [173] W. Wang, "**Face Recognition Based On Radial Basis Function Neural Networks**", in *International Seminar on Future Information Technology and Management Engineering, FITME'08*, pp. 41-44, 2008.
- [174] M.-S. Charifa, A. Suliman, and M. Bikdash, "**Face Recognition Using a Hybrid General Backpropagation Neural Network**", in *IEEE International Conference on Granular Computing, GRC 2007*, pp. 510-515, 2007.
- [175] M.-C. Zou, "**Face Recognition Based on FastICA and RBF Neural Networks**", in *International Symposium on Information Science and Engineering , ISISE'08*, vol. 1, pp. 588-592, 2008.

- [176] M. Li, F. Wu, and X. Liu, "**Face recognition based on WT, fastICA and RBF neural network**", in Third International Conference on Natural Computation ICNC 2007, vol. 2, pp. 3-7, 2007.
- [177] R. Hecht-Nielsen, "**Self-Organization and Associative Memory**", *IEEE Journal of Quantum Electronics*, vol. 25, no. 2, p. 237, 1989.
- [178] R. Brunelli, and T. Poggio, "**Face recognition: Features versus templates**", *IEEE transactions on pattern analysis and machine intelligence*, vol. 15, no.10, pp. 1042-1052, 1993.
- [179] P. Viola, and M. Jones, "**Rapid object detection using a boosted cascade of simple features**", in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001., vol. 1, 2001.
- [180] Y.-B. Jemaa, and S. Khanfir, "**Automatic Gabor Features Extraction for Face Recognition using Neural Networks**", in IEEE First Workshops on Image Processing Theory, Tools and Applications , pp. 1-6, 2008.
- [181] A. Pentland, B. Moghaddam, and T. Starner, "**View-based and modular eigenspaces for face recognition**", in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'94, pp. 84-91, 1994.
- [182] J.-R. Price, and T.-F. Gee, "**Face recognition using direct, weighted linear discriminant analysis and modular subspaces**", *Pattern Recognition*, vol. 38, no.2, pp. 209-219, 2005.
- [183] A.-M. Martínez, "**Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class**", *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 24, no. 6, pp. 748-763, 2002.
- [184] X. Tan, and S. Chen, and Z.-H. Zhou, and F. Zhang, "**Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble**", *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 875-886, 2005.
- [185] T. Kohonen, "**The self-organizing map**", *Proceedings of the IEEE*, vol. 78, no.9, pp. 1464-1480, 1990.
- [186] F. Samaria, "**Face segmentation for identification using hidden Markov models**", Olivetti Research Limited, pp. 399-408, 1993.
- [187] B.-S. Manjunath, R. Chellappa, and C. von der Malsburg, "**A feature based approach to face recognition**", In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'92, pp. 373-378, 1992.
- [188] Y. Gao, and Y. Qi, "**Robust visual similarity retrieval in single model face databases**", *Pattern Recognition*, vol. 38, no. 7, pp. 1009-1020, 2005.

- [189] M. Lades, J.-C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen, "***Distortion invariant object recognition in the dynamic link architecture***", *IEEE Transactions on computers*, vol. 42, no. 3, pp.300-311, 1993.
- [190] H. Shin, S.-D. Kim, and H.-C. Choi, "***Generalized elastic graph matching for face recognition***", *Pattern Recognition Letters*, vol. 28, no. 9, pp. 1077-1082, 2007.
- [191] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. Von Der Malsburg, "***Face recognition by elastic bunch graph matching***", *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 775-779, 1997.
- [192] J. Zhang, Y. Yan and M. Lades, "***Face recognition: eigenface, elastic matching and neural nets***", *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1423–1435, 1997.
- [193] S. Arca, P. Campadelli and R. Lanzarotti, "***A Face Recognition System Based On Automatically Determined Facial Fiducial Points***", *Pattern recognition*, vol. 39, no. 3, pp. 432-443, 2005.
- [194] C. Podilchuk and X. Zhang, "***Face recognition using DCT-based feature vectors***", in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2144-2147, 1996.
- [195] V.-V. Kohir and U.-B. Desai, "***Face recognition using a DCT-HMM approach***", in *Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision*, pp. 226-231, 1998.
- [196] Z. Pan and H. Bolouri, "***High speed face recognition based on discrete cosine transforms and neural networks***", technical report, University of Hertfordshire, UK, 1999.
- [197] A. Nefian, "***A hidden Markov model-based approach for face detection and recognition***", Thèse de doctorat. School of Electrical and Computer Engineering, Georgia Institute of Technology, 1999.
- [198] S. Eickeler, S. Muller, and G. Rigoll, "***Recognition of JPEG compressed face images based on statistical methods***", *Image and Vision Computing*, vol. 18, no. 4, pp. 279-287, 2000.
- [199] Z.-M. Hafed and M.-D. Levine, "***Face recognition using the discrete cosine transform***", *International Journal of Computer Vision*, vol. 43, no. 3, pp. 167-188, 2001.
- [200] S. Zhao and R.-R. Grigat, "***Multiblock-fusion scheme for face recognition***", in *Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004*, vol. 1, pp. 309-312, 2004.
- [201] B. Moghaddam, T. Jebara, and A. Pentland, "***Bayesian face recognition***", *Pattern Recognition*, vol. 33, no. 11, pp. 1771-1782, 2000.

- [202] J. Kittler, M. Hatef, R.-P. Duin, and J. Matas, "**On combining classifiers**", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 3, pp. 226-239, 1998.
- [203] W. Chen, M.-J. Er, and S. Wu, "**PCA and LDA in DCT domain**", Pattern Recognition Letters, vol. 26, no. 15, pp. 2474-2482, 2005.
- [204] M.-J. Er, W. Chen, and S. Wu, "**High-speed face recognition based on discrete cosine transform and RBF neural networks**", IEEE Transactions on Neural Networks, vol. 16, no. 3, pp. 679-691, 2005.
- [205] W. Chen, M.-J. Er, and S. Wu, "**Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain**", IEEE Transactions on Systems, Man, and Cybernetics, Part B(Cybernetics), vol. 36, no. 2, pp. 458- 466, 2006.
- [206] V. Govindarajan, S.-K. VVS, and S. Ramachandran, "**Face recognition using block-based DCT feature extraction**", Journal of Advanced Computer Science & Technology, vol. 1, no. 4, pp. 266-283, 2012.
- [207] M. Mazloom, and S. Ayat, "**Combinational method for face recognition: wavelet, PCA and ANN**", In *Digital Image Computing: Techniques and Applications (DICTA)*, pp. 90-95, 2008.
- [208] I. Paliy, A. Sachenko, V. Koval, and Y. Kurylyak, "**Approach to face recognition using Neural Networks**", In *IEEE Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, pp. 112-115, 2005.
- [209] J. Huang, X. Shao, and H. Wechsler, "**Face pose discrimination using support vector machines (SVM)**", in Proceedings of the fourteenth International Conference on Pattern Recognition, vol. 1, pp. 154-156, 1998.
- [210] R. Rioux, and T. Simon, "**Redondance de classes pour la reconnaissance de visages par SVM**", *CNRIUT 2008: XIV Colloque National de la Recherche dans les IUT. IUT A et IUT B, Université Claude Bernard LYON 1*, 2008.
- [211] C. Velardo, J.-L. Dugelay, L. Daniel, A. Dantcheva, N. Erdogmus, N. Kose, R. Min, and X. Zhao. "**Introduction to biometry**", in *Multimedia Image and Video Processing*, 2nd edition, CRC Press, pp. 397-418, 2012.
- [212] N. MORIZET, E.-A. Thomas, F. ROSSANT, F. AMIEL, and A. AMARA, "**Revue des algorithmes PCA, LDA et EBGGM utilisés en reconnaissance 2D du visage pour la biométrie** ", Institut Supérieur d'Electronique de Paris (ISEP), département d'Electronique, 2006.
- [213] J. Yang, D. Zhang, A.-F. Frangi, and J.-Y. Yang, "**Two-dimensional PCA: a new approach to appearance-based face representation and recognition**", IEEE transactions on pattern analysis and machine intelligence, vol. 26, no.1, pp. 131-137, 2004.

- [214] L. Boussaad, M. Benmohammed, R. Benzid , "*a comparative study of different subspace methods and metrics for age-invariant face recognition*", 2nd International Conference on Image, Vision and their Applications, Guelma, Algéria, 18 -20 November 2013.(Accepted paper)
- [215] Y. Fu and T.-S. Huang, "*Human age estimation with regression on discriminative aging manifold*", IEEE Transactions on Multimedia, vol. 10, no. 4, pp. 578–584, 2008.
- [216] X. Geng, Z.-H. Zhou, and K. Smith Miles, "*Automatic age estimation based on facial aging patterns*", IEEE Transactions on pattern analysis and machine intelligence, vol. 29, no. 12, pp. 2234–2240, 2007.
- [217] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "*Human age estimation using bio-inspired features*", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 112-119, 2009.
- [218] A. Lanitis, C. Draganova, and C. Christodoulou, "*Comparing different classifiers for automatic age estimation*", IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),, vol. 34, no. 1, pp. 621–628, 2004.
- [219] J. Wang, Y. Shang, G. Su, and X. Lin, "*Age simulation for face recognition*", in 18th International Conference on Pattern Recognition, ICPR'06, vol. 3, pp. 913–916, 2006.
- [220] A. Lanitis, C.-J. Taylor, and T. F. Cootes, "*Toward automatic simulation of aging effects on face images*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 442–455, 2002.
- [221] J. Suo, S.-C. Zhu, S. Shan, and X. Chen, "*A compositional and dynamic model for face aging*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 3, pp. 385-401, 2010.
- [222] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S.M. Seitz, "*Illumination-aware Age Progression*", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3334-3341, 2014.
- [223] H. Yang, D. Huang, Y. Wang, H. Wang, and Y. Tang, , "*Face Aging Effect Simulation using Hidden Factor Analysis Joint Sparse Representation*", IEEE Transactions on Image Processing, vol. 25, no. 6, pp. 2493-2507, 2016.
- [224] A. Lanitis, "*Comparative evaluation of automatic age-progression methodologies*", EURASIP Journal on Advances in Signal Processing, vol. 2008, pp. 1–10, 2008.
- [225] N. Ramanathan and R. Chellappa, "*Modeling age progression in young faces*", in IEEE Conference on Computer Vision and Pattern Recognition, CVPR, vol. 1, pp. 387–394, 2006.
- [226] N. Ramanathan and R. Chellappa, "*Modeling shape and textural variations in aging faces*", in the 8th IEEE International Conference on Face & Gesture Recognition, FG'08, pp. 1-8, 2008.

- [227] N. Ramanathan, R. Chellappa and S. Biswas, "**Age Progression in Human Face - a Survey**", Journal of Visual Languages and Computing, vol. 15, pp. 3349–3361, 2009.
- [228] U. Park, Y. Tong, and A.-K. Jain. "**Age-Invariant Face Recognition**", IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 5, pp. 947-954, 2010.
- [229] J.-T. Todd, L.-S. Mark, R.-E. Shaw, and J.-B. Pittenger, "**The perception of human growth**", Scientific American, vol. 242, no. 2, pp. 132–144, 1980.
- [230] L.-G. Farkas. "**Anthropometry of the Head and Face**", Raven Press, New York, 1994.
- [231] L.-S. Mark, J.-T. Todd, and R.-E. Shaw, "**Perception of growth : A geometric analysis of how different styles of change are distinguished**", Journal of Experimental Psychology : Human Perception and Performance, vol. 7, no. 4, pp. 855–868, 1981.
- [232] L.-G. Farkas and I.-R. Munro, "**Anthropometric Facial Proportions in Medicine**", Springfield: Charles C. Thomas, 1987.
- [233] K.Waters. "**A muscle model for animating three-dimensional facial expression**", Acm siggraph computer graphics, vol. 21, no. 4, pp.17–24, 1987.
- [234] J.-B. Pittenger and R.-E. Shaw. "**Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception**", Journal of Experimental Psychology: Human perception and performance, vol. 1, no. 4, pp.374-382, 1975.
- [235] H.-F. Chen, P.-N. Belhumeur and D.-W. Jacobs. "**In search of Illumination Invariants**", in proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 254-261, 2000.
- [236] T. Igarashi, K. Nishino, and S.-K. Nayar. "**The Appearance of Human Skin**", Technical Report CUCS-024-05, Dept. of Computer Science, Columbia University, 2005.
- [237] N. Tsumura, H. Haneishi, and Y. Miyake. "**Independent component analysis of skin color image**", Journal of Optical Society of America A, vol. 16, no. 9, pp.2169-2176, 1999.
- [238] H. Ling, S. Soatto, N. Ramanathan, and D. Jacobs, "**Face Verification Across Age Progression Using Discriminative Methods**", IEEE Transactions on Information Forensics and Security, vol. 5, no. 1, pp. 82-91, 2010.
- [239] G. Mahalingam, C. Kambhamettu, "**Face Verification With Aging Using Adaboost and Local Binary Patterns**", in Proceedings of the 7th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP), pp. 101-108, 2010.
- [240] R. Singh, M. Vatsa, A. Noore, , and S.-K. Singh, "**Age transformation for Improving Face Recognition Performance**", in Proceedings of 2nd international Conference on Pattern Recognition and Machine Intelligence, pp. 576-583, 2007.
- [241] S. Biswas, G. Aggarwal, N. Ramanathan, and R. Chellapa, "**A Non-Generative Approach For Face Recognition Across Aging**", in Proceedings of the 2nd International

- IEEE Conference on Biometrics: Theory , Applications and Systems (BTAS), pp. 1–6, 2008.
- [242] G. Mahalingam, and C. Kambhamettu. "**Age invariant face recognition using graph matching**", in Proceedings of the 4th International IEEE Conference on Biometrics: Theory , Applications and Systems (BTAS), pp. 1-7, 2010.
- [243] P.-S. Penev and J.-J. Atick, "**Local feature analysis: A general statistical theory for object representation**", Network: Computation in Neural Systems, vol. 7, no. 3, pp. 477-500, 1996.
- [244] H. Yang, D. Huang, and Y. Wang. "**Age invariant face recognition based on texture embedded discriminative graph model**", in IEEE International Joint Conference on Biometrics (IJCB), pp. 1-8, 2014.
- [245] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, " **Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition**", in the 10th IEEE International Conference on Computer Vision, vol. 1, pp. 786–791, 2005.
- [246] G. Guo, G. Mu and K. Ricanek, "**Cross-age face recognition on a very large database: the performance versus age intervals and improvement using soft biometric traits**", In 20th International Conference on Pattern Recognition (ICPR), pp. 3392-3395, 2010.
- [247] C. Meng, J. Lu and Y.-P. Tan, " **A comparative study of age-invariant face recognition with different feature representations**", in the 11th International IEEE Conference on Control Automation Robotics & Vision (ICARCV), pp. 890-895, 2010.
- [248] M. Bereta, P. Karczmarek, W. Pedrycz, and M. Reformat. "**Local descriptors in application to the aging problem in face recognition**", Pattern Recognition, vol. 46, no. 10, p. 2634-2646, 2013.
- [249] Z. Li, U. Park and A.-K. Jain, "**A discriminative model for age invariant face recognition**", *IEEE Transactions on Information Forensics and Security*, , vol. 6, no. 3, pp. 1028-1037, 2011.
- [250] D. Sungatullina, J. Lu, G. Wang, and P. Moulin, "**Multiview discriminative learning for age-invariant face recognition**", in Proceedings of the 10th International IEEE Conference and Workshops on Automatic Face and Gesture Recognition, pp. 1-6, 2013.
- [251] F. Juefei-Xu, K. Luu, M. Savvides, T.-D. Bui, and C.-Y. Suen, "**Investigating age invariant face recognition based on periocular biometrics**", in IEEE International Joint Conference on Biometrics (IJCB), pp. 1-7, 2011 .
- [252] J.-S. Nayak, M. Indiramma and N. Nagarathna, "**Modeling self-Principal Component Analysis for age invariant face recognition**", in International Conference on Computational Intelligence & Computing Research (ICCIC), pp. 1-5, 2012.

- [253] L. Boussaad, M. Benmohammed, R. Benzid, "***Age Invariant Face Recognition Based on DCT Feature Extraction and Kernel Fisher Analysis***", Journal of Information Processing Systems (JIPS), Vol. 12., No. 3, pp. 392-409, 2016.
- [254] X. Tan and B. Triggs, "***Enhanced local texture feature sets for face recognition under difficult lighting conditions***", IEEE transactions on image processing, vol. 19, no. 6, pp. 1635-1650, 2010.
- [255] V. ˇ Struc, "***Inface: A toolbox for illumination invariant face recognition***", 2009
- [256] T.-F. Cootes and C.-J. Taylor, " ***Statistical Models of Appearance for Computer Vision***", Technical Report, The University of Manchester School of Medicine, 2004.
- [257] I.-L. Dryden and K.-V. Mardia. "***Statistical Shape Analysis***", John Wiley & Sons, 1998.
- [258] M.-B. Stegmann and D.-D. Gomez, " ***A brief introduction to statistical shape analysis***", Informatics and mathematical modelling, Technical University of Denmark, DTU, vol. 15, no. 11, 2002.
- [259] T.-F. Cootes, C.-J. Taylor, D.-H. Cooper and J. Graham, "***Active shape models-their training and application***", Computer vision and image understanding, vol. 61, no. 1, pp. 38-59, 1995.
- [260] S. Le Gallou, "***Détection robuste des éléments faciaux par modèles actifs d'apparence***", Doctoral dissertation, Rennes 1, 2007.
- [261] K.-R. RAO and P. YIP, "***Discrete cosine transform: algorithms, advantages, applications***", Academic Press Professional, Inc. San Diego, CA, USA, 1990.
- [262] W.-B. Pennebaker, and J.-L. Mitchell, "***JPEG: Still Image Data Compression Standard***", Springer Science & Business Media, 1993.
- [263] W. Chen, E.-r. Meng Joo, and S. Wu, "***Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain***", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 36, no. 2, pp. 458-466, 2006.
- [264] Y. Wang and Q. Ruan, "***Kernel fisher discriminant analysis for palmprint recognition***", in the 18th International Conference on Pattern Recognition, pp. 457-460, 2006.
- [265] C. Liu, "***Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance***," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 5, pp. 725-737, 2006.
- [266] Q. Liu, H. Lu and S. Ma, "***Improving kernel Fisher discriminant analysis for face recognition***". IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 42-49, 2004.

- [267] M. Bengherabi, L. Mezai, F. Harizi, A. Guessoum and M. Cheriet, "***Fusion de la DCT-PCA et la DCT-RLDA pour la Reconnaissance de Visages***".
- [268] K. Jonsson, J. Matas, J. Kittler and Y.-P. Li, "***Learning support vectors for face verification and recognition***", in the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 208-213, 2000.
- [269] G. Panis, A. Lanitis, N. Tsapatsoulis and T.-F. Cootes, "***Overview of research on facial ageing using the FG-NET ageing database***", *IET Biometrics*, vol. 5, no. 2, pp. 37-46, 2016.
- [270] V. Štruc, and N. Pavešić, "***The Complete Gabor-Fisher Classifier for Robust Face Recognition***", *EURASIP Journal on Advances in Signal Processing*, Special issue on advanced image processing for defense and security applications, vol. 2010, 2010.
- [271] V. Štruc, and N. Pavešić, "***Gabor-Based Kernel Partial-Least-squares Discrimination Features for Face Recognition***", *Informatica*, vol. 20, no. 1, pp. 115-138, 2009.
- [272] L. Boussaad and M. Benmohammed, "***Effects of Aging on Face Recognition Algorithms: The Performance versus Age Intervals***", 5th International Conference on Software Engineering and New Technologies (ICSSENT'2016), Istanbul, Turkey, (Article accepté)
- [273] U. Park, Y. Tong, and A.-K. Jain, "***Face recognition with temporal invariance: A 3d aging model***", in Proceedings of the 8th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 1-7, 2008.

Références personnelles

- [1] L. Boussaad, M. Benmohammed, "*Pondération dynamiques de descripteurs par essaim de particules pour la recherche d'images par le contenu*", Conférence nationale sur la modélisation et la simulation informatique des systèmes industriels, MSISI' 2010, Skikda, Algérie,2010.
- [2] L. Boussaad, M. Benmohammed, R. Benzid , "*a comparative study of different subspace methods and metrics for age-invariant face recognition*", 2nd International Conference on Image, Vision and their Applications, Guelma, Algéria, 18 -20 November 2013.(Article accepté)
- [3] L. Boussaad, M. Benmohammed and R. Benzid, "*Face Recognition In The Presence Of Age Differences Using Wavelet Transform*", Conférence internationale sur le traitement de l'information multimédia (CITIM'2015), Mascara, Algéria, 12 -13 Mai 2015 (Article accepté + publié sur proceeding de la conférence).
- [4] L. Boussaad, M. Benmohammed and R. Benzid, "*Effects of Aging on Face Recognition Algorithms: The Performance versus Age Intervals*", 5th International Conference on Software Engineering and New Technologies (ICSSENT'2016), Istanbul, Turkey, (Article accepté).
- [5] L. Boussaad, M. Benmohammed, and R. Benzid, "*Age Invariant Face Recognition Based on DCT Feature Extraction and Kernel Fisher Analysis*", Journal of Information Processing Systems, vol. 12, no. 3, pp. 392-409, 2016.