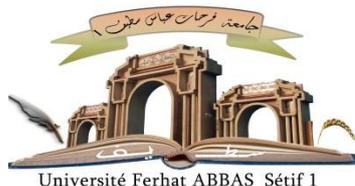


الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

Ministère de L'Enseignement Supérieur et de la Recherche Scientifique



UNIVERSITÉ FERHAT ABBAS - SETIF1

FACULTÉ DE TECHNOLOGIE

THÈSE

Présentée au Département d'électronique

Pour l'obtention du diplôme de

DOCTORAT

Domaine : Sciences et Technologie

Filière : Télécommunications

**Option : Réseaux et
Télécommunications**

Par

Mr. Alane Badreddine

THÈME

**Techniques et systèmes de reconnaissance
automatique de visages**

Soutenue le 23 / 06 / 2024 devant le Jury:

Zegadi Ameer	Professeur	Univ. Ferhat Abbas Sétif 1	Président
Bouguezel Saad	Professeur	Univ. Ferhat Abbas Sétif 1	Directeur de thèse
Terchi Younes	M.C.B.	Univ. Ferhat Abbas Sétif 1	Co-Directeur
Daachi Mohamed El Hossine	M.C.A.	Univ. Mohamed El Bachir El Ibrahimi BBA	Examineur
Amardjia Nourredine	Professeur	Univ. Ferhat Abbas Sétif 1	Examineur
Ziet Lahcene	Professeur	Univ. Ferhat Abbas Sétif 1	Examineur

Remerciements

Tout d'abord, je ne remercie personne avant de remercier Allah seul pour ce qu'il m'a aidé à faire, car cela n'aurait pas été réalisé sans sa grâce et sa gratitude, et je lui demande de faire de nous des clés de la bonté et de faire bénéficier les gens de nous et de faire de ceci un travail utile sincèrement pour la face du Tout-Puissant.

Les recherches exposées dans cette thèse font partie des projets menés au sein du Laboratoire de Croissance et Caractérisation de Nouveaux Semiconducteurs (LCCNS), dirigé par le Professeur Zouaoui Ahmed à l'Université Ferhat Abbas Sétif 1.

Je souhaite exprimer ma profonde gratitude envers mon directeur de thèse, le professeur Bouguzel Saad, du département d'électronique de l'Université Ferhat Abbas Sétif-1, pour ses conseils précieux et ses encouragements constants tout au long de la réalisation de ce travail. Je lui suis reconnaissant pour la confiance qu'il m'a accordée et pour ses conseils éclairés, qui ont enrichi ma réflexion.

Je tiens à exprimer ma gratitude envers mon co-encadrant de thèse, le docteur Terchi Younes, du département d'électronique de l'Université Ferhat Abbas Sétif-1, pour son soutien et ses conseils précieux. Ses explications ont grandement contribué à l'élaboration de la méthode de recherche adoptée dans ce travail.

Je souhaite exprimer ma gratitude envers monsieur Zegadi Ameer, professeur au département d'électronique de l'Université Ferhat Abbas Sétif-1, pour avoir présidé le jury de ma thèse.

Je souhaite exprimer ma reconnaissance envers monsieur Daachi Mohamed El Hossine, maître de conférences de classe A du département d'électronique de l'université Mohamed El Bachir El Ibrahimi BBA ainsi que messieurs Amardjia Nourredine et Ziet Lahcene, professeurs au département d'électronique de l'Université Ferhat Abbas Sétif-1. Leur participation en tant que membres du jury et leurs précieuses remarques ont grandement enrichi ce travail.

Je tiens à exprimer ma gratitude envers monsieur El-Bay Bourennane, professeur à l'Université de Bourgogne-Dijon, pour son accompagnement, ses explications claires et sa

générosité lors de mon stage de perfectionnement au sein du laboratoire ImVia à l'Université de Bourgogne-Dijon.

Je souhaite exprimer ma sincère reconnaissance envers messieurs Zebiri Chemseddine et Mayouf Abdelhalim, Professeurs à l'Université Ferhat Abbas Sétif-1, pour leurs encouragements et leur bienfait dont j'ai bénéficié.

Enfin, j'exprime mes vifs remerciements à mes chers parents, ma femme, ma petite fille, mon frère, mes sœurs, mes amis et tous les membres de ma famille, qui m'ont encouragé et aidé dans les périodes difficiles et pénibles durant toute la période de ma formation.

Dédicace

Je dédie ce modeste travail aux étoiles de ma vie : ma mère Merbouha et mon père,
Boubaker.

À ma femme, ma partenaire de vie, Cherifa, ainsi qu'à mon adorable petite princesse,
Yakine. À mon frère Nazim, mes sœurs Imene et Ines, ainsi qu'à ma tante Hadda, et à tous les
membres de ma famille.

À tous mes amis, les hommes qui m'ont toujours soutenu, et à tous ceux qui m'aiment.

Badreddine

INTRODUCTION 1

CHAPITRE I 6

1.1 INTRODUCTION..... 7
1.2 SYSTEME DE RECONNAISSANCE DES VISAGES EN GENERAL..... 7
1.2.1 ACQUISITION ET PRETRAITEMENT DES IMAGES POUR LA RECONNAISSANCE FACIALE 7
1.2.2 DETECTION DE VISAGE 8
1.2.3 EXTRACTION DES CARACTERISTIQUES 8
1.2.4 CLASSIFICATION..... 9
1.2.5 MISE EN CORRESPONDANCE DES CARACTERISTIQUES 9
1.3 BASES DE DONNEES DE REFERENCE CONNUES 9
1.4 PRINCIPAUX DEFIS POUR LA RECONNAISSANCE FACIALE..... 11
1.4.1 VARIATION DE POSE 11
1.4.2 VARIATION D'ECLAIRAGE 11
1.4.3 L'OCCLUSION 12
1.4.4 VIEILLISSEMENT 12
1.5 APPROCHES EN RECONNAISSANCE FACIALE..... 12
1.5.1 APPROCHES CLASSIQUES 13
1.5.2 RECONNAISSANCE DES VISAGES BASEE SUR L'APPRENTISSAGE PROFOND..... 15
1.5.3 LES APPROCHES HYBRIDES..... 16
1.6 RECONNAISSANCE DES VISAGES A L'AIDE DE LA PSYCHOLOGIE COGNITIVE 20
1.7 CONCLUSION 23

CHAPITRE II 24

2.1 INTRODUCTION..... 25
2.2 METHODES BASEES SUR LES SOUS-ESPACES 26
2.2.1 ANALYSE EN COMPOSANTES PRINCIPALES 26
2.2.2 ANALYSE DISCRIMINANTE LINEAIRE DE FISHER 31
2.3 TRANSFORMEES ET ANALYSE MULTI-RESOLUTION 36
2.3.1 TRANSFORMEE DISCRETE EN COSINUS..... 37
2.3.2 DECOMPOSITION EN ONDELETTES 38
2.3.3 DECOMPOSITION EN DCT-PYRAMID..... 41
2.4 RESEAUX DE NEURONES ARTIFICIELS..... 43
2.4.1 NEURONES ARTIFICIELS ET NEURONES BIOLOGIQUES 44
2.4.2 APPRENTISSAGE DES RESEAUX NEURONAUX ARTIFICIELS 46

2.4.3 TYPES DE RESEAUX NEURONAUX ARTIFICIELS	46
2.5 CONCLUSION	47
<u>CHAPITRE III</u>	<u>49</u>
3.1 INTRODUCTION.....	50
3.2 SYSTEME DE RECONNAISSANCE AUTOMATIQUE DES VISAGES PREPOSE	52
3.2.1 PRETRAITEMENT	52
3.2.2 EXTRACTION DES CARACTERISTIQUES	56
3.2.3 CLASSIFICATION :	61
3.3 CONCLUSION	63
<u>CHAPITRE IV</u>	<u>65</u>
4.1 INTRODUCTION.....	66
4.2 CRITERES D'EVALUATION DES SYSTEMES DE RECONNAISSANCE FACIALE.....	66
4.2.1 AJUSTEMENT EXCESSIF.....	67
4.2.2 ASPECTS DIVERS DES MESURES D'EVALUATION	68
4.2.3 ÉVALUATION DU SYSTEME DE RECONNAISSANCE DES VISAGES	74
4.3 ÉVALUATION DES PERFORMANCES DU SYSTEME DE RECONNAISSANCE FACIALE ..	75
4.3.1 EXPERIENCES MENEES SUR ORL	76
4.3.2 EXPERIENCES MENEES SUR YALE.....	81
4.3.3 EXPERIENCES MENEES SUR AR	84
4.3.4 COMPARAISON A L'AIDE DU TAUX DE RECONNAISSANCE MOYEN	87
4.4 CONCLUSION	90
<u>CONCLUSION</u>	<u>91</u>

CHAPITRE 1	
FIGURE 1.1 : DESCRIPTION D'UN SYSTEME DE RECONNAISSANCE AUTOMATIQUE DU VISAGE.	9
FIGURE 1.2 : VISAGES AVEC EFFET D'OCCLUSION TELS QUE PRESENTES DANS AR [7].	13
FIGURE 1.3 : REPARTITION DES METHODES DE RECONNAISSANCE DES VISAGES PAR CATEGORIES.	14
CHAPITRE 2	
FIGURE 2.1 : EXEMPLE D'EXTRACTION DE CARACTERISTIQUES A L'AIDE DE LA PCA.	26
FIGURE 2.2 : DONNEES COMPOSEES DE DEUX CLASSES REPRESENTÉES DANS UN FORMAT BIDIMENSIONNEL.	34
FIGURE 2.3 : DONNEES SOUSTRAITES DE LA MOYENNE DE DEUX CLASSES DIFFÉRENTES.	34
FIGURE 2.4 : RECONSTRUCTION DES DONNEES AVEC TOUS LES VECTEURS PROPRES (EIGENVECTORS).	35
FIGURE 2.5 : RECONSTRUCTION DES DONNEES AVEC UN VECTEUR PROPRE CORRESPONDANT A LA VARIANCE MAXIMALE.	35
FIGURE 2.6 : DONNEES DANS UN ESPACE TRANSFORME OU LE SYSTEME DE COORDONNEES EST BASE SUR LES VECTEURS PROPRES (EIGENVECTORS).	36
FIGURE 2.7 : DONNEES NON LINEAIREMENT SEPARABLES DANS UN ESPACE REDUIT.	36
FIGURE 2.8 : REPRESENTATION D'UNE TRANSFORMATION EN DCT D'UNE IMAGE FACIALE DANS AR [7].	39
FIGURE 2.9 : FONCTIONS DE BASE DE LA TRANSFORMEE EN ONDELETTES.	41
FIGURE 2.10 : A) UNE IMAGE DE VISAGE. (B) L'IMAGE DU VISAGE SOUMISE A UNE DECOMPOSITION EN ONDELETTES 2D. (C) L'IMAGE DU VISAGE SOUMISE A UNE DECOMPOSITION EN PAQUETS D'ONDELETTES 2D.	41
FIGURE 2.11 : DECOMPOSITION D'UNE PYRAMIDE DCT D'UN NIVEAU.	43
FIGURE 2.12 : ARCHITECTURE DES RESEAUX DE NEURONES.	44
FIGURE 2.13 : COMPARAISON DES NEURONES BIOLOGIQUES PAR RAPPORT AUX NEURONES ARTIFICIELS.	47

CHAPITRE 3	
FIGURE 3.1 : SCHEMA DESCRIPTIF DU SYSTEME CONÇU.	52
FIGURE 3.2 : TYPES DE CARACTERISTIQUES HAAR.	54
FIGURE 3.3 : ILLUSTRATION D'UN ALGORITHME POUR LE BOOSTING.	56
FIGURE 3.4 : RESEAU NEURONAL DE RETROPROPAGATION A DEUX COUCHES.	64
CHAPITRE 4	
FIGURE 4.1 : EXEMPLES D'IMAGES DE L'ENSEMBLE DE DONNEES ORL.	77
FIGURE 4.2 : HISTOGRAMME D'ERREUR DU RESEAU NEURONAL POUR LA DECOMPOSITION ORL 50%50.	80
FIGURE 4.3 : ETAT D'ENTRAINEMENT DU RESEAU NEURONAL POUR LA DECOMPOSITION ORL 50%50.	80
FIGURE 4.4 : ROC POUR LA DECOMPOSITION ORL 50%50.	81
FIGURE 4.5 : ÉCHANTILLONS DE L'ENSEMBLE DE DONNEES DE YALE.	82
FIGURE 4.6 : HISTOGRAMME D'ERREUR DU RESEAU NEURONAL POUR LA DECOMPOSITION YALE 50%50.	84
FIGURE 4.7 : ETAT D'ENTRAINEMENT DU RESEAU NEURONAL POUR LA DECOMPOSITION YALE 50%50.	84
FIGURE 4.8 : ROC POUR LA DECOMPOSITION YALE 50%50.	85
FIGURE 4.9 : HISTOGRAMME D'ERREUR DU RESEAU NEURONAL POUR LA DECOMPOSITION AR 70%30.	87
FIGURE 4.10 : ETAT D'ENTRAINEMENT DU RESEAU NEURONAL POUR LA DECOMPOSITION AR 70%30.	87
FIGURE 4.11 : ROC POUR LA DECOMPOSITION AR 70%30.	88

CHAPITRE 2	
TABLEAU 1.1 : COMPARAISON ENTRE NEURONES BIOLOGIQUES ET NEURONES ARTIFICIELS.	45
TABLEAU 1.2 : ANALOGIES ENTRE NEURONES BIOLOGIQUES ET ARTIFICIELS DANS LE PROCESSUS D'APPRENTISSAGE.	46
CHAPITRE 4	
TABLEAU 4.1 : MATRICE DE CONFUSION.	70
TABLEAU 4.2 : TAUX DE RECONNAISSANCE (%) SUR L'ENSEMBLE DES DONNÉES ORL EN UTILISANT DIFFÉRENTS TAUX DE DÉCOMPOSITION.	78
TABLEAU 4.3 : TAUX DE RECONNAISSANCE (%) SUR L'ENSEMBLE DES DONNÉES ORL EN UTILISANT DIFFÉRENTS TAUX DE DÉCOMPOSITION.	78
TABLEAU 4.4 : TAUX DE RECONNAISSANCE (%) EN UTILISANT DIFFÉRENTS TAUX DE DÉCOMPOSITION SUR L'ENSEMBLE DES DONNÉES DE YALE.	83
TABLEAU 4.5 : TAUX DE RECONNAISSANCE OBTENUS PAR DIFFÉRENTES MÉTHODES EN UTILISANT L'ENSEMBLE DE DONNÉES DE YALE.	83
TABLEAU 4.6 : TAUX DE RECONNAISSANCE (%) SUR AR EN UTILISANT DIFFÉRENTS TAUX DE DÉCOMPOSITION.	86
TABLEAU 4.7 : TAUX DE RECONNAISSANCE (%) OBTENUS PAR DIFFÉRENTES MÉTHODES À PARTIR DE AR.	86
TABLEAU 4.8 : TAUX DE RECONNAISSANCE MOYENS (%) OBTENUS PAR DIFFÉRENTES MÉTHODES EN UTILISANT L'ENSEMBLE DE DONNÉES ORL.	89
TABLEAU 4.9 : TAUX DE RECONNAISSANCE MOYENS (%) OBTENUS PAR LA MÉTHODE PROPOSÉE ET LES MÉTHODES PROPOSÉES DANS [50] ET [27] EN UTILISANT L'ENSEMBLE DE DONNÉES DE YALE.	90
TABLEAU 4.10 : TAUX DE RECONNAISSANCE MOYENS (%) OBTENUS PAR LA MÉTHODE PROPOSÉE ET LA MÉTHODE PROPOSÉE DANS [39] EN UTILISANT L'ENSEMBLE DE DONNÉES DE YALE.	90
TABLEAU 4.11 : TAUX DE RECONNAISSANCE MOYENS (%) OBTENUS PAR LA MÉTHODE PROPOSÉE ET LES MÉTHODES PROPOSÉES DANS [27] ET [51] EN UTILISANT L'ENSEMBLE DE DONNÉES AR.	90

DCT: Discrete Cosine Transform.

SIFT: Scale-Invariant Feature Transform.

SURF: Speed Up Robust Feature.

SVM: Support Vector Machine.

RF: Random Forest.

DT: Decision Table.

PCA: Principal Component Analysis.

KPCA: Kernel Principal Component Analysis.

ANN: Artificial Neural Networks.

CNN: Convolutional Neural Networks.

MDS: Multidimensional Scaling.

LDA: Linear Discriminant Analysis.

FLD: Fisher linear discriminant.

FLDA: Fishers linear discriminant analysis

SURF: Speeded Up Robust Features.

LBP: Local Binary Pattern.

HOG: Histogram of Oriented Gradients.

PSO: Particle Swarm Optimization.

BPNN: Back-Propagation Neural Network.

LBPH: Local Binary Patterns Histograms.

PNN: Probabilistic Neural Network.

IKLDA: Improved Kernel LDA.

IDCT: Inverse Discrete Cosine Transform.

RBF: Radial Basis Functions.

RNN: Recurrent Neural Network.

RGMS: Relative Gradient Magnitude Strength.

DNN: Deep Neural Networks.

ROC: Receiver Operating Characteristic.

PR: Precision and Recall.

TP: True Positive.

TN: True Negative.

FP: False Positive.

FN: False Negative.

TPR: True Positive Rate.

TNR: True Negative Rate.

FMR: False Match Rate.

FNMR: False Non-Match Rate.

GAR: Genuine Acceptance Rate.

FAR: False Acceptance Rate.

FRR: False Rejection Rate.

EER: Equal Error Rate.

CER: Crossover Error Rate.

FTE: Failure to Enroll.

FER: Failure Enroll Rate.

FTC: Failure to Capture.

تقترح هذه الأطروحة نظامًا جديدًا للتعرف على الوجوه يتضمن طريقة هجينة تعتمد على هرم تحويل جيب التمام المنفصل (DCT) لاستخراج السمات، ومقاييس إحصائية لتقليل أبعاد السمات، وشبكة عصبية ثنائية الطبقة للترحيل الخلفي للتصنيف. يتم استخدام هرم DCT لالتقاط المعلومات ذات التردد المنخفض والعالي بكفاءة من صور الوجه، مما يحسن قدرة النظام على التعرف على الوجوه بدقة. وفي الوقت نفسه، يساعد إدخال المقاييس الإحصائية لتقليل الأبعاد على تقليل التعقيد الحسابي ويوفر تمييزًا أفضل، مما يؤدي إلى معالجة أكثر كفاءة. بالإضافة إلى ذلك، تعمل الشبكة العصبية ثنائية الطبقة، التي تلعب دورًا أساسيًا في الإدارة الفعالة للأنماط المعقدة، على تعزيز قدرات النظام في التعرف على الوجوه. ويفضل هذه التطورات، حقق النظام معدل تعرف استثنائي بلغ 99% على مجموعة بيانات مختبر أوليفيتي للأبحاث (ORL)، و98.88% على YALE، و99.16% على AR. يوضح هذا الأداء متانة وإمكانات النظام المقترح لتطبيقات التعرف على الوجه الحقيقية.

الكلمات المفتاحية: التعرف على الوجه، التحويلات، الشبكات العصبية، هرم DCT، القياسات الإحصائية

Résumé :

Cette thèse propose un nouveau système de reconnaissance faciale intégrant une méthode hybride basée sur la pyramide de transformation en cosinus discrète (DCT) pour l'extraction des caractéristiques, des mesures statistiques pour la réduction de la dimensionnalité des caractéristiques, et un réseau de neurones à rétropropagation à deux couches pour la classification. La pyramide DCT est utilisée pour capturer efficacement les informations de basse et haute fréquence des images faciales, améliorant ainsi la capacité du système à reconnaître les visages avec précision. Parallèlement, l'introduction de mesures statistiques pour la réduction de la dimensionnalité aide à diminuer la complexité computationnelle et offre une meilleure discrimination, conduisant à un traitement plus efficace. De plus, le réseau de neurones à deux couches, qui joue un rôle essentiel dans la gestion efficace des motifs complexes, améliore davantage les capacités de reconnaissance du système. Grâce à ces avancées, le système atteint un taux de reconnaissance exceptionnel de 99 % sur l'ensemble de données du Laboratoire de Recherche Olivetti (ORL), 98,88 % sur YALE, et 99,16 % sur AR. Cette performance démontre la robustesse et le potentiel du système proposé pour des applications réelles en reconnaissance faciale.

Mots clés : Reconnaissance faciale, Transformées, Réseaux de neurones à rétropropagation, DCT-Pyramid, Mesures statistiques

Abstract:

This thesis proposes a new face recognition system incorporating a hybrid method based on the Discrete Cosine Transform (DCT) pyramid for feature extraction, statistical measures for feature dimensionality reduction, and a two-layer back-propagation neural network for classification. The DCT pyramid is used to efficiently capture low- and high-frequency information from facial images, improving the system's ability to recognize faces accurately. At the same time, the introduction of statistical measures for dimensionality reduction helps to reduce computational complexity and offers better discrimination, leading to more efficient processing. In addition, the two-layer neural network, which plays a key role in the efficient management of complex patterns, further enhances the system's recognition capabilities. Thanks to these advances, the system achieves an outstanding recognition rate of 99% on the Olivetti Research Laboratory (ORL) dataset, 98.88% on YALE, and 99.16% on AR. This performance demonstrates the robustness and potential of the proposed system for real facial recognition applications.

Key words: Face recognition, Transforms, Backpropagation neural networks, DCT-Pyramid, Statistical measurements

Introduction

Introduction

La reconnaissance faciale est devenue un pilier incontournable de notre société contemporaine, trouvant des applications variées allant de la sécurité à l'interaction quotidienne avec la technologie. La nécessité d'identifier et de vérifier rapidement les individus, que ce soit pour des raisons de sécurité, de gestion des données ou d'expérience utilisateur optimisée, a suscité un intérêt croissant pour le développement de systèmes de reconnaissance faciale toujours plus performants.

Takeo Kanade a créé le premier système de reconnaissance faciale automatique, qu'il a décrit dans sa thèse de doctorat en 1973 [1]. Il s'avère qu'aux premiers stades de la recherche, les systèmes de reconnaissance faciale offraient les avantages d'une approche physiologique sans être invasifs. Malheureusement, l'efficacité et la précision requises ne pouvaient être atteintes à l'époque en raison des limites de la capacité de traitement des grands ensembles de données sur les visages et des changements dans les traits du visage, même pour une seule personne. Ainsi, jusqu'aux travaux de Kirby et Sirovich [1-5] en 1990 sur la représentation des visages à faible dimension, générée à partir de la transformée de Karhunen-Loeve, l'activité est restée en sommeil. En outre, la recherche sur la reconnaissance des visages a été relancée par les travaux révolutionnaires de Turk et Pentland sur la surface propre [1-5] et, dans les décennies qui ont suivi, une grande variété d'algorithmes de reconnaissance automatique du visage a vu le jour.

Différents scénarios d'application peuvent être désignés par l'expression générale "reconnaissance faciale". Il existe deux types de scénarios : l'identification ou la reconnaissance et l'authentification ou la vérification. Dans les deux cas, le système commence par des photos de visages de personnes qu'il connaît déjà. Le terme "galerie" est parfois utilisé pour décrire ce groupe de personnes. Les photos suivantes de cette personne ou d'autres individus servent de sondes à comparer avec les photos de la galerie. Dans le cadre d'un scénario de reconnaissance, il y a correspondance d'un à plusieurs : une sonde est comparée à chaque image de la galerie afin d'identifier la meilleure correspondance au-delà d'un seuil prédéterminé. Lorsqu'une identité revendiquée est comparée à l'entrée de la galerie dans une situation d'authentification, la correspondance est biunivoque ; si la qualité de la correspondance dépasse un niveau prédéterminé, l'identité revendiquée est considérée comme authentique. D'un point de vue technique, le cas d'authentification est plus facile que le scénario de reconnaissance. Une galerie plus large dans un scénario de reconnaissance tend à donner plus de possibilités de reconnaissance inexacte, ce qui est l'une des raisons. Une autre

raison est qu'à chaque tentative de reconnaissance, l'ensemble de la galerie doit être fouillé d'une manière ou d'une autre. Dans un troisième scénario, le sujet testé peut se trouver ou non dans la base de données du système. Un score est généré en comparant l'image du visage de la requête à toutes les images de visage de la base de données. Une alarme de reconnaissance peut être déclenchée si le score est supérieur au seuil spécifié.

Les méthodes de reconnaissance automatique du visage peuvent être classées en deux opérations interconnectées : la détection des visages et la reconnaissance des visages [1-6] Afin de localiser et de séparer la région du visage de l'arrière-plan, la détection des visages est une première étape essentielle. Le défi est relevé en extrayant les visages et peut-être les traits du visage d'un arrière-plan incontrôlé à l'aide de la segmentation. La vérification et l'identification sont nécessaires pour les opérations de reconnaissance des visages. Dans cette étape, une base de données de visages reconnus précédemment enregistrés est comparée à l'image d'exploration qui a été prise de la scène au cours de l'étape de détection des visages. L'étape suivante consiste à trouver le visage le plus susceptible de correspondre en recherchant les images les plus proches.

L'identification et la vérification constituent la dernière phase de la reconnaissance faciale. L'identification consiste à comparer un visage à un ensemble de deux ou plusieurs visages afin de trouver la correspondance la plus probable. La procédure de vérification des visages consiste à comparer le visage testé à un autre visage reconnu dans la base de données. Sur la base de cette comparaison, le visage est soit accepté comme étant celui du client, soit rejeté comme étant un imposteur. Une image d'interrogation est une image qui doit être vérifiée ou identifiée. Les images de la galerie sont celles qui sont comparées à l'image d'interrogation.

Depuis l'introduction de l'approche Eigenface, la technologie de la reconnaissance faciale a considérablement progressé [2-5]. La reconnaissance automatisée des visages peut donner des taux de reconnaissance élevés dans des scénarios limités, tels que ceux où l'éclairage, la position, la position debout, les vêtements et l'expression du visage peuvent être contrôlés. Cela est particulièrement vrai lorsque la base de données (galerie) contient un grand nombre, voire un nombre modeste, d'images de visages. Cependant, le taux d'erreur déclaré est important même dans des situations d'imagerie très contrôlées, comme les images d'identité [2-4]. Les performances se dégradent encore plus dans des conditions non contrôlées [2-6]. Il peut être difficile de formuler une tâche qui nécessite une formation au sein d'un système dans des conditions d'imagerie spécifiques (telles qu'un éclairage, une position et un modèle de

mouvement uniques) et d'être capable d'effectuer une reconnaissance en cas de changements arbitraires de ces paramètres.

L'un des inconvénients typiques de ces systèmes est qu'ils nécessitent un processus rigide et fastidieux pour recueillir les données de formation. Il s'agit de recueillir un certain nombre d'images prédéterminées pour chaque participant et de les étiqueter en conséquence. Bien que les performances des systèmes disponibles dans le commerce soient acceptables, il n'est pas certain que le visage seul, sans aucune autre information, constitue une base satisfaisante pour identifier avec précision une personne parmi un grand nombre d'individus avec un niveau de certitude très élevé. Reconnaître un visage à partir de photos prises de deux points de vue très différents est une tâche difficile. En outre, les systèmes de reconnaissance faciale existants imposent diverses limites à l'acquisition d'images faciales, nécessitant souvent un arrière-plan uni ou des conditions d'éclairage spécifiques.

Pour que les systèmes de reconnaissance faciale soient largement acceptés, ils doivent être capables de détecter automatiquement la présence d'un visage dans l'image capturée, de localiser avec précision le visage s'il existe, et de reconnaître avec succès le visage dans une perspective large. La principale difficulté de la reconnaissance des visages basée sur la vision réside dans la variété considérable observée dans les images faciales humaines. Les variations intra-sujets peuvent potentiellement être importantes et résulter de facteurs tels que la posture de la tête en 3D, l'éclairage, l'expression du visage, la pilosité faciale et le vieillissement. En revanche, les variations inter-sujets tendent à être minimales, car les individus ont souvent des caractéristiques similaires.

Globalement, le processus d'identification et de reconnaissance des visages est influencé par divers facteurs tels que les composants externes et internes du visage, le caractère distinctif de ces composants, leur configuration et la texture locale. D'autre part, les personnes sont capables de combiner sans effort et d'effectuer de manière autonome des activités liées à la détection et à la reconnaissance basées sur l'apparence et la géométrie avec le plus haut niveau d'efficacité.

Dans ce contexte, cette thèse se concentre sur la conception et l'implémentation d'un système de reconnaissance faciale novateur, cherchant à combiner judicieusement plusieurs techniques de pointe pour obtenir une précision et une robustesse accrues. Contrairement aux approches conventionnelles qui se limitent souvent à l'extraction de caractéristiques basées sur des méthodes spécifiques, notre système adopte une approche plus holistique en combinant

plusieurs méthodes pour exploiter au mieux les informations disponibles dans les images faciales.

La première composante clé de notre approche est la pyramide de transformée en cosinus discret (DCT : Discrete Cosine Transform), une technique bien établie dans le domaine du traitement d'images. La DCT permet de représenter efficacement les caractéristiques du visage en capturant à la fois les informations de basses et hautes fréquences, ce qui permet une représentation plus robuste et discriminative des visages. Cette représentation multi-échelle est cruciale pour la capacité du système à reconnaître les visages sous des conditions variables telles que l'éclairage et l'angle de vue.

En complément de la DCT, notre approche intègre des mesures statistiques pour la réduction de la dimensionnalité des caractéristiques extraites. Cette étape est essentielle pour réduire la complexité computationnelle du système tout en préservant les informations les plus discriminantes. En réduisant la dimensionnalité de l'espace des caractéristiques, notre système devient plus efficace en termes de temps de calcul tout en maintenant des performances élevées en termes de précision de reconnaissance.

Enfin, la troisième composante majeure de notre système est un réseau de neurones à deux couches avec rétropropagation. Ce type de réseau neuronal est choisi pour sa capacité à apprendre des caractéristiques hautement abstraites et à généraliser à partir des données d'entraînement. En exploitant la puissance de calcul et la flexibilité offertes par les réseaux de neurones, notre système est capable d'effectuer une classification précise des visages, même en présence de variations subtiles dans les conditions d'acquisition des images.

Afin de mieux comprendre ces capacités, le premier chapitre passe en revue les développements récents dans le domaine de la reconnaissance faciale, en mettant en lumière les différentes approches, les méthodes traditionnelles ainsi que les avancées les plus récentes. Il examine les principaux défis rencontrés et présente les tendances actuelles de la recherche.

Le deuxième chapitre explore les concepts fondamentaux des transformées et des réseaux de neurones, en mettant en évidence leur importance dans le domaine de la reconnaissance faciale. Il examine les différentes techniques de transformées utilisées pour l'extraction de caractéristiques et les principes de fonctionnement des réseaux de neurones utilisés dans ce contexte.

Le troisième chapitre décrit en détail le système de reconnaissance faciale proposé, qui intègre la pyramide DCT pour l'extraction des caractéristiques et les réseaux de neurones pour

la classification. Il explique comment ces deux composantes sont combinées de manière synergique pour améliorer la précision et la robustesse du système de reconnaissance faciale.

Le quatrième chapitre propose une évaluation détaillée du système de reconnaissance faciale développé, en utilisant les ensembles de données ORL, Yale et AR, largement reconnus et utilisés dans le domaine de la recherche. Il analyse les performances du système en termes de précision et d'autres métriques pertinentes, et compare ses résultats avec ceux des méthodes existantes pour évaluer son efficacité et sa fiabilité.

En résumé, cette thèse présente une approche intégrée et novatrice pour la reconnaissance faciale, qui combine la pyramide DCT pour l'extraction des caractéristiques, des mesures statistiques pour la réduction de la dimensionnalité, et un réseau de neurones à deux couches pour la classification. En mettant l'accent sur l'exploitation synergique de ces techniques, notre système vise à fournir des performances de reconnaissance faciale de pointe, ouvrant ainsi la voie à une large gamme d'applications dans le monde réel.

Chapitre I

Etat de l'art sur la reconnaissance faciale

1 Chapitre I

1.1 Introduction

Dans ce premier chapitre, nous examinerons les principes essentiels ainsi que les principales difficultés rencontrées dans le domaine de la reconnaissance faciale. Nous débuterons par établir les fondements d'un système de reconnaissance des visages, en soulignant son fonctionnement global et ses éventuels usages. Par la suite, nous examinerons l'importance des données de référence dans la création et l'évaluation des systèmes de reconnaissance faciale.

Les défis inhérents à la reconnaissance faciale seront également discutés, notamment la variabilité des conditions d'acquisition des images et les questions liées à la sécurité et à la protection de la vie privée.

On abordera les diverses méthodes employées dans le domaine de la reconnaissance faciale, en soulignant les méthodes classiques ainsi que les progrès récents, en particulier ceux qui reposent sur l'apprentissage profond.

Finalement, ce chapitre examinera les progrès récents et les difficultés persistantes dans le domaine de la reconnaissance faciale, mettant en évidence l'importance grandissante de cette discipline et les problématiques éthiques et sociétales qui en résultent.

Dans cette exploration, notre objectif est de présenter de manière claire et succincte la reconnaissance faciale, en mettant en évidence son importance, ses difficultés et ses perspectives à venir.

1.2 Système de reconnaissance des visages en général

De manière générale, un système de reconnaissance faciale aborde l'image d'entrée sous l'angle d'un problème de classification. Une synthèse d'un système de reconnaissance des visages est ensuite proposée et illustrée en Figure 1.1.

1.2.1 Acquisition et prétraitement des images pour la reconnaissance faciale

Dans le processus de reconnaissance faciale, l'acquisition et le prétraitement des images sont cruciaux. L'acquisition garantit des images de haute qualité, tandis que le prétraitement les

prépare en supprimant les imperfections et en normalisant l'éclairage pour une meilleure reconnaissance.

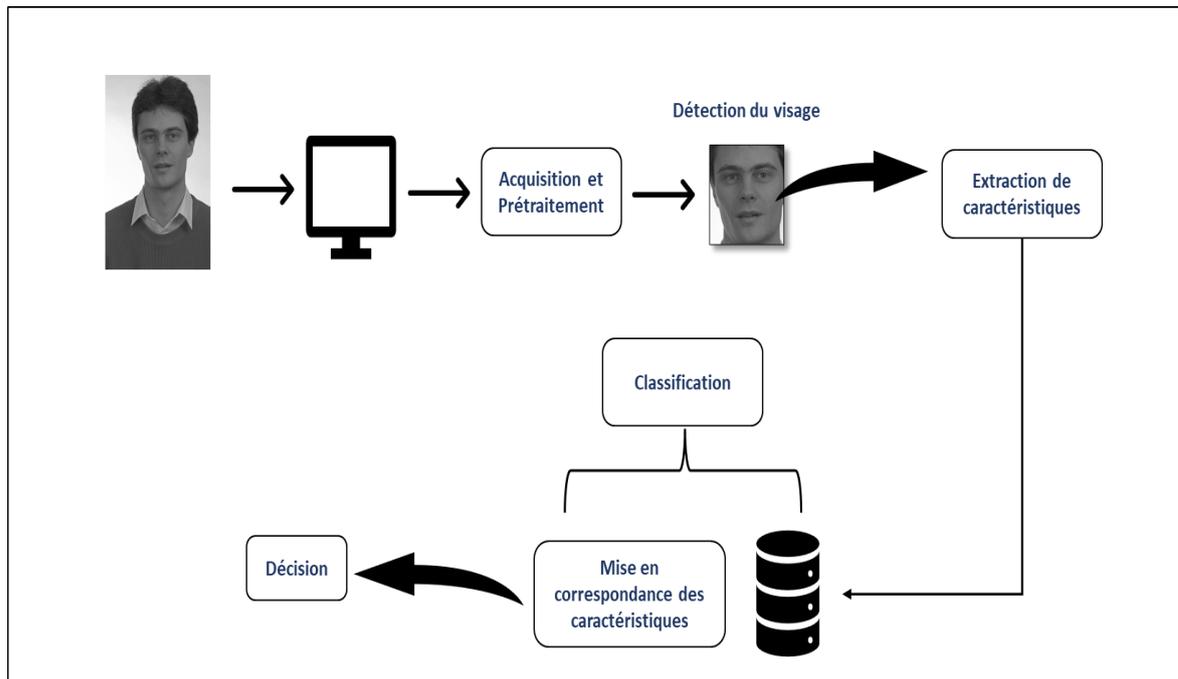


Figure 1.1 : Description d'un système de reconnaissance automatique du visage.

1.2.2 Détection de visage

La détection de visage est une étape essentielle de la reconnaissance faciale, consistant à repérer et extraire les zones d'une image contenant des visages. Cela permet d'isoler les informations pertinentes pour l'identification des individus. Les méthodes traditionnelles utilisent des techniques de traitement d'images comme la détection de contours, tandis que les approches d'apprentissage automatique reposent sur des modèles statistiques ou des réseaux neuronaux pour apprendre à détecter les visages. Une fois détectés, les visages peuvent être alignés pour garantir une analyse précise. En résumé, la détection de visage est une étape cruciale pour extraire les données nécessaires à la reconnaissance faciale.

1.2.3 Extraction des caractéristiques

L'extraction des caractéristiques joue un rôle crucial dans les performances globales des systèmes de reconnaissance faciale. Plusieurs modèles ont été développés pour cette tâche, tels que SIFT (Scale-Invariant Feature Transform) et SURF (Speed Up Robust Feature) [6] [7]. Nous abordons ces modèles ainsi que de nombreux autres extracteurs et descripteurs de caractéristiques récents dans la suite de ce document. La classification des méthodes d'extraction de caractéristiques s'est faite selon différents aspects, notamment l'utilisation de

méthodes globales ou locales, artisanales ou basées sur l'apprentissage, ainsi que l'extraction en 2D ou en 3D [5].

Souvent, l'extracteur de caractéristiques ou le descripteur produit un large ensemble de caractéristiques par rapport à une seule image, ce qui devient encore plus complexe lorsqu'il s'agit de reconnaître une personne à partir d'un flux vidéo. Ce vaste ensemble de caractéristiques est ensuite soumis à des techniques telles que PCA (Principal Component Analysis), SVD (Singular Value Decomposition), MDS (Multidimensional Scaling) et LDA (Linear Discriminant Analysis) [4] [5] [9-11] pour sélectionner les caractéristiques les plus pertinentes et réduire la dimensionnalité. Cette réduction de la dimensionnalité a un effet bénéfique sur les performances du système en réduisant le coût global et en optimisant l'utilisation du temps du système pour la reconnaissance.

1.2.4 Classification

L'étape de classification des caractéristiques consiste à classifier les données recueillies sur les caractéristiques des visages dans un nombre spécifique de classes, en fonction de la tâche à accomplir. Pour ce faire, plusieurs techniques ont été proposées dans différentes recherches, telles que SVM (Support Vector Machine), RF (Random Forest), DT (Decision Table) et NB (Naïve Bayes) [8].

1.2.5 Mise en correspondance des caractéristiques

Après avoir extrait les caractéristiques principales de l'image faciale initiale, nous les comparons de manière itérative à une base de données préexistante de caractéristiques, afin d'atteindre l'objectif visé.

1.3 Bases de données de référence connues

L'évaluation des performances d'un système de reconnaissance faciale représente l'un des principaux défis de la recherche dans ce domaine. Un large éventail de bases de données d'images pour l'évaluation des systèmes de reconnaissance faciale a été établi et est rendu public pour faciliter l'évaluation des nouveaux algorithmes proposés [9][10].

MS-Celeb-1M [9] est une base de données d'images faciales constituée d'environ 8,2 millions d'images collectées sur Internet dans le but de développer des technologies de reconnaissance faciale. Elle représente l'une des plus vastes collections d'images faciales accessibles au public, englobant environ 100 000 individus, principalement des acteurs américains, britanniques et des personnalités publiques en ligne. Étant donné que les images

sources n'ont pas été capturées à des fins spécifiques, cette base de données présente des limitations significatives en termes de variations d'éclairage et de pose.

CMU-PIE [11] : L'ensemble de données de l'Université Carnegie Mellon sur l'illumination et l'expression de la pose (CMU-PIE) a été rassemblé entre octobre et décembre 2000. Chaque sujet de l'ensemble de données présente 13 poses différentes, 43 conditions d'éclairage différentes et 4 expressions faciales variables. Dans sa version initiale, toutes les images ont été capturées lors d'une seule session d'enregistrement, ce qui signifie que l'ensemble de données ne tient pas compte de la variation due au vieillissement.

FERET [12] : L'ensemble de données FERET a été initialement recueillie dans le cadre du protocole de l'Université George Mason, avec la collaboration du Dr Harry Wechsler de l'Université George Mason et du Dr. Jonathan Phillips du laboratoire de recherche de l'armée à Adelphi, Maryland. Elle contient 11 338 images couleur de 512×768 pixels capturées dans un environnement semi-contrôlé avec 13 poses différentes de 994 sujets.

AR [13] : La base de données de visages AR a été développée par Aleix Martinez et Robert Benavente au Computer Vision Center (CVC) de l'U.A.B. Elle comprend plus de 4 000 images en couleur de 126 individus, soit 70 hommes et 56 femmes. Les images représentent des visages en vue frontale avec différentes expressions faciales, conditions d'éclairage et occlusions telles que des lunettes de soleil et des foulards. Chaque image a été capturée deux fois lors de deux sessions distinctes, avec un intervalle de 14 jours entre les deux.

Yale B [14] : Les ensembles de données Yale et Yale B sont largement utilisés dans la littérature académique pour évaluer les algorithmes de reconnaissance faciale. La version initiale de l'ensemble de données Yale sur les visages ne comprend que 165 images en niveaux de gris au format GIF de 15 personnes. La version étendue, Yale-B, inclut un nombre plus important de participants, avec 64 conditions d'éclairage par sujet et 9 angles de pose différents.

AT&T/ORL [10] : La base de données de visages AT&T, anciennement connue sous le nom de base de données de visages ORL, comprend 40 sujets distincts, avec 10 images disponibles pour chaque sujet. Les images ont été capturées entre avril 1992 et avril 1994 par le groupe Speech, Vision and Robotics du laboratoire du département d'ingénierie de l'université de Cambridge. Elles ont été prises à différents moments, présentant des variations d'éclairage, d'expressions faciales (yeux ouverts ou fermés, souriant ou non) et de détails du

visage (avec ou sans lunettes). Toutes les images ont été prises sur un fond sombre et homogène, avec les sujets en position verticale et frontale.

1.4 Principaux défis pour la reconnaissance faciale

Dans un cadre contrôlé, la recherche sur la reconnaissance faciale a connu un succès significatif [2] , permettant d'atteindre les objectifs essentiels pour les applications sociales. Cependant, malgré ces progrès, des défis substantiels persistent dans les environnements non contrôlés, caractérisés par l'évolution constante des sujets et la complexité des variations temporelles. La dynamique de ces contextes rend ardue l'observation des changements et la calibration des perspectives des systèmes d'intelligence artificielle. Cette section s'attache à décrire les multiples obstacles rencontrés par les technologies de reconnaissance faciale dans de tels contextes, soulignant ainsi la nécessité d'approches plus adaptatives et sophistiquées.

1.4.1 Variation de pose

La reconnaissance de la pose représente un défi particulièrement complexe, surtout dans des environnements peu coopératifs, tels que la détection d'espions au sein d'une foule, la sécurité dans les aéroports contre les menaces terroristes, ou encore l'identification de voleurs dans des marchés animés. Idéalement, la collecte de multiples images de référence du sujet serait préférable pour aborder cette tâche. Cependant, dans la pratique, cette approche s'avère peu réalisable dans la plupart des applications du monde réel. Par exemple, en cas de disponibilité limitée de photos d'identité pour chaque individu dans une base de données d'images, la tâche de repérer un sujet spécifique devient considérablement plus difficile. Comment pourrait-on alors détecter un individu terroriste parmi les spectateurs d'un stade de football ? D'un point de vue technique, la disparité entre deux images identiques présentant des poses différentes est souvent plus prononcée que celle observée entre deux images de personnes différentes.

1.4.2 Variation d'éclairage

Les systèmes de reconnaissance faciale sont confrontés à un autre obstacle de taille : la variation de l'éclairage ou l'effet d'éclairage. En raison des caractéristiques de réflexion de la peau, de la diversité des capteurs de caméra, des effets de résolution et des conditions environnementales, l'éclairage représente un paramètre totalement incontrôlable pour l'intelligence artificielle. Les méthodes de classification rencontrent souvent des difficultés à gérer de manière efficace les modifications d'éclairage. En général ces approches souvent appelées approche traditionnelles [6] [15] présentent un faible capacité d'adaptation aux

variations des conditions d'éclairage, ce qui peut entraîner une baisse significative de la performance et de la fiabilité de leurs résultats.

1.4.3 L'occlusion

La présence d'autres objets dans l'image faciale tels que la haine, le casque, les lunettes de soleil, la main, l'écharpe, le masque, etc., peut entraîner l'occlusion du visage, qui consiste à dissimuler des caractéristiques faciales essentielles.

L'occlusion constitue un autre élément essentiel qui a un impact significatif sur les performances de la méthode de reconnaissance faciale. Comme le montre la Figure 1.2, il est fréquent dans la réalité de dissimuler certaines caractéristiques du visage à l'aide de lunettes ou d'une écharpe.

1.4.4 Vieillesse

Le processus de vieillissement facial est intrinsèquement complexe, caractérisé par des transformations imprévisibles et parfois divergentes au fil du temps. Cette évolution, bien que non linéaire, présente des variations qui peuvent sembler incohérentes d'un individu à l'autre. Il revêt une importance capitale pour les systèmes d'intelligence humaine et artificielle d'être en mesure de détecter et de comprendre ces fluctuations dans le processus de vieillissement du visage.

1.5 Approches en reconnaissance faciale

Depuis trois décennies, la reconnaissance faciale a été un domaine de recherche essentiel pour diverses communautés scientifiques telles que l'apprentissage automatique, l'intelligence artificielle, le traitement d'images et la vision par ordinateur [2-4]. Les méthodes développées pour la reconnaissance faciale relèvent de domaines scientifiques variés et étendus, ce qui rend difficile leur classification standard. De plus, l'utilisation croissante de modèles hybrides complique encore davantage la catégorisation de ces approches dans des branches standard pour la représentation des caractéristiques ou la classification. Néanmoins, en se référant à la littérature récente, nous avons tenté de présenter une catégorisation claire et générale des approches de reconnaissance faciale. La Figure 1.3 illustre la distribution catégorielle des différentes approches de reconnaissance faciale.



Figure 1.2 : Visages avec effet d'occlusion tels que présentés dans AR [13].

1.5.1 Approches classiques

La recherche sur la reconnaissance faciale trouve ses racines dans des domaines historiques tels que la psychologie des années 1950 et la littérature technique des années 1960 [2-4] [16]. Ces premiers concepts étaient essentiellement dérivés des systèmes de reconnaissance des formes, comme exposé par Lawrence Gilman dans sa thèse de doctorat au MIT [2-4] [17]. Il a établi les bases en montrant comment une caractéristique 2D extraite d'une photographie peut être mise en relation avec sa représentation 3D. Cependant, les recherches ultérieures ont rapidement révélé les défis pratiques posés par les conditions environnementales variables, défis qui demeurent encore aujourd'hui, malgré les avancées en matière de puissance de calcul. Ces premières méthodes étaient principalement centrées sur les relations géométriques entre les points du visage, et leur efficacité dépendait largement de la détection précise de ces points dans des environnements complexes, ainsi que de la cohérence de ces relations malgré les variations. Ces questions continuent de représenter un défi majeur pour la communauté de recherche. Une tentative précoce de développement d'un système de reconnaissance faciale a été entreprise par Kanade et ses collaborateurs [1], qui ont utilisé des techniques simples de traitement d'images pour extraire un vecteur de 16 paramètres faciaux. Leur méthode, bien que rudimentaire, a donné des résultats prometteurs avec un taux de précision de 75 % sur une

base de données préétablie de 20 personnes. Plus récemment, en 2003, Zhao et ses collaborateurs ont dressé un panorama complet des techniques utilisées dans la détection et la reconnaissance faciales au cours des 30 dernières années [16]. Leur revue couvre de nombreux aspects psychologiques et neuroscientifiques de la reconnaissance faciale, et classe initialement les techniques en trois grandes catégories : les méthodes holistiques, les méthodes basées sur la géométrie pure et les méthodes hybrides.

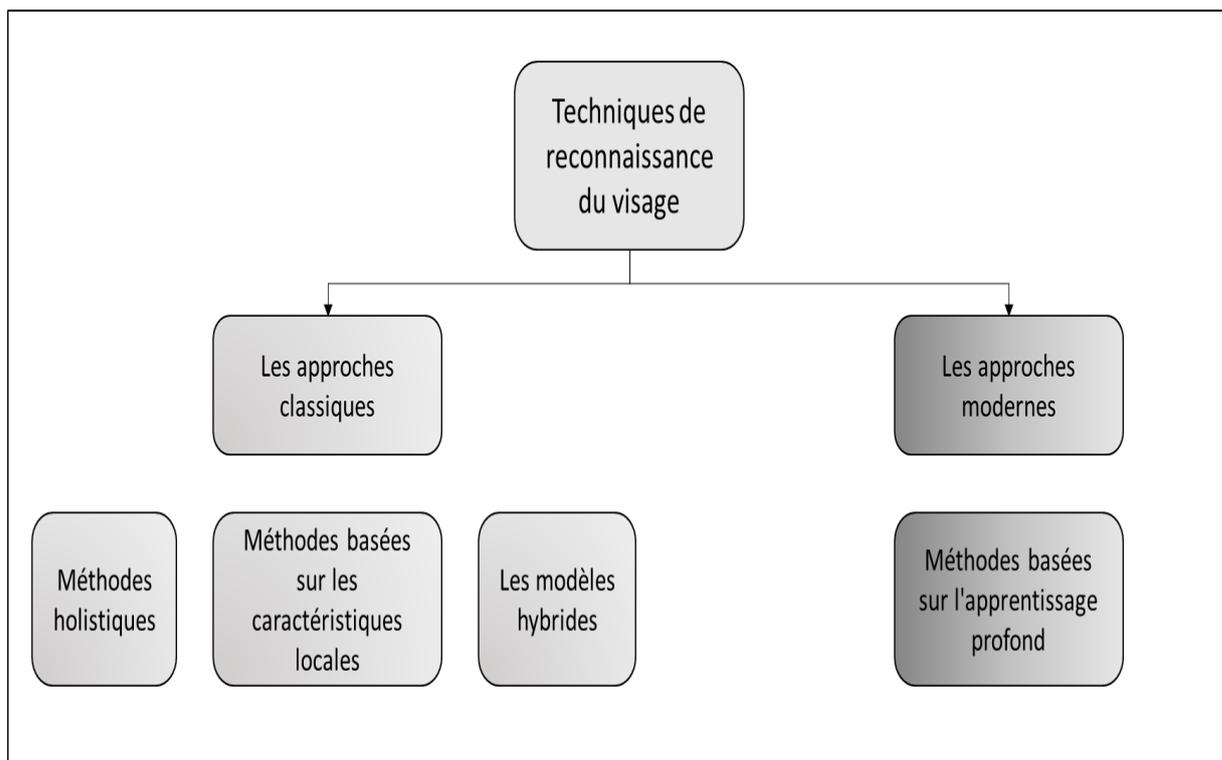


Figure 1.3 : Répartition des méthodes de reconnaissance des visages par catégories.

Les méthodes classiques de reconnaissance des visages englobent diverses approches, chacune ayant ses propres caractéristiques et techniques. Les méthodes holistiques représentent une catégorie dans laquelle l'image entière du visage est considérée comme un tout pour la reconnaissance. Ces méthodes font généralement appel à la PCA, à l'analyse discriminante linéaire (LDA) ou au discriminant linéaire de Fisher (FLD : Fisher linear discriminant) pour extraire les caractéristiques essentielles de l'image du visage et réduire sa dimensionnalité. Les méthodes basées sur les caractéristiques locales, quant à elles, se concentrent sur des régions ou des caractéristiques faciales spécifiques, telles que les yeux, le nez ou la bouche. Des techniques telles que le SIFT ou le SURF sont couramment utilisées pour détecter et décrire ces caractéristiques locales, qui sont ensuite mises en correspondance sur différentes images de visage. Les méthodes hybrides combinent des aspects des approches holistiques et des approches basées sur les caractéristiques locales pour obtenir des systèmes

de reconnaissance faciale robustes. Ces méthodes intègrent souvent plusieurs algorithmes ou techniques d'extraction de caractéristiques afin de tirer parti des points forts de chaque approche et d'atténuer leurs limites respectives. En combinant les informations holistiques et locales, les méthodes hybrides visent à améliorer la précision et la robustesse des systèmes de reconnaissance faciale, ce qui les rend adaptées à un large éventail d'applications dans les domaines de la sécurité, de la surveillance et de la biométrie (Figure 1.3).

1.5.2 Reconnaissance des visages basée sur l'apprentissage profond

Il a été constaté que les réseaux neuronaux possèdent une capacité de calcul considérable pour la reconnaissance d'objets, et ils ont profondément transformé l'environnement de l'apprentissage automatique ces dernières années [2]. Ces réseaux ont suscité un intérêt généralisé parmi les chercheurs de divers domaines, notamment les sciences sociales, l'ingénierie et les sciences de la vie [2] [18-24]. Dans de nombreux cas, les cadres profonds sont intégrés aux modèles existants pour créer des hybrides innovants, générant ainsi des résultats révolutionnaires. Plusieurs chercheurs, en particulier dans le domaine de la reconnaissance faciale, ont souligné leur capacité remarquable de calcul, leur précision exceptionnelle et leur orientation vers des résultats concrets.

La recherche actuelle sur la reconnaissance faciale à l'aide de l'apprentissage profond se concentre principalement sur l'hybridation des modèles existants avec des réseaux profonds afin d'introduire des techniques améliorées pour une reconnaissance des visages plus performante. Les chercheurs reconnaissent que les réseaux neuronaux et les techniques associées sont particulièrement bien adaptés pour atteindre des performances élevées en termes de précision et de robustesse [18]. Leur capacité à classifier efficacement un grand nombre d'images de visages non étiquetées les positionne en avance par rapport à d'autres approches. Les recherches en cours se concentrent sur l'utilisation de modèles profonds dans des environnements non contraints, où les visages présentent des variations importantes en termes de pose, d'illumination et de couleur. Bien que des progrès aient été réalisés dans l'évaluation des techniques d'apprentissage profond pour la reconnaissance faciale, il reste encore un écart significatif à combler entre les différentes variations [18].

L'intégration des réseaux neuronaux artificiels (ANN : Artificial Neural Networks) et des réseaux neuronaux convolutifs (CNN : Convolutional Neural Networks) a révolutionné le domaine de la classification de la reconnaissance faciale dans le cadre de l'apprentissage en profondeur. Inspirés des systèmes neuronaux biologiques, les systèmes ANN utilisent des

neurones artificiels interconnectés pour apprendre de l'expérience et répondre efficacement aux tâches de reconnaissance en détectant des motifs complexes à l'intérieur des motifs [8]. D'autre part, le CNN s'inspire du cortex visuel du cerveau des mammifères et utilise différentes couches telles que les couches convolutives, de mise en commun, non linéaires et entièrement connectées pour extraire des caractéristiques et classer efficacement les images faciales [8]. Les modèles d'apprentissage profond pour la classification de la reconnaissance faciale peuvent apprendre automatiquement des extracteurs de caractéristiques locaux, s'adapter aux variations des expressions faciales, aux conditions d'éclairage et à d'autres facteurs, améliorant ainsi considérablement la précision et la robustesse de la reconnaissance dans les scénarios du monde réel [8]. Cette intégration comble les lacunes des méthodes classiques en améliorant la capacité de performance et en réduisant le risque de surajustement, ouvrant ainsi la voie à des systèmes de reconnaissance faciale plus précis et plus fiables [8].

1.5.3 Les approches hybrides

La biométrie a fait l'objet d'une attention particulière en tant que domaine important ayant des implications significatives dans de multiples secteurs [8] [18]. Elle consiste à analyser les caractéristiques physiques ou comportementales uniques des individus afin de les identifier et de les authentifier. L'identification précise des individus joue un rôle crucial dans le renforcement des mesures de sécurité, l'optimisation des systèmes de contrôle d'accès et l'atténuation des activités frauduleuses. Parmi les différentes modalités biométriques, la reconnaissance faciale est apparue comme une approche particulièrement remarquable et prometteuse en raison de sa nature non intrusive, de sa fonctionnalité en temps réel et de son adaptabilité à divers environnements [8] [18]. Les techniques traditionnelles de reconnaissance faciale comportent deux étapes principales, à savoir l'extraction des caractéristiques et la classification [2], [3], [18]. La première vise à capturer les caractéristiques distinctives d'un visage à partir d'une image, tandis que la seconde utilise ces caractéristiques extraites pour identifier l'individu [3], [18]. Diverses méthodes d'extraction de caractéristiques ont été employées dans la littérature, telles que les motifs binaires locaux (LBP : Local Binary Pattern) [25], [26], les faces propres (Eigenfaces) [27], [28], les filtres de Gabor [29] et l'histogramme des gradients orientés (HOG : Histogram of Oriented Gradients) [30]. Le LBP est un descripteur de texture qui code les motifs locaux en comparant l'intensité du pixel central à celle de ses voisins [25]. Il est efficace pour capturer les variations de texture dans le visage. Le HOG se concentre sur la distribution des gradients locaux dans une image, ce qui peut représenter des informations sur la forme [30]. Les visages propres utilisent la PCA pour extraire les caractéristiques les plus discriminantes d'un ensemble

d'images de visages [27], [28]. Enfin, les filtres de Gabor sont un ensemble de filtres qui capturent les informations de fréquence spatiale et peuvent représenter les textures faciales [29]. Une autre approche pour l'extraction de caractéristiques dans la reconnaissance faciale implique l'utilisation de transformées fréquentielles, telles que la DCT [31-36] et la transformée en ondelettes discrète (DWT) [19], [33], [34], [37-39]. D'une part, la transformée en ondelettes discrètes offre l'avantage d'une analyse multi-résolution, ce qui permet de capturer les détails du visage à différentes échelles [36]. Dans [33], les auteurs ont proposé une hybridation de la DCT fractionnaire et de la DWT pour normaliser les variations de lumière dans les images de visage. Cette combinaison facilite considérablement le processus d'extraction des caractéristiques des images. Une autre approche a été introduite dans [34] où deux ensembles de coefficients de caractéristiques extraits ont été combinés à l'aide de deux techniques distinctes d'extraction de coefficients de caractéristiques : filtres gaussiens basés sur le DWT et filtres de Gabor basés sur le DCT. L'étiquette de classe souhaitée pour les visages individuels a été générée en employant un classificateur à vecteur de support multiclasse, qui utilise les deux caractéristiques combinées. Une méthode combinant les K-voisins les plus proches et l'optimisation des essaims de particules a été proposée dans [35]. Le classificateur du plus proche voisin a été amélioré à l'aide de l'algorithme PSO (Particle Swarm Optimization), une métaheuristique basée sur la population. Par exemple, Ari Setiawan, Riyanto Sigit et Rika Rokhana [36] ont démontré l'efficacité de l'utilisation de la compression d'images basée sur le DCT pour réduire le temps d'apprentissage de leur système de reconnaissance faciale basé sur les réseaux neuronaux convolutionnels (CNN : Convolutional Neural Network) proposé sans compromettre de manière significative sa précision. En outre, la combinaison de la cohérence DWT avec diverses méthodes, notamment la PCA, LDA et CNN, a permis d'améliorer les performances de manière prometteuse, ce qui se traduit par de meilleurs taux de reconnaissance [37]. Dans ce contexte, une contribution remarquable a été apportée par Randa Atta et Mohammad Ghanbari dans [40], qui ont introduit une nouvelle méthodologie connue sous le nom de pyramide DCT. Cette approche combine les caractéristiques avantageuses de la DCT et de la stratégie multirésolution. Initialement développée pour le codage d'images compactes [41], la pyramide DCT a donné des résultats prometteurs dans le domaine de la reconnaissance faciale [40]. Pour l'étape de classification, plusieurs algorithmes sont couramment utilisés dans la reconnaissance faciale traditionnelle, notamment le plus proche voisin [42], les SVM [20] et les arbres de décision [43]. Les algorithmes du plus proche voisin classent une image en fonction de sa similarité avec des visages connus dans l'ensemble d'apprentissage [35], [44], tandis que les SVM créent un hyperplan pour séparer les différentes classes, en maximisant la

marge entre elles [20]. D'autre part, les arbres de décision utilisent une structure hiérarchique pour classer les images en fonction des différentes caractéristiques du visage [43]. Néanmoins, les techniques traditionnelles sont moins robustes lorsqu'elles sont confrontées à des occlusions telles que des lunettes de soleil, des chapeaux ou des poils du visage, car elles peuvent interférer avec le processus d'extraction des caractéristiques et affecter la précision de la classification [3], [18], [45]. En outre, les méthodes traditionnelles peuvent avoir du mal à gérer les images bruyantes ou de faible qualité, ce qui entraîne une baisse des performances de reconnaissance dans de tels cas [3], [18]. L'apprentissage profond a révolutionné le domaine de la reconnaissance faciale en offrant des améliorations significatives en termes de précision [3], [18]. Les réseaux neuronaux profonds, tels que les CNN, ont montré des performances remarquables dans l'extraction et la classification des caractéristiques faciales [35], [46], [47]. En effet, les CNN peuvent apprendre automatiquement des représentations de haut niveau des visages en utilisant plusieurs couches d'opérations convolutives et de mise en commun. L'avantage de l'apprentissage profond dans la reconnaissance faciale réside dans sa capacité à apprendre des caractéristiques complexes directement à partir de données brutes, ce qui élimine la nécessité de recourir à des techniques d'extraction de caractéristiques artisanales [3], [18], [45]. En outre, les modèles d'apprentissage profond peuvent capturer les variations subtiles de l'apparence du visage et gérer les variations de la pose, des conditions d'éclairage et des expressions faciales. Cependant, le principal inconvénient des approches d'apprentissage profond est leur besoin de grandes quantités de données d'entraînement étiquetées [3], [18], [45]. L'entraînement des réseaux neuronaux profonds nécessite un ensemble substantiel de données de visages étiquetés afin d'obtenir des performances optimales. La collecte et l'annotation d'un tel ensemble de données peuvent être longues et coûteuses. Pour résoudre ce problème, une approche courante consiste à combiner les techniques traditionnelles d'extraction de caractéristiques avec l'apprentissage profond. Ce couplage implique l'utilisation de méthodes traditionnelles pour extraire les caractéristiques des images faciales, qui sont ensuite introduites dans un réseau neuronal profond pour la classification [11]. Cette approche hybride bénéficie de la puissance de représentation de l'apprentissage profond tout en utilisant la robustesse et l'interprétabilité des méthodes traditionnelles d'extraction de caractéristiques [18-24]. Plus précisément, dans [21], les auteurs ont proposé une approche améliorée pour la reconnaissance des visages humains à l'aide d'un réseau neuronal à rétropropagation (BPNN : Back-Propagation Neural Network) et d'une extraction de caractéristiques basée sur la corrélation. L'utilisation du descripteur d'histogramme binaire local (LBPH : Local Binary Patterns Histograms) et la génération d'un nouvel ensemble de données (T-Dataset) y sont également envisagées. Bien que l'article

revendique une plus grande précision et un coût de calcul réduit grâce à l'utilisation de caractéristiques d'image réduites, l'utilisation de cinq algorithmes combinés de mesure de la distance pour générer le T-Dataset pourrait poser un problème en termes d'efficacité de calcul réduite lors de la réalisation d'essais sur des ensembles de données étendus. Les auteurs de [19] se sont concentrés sur une technique qui combine les réseaux neuronaux probabilistes (PNN : Probabilistic Neural Network) avec une version améliorée de l'analyse discriminante linéaire à noyau à des fins de reconnaissance faciale (IKLDA : Improved Kernel LDA). Ari Setiawan, Riyanto Sigit et Rika Rokhana explorent dans [14] l'utilisation de l'apprentissage profond CNN et du DCT pour la reconnaissance faciale. Bien qu'ils discutent de l'impact des fonctions d'activation et des époques, l'utilisation des CNN pourrait conduire à des structures de données complexes et à une augmentation de la charge de calcul pour les tâches de classification. Les chercheurs ont introduit un cadre dans la référence [23], qui implique l'utilisation de lunettes intelligentes pour la reconnaissance faciale. Ils ont exploité des techniques de traitement d'images et appliqué des CNN, en employant spécifiquement l'apprentissage par transfert avec le réseau multicouche AlexNet pré-entraîné. Bien que ce cadre offre des avantages attrayants tels que la portabilité et la capture efficace de la vue frontale, il convient de noter que l'utilisation de la structure complexe du réseau AlexNet nécessiterait des spécifications matérielles robustes. Dans notre système, nous proposons une nouvelle approche pour améliorer la précision et l'efficacité des systèmes de reconnaissance faciale. Elle combine la technique de décomposition DCT-pyramide avec un réseau neuronal artificiel (ANN : Artificial Neural Network) utilisant l'algorithme de rétropropagation largement répandu. Ainsi, nous visons à exploiter efficacement les avantages offerts par chacune de ces deux techniques. Nous montrons que le système de reconnaissance des visages qui en résulte est très robuste et capable de fonctionner avec des exigences minimales en matière de mémoire. La technique de décomposition pyramidale DCT sert de base à notre approche. Elle décompose l'image du visage en plusieurs sous-bandes de fréquences pour capturer les composantes à basse et haute fréquence. En utilisant cette décomposition, nous extrayons un riche ensemble de caractéristiques qui peuvent améliorer de manière significative le pouvoir discriminant du système de reconnaissance faciale. En outre, la structure pyramidale nous permet de représenter efficacement les informations faciales à différentes échelles, ce qui permet de résister aux variations de la pose, de l'éclairage et des expressions faciales. Toutefois, la représentation efficace des informations faciales fournie par la décomposition pyramidale DCT se traduirait par des vecteurs de grande dimension. Afin de surmonter cet inconvénient, des mesures statistiques telles que la moyenne, la variance et l'entropie ont été utilisées pour réduire efficacement la dimensionnalité sans compromettre les

informations essentielles contenues dans les vecteurs de caractéristiques. Pour améliorer encore la précision de la reconnaissance, nous intégrons un ANN dans notre système. Plus précisément, nous utilisons une architecture avec rétropropagation, un algorithme d'apprentissage efficace [3], [18-20]. Le réseau neuronal est formé pour apprendre les relations complexes entre les caractéristiques extraites et les identités correspondantes. Grâce à un processus d'apprentissage itératif, le réseau adapte ses paramètres internes pour optimiser les performances de reconnaissance, améliorant ainsi la précision des tâches d'identification et de vérification des visages. En concevant soigneusement l'intégration de la décomposition pyramidale DCT et du réseau neuronal artificiel, nous nous assurons que le système utilise efficacement les ressources disponibles sans sacrifier la précision. Ainsi, un autre avantage de notre approche est son faible besoin en mémoire. Ceci est particulièrement important dans les applications réelles, où les contraintes de mémoire sont cruciales [3], [18], comme dans les systèmes embarqués ou les appareils mobiles.

1.6 Reconnaissance des visages à l'aide de la psychologie cognitive

La question fondamentale de la perception des visages par les individus est abordée par plusieurs études menées dans les domaines de la psychologie cognitive et de la neurophysiologie liées à la reconnaissance des visages [48-51]. Par ailleurs, quels sont les mécanismes qui sous-tendent la reconnaissance des visages ? L'analyse théorique et empirique des mécanismes cognitifs qui sous-tendent la connaissance conceptuelle et pratique de la reconnaissance des visages chez l'homme est au cœur de la discussion. Ces études devraient permettre d'améliorer la traduction automatique du processus de reconnaissance des visages, sans ambiguïté ni erreur de classification dans diverses conditions. Ces recherches pourraient permettre d'étudier la physiologie de la prosopagnosie, une maladie dans laquelle les patients perdent la capacité de reconnaître les visages, y compris ceux de personnes qu'ils ont déjà rencontrées, et les effets de la spécificité du visage sur la reconnaissance des visages. De nombreux travaux de recherche ont également examiné l'application des caricatures de visages, qui modifient les visages dans le but de les faire sortir du lot [48-51]. On pense que les motifs ressemblant à des visages ont une préférence primitive dans le système visuel humain. Des recherches sur le développement de la reconnaissance des visages dans la petite enfance ont montré qu'un bébé développe une préférence pour les visages et les motifs ressemblant à des visages quelques heures après sa naissance, en commençant par le visage de sa mère [48-51]. L'hypothèse de la redondance inter sensorielle prédit qu'au cours du développement précoce, une stimulation visuelle unimodale - c'est-à-dire un visage

dynamique silencieux - améliore davantage la perception des visages qu'une stimulation audiovisuelle bimodale - c'est-à-dire un visage dynamique accompagné d'un discours synchrone [48-51]. On ne sait toujours pas comment les compétences de reconnaissance des visages chez l'adulte se construisent à partir des préférences pour les visages et les motifs ressemblant à des visages jusqu'aux capacités de perception et de reconnaissance des visages en reliant les années d'expérience en matière de reconnaissance des visages.

Des recherches ultérieures ont montré que la capacité à traiter et à identifier les visages en utilisant des modalités de traitement efficaces, notamment faciales, contextuelles et holistiques, est corrélée à l'expertise humaine en matière de reconnaissance faciale [48-51]. La capacité à voir et à identifier les visages en se basant uniquement sur leurs caractéristiques spécifiques, telles que la taille du nez ou la forme des yeux, est connue sous le nom de traitement facial. Plus sophistiqué que le traitement des traits, le traitement congruent [3] décrit la perception et l'identification des visages sur la base de la longueur et de l'espacement des traits. Le terme "traitement holistique" décrit la capacité de percevoir et d'identifier les visages dans leur ensemble plutôt que de se limiter à certaines caractéristiques. La plupart des gens s'accordent à dire que les bébés perçoivent les visages de manière immature, comme le traitement des traits [48-51]. Les adolescents, quant à eux, traiteraient les visages sur la base d'informations contextuelles plutôt que sur les seules informations faciales, parce qu'ils sont plus familiers avec eux. Les adultes possèdent d'excellentes capacités de traitement des visages et sont capables de les voir et de les identifier. Cependant, ils font preuve d'une imprécision notable lorsqu'il s'agit de reconnaître des visages inversés. La capacité de reconnaissance robuste des visages est progressivement renforcée par une approche holistique plus complexe impliquant des informations congruentes. Ce modèle de comportement indique qu'un changement d'approche s'opère sur plusieurs années. L'effet "autre race" est un problème rencontré au cours du développement de la technologie de reconnaissance faciale. L'expression "effet d'appartenance à une autre race" décrit les personnes qui sont plus aptes à différencier les membres de leur propre race de ceux d'une autre race. La discrimination à l'égard des visages d'autres races diminue au fur et à mesure que l'on se familiarise avec les visages de sa propre race. L'expérience accrue des visages conduit à un rétrécissement de la perception des visages, un phénomène connu sous le nom de rétrécissement perceptif [48-51]. Cet effet aboutit finalement à la création d'un "prototype de visage". La perception par un adulte d'un certain nombre de visages dont la moyenne est calculée ensemble est appelée prototype de visage, et l'on pense qu'elle a une plus grande capacité de discrimination que n'importe quel visage pris isolément.

D'un point de vue neurophysiologique, il semble que le système visuel humain consacre des ressources cérébrales spécifiques à la perception et à l'identification des visages. On soupçonne que le système visuel humain traite les visages par l'intermédiaire de systèmes cognitifs et neurologiques distincts. En fait, une pléthore de données suggère que le cortex visuel extrastrié pourrait être le site principal de la compréhension des visages humains. Le fait que les visages d'animaux suscitent une bonne réaction mais que les visages schématiques ne stimulent pas beaucoup d'activité indique un modèle intéressant de sélectivité et de généralité dans cette région. Le concept d'un module propre de traitement des visages semble très plausible.

Il est également proposé que des systèmes neurophysiologiques distincts soient impliqués dans le traitement de l'identité et de l'expression faciales. Il est possible d'extraire l'expression d'un visage indépendamment de l'identité et vice versa si ce corollaire est prouvé. La question de savoir si une implémentation basée sur la biologie pourrait identifier une personne sans tenir compte de son expression dépendrait des conséquences informatiques de ce sujet. Dans ce type de situation, l'évaluation des émotions sur un visage dans une application d'interface homme-machine ne nécessitera pas l'extraction d'une représentation de l'identité. Toutefois, il convient de noter que la plupart des études autorisent un certain chevauchement, éventuellement au stade de la représentation, même s'il semble y avoir un degré important de dissociation entre l'identité et l'expression. Par exemple, une petite fraction de neurones réagit aux deux éléments, même si certains ne réagissent qu'à l'identité et d'autres qu'à l'expression.

Cependant, les chercheurs évaluent fréquemment la perception des relations de second ordre pour étudier l'expertise des visages lors de la conception d'un système de reconnaissance automatique des visages. Les relations de premier ordre décrivent les caractéristiques fondamentales d'un visage, tandis que les relations de second ordre décrivent la distance entre les caractéristiques d'un visage. Alors que la perception ou la reconnaissance d'un visage basée sur les relations de second ordre est plus fiable, la reconnaissance d'un visage basée sur les relations de premier ordre est plus sujette aux erreurs [3].

1.7 Conclusion

Dans ce chapitre, il est mis en évidence l'importance capitale de la reconnaissance faciale dans notre société moderne, ainsi que les nombreux défis auxquels cette technologie fait face. Les principes fondamentaux qui sous-tendent les systèmes de reconnaissance faciale ont été étudiés en profondeur, mettant en évidence leur fonctionnement global et leurs différentes applications potentielles. En outre, nous avons étudié l'importance cruciale des données de référence dans la création et l'évaluation de ces systèmes, mettant en évidence leur rôle essentiel dans leur efficacité et leur fiabilité.

Nous avons abordé en détail les défis liés à la reconnaissance faciale, tels que la diversité des conditions d'acquisition des images et les inquiétudes concernant la sécurité et la protection de la vie privée. Ces difficultés mettent en évidence l'importance d'adopter une approche méticuleuse et éthique lors du développement et de l'utilisation de cette technologie

Nous avons aussi étudié les diverses techniques employées dans le domaine de la reconnaissance faciale, en soulignant à la fois les méthodes classiques et les progrès récents, notamment ceux qui s'appuient sur l'apprentissage profond. Grâce à cette exploration, nous avons pu saisir la progression constante de la technologie de reconnaissance faciale et son potentiel pour faire face aux défis actuels et futurs.

Enfin, ce chapitre a souligné les avancées récentes et les obstacles persistants dans le domaine de la reconnaissance faciale, mettant en évidence l'importance grandissante de cette discipline dans différents domaines et les problématiques éthiques et sociétales qui en découlent. De nos jours, il est clair que la reconnaissance faciale est une technologie qui a un impact considérable, ce qui demande une réflexion approfondie sur ses conséquences et son utilisation responsable dans la société.

Globalement, ce chapitre a présenté une vision globale et succincte de la reconnaissance faciale, soulignant son importance, ses défis et les perspectives à venir. Celui-ci établit les fondements d'une exploration plus approfondie dans les chapitres qui suivent, où nous étudierons de manière plus approfondie les progrès technologiques, les applications particulières et les conséquences éthiques de la reconnaissance automatique des visages.

Chapitre II

Etat de l'art sur les transformées et les réseaux de neurones

2 Chapitre II

2.1 Introduction

Dans le domaine de la reconnaissance faciale, l'analyse et l'interprétation des données visuelles sont des aspects essentiels pour élaborer des systèmes de reconnaissance précis. Nous aborderons ici trois concepts fondamentaux sous-jacents à cette discipline : les méthodes basées sur les sous-espaces, l'analyse multi-résolution et les réseaux de neurones.

Les méthodes basées sur les sous-espaces, telles que la PCA et LDA, revêtent une importance cruciale dans le domaine de la reconnaissance faciale. Elles permettent la réduction de la dimensionnalité des données et l'extraction des caractéristiques les plus discriminantes des images faciales. Nous explorerons ici en détail l'application de ces techniques dans le contexte de la reconnaissance faciale, ainsi que leurs avantages, limites et applications pratiques.

La capacité de la reconnaissance faciale à extraire des informations discriminantes à partir de données visuelles est essentielle, et les conversions fréquentielles permettent de représenter de manière efficace les caractéristiques essentielles des images faciales dans le domaine des fréquences. En même temps, l'utilisation de l'analyse multi-résolution permet de saisir les détails précis des visages tout en préservant une vision globale, ce qui est essentiel pour détecter et reconnaître les visages dans différentes situations.

Finalement, Les réseaux de neurones artificiels ont progressivement gagné en pertinence en raison de leur aptitude à assimiler des représentations hiérarchiques à partir de données non structurées, favorisant ainsi l'avancement de la technologie de reconnaissance faciale. Ces systèmes peuvent être entraînés à extraire de manière automatique des caractéristiques distinctives des visages, entraînant une amélioration substantielle des performances de ces réseaux.

Ce chapitre se consacre à l'exposition de principes fondamentaux, en mettant l'accent sur leur application spécifique à la reconnaissance faciale. L'accent est mis sur l'importance et la pertinence de ces principes dans le développement de systèmes de reconnaissance faciale avancés et robustes.

2.2 Méthodes basées sur les sous-espaces

2.2.1 Analyse en composantes principales

La PCA est une technique fondamentale largement appliquée dans le domaine de la reconnaissance faciale pour l'extraction de caractéristiques, la réduction de la dimensionnalité et d'apprentissage automatique utilisée pour condenser un énorme ensemble de données en un ensemble plus petit tout en préservant les modèles et les tendances importants.

L'analyse de la composante principale comporte cinq phases :

- Normaliser l'intervalle des variables initiales continues.
- Calculer les corrélations en calculant la matrice de covariance.
- Déterminer les composantes primaires de la matrice de covariance en calculant ses vecteurs propres et ses valeurs propres.
- Créer un vecteur de caractéristiques pour décider des composantes principales à conserver.
- Transformer les données en utilisant l'axe des composants primaires.

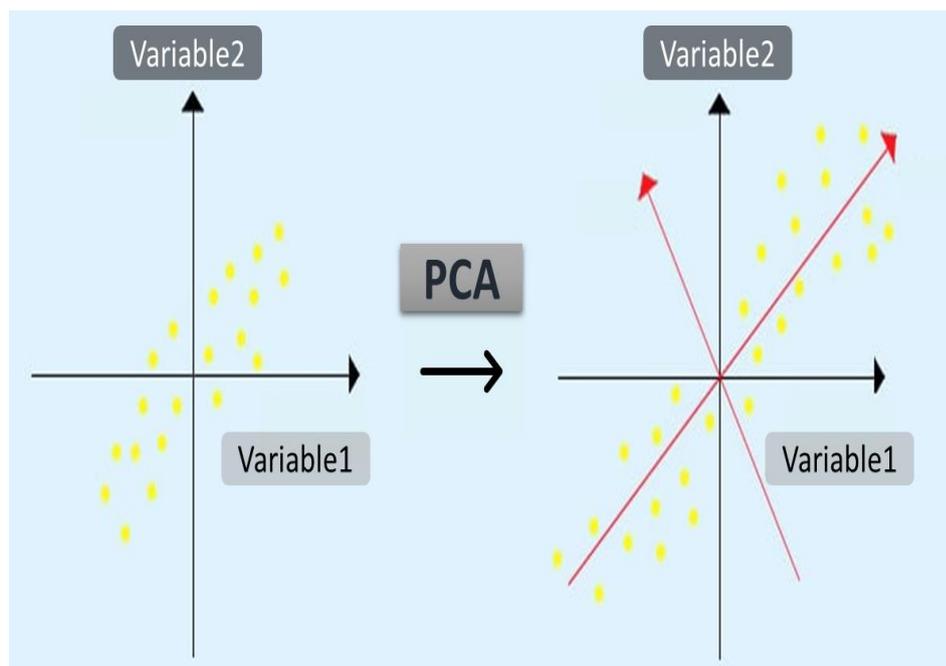


Figure 2.1 : Exemple d'extraction de caractéristiques à l'aide de la PCA.

L'analyse en composantes principales est une méthode couramment utilisée pour réduire la dimensionnalité des ensembles de données volumineux, en réduisant une grande collection de variables à une plus petite tout en conservant la majorité des informations de l'ensemble initial. La réduction de la dimensionnalité implique un compromis entre la précision et la simplicité, car une diminution du nombre de variables entraîne naturellement une perte de précision, mais facilite également l'analyse des données pour les algorithmes d'apprentissage automatique. En utilisant PCA, les variables originales sont combinées linéairement pour créer de nouvelles variables non corrélées appelées composantes principales, qui capturent l'essentiel de l'information contenue dans les données. En sélectionnant les composantes principales les plus informatives et en les utilisant comme nouvelles variables, la dimensionnalité des données peut être réduite tout en préservant la majeure partie des informations Figure 2.1. Cependant, il est important de noter que les composantes principales ne sont pas toujours facilement interprétables et n'ont pas nécessairement de signification intrinsèque. Elles représentent simplement les directions des données qui expliquent le mieux la variance observée. Par conséquent, PCA offre une perspective efficace pour examiner et évaluer les données en mettant en évidence les variations entre les observations. Ce processus permet de déterminer un ensemble de composantes principales qui capturent la variance maximale dans les données, avec chaque composante successive représentant la variance restante non capturée et étant orthogonale à toutes les composantes précédentes. Ce processus se répète jusqu'à ce que toutes les variables initiales soient représentées par les composantes principales.

La description détaillée de la PCA est la suivante :

- **NORMALISATION :** L'objectif de cette étape est de normaliser les plages de variables initiales continues afin que chacune d'entre elles apporte une contribution égale à l'analyse. Plus précisément, la raison pour laquelle la PCA est extrêmement sensible aux variances des variables initiales est la raison pour laquelle la normalisation doit être effectuée en premier. En d'autres termes, si les intervalles des variables initiales diffèrent de manière significative, les variables dont l'intervalle est le plus grand auront la priorité sur les variables dont l'intervalle est le plus petit (par exemple, une variable allant de 0 à 100 aura la priorité sur une variable allant de 0 à 1), ce qui produira des résultats biaisés. Ce problème peut donc être évité en convertissant les données à des échelles comparables.

Cela peut être réalisé mathématiquement pour chaque valeur de chaque variable en retirant la moyenne et en divisant par l'écart-type.

$$z = \frac{\text{valeur-moyenne}}{\text{écart-type}} \quad (2.1)$$

Toutes les variables seront converties sur la même échelle une fois la standardisation terminée.

- **CALCUL DE LA MATRICE DE COVARIANCE :** L'objectif de cette étape est de déterminer s'il existe un lien entre les variables de l'ensemble des données d'entrée et comment elles diffèrent de la moyenne les unes par rapport aux autres. En effet, les variables peuvent parfois inclure des informations redondantes en raison de fortes corrélations. Par conséquent, nous calculons la matrice de covariance pour trouver ces relations.

La matrice de covariance a pour entrées toutes les paires possibles de variables initiales et est une matrice symétrique $p \times p$ (où p est le nombre de dimensions). Par exemple, la matrice de covariance d'un ensemble de données tridimensionnel avec trois variables, x, y et z , est une matrice de données 3×3 obtenue à partir de :

$$\begin{bmatrix} Cov(x, x) & Cov(x, y) & Cov(x, z) \\ Cov(y, x) & Cov(y, y) & Cov(y, z) \\ Cov(z, x) & Cov(z, y) & Cov(z, z) \end{bmatrix} \quad (2.2)$$

Nous obtenons réellement les variances de chaque variable de départ sur la diagonale principale (du haut à gauche au bas à droite), puisque la variance d'une variable par rapport à elle-même est sa covariance ($Cov(a, a) = Var(a)$). En outre, les entrées de la matrice de covariance sont symétriques par rapport à la diagonale principale puisque la covariance est commutative ($Cov(a, b) = Cov(b, a)$), ce qui signifie que les sections triangulaires supérieure et inférieure sont égales.

- **CALCULER LES VECTEURS PROPRES ET LES VALEURS PROPRES DE LA MATRICE DE COVARIANCE POUR IDENTIFIER LES COMPOSANTES PRINCIPALES :** Pour trouver les composantes primaires des données, nous devons calculer deux concepts d'algèbre linéaire en utilisant la matrice de covariance : les vecteurs propres et les valeurs propres. La première chose à

comprendre à propos des valeurs propres et des vecteurs propres est qu'ils sont toujours appariés, ce qui signifie que chaque vecteur propre a des valeurs propres correspondantes. En outre, leur nombre correspond au nombre de dimensions des données. Par exemple, s'il y a trois variables dans un ensemble de données tridimensionnel, il y a également trois vecteurs propres et trois valeurs propres correspondantes.

Le secret de l'attrait des composantes principales réside dans les valeurs propres et les vecteurs propres, puisque les vecteurs propres de la matrice de covariance représentent les orientations de l'axe présentant la variance (ou l'information) la plus élevée, que nous appelons composantes principales. En outre, les valeurs propres, qui indiquent le degré de variation de chaque composante principale, ne sont que les coefficients annexés aux vecteurs propres.

Les composantes primaires sont obtenues en classant les vecteurs propres par ordre d'importance, de la valeur propre la plus élevée à la plus faible.

- **CRÉER UN VECTEUR DE CARACTÉRISTIQUES :** Comme nous l'avons vu dans la phase précédente, nous pouvons déterminer les principales composantes par ordre d'importance en calculant les vecteurs propres et en les classant par ordre décroissant de leurs valeurs propres. À ce stade, nous décidons quelles sont les composantes les plus importantes à conserver ou à rejeter (celles dont les valeurs propres sont faibles), puis nous combinons les composantes restantes pour créer une matrice de vecteurs connue sous le nom de vecteur de caractéristiques.

Les vecteurs propres des composantes que nous choisissons de retenir sont les colonnes d'une matrice qui constitue le vecteur de caractéristiques. Étant donné que l'ensemble de données final n'aura que p dimensions si nous décidons de ne conserver que p vecteurs propres (composantes) sur n , il s'agit de la première étape vers la réduction de la dimensionnalité.

- **REFONDRE LES DONNÉES SELON LES AXES DES COMPOSANTES PRINCIPALES :** À l'exception de la normalisation, aucune donnée n'est modifiée au cours des étapes précédentes ; au lieu de cela, l'ensemble des données d'entrée est toujours exprimé en termes d'axes originaux, ou de variables initiales. Il suffit de choisir les composantes primaires et de créer le vecteur caractéristique.

L'objectif de cette dernière étape est de réorienter les données de l'axe d'origine vers celles que les composantes principales (d'où le nom d'analyse des composantes principales) représentent en utilisant le vecteur de caractéristiques créé à l'aide des vecteurs propres de la matrice de covariance. Pour ce faire, on multiplie la transposition du vecteur caractéristique par la transposition de l'ensemble de données d'origine.

$$\text{Ensemble final de données} = \text{Vecteur de caractéristiques}^T * \text{Ensemble de données originales standardisées}$$

Dans le contexte de la reconnaissance faciale, la PCA transforme une vaste collection d'images faciales de haute dimension en un sous-espace de dimension inférieure, tout en conservant les informations essentielles qui distinguent un visage d'un autre. En examinant minutieusement un ensemble complet de données d'images faciales, la PCA identifie les composantes principales, qui sont essentiellement les axes de l'espace de données qui capturent les variations les plus significatives entre les images Figure 2.1. Ces composantes principales codent des caractéristiques faciales essentielles, englobant les variations dans la pose du visage, les conditions d'éclairage et les expressions. En condensant la dimensionnalité de l'espace des caractéristiques, La PCA simplifie les exigences informatiques des algorithmes de reconnaissance faciale tout en préservant les détails discriminatoires essentiels nécessaires à une classification précise des visages. Par conséquent, la PCA sert de pierre angulaire au développement de systèmes de reconnaissance faciale efficaces et résistants, capables d'identifier avec précision les visages dans une multitude de conditions et de scénarios.

L'application de la fonction noyau dans les algorithmes linéaires a permis de mettre au point des méthodes efficaces de reconnaissance des visages. La PCA à noyau (KPCA : Kernel Principal Component Analysis) est une version modifiée de la PCA linéaire qui opère dans un espace à haute dimension et est créée à l'aide d'une fonction à noyau. L'algorithme calcule les vecteurs propres primaires de la matrice du noyau, par opposition à ceux de la matrice de covariance.

La PCA à noyau utilise la PCA dans l'espace à noyau pour permettre la création de correspondances non linéaires. En raison de l'augmentation de la dimensionnalité, le processus de mise en correspondance de grandes quantités de données est rendu implicite (et rentable) par l'utilisation de fonctions de noyau.

Où les évaluations du noyau dans l'espace d'entrée correspondent à des produits de points dans l'espace de caractéristiques à plus haute dimension.

La PCA à noyau étant une méthode basée sur le noyau, la mise en correspondance effectuée par la PCA à noyau dépend du choix de la fonction noyau. Les choix possibles pour la fonction noyau comprennent le noyau linéaire (rendant la PCA à noyau égale à la PCA traditionnelle), le noyau polynomial tel que : $k(x, y) = (x^T y + 1)^2$ et le noyau gaussien tel

$$\text{que : } k(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}.$$

2.2.2 Analyse discriminante linéaire de Fisher

L'analyse en PCA est une technique qui permet de trouver une transformation linéaire de données de haute dimension dans un espace de dimension inférieure. Cette transformation est choisie de manière à maximiser la variance conservée et à minimiser l'erreur de reconstruction en utilisant les moindres carrés. Les moindres carrés sont utilisés pour minimiser l'erreur de reconstruction lors de la projection des données dans un sous-espace de dimension réduite [39] [52]. Cette approche vise à trouver une transformation qui conserve le maximum de variance des données tout en minimisant la distance entre les données originales et leurs projections dans le nouvel espace. En intégrant les moindres carrés dans le processus de PCA, nous cherchons à optimiser la conservation de la variance tout en réduisant l'erreur de reconstruction, ce qui permet d'obtenir une représentation compacte des données tout en préservant au mieux leur structure [39] [52].

Néanmoins, l'orientation de la variance la plus élevée peut ne pas être utile à des fins de classification, étant donné que les composantes identifiées par une PCA n'incluent pas nécessairement d'informations distinctives. Par conséquent, les échantillons projetés sont flous et il est impossible de les classer. Lorsque l'éclairage et l'expression faciale changent de manière significative, une grande partie de la fluctuation des données peut être attribuée à ces changements. Les approches PCA choisissent effectivement un sous-ensemble de dimensions qui préservent une part importante de la variance, de sorte que la similarité dans le sous-espace du visage n'est pas uniquement contrôlée par l'identité.

Pour pallier les défis inhérents à la PCA, une autre méthode de sous-espace, nommée Fisherface, est employée dans le cadre de l'identification faciale. Contrairement à l'approche basée sur les eigenfaces (les eigenfaces sont les composantes principales extraites des données d'images faciales à l'aide de la PCA. Elles représentent les directions dans l'espace de haute

dimension qui capturent la plus grande variation dans les images faciales [53]), Fisherface met en évidence des sous-espaces linéaires spécifiques à chaque classe, fournissant ainsi une méthode plus appropriée pour la classification des visages [54]. L'approche de réduction des dimensions employée dans Fisherface est l'analyse discriminante linéaire de Fishers (FLDA : Fishers linear discriminant analysis), qui optimise simultanément la dispersion entre les classes et la dispersion à l'intérieur des classes des données relatives aux visages. Dans l'ensemble, Fisherface est plus performant que eigenface car il utilise efficacement les informations discriminantes [54].

Les Figures 2.2 à 2.7 illustrent les limites de l'analyse en composantes principales dans le contexte de la classification. La Figure 2.2 présente deux ensembles de données bidimensionnelles, l'un appartenant à la classe 1 et l'autre à la classe 2. La Figure 2.3 montre le tracé des données soustraites de la moyenne. La Figure 2.3 présente deux vecteurs propres perpendiculaires qui correspondent à la matrice de covariance des ensembles de données. Le premier vecteur propre représente la direction dans laquelle les ensembles de données présentent la plus grande variation.

$$\text{Classe-1 donnés} = ([1., 2.], [2., 3.], [3., 3.], [4., 5.], [5., 5.])$$

$$\text{Classe-2 donnés} = ([1., 0.], [2., 1.], [3., 1.], [3., 2.], [5., 3.], [6., 5.])$$

La Figure 2.4 illustre le processus de reconstruction des données à l'aide des deux vecteurs propres, qui aboutit à une reconstruction parfaite des données d'origine. Les données présentent une séparabilité linéaire avec cette reconstruction. L'imprécision de la reconstruction devient évidente lorsqu'un seul vecteur propre, qui correspond à la variance la plus élevée, est utilisé, comme le montre la Figure 2.5. Les Figures 2.6 et 2.7 montrent que les données projetées dans un espace réduit (de 2D à 1D) ne sont pas linéairement séparables, mais que les ensembles de données originaux sont linéairement séparables. Cette observation est intrigante. De ce point de vue, on peut déduire que la PCA ne parvient pas à classer ces données parce que l'orientation de la composante principale s'aligne sur la variance la plus importante. La PCA est efficace lorsque la variance des ensembles de données contient l'information discriminante de la classe.

Toutefois, dans cet exemple, l'information discriminante de la classe est préservée dans leurs valeurs moyennes respectives, ce qui fait que la PCA ne parvient pas à séparer linéairement les données.

L'analyse discriminante linéaire maximise le rapport entre la dispersion entre les classes et la dispersion à l'intérieur des classes afin d'identifier la combinaison de caractéristiques qui distingue le mieux les classes. Le concept est fondé sur le principe selon lequel les classes similaires doivent former des groupes compacts, tandis que les classes dissemblables doivent être nettement séparées les unes des autres. L'objectif de la LDA est de réduire le nombre de dimensions dans les données tout en conservant autant que possible les informations qui distinguent les différentes classes.

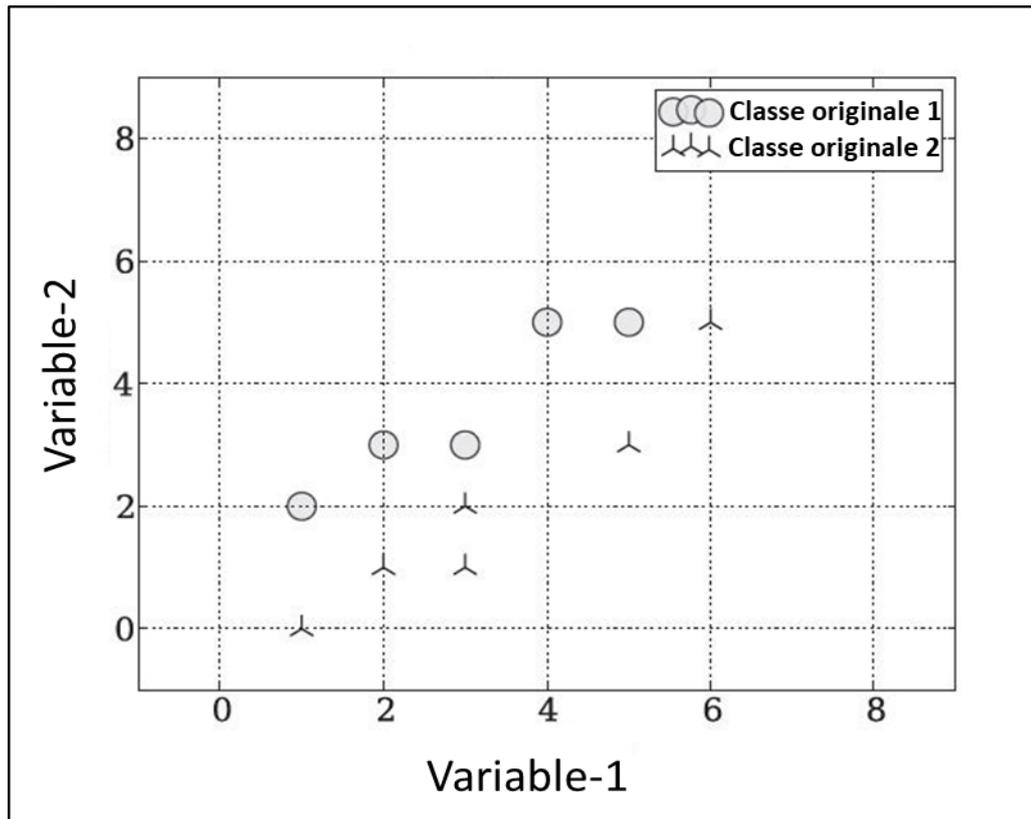


Figure 2.2 : Données composées de deux classes représentées dans un format bidimensionnel.

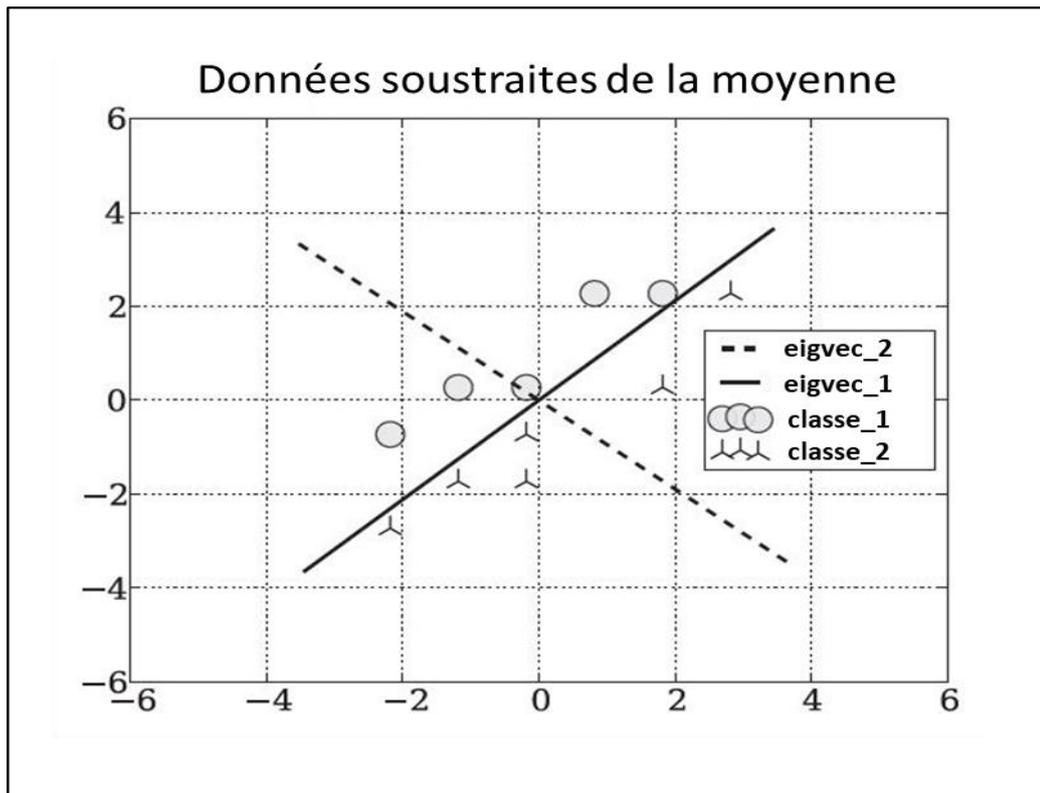


Figure 2.3 : Données soustraites de la moyenne de deux classes différentes.

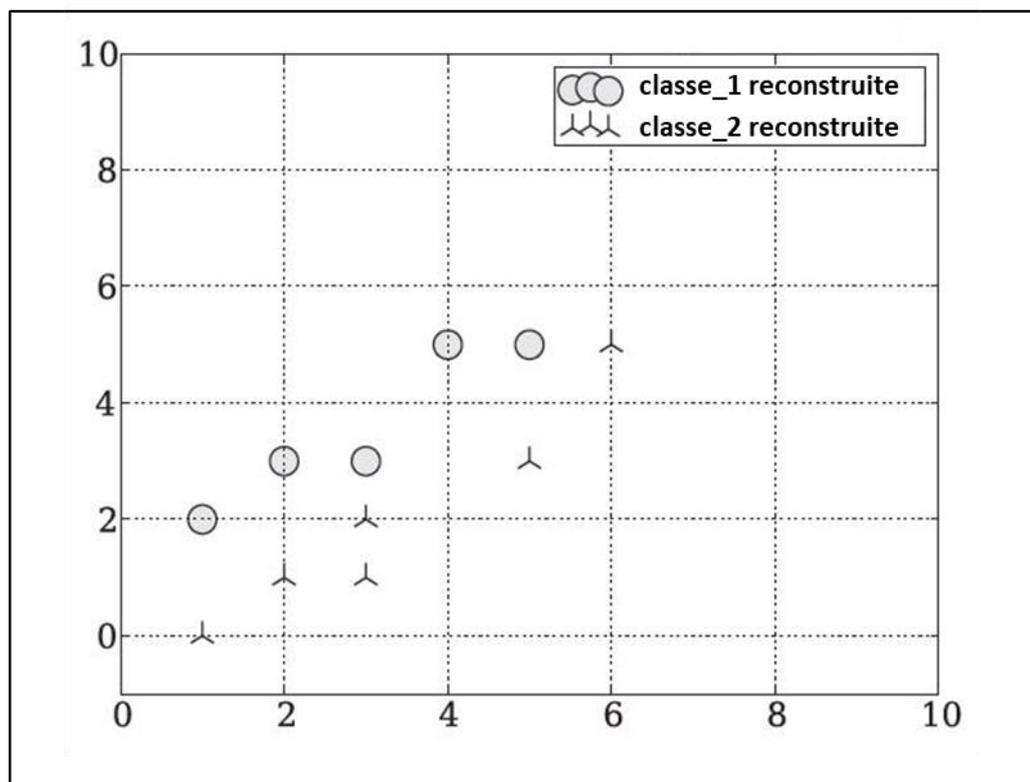


Figure 2.4 : Reconstruction des données avec tous les vecteurs propres (eigenvectors).

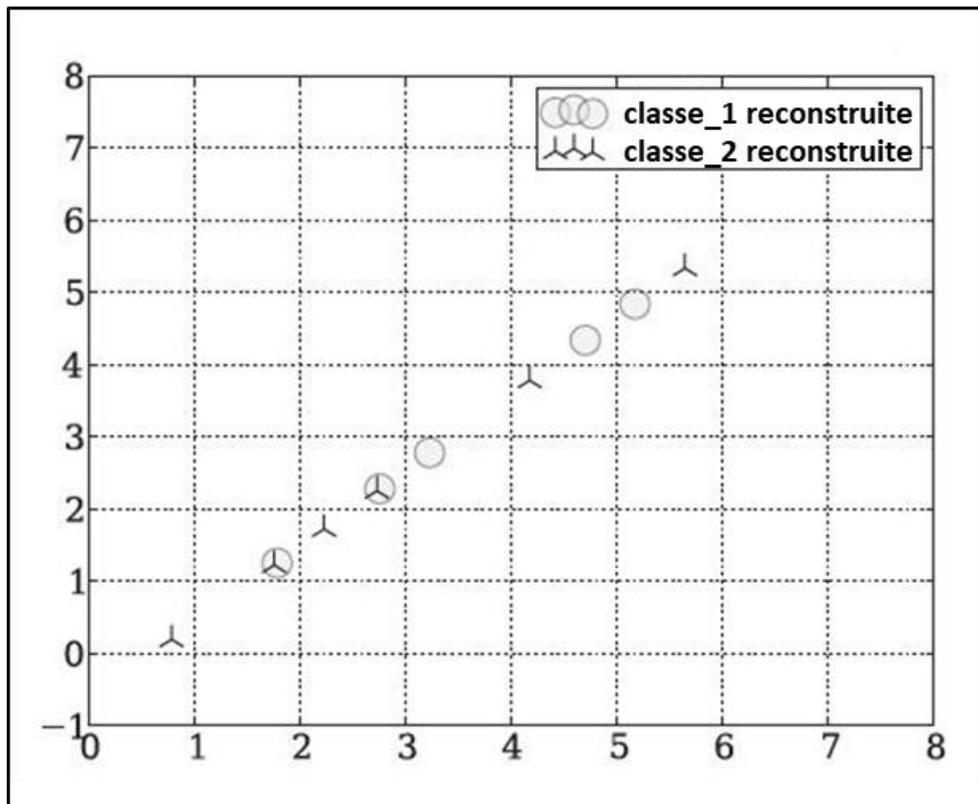


Figure 2.5 : Reconstruction des données avec un vecteur propre correspondant à la variance maximale.

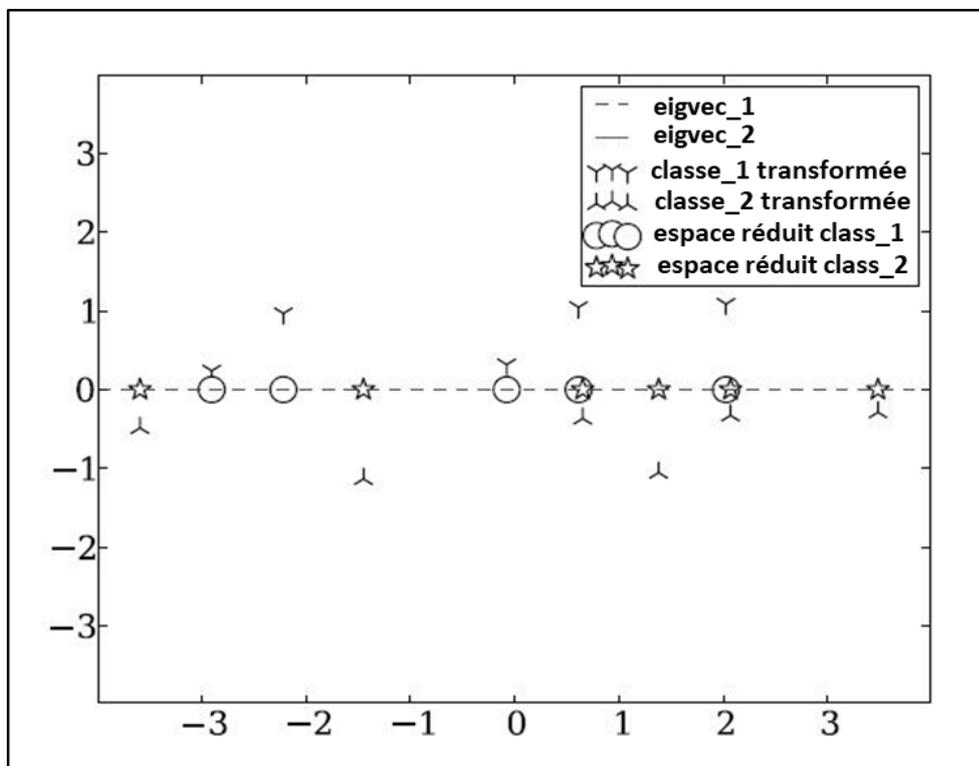


Figure 2.6 : Données dans un espace transformé où le système de coordonnées est basé sur les vecteurs propres (eigenvectors).

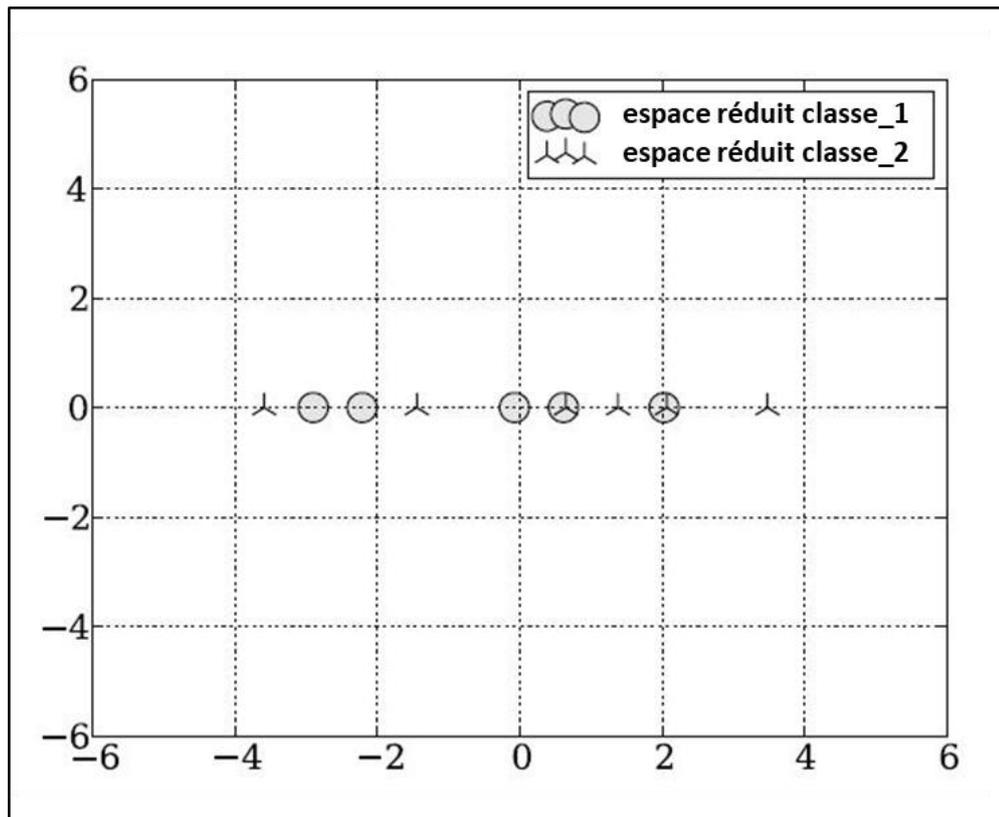


Figure 2.7 : Données non linéairement séparables dans un espace réduit.

2.3 Transformées et analyse multi-résolution

L'analyse fréquentielle multirésolutionnelle constitue une approche essentielle dans le traitement du signal et de l'image, offrant une représentation hiérarchique des informations à différentes échelles de résolution. Parmi les techniques les plus utilisées figurent la DCT, DWT et la DCT pyramidale. Ces méthodes présentent des avantages distincts pour la représentation et l'analyse des signaux et des images. La DWT permet une décomposition efficace des signaux en différentes fréquences et localisations spatiales, tandis que la DCT pyramidale offre une représentation compacte basée sur les coefficients de cosinus, utilisée dans une variété d'applications, notamment la compression d'images et la reconnaissance automatique des visages.

Dans cette section, nous explorerons en détail ces techniques d'analyse fréquentielle multirésolutionnelle, en mettant en lumière leurs principes fondamentaux, leurs applications et leurs implications dans le domaine de la reconnaissance automatique du visage.

2.3.1 Transformée discrète en cosinus

La DCT permet de séparer l'image en parties (ou sous-bandes spectrales) d'importance différente (en ce qui concerne la qualité visuelle de l'image). La DCT transforme un signal ou une image du domaine spatial au domaine fréquentiel.

La représentation générale d'une DCT 1D (N éléments de données) est définie par l'équation suivante :

$$F(u) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1} A(i) \cos \left[\frac{\pi \cdot u}{2 \cdot N} (2i+1) \right] \cdot f(i) \quad (2.3)$$

où :

$$A(\xi) = \begin{cases} \frac{1}{\sqrt{2}} & \text{pour } \xi = 0 \\ 1 & \text{ailleurs} \end{cases} \quad (2.4)$$

On peut définir l'équation générale d'une DCT 2D (image N par M) comme suit :

$$F(u, v) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \left(\frac{2}{M}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} A(i) A(j) \cdot \cos \left[\frac{\pi \cdot u}{2 \cdot N} (2i+1) \right] \cos \left[\frac{\pi \cdot v}{2 \cdot M} (2j+1) \right] \cdot f(i, j) \quad (2.5)$$

où :

- $f(i, j)$ est l'intensité du pixel de la ligne i et la colonne j .
- $F(u, v)$ est le coefficient DCT de la matrice DCT.

La DCT joue un rôle important dans la reconnaissance des visages en raison de sa capacité à représenter efficacement les images dans le domaine des fréquences. Dans les systèmes de reconnaissance faciale, la DCT est souvent utilisée comme technique d'extraction de caractéristiques pour capturer les caractéristiques essentielles des images faciales comme illustré dans Figure 2.8. En transformant les images faciales en coefficients DCT, qui représentent les composantes de fréquence de l'image, il est possible d'encoder efficacement les caractéristiques faciales importantes. Les coefficients DCT sont ensuite généralement utilisés comme entrée dans les algorithmes de classification pour les tâches de reconnaissance faciale. En outre, la DCT offre des avantages tels que la représentation compacte des données d'image, les propriétés de compactage de l'énergie et la robustesse au bruit. En outre, les

techniques de normalisation appliquées aux coefficients DCT peuvent améliorer la capacité du système à gérer les variations de l'expression faciale, de la pose et de l'éclairage, contribuant ainsi à améliorer la précision de la reconnaissance des visages. La DCT est donc un outil fondamental pour le développement de systèmes de reconnaissance faciale robustes et efficaces.

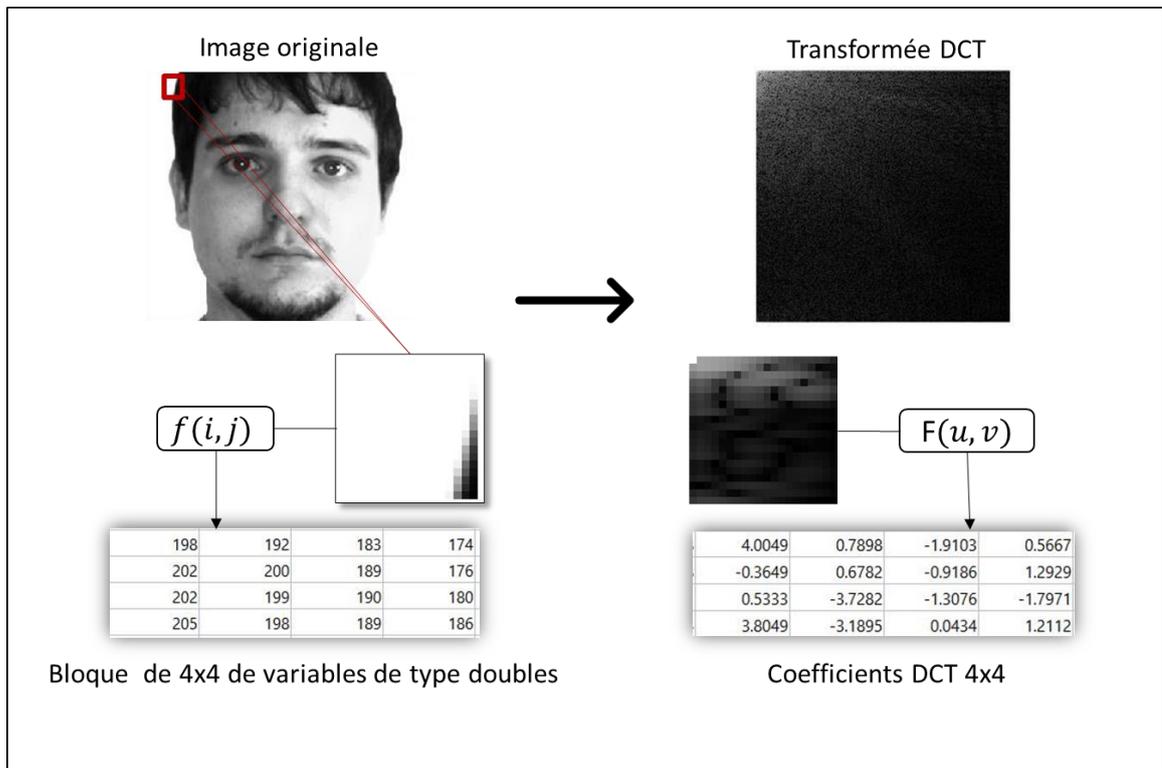


Figure 2.8 : Représentation d'une transformation en DCT d'une image faciale dans AR[13].

2.3.2 Décomposition en ondelettes

La conversion en ondelettes est un cas général de la transformée de Fourier. La transformée de Fourier décompose une fonction en ondes sinusoïdales et cosinusoidales (qui sont utilisées comme fonctions de base) avec des périodes de longueur différentes, mais les mêmes fonctions de base couvrent l'ensemble de l'intervalle. Toutefois, l'ondelette permet à la fois de mesurer et de décaler les fonctions de base. De plus, il n'est pas obligatoire que les fonctions de base soient les ondes sinusoïdales et cosinusoidales, même s'il existe des fonctions de base célèbres de ce genre. N'importe quelle fonction peut être sélectionnée comme fonction de base, à condition qu'elle réponde à la condition de base d'une ondelette, c'est-à-dire qu'elle possède une énergie finie.

En général, la transformée en ondelettes peut être exprimée par l'équation suivante :

$$F(a,b) = \int_{-\infty}^{+\infty} f(x)\psi_{(a,b)}^*(x) dx \quad (2.6)$$

où le * est le symbole du complexe conjugué et la fonction ψ est une fonction quelconque. Cette fonction peut être choisie arbitrairement à condition qu'elle obéisse à certaines règles.

La transformée en ondelettes est une collection infinie de transformées qui varient en fonction de la fonction de mérite utilisée pour le calcul. Le terme « transformée en ondelettes » est couramment utilisé dans divers contextes et applications, ce qui est la principale raison de sa large reconnaissance. Il existe plusieurs méthodes pour classer les différents types de transformées en ondelettes (Figure 2.9).

La transformée en ondelettes discrète est une méthode de transformation en ondelettes utilisant une collection discrète d'échelles et de translations d'ondelettes qui adhèrent à des contraintes spécifiques. En d'autres termes, cette transformée décompose le signal en un groupe d'ondelettes mutuellement orthogonales. C'est la principale différence avec la transformée en ondelettes continue ou sa version pour les séries temporelles discrètes, souvent connue sous le nom de transformée en ondelettes continue à temps discret. L'ondelette peut être dérivée d'une fonction d'échelle qui caractérise ses propriétés d'échelle. La nécessité pour les fonctions d'échelle d'être orthogonales à leurs translations discrètes leur impose certaines contraintes mathématiques, qui sont largement discutées, telles que l'équation de dilatation :

$$\phi(x) = \sum_{k=-\infty}^{+\infty} a_k \phi(S_x - k) \quad (2.7)$$

où S est un facteur d'échelle (généralement choisi comme 2). En outre, la zone entre la fonction doit être normalisée et la fonction de mise à l'échelle doit être orthogonale à ses translations entières, à savoir

$$\int_{-\infty}^{+\infty} \phi(x)\phi(x+l) dx = \delta_{0,l} \quad (2.8)$$

Les méthodes multirésolutionnelles sont largement utilisées pour l'extraction de caractéristiques ainsi que dans les applications de compression d'images et de vidéos. Les méthodes multirésolutionnelles largement adoptée dans la reconnaissance automatique des

visages et le traitement d'images en général. Elle consiste à décomposer une image en une structure de sous-bandes multi résolution. Cette décomposition bidimensionnelle en ondelettes est réalisée en appliquant un ensemble de filtres à deux canaux à l'image, filtrant les fréquences basses et hautes de manière répétée et sous-échantillonnant l'image dans les directions horizontale et verticale. Cette méthode produit des images d'approximation et des images de détail, représentant respectivement les informations basse-fréquence et les détails dans les directions horizontale, verticale et diagonale.

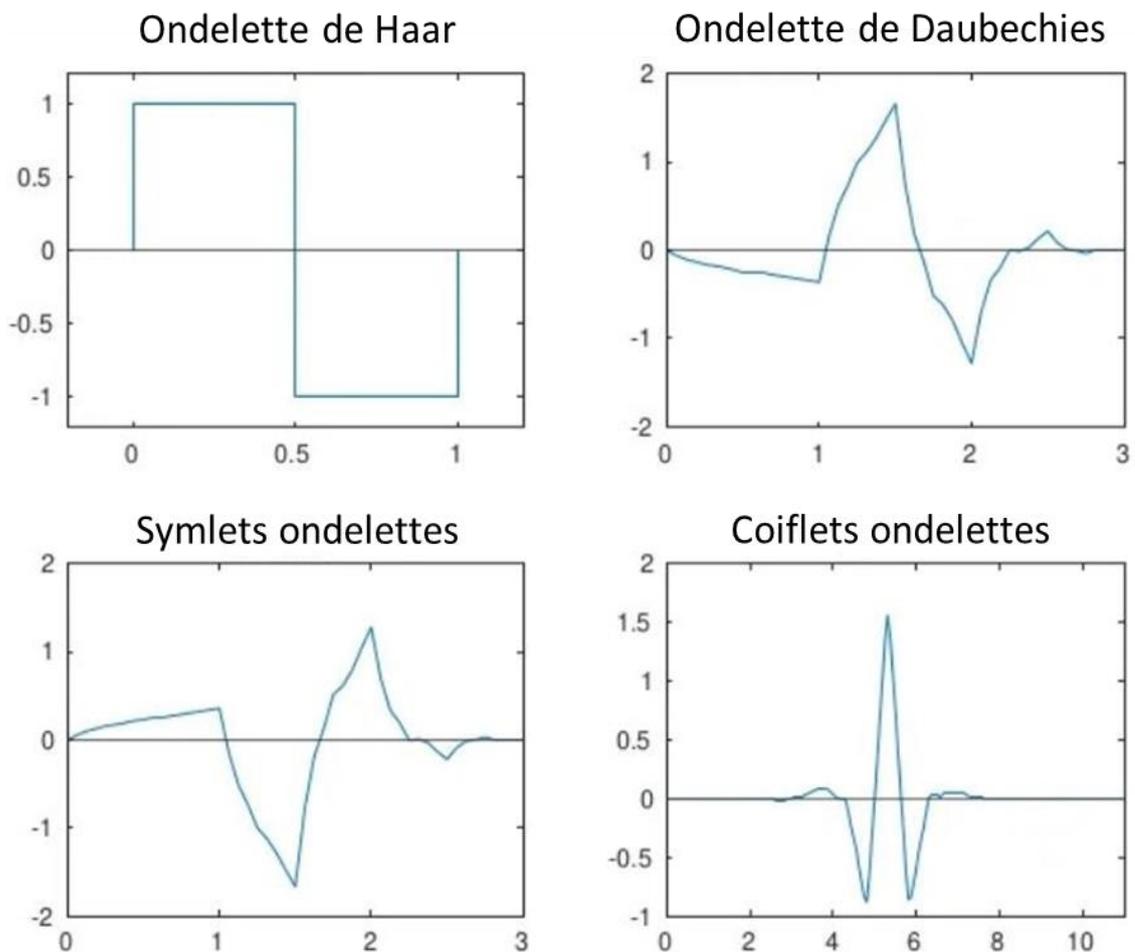


Figure 2.9 : Fonctions de base de la transformée en ondelettes.

Dans la décomposition classique en ondelettes, seules les sous-bandes filtrées par un filtre passe-bas sont prises en compte, tandis que dans la décomposition en paquets d'ondelettes, une généralisation de la méthode classique, les filtres sont itérés sur toutes les bandes de fréquences. La Figure 2.10 illustre un exemple d'une image faciale décomposée à l'aide de ces deux méthodes.

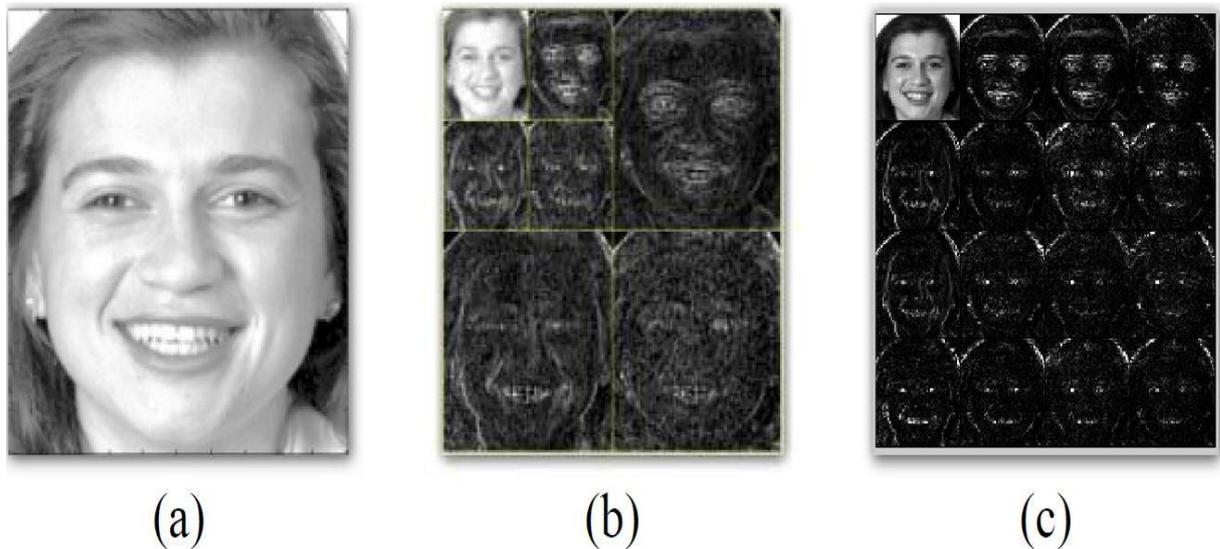


Figure 2.10 : a) Une image de visage. (b) L'image du visage soumise à une décomposition en ondelettes 2D. (c) L'image du visage soumise à une décomposition en paquets d'ondelettes 2D.

2.3.3 Décomposition en DCT-PYRAMID

Les méthodes multi résolution sont largement utilisées dans divers domaines, notamment l'extraction de caractéristiques, la compression et le codage [55], [56], [57]. La pyramide DCT, qui décompose une image d'entrée en plusieurs niveaux d'approximations et de détails, constitue l'une des approches permettant d'obtenir une décomposition multi résolution. Une décomposition pyramidale DCT à un niveau est essentiellement obtenue en employant une décimation DCT par bloc comme suit : Tout d'abord, l'image d'entrée est divisée en blocs de taille $N \times N$. Ensuite, chaque bloc est soumis à une transformation via la DCT directe $N \times N$. La DCT inverse (IDCT), mise à l'échelle par un facteur M/N , est ensuite appliquée aux composantes basses fréquences $M \times M$ du bloc transformé. Ici, $M < N$ représente la taille d'échantillonnage descendante souhaitée. Ce processus permet d'obtenir une version à plus faible résolution du bloc original, appelée "coin", c'est-à-dire que le bloc original de taille $N \times N$ est décimé en un coin à plus faible résolution de taille $M \times M$. En outre, les composantes hautes fréquences restantes, qui représentent les détails fins du bloc original, forment une structure en L inversée. En rassemblant les coins dans le même ordre que les blocs originaux correspondants, la sous-bande d'approximation, qui est une version à plus faible résolution de l'image d'entrée, est construite. Inversement, les formes en L inversées qui en résultent constituent la sous-bande des détails, qui contient les composantes à hautes fréquences de l'image d'entrée. Par conséquent, la pyramide DCT génère une version à faible résolution dont

la dimensionnalité est réduite d'un facteur M/N , ainsi qu'une sous-bande de détails comprenant toutes les formes L inversées qui en résultent.

La Figure 2.11 illustre la mise en œuvre pratique d'une décomposition pyramidale DCT à un niveau 80×80 d'une image faciale. Dans cet exemple, en utilisant une taille de bloc 8×8 et une taille d'échantillonnage $M = 4$, l'image est divisée en 100 blocs.

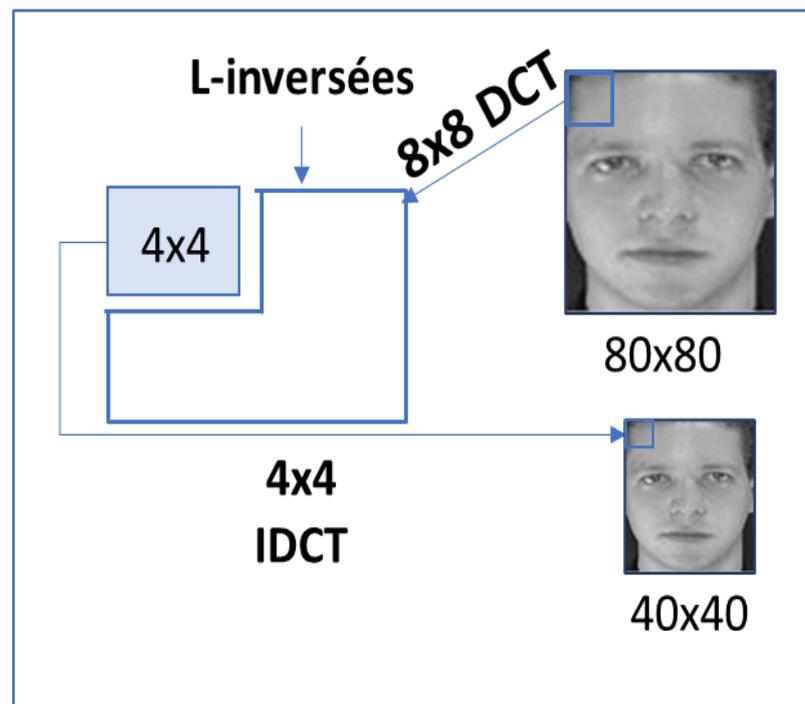


Figure 2.11 : Décomposition d'une pyramide DCT d'un niveau.

Dans un premier temps, la DCT directe est appliquée à chaque bloc. Ensuite, les composantes à basse fréquence dans les blocs transformés 4×4 subissent une transformation inverse à l'aide de 4×4 -IDCT. Les valeurs résultant de l'IDCT sont ensuite mises à l'échelle de manière appropriée par un facteur de $M/N = 1/2$, ce qui donne les coins correspondants. Simultanément, les composantes hautes fréquences restantes se manifestent sous formes L inversées. Enfin, les coins sont combinés dans l'ordre approprié pour former la sous-bande d'approximation de taille 40×40 , tandis que les formes en L inversées sont rassemblées pour constituer la sous-bande de détail. En appliquant de manière itérative la pyramide DCT à l'approximation obtenue au niveau précédent, on obtient plusieurs niveaux de décomposition. Chaque niveau capture des détails plus fins et fournit une représentation plus détaillée de l'image.

2.4 Réseaux de neurones artificiels

Les unités, ou neurones artificiels, sont des composants des réseaux neuronaux artificiels. Le réseau de neurones artificiels d'un système est constitué de ces unités regroupées dans une séquence de couches. Le nombre d'unités dans une couche peut varier de quelques dizaines à des millions, en fonction du nombre de réseaux neuronaux complexes nécessaires pour découvrir les modèles cachés de l'ensemble de données. Les réseaux neuronaux artificiels se composent généralement de couche cachée, de couche de sortie et de couche d'entrée [58]. La couche d'entrée est l'endroit où les données externes sont introduites dans le réseau neuronal à des fins d'analyse ou d'éducation. Ensuite, les données passent par une ou plusieurs couches cachées, qui convertissent l'entrée en données utiles pour la couche de sortie. Enfin, la réaction du réseau neuronal artificiel aux données d'entrée fournies est présentée comme une sortie par la couche de sortie.

Dans la plupart des réseaux neuronaux, les unités sont connectées d'une couche à l'autre, Figure 2.12. Les poids attribués à chacune de ces relations indiquent l'effet d'une unité sur les autres. Le réseau neuronal acquiert de plus en plus de connaissances sur les données au fur et à mesure qu'il passe d'une unité à l'autre, pour finalement produire une sortie à partir de la couche de sortie.

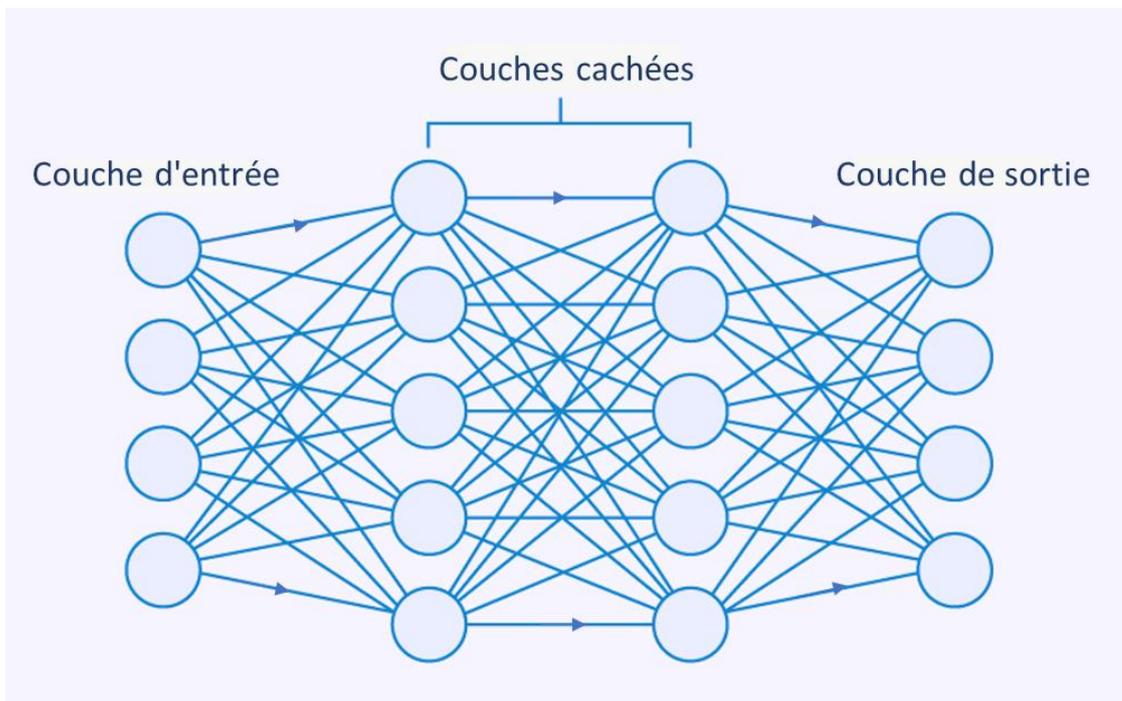


Figure 2.12 : Architecture des réseaux de neurones.

Les réseaux neuronaux artificiels sont basés sur l'architecture et les fonctions des neurones humains. La première couche d'un réseau neuronal artificiel, appelée couche d'entrée, transfère des données provenant de sources extérieures à la deuxième couche, appelée couche cachée. Chaque neurone de la couche cachée reçoit des informations des neurones de la couche supérieure, calcule la somme pondérée et la transmet aux neurones de la couche inférieure. Comme ces connexions sont pondérées, les effets des entrées de la couche précédente sont essentiellement maximisés en donnant à chaque entrée un poids unique, qui est ensuite modifié au cours de l'apprentissage pour améliorer les performances du modèle.

2.4.1 Neurones artificiels et neurones biologiques

Les réseaux neuronaux artificiels et les neurones biologiques que l'on trouve dans le cerveau humain sont comparables dans leur forme et leur fonction, d'où l'origine du concept de réseaux neuronaux artificiels (Figure 2.13 et Tableau 1.1).

Tableau 1.1 : Comparaison entre Neurones Biologiques et Neurones Artificiels

Neurones biologiques	Neurones artificiels
Dendrites	Entrées
Noyau cellulaire ou Soma	Nœuds
Les synapses	Poids
L'axone	Sortie

- **Structure :** Les neurones biologiques servent d'inspiration pour l'architecture des réseaux neuronaux artificiels. Un axone envoie des impulsions aux neurones voisins à partir du corps cellulaire, ou soma, d'un neurone biologique, qui traite les impulsions. Les réseaux neuronaux artificiels sont composés de nœuds d'entrée qui acceptent les signaux d'entrée, de nœuds de couche cachée qui calculent ces signaux d'entrée et de nœuds de couche de sortie qui utilisent des fonctions d'activation pour traiter les résultats de la couche cachée afin de calculer la sortie finale.

- **Les synapses :** Les connexions entre les neurones biologiques, appelées synapses, permettent aux impulsions de passer des dendrites au corps cellulaire. Dans les neurones artificiels, les synapses sont les poids qui relient les nœuds d'une couche aux nœuds de la couche suivante. La valeur du poids détermine la force des liens.
- **Apprentissage :** Le soma, ou noyau du corps cellulaire, des neurones biologiques est le lieu de l'apprentissage. Il contient un noyau qui aide au traitement des impulsions. Si les impulsions sont suffisamment fortes pour franchir le seuil, un potentiel d'action est créé et se déplace dans les axones. La plasticité synaptique, c'est-à-dire la capacité des synapses à modifier leur force au fil du temps en réponse aux variations de leur activité, rend ce processus possible. La rétropropagation est une méthode d'apprentissage dans les réseaux neuronaux artificiels qui modifie les poids entre les nœuds en fonction de l'erreur ou des variations entre les résultats attendus et les résultats réels.
- **Activation :** Lorsqu'une impulsion est suffisamment forte pour franchir le seuil et provoquer l'activation d'un neurone, on parle d'activation dans les neurones biologiques. Une fonction d'activation est une fonction mathématique qui fait correspondre l'entrée à la sortie et effectue des activations dans les réseaux neuronaux artificiels.

Tableau 1.2 : Analogies entre Neurones Biologiques et Artificiels dans le Processus d'Apprentissage

Neurones biologiques	Neurones artificiels
La plasticité synaptique	Rétropropagations

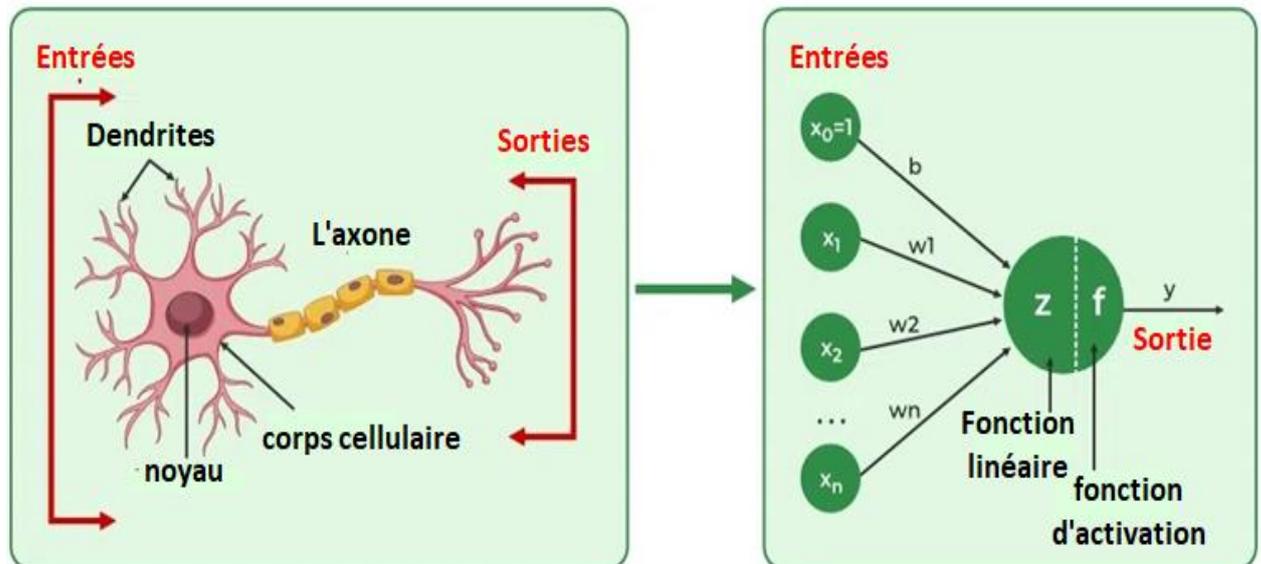


Figure 2.13 : Comparaison des neurones biologiques par rapport aux neurones artificiels.

2.4.2 Apprentissage des réseaux neuronaux artificiels

Un ensemble d'apprentissage est utilisé pour former les réseaux neuronaux artificiels. Supposons qu'on veut entraîner un réseau neuronal artificiel à identifier un chat. Pour entraîner le réseau à reconnaître un chat, des milliers de photos de chats différentes lui sont présentées. On doit vérifier que le réseau neuronal est capable d'identifier avec précision des photos de chats une fois qu'il a été suffisamment entraîné avec des images de chats. Pour ce faire, l'ANN est entraîné à classer les images données en déterminant s'il s'agit ou non de photo de chats. Une explication de l'image d'un chat fournie par un humain valide la sortie produite par l'ANN. La rétropropagation est utilisée pour modifier les leçons acquises par l'ANN au cours de l'entraînement dans le cas où il effectue une identification inexacte [58]. La rétropropagation s'effectue en ajustant les poids des connexions dans les unités du réseau neuronal artificiel en fonction du taux d'erreur constaté. Ce processus est répété jusqu'à ce que le réseau neuronal artificiel puisse identifier avec précision un chat sur une image avec le moins de taux d'erreur possible.

2.4.3 Types de réseaux neuronaux artificiels

- **Réseau neuronal en amont (Feedforward Neural Network) :** L'un des types les plus fondamentaux de réseaux neuronaux artificiels est le réseau neuronal feedforward. Les données ou entrées fournies à ce ANN ne se déplacent que dans une seule direction. Il peut avoir ou non des couches cachées ; les données entrent

dans l'ANN par la couche d'entrée et en sortent par la couche de sortie. Ainsi, le réseau neuronal d'anticipation est généralement dépourvu de rétropropagation et ne présente que des ondes à propagation frontale [58].

- **Réseau de neurones Convolutionnels :** Les réseaux neuronaux convolutifs et les réseaux neuronaux feed-forward sont similaires en ce sens qu'ils utilisent tous deux des connexions pondérées entre les unités pour déterminer l'influence de chaque unité sur les autres. Toutefois, un CNN comprend une ou plusieurs couches convolutives, chacune d'entre elles effectuant une opération de convolution sur l'entrée avant d'envoyer la sortie, ou le résultat, à la couche suivante. Le CNN peut être utilisé pour le traitement de la voix et de l'image, ce qui est très utile pour les applications de vision par ordinateur [58].
- **Réseau neuronal à fonction de base radiale :** Les fonctions qui prennent en compte la distance d'un point par rapport au centre sont appelées fonctions à base radiale (RBF : Radial Basis Functions). Les fonctions RBF sont des fonctions bicouches. L'entrée est traduite dans chacune des fonctions de base radiale de la couche cachée dans la première couche, et la couche de sortie calcule la sortie dans l'étape suivante. Pour modéliser les données qui indiquent une tendance ou une fonction sous-jacente, on utilise généralement des réseaux de fonctions à base radiale [58].
- **Réseau de neurones récurrents :** Afin d'améliorer la prédiction des couches, le réseau neuronal récurrent (RNN : Recurrent Neural Network) stocke la sortie de la couche et la réinjecte dans l'entrée. La première couche du RNN est très similaire à celle d'un réseau neuronal feed-forward et, une fois la sortie de la première couche calculée, le réseau neuronal récurrent commence. Après cette couche, chaque unité conserve une partie des données de la phase précédente, ce qui lui permet de fonctionner comme une cellule de mémoire pendant les calculs [58].

2.5 Conclusion

Au cours de ce chapitre, nous avons examiné trois concepts essentiels qui sont à la base de la reconnaissance faciale contemporaine : les méthodes basées sur les sous-espaces, l'analyse multi-résolution et les réseaux de neurones artificiels. Ces idées constituent des fondements fondamentaux de l'étude, fournissant des outils et des perspectives essentiels pour l'élaboration de systèmes de reconnaissance faciale précis et solides.

On a étudié en profondeur les méthodes utilisant des sous-espaces, comme le PCA et le LDA, afin de leur aptitude à diminuer la dimensionnalité des données et à extraire les caractéristiques les plus distinctives des images faciales. Ils sont indispensables pour la classification précise des visages, car ils permettent de simplifier la représentation des données tout en préservant les informations les plus importantes.

Par la suite, nous nous sommes penchés sur l'analyse multi-résolution, soulignant son importance capitale dans la prise en compte à la fois des détails précis des visages et d'une perspective globale. Il est essentiel d'avoir cette capacité pour repérer et reconnaître les visages dans diverses situations, ce qui permet une adaptation efficace à des conditions variables comme l'éclairage et l'angle de vue.

Finalement, nous avons analysé l'importance des réseaux de neurones artificiels dans le domaine de la reconnaissance faciale. Ces réseaux permettent d'acquérir des représentations hiérarchiques à partir de données non structurées, ce qui améliore considérablement les performances des systèmes de reconnaissance faciale en leur permettant d'extraire automatiquement des caractéristiques spécifiques des visages.

En bref, ce chapitre a mis en évidence l'importance et la pertinence de ces principes fondamentaux dans le contexte bien particulier de la reconnaissance faciale. En ayant une compréhension et une maîtrise de ces concepts, nous sommes davantage aptes à élaborer et à mettre en place des systèmes de reconnaissance faciale avancés et solides, ce qui ouvre la voie à une variété d'applications pratiques dans différents domaines, tels que la sécurité, la technologie et l'interaction entre les hommes et les machines.

Chapitre III

Reconnaissance faciale par intégration de la DCT pyramide et les Réseaux de Neurones

3 Chapitre III

3.1 Introduction

Ce chapitre met en avant un système de reconnaissance faciale novateur qui vise à prendre en compte de manière efficace les caractéristiques faciales pertinentes et à permettre une identification précise des personnes.

Nous utilisons différentes approches pour classifier les visages, telles que la détection des visages à l'aide de la cascade de Haar, l'extraction de caractéristiques à l'aide de la pyramide DCT, la réduction de dimensionnalité du vecteur de caractéristiques par des mesures statistiques, et l'utilisation d'un réseau de neurones à rétropropagation (Backpropagation) à deux couches.

Dans un premier temps, nous exposons la technique de Haar cascade pour détecter les visages. Cette méthode se base sur des éléments visuels afin de repérer les zones potentielles du visage dans une image, ce qui constitue une étape essentielle dans le processus de reconnaissance faciale.

Ensuite, nous abordons les méthodes statistiques employées afin de diminuer la taille du vecteur de caractéristiques extrait par la DCT pyramide. L'objectif de cette étape de diminution de la dimensionnalité est de supprimer les répétitions et de conserver uniquement les informations discriminantes indispensables à la classification.

Finalement, nous exposons notre système de réseau de neurones à rétropropagation à deux couches, qui exploite les caractéristiques limitées pour identifier les visages. La capacité de ce réseau à reconnaître et à identifier les personnes à partir des images faciales lui permet de proposer une solution solide et évolutive pour la reconnaissance automatique des visages Figure 3.1.

Ce chapitre souligne l'efficacité et la solidité de notre approche intégrée en mettant en évidence les bénéfices de chaque élément du système et en offrant une évaluation empirique de sa performance dans des situations concrètes.

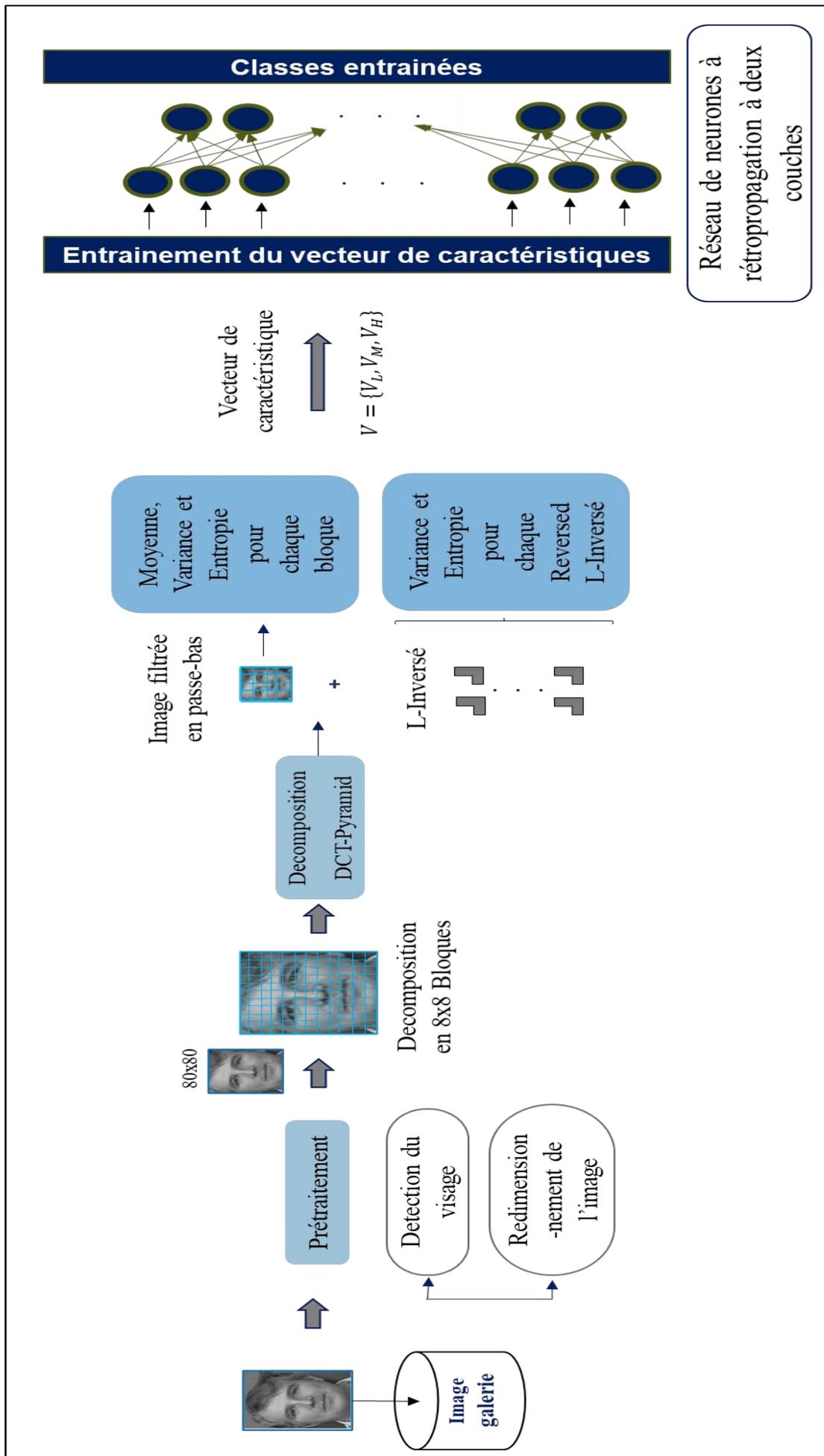


Figure 3.1 : Schéma Descriptif du système conçu.

3.2 Système de reconnaissance automatique des visages proposé

3.2.1 Prétraitement

Lors de l'étape de prétraitement, les images des visages sont recadrées à la taille du visage à l'aide de la méthode de détection des visages en cascade de Haar [59], qui associe un calcul efficace à l'apprentissage automatique afin d'obtenir des résultats précis. L'algorithme utilise des caractéristiques de type Haar, qui sont des filtres rectangulaires appliqués à l'image à différentes échelles et positions. Ces caractéristiques capturent des modèles de variations d'intensité dans l'image, tels que les bords, les coins et la texture. Ensuite, pour calculer efficacement les caractéristiques de type Haar à l'aide d'opérations arithmétiques simples et rapides, l'algorithme utilise des images intégrales. Une image intégrale est une représentation de l'image originale, où la valeur de chaque pixel est la somme des intensités des pixels dans une région rectangulaire à partir du coin supérieur gauche de l'image [59]. L'algorithme utilise également une technique d'apprentissage automatique appelée Adaboost pour former un classificateur [59]. Adaboost est une méthode d'apprentissage d'ensemble qui combine plusieurs classificateurs faibles pour former un classificateur fort [59]. Au cours de l'apprentissage, l'algorithme sélectionne les meilleures caractéristiques de type Haar et leurs seuils qui peuvent discriminer avec précision les visages et les non-visages dans les données d'apprentissage. Le classificateur formé est organisé en une cascade de classificateurs, où chaque classificateur est formé pour être plus complexe et plus précis que le précédent. La cascade est conçue pour rejeter rapidement les régions de l'image qui ne sont pas susceptibles de contenir un visage. L'algorithme applique la cascade entraînée à l'image d'entrée en la balayant avec une fenêtre coulissante à différentes échelles et positions. À chaque fenêtre, les caractéristiques de type Haar sont calculées à l'aide d'images intégrales et la cascade de classificateurs est appliquée pour déterminer si la fenêtre contient un visage ou non. Si une fenêtre est classée comme un visage par tous les classificateurs de la cascade, elle est considérée comme détectée [59]. Après l'étape de détection des visages, chaque image est redimensionnée à 80×80 pour la décomposition DCT-pyramide évoquée précédemment.

Quatre étapes peuvent être utilisées pour expliquer l'algorithme :

- Calcul des caractéristiques de Haar.
- Création d'images intégrales.

- Utilisation d'Adaboost.
- Implémentation de classificateurs en cascade.

Il est essentiel de garder à l'esprit que, comme d'autres modèles d'apprentissage automatique, cet algorithme a besoin d'un grand nombre d'images positives de visages et d'images négatives de non-visages afin d'entraîner le classificateur.

- **Calcul des caractéristiques de Haar :** La collecte des caractéristiques de Haar constitue l'étape initiale. Par essence, une caractéristique de Haar est un ensemble de calculs effectués sur des régions rectangulaires voisines en un point particulier d'une fenêtre de détection. Les intensités des pixels de chaque région sont additionnées et les différences entre les totaux sont ensuite calculées. Vous trouverez ci-dessous, Figure 3.2, quelques illustrations des caractéristiques de Haar.

La détermination de ces éléments dans une image complexe peut s'avérer difficile. C'est là que les images intégrales deviennent utiles, car elles minimisent le nombre d'opérations nécessaires.

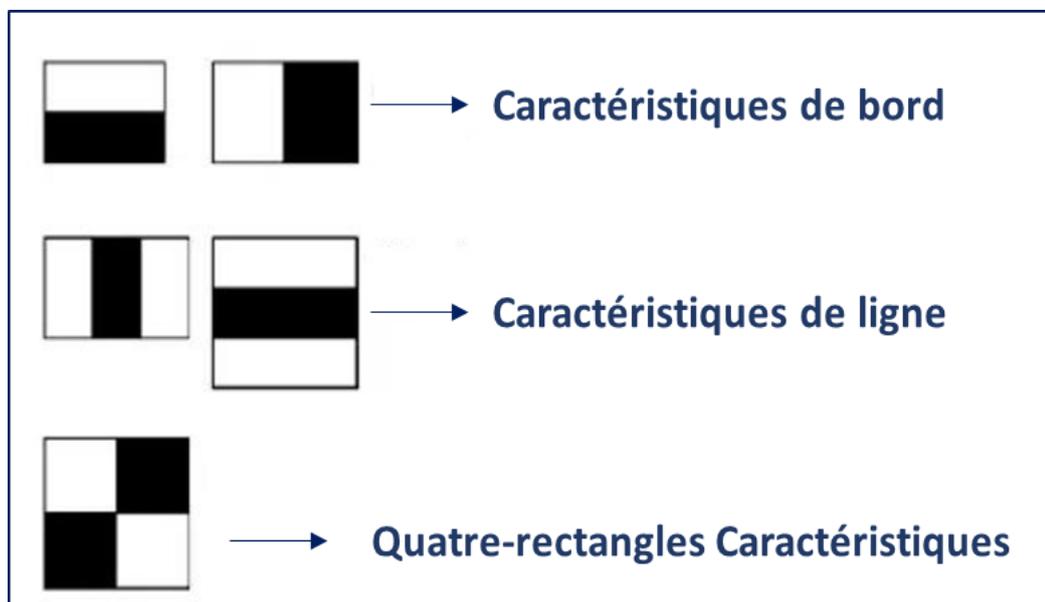


Figure 3.2 : Types de caractéristiques Haar.

- **Création d'images intégrales :** Les images intégrales accélèrent le processus de calcul des caractéristiques de Haar. Plutôt que d'effectuer des calculs à chaque pixel,

elles génèrent des références de tableau pour chaque sous-rectangle et sous-rectangle. Les caractéristiques de Haar sont ensuite calculées à partir de ces références.

Il est important de se rappeler que, lors de la détection d'un objet, presque toutes les caractéristiques de Haar sont insignifiantes - seules les caractéristiques de l'objet comptent. Mais comment trier des centaines de milliers de caractéristiques de Haar et obtenir les meilleures pour décrire un objet ? C'est là qu'Adaboost devient utile.

- **Entraînement Adaboost :** En termes simples, Adaboost sélectionne les meilleures caractéristiques et apprend aux classificateurs à les utiliser. L'algorithme peut identifier des éléments en utilisant un "classificateur fort" qui est produit en combinant un certain nombre de "classificateurs faibles". En faisant glisser une fenêtre sur l'image d'entrée et en calculant les caractéristiques Haar pour chaque zone de l'image, on obtient des apprenants faibles. Ce contraste est opposé à un seuil qui est appris pour distinguer les objets des non-objets. Il s'agit de "classificateurs faibles", ce qui signifie que pour créer un classificateur fort et précis, un nombre important de caractéristiques Haar est nécessaire.

En utilisant des classificateurs en cascade, la dernière phase fusionne ces apprenants faibles en un apprenant fort.

- **Utilisation pratique des classificateurs en cascade :** Le classificateur en cascade se compose de plusieurs étapes, chacune d'entre elles étant constituée d'un groupe d'apprenants faibles. L'amplification est utilisée pour former les apprenants faibles, ce qui permet de créer un classificateur très précis à partir de la prédiction moyenne de tous les apprenants faibles.

Le classificateur utilise cette prédiction pour déterminer s'il faut passer à la région suivante (négatif) ou signaler si un objet a été trouvé (positif). Des étapes sont franchies pour rejeter rapidement les échantillons négatifs, car la plupart des fenêtres ne contiennent rien d'intéressant.

Il est essentiel d'obtenir un taux de faux négatifs le plus bas possible, car cela aura un impact négatif sur le système de détection d'objets lorsqu'un objet est classé comme non-objectif.

Les cascades de Haar constituent un sous-ensemble de méthodes utilisées pour la détection d'objets. Il est crucial de minimiser l'apparition de faux négatifs lorsque l'on travaille avec des

cascades de Haar. Il est donc essentiel d'ajuster les hyperparamètres de manière appropriée au cours du processus d'apprentissage du modèle.

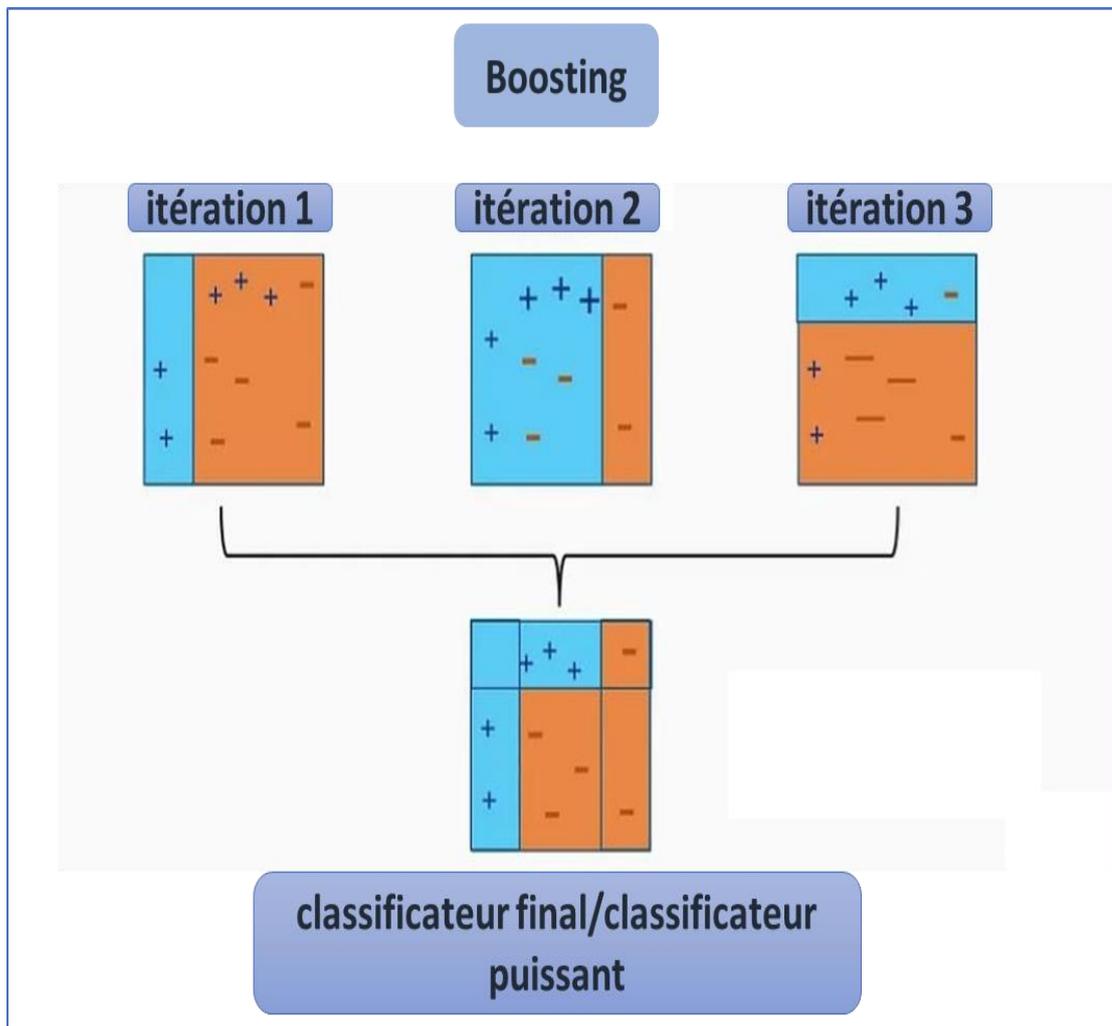


Figure 3.3 : Illustration d'un algorithme pour le boosting.

- Utilisations pratiques des cascades de Haar :

La technologie de reconnaissance faciale : peut être employée par de nombreux appareils électroniques et systèmes de sécurité pour vérifier l'authenticité d'un utilisateur pendant le processus de connexion à l'aide des cascades de Haar.

Robotique : Les robots ont la capacité de percevoir visuellement leur environnement afin d'effectuer des tâches grâce au processus de reconnaissance des objets. Ils peuvent par exemple être utilisés pour mécaniser les activités de fabrication.

Véhicules autonomes : Pour améliorer la sécurité et prendre des décisions plus éclairées, les véhicules autonomes s'appuient sur leur capacité à percevoir leur environnement. Les

cascales de Haar, peuvent être utilisées pour distinguer divers éléments tels que les piétons, les feux de circulation et les trottoirs.

La recherche d'images et la reconnaissance d'objets : impliquent l'utilisation d'algorithmes de vision artificielle, tels que les cascales de Haar, pour rechercher divers objets, en plus de la reconnaissance faciale.

Agriculture : Les classificateurs de Haar peuvent être utilisés pour vérifier la présence d'insectes nuisibles en vol, ce qui permet d'atténuer la pénurie de nourriture résultant des infestations de ravageurs.

Applications industrielles : Les classificateurs de Haar permettent aux machines de détecter et d'identifier des objets spécifiques, automatisant ainsi des tâches qui étaient auparavant limitées aux capacités humaines.

3.2.2 Extraction des caractéristiques

Les mesures statistiques sont une technique d'analyse descriptive utilisée pour résumer les caractéristiques d'un ensemble de données. Cet ensemble de données peut représenter l'ensemble de la population ou un échantillon de celle-ci. Les mesures statistiques peuvent être classées en mesures de tendance centrale et de diffusion.

Les mesures de tendance centrale décrivent certaines caractéristiques clés de l'ensemble de données en fonction des valeurs moyennes ou moyennes, car elles décrivent le centre des données. Les mesures de la tendance centrale que nous examinerons sont la moyenne, la mode et la médiane.

- **Moyenne :** La moyenne, également appelée moyenne mathématique d'un ensemble de données donné, peut être trouvée en additionnant toutes les valeurs de l'ensemble de données et en divisant par le nombre de valeurs n . Nous pouvons utiliser une

formule mathématique pour décrire cela : $\mu = \frac{\sum_{i=1}^n x_i}{n}$ où μ est utilisé pour

représenter la moyenne.

- **Mode :** La valeur qui apparaît le plus fréquemment dans un ensemble de données est le mode. Dans certains cas, il peut s'agir de plusieurs valeurs dans un ensemble de données donné. Elles sont toutes considérées ici comme le mode.

- **Médiane :** La valeur centrale d'une collection de données donnée est appelée médiane. Lorsqu'il y a deux valeurs médianes (un nombre pair de points de données), on doit calculer la moyenne des deux valeurs médianes. Il convient de réorganiser les valeurs par ordre croissant afin de trouver la médiane. Si le nombre de points de données est impair, on prend la valeur de $\frac{n+1}{2}$. Si le nombre de points de données est pair, on prend la valeur de $\frac{n}{2}$.

Les mesures de dispersion sont des mesures statistiques utilisées pour caractériser la diversité et la similarité des valeurs au sein d'un ensemble de données particulier. Étant donné que les mesures de tendance centrale ne prennent pas en compte les valeurs extrêmes, s'appuyer uniquement sur elles pour fournir une description complète d'un ensemble de données peut s'avérer très trompeur. Nous pouvons y parvenir en utilisant des mesures de dispersion telles que l'écart type, la variance et l'étendue.

- **Etendue :** est la différence entre les valeurs les plus élevées et les plus basses d'un ensemble de données donné. Elle permet de connaître l'étendue des données. Pour déterminer l'étendue, la valeur la plus faible des données est soustraite de la valeur la plus élevée. L'étendue ne peut pas être utilisée pour les distributions ouvertes et elle est affectée par les valeurs aberrantes. Ce sont là ses inconvénients mineurs.
- **Quartiles et intervalle interquartile :** Un quantile qui divise une collection de données ordonnées en quatre moitiés égales s'appelle un quartile (quarts). Il n'est pas possible de diviser un ensemble de nombres en quartiles. C'est le point limite de la division. L'écart entre les valeurs du quartile supérieur et du quartile inférieur s'appelle l'intervalle interquartile.
- **Variance et écart-type :** La variance et l'écart-type sont deux autres moyens d'évaluer la variabilité. L'écart-type est la racine carrée de la variance, qui est une mesure de la différence entre les points de données et la moyenne. Cela indique que la variance est la source à partir de laquelle l'écart-type est obtenu.

La variance est désignée par σ^2 .

L'écart-type est désigné par σ .

La formule de la variance de la population est la suivante : $\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$

où σ^2 est la variance de la population

N : la taille de la population

x_i : chaque valeur de la population

μ : la moyenne de la population.

La formule de la variance de l'échantillon est la suivante : $s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1}$

où s^2 : la variance de l'échantillon

n : la taille de l'échantillon

x_i : chaque valeur de l'échantillon

\bar{x} : la moyenne de l'échantillon

La formule de l'écart-type de la population est la suivante : $\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$

où σ est l'écart-type de la population.

La formule de l'écart-type de l'échantillon est donnée par : $s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}}$

où s écart-type de l'échantillon.

- **Entropie** : Selon la théorie de l'information, le degré moyen d'information ou d'incertitude concernant les résultats potentiels d'une variable aléatoire est connu sous le nom d'entropie. Dans le cas d'une variable aléatoire discrète X , distribuée sur la base de $p: \mathcal{X} \rightarrow [0,1]$, l'entropie est la suivante : $H(X) := -\sum_{x \in \mathcal{X}} p(x) \log p(x)$,

Claude Shannon a proposé l'idée de l'entropie de l'information, communément appelée entropie de Shannon, dans son article intitulé "A Mathematical Theory of Communication" [60]. Selon la théorie de Shannon, un système de communication de données se compose de trois éléments : une source de données, un canal de communication et un récepteur. Selon Shannon, le "problème fondamental de la communication" est la capacité du récepteur à déterminer, à partir du signal qu'il reçoit sur le canal, quelles données proviennent de la source. Dans sa célèbre théorie du codage des sources, Shannon a démontré que l'entropie reflète une limite mathématique absolue quant à l'efficacité avec laquelle les données de la source peuvent être compressées sans perte sur un canal totalement silencieux. Shannon a pris en compte diverses méthodes de codage, de compression et de transmission de messages à partir d'une source de données. Le théorème de Shannon sur le codage des canaux bruyants a considérablement renforcé ce résultat pour les canaux bruyants [60].

Extraction des caractéristiques pour le système proposé :

Pour l'extraction des caractéristiques, nous utilisons une pyramide DCT à deux niveaux, qui est réalisée en deux étapes de décomposition et appliquée directement à l'image 80×80 du visage obtenu lors de l'étape de prétraitement précédente. Cette décomposition est réalisée en utilisant la dimension du bloc $N = 8$ et la taille d'échantillonnage $M = 4$. La première étape de décomposition produit une bande d'approximation comprenant $100 (M \times M)$ -coins, ainsi que les formes en L inversées correspondantes. Ensuite, la deuxième étape de décomposition fournit un autre ensemble de $25 (M \times M)$ -coins et de 25 formes en L inversées. Ces 25-coins décrivent effectivement la sous-bande basse fréquence de l'image du visage, tandis que les formes en L inversées correspondantes encapsulent la sous-bande moyenne fréquence. Inversement, les formes en L inversées obtenues lors de la première étape de décomposition englobent la sous-bande haute fréquence de l'image du visage. Les sous-bandes basses, moyennes et hautes fréquences obtenues couvrent ensemble l'ensemble du spectre de fréquences de l'image d'entrée, ce qui permet une compréhension globale de ses caractéristiques. Cependant, l'utilisation directe de ces sous-bandes comme entrée d'un algorithme de classification impliquerait la manipulation d'une quantité substantielle de données. Pour relever ce défi, nous proposons des mesures statistiques telles que la moyenne, la variance et l'entropie des coins individuels afin de réduire la dimensionnalité de la sous-bande basse fréquence. En conséquence, nous calculons la moyenne, la variance et l'entropie correspondantes du j^{th} angle \mathbf{C}_j :

$$\boldsymbol{\mu}_j = \frac{1}{M^2} \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \mathbf{C}_j(m,n) \quad (3.1)$$

$$\boldsymbol{\sigma}_j^2 = \frac{1}{M^2} \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} (\mathbf{C}_j(m,n) - \boldsymbol{\mu}_j)^2 \quad (3.2)$$

et

$$\mathbf{H}_j = - \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} p(\mathbf{C}_j(m,n)) \times \log_2(p(\mathbf{C}_j(m,n))) \quad (3.3)$$

respectivement, où $p(\mathbf{C}_j(m,n))$ désigne la probabilité associée à chaque valeur distincte $\mathbf{C}_j(m,n)$ présent dans le coin \mathbf{C}_j . Ce processus est appliqué à tous les coins de la sous-bande basse fréquence, ce qui aboutit à la création du sous-vecteur des caractéristiques basses fréquences \mathbf{v}_L donné par :

$$\mathbf{v}_L = \bigcup_{j=0}^{L_2-1} \{ \boldsymbol{\mu}_j, \boldsymbol{\sigma}_j^2, \mathbf{H}_j \} \quad (3.4)$$

où $L_2 = 25$ est le nombre total de coins au deuxième stade de la décomposition pyramidale DCT à deux niveaux. Afin de réduire la dimensionnalité des sous-bandes à moyennes et hautes fréquences, nous tentons d'utiliser une stratégie similaire à celle utilisée pour réduire la dimensionnalité de la sous-bande à basses fréquences. Cependant, on constate que les sous-bandes de moyennes et hautes fréquences sont constituées de formes en L inversées, qui représentent intrinsèquement le contenu en hautes fréquences des blocs, ce qui fait que leurs valeurs moyennes convergent vers zéro. Par conséquent, nous nous concentrons uniquement sur l'utilisation des mesures de variance et d'entropie pour les moyennes et hautes fréquences, et donc sur les sous-vecteurs de caractéristiques correspondants \mathbf{v}_M et \mathbf{v}_H sont respectivement données par

$$\mathbf{v}_M = \bigcup_{j=0}^{L_2-1} \{ \hat{\boldsymbol{\sigma}}_{2j}^2, \hat{\mathbf{H}}_{2j} \} \quad (3.5)$$

et

$$\mathbf{v}_H = \bigcup_{j=0}^{L_1-1} \{ \hat{\boldsymbol{\sigma}}_{1j}^2, \hat{\mathbf{H}}_{1j} \} \quad (3.6)$$

3.2.3 Classification :

Nous avons mis en œuvre et testé différentes méthodes pour l'étape de classification. En fin de compte, nous avons décidé d'utiliser l'ANN avec rétropropagation, qui s'est avéré performant, en particulier avec des structures plus complexes. La rétropropagation est un algorithme d'apprentissage supervisé très répandu, utilisé dans les réseaux neuronaux pour former des ANN à plusieurs couches de type feedforward. Un algorithme de rétropropagation pour un réseau neuronal feedforward simple tel qu'illustré à la Figure 3.4 peut être décrit comme suit :

- Initialiser les poids $\mathbf{w}_{ji}^{[l]}$ et les biais $\mathbf{b}_j^{[l]}$ du réseau neuronal de manière aléatoire ou avec de petites valeurs.
- Introduire l'ensemble des caractéristiques d'entrée $\mathbf{x}_j, j = 1, \dots, n$, dans la couche d'entrée du réseau neuronal, et calculer la somme pondérée des entrées et des biais pour chaque neurone dans chaque couche.

Faire passer le résultat par une fonction d'activation pour obtenir la sortie de chaque neurone dans les couches cachées et de sortie. C'est ce qu'on appelle la propagation vers l'avant, car les signaux d'entrée circulent de la couche d'entrée à la couche de sortie en passant par les couches cachées. Elle est réalisée à l'aide de :

$$\mathbf{a}_j^{[l+1]} = f \left(\sum_{i=1}^n \mathbf{w}_{ji}^{[l+1]} \mathbf{a}_i^{[l]} + \mathbf{b}_j^{[l]} \right) \quad (3.7)$$

où $\mathbf{a}_j^{[0]} = \mathbf{x}_j$ et f est une fonction d'activation.

Dans nos simulations, nous avons expérimenté plusieurs fonctions d'activation, notamment Rectified Linear Unit (ReLU), Sigmoid, Softmax et Hyperbolic Tangent Sigmoid (Tansig). Après une expérimentation approfondie, nous avons opté pour la fonction d'activation Tansig pour la couche cachée et la fonction d'activation Softmax pour la couche de sortie. Cette combinaison a permis d'obtenir des meilleures performances dans notre étude.

- Comparer la sortie du réseau neuronal aux valeurs de vérité terrain (cible) à l'aide d'une fonction de perte, qui quantifie la différence entre la sortie prédite et la sortie réelle, et sert de mesure de la performance du réseau neuronal. L'entropie croisée est utilisée comme fonction de perte dans les réseaux neuronaux de rétropropagation. Elle mesure la différence entre les probabilités prédites (sortie) et les étiquettes de classe

réelles (cible) de manière probabiliste et sert à guider le processus d'apprentissage en mettant à jour les poids et les biais du réseau neuronal pendant la rétropropagation. La fonction de perte d'entropie croisée est définie comme suit

$$E = -\sum_i y_i \log(p_i) \quad (3.8)$$

où y_i est la vérité de base pour la classe i^{th} dans la classification multi-classes, et p_i est la probabilité de la classe i^{th} prédite par le réseau neuronal.

- Utilisez un algorithme d'optimisation, tel que la descente de gradient ou l'une de ses variantes, pour mettre à jour les poids et les biais du réseau neuronal afin de minimiser la perte sur la base du gradient calculé. Les poids et les biais sont ajustés dans la direction opposée du gradient, avec un taux d'apprentissage contrôlant la taille du pas des mises à jour.
- Itérer la propagation vers l'avant, le calcul de la perte, la propagation vers l'arrière et les mises à jour des poids et des biais pendant un certain nombre d'époques ou jusqu'à ce qu'un critère de convergence soit satisfait. Cela permet au réseau neuronal d'apprendre à partir des données d'apprentissage et de mettre progressivement à jour ses poids et ses biais afin de minimiser la perte et d'améliorer la précision de la prédiction.

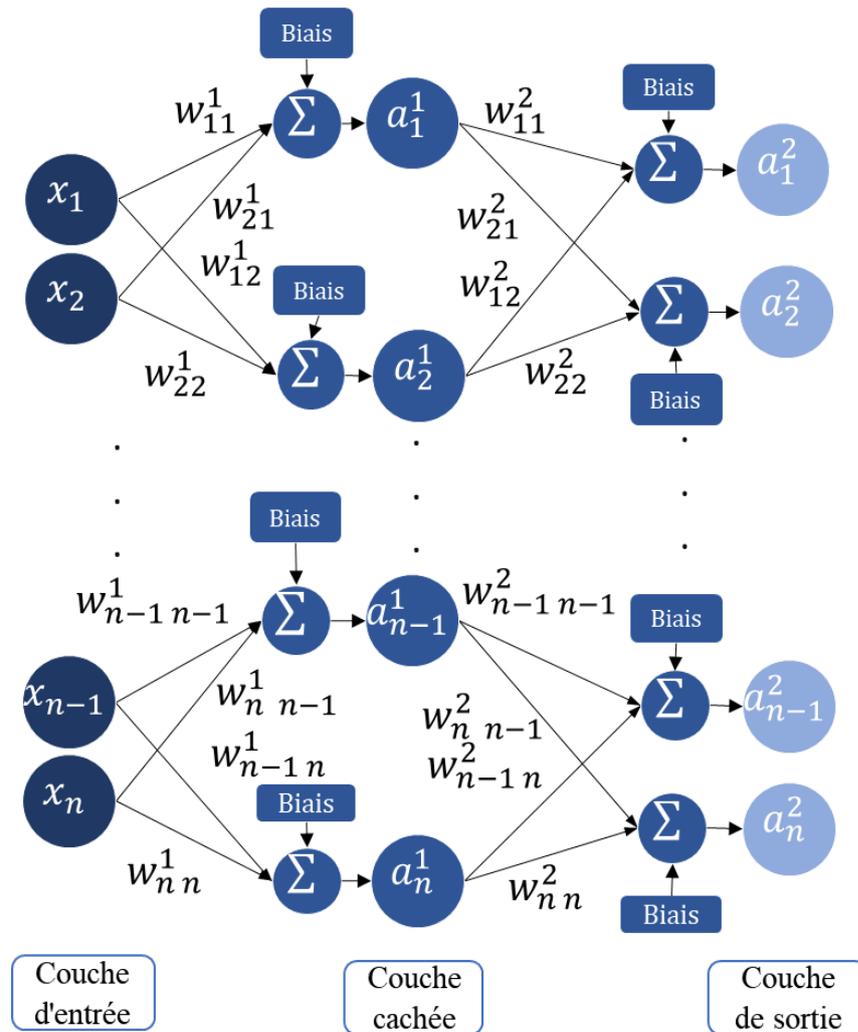


Figure 3.4 : Réseau neuronal de rétropropagation à deux couches.

3.3 Conclusion

Dans ce chapitre, nous avons exposé un nouveau système de reconnaissance faciale qui intègre diverses méthodes pour effectuer une classification efficace des visages. Nous utilisons la cascade de Haar pour détecter les visages, extraire les caractéristiques à l'aide de la pyramide DCT, réduire la dimensionnalité du vecteur de caractéristiques par des mesures statistiques, et exploiter un réseau de neurones à rétropropagation à deux couches.

Dans un premier temps, nous avons présenté la technique de la cascade de Haar pour détecter les visages, mettant en évidence son importance capitale dans l'identification des zones pertinentes du visage dans une image, une étape préliminaire cruciale dans le processus de reconnaissance faciale.

Par la suite, nous avons examiné les méthodes statistiques employées afin de diminuer la taille du vecteur de caractéristiques extrait à l'aide de la méthode DCT. La réduction de

dimensionnalité de cette étape a pour objectif de supprimer les redondances tout en maintenant les informations discriminantes indispensables à la classification.

Finalement, nous avons exposé notre réseau de neurones à rétropropagation à deux couches, qui tire parti des caractéristiques extraites afin d'identifier les visages. Ce réseau est capable de reconnaître et d'identifier les individus à partir des images faciales, ce qui en fait une solution solide et évolutive pour la reconnaissance automatique des apparences faciales.

Ce chapitre souligne la performance et la solidité de notre approche intégrée en exposant en détail les bénéfices de chaque élément du système. En outre, une évaluation pratique de la performance du système dans des situations réelles a été exposée, ce qui a confirmé sa fiabilité et son potentiel dans le domaine de la reconnaissance faciale.

En bref, notre système de reconnaissance faciale offre une solution complète et performante, capable de faire face aux défis complexes liés à l'identification précise des individus à partir d'images faciales. Ces progrès offrent de multiples applications concrètes dans différents secteurs, tels que la sécurité, la technologie et la gestion des identités.

Chapitre IV

Evaluation de notre Système de Reconnaissance Faciale

4 Chapitre IV

4.1 Introduction

Ce chapitre se concentre sur l'évaluation de notre système de reconnaissance faciale proposé, qui repose sur une combinaison innovante de la pyramide DCT et des ANN. Nous nous pencherons spécifiquement sur trois ensembles de données largement reconnus dans le domaine de la recherche en reconnaissance faciale : ORL, Yale et AR.

La diversité, la taille et la représentativité de ces ensembles de données en ont fait des références incontournables dans le domaine. L'objectif de notre système est d'évaluer sa capacité à détecter de manière précise et fiable les visages dans diverses situations, notamment en présence de variations d'éclairage, d'expressions faciales et de positions.

Dans les sections à suivre, nous analyserons en détail les résultats expérimentaux obtenus en utilisant ces ensembles de données. Nous comparerons également les performances de notre système avec celles de méthodes pertinentes et récemment publiées. En mettant l'accent sur des critères tels que la précision et la fiabilité, notre objectif est d'analyser les forces et les limites de notre méthode dans des contextes réels de reconnaissance faciale.

Nous examinons en profondeur différents résultats expérimentaux et les comparons à des méthodes pertinentes et récemment publiées [19], [21], [22], [26], [61-68].

4.2 Critères d'évaluation des systèmes de reconnaissance faciale

Le domaine de la reconnaissance faciale englobe des universitaires de plusieurs domaines tels que l'exploration de données, le traitement d'images, la reconnaissance de formes, les neurosciences, la psychologie, la vision par ordinateur et l'apprentissage automatique, entre autres. Le système de reconnaissance faciale est capable d'identifier un ou plusieurs individus à partir de photos ou de vidéos en utilisant une base de données préexistante de visages. Il s'agit d'une tâche impliquant la catégorisation d'images, notamment dans le but d'identifier et de distinguer automatiquement les visages. L'élément principal des systèmes de reconnaissance faciale est l'entraînement du système à l'aide d'images d'individus dont l'identité est déjà connue, puis la catégorisation d'images de test supplémentaires dans l'une des classes établies.

L'approche de l'évaluation des performances sert de référence pour évaluer l'efficacité de tout système de reconnaissance faciale. L'évaluation est cruciale pour comprendre la qualité du modèle ou de la technique, pour affiner les paramètres tout au long du processus d'apprentissage itératif et pour choisir le modèle ou la stratégie le plus approprié parmi un ensemble donné d'options. Différents critères sont utilisés pour évaluer les modèles pour diverses tâches. Cette section donne un aperçu des concepts fondamentaux et des méthodologies employées dans l'évaluation des systèmes de reconnaissance faciale.

4.2.1 Ajustement excessif

Les techniques modernes de détection des visages et les technologies de reconnaissance reposent principalement sur l'apprentissage automatique et la vision par ordinateur. Le surajustement est un phénomène courant dans le domaine de l'apprentissage automatique. Il s'agit d'une situation dans laquelle le système fait preuve de performances élevées sur les données d'apprentissage, mais ne parvient pas à maintenir son efficacité lorsqu'il est appliqué à un autre ensemble de données.

Pour illustrer ce concept à l'aide d'un exemple concret : considérons un scénario dans lequel on souhaite un système qui accorde l'accès à une porte exclusivement à des personnes spécifiques dans une entreprise, authentifiées par reconnaissance faciale. Pour atteindre cet objectif, on recueille des images des personnes auxquelles on souhaite donner accès, puis entraîne un algorithme à différencier ces personnes des autres. Après des essais réussis sur le terrain et le déploiement du système en production, les économies réalisées grâce à l'automatisation sécurisée de ce point d'accès peuvent être affectées à d'autres objectifs de l'entreprise.

De manière inattendue, le système cesse d'accorder l'accès aux nouveaux employés. Le système a acquis la capacité d'identifier les anciens employés, mais il n'a pas été soumis à des tests avec de nouvelles données. Il s'agit d'un exemple hyperbolique de surajustement, mais il illustre bien la notion.

Le processus de détection de l'ajustement excessif n'est pas simple ; il devient plus complexe lorsqu'un système de reconnaissance faciale est formé sur un ensemble de données qui comprend principalement des individus d'une seule ethnie. Le déploiement de ce système biométrique dans une zone multinationale entraînera probablement une diminution de la précision de la reconnaissance faciale du système. Le point commun de ces cas est la tendance à surestimer les capacités du logiciel en raison de tests inadéquats. Le principal défaut de

performance réside dans l'ensemble de données utilisé pour l'apprentissage, plutôt que dans l'ensemble de données qu'il rencontrerait dans des scénarios réels.

Comment éviter ces situations ? La principale leçon à tirer est qu'il faut éviter d'utiliser un ensemble de données d'apprentissage pour tester le logiciel. Pour garantir la sécurité et l'efficacité maximales de la classification, il est recommandé d'utiliser un ensemble de données de test dédié qui n'a pas été vu auparavant par le système. Avant de tester l'ensemble de données public, il faut s'assurer que le fournisseur ne l'a pas utilisé à des fins d'apprentissage et d'étalonnage. On se renseigne sur les sources de données spécifiques utilisées pour l'apprentissage, puis on sélectionne un ensemble de données qui ne figure pas dans la liste de classement. On examine l'ensemble de données avant de procéder au test et on évalue sa ressemblance avec les données que le système rencontrera au cours de la phase de production.

4.2.2 Aspects divers des mesures d'évaluation

Actuellement, plusieurs méthodes sont utilisées pour évaluer les performances du système de reconnaissance des visages. Cette section fournit des détails supplémentaires sur certaines d'entre elles. La méthode conventionnelle d'évaluation des systèmes de reconnaissance faciale repose sur le concept de vérité terrain, qui consiste à déterminer la précision des détections positives et négatives. La matrice de confusion est présentée dans le tableau 4.1. Les expressions "positif" et "négatif" indiquent la nature asymétrique des tâches de détection, où une classe représente la classe de motifs pertinents et l'autre la classe non pertinente.

Pour l'identification binaire ou la reconnaissance de deux classes, le système doit faire la distinction entre les critères d'identification des visages et les critères d'identification des non-visages. Le terme "vrai positif ou true positive (TP)" désigne la proportion d'images de visages correctement détectées par le système, tandis que le terme "faux positif ou false positive (FN)" désigne la proportion d'images de non-visages incorrectement identifiées comme des visages. Dans ce contexte, l'expression "vrai positif ou true positive" est synonyme de "taux de détection" et de "rappel". Les "faux positif ou false positive" correspondent à l'identification incorrecte de personnes en faisant correspondre leurs images avec celles de la base de données, tandis que les faux négatifs indiquent l'impossibilité d'identifier des personnes même lorsque leur photo est présente dans la base de données. Les deux principales courbes d'évaluation sont la courbe des caractéristiques de fonctionnement du récepteur (ROC : Receiver Operating Characteristic) et la courbe de précision et de rappel (PR : Precision and Recall). La courbe ROC analyse la corrélation entre le taux de vrais

positifs et le taux de faux positifs, tandis que la courbe PR évalue la relation entre le taux de détection (rappel) et la précision de détection.

Tableau 4.1 : Matrice de confusion.

Vérité terrain ou Détection	Déecté (positif)	Rejeté (négatif)
Pertinent	Vrai positif (TP)	Faux négatif (FN)
Non pertinent	Faux positif (FP)	Vrai négatif (TN)

4.2.2.1 La précision

La précision est le rapport entre les images pertinentes détectées et les exigences de l'utilisateur. Elle est également connue sous le nom de fiabilité ou de répétabilité et désigne la mesure dans laquelle des mesures répétées dans des conditions inchangées donnent les mêmes résultats. L'équation (4.1) représente les entités données.

$$\text{Précision} = \frac{\text{Nombre des vrais positifs}}{\text{Nombre de modèles détectés}} \quad (4.1)$$

Dans le contexte de la classification binaire, la précision est souvent appelée valeur prédictive positive. L'équation (4.2) la représente.

$$\text{Précision} = \frac{TP}{TP + FP} \quad (4.2)$$

4.2.2.2 Rappel

Le rappel fait référence au ratio de cas positifs correctement détectés. Il représente la proportion d'images pertinentes identifiées avec précision. Il est également connu sous le nom de taux de vrais positifs. Le calcul du rappel est déterminé à l'aide de l'équation (4.3).

$$\text{Rappel} = \frac{\text{Nombre de vrais positifs}}{\text{Nombre de modèles pertinents}} \quad (4.3)$$

Dans la classification binaire, le rappel est communément appelé sensibilité. Il est représenté par l'équation (4.4)

$$\text{Rappel} = \frac{TP}{TP + FN} \quad (4.4)$$

4.2.2.3 La retombée

La chute est la proportion d'images non pertinentes qui sont détectées comme positives, sur l'ensemble des images non pertinentes (équation 4.5).

$$\text{les retombées} = \frac{|\{non - pertinente\} \cap \{détecté\}|}{non\ pertinent} \quad (4.5)$$

Dans le contexte d'une catégorie binaire, les retombées sont directement liées à la spécificité et peuvent être calculées comme (1 - spécificité). La probabilité que des images non pertinentes soient identifiées comme positives (équation 4.6) est souvent examinée.

$$\text{les retombées} = \frac{FN}{FN + TP} \quad (4.6)$$

4.2.2.4 F-Mesure

La mesure F est également connue sous le nom de score F ou mesure F1. Elle combine la précision et la récupération. L'algorithme calcule la moyenne des valeurs de précision et de rappel. La mesure F est calculée comme la moyenne harmonique de la précision et du rappel. Ce score est utilisé pour fournir un aperçu concis de la courbe précision-rappel. Elle sera représentée comme indiqué dans l'équation (4.7).

$$\text{F-mesure} = \frac{2 \times \text{précision} \times \text{rappel}}{\text{précision} + \text{rappel}} \quad (4.7)$$

Dans la classification binaire, elle est représentée par l'équation 8.

$$\text{F-mesure} = \frac{2 \times TP}{(2 \times TP + FP + FN)} \quad (4.8)$$

4.2.2.5 Taux de reconnaissance

Le taux de reconnaissance est le rapport entre le nombre de catégories identifiées avec succès et le nombre total d'exemples, noté N. Le taux de reconnaissance est le rapport entre le nombre d'échantillons correctement classés et le nombre total d'échantillons. Le nombre de prédictions dans les approches de classification est déterminé par le nombre d'enregistrements de test qui sont correctement ou faussement prédits par le modèle. Les chiffres sont organisés et enregistrés dans une matrice de confusion, également connue sous le nom de tableau de

contingence (tableau 4.1). La matrice de confusion fournit une vue d'ensemble des performances du classificateur pour chaque catégorie particulière.

$$\text{La précision} = \frac{\text{Nombre de motifs correctement détectés}}{\text{Nombre total d'ensembles de validation}} \quad (4.10)$$

$$\text{La précision} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.11)$$

4.2.2.6 Le taux d'erreur

La fraction est le rapport entre les échantillons mal classés et le nombre total d'échantillons de validation. L'introduction du taux d'erreur est motivée par la réaction du système aux réponses incorrectes. Étant donné la présence d'ensembles de données équilibrés, cette mesure de performance est jugée appropriée pour comparer les techniques de classification. La précision, le rappel et la mesure F sont des mesures appropriées pour évaluer les performances des ensembles de données non équilibrés, comme l'indiquent les équations 12 et 13.

$$\text{Taux d'erreur} = \frac{\text{Nombre d'erreurs de classification}}{\text{Nombre total d'ensembles de validation}} \quad (4.12)$$

$$e = \frac{FP + FN}{TP + TN + FP + FN} \quad (4.13)$$

4.2.2.7 La sensibilité

Le taux de vrais positifs (TPR : True Positive Rate) est souvent appelé sensibilité, taux de réussite et rappel. Le terme "probabilité de condition" fait référence à la probabilité d'identifier correctement les membres d'une classe spécifique dans un test de classification binaire. L'équation 4.14 est une mesure mathématique appliquée qui quantifie l'efficacité de ce test à étiqueter avec précision la classe cible.

$$\text{sensibilité} = \frac{TP}{TP + FN} \quad (4.14)$$

4.2.2.8 La spécificité

La spécificité est le terme utilisé pour désigner le taux de vrais négatifs (TNR : True Negative Rate). L'équation 4.15 est une mesure de mathématiques appliquées qui quantifie la précision d'un test de classification binaire dans l'identification correcte des situations négatives.

$$\text{TNR} = \frac{TN}{TN + FP} \quad (4.15)$$

Le taux de faux positifs (FPR : False Positive Rate), souvent appelé taux d'alarme, est représenté par les équations 4.16 et 4.17 :

$$\text{FPR} = \frac{FP}{TN + FP} \quad (4.16)$$

$$\text{Spécificité} = \frac{TN}{TN + FP} + 1 - \text{Fausse alerte} \quad (4.17)$$

4.2.2.9 Caractéristiques de fonctionnement du récepteur

Les caractéristiques opérationnelles du récepteur (ROC : Receiver Operating Characteristics) sont un outil graphique utilisé pour organiser et visualiser les performances d'un système. Il s'agit d'un choix unique pour les courbes de précision-rappel. Une représentation graphique est utilisée pour afficher visuellement la transition entre le taux de vrais positifs (TPR) et le taux de faux positifs (FPR). Le TPR représente le nombre de valeurs positives correctement identifiées et est représenté sur l'axe des y, tandis que le FPR représente le nombre de valeurs négatives catégorisées par erreur et est représenté sur l'axe des x.

Les points de données situés dans la partie supérieure gauche de la courbe ROC présentent un taux élevé de vrais positifs (TP) et un faible taux de faux positifs (FP), ce qui indique la présence de classificateurs intelligents. Les graphiques ROC sont particulièrement utiles dans les situations où il existe un déséquilibre important dans la distribution des catégories et où les coûts d'une mauvaise classification ne sont pas égaux. Les graphiques ROC sont largement préférés à la précision et au taux d'erreur pour évaluer cette capacité. Le graphique ROC permet de visualiser les changements dans la caractérisation entre le taux de fausses correspondances (FMR : False Match Rate) et le taux de fausses non-correspondances (FNMR : False Non-Match Rate).

En règle générale, la technique de mise en correspondance prend une décision en comparant l'image à un modèle et en utilisant un seuil pour déterminer le niveau de similitude. L'abaissement du seuil entraîne une diminution du nombre de non-correspondances incorrectes, mais une augmentation du nombre d'acceptations incorrectes. De même, l'augmentation du seuil entraîne une diminution du taux FMR, mais une augmentation du taux

FNMR. Ce graphique à tendance linéaire met clairement en évidence les variations de performance à des niveaux plus élevés (avec moins d'erreurs).

Il existe plusieurs méthodes supplémentaires pour évaluer l'efficacité des systèmes d'identification faciale : L'article cité en référence [69] examine plusieurs paramètres liés aux systèmes biométriques, notamment le taux de reconnaissance, le taux de vérification, le taux de demi-erreur totale, le taux d'acceptation réelle (GAR : Genuine Acceptance Rate), le taux de fausse acceptation (FAR : False Acceptance Rate) et le taux de faux rejet (FRR : False Rejection Rate).

Une autre mesure d'évaluation est le taux de vérification, comme indiqué dans la référence [70]. Le processus repose sur une collection de paires d'images, où les paires d'identités identiques et les paires d'identités différentes sont comparées. En utilisant les listes de similitudes entre les types, il est possible de calculer le graphique ROC et, en fin de compte, de déterminer le taux de vérification. D'autres mesures, telles que le taux d'erreur total et comparable, dépendent de l'utilisation d'ensembles de développement et d'évaluation distincts pour leur calcul. Le test de validation est un type d'examen spécifique utilisé pour reconnaître et vérifier les traits du visage des individus. Le système de vérification utilise certaines mesures, telles que le taux d'erreur égal, tandis que les systèmes de reconnaissance utilisent généralement d'autres mesures, telles que le taux de reconnaissance.

4.2.2.10 Le taux de reconnaissance

Le taux de reconnaissance est une mesure simple et directe. Le processus dépend d'une collection d'images de la galerie, généralement une image par identité, ainsi que d'une collection d'images d'essai correspondant aux mêmes identités. Le taux de reconnaissance est calculé en divisant le nombre total d'images de sondage identifiées avec succès par le nombre total de N.

4.2.2.11 Taux de fausses correspondances

Il peut également être appelé FNMR ou taux de faux rejets (FAR). Le terme "taux de faux négatifs" désigne la probabilité que le système n'identifie pas de correspondance entre le modèle d'entrée et un modèle correspondant dans la base de données. Il quantifie le pourcentage d'entrées valides qui sont rejetées par erreur.

4.2.2.12 Taux d'erreur équivalent

Le taux d'erreur de croisement (EER : Equal Error Rate ou CER : Crossover Error Rate) est le point où les taux d'erreur d'acceptation et de rejet sont égaux. La valeur EER peut être

dérivée de la courbe ROC. L'EER, ou taux d'erreur égal, est une méthode pratique pour comparer la précision des dispositifs qui ont des courbes ROC distinctes. En règle générale, le dispositif présentant le taux d'efficacité énergétique (EER) le plus bas est le plus précis.

4.2.2.13 Taux d'échec à l'inscription

Le terme 'FTE : Failure To Enroll Rate ou FER : Failure Enroll Rate' fait référence au taux d'échec des tentatives de génération d'un modèle à partir d'une entrée. Ce problème est généralement dû à des données d'entrée de qualité inférieure.

4.2.2.14 Taux d'échec à la capture

Le 'FTC : Failure to Capture' désigne la probabilité que le système n'identifie pas une entrée même lorsqu'elle est fournie correctement.

4.2.3 Évaluation du système de reconnaissance des visages

La reconnaissance des visages dépend de la capacité du système à s'adapter à différentes poses. N'utilisez que quelques classificateurs et fonctions si l'objectif du système est de reconnaître uniquement les visages frontaux. Le nombre d'images de chaque visage est déterminé en s'entraînant sur une image et en testant sur les autres. Un minimum d'une image par posture pour chaque individu est utilisé dans l'ensemble d'apprentissage, en fonction du type de photographies et de la manière dont elles sont reconnues sous différents angles. L'application du système détermine le nombre d'images de l'ensemble d'apprentissage et le nombre d'images testées sur l'ensemble restant.

Il existe trois façons d'évaluer la précision d'une tâche de reconnaissance faciale. La meilleure option peut varier quelque peu en fonction de l'objectif final :

Dans quelle mesure l'algorithme reconnaît-il une personne à partir d'un ensemble de données comprenant de nombreuses images de la même personne ainsi qu'une variété d'images de personnes différentes ?

Dans quelle mesure l'algorithme apprend-il à partir d'un ensemble de visages provenant d'ensembles de données d'entraînement et de test comprenant des images des mêmes personnes ?

Dans quelle mesure l'algorithme peut-il distinguer plusieurs personnes à partir d'un ensemble de données comprenant des photographies de ces personnes mélangées à d'autres ?

Dans le 1^{er} cas, l'algorithme doit être entraîné à l'aide d'une collection d'images représentant le visage d'une personne spécifique. Les performances de l'algorithme doivent ensuite être testées à l'aide d'une autre collection d'images comprenant plusieurs images du même sujet, ainsi qu'un nombre égal d'images représentant d'autres personnes. Cette tâche peut être classée comme une classification binaire, et la précision peut être testée efficacement à l'aide de la précision et du rappel. Pour obtenir des résultats plus généralisés, il serait utile de reproduire ce test avec un éventail diversifié d'individus.

Pour le 2^{eme} cas, le processus d'apprentissage implique l'utilisation de nombreuses photographies de différents individus, suivie d'un test utilisant d'autres images des mêmes individus. Dans les cas où l'ensemble de données comporte un nombre restreint d'individus, la méthodologie "leave-one-out : laisser-un-out " peut être employée. L'évaluation de ce problème de classification multi classe peut se faire à l'aide de matrices de confusion, qui sont utiles pour évaluer ce type de test.

Dans le 3^{eme} cas, l'algorithme doit être entraîné à l'aide d'un ensemble d'images catégorisées comportant plusieurs individus. Il doit ensuite être testé sur un ensemble d'images comprenant diverses images des mêmes personnes mélangées à d'autres images de visages. Cet ensemble de données de test peut être utilisé pour identifier des individus dans une foule et peut contenir un grand nombre d'images de personnes différentes. Cette tâche peut être abordée comme un problème de classification binaire, où les individus sont catégorisés comme des personnes d'intérêt ou non. Elle peut également être traitée comme un problème multi classe, où chaque individu est considéré comme une classe distincte parmi les autres. Si l'ensemble de test est constitué d'images déséquilibrées, on peut utiliser différentes méthodes de précision qui prennent en compte les vrais négatifs.

4.3 Évaluation des performances du système de reconnaissance faciale

Dans le cadre de cette étude, nous utilisons trois ensembles de données largement reconnus pour évaluer les performances du système de reconnaissance faciale que nous proposons. Ces ensembles de données incluent l'ORL, Yale et AR. Le système de reconnaissance faciale repose sur une approche qui combine la transformation DCT pyramidale avec un ANN. Les sections suivantes de cette étude sont dédiées à une analyse approfondie et à une comparaison détaillée des résultats expérimentaux obtenus, en mettant l'accent sur la précision et l'efficacité de notre méthode. Ces résultats sont comparés aux méthodes récentes afin de situer notre contribution dans le contexte actuel des recherches en reconnaissance faciale.

4.3.1 Expériences menées sur ORL

L'ORL est une vaste compilation d'images faciales, composée de dix images uniques pour chacune des 40 personnes incluses dans l'ensemble de données. Dans un souci de diversité et d'imprévisibilité, les images de certaines personnes ont été prises à des moments différents, ce qui entraîne des différences dans les conditions d'éclairage, les expressions faciales (yeux ouverts ou fermés, sourire ou absence de sourire) et la présence ou l'absence de lunettes.

L'ensemble des données utilise systématiquement un arrière-plan uniforme et sombre, les sujets étant régulièrement placés en position verticale et regardant vers l'avant, tout en autorisant de modestes déplacements latéraux dans une certaine limite de tolérance. Afin de simplifier le traitement et l'analyse, toutes les photos ont été réduites à une résolution de 80x80 en utilisant un rapport de décimation de la pyramide DCT de 1/2. L'utilisation de ce format standardisé garantit l'uniformité et simplifie les calculs ultérieurs et les procédures d'extraction des caractéristiques. L'ensemble de données ORL est réputé pour sa structure diversifiée et bien organisée, ce qui en fait une ressource inestimable pour la recherche et le développement dans le domaine de la reconnaissance faciale et des domaines connexes [19], [21], [22],[26] , [61-68]

La performance du système proposé est évaluée par des tests multiples, où les ratios de décomposition pour l'entraînement et le test de l'ensemble de données sont variés. L'ensemble de données est fréquemment utilisé dans diverses études de cas et est séparé en plusieurs proportions pour l'entraînement et le test. Ces proportions comprennent 50 % pour l'entraînement et 50 % pour le test, 70 % pour l'entraînement et 30 % pour le test, et 90 % pour l'entraînement et 10 % pour le test. Chaque vecteur de l'ensemble de données est étiqueté et se compose de 325 caractéristiques extraites en appliquant la pyramide DCT aux photos des patients. Ces caractéristiques sont ensuite utilisées pour le processus de classification au sein d'un réseau ANN. L'objectif de ces expériences est d'étudier



Figure 4.1 : Exemples d'images de l'ensemble de données ORL.

l'arrangement idéal pour les tâches de reconnaissance faciale, afin de déterminer le rapport de décomposition et l'ensemble de caractéristiques les plus efficaces pour obtenir une précision et une fiabilité élevées dans la catégorisation des sujets. Le tableau 4.2 présente un résumé complet des performances sans faille atteintes par le système proposé dans l'ensemble de données ORL. Afin de démontrer les performances exceptionnelles du système proposé sur l'ensemble de données ORL, des comparaisons approfondies sont effectuées avec diverses méthodes de pointe, notamment celles documentées dans [19], [21], [22], [26] , [61-68]. Le tableau 4.3 présente une analyse complète des taux d'identification obtenus par chacune des approches susmentionnées. La méthode proposée, qui a utilisé cinq images pour l'apprentissage, affiche un taux de reconnaissance exceptionnel de 99 %. Un examen plus approfondi de ce tableau montre que l'approche proposée surpasse toutes les méthodes précédentes, même lorsqu'elle utilise un vecteur d'une dimension de 325 caractéristiques seulement. Les résultats recueillis mettent en évidence la supériorité du système proposé pour les tâches de reconnaissance faciale, en montrant sa capacité à reconnaître précisément les sujets dans l'ensemble de données donné. L'augmentation significative des taux de reconnaissance obtenue en combinant la pyramide DCT et l'ANN démontre l'efficacité de cette approche dans le développement d'une solution solide et fiable pour les applications de reconnaissance faciale. En outre, la petite taille du vecteur de caractéristiques, qui n'est que de 325 dimensions, souligne l'efficacité de cette méthode. Cela en fait une option très

prometteuse pour une mise en œuvre dans le monde réel. Par conséquent, ces découvertes apportent une contribution substantielle au progrès de la technologie de la reconnaissance faciale et à sa mise en œuvre possible dans de nombreuses situations pratiques.

Tableau 4.2 : TAUX DE RECONNAISSANCE (%) SUR L'ENSEMBLE DES DONNÉES ORL EN UTILISANT DIFFÉRENTS TAUX DE DÉCOMPOSITION.

Taux correspondants	50:50	70:30	80:20	90:10
Système proposé	99	100	100	100

Tableau 4.3 : TAUX DE RECONNAISSANCE (%) SUR L'ENSEMBLE DES DONNÉES ORL EN UTILISANT DIFFÉRENTS TAUX DE DÉCOMPOSITION.

Méthode	Échantillons d'entraînement	Taux de reconnaissance (%)
[62]	5	91
[67]	5	93.10
[64]	5	95.55
[26]	-	95.95
[22]	5	96.5
[61]	5	97.8
[21]	5	98
[68]	5	98.5
Système Proposé	5	99

La figure 4.2 représente la distribution des erreurs de prédiction du réseau de neurones où l'axe des X (Erreurs) représente les valeurs des erreurs de prédiction. L'erreur est définie comme la différence entre la valeur prédite par le réseau de neurones et la valeur réelle. Les erreurs peuvent être positives (sur-estimation) ou négatives (sous-estimation). L'axe des Y (Fréquence) : représente le nombre de cas (ou la fréquence) pour chaque intervalle d'erreur. D'après la figure 4.2 les erreurs sont situées autour de zéro avec une distribution étroite, cela indique que les prédictions du réseau de neurones est proche des valeurs réelles et la distribution étroite signifie que les erreurs de prédiction du réseau sont faibles et consistant. la figure 4.3 affiche trois sous-graphes qui montrent diverses métriques et paramètres au cours de l'entraînement. Principalement les deux composants typiques sont : 'Gradient', ce graphique montre la valeur du gradient de la fonction de perte. Le gradient est une mesure de

combien les poids du réseau sont ajustés à chaque itération. De la figure, le gradient qui diminue indique que l'algorithme de rétropropagation (backpropagation) converge vers un minimum de la fonction de perte. 'Validation Checks', ce graphique indique le nombre de fois où l'erreur de validation a cessé de diminuer et a commencé à augmenter. C'est un indicateur que l'entraînement devrait peut-être s'arrêter pour éviter le surajustement. Un critère d'arrêt courant est d'arrêter l'entraînement après un certain nombre de "validation checks" consécutifs où l'erreur de validation ne s'améliore pas. La figure 4.4 (ROC) illustre la capacité du modèle de classification à distinguer entre les classes positives et négatives où l'axe des X (False Positive Rate - Taux de Faux Positifs) : représente la proportion de faux positifs sur le nombre total de véritables négatifs. Axe des Y (True Positive Rate - Taux de Vrais Positifs) : représente la proportion de vrais positifs sur le nombre total de véritables positifs. On remarque la proximité de la courbe à l'angle supérieur gauche : une courbe qui se rapproche de l'angle supérieur gauche du graphique indique un modèle avec une haute sensibilité et une faible proportion de faux positifs, ce qui est désirable.

À noter que l'interprétation reste la même pour les sections suivantes (les figures de 4.6 à 4.11) vu que le réseau de neurones démontre les mêmes performances pour les différentes bases de données testées.

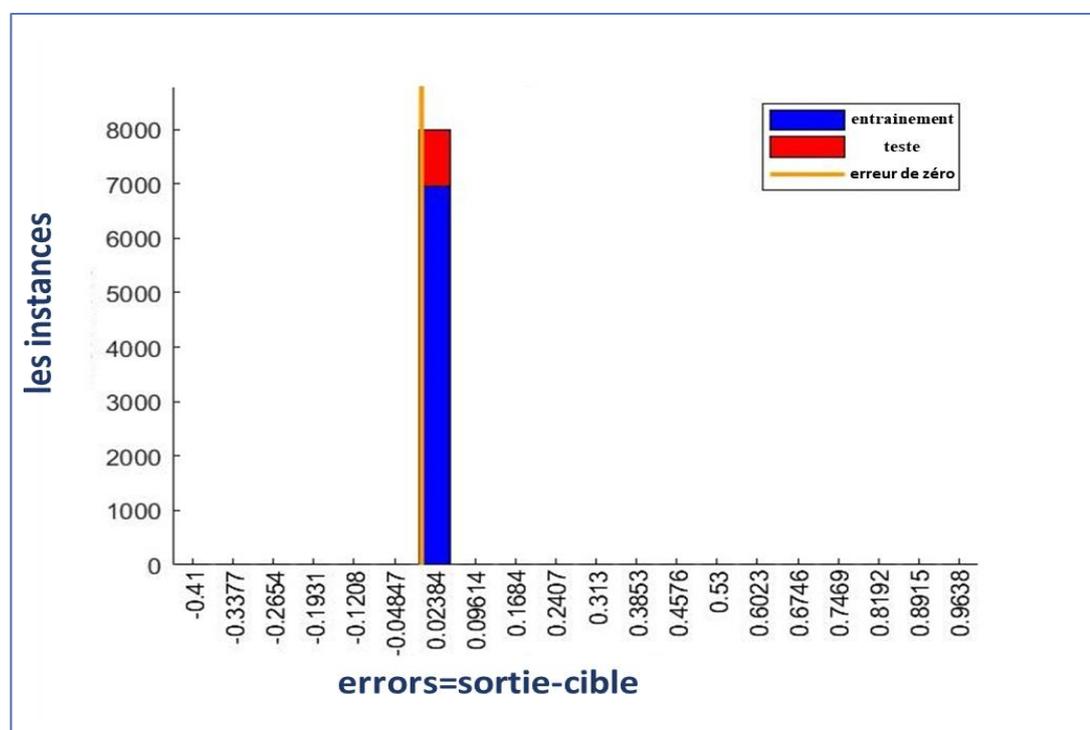


Figure 4.2 : Histogramme d'erreur du réseau neuronal pour la décomposition ORL 50%50.

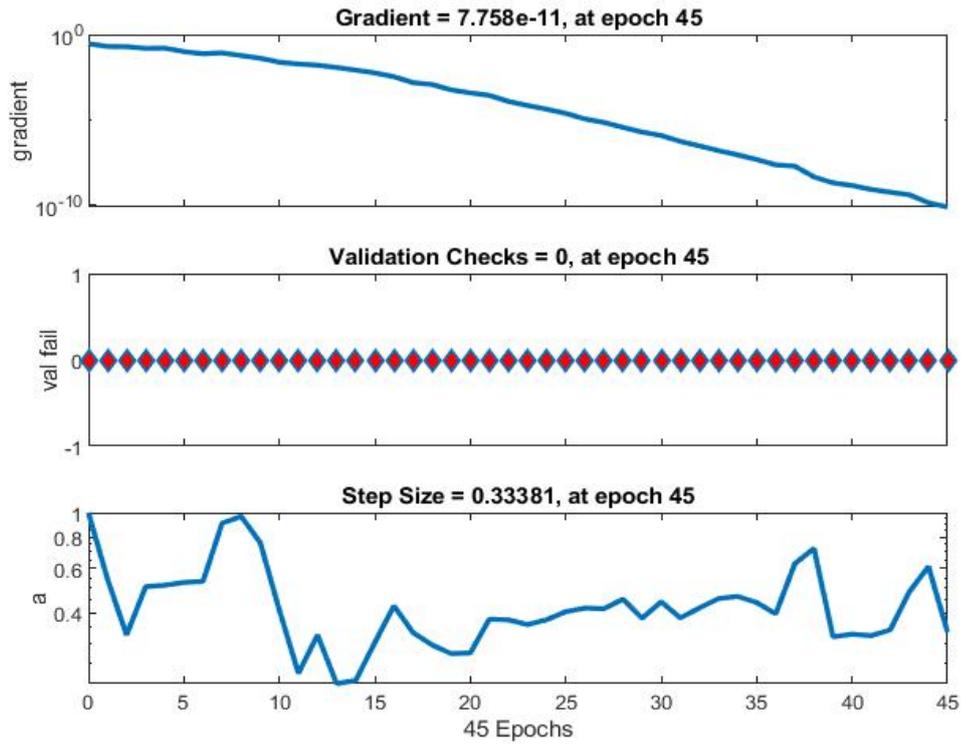


Figure 4.3 : Etat d'entraînement du réseau neuronal pour la décomposition ORL 50%50.

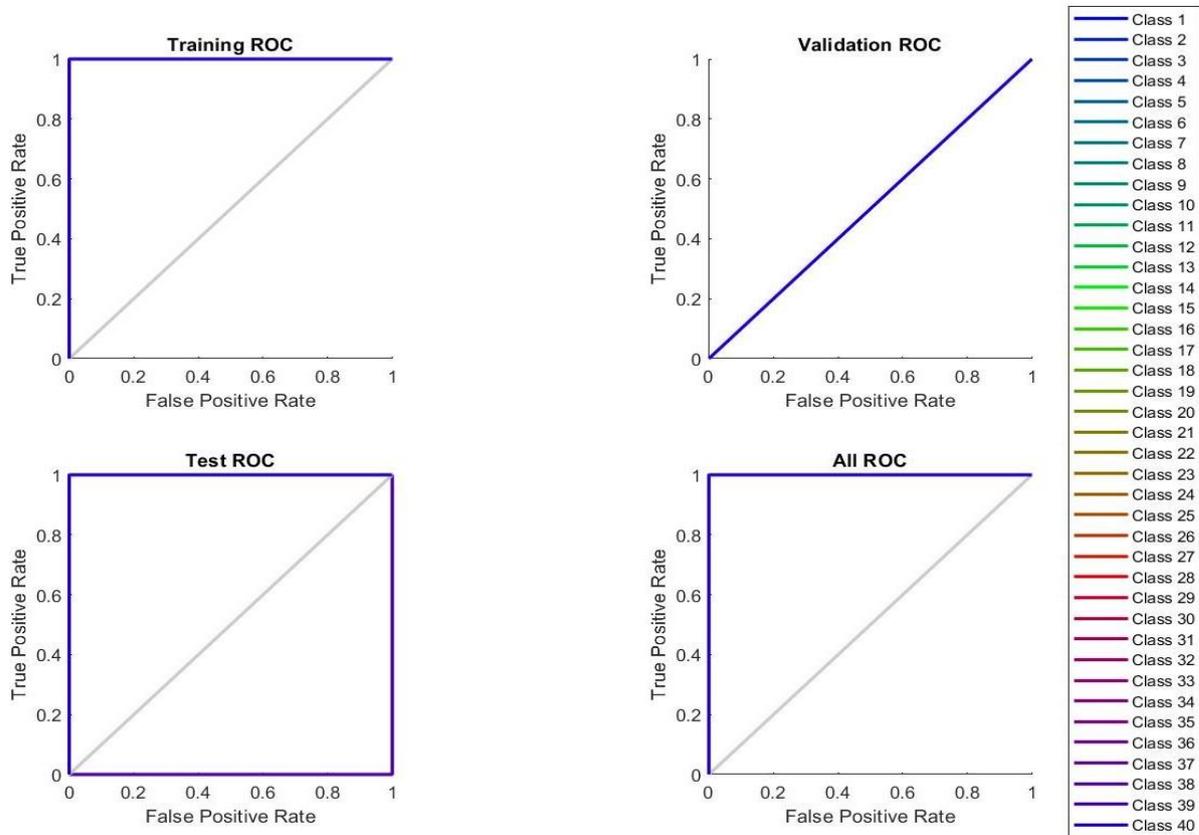


Figure 4.4 : ROC pour la décomposition ORL 50%50.

4.3.2 Expériences menées sur Yale

Dans cette partie, nous menons des expériences en utilisant l'ensemble de données sur les visages de Yale, qui se compose de 165 images en niveaux de gris représentant 15 individus uniques [14]. Il y a 11 photos pour chaque sujet, chacune représentant une émotion ou une configuration faciale distincte. Il s'agit notamment de la lumière centrale, des lunettes, du bonheur, de la lumière gauche, de l'absence de lunettes, de l'expression neutre, de la lumière droite, de la mélancolie, de la somnolence, de la surprise et du clin d'œil. La diversité des expressions et des configurations présentes dans l'ensemble de données permet d'évaluer l'efficacité et la réussite des approches dans différents scénarios. En ce qui concerne l'ensemble de données de Yale, nos études comprennent également différentes proportions d'ensembles d'apprentissage et de test, à savoir 50 % d'apprentissage et 50 % de test, 60 % d'apprentissage et 40 % de test, 70 % d'apprentissage et 30 % de test, et 90 % d'apprentissage et 10 % de test. Le tableau 4.4 présente un résumé complet des performances sans faille obtenues par le système proposé dans l'ensemble de données de Yale.



Figure 4.5 : Échantillons de l'ensemble de données de Yale.

Pour mieux évaluer le système dans cet ensemble de données, nous le comparerons à d'autres méthodes récemment publiées. Ces méthodes comprennent le descripteur LBPH avec

multi-KNN et réseau neuronal de rétropropagation [21], LBPH pour l'extraction de caractéristiques avec CNN pour la classification [63], et l'intensité relative de la magnitude du gradient (RGMS : Relative Gradient Magnitude Strength) avec des réseaux neuronaux profonds (DNN : Deep Neural Networks) [22]. Les résultats comparatifs sont présentés dans le tableau 4.5, qui démontre clairement les performances exceptionnelles du système proposé. La phase d'apprentissage a porté sur cinq photographies, tandis que six images ont été utilisées pour les tests. Pour ce faire, la dimensionnalité du vecteur de caractéristiques, obtenue en décomposant la pyramide DCT avec des caractéristiques statistiques, a été réduite [40]. Les résultats de cette étude montrent clairement que la combinaison de la pyramide DCT et de l'ANN est très efficace et présente un grand potentiel pour les tâches de reconnaissance faciale, en particulier lorsque l'on travaille avec les photos de l'ensemble de données de Yale. L'utilisation de caractéristiques statistiques dans la décomposition de la pyramide DCT améliore la fiabilité et la précision de la représentation des caractéristiques faciales, ce qui permet d'améliorer les performances de la reconnaissance faciale et de repousser les limites de la recherche actuelle dans ce domaine.

Tableau 4.4 : TAUX DE RECONNAISSANCE (%) EN UTILISANT DIFFÉRENTS TAUX DE DÉCOMPOSITION SUR L'ENSEMBLE DES DONNÉES DE Yale.

Taux correspondants	50:50	70:30	90:10
Système proposé	98.88	100	100

Tableau 4.5 : TAUX DE RECONNAISSANCE OBTENUS PAR DIFFÉRENTES MÉTHODES EN UTILISANT L'ENSEMBLE DE DONNÉES DE Yale.

Méthode	Échantillons d'entraînement	Taux de reconnaissance (%)
[67]	8	90.07
[64]	5	92
[26]	-	94.09
[68]	5	94.44
[21]	5	97.7
[22]	5	98.67
Système Proposé	5	98.88

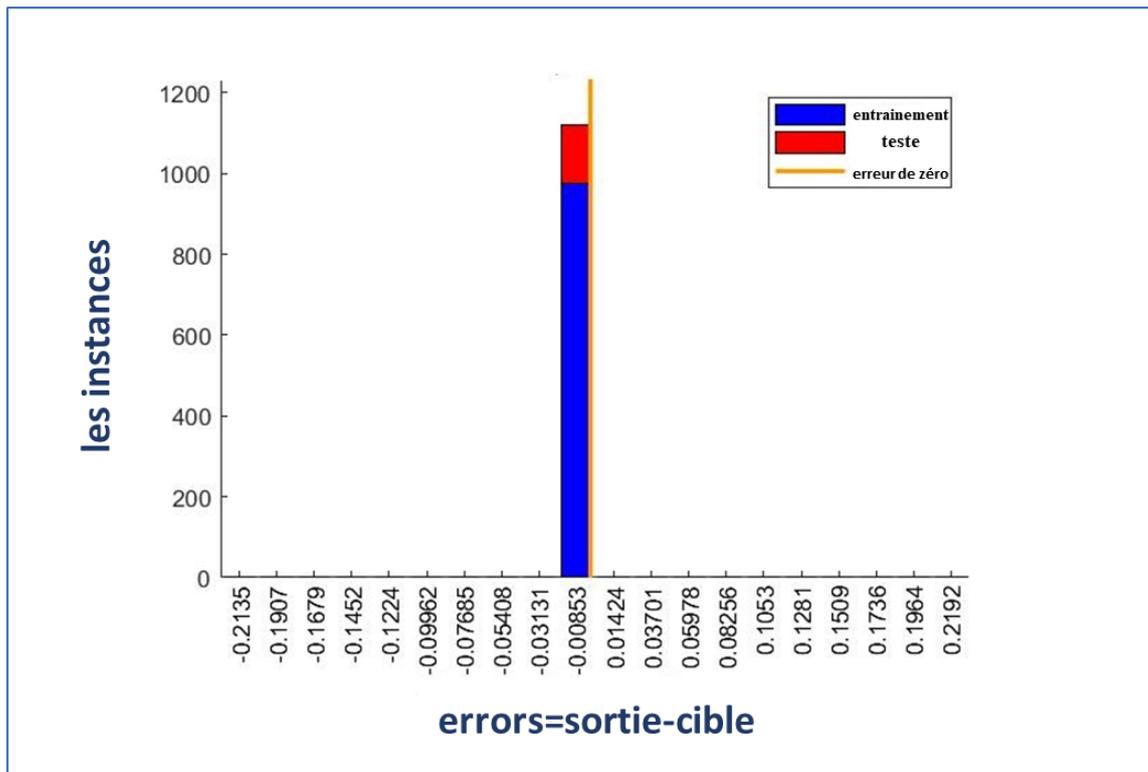


Figure 4.6 : Histogramme d'erreur du réseau neuronal pour la décomposition Yale 50%50.

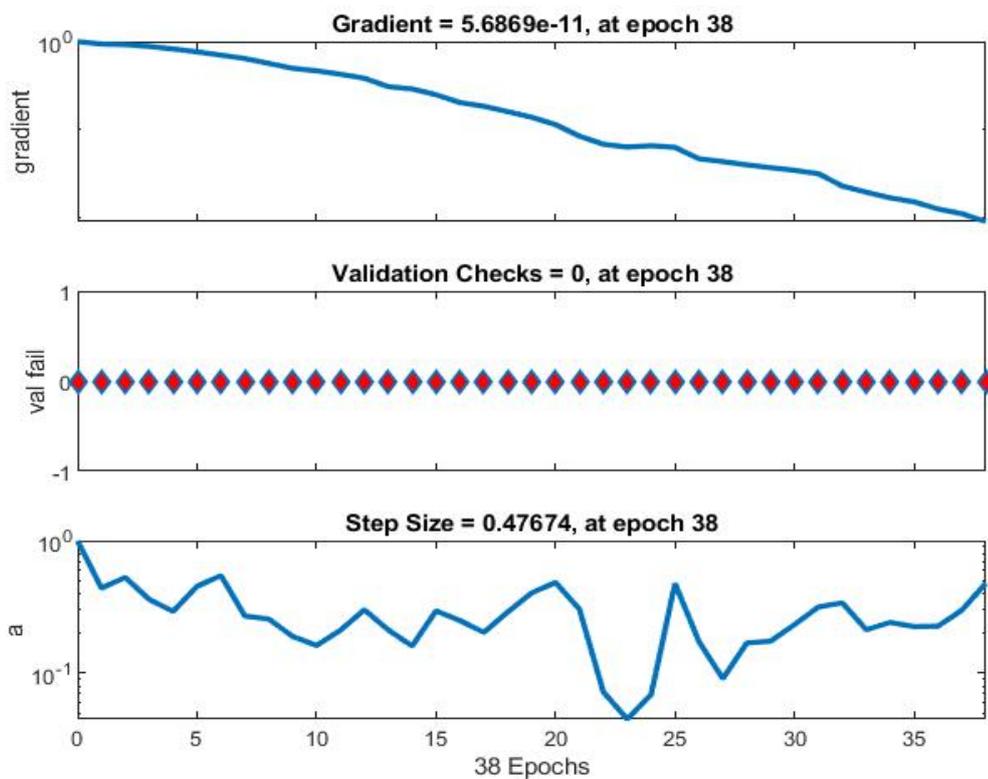


Figure 4.7 : Etat d'entraînement du réseau neuronal pour la décomposition Yale 50%50.

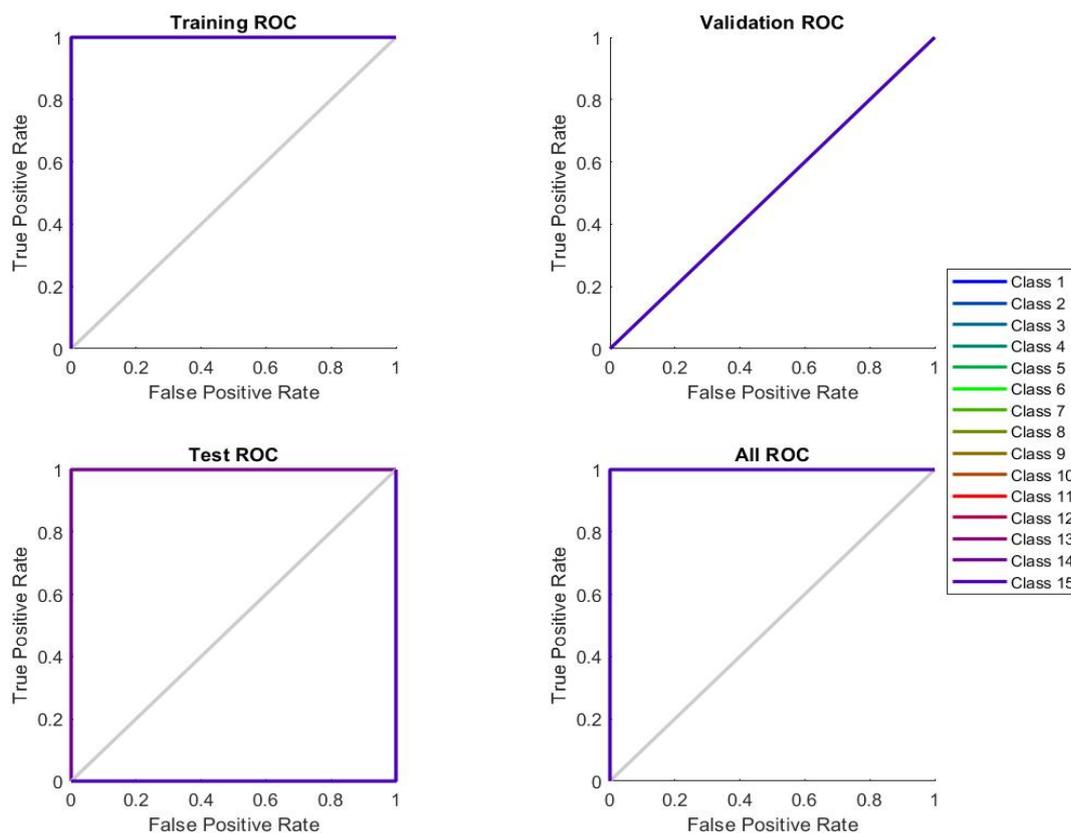


Figure 4.8 : ROC pour la décomposition Yale 50%50.

4.3.3 Expériences menées sur AR

Cette partie se concentre sur l'analyse de l'ensemble de données faciales présenté par Alex Martinez et Robert Benavente [13]. Cet ensemble de données comprend une compilation complète de près de 4000 images en couleur représentant les visages de 126 personnes, dont 70 hommes et 56 femmes. Les images fournissent un affichage complet des visages frontaux, capturant un large éventail d'expressions faciales, différentes conditions d'éclairage et des obstacles tels que des lunettes de soleil et des écharpes. Le processus de collecte des données a été entrepris avec un grand souci du détail et de la précision, dans un environnement hautement réglementé, afin de garantir l'acquisition de données fiables et de haute qualité. Les volontaires n'ont été soumis à aucune restriction concernant leur tenue vestimentaire, leurs lunettes, leur maquillage ou leur coupe de cheveux pendant la prise de vue. Pour nos expériences avec l'ensemble de données AR, nous utilisons deux proportions différentes pour les ensembles d'apprentissage et de test, à savoir 70 % pour l'apprentissage et 30 % pour le test, et 90 % pour l'apprentissage et 10 % pour le test. Cela nous permet d'effectuer une évaluation complète du système proposé en utilisant différents ratios de données d'apprentissage et de test. Les résultats de ces tests sont présentés dans le tableau 4.6 et

démontrent sans équivoque les performances supérieures et l'efficacité de notre méthode pour la reconnaissance faciale dans l'ensemble de données AR. En outre, ces résultats donnent un aperçu significatif de la résistance et de la précision de la méthode lorsqu'elle est confrontée à différentes circonstances et obstructions faciales.

Tableau 4.6 : TAUX DE RECONNAISSANCE (%) SUR AR EN UTILISANT DIFFÉRENTS TAUX DE DÉCOMPOSITION.

Taux correspondants	70:30	90:10
Système proposé	99.16	100

En outre, nous effectuons une comparaison séparée du système proposé avec [22], en utilisant le ratio d'entraînement et de test de 70%30%. Cette comparaison est présentée dans le tableau 4.7, qui montre clairement que le système proposé est nettement plus performant que la méthode proposée dans [22] sur l'ensemble de données AR. Ces résultats comparatifs soulignent la capacité du système proposé à gérer les complexités présentes dans l'ensemble de données AR, y compris les variations dans les expressions faciales, les conditions d'éclairage et les occlusions. L'habileté de la méthode à extraire les caractéristiques pertinentes des images faciales, combinée à son architecture de réseau neuronal sophistiquée, contribue de manière significative à ses performances exceptionnelles.

Tableau 4.7 : TAUX DE RECONNAISSANCE (%) OBTENUS PAR DIFFÉRENTES MÉTHODES À PARTIR DE AR.

Méthode	Échantillons d'entraînement	Taux de reconnaissance (%)
[22]	8	92.3
Système Proposé	7	99.16

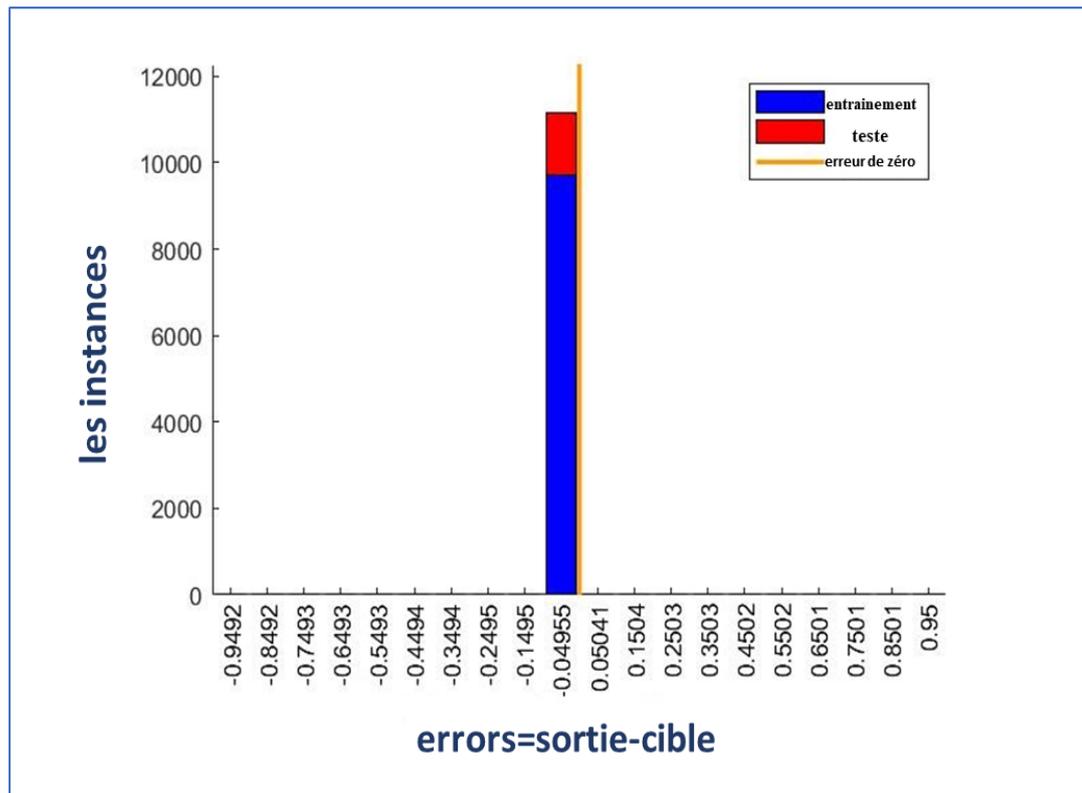


Figure 4.9 : Histogramme d'erreur du réseau neuronal pour la décomposition AR 70%30.

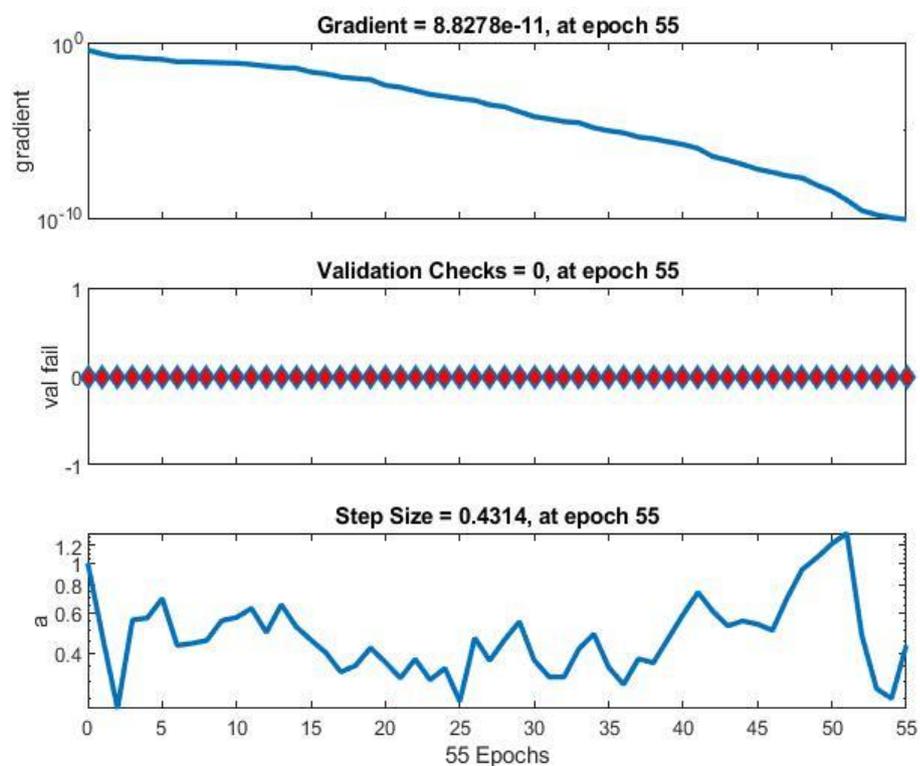


Figure 4.10 : Etat d'entraînement du réseau neuronal pour la décomposition AR 70%30.

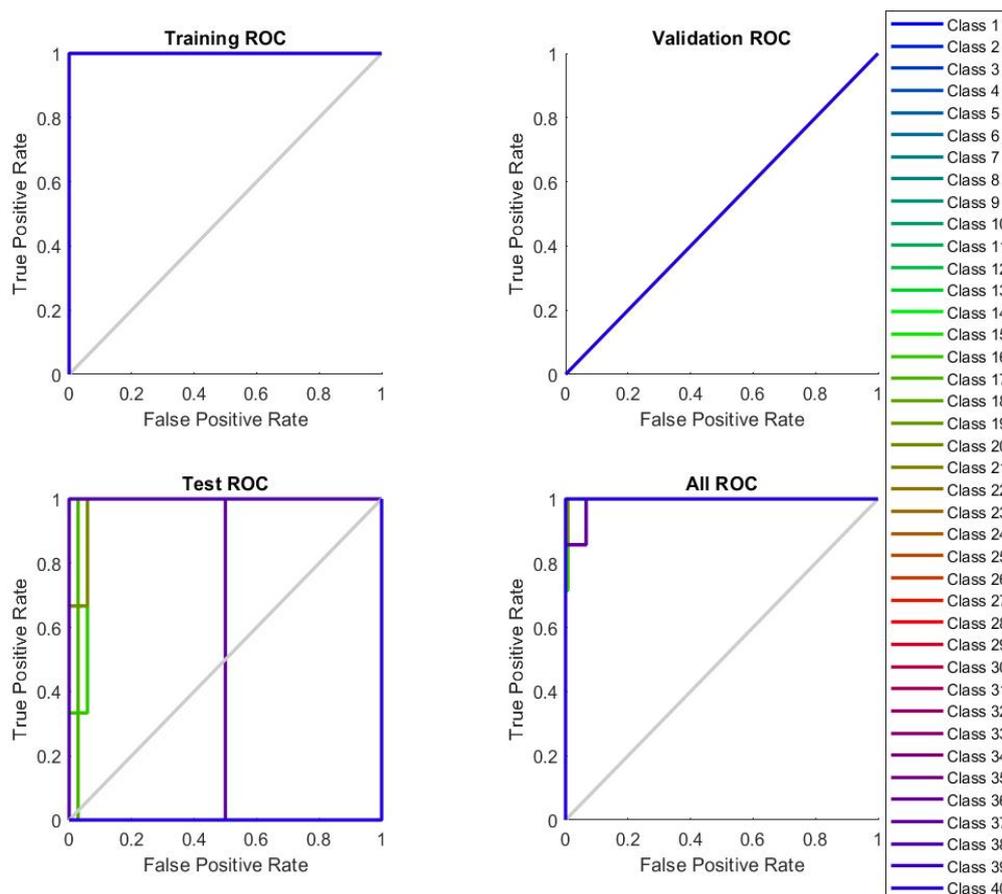


Figure 4.11 : ROC pour la décomposition AR 70%30.

4.3.4 Comparaison à l'aide du taux de reconnaissance moyen

Dans cette partie, nous effectuons une recherche comparative détaillée en utilisant le taux de reconnaissance moyen pour évaluer complètement la performance de notre approche. Notre analyse se concentre spécifiquement sur les ensembles de données ORL et Yale. L'objectif est d'évaluer l'efficacité de la technique que nous proposons par rapport à de nombreuses méthodes de reconnaissance faciale bien connues. Plus précisément, nous comparons notre approche à la méthode d'analyse IKLDA + PNN [19], à la méthode qui s'appuie sur des chaînes de valeurs successives [65], et à l'approche qui utilise un réseau neuronal basé sur une fonction de base radiale [66].

Pour réaliser cette évaluation, nous procédons à de multiples itérations d'entraînement du réseau à l'aide des ensembles de données. Nous modifions le rapport entre l'entraînement et le test pour mesurer son influence sur la précision de la reconnaissance. Les taux de reconnaissance moyens obtenus par les différentes approches sont présentés dans le tableau 4.8 pour l'ensemble de données ORL. Le tableau 4.9 fournit une comparaison entre la méthode que nous proposons et les méthodes rapportées dans [19] et [65] en utilisant

l'ensemble de données de Yale. En outre, le tableau 4.10 présente une juxtaposition de l'approche que nous proposons et de l'approche suggérée dans [23], en utilisant des échantillons d'entraînement et de test distincts.

En outre, afin d'offrir une évaluation complète, nous étendons notre étude pour incorporer l'ensemble de données AR. Le tableau 4.11 présente une juxtaposition des taux de reconnaissance moyens entre l'approche que nous proposons et celles suggérées dans les références [66] et [19]. Les résultats de ces tableaux montrent systématiquement que la stratégie que nous proposons est plus performante que les solutions existantes dans la plupart des cas.

Ces données soulignent l'efficacité de notre système à capturer des caractéristiques distinctives et à atteindre de meilleurs taux d'identification, confirmant sa promesse d'être une solution solide pour les tâches de reconnaissance des visages.

Tableau 4.8 : TAUX DE RECONNAISSANCE MOYENS (%) OBTENUS PAR DIFFÉRENTES MÉTHODES EN UTILISANT L'ENSEMBLE DE DONNÉES ORL.

Méthode	50:50	70:30	80:20
[19]	93.95 ± 0.02	96.35 ± 0.020	97.22 ± 0.02
[65]	95.65 ± 0.6	98.37 ± 0.47	98.75 ± 0.4
[66]	97.75 ± 2.31	-	-
Système proposé	97.78 ± 1.29	99.68 ± 0.72	99.78 ± 0.50

Tableau 4.9 : TAUX DE RECONNAISSANCE MOYENS (%) OBTENUS PAR LA MÉTHODE PROPOSÉE ET LES MÉTHODES PROPOSÉES DANS [65] ET [19] EN UTILISANT L'ENSEMBLE DE DONNÉES DE YALE.

Méthode	Ensemble d'entraînement	Ensemble de teste	Taux de reconnaissance moyen
[19]	7	4	82.95 ± 0.07
[65]	7	4	96.88 ± 1.55
Système proposé	7	4	99.91 ± 0.54

Tableau 4.10 : TAUX DE RECONNAISSANCE MOYENS (%) OBTENUS PAR LA MÉTHODE PROPOSÉE ET LA MÉTHODE PROPOSÉE DANS [23] EN UTILISANT L'ENSEMBLE DE DONNÉES DE YALE.

Méthode	Ensemble d'entraînement	Ensemble de teste	Taux de reconnaissance moyen
[23]	6	4	99.83 ± 0.53
Système proposé	5	6	98.01 ± 1.29
Système proposé	6	4	99.58 ± 0.86

Tableau 4.11 : TAUX DE RECONNAISSANCE MOYENS (%) OBTENUS PAR LA MÉTHODE PROPOSÉE ET LES MÉTHODES PROPOSÉES DANS [19] ET [66] EN UTILISANT L'ENSEMBLE DE DONNÉES AR.

Méthode	70:30	80:20
[66]	93.15 ± 3.25	-
[19]	98.85 ± 1.08	99.16 ± 0.8
Système proposé	99.2 ± 0.8	99.57 ± 0.645

4.4 Conclusion

En résumé, ce chapitre a donné l'occasion d'examiner en profondeur notre nouveau système de reconnaissance faciale, qui repose sur une combinaison novatrice de la pyramide DCT et des réseaux de neurones artificiels. Nous avons examiné la capacité de notre système à détecter et à reconnaître les visages dans différentes situations, y compris les variations d'éclairage, d'expressions faciales et de positions, en nous appuyant sur trois ensembles de données de référence largement reconnus dans le domaine de la recherche en reconnaissance faciale, ORL, Yale et AR.

Les résultats expérimentaux ont été examinés de manière approfondie, ce qui a permis de comparer les performances de notre système avec celles de méthodes pertinentes et récemment publiées. En accordant une attention particulière aux critères tels que la précision, nous avons réussi à mettre en évidence les atouts et les limites de notre approche dans des situations concrètes de reconnaissance faciale.

Les études effectuées ont montré que notre système offre des résultats prometteurs, en particulier en ce qui concerne la précision et la capacité à gérer des variations significatives dans les conditions d'acquisition des images. Toutefois, il reste des difficultés, notamment en ce qui concerne la capacité du système à être généralisé à des situations non contrôlées et la résistance à des variations extrêmes.

En résumé, ce chapitre a permis d'approfondir notre compréhension des capacités et des limites de notre système de reconnaissance faciale, établissant ainsi les fondations pour des améliorations futures et des avancées plus approfondies dans ce domaine en perpétuelle évolution.

Conclusion

Conclusion

La reconnaissance faciale est devenue un élément important de notre société contemporaine, intégrée dans divers domaines allant de la sécurité à la technologie grand public. Au fil des décennies, depuis les premiers travaux pionniers de Kanade jusqu'aux avancées actuelles, cette technologie a connu des évolutions notables tout en rencontrant des défis persistants.

Cette thèse a cherché à contribuer à cette évolution en proposant une approche intégrée et novatrice pour la reconnaissance faciale. À travers une combinaison de la pyramide DCT, des mesures statistiques pour la réduction de la dimensionnalité, et un réseau de neurones à deux couches, notre système vise à relever les défis de précision et de robustesse qui persistent dans ce domaine.

Le premier chapitre a posé les bases en explorant les développements récents dans le domaine, mettant en lumière les différentes approches et les tendances actuelles de la recherche. Nous avons identifié les principaux défis rencontrés par les systèmes de reconnaissance faciale et présenté les avancées récentes pour les surmonter.

Dans le deuxième chapitre, nous avons examiné en détail les concepts fondamentaux des transformées et des réseaux de neurones, soulignant leur importance dans le contexte de la reconnaissance faciale. Nous avons analysé les différentes techniques de transformées utilisées pour l'extraction des caractéristiques et les principes de fonctionnement des réseaux de neurones pour la classification.

Le troisième chapitre a présenté notre système de reconnaissance faciale proposé, décrivant comment la pyramide DCT et les réseaux de neurones sont intégrés pour améliorer la précision et la robustesse du système. Nous avons expliqué comment ces composantes sont combinées de manière synergique pour obtenir des performances optimales.

Enfin, le quatrième chapitre a évalué en profondeur notre système en utilisant des ensembles de données bien établis dans le domaine de la recherche. Nous avons analysé les performances du système en termes de précision, de rappel et d'autres métriques pertinentes, et comparé ses résultats avec ceux des méthodes existantes pour évaluer son efficacité et sa fiabilité.

En conclusion, cette thèse contribue au domaine de la reconnaissance faciale en offrant une approche intégrée pour relever les défis de précision et de robustesse. En mettant l'accent sur l'exploitation synergique des techniques de pointe, notre système propose une base pour une

nouvelle génération de solutions de reconnaissance faciale, avec des implications potentielles dans une large gamme d'applications réelles. En particulier, le développement d'algorithmes pour gérer les variations d'expressions, d'éclairage, et de postures est une avancée importante. De plus, l'adaptation de notre système pour des applications de surveillance en temps réel, de contrôles d'accès sécurisés, et de systèmes de paiement biométriques est en cours. Enfin, nous envisageons des collaborations avec des industries et des institutions pour favoriser l'intégration commerciale et publique de notre technologie.

En somme, cette recherche apporte des solutions pour répondre aux défis croissants dans le domaine de la reconnaissance faciale.

RÉFÉRENCES

- [1] « Picture Processing System by Computer Complex and Recognition of Human Faces - Robotics Institute Carnegie Mellon University ». Consulté le: 3 mai 2024. [En ligne]. Disponible sur: <https://www.ri.cmu.edu/publications/picture-processing-system-by-computer-complex-and-recognition-of-human-faces/>
- [2] D. Zeng, R. Veldhuis, et L. Spreeuwens, « A survey of face recognition techniques under occlusion », *IET Biom.*, vol. 10, n° 6, p. 581-606, nov. 2021, doi: 10.1049/bme2.12029.
- [3] A. Parashar, A. Parashar, W. Ding, M. Shabaz, et I. Rida, « Data preprocessing and feature selection techniques in gait recognition: A comparative study of machine learning and deep learning approaches », *Pattern Recognit. Lett.*, vol. 172, p. 65-73, août 2023, doi: 10.1016/j.patrec.2023.05.021.
- [4] W. Ali, W. Tian, S. U. Din, D. Iradukunda, et A. A. Khan, « Classical and modern face recognition approaches: a complete review », *Multimed. Tools Appl.*, vol. 80, n° 3, p. 4825-4880, janv. 2021, doi: 10.1007/s11042-020-09850-1.
- [5] D. N. Parmar et B. B. Mehta, « Face Recognition Methods & Applications », 2014, doi: 10.48550/ARXIV.1403.0485.
- [6] Y. Kortli, M. Jridi, A. Al Falou, et M. Atri, « Face Recognition Systems: A Survey », *Sensors*, vol. 20, n° 2, p. 342, janv. 2020, doi: 10.3390/s20020342.
- [7] Q. He, B. He, Y. Zhang, et H. Fang, « Multimedia based fast face recognition algorithm of speed up robust features », *Multimed. Tools Appl.*, vol. 78, n° 17, p. 24035-24045, sept. 2019, doi: 10.1007/s11042-019-7209-0.
- [8] M. O. Oloyede, G. P. Hancke, et H. C. Myburgh, « A review on face recognition systems: recent approaches and challenges », *Multimed. Tools Appl.*, vol. 79, n° 37-38, p. 27891-27922, oct. 2020, doi: 10.1007/s11042-020-09261-2.
- [9] Y. Guo, L. Zhang, Y. Hu, X. He, et J. Gao, « MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition », 2016, doi: 10.48550/ARXIV.1607.08221.
- [10] F. S. Samaria et A. C. Harter, « Parameterisation of a stochastic model for human face identification », in *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, USA: IEEE Comput. Soc. Press, 1994, p. 138-142. doi: 10.1109/ACV.1994.341300.
- [11] M. Afifi et A. Abdelhamed, « AFIF4: Deep Gender Classification based on AdaBoost-based Fusion of Isolated Facial Features and Foggy Faces », 2017, doi: 10.48550/ARXIV.1706.04277.
- [12] P. J. Phillips, Hyeonjoon Moon, P. Rauss, et S. A. Rizvi, « The FERET evaluation methodology for face-recognition algorithms », in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico: IEEE Comput. Soc, 1997, p. 137-143. doi: 10.1109/CVPR.1997.609311.
- [13] « AR Face Database Webpage ». Consulté le: 9 mars 2024. [En ligne]. Disponible sur: <https://www2.ece.ohio-state.edu/~aleix/ARdatabase.html>
-

- [14] A. S. Georghiades, P. N. Belhumeur, et D. J. Kriegman, « From few to many: illumination cone models for face recognition under variable lighting and pose », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, n° 6, p. 643-660, juin 2001, doi: 10.1109/34.927464.
- [15] I. Adjabi, A. Ouahabi, A. Benzaoui, et A. Taleb-Ahmed, « Past, Present, and Future of Face Recognition: A Review », *Electronics*, vol. 9, n° 8, p. 1188, juill. 2020, doi: 10.3390/electronics9081188.
- [16] W. Zhao, R. Chellappa, P. J. Phillips, et A. Rosenfeld, « Face recognition: A literature survey », *ACM Comput. Surv.*, vol. 35, n° 4, p. 399-458, déc. 2003, doi: 10.1145/954339.954342.
- [17] L. Roberts, « Machine perception of three-dimensional solids. », thesis Massachusetts Institute of Technology. [En ligne]. Disponible sur: <https://dspace.mit.edu/bitstream/handle/1721.1/11589/33959125-MIT.pdf?sequence%3D2>
- [18] M. Adel, S. H. El-Ramly, et B. Abdelhamid, « Adaptive Rate Polar Code for DCT-Based Digital Image Transmission », *IEEE Access*, vol. 11, p. 56196-56213, 2023, doi: 10.1109/ACCESS.2023.3281197.
- [19] A. Ouyang, Y. Liu, S. Pei, X. Peng, M. He, et Q. Wang, « A hybrid improved kernel LDA and PNN algorithm for efficient face recognition », *Neurocomputing*, vol. 393, p. 214-222, juin 2020, doi: 10.1016/j.neucom.2019.01.117.
- [20] L. Li, X. Mu, S. Li, et H. Peng, « A Review of Face Recognition Technology », *IEEE Access*, vol. 8, p. 139110-139120, 2020, doi: 10.1109/ACCESS.2020.3011028.
- [21] M. A. Abuzneid et A. Mahmood, « Enhanced Human Face Recognition Using LBPH Descriptor, Multi-KNN, and Back-Propagation Neural Network », *IEEE Access*, vol. 6, p. 20641-20651, 2018, doi: 10.1109/ACCESS.2018.2825310.
- [22] M. N. ElBedwehy, G. M. Behery, et R. Elbarougy, « Face Recognition Based on Relative Gradient Magnitude Strength », *Arab. J. Sci. Eng.*, vol. 45, n° 12, p. 9925-9937, déc. 2020, doi: 10.1007/s13369-020-04538-y.
- [23] S. Khan, M. H. Javed, E. Ahmed, S. A. A. Shah, et S. U. Ali, « Facial Recognition using Convolutional Neural Networks and Implementation on Smart Glasses », in *2019 International Conference on Information Science and Communication Technology (ICISCT)*, Karachi, Pakistan: IEEE, mars 2019, p. 1-6. doi: 10.1109/CISCT.2019.8777442.
- [24] I. William, D. R. Ignatius Moses Setiadi, E. H. Rachmawanto, H. A. Santoso, et C. A. Sari, « Face Recognition using FaceNet (Survey, Performance Test, and Comparison) », in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, Semarang, Indonesia: IEEE, oct. 2019, p. 1-6. doi: 10.1109/ICIC47613.2019.8985786.
- [25] T. Ahonen, A. Hadid, et M. Pietikainen, « Face Description with Local Binary Patterns: Application to Face Recognition », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, n° 12, p. 2037-2041, déc. 2006, doi: 10.1109/TPAMI.2006.244.
- [26] S. Najafi Khanbebin et V. Mehrdad, « Local improvement approach and linear discriminant analysis-based local binary pattern for face recognition », *Neural Comput. Appl.*, vol. 33, n° 13, p. 7691-7707, juill. 2021, doi: 10.1007/s00521-020-05512-3.
-

- [27] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes, et R. M. Mersereau, « Eigenface-domain super-resolution for face recognition », *IEEE Trans. Image Process.*, vol. 12, n° 5, p. 597-606, mai 2003, doi: 10.1109/TIP.2003.811513.
- [28] Jun Zhang, Yong Yan, et M. Lades, « Pace recognition: eigenface, elastic matching, and neural nets », *Proc. IEEE*, vol. 85, n° 9, p. 1423-1435, sept. 1997, doi: 10.1109/5.628712.
- [29] J. Wu, G. An, et Q. Ruan, « Independent Gabor Analysis of Discriminant Features Fusion for Face Recognition », *IEEE Signal Process. Lett.*, vol. 16, n° 2, p. 97-100, févr. 2009, doi: 10.1109/LSP.2008.2010070.
- [30] C. Shu, X. Ding, et C. Fang, « Histogram of the oriented gradient for face recognition », *Tsinghua Sci. Technol.*, vol. 16, n° 2, p. 216-224, avr. 2011, doi: 10.1016/S1007-0214(11)70032-3.
- [31] S. Naveen et R. S. Moni, « A Robust Novel Method for Face Recognition from 2D Depth Images Using DWT and DCT Fusion », *Procedia Comput. Sci.*, vol. 46, p. 1518-1528, 2015, doi: 10.1016/j.procs.2015.02.072.
- [32] B. Nassih, A. Amine, M. Ngadi, et N. Hmina, « DCT and HOG Feature Sets Combined with BPNN for Efficient Face Classification », *Procedia Comput. Sci.*, vol. 148, p. 116-125, 2019, doi: 10.1016/j.procs.2019.01.015.
- [33] V. P. Vishwakarma et S. Dalal, « Generalized DCT and DWT hybridization based robust feature extraction for face recognition », *J. Inf. Optim. Sci.*, vol. 41, n° 1, p. 61-72, janv. 2020, doi: 10.1080/02522667.2020.1721593.
- [34] V. K. H.R. et M. M., « A novel hybrid biometric software application for facial recognition considering uncontrollable environmental conditions », *Healthc. Anal.*, vol. 3, p. 100156, nov. 2023, doi: 10.1016/j.health.2023.100156.
- [35] B. M. AlFawwaz, A. AL-Shatnawi, F. Al-Saqqar, et M. Nusir, « Multi-Resolution Discrete Cosine Transform Fusion Technique Face Recognition Model », *Data*, vol. 7, n° 6, p. 80, juin 2022, doi: 10.3390/data7060080.
- [36] A. Setiawan, R. Sigit, et R. Rokhana, « Implementation of Face Recognition Using Discrete Cosine Transform on Convolutional Neural Networks », in *2022 International Electronics Symposium (IES)*, Surabaya, Indonesia: IEEE, août 2022, p. 502-505. doi: 10.1109/IES55876.2022.9888295.
- [37] F. Tabassum, Md. Imdadul Islam, R. Tasin Khan, et M. R. Amin, « Human face recognition with combination of DWT and machine learning », *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, n° 3, p. 546-556, mars 2022, doi: 10.1016/j.jksuci.2020.02.002.
- [38] Z.-H. Huang, W.-J. Li, J. Shang, J. Wang, et T. Zhang, « Non-uniform patch based face recognition via 2D-DWT », *Image Vis. Comput.*, vol. 37, p. 12-19, mai 2015, doi: 10.1016/j.imavis.2014.12.005.
- [39] Z. B. Lahaw, D. Essaidani, et H. Seddik, « Robust Face Recognition Approaches Using PCA, ICA, LDA Based on DWT, and SVM Algorithms », in *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, Athens: IEEE, juill. 2018, p. 1-5. doi: 10.1109/TSP.2018.8441452.
-

- [40] R. Atta et M. Ghanbari, « Low-memory requirement and efficient face recognition system based on DCT pyramid », *IEEE Trans. Consum. Electron.*, vol. 56, n° 3, p. 1542-1548, août 2010, doi: 10.1109/TCE.2010.5606295.
- [41] K. H. Tan et M. Ghanbari, « Compact image coding using two-dimensional DCT pyramid », *Electron. Lett.*, vol. 28, n° 8, p. 791, 1992, doi: 10.1049/el:19920499.
- [42] S.-H. Hsiao et J.-S. R. Jang, « Improving ResNet-based Feature Extractor for Face Recognition via Re-ranking and Approximate Nearest Neighbor », in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Taipei, Taiwan: IEEE, sept. 2019, p. 1-8. doi: 10.1109/AVSS.2019.8909884.
- [43] A. Abozaid, A. Haggag, H. Kasban, et M. Eltokhy, « Multimodal biometric scheme for human authentication technique based on voice and face recognition fusion », *Multimed. Tools Appl.*, vol. 78, n° 12, p. 16345-16361, juin 2019, doi: 10.1007/s11042-018-7012-3.
- [44] K. Sasirekha et K. Thangavel, « Optimization of K-nearest neighbor using particle swarm optimization for face recognition », *Neural Comput. Appl.*, vol. 31, n° 11, p. 7935-7944, nov. 2019, doi: 10.1007/s00521-018-3624-9.
- [45] S. D. Lin, L. Chen, et W. Chen, « Thermal face recognition under different conditions », *BMC Bioinformatics*, vol. 22, n° S5, p. 313, nov. 2021, doi: 10.1186/s12859-021-04228-y.
- [46] H. Ben Fredj, S. Bouguezzi, et C. Souani, « Face recognition in unconstrained environment with CNN », *Vis. Comput.*, vol. 37, n° 2, p. 217-226, févr. 2021, doi: 10.1007/s00371-020-01794-9.
- [47] G. Gao, Y. Yu, J. Yang, G.-J. Qi, et M. Yang, « Hierarchical Deep CNN Feature Set-Based Representation Learning for Robust Cross-Resolution Face Recognition », *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, n° 5, p. 2550-2560, mai 2022, doi: 10.1109/TCSVT.2020.3042178.
- [48] L. E. Bahrick et R. Lickliter, « Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. », *Dev. Psychol.*, vol. 36, n° 2, p. 190-201, 2000, doi: 10.1037/0012-1649.36.2.190.
- [49] C. Turati, E. Valenza, I. Leo, et F. Simion, « Three-month-olds' visual preference for faces and its underlying visual processing mechanisms », *J. Exp. Child Psychol.*, vol. 90, n° 3, p. 255-273, mars 2005, doi: 10.1016/j.jecp.2004.11.001.
- [50] C. A. Nelson, « The development and neural bases of face recognition », *Infant Child Dev.*, vol. 10, n° 1-2, p. 3-18, mars 2001, doi: 10.1002/icd.239.
- [51] R. Diamond et S. Carey, « Why faces are and are not special: An effect of expertise. », *J. Exp. Psychol. Gen.*, vol. 115, n° 2, p. 107-117, 1986, doi: 10.1037/0096-3445.115.2.107.
- [52] F. De La Torre, « A Least-Squares Framework for Component Analysis », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, n° 6, p. 1041-1055, juin 2012, doi: 10.1109/TPAMI.2011.184.
- [53] Army Public College of Management & Sciences, Rawalpindi, Pakistan., M. S. Khane*, A. Ware, University of South Wales, United Kingdom., A. Khan, et Army Public College of Management & Sciences, Rawalpindi, Pakistan., « Face Acknowledgment using
-

- Principle Component Analysis (PCA) of Eigenfaces », *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, n° 7, p. 1196-1200, mai 2020, doi: 10.35940/ijitee.E2861.059720.
- [54] P. N. Belhumeur, J. P. Hespanha, et D. J. Kriegman, « Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection », in *Computer Vision — ECCV '96*, vol. 1064, B. Buxton et R. Cipolla, Éd., in Lecture Notes in Computer Science, vol. 1064. , Berlin, Heidelberg: Springer Berlin Heidelberg, 1996, p. 43-58. doi: 10.1007/BFb0015522.
- [55] Kuan Hui Tan et M. Ghanbari, « Layered image coding using the DCT pyramid », *IEEE Trans. Image Process.*, vol. 4, n° 4, p. 512-516, avr. 1995, doi: 10.1109/83.370682.
- [56] R. Atta et M. Ghanbari, « An efficient layered video codec based on DCT pyramid », in *IEEE International Conference on Acoustics Speech and Signal Processing*, Orlando, FL, USA: IEEE, mai 2002, p. IV-3405-IV-3408. doi: 10.1109/ICASSP.2002.5745385.
- [57] Q. Wang, « Pyramidal coding of super high definition images », in *Fifth International Conference on Image Processing and its Applications*, Edinburgh, UK: IEE, 1995, p. 55-59. doi: 10.1049/cp:19950619.
- [58] Z. Zhang, « Artificial Neural Network », in *Multivariate Time Series Analysis in Climate and Environmental Research*, Cham: Springer International Publishing, 2018, p. 1-35. doi: 10.1007/978-3-319-67340-0_1.
- [59] P. Viola et M. Jones, « Rapid object detection using a boosted cascade of simple features », in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA: IEEE Comput. Soc, 2001, p. I-511-I-518. doi: 10.1109/CVPR.2001.990517.
- [60] C. E. Shannon, « A Mathematical Theory of Communication », *Bell Syst. Tech. J.*, vol. 27, n° 3, p. 379-423, juill. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.
- [61] M. Alajmi, K. Awedat, A. Essa, F. Alassery, et O. S. Faragallah, « Efficient Face Recognition Using Regularized Adaptive Non-Local Sparse Coding », *IEEE Access*, vol. 7, p. 10653-10662, 2019, doi: 10.1109/ACCESS.2019.2890845.
- [62] S. Liu, L. Li, M. Jin, S. Hou, et Y. Peng, « Optimized Coefficient Vector and Sparse Representation-Based Classification Method for Face Recognition », *IEEE Access*, vol. 8, p. 8668-8674, 2020, doi: 10.1109/ACCESS.2019.2960928.
- [63] V. B. T. Shoba et I. S. Sam, « Face Recognition Using LBPH Descriptor and Convolution Neural Network », in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India: IEEE, juin 2018, p. 1439-1444. doi: 10.1109/ICCONS.2018.8663009.
- [64] A. Singha, M. K. Bhowmik, et D. Bhattacharjee, « Akin-based Orthogonal Space (AOS): a subspace learning method for face recognition », *Multimed. Tools Appl.*, vol. 79, n° 47-48, p. 35069-35091, déc. 2020, doi: 10.1007/s11042-020-08892-9.
- [65] H. Zaaraoui, S. El Kaddouhi, A. Saaidi, et M. Abarkan, « Face recognition with a new local descriptor based on strings of successive values », *Multimed. Tools Appl.*, vol. 80, n° 18, p. 27017-27044, juill. 2021, doi: 10.1007/s11042-020-09400-9.
- [66] V. Agarwal et S. Bhanot, « Radial basis function neural network-based face recognition using firefly algorithm », *Neural Comput. Appl.*, vol. 30, n° 8, p. 2643-2660, oct. 2018, doi: 10.1007/s00521-017-2874-2.
-

- [67] T. S. Akheel, V. U. Shree, et S. A. Mastani, « Stochastic gradient descent linear collaborative discriminant regression classification based face recognition », *Evol. Intell.*, vol. 15, n° 3, p. 1729-1743, sept. 2022, doi: 10.1007/s12065-021-00585-y.
- [68] D. Ali, I. Touqir, A. M. Siddiqui, J. Malik, et M. Imran, « Face Recognition System Based on Four State Hidden Markov Model », *IEEE Access*, vol. 10, p. 74436-74448, 2022, doi: 10.1109/ACCESS.2022.3188717.
- [69] D. White, J. D. Dunn, A. C. Schmid, et R. I. Kemp, « Error Rates in Users of Automatic Face Recognition Software », *PLOS ONE*, vol. 10, n° 10, p. e0139827, oct. 2015, doi: 10.1371/journal.pone.0139827.
- [70] P. Ahlgren et L. Grönqvist, « Retrieval evaluation with incomplete relevance data: a comparative study of three measures », in *Proceedings of the 15th ACM international conference on Information and knowledge management - CIKM '06*, Arlington, Virginia, USA: ACM Press, 2006, p. 872. doi: 10.1145/1183614.1183773.
-