



RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ DJILLALI LIABES SIDI BEL ABBES
FACULTÉ GÉNIE ÉLECTRIQUE
DÉPARTEMENT D'ÉLECTRONIQUE



THÈSE

En vue de l'obtention du
Diplôme de **Doctorat en Sciences**

SPÉCIALITÉ : ÉLECTRONIQUE

OPTION : TRAITEMENT DE SIGNAL ET DE L'IMAGE

PRÉSENTÉE PAR

BENKADDOUR MOHAMMED KAMEL

INTITULÉ :

**Reconnaissance des formes et classification automatique :
Application à l'identification biométrique**

Le jury composé de :

Nom & Prénom(s)	Grade	Qualité	Etablissement de rattachement
TALEB Nasreddine	Professeur	Président	UDL-SBA
BOUNOUA Abdennacer	Professeur	Directeur de thèse	UDL-SBA
BENTOUTOU Youcef	Directeur de Recherche	Examineur	Centre de Développement des Satellites (CDS) - Oran
HADJ ABDERRAHMANE Lahcene	Directeur de Recherche	Examineur	Centre de Développement des Satellites (CDS) - Oran

Laboratoire 'Réseaux de Communication Architecture et Multimédia (RCAM)'

Année Universitaire 2019-2020

REMERCIEMENTS

Mes louanges et gratitudes intarissable vont en premier à Dieu le Tout-Puissant qui m'a prodigué la santé, et le courage afin d'accomplir ce travail. La réalisation de cette thèse fut une occasion merveilleuse de rencontrer et d'échanger avec de nombreuses personnes. Je ne saurais pas les citer toutes sans dépasser le nombre de pages raisonnablement admis dans ce genre de travail. Je reconnais que chacune à des degrés divers, mais avec une égale bienveillance, pour apporter une contribution positive à sa finalisation. Mes dettes de reconnaissance sont, à ce point de vue, énormes à leur égard.

Mes plus vifs remerciements vont à mon encadreur Monsieur BOUNOUA Abennacer, qui avec son noble esprit scientifique, sa modestie, et sa patience m'a soutenu tout au long de la réalisation de ce travail.

Je tiens également à remercier tous les membres du jury qui m'ont fait l'honneur d'accepter de lire et de juger mon travail de thèse.

Je ne peux pas omettre d'exprimer ma gratitude à ma Mère en particulier et envers toutes ma famille, pour leur soutien indéfectible, leurs encouragements et leurs sacrifices, que ce travail leur soit dédié avec toute mon affection.

À tous mes amis, à toute personne que je n'ai pas su citer ici et qui m'ont apporté de la joie, du support, de l'aide, quand j'en avais besoin, merci.

Une pensée pour terminer ces remerciements pour toi qui n'a pas vu l'aboutissement de mon travail mais je sais que tu en aurais été très fier de ton fils, **PAPA**.

RÉSUMÉ

Récemment, la reconnaissance faciale a suscité l'intérêt de la communauté scientifique. Cette technique de biométrie efficace, non intrusive et sans contact occupe une place de plus en plus importante dans le domaine de la recherche. Les récents progrès dans le domaine des réseaux de neurones artificiels plus connu actuellement sous le nom d'apprentissage profond, ont permis d'améliorer l'état de l'art dans plusieurs domaines de reconnaissance de forme en offrant une possibilité de s'attaquer à des problématiques qui étaient difficilement traitables par les méthodes d'apprentissage automatique conventionnelles. Les travaux effectués dans le cadre de cette thèse portent sur l'identification automatique de visages sur des bases de données d'images numériques. Dans un premier temps, notre approche a consisté à explorer des techniques d'apprentissage automatique et les possibles gains par l'intégration de modèle pré-entraîné de réseaux neurones. Nous décrivons aussi des sous-problèmes spécifiques tels que l'extraction des caractéristiques, la classification et la stratégie d'apprentissage des réseaux de neurones profond pour la reconnaissance de visage.

L'objectif de cette thèse est de développer un système de reconnaissance de visage qui combine les réseaux de neurones convolutifs (CNN) avec des approches de classification. Dans ce travail nous proposons une méthode de reconnaissance faciale et de classification basée sur un apprentissage profond, en particulier le réseau neuronal convolutif pour l'extraction des caractéristiques faciale, qui s'avère être un outil puissant qui a connu un grand succès dans la classification des images et la reconnaissance des formes. Afin réduire la dimension de ces caractéristiques, une technique d'analyse en composantes principales est utilisée, suivi par machine à vecteurs de support (SVM) pour une classification plus élevée. Nous avons effectué une évaluation approfondie de nos méthodes sur l'ensemble de données de la technologie de reconnaissance faciale (FERET). Les résultats obtenus ont montré que les réseaux CNNs associés aux solutions d'analyse en composantes principales et de SVM améliorent considérablement les performances ainsi que la précision de la reconnaissance.

Mots-Clés : biométrie, reconnaissance de visage, extraction de caractéristique, réseau de neurone convolutif, machines à vecteurs de support, FERET, analyse en composantes principale.

RESUME

Recently, face recognition has attracted the interest of the scientific community. This efficient, non-intrusive and non-contact biometrics technique is becoming increasingly important in scientific research. Recent advances in the field of artificial neural networks, now known as deep learning, have improved the state of the art in several areas of pattern recognition by providing an opportunity to tackle problems that were difficult to deal with by conventional machine learning methods. In this thesis, we propose face recognition system on digital image databases, with automatic learning techniques and we describe specific sub problems such as the extraction of faces features, classification and the learning strategy of deep neural networks. In order to address this problem, we are focusing on a technique called Convolutional Neural Network (CNN).

The aim of this thesis is to evaluate the problem of face identification using deep convolutional neural networks. In this work we propose a method of facial recognition and classification based on deep learning, in particular the convolutional neural network for the extraction of facial features which proves to be a powerful tool and has been very successful in the field of image classification and pattern recognition. In order to reduce the size of these features, a principal component analysis technique is used, followed by a support vector machine (SVM) for a higher classification. We conduct a thorough evaluation of our methods on the facial recognition technology (FERET) dataset. The results obtained showed that the CNNs networks associated with the principal component analysis and SVM solutions, considerably improve the performance as well as the accuracy of the recognition.

Keywords: biometrics, face recognition, feature extraction, convolutional neural network, CNN, support vector machines (SVM), FERET, principal component analysis, PCA.

ملخص

في الآونة الأخيرة، جذب التعرف البيو مترى على الوجه اهتمام المجتمع العلمي. تزداد أهمية تقنية القياسات الحيوية الفعالة وغير التداخلية وغير المتصلة في مجال البحث. أدت التطورات الحديثة في مجال الشبكات العصبية الاصطناعية، والمعروفة الآن باسم التعلم العميق، إلى تحسين الحالة الحديثة في العديد من مجالات التعرف على الأنماط من خلال توفير فرصة لمعالجة المشاكل التي كان من الصعب التعامل معها عن طريق أساليب التعلم الآلي التقليدية. في هذه الأطروحة، نقترح نظام التعرف على الوجوه مع تقنيات التعلم التلقائي ونصف المشاكل الفرعية المحددة مثل استخراج ملامح الوجه واستراتيجية التعلم للشبكات العصبية العميقة. من أجل معالجة هذه المشكلة، نركز على تقنية تسمى الشبكة العصبية التلافيفية (CNN).

الهدف من هذه الأطروحة هو تطوير نظام التعرف على الوجوه باستخدام الشبكات العصبية التلافيفية العميقة. في هذا العمل، نقترح طريقة للتعرف على الوجه وتصنيفه على أساس التعلم العميق، وخاصة الشبكة العصبية التلافيفية لاستخراج ملامح الوجه التي اتبنت أنها أداة قوية والتي كانت ناجحة للغاية في تصنيف الصورة والتعرف على الأنماط. تليها آلة ناقلات الدعم (SVM) للتصنيف العالي. لتقليل حجم هذه الميزات، يتم استخدام تقنية تحليل المكون الرئيسي. تجري تقيماً كاملاً لأساليبنا على مجموعة بيانات تقنية التعرف على الوجوه (FERET). أظهرت النتائج التي تم الحصول عليها أن شبكات CNN المرتبطة بتحليل المكونات الرئيسية وحلول SVM، تعمل على تحسين الأداء وكذلك دقة التعرف.

الكلمات المفتاحية: القياسات الحيوية ، التعرف على الوجه ، استخلاص المعالم ، الشبكة العصبية التلافيفية ، FERET، آلات ناقلات الدعم، SVM ، تحليل المكون الرئيسي.

Publications et Communications

Les publications et les communications relatives à cette thèse sont les suivantes :

Publications

1. **M.K. Benkaddour, A. Bounoua.** *Feature extraction and classification using deep convolutional neural networks, PCA and SVC for face recognition, Traitement du Signal, Vol. 34, No. 1-2, pp. 77-91, 2017. <https://doi.org/10.3166/TS.34.77-91>.*
2. **M.K. BENKADDOUR .** *Applying DWT and PCA with Artificial Neural Network for FACE RECOGNITION ,Revue ElWahat pour les Recherches et les Etudes Vol.7n°2: 06 – 12,2014.*

Communications

1. **M.K. Benkaddour A. Benyettou, R.Imine .** *Reconnaissance de visage par un technique hybride Neuro-Acp , Conférence internationale ISPA 12 Image and Signal Processing Aplication Mostaganem ,Algerie Decembre 2012.*
2. **M.K. Benkaddour and A.Bounoua .** *Face Recognition Using Artificial Neural Network And Eignefaces , CGE 08 école polytechnique militaire bordj el bahri Alger ,Algerie Avril 2013.*
3. **M.K. Benkaddour and A.Bounoua .** *Improving Face Recognition using DWT, PCA and Neural Network , Conférence Internationale sur l'Intelligence Artificielle et les Technologies de l'Information (ICA2IT'14) Ouargla, Algerie, 10-12 Mars 2014 .*
4. **M.K. BENKADDOUR , A. BOUNOUA .** *DWT and PCA Based Feature Extraction with Neural Network Classification for Face Recognition ,IWMCS'2014 Deuxième Workshop International sur les Mathématiques et Informatique Tiaret, Algérie 1-2 Décembre ,2014.*
5. **Benouis Mohamed , benkaddour Mohamed kamel and Senouci Mohamed .** *A Novel technique for Human face recognition using fractal code and bi-dimensional subspace, The 5th International Conference on Computer Science and its Applications (CIIA'2015) in Saida, University Of Saida ,Algeria , 20-21 may 2015.*
6. **M. K. BENKADDOUR , A.BOUNOUA ,M. Benouis .** *Face Recognition Based WPD & LBP &Bi-dimensional Subspace Feature Extraction and neural network classification , The TASSILI SCCIBOV'2015 Workshop universite djillali liabes sidi bel abbes , Algeria 02-03 decembre 2015.*

Table des matières

INTRODUCTION GENERALE	01
-----------------------------	----

Chapitre 1 **Biométrie & systèmes de reconnaissance de visage**

1.1	Introduction	06
1.2	La Biométrie.....	06
1.2.1	Définition.....	06
1.2.2	Exigences de la biométrie.....	06
1.2.3	Les modes opératoires d'un système biométrique.....	07
1.2.4	Structure d'un Système biométrique	08
1.3	Les différentes modalités biométriques	09
1.3.1	Les technologies biométriques	11
1.3.2	Evaluation des performances des systèmes biométriques.....	13
1.3.3	Applications des systèmes biométriques.....	15
1.3.4	Mesure de la performance d'un système biométrique.....	16
1.4	La reconnaissance de visages	17
1.4.1	Motivation : (pourquoi la reconnaissance de visages ?).....	17
1.5	Les classes des techniques de reconnaissance de visages.....	18
1.5.1	Méthodes locales	18
1.5.2	Méthodes globales	19
1.5.3	Méthodes hybrides	20
1.6	Systèmes de reconnaissance de visage.....	21
1.6.1	Acquisition de l'image.....	22
1.6.2	Prés traitements.....	23
1.6.3	Détection puis localisation.....	23
1.6.4	Extraction des paramètres et classification.....	23
1.6.5	Apprentissage.....	23
1.6.6	Décision.....	24
1.7	Principales difficultés de la reconnaissance de visage	24
1.7.1	Changement d'illumination.....	24
1.7.2	Variation de pose.....	25
1.7.3	Expressions faciales.....	25
1.7.4	Présence ou absence des composants structurels.....	25
1.7.5	Occultations partielles.....	26
1.7.6	Vieillesse et le changement d'aspect.....	26
1.8	Conclusion.....	26

Chapitre 2 **Etat de l'art des techniques de reconnaissance de visages**

2.1	Introduction	27
2.2	État de l'art des techniques de reconnaissance	27
2.2.1	Approches globales (ou holistiques).....	28
A.	Analyse en Composantes Principales.....	28
B.	Analyse Discriminante Linéaire.....	32
C.	Analyse en Composantes Indépendantes.....	33
D.	Les réseaux de neurones artificiels	34
2.2.2	Les approches basées sur les caractéristiques locales.....	35
A.	Modèle de Markov caché	35

B.	Machine à vecteurs de support.....	35
C.	Approches basées sur les graphes	38
D.	Local Binary Pattern (LBP).....	39
2.2.3	Les approches hybrides	40
2.3	Systèmes biométriques existants pour la reconnaissance des visages	41
2.4	Conclusion.....	42

Chapitre 3 Apprentissage automatique et réseaux de neurones convolutifs

3.1	Introduction.....	43
3.2	Apprentissage automatique	43
3.2.1	En quoi consiste l'apprentissage automatique ?.....	43
3.2.2	Définition	44
3.2.3	Intelligence artificielle	44
3.2.4	Pourquoi l'apprentissage automatique ?.....	45
3.2.5	Les données d'apprentissage	46
3.2.6	L'histoire des systèmes d'apprentissage automatique.....	46
3.3	Les types d'apprentissage	47
3.3.1	Apprentissage supervisé	48
3.3.2	Apprentissage non supervisé	48
3.3.3	Apprentissage semi-supervisé	49
3.3.4	Apprentissage par renforcement.....	49
3.3.5	Apprentissage par transfert.....	49
3.4	Les algorithmes d'apprentissage.....	49
3.5	Apprentissage Profond (Deep Learning)	50
3.6	Réseaux de neurones artificiels : les bases.....	51
3.6.1	Neurone formel et perceptron.....	51
3.6.2	Les fonctions d'activation.....	53
3.6.3	Le perceptron multicouche.....	54
3.6.4	L'algorithme de la rétro-propagation du gradient (back-propagation)..	55
3.6.5	Les différents types de réseaux de neurones.....	56
3.6.6	Classification et reconnaissance des visages par réseaux de neurones..	58
3.7	Les réseaux de neurones profonds.....	60
3.8	Réseaux de neurones convolutifs.....	63
3.8.1	Différents modules d'un réseau de neurones convolutif.....	65
3.8.2	Outils d'optimisation des réseaux convolutifs	70
3.9	Utilisation de modèles pré-entraînés pour l'apprentissage par transfert	72
3.10	Les architectures neuronales convolutifs	74
3.11	Conclusion.....	75

Chapitre 4 Expérimentations, Résultats et discussions

4.1	Introduction.....	76
4.2	Les données biométriques	76
4.2.1	ORL.....	77
4.2.2	FERET	78
4.2.3	CASIA-WebFace DATABASE	78
4.2.4	Séparation des bases de données et matériel utilisé	79
4.3	Les différentes étapes de notre approche de reconnaissance de visage..	80
4.3.1	Extraction des caractéristiques faciales.....	82

4.3.1.1	Convolution et pooling	82
4.3.1.2	La structure du réseau de neurones convolutifs CNN proposée	83
4.3.1.3	Le pré-entraînement du réseau CNN	85
4.3.2	Réduction des dimensions par analyse en composantes principales....	86
4.3.3	Classification par support vecteur machine SVM.....	89
4.4	Expérimentations	91
4.4.1	Données utilisées.....	91
4.4.2	Méthodologies des tests, résultats et discussions.....	92
4.5	Conclusion.....	99
CONCLUSION GENERALE		100

Table des Figures

1.1	- Les différents modules d'un système biométrique (exemple d'empreinte digitale)	08
1.2	- La Classification De La Biométrie.....	10
1.3	- Différentes modalités biométriques	10
1.4	-Le visage occupe la deuxième place en biométrie.....	14
1.5	- Illustration du FRR et du FAR	16
1.6	- Distances entre les points caractéristiques du visage	19
1.7	- L'utilisation de la région entière du visage comme entrée : méthodes globales.....	20
1.8	-les principaux algorithmes principaux utilisés en reconnaissance faciale.....	21
1.9	- Fonctionnement d'un système de reconnaissance de visages.....	22
1.10	-Exemple de variation d'éclairage.....	24
1.11	-Exemples de variation de poses.....	25
1.12	-Exemples de variation d'expressions.....	25
2.1	-Système de reconnaissance (apprentissage) faciale utilisant PCA et LDA	33
2.2	-Le modèle de synthèse d'image de l'architecture d'ICA.....	34
2.3	- Support vecteur machine (SVM) : calcul de l'hyperplan optimal.....	37
2.4	-La représentation des visages par graphe rectangulaire.....	39
2.5	- Représentation d'un visage par les histogrammes du code LBP	40
3.1	- Les différentes techniques d'apprentissage automatique.....	47
3.2	- Le procédé d'apprentissage classique comparé à celui du Deep Learning.....	51
3.3	-Comparaison entre un neurone biologique et le modèle mathématique d'un Neurone artificiel (formel).....	52
3.4	- Représentation d'un réseau de neurone perceptron.....	52
3.5	- Fonctions d'activation couramment utilisées	53
3.6	-Architecture d'un perceptron multicouche. (Constitué de deux couches cachées. Chaque cercle représente un neurone formel).....	54
3.7	-Le principe général de l'algorithme de la retro-propagation de gradient.....	56

3.8	- Réseau de Neurons Artificiels discriminant pour la reconnaissance de visages.....	59
3.9	- Architecture d'un auto-encodeur (AE).....	62
3.10	- À gauche : Topologie d'une machine de Boltzmann classique. À droite : Topologie d'une machine de Boltzmann restreinte (RBM).....	62
3.11	- Architecture d'un réseau de neurones convolutifs LeNet-5.....	64
3.12	- Les différentes couches d'un réseau de neurones convolutif standard.....	65
3.13	- Illustration de la convolution.....	66
3.14	- Illustration du Max Pooling	67
3.15	- les différents types de Pooling avec un filtre 2x2 et un pas de	68
3.16	- Exemple de fonctionnement d'une couche Relu.....	69
3.17	- Architecture standard d'une réseaux convolutifs.....	70
4.1	- Extrait de la base ORL. Pour chacune des 40 personnes enregistrées, on dispose de 10 vues avec des changements de pose, d'expression et d'éclairage.....	77
4.2	- Extraits d'images de visage de la base de données FERET.....	78
4.3	- Extraits d'images faciale de la base de données de CASIA-WebFace.....	78
4.4	-Exemples de séparation de données d'apprentissage et de test couramment utilisés.....	79
4.5	- L'architecture générale de notre approche de reconnaissance faciale....	80
4.6	- L'ensemble du système d'apprentissage.....	81
4.7	- Phase de reconnaissance (décision).....	81
4.8	- Le modèle de réseaux CNN utilisée dans ce travail.....	83
4.9	-Extraction des caractéristiques faciale par pré-apprentissage d'un réseau CNN	86
4.10	- Schéma de base d'un système basé sur l'ACP.....	87
4.11	- Le nombre optimal de composantes principales	88
4.12	- Prétraitement et redimensionnement des image de la base FERET.....	91
4.13	- Taux d'identification avec les modèles pré-entraînées.....	93
4.14	- Erreur de validation moyenne pour chaque modèle du réseaux CNN....	95

Liste des tableaux

1.1	- <i>Comparaison entre les technologies biométriques.....</i>	14
1.2	- <i>Avantages et inconvénients de la Reconnaissance de Visage.....</i>	18
4.1	- <i>Architecture détaillée de notre CNN utilisée pour l'extraction des caractéristiques.....</i>	84
4.2	- <i>Paramètres d'apprentissage du CNN.....</i>	92
4.3	- <i>Performances de classification des modèles testé sur la base de données FERET</i>	97
4.4	- <i>Configuration expérimentale et performance pour la base de données ORL.....</i>	97
4.5	- <i>Taux de reconnaissance des différents approches utilisés dans ce travail (FERET).....</i>	98
4.6	- <i>Comparaison des performances de certaines méthodes citées dans l'états de l'arts</i>	99

Liste des acronymes

- **2D** : Représentation à 2 Dimensions
- **3D** : Représentation à 3 Dimensions
- **ACI** : Analyse en Composantes Indépendantes
- **ACP** : Analyse en Composantes Principales
- **ANN** : Artificial Neural Network
- **AI** : Artificial Intelligence
- **BDD** : Base De Donnée
- **CNN** : Convolutional Neural Network
- **DBN** : Deep Belief Network
- **DL** : Deep Learning
- **GPU** : Graphical Processor Unit
- **FAR** : False Acceptance Rate
- **FERET** : Face REcognition Technologie
- **FRR** : False Rejection Rate
- **LBP** : Local Binary Pattern
- **LDA** : Linear Discriminant Analysis
- **ML** : Machine Learning
- **MLP** : Multi-Layer Perceptron
- **NIST** : National Institute of Standards and Technology
- **LSTM** : Long Short-Term Memory
- **ORL** : Olivetti Research Laboratory
- **PCA** : Principal Component Analysis
- **PMC** : Perceptron Multi-Couche
- **RBF** : Radial Basis Fonction
- **ReLU** : Rectified linear unit
- **RNA** : Réseau de Neurones Artificiels
- **RNN** : Recurrent Neural Network
- **ROI** : Région d'intérêt (Region Of Interest)
- **SVC** : Support Vectors clustering
- **SVM** : Support Vectors Machines

Introduction Générale

Contexte de la thèse

Les travaux réalisés au cours de cette thèse s'inscrivent dans le contexte d'identification biométrique. Nous nous intéressons ici à une seule modalité : le visage. Même si nous sommes focalisés sur cet objet d'intérêt, les algorithmes présentés sont facilement transférables à d'autres modalités biométriques pour des applications de surveillance vidéo, de sécurité ou de contrôle d'accès. La reconnaissance de visages humains est considérée actuellement comme l'une des tâches les plus importantes pour relever le défi en reconnaissances des formes. La facilité et la précision avec lesquelles nous pouvons identifier nos amis et nos ennemis, même dans des conditions défavorables, sont des capacités les plus étonnantes du système visuel humain, le but de la reconnaissance de visages est de concevoir des systèmes informatiques capables d'égaliser les êtres humains dans ce domaine.

S'investir dans le domaine de la reconnaissance de visages est sans doute motivé par la multiplicité et la variété des champs d'applications de celui-ci (télésurveillance et vérification d'identités, chirurgie plastique, langage des signes...). L'intérêt pour ce domaine s'explique aussi par le fait que par rapport aux autres filières de la reconnaissance des formes (reconnaissance de l'écriture, de la voix, etc.), Le visage offre un potentiel plus important que la plupart des autres biométries. En effet, le visage est facilement accessible (ce qui n'est pas le cas de l'iris), bien acceptée par les utilisateurs, sans contact (ce qui n'est pas le cas des empreintes digitales) et offre un potentiel énorme dans les années à venir si elle pouvait fonctionner avec les systèmes de vidéosurveillance existants dans les stades, les aéroports, les rues, etc. Parmi les différentes biométries, le visage est la caractéristique la plus utilisée par l'homme pour identifier un membre de son entourage. Afin de simuler ce comportement biologique, les chercheurs se sont mis à concevoir des systèmes de plus en plus performants. Dès lors, la reconnaissance des visages a connu un fort développement et elle reste un domaine qui suscite toujours des interrogations et un engouement par les chercheurs.

Une large variété de systèmes intègre des techniques de reconnaissance de personnes fiables pour vérifier ou trouver l'identité des individus. Le but de ces techniques est de s'assurer que les services rendus sont uniquement exploités par leurs propres utilisateurs (utilisateurs légitimes). Cependant, ces systèmes sont vulnérables aux ruses d'imposteurs s'ils ne sont pas dotés de techniques de sécurité et de reconnaissance robustes.

La biométrie est un domaine émergent, où la technologie améliore notre capacité à identifier une personne. La protection des gens contre la fraude ou le vol est l'un des buts de la reconnaissance par la biométrie. L'avantage de la reconnaissance biométrique est que chaque individu a ses propres caractéristiques physiques qui ne peuvent être changées, perdues ou volées. Plusieurs raisons peuvent motiver l'usage de la reconnaissance par la biométrie :

- Une haute sécurité : en l'associant à d'autres technologies comme le cryptage.
- Confort : en remplaçant juste le mot de passe, exemple pour l'accès à un système d'exploitation.
- Sécurité/Psychologie : dans certains cas, particulièrement pour le commerce électronique, l'usager n'a pas confiance. Il est important pour les acteurs de ce marché de convaincre le consommateur de faire des transactions.

Les systèmes d'authentification biométriques suppriment donc :

- La duplication.
- Le vol.
- L'oubli.
- La perte.

En effet, l'analyse biométrique automatique du visage humain est un domaine de recherche très demandé aujourd'hui. Le développement inédit des caméras digitales, des calculateurs et d'internet ont augmenté de manière significative le nombre de photos qui sont constamment créés et partagés entre les gens partout dans le monde. Récemment, une famille de méthodes d'apprentissage automatique qui est basée sur les réseaux de neurones artificiels profonds s'est avérée être particulièrement adaptée aux larges quantités de données d'apprentissage et peut se généraliser remarquablement mieux que les approches alternatives. Cette famille est connue aujourd'hui sous le nom "d'apprentissage profond" ("deep learning" en anglais). Ainsi, la présente thèse est dédiée à l'exploration et à la conception des méthodes d'apprentissage profond dans le cadre de la reconnaissance de visage sur des bases de données faciale.

Problématique et Objectifs

Les travaux de cette thèse s'intéressent aux problèmes de reconnaissance de forme plus particulièrement appliquées à la reconnaissance biométrique du visage faiblement résolu.

L'objectif de ce travail de thèse est de développer l'un des systèmes biométriques les plus répandus dans le domaine de reconnaissance de formes, et cela grâce à sa simplicité et son efficacité : il s'agit du système de reconnaissance de visage. En effet, le visage est certainement la caractéristique biométrique que l'on utilise le plus naturellement pour s'identifier. Les méthodes classiques pour la reconnaissance faciale suivent un procédé typique consistant en deux étapes : l'extraction des caractéristiques d'image de visage en question, et ensuite, la classification à partir de ces descripteurs pour obtenir le résultat final. Dans ce schéma, l'étape d'extraction des caractéristiques a une importance primordiale car elle définit quelles informations sur le visage en question sont données à l'entrée de la technique de classification. Contrairement aux méthodes classiques antérieures, nous fusionnons les étapes d'extraction des caractéristiques et de classification dans un seul modèle en employant les réseaux de neurones convolutifs CNNs (Convolutionel Neural Network). Plus particulièrement, notre objectif est de trouver des méthodes optimales pour la conception, l'apprentissage de CNNs et la reconnaissance faciale.

Un autre objectif majeur de ce travail est d'étudier l'utilité du préapprentissage dans la formation de réseaux de neurones convolutifs pour des tâches de reconnaissance. Pour automatiser cette tâche, plusieurs méthodes et approches ont été adoptées. Nous proposons une nouvelle approche de reconnaissance de visage basée sur une architecture spécifique de CNN suivie d'une étape de prétraitement par l'Analyse en Composantes Principales (ACP) et classification par les supports vecteurs machines (SVMs). Nous appliquons ces techniques à plusieurs bases de données publiques de visage.

Contributions et Organisation de la Thèse

Les principales Contributions de cette thèse peuvent se résumer aux points suivants :

-D'abord, nous effectuons des études préliminaires qui illustrent les caractéristiques les plus importantes des CNNs. En particulier, nous démontrons que l'efficacité des caractéristiques appris avec un CNN dépend non seulement de la taille de l'architecture employée, la quantité de données d'apprentissages, mais aussi de la complexité du problème cible.

- La construction d'un modèle mathématique permettant d'adapter l'architecture (nombre de neurones par couches, taille des noyaux, connexions entre couches, ...) d'un réseau de neurones convolutifs à un problème donné (dimension du problème, nombre d'échantillons, ...) est un sujet nécessitant encore de nombreux travaux. En effet, les architectures utilisées sont construites bien souvent empiriquement, ce qui est coûteux en temps. L'utilisation de méthodes de préapprentissage dans ce travail permet de résoudre, en partie, ce problème. Nous aussi proposons une stratégie d'apprentissage de CNNs pour la reconnaissance de visage qui est validée par un préapprentissage du réseau sur des données très volumineuse. Elle nous permet de concevoir les CNNs qui surpassent l'état de l'art actuel sur les problèmes en question.

-Ensuite, nous présentons une méthode permettant d'appliquer un prétraitement par l'analyse en composante principale sur les images de visage et la proposition d'une méthode de classification par SVM. Afin de démontrer l'utilité pratique de notre approche, nous l'appliquons à plusieurs bases publiques de visage

Le manuscrit se décompose en trois parties, organisées sur quatre chapitres :

– la première partie introduit la biométrie et dresse un état de l'art des techniques utilisées pour la reconnaissance de visages. Nous insistons notamment sur l'apprentissage profond et les réseaux de neurones convolutifs.

– la seconde partie explore les techniques de représentation et reconnaissance de visage utilisées tout au long de la thèse.

– la troisième partie détaille les résultats expérimentaux obtenus à l'aide des méthodes mises en œuvre.

- Le premier chapitre est consacré à la présentation générale de la biométrie. Il décrit tout d'abord le principe de fonctionnement des systèmes biométriques, puis définit les outils utilisés pour évaluer leurs performances. Nous abordant aussi la place de la reconnaissance faciale parmi les autres techniques biométriques. A travers ce chapitre, nous voulons positionner le problème de la reconnaissance faciale et présenter ses enjeux et intérêts par rapport aux autres techniques. Nous étudions également les principales composantes d'un système de reconnaissance faciale, notamment la détection, l'extraction de caractéristiques et la phase de reconnaissance. Enfin, nous mettons en lumière les difficultés rencontrées par les systèmes de reconnaissance de visage.

- Dans le second chapitre une analyse détaillée des différentes techniques développées au cours de ces dernières années dans le domaines de la reconnaissance du visage est présentée. Notre objectif est de classifier ces méthodes de manière efficace afin de mettre en évidence les particularités ainsi que les avantages et les inconvénients de chacune d'entre elles.
- Le troisième chapitre est dédié à notre approche d'extraction des caractéristiques du visage et de classification basée sur les réseaux de neurones convolutifs. Dans ce chapitre nous indiquerons aussi les ajustements pratiques nécessaires à l'implantation de la technique d'identification hybride dite CNN-ACP-SVM.
- Dans le quatrième chapitre, Nous présentons les résultats expérimentaux obtenus par chaque méthode en analysons leurs performances, suivi d'une discussion avec interprétation des résultats.

Enfin, la conclusion générale résumera les résultats obtenus par les différentes approches et donnera quelques perspectives sur les travaux futurs.

CHAPITRE : 1

Biométrie et système de reconnaissance de visage

-
- 1.1 Introduction
 - 1.2 La Biométrie
 - 1.2.1 Définition
 - 1.2.2 Exigences de la biométrie
 - 1.2.3 Les modes opératoires d'un système biométrique
 - 1.2.4 Structure d'un Système biométrique
 - 1.3 Les différentes modalités biométriques
 - 1.3.1 Les technologies biométriques
 - 1.3.2 Evaluation des performances des systèmes biométriques
 - 1.3.3 Applications des systèmes biométriques
 - 1.3.4 Mesure de la performance d'un système biométrique
 - 1.4 La reconnaissance de visages
 - 1.4.1 Motivation : (pourquoi la reconnaissance de visages ?)
 - 1.5 Les classes des techniques de reconnaissance de visages
 - 1.5.1 Méthodes locales
 - 1.5.2 Méthodes globales
 - 1.5.3 Méthodes hybrides
 - 1.6 Systèmes de reconnaissance de visage
 - 1.6.1 Acquisition de l'image
 - 1.6.2 Prés-traitements
 - 1.6.3 Détection puis localisation
 - 1.6.4 Extraction des paramètres et classification
 - 1.6.5 Apprentissage
 - 1.6.6 Décision
 - 1.7 Principales difficultés de la reconnaissance de visage
 - 1.7.1 Changement d'illumination
 - 1.7.2 Variation de pose
 - 1.7.3 Expressions faciales
 - 1.7.4 Présence ou absence des composants structurels
 - 1.7.5 Occultations partielles
 - 1.7.6 Vieillesse et le changement d'aspect
 - 1.8 Conclusion
-

1.1 Introduction

La biométrie s'invite progressivement dans notre vie quotidienne, elle fait partie des grands enjeux pour un monde plus sûr. Dans son environnement quotidien, un individu a besoin de s'identifier dans une multitude de contextes : pour accéder à son lieu de travail ou retirer de l'argent à un distributeur, l'accès à des locaux sécurisés, le vote électronique, la sécurité des paiements bancaires ou des transactions via Internet, autant de codes et de mots de passe à mémoriser et à protéger. La biométrie est une alternative aux mots de passe et autres identifiants pour supprimer le doute sur l'identité. Elle permet de vérifier que l'utilisateur est bien la personne qu'il prétend être.

Un système biométrique est essentiellement un système de reconnaissance de formes qui reconnaît une personne en déterminant l'authenticité d'un trait physiologique et/ou comportemental spécifique possédé par cette personne. Il fonctionne comme suit : acquiert des données biométriques d'un individu, il extrait un ensemble de caractéristiques à partir des données acquises, et compare cet ensemble avec l'ensemble de modèles (*template*) stockés dans la base de données.

1.2 La Biométrie

1.2.1 Définition

Le terme biométrie vient des mots Grecs bios (vie) et metrikos (mesure ou distance) qui veut dire distances biophysiques de l'être humain, ou en d'autre terme, c'est l'anthropométrie. Il existe plusieurs définitions de la biométrie, en voici quelques-unes :

- Définition 1 : La biométrie est la science qui étudie à l'aide des mathématiques (statistiques, probabilités) les variations biologiques ou biophysiques à l'intérieur d'un groupe déterminé de personnes [Robert,1987].
- Définition 2 : La biométrie permet l'identification d'une personne sur la base de caractères physiologiques ou de traits comportementaux automatiquement reconnaissables et vérifiables [Larousse,2002].

1.2.2 Exigences de la biométrie

Dans la théorie tout trait de l'être humain (physiologique ou comportemental) peut être employé comme caractéristique biométrique pour la reconnaissance des personnes aussi longtemps qu'il répond aux conditions suivantes [1] :

- *Universalité* : ceci signifie que chaque personne devrait avoir le trait.

- *Unicité* : ceci indique que deux personnes quelconques devraient être suffisamment différentes en termes de leurs traits biométriques.
- *Permanence* : ceci signifie que le trait devrait être suffisamment invariable dans le temps.
- *Collecte* : ceci indique que le trait peut être mesuré quantitativement.

Dans la pratique, il y a d'autres problèmes importants qui doivent être pris en considération, incluant :

- *Performance* : qui se rapporte à l'exactitude et la vitesse d'identification, les ressources requises pour obtenir l'exactitude d'identification et la vitesse désirées, aussi bien que les facteurs opérationnels et environnementaux qui les affectent.
- *Acceptabilité* : ce qui indique le point auquel les gens sont disposés à accepter l'utilisation d'un trait biométrique particulier dans leurs vies quotidiennes.
- *Vulnérabilité (Mise en échec)* : ce qui reflète à quel point il est facile de tromper le système par des méthodes frauduleuses.

1.2.3 Les modes opératoires d'un système biométrique

On trouve deux grands axes dans la biométrie : l'identification et la vérification. Avec l'identification ou la reconnaissance, le système biométrique pose et essaye de répondre à la question, " Qui suis-je ? ". Dans une application d'identification, le dispositif biométrique requiert une information biométrique et la compare avec chaque information stockée dans la base de données, c'est une comparaison un à plusieurs (1:N). Le but des applications d'identification est d'identifier des criminels et des terroristes en utilisant les données des surveillances [1].

Dans la vérification ou l'authentification, le système biométrique demande à l'utilisateur son identité et essaye de répondre à la question, " Suis-je bien la personne que je prétends être ? ".

Dans une application de vérification l'utilisateur annonce son identité par l'intermédiaire d'un mot de passe, d'un numéro d'identification, d'un nom d'utilisateur, ou toute combinaison des trois. Le système sollicite également une information biométrique provenant de l'utilisateur, et compare la donnée caractéristique obtenue à partir de l'information entrée, avec la donnée enregistrée correspondante à l'identité prétendue, c'est une comparaison un à un (1:1). Le système trouvera ou ne trouvera pas d'appariement entre les deux [1] [2].

La vérification est communément employée dans des applications de contrôle d'accès et de paiement par authentification. La biométrie offre beaucoup plus d'avantages que les méthodes existantes d'authentification personnelle telles que les clefs, les numéros d'identification (ID), les mots de passe et les cartes magnétiques. En effet elle fournit encore plus de sûreté et de convenance ce qui engendre d'énormes avantages économiques et elle comble les grandes failles de sécurité des mots de passe [2].

1.2.4 Structure d'un système biométrique

Un système biométrique est conçu à l'aide des quatre modules principaux suivants [3] [4], (voir figure 1.1) :

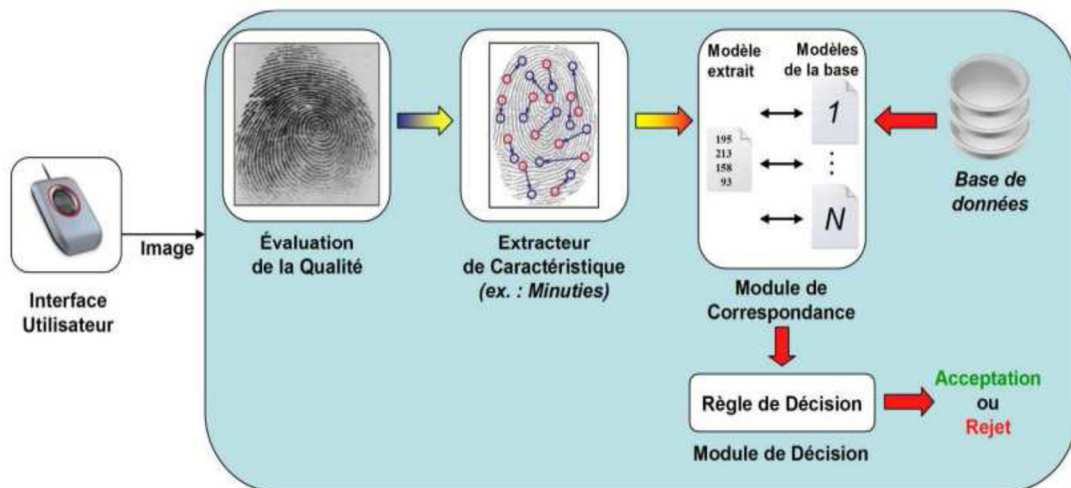


Fig.1.1- Les différents modules d'un système biométrique (exemple d'empreinte digitale).

- 1) **Module de capture / interface utilisateur** : Qui acquiert la donnée biométrique d'un utilisateur (peut être un appareil photo, un lecteur d'empreintes digitales, une caméra de sécurité, etc.).
- 2) **Module d'extraction des caractéristiques** : Qui prend en entrée les données biométriques acquises par le module de capture et extrait seulement l'information pertinente afin de former une nouvelle représentation des données. Cette nouvelle représentation doit être pertinente, idéalement unique pour chaque personne et invariante aux modifications qui peuvent intervenir sur la capture.
- 3) **Module de comparaison / de correspondance** : Dans lequel les caractéristiques extraites par le module précédent sont comparées contre les modèles stockés pour produire des scores de ressemblance. Ce module détermine le niveau de similitude (ou de divergence) entre deux empreintes biométriques.

- 4) **Module de décision** : Qui vérifie l'identité affirmée par un utilisateur, il détermine si le degré de similitude retourné par le module de similarité est suffisant pour déterminer l'identité d'un individu.

Le module d'inscription ou d'enregistrement est responsable de l'inscription des individus dans la base de données du système biométrique. Pendant la phase d'inscription, la caractéristique biométrique d'un individu est d'abord scannée par un lecteur biométrique pour produire une représentation numérique de la caractéristique. La capture de données pendant le processus d'inscription peut ou peut ne pas être dirigée par un humain selon l'application. Un contrôle de qualité est généralement exécuté pour garantir que l'échantillon acquis peut être sûrement traité par les étapes successives [4].

Afin de faciliter la comparaison, la représentation numérique d'entrée est de plus traitée par un extracteur de caractéristique pour produire une représentation compacte mais expressive, appelée un modèle ou pattern. Selon l'application, le modèle peut être stocké dans la base de données centrale du système biométrique ou enregistré sur une carte à puce livrée à l'utilisateur. Généralement, les modèles dans la base de données peuvent être mis à jour dans le temps.

1.3 Les différentes modalités biométriques

La biométrie est la reconnaissance automatique d'une personne en utilisant des traits distinctifs. Une autre définition de la biométrie est toutes caractéristiques physiques ou traits personnels automatiquement mesurables, robustes et distinctives qui peuvent être utilisées pour identifier un individu ou pour vérifier l'identité prétendue d'un individu.

Une gamme de systèmes biométriques monomodal est à l'étude ou sur le marché. Les différences en développant ces systèmes impliquent le coût, la fiabilité, le manque de confort en utilisant l'appareil, et la quantité de données requise. Les empreintes digitales, par exemple, font très peu d'erreurs de classification, mais la quantité de données qui doit être stockée pour décrire une empreinte digitale (le modèle) avait tendance à être assez grande. En revanche, le matériel utilisé pour capturer la voix est moins cher (microphone à bon marché), mais elle change quand les émotions et les états de santé changent. La biométrie comprend des caractéristiques physiologiques et comportementales [5]. La figure 1.2 illustre cela pour un nombre de technologie biométriques les plus fréquemment utilisées.

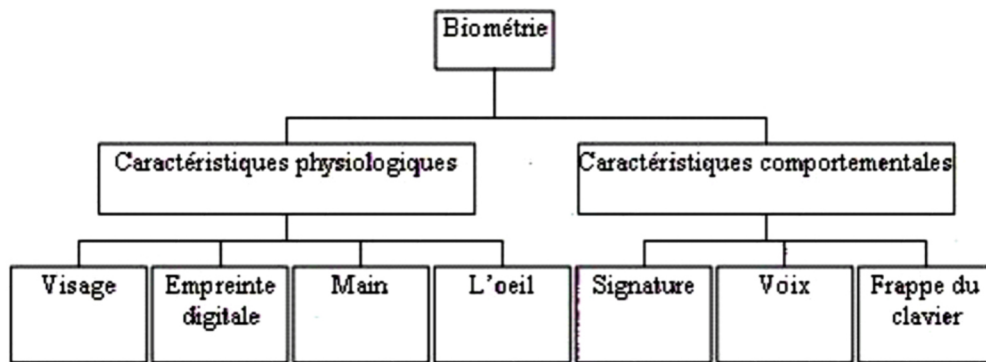


Fig.1.2- La classification de la biométrie.

Un trait physiologique est relativement stable tel que l’empreinte digitale, la géométrie de la main, l’iris, la rétine. En effet, toutes ces caractéristiques sont fondamentalement invariables sans trauma (blessure) de l’individu. Un trait comportemental d’autre part, a une certaine base physiologique, mais reflète également l’état d’une personne [6].

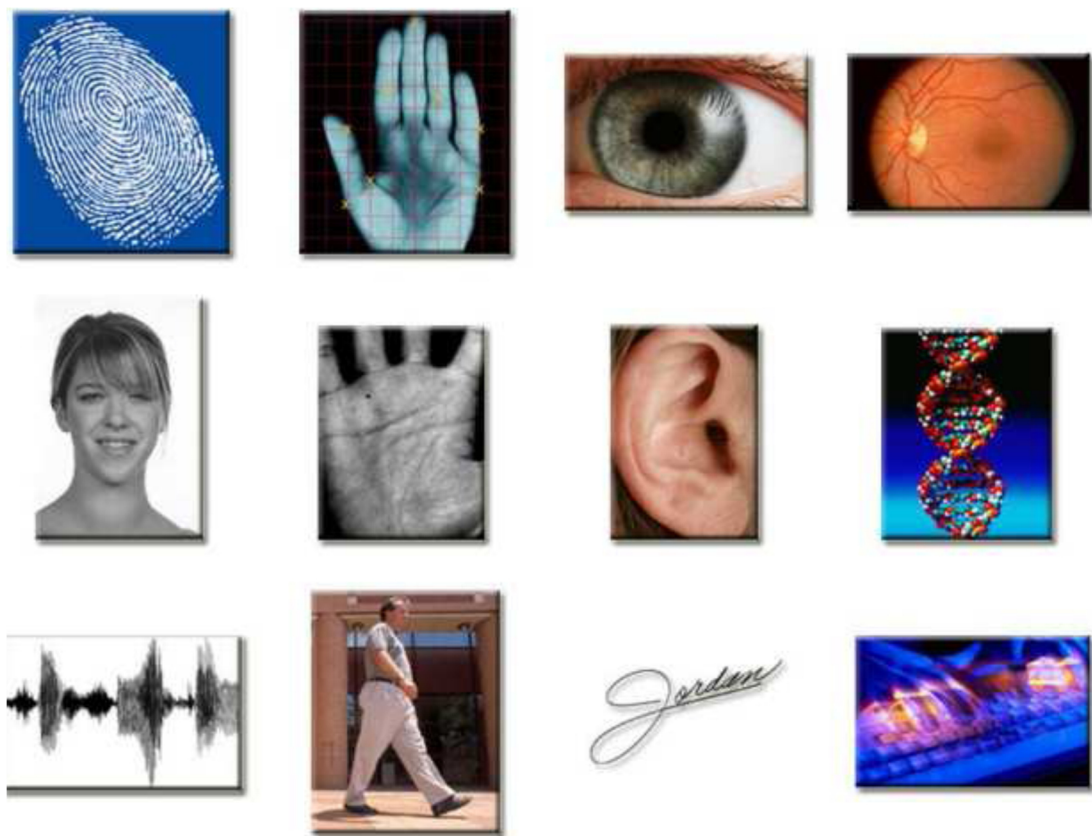


Fig.1.3- Différentes modalités biométriques [6].

1.3.1 Les technologies biométriques

Il existe différentes caractéristiques par lesquelles il est possible d'identifier un individu sont appelées modalités biométriques (Figure 1.3). Les plus couramment utilisées sont :

- *L'iris* : c'est la région annulaire située entre la pupille et le blanc de l'œil. La reconnaissance de l'iris est une technologie plus récente puisqu'elle ne s'est véritablement développée que dans les années 80, principalement grâce aux travaux de J.Daugman [7]. La reconnaissance par iris est très utilisée dans les applications d'identification et de vérification, car il est hautement distinctif, unique, sa forme est stable et il est protégé et très robuste, mais toutefois les équipements d'acquisition coûtent chères [2].
- *La rétine* : la rétine est la couche sensorielle de l'œil qui permet la vision, cette zone est parcourue par des vaisseaux sanguins dont leurs positions est inchangeable durant toute la vie de la personne. On mesure dans cette technique la forme des vaisseaux sanguins qui se trouvent dans l'arrière de l'œil. Puisque les utilisateurs perçoivent cette technologie comme étant intrusive, la reconnaissance par rétine reste peu populaire et actuellement il n'existe peu dispositif commercial [1].
- *La voix* : la reconnaissance par voix utilise les caractéristiques vocales pour identifier les personnes. En 1962 Lawrence Kersta [8] a prouvé que la voix de chaque personne est unique et qu'il est possible de la présenter graphiquement. Un téléphone ou un microphone peut être utilisé comme dispositif d'acquisition, ce qui rend cette technologie relativement économique et facilement réalisable, cependant elle peut être perturbée par des facteurs extérieurs comme le bruit de fond [1].
- *L'empreinte digitale* : représente le modèle de relief cutané des doigts. Après la capture de l'image de l'empreinte, on fait un rehaussement de l'image. Ensuite on identifie et on extrait les minuties, qui vont être comparées avec l'ensemble des minuties sauvegardées des autres utilisateurs. C'est l'une des premières (sinon la plus anciennes) modalités utilisées. Les avantages de cette technique sont : le non changement des empreintes, la haute stabilité et fiabilité, en plus du prix accessible des équipements. Les inconvénients sont : la fragilité par rapport aux saletés et la possibilité d'endommager ou d'élimination des empreintes [1] [2].
- *La géométrie de la main* : la géométrie de la main ou du doigt est une mesure automatisée de plusieurs dimensions, notamment la largeur de la main, des doigts et la longueur des

doigts. Cette technique est récente, simple, et bien acceptée par les utilisateurs qui suivent des guides des capteurs (infrarouge, des appareils photos numériques) pour qu'ils bien positionner leurs doigts, ce qui rend ainsi la détection et la segmentation plus aisée, cependant ce genre de système peut être trompé par de vrais jumeau ou même par des personnes ayant des formes de la main proches [1].

- *La signature* : la vérification par signature est une méthode automatique de mesure des modèles d'écritures des personnes. Cette technologie examine un ensemble de dynamiques comme la vitesse, la direction, et la pression de l'écriture, le temps pendant lequel le stylo est en contact avec le papier, le temps pris pour faire la signature et les positions où le stylo est relevé et abaissé sur le papier [2]. Bien que la signature soit bien acceptée par les utilisateurs, sa variabilité (à cause de l'état de santé ou l'état émotionnel de l'individu) pose un grand problème.

- *Le visage* : la reconnaissance basée sur le visage vérifie l'identité, en comparant les données caractéristiques extraites des visages des différentes personnes à partir de vidéos ou d'images fixes. La fabrication des caméras connaît actuellement d'énormes progrès, on est capable de prendre des photos de personnes distantes sans qu'ils ne sachent qu'ils sont observés. C'est pour cette raison que la reconnaissance de visages est souvent utilisée dans les applications de contrôle de frontières, dans la sécurité des établissements, des zones urbaines et dans l'identification des conducteurs. C'est une technique commune, populaire, simple et qui a beaucoup d'avantages à savoir : l'utilisation des visages qui sont des données publiques, offrant la possibilité de s'intégrer aux systèmes de surveillance existants et elle ne nécessite pas des équipements chères. Cependant, des éléments externes comme l'éclairage, le bruit, l'expression faciale et la posture dégradent les taux de reconnaissance [1] [2].

Le visage peut être vu comme le moyen le plus naturel de reconnaître une personne. De nombreuses recherches sont effectuées dans ce domaine, mais les problèmes sont nombreux. Cette technologie de reconnaissance faciale est abordée en détails dans la suite du manuscrit. Notons également que d'autres modalités biométriques ont émergé depuis peu : ainsi il a été montré que des modalités comme l'articulation des doigts, la dynamique de la frappe au clavier, le réseau veineux, le rythme cardiaque, et la géométrie de l'oreille peuvent être utilisées pour la reconnaissance d'un individu [9] et d'autres se base sur l'analyses biologiques comme l'A.D.N (Support matériel de l'hérédité), l'odeur corporelle,...etc.

1.3.2 Évaluation des performances des systèmes biométriques

Chaque caractéristique (ou modalité) biométrique a ses forces et ses faiblesses, et le choix dépend de l'application visée. On ne s'attend à ce qu'aucune modalité biométrique ne réponde efficacement aux exigences de toutes les applications. En d'autres termes, aucun système biométrique n'est "optimal". Faire correspondre un système biométrique spécifique à une application dépend du mode opérationnel de l'application et des caractéristiques biométriques choisies. Plusieurs études ont été menées afin d'évaluer les performances de ces systèmes [3]. La société américaine International Biometric Group [IBG] a par exemple effectué une étude basée sur quatre critères d'évaluation [10] :

- *Intrusivité* : ce critère permet de classifier les systèmes biométriques en fonction de l'existence d'un contact direct entre le capteur utilisé et l'individu à reconnaître. La reconnaissance faciale est une technique « non intrusive », car il n'existe aucun contact entre le capteur (la caméra) et le sujet, elle est bien acceptée par les utilisateurs à l'inverse d'autres techniques « intrusives » comme l'iris où un contact direct est nécessaire entre le capteur et l'œil.
- *Fiabilité* : dépend de la qualité de l'environnement (éclairage par exemple) dans lequel l'utilisateur se trouve. Ce critère influe sur la reconnaissance de l'utilisateur par le système. Nous verrons ce point en détail dans les sections suivantes.
- *Coût* : doit être modéré, à cet égard nous pouvons dire que la reconnaissance faciale ne nécessite pas une technologie coûteuse. En effet, la plupart des systèmes fonctionnent en utilisant un appareil à photo numérique de qualité standard.
- *Effort* : requis par l'utilisateur lors de la saisie de mesures biométriques, et qui doit être réduit le plus possible. La reconnaissance faciale est la technique biométrique la plus facile à utiliser car elle est non contraignante.

Une comparaison (avantages / inconvénients) des principales technologies biométriques (tableau 1.1) en se basant sur la facilité ou l'ergonomie d'utilisation, la vulnérabilité aux attaques et aux contournements et la fiabilité relative à la précision et à l'efficacité de la reconnaissance [10].

Tab.1.1- Comparaison entre les technologies biométriques [10].

Techniques	Avantages	Inconvénients
Empreintes Digitales	Coût, ergonomie moyenne, facilité de mise en place, taille du capteur	Fiabilité des appareils de mesure, acceptabilité moyenne, possibilité d'attaques (rémanence de l’empreinte...)
Géométrie de la Main	Très ergonomique, bonne acceptabilité	Système encombrant, coût, perturbation par des blessures et l'authentification des membres d'une même famille, permanence des données.
Visage	Coût, peu encombrant, bonne acceptabilité	Jumeaux, psychologie, déguisement, vulnérabilité aux attaques
Rétine	Fiabilité, pérennité	Coût, acceptabilité faible, installation difficile
Iris	Fiabilité	Acceptabilité très faible, contrainte d'éclairage
Voix	Coût, facilité	Vulnérable aux attaques
Signature	Coût, ergonomie	Dépendant de l'état émotionnel de la personne, fiabilité

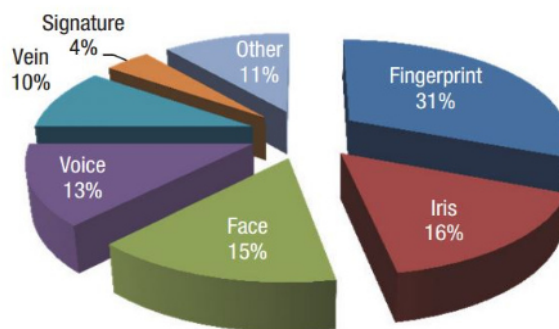
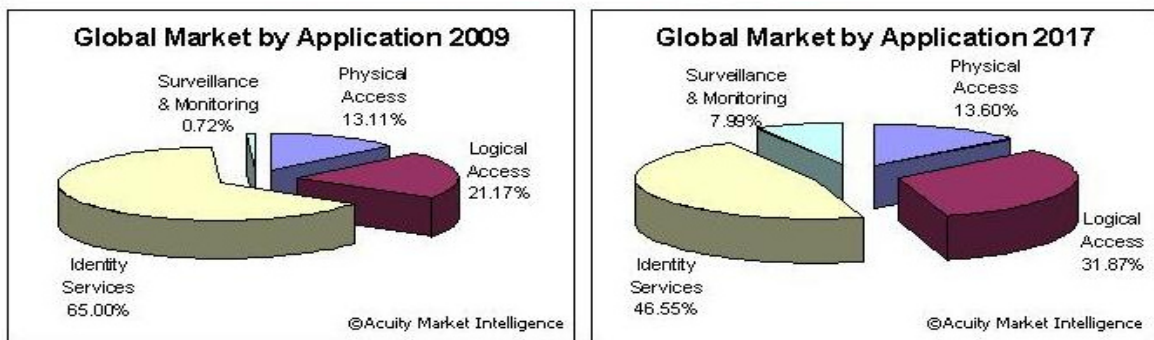


Fig.1.4- Le visage occupe la deuxième place en biométrie [3].

1.3.3 Applications des systèmes biométriques

Le marché de la biométrie ne cesse d'évoluer depuis son apparition, le chiffre d'affaire de l'industrie biométrique est en forte croissance [10]. Une partie importante de cette progression s'attache au contrôle d'accès aux systèmes d'information (ordinateur / réseau) et au commerce électronique, quoique les applications du secteur public continuent à être une partie essentielle de l'industrie. Le champ d'application de la biométrie couvre potentiellement tous les domaines de la sécurité où il est nécessaire de connaître l'identité des personnes. Aujourd'hui, les principales applications sont la production de titres d'identité, le contrôle d'accès à des sites sensibles, le contrôle des frontières, l'accès aux réseaux, systèmes d'information, stations de travail, le paiement électronique, la signature électronique et même le chiffrement de données. Cette liste n'est pas exhaustive, et de nouvelles applications vont très certainement voire rapidement le jour.

Les applications de la biométrie peuvent être divisées en plusieurs groupes principaux :

- *Applications commerciales* : telles que l'ouverture de réseau informatique, la sécurité de données électroniques, l'e-commerce, l'accès Internet, la carte de crédit, le contrôle d'accès physique, le téléphone cellulaire, la gestion des registres médicaux, l'étude à distance, etc.
- *Applications gouvernementales* : telles que la carte d'identité nationale, le permis de conduire, la sécurité sociale, le contrôle des frontières, le contrôle des passeports, etc.
- *Applications légales* : telles que l'identification de corps, la recherche criminelle, l'identification de terroriste, etc.

De nos jours les systèmes biométriques sont de plus en plus utilisés dans des applications civiles. Par exemple, le dispositif de Schiphol Privium à l'aéroport d'Amsterdam [10], utilise un capteur d'iris pour accélérer la procédure de contrôle des passeports et des visas. Les passagers insèrent leur carte dans un lecteur et se mettent en face d'un appareil photo, ce dernier acquiert l'image de l'œil. Un autre exemple d'utilisation à grande échelle de la reconnaissance faciale est celui de la ville de Newham, dans la banlieue de Londres [9]. Il s'est vu décerner le trophée Big Brother Award en 1998 par l'organisation non gouvernementale Privacy International. Aujourd'hui pour régler un achat avec votre carte bancaire, vous pouvez composer votre code secret ou utiliser le paiement sans contact. Bientôt, vous pourrez peut-être simplement toucher votre carte avec votre pouce. Des

moyens de paiements intégrant un capteur d'empreinte digitale sont en effet actuellement en cours d'expérimentation.

1.3.4 Mesure de la performance d'un système biométrique

Tout d'abord, afin de comprendre comment déterminer la performance d'un système biométrique, il nous faut définir clairement trois critères principaux [11] :

1. Le premier critère s'appelle le taux de faux rejet ("False Reject Rate" ou FRR). Ce taux représente le pourcentage de personnes censées être reconnues mais qui sont rejetées par le système. Le système classe alors deux caractéristiques biométriques provenant de la même personne comme provenant de deux personnes différentes.
2. Le deuxième critère est le taux de fausse acceptation ("False Accept Rate" ou FAR). Ce taux représente le pourcentage de personnes censées ne pas être reconnues mais qui sont tout de même acceptées par le système. Le système classe alors deux caractéristiques provenant de deux personnes différentes comme appartenant à la même personne.
3. Le troisième critère est connu sous le nom de taux d'égale erreur ("Equal Error Rate" ou EER). Ce taux est calculé à partir des deux premiers critères et constitue un point de mesure de performance courant. Ce point correspond à l'endroit où $FRR = FAR$, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations.

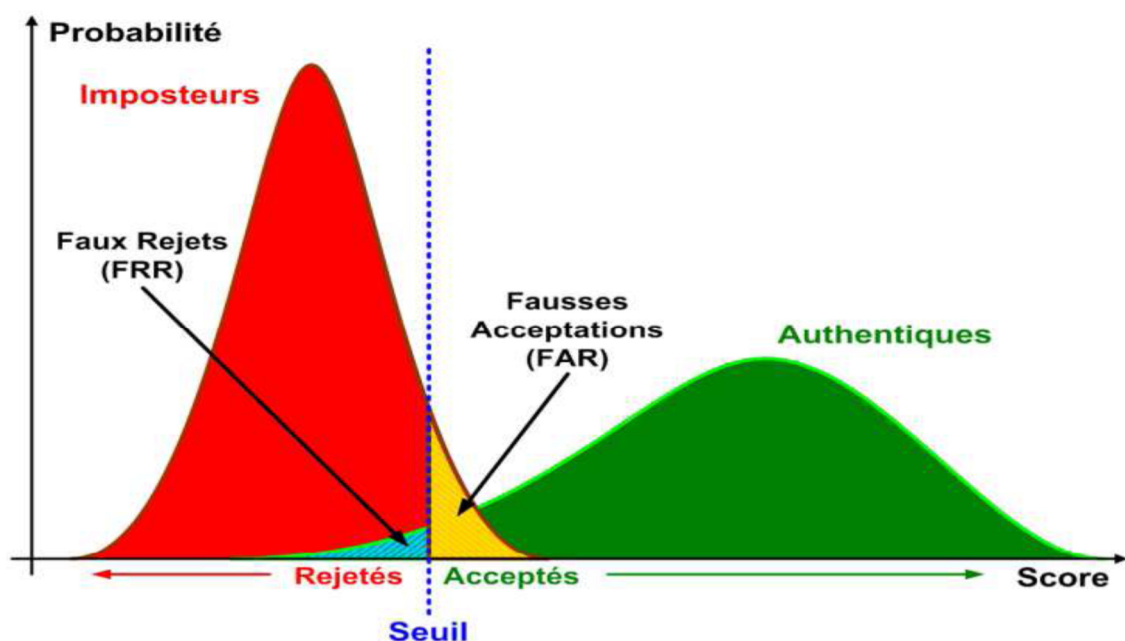


Fig.1.5- Illustration du FRR et du FAR [11].

1.4 La reconnaissance de visages

La reconnaissance faciale est une tâche que les humains effectuent naturellement et sans effort dans leurs vies quotidiennes. La grande disponibilité d'ordinateurs puissants et peu onéreux ainsi que des systèmes informatiques embarqués ont suscité un énorme intérêt dans le traitement automatique des images et des vidéos numériques au sein de nombreuses applications, incluant l'identification biométrique, la surveillance, l'interaction homme-machine et la gestion de données multimédia [11].

La reconnaissance faciale, en tant qu'une des technologies biométriques de base, a pris une part de plus en plus importante dans le domaine de la recherche, ceci étant dû aux avancées rapides dans des technologies telles que les appareils photo numériques, Internet et les dispositifs mobiles, le tout associé à des besoins en sécurité sans cesse en augmentation.

La performance des systèmes de reconnaissance faciale s'est significativement améliorée depuis les premiers travaux qui ont été menés et de nombreux algorithmes de reconnaissance du visage ont été proposés depuis. Certaines méthodes se basent sur une photographie (2D) du visage tandis que d'autres méthodes prennent en compte des informations 3D de celui-ci. On peut également noter qu'il existe d'autres méthodes (parfois appelées 2.5D) qui effectuent la reconnaissance du visage en se basant la plupart du temps sur l'information de profondeur [3]. La reconnaissance de visages est un gros challenge tellement intéressant, qu'il a attiré les chercheurs des différents domaines : psychologie, identification de modèles, réseaux de neurones, vision d'ordinateur, infographie... .

C'est pour cela, que la littérature est si vaste et si diverse , qu'on peut répartir l'ensemble des techniques de reconnaissance de visages basées sur les images fixes et la reconnaissance à partir de séquence d'images (vidéo) en trois grandes catégories [2]: les méthodes globales, les méthodes locales et les méthodes hybrides. La reconnaissance à partir du visage se base sur les caractéristiques jugées significatives comme l'écart entre les yeux, la forme de la bouche, le tour du visage, la position des oreilles. Il existe plus de 60 critères fondamentaux. La plupart des systèmes d'identification du visage utilisent du matériel classique du marché, un ordinateur et une caméra pour capturer l'image.

1.4.1 Motivation : (pourquoi la reconnaissance de visages ?)

Durant les vingt dernières années, la reconnaissance automatique des visages est devenue un enjeu primordial, notamment dans les domaines de l'indexation de documents

multimédias et surtout dans la sécurité, ceci est dû aux besoins du monde actuel mais aussi à ses caractéristiques avantageuses dont on peut citer [12] :

- La disponibilité des équipements d'acquisition, leur simplicité et leurs coûts faibles.
- Passivité du système : un système de reconnaissance de visages ne nécessite aucune coopération de l'individu, du genre : mettre le doigt ou la main sur un dispositif spécifique ou parler dans un microphone. En effet, la personne n'a qu'à rester ou marcher devant une caméra pour qu'elle puisse être identifiée par le système.

En plus, cette technique est très efficace pour les situations non standards, c'est les cas où on ne peut avoir la coopération de l'individu à identifier, par exemple lors d'une arrestation des criminels. Certes que la reconnaissance des visages n'est pas la plus fiable comparée aux autres techniques de biométrie, mais elle peut être ainsi si on utilise des approches plus efficaces en plus du bon choix des caractéristiques d'identification représentant le visage en question.

Tab.1.2- *Avantages et inconvénients de la Reconnaissance de Visage [12].*

Avantages	Inconvénients
-Bien accepté par le public. -Aucune action de l'utilisateur (peu intrusive). -Pas de contact physique. -Technique peu coûteuse.	-Technologie sensible à l'environnement (éclairage, position, expression du visage...) -Difficultés de différencier de vrais jumeaux. -Sensible aux changements. (barbe, lunettes, moustache, piercing, chirurgie...)

1.5 Les classes des techniques de reconnaissance de visages

On distingue trois catégories de méthodes : les méthodes globales, les méthodes locales et les méthodes hybrides.

1.5.1 Méthodes locales

Ce sont des méthodes géométriques, on les appelle aussi les méthodes à traits, à caractéristiques locales, ou analytiques [13]. L'analyse du visage humain est donnée par la description individuelle de ses parties et de leurs relations. Ce modèle correspond à la manière avec laquelle l'être humain perçoit le visage, c'est à dire, à nos notions de traits de visage et de parties comme les yeux, le nez, la bouche, etc.

La plus part des travaux réalisés se sont concentrés sur l'extraction des traits à partir d'une image du visage et sur la définition d'un modèle adéquat pour représenter ce visage. Un certain nombre de stratégies automatiques et semi-automatiques ont modélisé et classé les visages sur la base de distances normalisées et angles entres points caractéristiques.

La stratégie est qualifiée d'automatique ou non selon que la phase d'extraction des points est faite par l'ordinateur ou qu'elle est assistée par un opérateur. Cette phase constitue l'étape clé du processus, car la performance du système entier dépend de la précision avec laquelle les informations utiles sont extraites [13] [14].

L'avantage de ces méthodes réside dans la prise en compte de la particularité du visage tant que forme naturelle à reconnaître, en exploitant les résultats de la recherche en neuropsychologie et psychologie cognitive sur le système visuel humain. La difficulté éprouvée lors de la considération de plusieurs vues du visage, ainsi que le manque de précision dans la phase extraction des points, constituent leur inconvénient majeur. En plus ces techniques sont facilement affectées par l'information non pertinente.

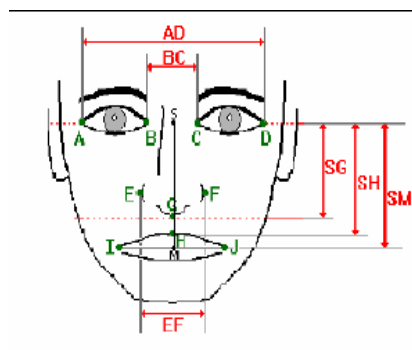


Fig.1.6- Distances entre les points caractéristiques du visage [14].

Dans cette catégorie, on trouve plusieurs méthodes comme [14] : Les filtres de Gabor, Modèles de Markov Cachés (Hidden Markov Models (HMM)), L'Algorithme Elastic Bunch Graph Matching (EBGM), Eigen Object (EO)...

1.5.2 Méthodes globales

Les méthodes holistiques appelées aussi méthodes globales, sont des techniques les plus utilisées dans la reconnaissance de visage, ces méthodes utilisent la région entière du visage comme entrée à l'algorithme de reconnaissance. Ce sont des techniques très réussies et bien étudiées. Ces méthodes offrent de meilleures performances, mais le problème de

stockage des informations extraites lors de la phase d'apprentissage reste l'inconvénient majeur [14] [15]. Parmi les approches les plus importantes réunies au sein de cette classe on trouve :

- L'Analyse en Composantes Principales (PCA ou Eigen Faces).
- L'Analyse Discriminante Linéaire (LDA).
- Machine à Vecteurs de Support (SVM).
- Les Réseaux de Neurones Artificiels (RNA).
- Mélange de Gaussiennes (GMM).
- Modèle Surfamique du Visage (3D).
- L'approche statistique et probabiliste.

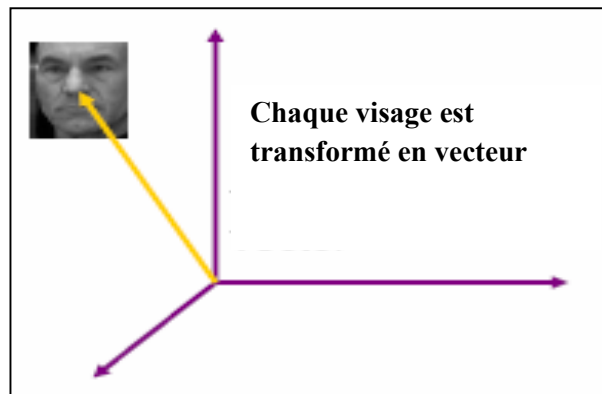


Fig.1.7- L'utilisation de la région entière du visage comme entrée : méthodes globales.

1.5.3 Méthodes hybrides

Comme on a vu précédemment plusieurs approches ont été proposées pour la reconnaissance de visages, sauf qu'aucune d'elle n'est capable de s'adapter aux changements d'environnements tels que la pose, expression du visage, éclairage, etc. Les techniques hybrides combinent les deux types de méthodes, offrant ainsi potentiellement le meilleur des deux types. Elles se basent sur le fait que le système de perception humain emploie les caractéristiques locales et la région entière du visage pour l'identification [14] [15].

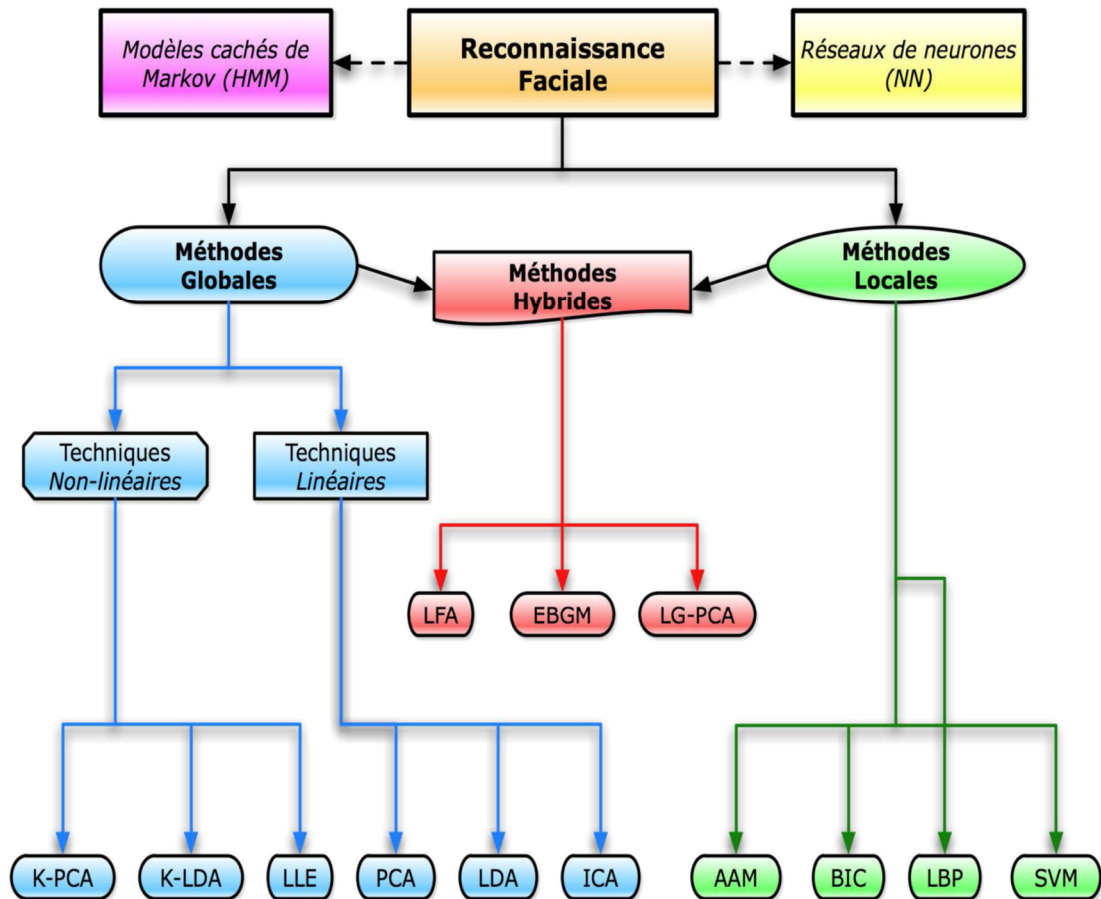


Fig.1.8- les principaux algorithmes principaux utilisés en reconnaissance faciale [12].

1.6 Systèmes de reconnaissance de visage

Idéalement, un système de reconnaissance faciale doit pouvoir identifier des visages présents dans une image ou une vidéo de manière automatique. Le système peut opérer dans les deux modes suivants : authentification ou identification [12].

Un système de reconnaissance de visage est un système d'identification et de vérification d'individus, qui permet de vérifier si une personne appartient à la base de données du système, et de l'identifier si c'est le cas. On peut représenter les systèmes de reconnaissance par le diagramme suivant :

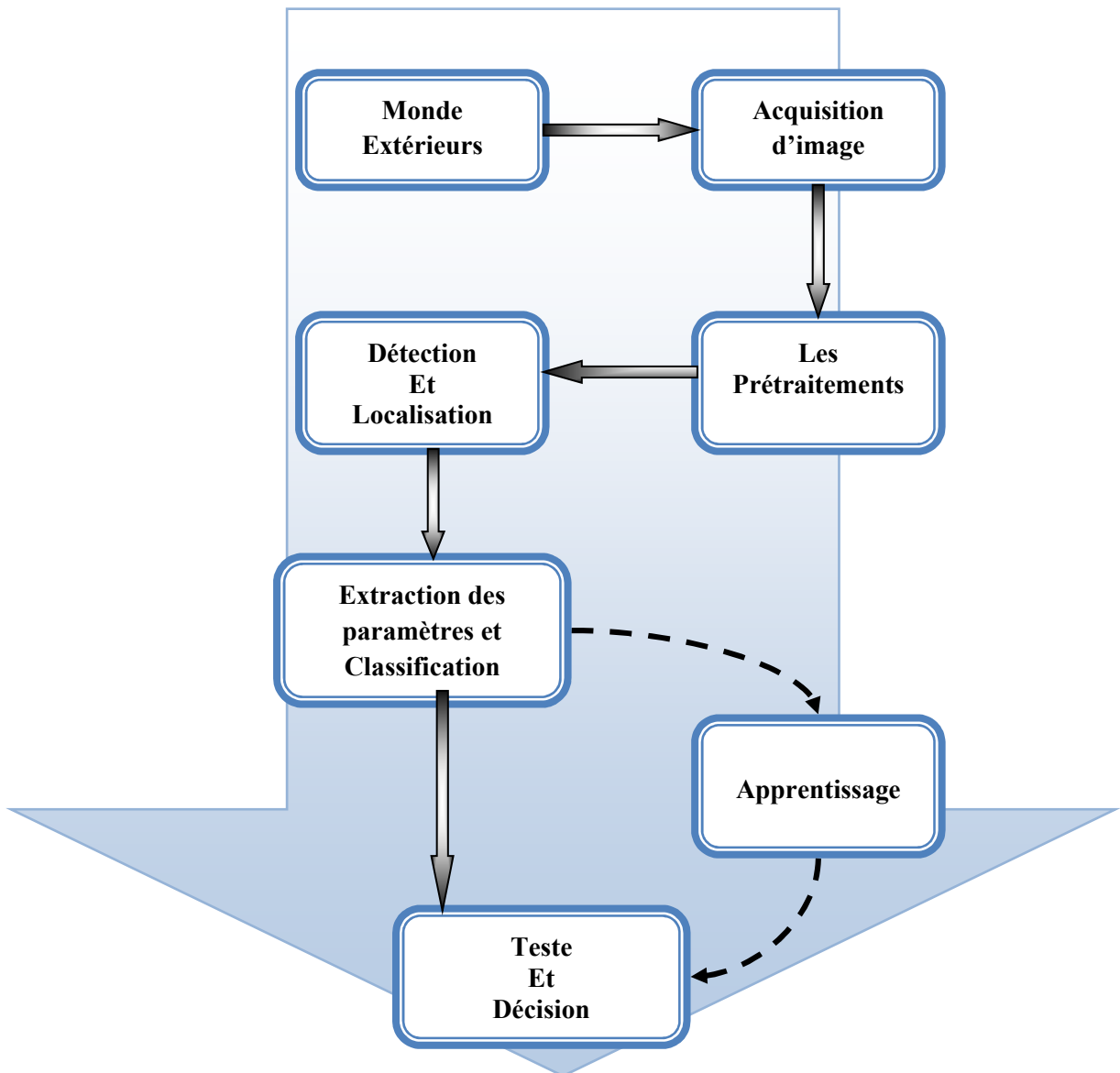


Fig.1.9- Fonctionnement d'un système de reconnaissance de visages.

Comme l'indique la figure 1.9, un système de reconnaissance faciale englobe plusieurs étapes qui sont :

1.6.1. Acquisition de l'image

C'est l'opération qui permet d'extraire du monde réel une représentation bidimensionnelle pour des objets en 3D, cette opération peut être statique (Appareil Photo, Scanner...etc.) ou dynamique (Caméra, Webcam), dans ce cas on aura une séquence vidéo, mais ce niveau on aura juste une image brute [14].

1.6.2. Prés traitements

Les données brutes issues des capteurs sont les représentations initiales des données, à partir desquelles des traitements permettent de construire celles qui seront utilisées pour la reconnaissance. L'image brute peut être affectée par différents facteurs causant ainsi sa détérioration, elle peut être bruitée, c'est à dire contenir des informations parasites à cause des dispositifs optiques ou électroniques [16]. Pour pallier à ces problèmes, il existe plusieurs méthodes de traitement et d'amélioration des images, telle que : la normalisation, l'égalisation de l'histogramme, etc.

1.6.3. Détection puis localisation

Les systèmes de reconnaissance de visages sont complexes. La difficulté réside notamment dans la partie détection automatique du visage, bien que nous développons surtout la partie reconnaissance, il est intéressant de parler de l'étape de détection automatique du visage qui est très importante dans un système de reconnaissance. Ce qui rend la détection de visages dans une image très difficile, c'est surtout la complexité du décor, les variations de poses, les conditions de lumières généralement inconnues, etc. Il existe plusieurs méthodes qui peuvent être appliquées à la détection automatique des visages [16]. Il faut détecter la présence d'un visage dans l'image, ensuite le localiser en vue d'extraire les traits pour le caractériser et le différencier des autres. Le résultat de cette étape est l'obtention de la partie d'image à traiter.

1.6.4 Extraction des paramètres et classification

Dans cette étape on extrait de l'image les informations qui seront sauvegardées en mémoire pour être utilisées plus tard dans la phase de Décision. Le choix de ces informations utiles revient à établir un modèle pour le visage, elles doivent être discriminantes et non redondantes. Ces informations seront ensuite classées, en d'autres termes, affectés à la classe la plus proche, les individus ayant des similarités sont regroupés dans la même classe. Ces classes varient selon le type de décision [16].

1.6.5 Apprentissage

L'apprentissage consiste à mémoriser les modèles calculés dans la phase analyse pour les individus connus. Un modèle est une représentation compacte des images qui permet de faciliter la phase de reconnaissance mais aussi de diminuer la quantité de données à stocker en quelque sorte l'apprentissage est la mémoire du système.

1.6.6 Décision

La décision est la partie du système où on tranche sur l'appartenance d'un individu à l'ensemble des visages ou pas, et si oui quelle est son identité. Donc la décision c'est l'aboutissement du processus. On peut le valoriser par taux de reconnaissance (fiabilité) qui est déterminé par le taux de justesse de la décision. Nous présentons dans ce qui suit l'état de l'art de la reconnaissance des individus.

1.7 Principales difficultés de la reconnaissance de visages

Pour le cerveau humain, le processus de la reconnaissance de visages est une tâche visuelle de haut niveau. Bien que les êtres humains puissent détecter et identifier des visages dans une scène sans beaucoup de peine, construire un système automatique qui accomplit de telles tâches représente un sérieux défi. Ce défi est d'autant plus grand lorsque les conditions d'acquisition des images sont très variables. Il existe deux types de variations associées aux images de visages : inter et intra sujet. La variation inter-sujet est limitée à cause de la ressemblance physique entre les individus. Par contre la variation intra-sujet est plus vaste. Elle peut être attribuée à plusieurs facteurs que nous analysons ci-dessous [11].

1.7.1 Changement d'illumination

L'apparence d'un visage dans une image varie énormément en fonction de l'illumination de la scène lors de la prise de vue. Les variations d'éclairage rendent la tâche de reconnaissance de visage très difficile. En effet, le changement d'apparence d'un visage dû à l'illumination, se révèle parfois plus critique que la différence physique entre les individus, et peut entraîner une mauvaise classification des visages. Les évaluations FRVT [17] ont révélé que le problème de variation d'illumination constitue un défi majeur pour la reconnaissance faciale.

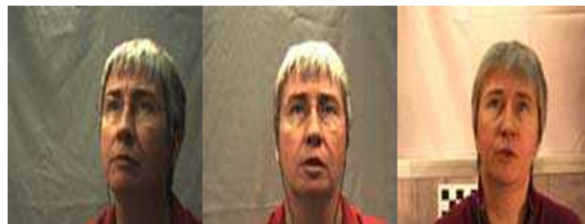


Fig.1.10- Exemple de variation d'éclairage.

1.7.2 Variation de pose

Le taux de reconnaissance de visage baisse considérablement quand des variations de pose sont présentes dans les images. La variation de pose est considérée comme un problème majeur pour les systèmes de reconnaissance faciale. Quand le visage est de profil dans le plan image (orientation $< 30^\circ$), il peut être normalisé en détectant au moins deux traits faciaux (passant par les yeux). Cependant, lorsque la rotation est supérieure à 30° , la normalisation géométrique n'est plus possible [11] (voir figure 1.11).



Fig.1.11- Exemples de variation de poses.

1.7.3 Expressions faciales

Un autre facteur qui affecte l'apparence du visage est l'expression faciale. La déformation du visage qui est due aux expressions faciales est localisée principalement sur la partie inférieure du visage. L'information faciale se situant dans la partie supérieure du visage reste quasi invariable. Toutefois, étant donné que l'expression faciale modifie l'aspect du visage, elle entraîne forcément une diminution du taux de reconnaissance. L'identification de visage avec expression faciale est un problème difficile qui est toujours d'actualité [11] [17].



Fig.1.12- Exemples de variation d'expressions.

1.7.4 Présence ou absence des composants structurels

La présence des composants structurels telle que la barbe, la moustache, ou bien les lunettes peut modifier énormément les caractéristiques faciales telles que la forme, la couleur, ou la taille du visage. De plus, ces composants peuvent cacher les caractéristiques faciales de base causant ainsi une défaillance du système de reconnaissance [16]. Par

exemple, des lunettes opaques ne permettent pas de bien distinguer la forme et la couleur des yeux, et une moustache ou une barbe modifie la forme du visage.

1.7.5 Occultations partielles

Le visage peut être partiellement masqué par des objets dans la scène, ou par le port d'accessoire tels que lunettes, écharpe etc. Dans le contexte de la biométrie, les systèmes proposés doivent être non intrusifs c'est-à-dire qu'on ne doit pas compter sur une coopération active du sujet. Par conséquent, il est important de savoir reconnaître des visages partiellement occultés. L'impact du port de lunettes de soleil, et du cache-nez occultant la partie inférieure du visage sur la reconnaissance faciale rend les performances des algorithmes de reconnaissance faibles [17].

1.7.6 Vieillessement et le changement d'aspect

Les visages changent d'apparence au fil du temps. Les modifications concernent la tension des muscles, l'apparence de la peau (apparition de rides), le port de lunettes, éventuellement le maquillage ou la présence d'une frange occultant une partie du front.

1.8 Conclusion

Dans ce chapitre, nous avons présenté les différentes technologies utilisées dans les systèmes biométriques pour l'identification de personnes, on a donné aussi un aperçu sur les techniques de mesure des performances des systèmes biométriques. Cette étude nous a permis de constater que la reconnaissance de visage suscite de plus en plus l'intérêt de la communauté scientifique, car elle présente plusieurs challenges et verrous technologiques. Enfin, nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance automatique de visages. Les techniques utilisées aux différentes étapes de la reconnaissance de visage seront détaillées dans le chapitre suivant.

CHAPITRE : 2

Etat de l'art des techniques de reconnaissance de visages

2.1	Introduction
2.2	État de l'art des techniques de reconnaissance
2.2.1	Approches globales (ou holistiques).
A.	Analyse en Composantes Principales
B.	Analyse Discriminante Linéaire
C.	Analyse en Composantes Indépendantes
D.	Les réseaux de neurones artificiels
2.2.2	Les approches basées sur les caractéristiques locales
A.	Modèle de Markov caché
B.	Machine à vecteurs de support
C.	Approches basées sur les graphes
D.	Local Binary Pattern (LBP)
2.2.3	Les approches hybrides
2.3	Systèmes biométriques existants pour la reconnaissance des visages
2.4	Conclusion

2.1 Introduction

Comme nous l'avons évoqué au chapitre précédent, un système automatique de reconnaissance de visages se décompose globalement en trois sous-systèmes : détection de visage, extraction des caractéristiques et l'identification (décision). La mise en œuvre d'un système automatique fiable de reconnaissance faciale est un verrou technologique qui est très dure à réaliser. Différentes méthodes et logiciels commerciaux ont vu le jour pour effectuer ce type de reconnaissance. Les performances de ses systèmes dépendent en grande partie de l'efficacité de l'approche d'identification utilisée. Dans cette description de l'état de l'art vu le grand nombre de méthodes appliquées à la reconnaissance du visage, la citation de la totalité des méthodes ne peut être complète. On s'est limité à décrire la plupart des grandes approches classiques et la majorité des techniques qui ont apporté une vraie originalité et une avancée sensible dans le domaine.

Ce chapitre est organisé comme suit. Après avoir positionné notre champ de traitement sur la reconnaissance faciale 2D parmi tant d'autres domaines de reconnaissance de visage dans la section suivante, nous présentons dans un premier temps l'état de l'art des méthodes de reconnaissance de visage, afin de tirer les meilleurs enseignements des différentes approches utilisées dans ce domaine. La seconde partie met en lumière les systèmes mis en œuvre ainsi que les différents domaines d'utilisation de cette technologie en pleine expansion.

2.2 État de l'art des techniques de reconnaissance

Nous commençons la description de l'évolution de la recherche par une description de l'objet à reconnaître. Le visage est un objet à nature tridimensionnel, cet objet peut être sujet à diverses rotations, non seulement planes mais aussi spatiales, et également sujet à des déformations dues aux expressions faciales. La forme et les caractéristiques de cet objet évoluent aussi dans le temps. Les informations de cet objet tridimensionnel sont capturées à l'aide de caméras et à la sortie, nous obtenons une image bidimensionnelle qui le caractérise. Plusieurs méthodes de reconnaissance de visage ont été proposées durant les trois dernières décennies. La reconnaissance du visage est devenue un challenge intéressant qui a attiré les chercheurs de différentes spécialités [12] [18] : psychologie, reconnaissance de forme, réseaux de neurones, vision par ordinateur, infographie...etc. Suite à cela la littérature en reconnaissance du visage est vaste et diverse.

Dans l'état de l'art, trois familles d'algorithmes ressortent. Une première famille, la plus populaire, est basée sur une approche globale qui s'appuie sur des méthodes de réduction d'espace. Une deuxième famille d'algorithmes est basée sur des approches locales, appelées aussi techniques par points caractéristiques et enfin, des algorithmes basés sur des approches hybrides [2].

2.2.1 Approches globales (ou holistiques)

Les techniques de classification par réduction d'espace restent parmi les techniques les plus utilisées et les plus populaires dans le domaine de la reconnaissance du visage [18]. Ces méthodes considèrent l'image entière du visage comme un vecteur dans un espace multidimensionnel. L'objectif de ces méthodes est de trouver une base réduite de projections qui optimise un certain critère et l'espace original est ensuite projeté sur cette base de dimension inférieure. De ce fait on réduit considérablement la complexité des calculs tout en conservant l'information discriminante entre les images. La classification est habituellement effectuée selon une mesure de distance simple dans l'espace multidimensionnel. Ainsi, l'utilisation de distances classiques dans l'espace projeté ne permet pas toujours de réaliser une bonne classification entre les classes. Les méthodes globales (de réduction d'espace) peuvent se décomposer en deux types de techniques : les techniques linéaires et les techniques non linéaires. Parmi les approches les plus importantes réunies au sein de cette classe on trouve :

A. Analyse en Composantes Principales

Une des méthodes les plus anciennes et les mieux étudiées pour la représentation des visages dans un espace de dimension réduite, reste la technique dite des Eigenfaces présentée par Turk et Pentland en 1991 [19] basé sur l'analyse en composante principale (ACP). Cette représentation est réalisée sur un ensemble d'images d'apprentissage de visages. Les principaux vecteurs propres résultant de l'ACP définissent le nouvel espace. Les images de visages sont ensuite projetées sur cet espace, et les vecteurs obtenus sont utilisés pour la classification. L'ACP appliquée aux visages permet la réduction de la dimension de représentation du visage [20]. C'est une approche rapide et très facile à mettre en œuvre, néanmoins elle présente quelques inconvénients tels qu'une baisse des performances (temps d'exécution plus long) dans le cas où la base d'image est volumineuse. De plus, pour l'approche Eigenfaces, l'ajout ou la suppression d'un individu de la base

engendre le recalcul des visages propres. Aussi, cette méthode n'est pas robuste aux variations :

- d'échelle, mais un changement d'échelle préalable est toujours possible.
- de l'orientation du visage (visage plus ou moins de profile, visage incliné).
- d'expression du visage qui correspondent aux sept émotions suivantes : la neutralité, la joie, la tristesse, la surprise, la peur, la colère et le dégoût.
- de la lumière, étant donné que le visage peut être exposé à la lumière lors de sa prise en photo, des zones d'ombre et de lumière peuvent apparaître sur le visage faussant ainsi l'authentification. En effet, deux images d'un même individu prises dans des conditions d'éclairage différentes, pourraient se retrouver éloignées dans l'espace propre. En plus les systèmes utilisant l'ACP sont souvent confrontés à un problème majeur qui est la difficulté de l'estimation du seuil de décision et le nombre des vecteurs propres à retenir [20].

De nombreux travaux ont été réalisés sur le choix des vecteurs propres à retenir pour définir le nouvel espace. Ainsi, Kirby et al. [21] proposent un critère basé sur l'énergie des valeurs propres associées aux vecteurs propres. Martinez et al. [22] montrent que les taux de reconnaissance peuvent être améliorés en ignorant les premiers vecteurs propres (ceux dont les valeurs propres associées sont les plus grandes), ceux-ci encodant souvent les variations d'illumination. En 1996, l'approche de l'ACP a été étendue à la version non linéaire par l'introduction des fonctions à noyaux non linéaires, appelée Kernel Principal Component Analysis (KPCA) [23]. L'hypothèse est que si les visages ne sont pas linéairement séparables dans l'espace d'entrée, ils seront linéairement séparables dans un espace non linéaire, d'où l'introduction de notion de noyau de projection non linéaire.

La méthode des Eigen Faces repose sur une utilisation des premiers vecteurs propres comme visages propres, d'où le terme Eigen Faces. La base formée par ces vecteurs constitue un espace qu'on utilise pour représenter les images des visages. Les personnes se voient donc attribuer un vecteur d'appartenance pour chacune de leur image. Cela étant dit, la reconnaissance est réalisée en comparant les coefficients de projection d'un visage test avec ceux appartenant aux visages d'entraînement. Après le choix des visages sur lesquels on va travailler, l'idée de cette méthode est de représenter chaque image sous forme de vecteur, puis regrouper ses derniers pour former une matrice de vecteur qu'on appellera matrice d'image, soit Γ cette matrice. On suppose Γ_i un vecteur de $N^2 \times 1$ correspondant à une image I_i de taille $N \times N$. Le but est de représenter Γ dans un espace de dimension inférieur, ce dernier devra être orthogonale (vecteurs composant sa base sont orthogonaux

deux à deux), pour pouvoir discriminer les images. Nous présentons dans ce qui suit les principales étapes de la méthode Eigen Face.

Etape1 :

Cette étape consiste à définir les images des personnes, soit M le nombre d'image allant de I_1, I_2, \dots, I_M . Ces images doivent être centré et de même taille.

Etape2 :

Cette étape consiste à représenter chaque image I_i par un vecteur Γ_i , pour cela on va superposer les colonnes de chaque image.

Exemple :

Soit I_i une image de taille 3×3

$$I_i = \begin{bmatrix} 5 & 3 & 0 \\ 2 & 7 & 9 \\ 1 & 1 & 4 \end{bmatrix} \quad \text{On obtiendra} \quad \Gamma_i = \begin{bmatrix} 5 \\ 3 \\ 0 \\ 2 \\ 7 \\ 9 \\ 1 \\ 1 \\ 4 \end{bmatrix}$$

Etape 3 :

Cette étape consiste à calculer la moyenne des visages et de les représenter sous forme de vecteur Ψ .

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \dots\dots\dots 2.1$$

Etape 4 :

Cette étape consiste à enlever la moyenne de la matrice d'image, en d'autres termes : enlever tous ce qui est commun aux individus. Soit Φ la matrice résultante obtenue comme suit :

$$\Phi_i = \Gamma_i - \Psi \dots\dots\dots 2.2$$

Etape 5 :

Cette étape consiste à construire la matrice de covariance C de la matrice Φ . La matrice de covariance représente l'interaction entre les individus.

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = A \times A^T \quad (N^2 \times N^2) \dots \dots \dots 2.3$$

Où

$$A = [\Phi_1 \Phi_2 \dots \Phi_M] \quad (N^2 \times M) \dots \dots \dots 2.4$$

Etape6 :

Cette étape consiste à calculer les vecteurs propres qui constitueront notre espace d'étude. Ces vecteurs u_i seront tirés de la matrice de covariance $C = AA^T$. Cependant, cette matrice est très grande donc non praticable, à ce niveau, nous allons procéder à une déviation qui nous permettra de détourner le problème.

Considérons la matrice $A^T A (M \times M)$, si on essaye de calculer les valeurs propres de cette dernière, on remarquera que ce sont les mêmes que celles de la matrice C . De plus il existe un lien entre les vecteurs propres de chacune des deux matrices qui est le suivant : $u_i = Av_i$

On notera que pour la matrice AA^T on aura N^2 valeurs et vecteurs propres. Et pour la matrice $A^T A$ on aura M valeurs et vecteurs propres. Ainsi, les M valeurs propres de $A^T A$ correspondent aux M plus grandes valeurs de la matrice AA^T (en correspondance avec leurs vecteurs propres).

Etape 7 :

Cette étape est assez simple à réaliser, elle consiste à prendre K vecteurs propres correspondant aux K plus grandes valeurs propres.

Une fois les vecteurs propres trouvés, il ne reste plus qu'à déterminer la représentation des visages dans notre nouvel espace, pour cela on procède comme suit : Chaque visage (moins la moyenne) sera représenté comme étant une combinaison linéaire des K vecteurs propres choisis.

$$\hat{\phi}_i - mean = \sum_{j=1}^K w_j u_j \dots \dots \dots 2.5$$

$$w_j = u_j^T \phi_i \dots \dots \dots 2.6$$

Donc chaque visage d'apprentissage ϕ_i sera représenté dans l'espace comme suit :

$$\Omega_i = \begin{bmatrix} w_1^i \\ w_2^i \\ \cdot \\ \cdot \\ w_K^i \end{bmatrix} \quad i = 1, 2, \dots, M$$

Le problème qui reste à résoudre est le choix de K , la dimension de l'espace, Pour cela on aura besoin d'un seuil (pourcentage) dit de quantité d'information. Le but est de pouvoir représenter une certaine quantité d'information en un minimum de vecteurs base.

$$\frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^N \lambda_i} > SEUIL \dots\dots\dots 2.7$$

B. Analyse Discriminante Linéaire

L'analyse discriminante linéaire (LDA) appelée également "Fisher Linear Discriminant" (FLD), est l'une des méthodes les plus utilisées pour la reconnaissance de visage [24]. Elle est utilisée pour trouver la combinaison linéaire des caractéristiques qui séparent le mieux les classes d'objet ou d'événement. Les combinaisons résultantes peuvent être employées comme classificateur linéaire, ou généralement dans la réduction de caractéristiques avant la classification postérieure. LDA est étroitement lié à l'ACP, du fait que tous les deux recherchent des combinaisons linéaires des variables qui représentent au mieux les données. Les combinaisons linéaires obtenues en utilisant FLD s'appellent les Fisherfaces, en analogie avec les Eigenfaces [24]. Récemment, des algorithmes pour l'application directe de la LDA pour l'extraction des informations discriminantes, ont été proposés. Ces solutions ont pour but d'éviter la perte d'information discriminante qui se produit après une réduction par ACP. On cite la DLDA (Direct LDA) [25], la RLDA (Regression LDA) [26], GDA (General Discriminant Analysis) [27] et l'analyse non linéaire discriminante de Fisher KFDA (Kernel Fisher Discriminant Analysis) [28].

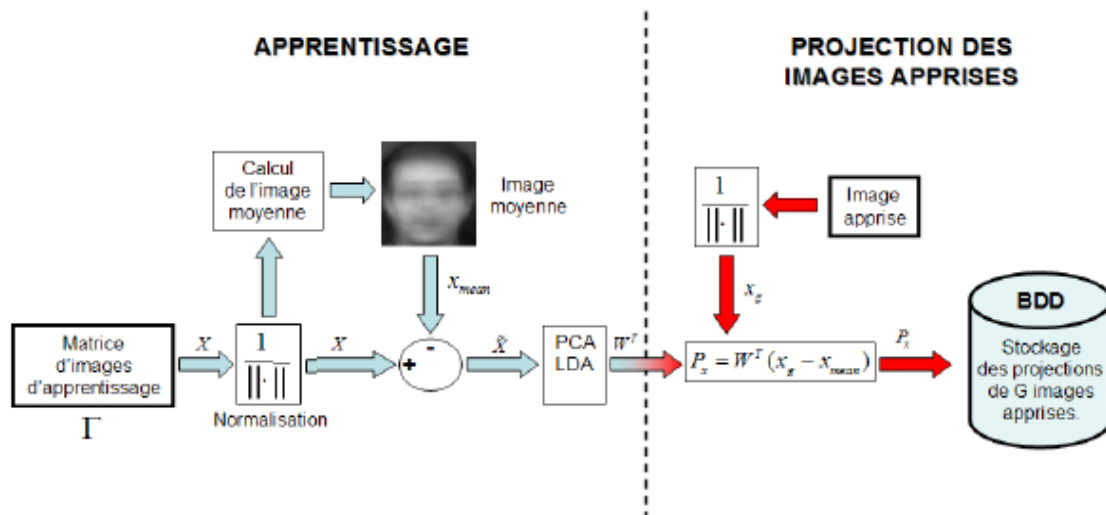


Fig.2.1- Système de reconnaissance (apprentissage) faciale utilisant ACP et LDA [29].

D'autres techniques linéaires ont également été utilisées pour le calcul de vecteurs caractéristiques. On peut citer l'analyse en composantes indépendantes (ICA) [30], la factorisation de matrices non négatives (NMF) [31], l'analyse discriminante bilinéaire (BDA) dans [32], et la technique dite de Vecteurs communs discriminants (DCV) [33]. Bien que ces méthodes linéaires soient assez efficaces, elles manquent de précision dès lors que les images de visages subissent des transformations non linéaires. Une simple modification de la luminosité transforme celui-ci de façon non linéaire étant donné la complexité de la forme.

C. Analyse en Composantes Indépendantes

Basé sur l'argument qu'une grande partie de l'information faciale est contenue dans les statistiques d'ordre élevé, Bartlett et al, ont proposé dans [34] d'utiliser l'Analyse en Composantes Indépendantes (en français *ACI*, en anglais : *Independent Component Analysis* : *ICA*) pour extraire des caractéristiques pour l'identification de visage. Au lieu de simplement décorréler les données comme le font la *PCA* et la *LDA*, l'*ICA* cherche à les rendre statistiquement indépendant. La méthode *ICA* proposé dans [34] repose sur une étape préliminaire de l'analyse en composantes principales *ACP*. Deux architectures ont été proposées pour l'identification de visage : la première architecture vise à obtenir des vecteurs *Eigenfaces* qui soient statistiquement indépendant deux à deux. La deuxième architecture quant à elle, cherche à rendre les coefficients de projection (variables) mutuellement statistiquement indépendant [30].

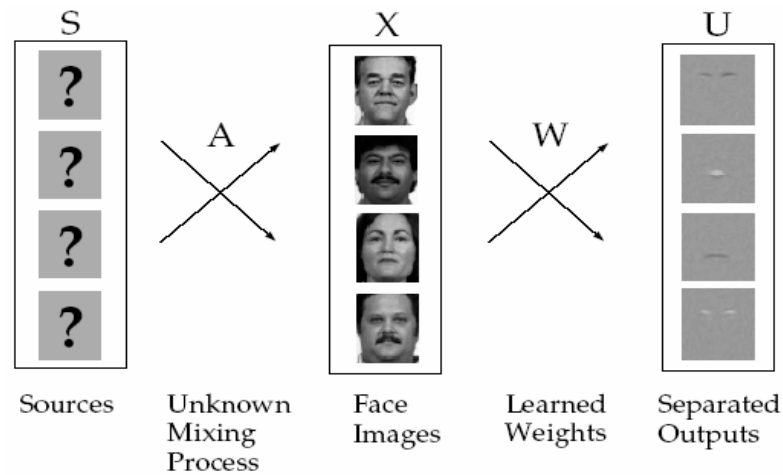


Fig.2.2- Le modèle de synthèse d'image de l'architecture d'ICA [34].

D. Les réseaux de neurones artificiels

L'application des réseaux de neurones artificiels dans la reconnaissance de visages a visé plusieurs problèmes : la classification homme et femme, l'estimation de la pose et la classification des expressions faciales. Les RNA (Réseaux de Neurones Artificiels) ont été utilisés dans nombreuses applications, particulièrement pour la classification de données, la modélisation des processus complexes et le traitement non-linéaire des signaux. Ces recherches ont engendré une panoplie d'architectures de réseaux dont chacune répond parfaitement à une application donnée. On trouve plusieurs types de réseaux de neurones : On note en particulier l'architecture MLP (Multi-Layer Perceptron, Perceptron Multi-Couches : PMC), l'architecture RBF (Radial Basis Function, Fonctions à base radiale) [12].

Dans le cas de la reconnaissance de visages, on constate que les réseaux de neurones ont été employés dans tous les modules intervenant dans la chaîne de traitements. Ils sont utilisés pour la détection de visages, l'extraction de signatures et pour la classification. Une classification par réseau de neurones comporte plusieurs étapes, tout d'abord une phase de prétraitement des images d'apprentissage et l'association à chaque image d'apprentissage (entrée du réseau) un vecteur de sortie, puis vient l'étape d'initialisation (création des couches du réseau). On fait l'apprentissage du réseau, jusqu'à atteindre une certaine erreur minimale (le réseau apprend à bien classifier les images d'apprentissage) [35].

On présente ensuite au réseau une nouvelle image à identifier (phase de reconnaissance ou de simulation ou d'activation du réseau) qui sera finalement affectée à une classe donnée. L'avantage de ce modèle est le gain de temps considérable, cependant l'utilisation

d'exemples pour apprendre apporte le risque de ne pouvoir résoudre que des situations déjà rencontrées, où un phénomène de sur-apprentissage qui spécialiserait le réseau uniquement sur les exemples connus sans généraliser [35].

2.2.2 Les approches basées sur les caractéristiques locales

Dans cette section, nous présentons les principales approches locales basées sur l'étude de caractéristiques extraites de différentes régions des visages. Nous exposons également les techniques hybrides, en ce sens qu'elles utilisent conjointement des caractéristiques globales et locales des visages. Plusieurs méthodes dans la catégorie d'appariement structurel de caractéristiques locales ont été proposées, incluant plusieurs anciennes méthodes basées sur la géométrie des caractéristiques locales.

A. *Modèle de Markov caché*

Les modèles de Markov cachés (HMM) sont massivement utilisés en reconnaissance de formes, en intelligence artificielle ou encore en traitement automatique du langage naturel. Les HMM sont généralement employés pour la modélisation statistique de séries de vecteur de temps non stationnaire. Les caractéristiques faciales les plus significatives d'une image de visage frontale, à savoir les cheveux, le front, les yeux, le nez et la bouche, se présentent dans un ordre naturel de haut en bas, même dans le cas où il y'a des petites rotations [36].

En se basant sur cette observation, l'image d'un visage peut être modélisée en utilisant un HMM unidimensionnel en assignant à chacune de ces régions un état. Le HMM classe une caractéristique par la propriété de la chaîne de Markov. Dans une chaîne de Markov, la transition d'un état à un autre est probabiliste ce qui donne une distribution de probabilité de l'ensemble des sorties de chaque état. Ce résultat est employé pour faire la comparaison entre deux visages [37]. L'approche HMM a des performances de reconnaissance satisfaisantes, cependant il nécessite des calculs intenses ce qui augmente les temps d'exécution.

B. *Machine à vecteurs de support*

Une machine à vecteurs de support (SVM) est une technique qui a été proposée par V.Vapnik en 1995 [38], elle est utilisée dans plusieurs domaines (classement, statistiques, régression, fusion,...). Depuis son introduction dans le domaine de la reconnaissance de formes, plusieurs travaux ont pu montrer l'efficacité de cette technique principalement en

traitement d'image. L'idée essentielle de cette approche consiste à projeter les données de l'espace d'entrée (appartenant à des classes différentes) non linéairement séparables, dans un espace de plus grande dimension appelé espace de caractéristiques, de façon à ce que les données deviennent linéairement séparables [38] [39].

Le SVM est une méthode de discrimination de données basé sur un apprentissage supervisé, utilisée pour la classification et la régression. Elle consiste à séparer deux ou plusieurs ensembles de points par un hyperplan. Selon les cas et la configuration des points, la performance de la machine à vecteurs de support peut être supérieure à celle d'un réseau de neurones ou d'un modèle de mixture gaussienne [12]. Le but du SVM est donc de trouver une séparatrice qui minimise l'erreur de classification sur l'ensemble d'apprentissage mais qui sera également performante en généralisation sur des données non utilisées en apprentissage. Pour cela le concept utilisé est celui de marge (d'où le nom de séparateurs à vaste marge). La marge est la distance quadratique moyenne entre la séparatrice et les éléments d'apprentissage les plus proches de celle-ci appelés vecteurs de support (figure 2.3). Ces éléments sont appelés vecteurs de support car c'est uniquement sur ces éléments de l'ensemble d'apprentissage qu'est optimisée la séparatrice. Dans cette section, nous présenterons quelques notions de base qui sont Hyperplan, marge et support vecteur. Nous mettrons également en évidence l'utilisation des SVM dans les deux cas où les données sont (linéairement ou non linéairement) séparables [12].

Etant donné un ensemble de données d'apprentissage $S = \{(x_1, u_1), \dots, (x_i, u_i), \dots, (x_m, u_m)\}$, avec $x_i \in \mathcal{R}^d$, $i = 1, \dots, n$ et $u_i \in \{-1, +1\}$. Où : x est l'ensemble des objets déjà classés (espace d'entrée), u est l'ensemble des classes (espace d'arrivée). On suppose dans un premier temps qu'il existe un hyperplan linéaire permettant de discriminer les données positives (étiquetées +1) des données négatives (étiquetées -1). L'identification d'un tel hyperplan dans x revient à trouver une fonction hypothèse $h(x) = w^t x + w_0$ telle que la paire (w, w_0) définie l'équation de l'hyperplan $w \cdot x + w_0 = 0$, nommé l'hyperplan séparatrice [40]. La fonction $h(x)$ correspond à l'équation d'un hyperplan dans x de vecteur normal w . la distance d'un point x à l'hyperplan d'équation $h(x) = w^t x + w_0$ est égale à $\frac{w \cdot x_i + w_0}{\|w\|}$, où

$\|w\|$ est la norme euclidienne du vecteur w .

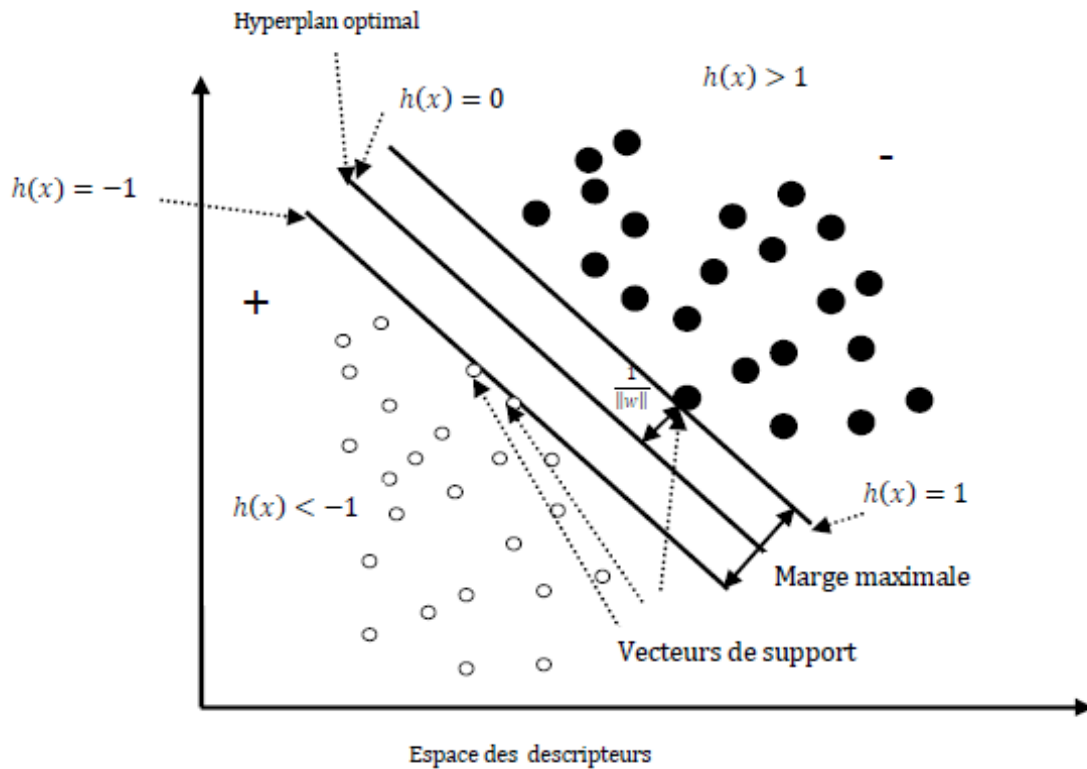


Fig.2.3- Support vecteur machine (SVM) : calcul de l'hyperplan optimal [40].

En général, il existe une infinité des séparateurs linéaires entre les données d'apprentissage. On peut alors chercher parmi ces hyperplans, un hyperplan séparateur optimal celui qui est se trouve au milieu des deux ensembles de données positives et de données négatives.

Pour maximiser la marge, on doit maximiser $\frac{2}{\|w\|}$, ce qui revient à minimiser $\frac{1}{2}\|w\|^2$.

Trouver l'hyperplan optimal revient à résoudre le problème quadratique suivant qui porte sur les paramètres w et w_0 :

$$\begin{cases} \min \frac{1}{2}\|w\|^2 \\ \forall i, y_i(w \cdot x_i + w_0) \geq 1 \end{cases} \dots\dots\dots 2.8$$

Les SVMs sont des algorithmes d'apprentissage utilisés pour optimiser le classement de données, dont l'objectif est de prédire leur classe d'appartenance, pour cela deux types de régularisations peut être utilisées pour la recherche de l'hyperplan optimale

$$\min \frac{1}{P} w^T w + C \sum_{i=1}^P \max(0, 1 - y_i' (w^T x_i + b)) \dots\dots\dots 2.9$$

Où $w^T w$ est la norme de Manhattan (également appelée norme L1), et C représente le paramètre de pénalité (peut être une valeur arbitraire ou une valeur sélectionnée à l'aide du réglage hyper-paramètre), y' est l'étiquette réelle et $w^T x + b$ est la fonction de prédiction. L'équation (2.9) est connu sous le nom de L1-SVM, avec la perte standard. Son équivalent différentiable, L2-SVM (Eq. 2.10) qui fournit des résultats plus stables [39].

$$\min \frac{1}{p} \|w\| + C \sum_{i=1}^p \max(0, 1 - y'_i (w^T x_i + b))^2 \dots\dots\dots 2.10$$

Où $\|w\|_2$ est la norme euclidienne (également appelée norme L2), avec la perte quadratique.

Dans le cas où les données sont non linéairement séparables, c'est-à-dire la surface séparatrice est non linéaire, on transpose le problème dans un autre espace de dimension plus élevé pour rendre les nuages de points des deux classes linéairement séparables au moyen d'une transformation. A ce stade, le problème se situe dans le choix de la transformation ou plus généralement à la fonction noyau K (Kernel). Il existe plusieurs types de fonction noyau utilisés avec les SVM, nous citons les plus utilisés : [40]

- Noyau linéaire : $K(x,y) = x.y \dots\dots\dots 2.11$

- Noyau polynomiale : $K(x,y) = (x.y)^n \dots\dots\dots 2.12$

- Noyau gaussien : $K(x,y) = \exp\left(-\frac{\|x-y\|^n}{2\sigma}\right) \dots\dots\dots 2.13$

- Sigmoidale : $K(x,y) = \tanh(a(x.y) - b) \dots\dots\dots 2.14$

L'avantage de la création d'une fonction de décision avec l'algorithme des SVM est que la solution produite correspond à l'optimum d'une fonction convexe. Elle ne possède donc pas plusieurs optima locaux comme pour les réseaux de neurones (dans leur formulation classique), mais un optimum global.

C. Approches basées sur les graphes

Plutôt que d'utiliser des méthodes purement géométriques, certains chercheurs ont choisi de représenter les caractéristiques locales du visage sous forme de graphes basés sur la décomposition en ondelettes de Gabor [41]. La reconnaissance de visages est alors formulée comme un problème de mise en correspondance de graphes. Cependant, une fois construit, le graphe topologique ne peut pas être modifié. Or, les images de visage changent facilement d'apparence en raison des différentes variations (illumination, expression, pose, etc.), et du coup un schéma de graphe topologique fixe n'est plus adéquat.

Les techniques d'identification présentées dans cette partie sont basées sur la mise en correspondance de graphes. Nous nous intéressons particulièrement aux approches appelées *Elastic Graph Matching (EGM)* et *Elastic Bunch Graph Matching (EBGM)* [41] [42]. Ces méthodes sont basées sur la méthodologie d'*Architecture de Lien Dynamique (DLA)*. Cette dernière est étroitement liée à la théorie des réseaux de neurones. Elle permet de structurer, par le biais de graphes et des nœuds caractérisant les propriétés locales des visages. Une représentation typique des caractéristiques locales s'effectue en affectant des coefficients d'ondelettes pour différentes échelles et rotations appelées jets [42]. Ces coefficients d'ondelettes, localement estimés, sont robustes au changement d'éclairage, à la translation, à la déformation, à la rotation, et aux facteurs d'échelles. L'approche d'*EBGM* est performante dans le cadre de la reconnaissance de visages. Elle est néanmoins très coûteuse en termes de temps de calcul, tant pour la construction du modèle que pour la mise en correspondance de graphes.

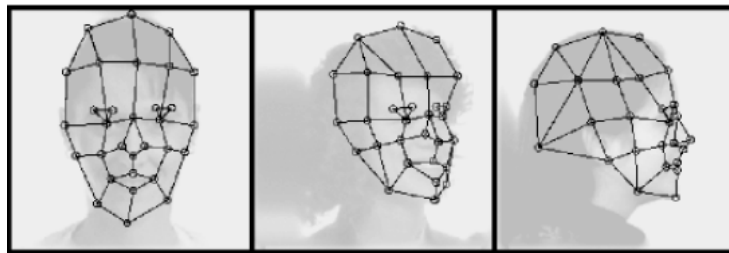


Fig. 2.4- La représentation des visages par graphe rectangulaire [37].

D. Local Binary Pattern (LBP)

L'opérateur LBP a été proposé à la fin des années 90 par Ojala et al, 1996 [43]. L'idée de cet opérateur de texture est de donner à chaque pixel un code dépendant des niveaux de gris de son voisinage. Le niveau de gris du pixel central est comparé à ceux de ses voisins. Le code LBP du pixel courant est alors produit en concaténant ces 8 valeurs pour former un code binaire. La figure (fig.2.5) donne un exemple de traitement de l'opérateur LBP. On obtient donc pour une image en niveaux de gris, une matrice des valeurs LBP contenant des valeurs d'intensité comprises entre 0 et 255. Une fois le code LBP calculé pour tous les pixels de l'image, on calcule l'histogramme de cette image LBP pour former un vecteur de caractéristiques représentant l'image [43].

Les LBP ont également été utilisés pour la reconnaissance faciale, notamment dans [44]. Le visage est subdivisé en sous-régions carrées de taille égale sur lesquelles sont calculées

les caractéristiques LBP. Les vecteurs obtenus sont ensuite concaténés pour obtenir le vecteur de caractéristiques final. Des extensions des LBP comme les MB-LBP (pour Multi-Scale Block Binary Pattern), local ternary patterns (LTP), et center symmetric LBP (CS-LBP), ont été proposées et appliquées aux visages [45].

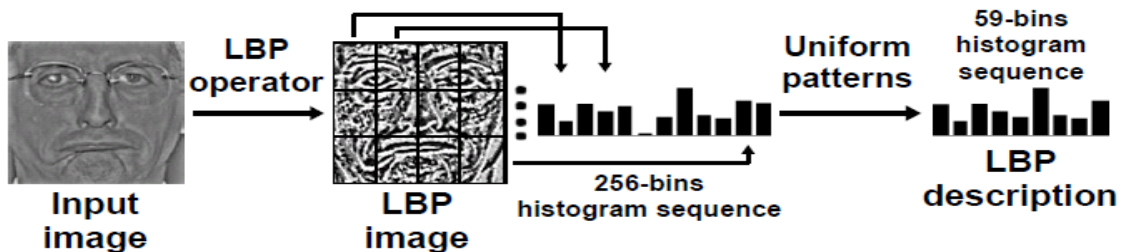


Fig.2.5- Représentation d'un visage par les histogrammes du code LBP [43].

L'avantage des méthodes locales de reconnaissance de visages est qu'elles peuvent modéliser facilement les variations de pose, d'illumination ou encore d'expressions que peut subir un visage. Cependant, elles nécessitent souvent le placement manuel de nombreux points d'intérêts pour une bonne précision, et sont donc lourdes à mettre en œuvre.

2.2.3 Les approches hybrides

Plusieurs techniques peuvent parfois s'appliquer afin de résoudre un problème de reconnaissance des formes. Chacune d'entre elles possède évidemment ses points forts et ses points faibles qui, dans la majorité des cas, dépendent des situations (pose, éclairage, expressions faciales, ...etc.). Il est par ailleurs possible d'utiliser une combinaison de classificateurs basés sur des techniques variées dans le but d'unir les forces de chacun et ainsi pallier à leurs faiblesses.

Les méthodes hybrides résultent de l'association des méthodes locales et globales. Elles combinent la détection de caractéristiques locales avec l'extraction de caractéristiques globales. Ces techniques essaient finalement de tirer parti des avantages des deux types de méthodes citées plus haut. Dans cette partie nous allons citer quelques méthodes hybrides connus dans le domaine de reconnaissance faciale. L'approche appelée Analyse en Composantes Locales (LCA pour Local Component Analysis) a été proposée par Penev et Atick [46]. Plusieurs analyses en composantes principales sont réalisées pour extraire différentes caractéristiques locales. Celles-ci sont ensuite combinées et une procédure

minimisant l'erreur de reconstruction avec une contrainte parcimonieuse permet de réaliser la reconnaissance. Perlibakas présente plus récemment l'algorithme LogGabor PCA dans [47]. Une convolution par des ondelettes de Gabor orientées est réalisée autour de certains points caractéristiques du visage. Les vecteurs ainsi créés contiennent à la fois la localisation ainsi que les amplitudes des énergies locales. Une Analyse en Composantes Principales est ensuite réalisée afin de réduire la dimension de ces vecteurs. L'approche proposée par Cootes et al. [48] met en œuvre les Modèles Actifs d'Apparence (AAM pour Active Appearance Models). Cette méthode modélise indépendamment la forme et la texture d'un visage en appliquant une analyse en composantes principales. Les vecteurs obtenus pour la forme et la texture sont ensuite utilisés pour la reconnaissance et le nouveau visage qui doit être identifié est adapté au modèle par un processus d'optimisation itératif. Les paramètres de forme et de texture obtenus sont ensuite comparés à ceux de la base. Latinis et al. [49] appliquent cette méthode pour la première fois à la reconnaissance de visages. Edwards et al. [50] proposent des méthodes basées sur les AAMs pour la reconnaissance faciale.

Nous avons présenté dans cette partie les principales approches utilisées dans la littérature pour la reconnaissance faciale automatique. Les méthodes peuvent principalement se classer en deux catégories : les méthodes globales et les méthodes locales (les méthodes hybrides essayant de faire un lien entre ces types d'approches).

2.3 Systèmes biométriques existants pour la reconnaissance des visages

Les systèmes biométriques de reconnaissance des visages se trouvent être de plus en plus répandus. Voici cités ci-dessous quelques systèmes mis en œuvre ainsi que les différents domaines d'utilisation de cette technologie en pleine expansion. Le 14 octobre 1998, le *Borough de Newham* de Londres met en service un système qui a diminué le nombre de crimes et délits de 10 % en 6 mois, grâce à l'utilisation du logiciel de reconnaissance de visage appelé « Mandrake » [2]. Le système alertait les opérateurs de caméra dès qu'il y avait 80 % de concordances entre l'image préalablement numérisée d'un délinquant et ce que capturaient les caméras. Les systèmes de reconnaissance des visages sont déployés dans le transport aérien, le système « SmartGate » par exemple a été mis en œuvre afin d'effectuer une vérification automatique de l'identité pour l'équipage d'Aéronef traversant la frontière de l'Australie. Ce dernier effectue une comparaison entre le visage d'une personne à sa photographie de passeport [51].

En Janvier 2002, « Visage Technology » le fournisseur de technologie et services de reconnaissance des visages a annoncé l'installation du premier système de reconnaissance des visages en Floride dans l'aéroport international St. Petersburg- Clearwater [52]. Cette technologie est aussi utilisée afin d'identifier les personnes recherchées en comparant les photos des passeports avec une base de personnes recherchées. Il existe encore toute une panoplie d'utilisations de la reconnaissance des visages. Aujourd'hui Les ATM's (Automatic Teller Machine : Distributeurs automatiques de billet mis en services par les banques ou autres institutions financières) identifient les utilisateurs non grâce à leurs numéros de carte bancaire, mais en se référant en plus à leur visage. En effet, l'ATM capture une image d'un visage et compare celle-ci avec la photo de la base afin de confirmer son identité. Plusieurs entreprises ont orienté leurs activités vers cette technologie, on retrouve par exemple L'Entreprise « Widget » qui a mis au point le système « Snappy Face » qui permet d'identifier le visage du propriétaire de l'ordinateur pour sécuriser son accès grâce à une webcam ou encore « Titanium Technology Entreprise » basée à Pékin qui a développé un logiciel de reconnaissance automatique de visages pour la surveillance (Automatic Face Recognition Systems ou AFRS) nommé « ProFacer iDVR » [52].

L'engouement pour les systèmes biométriques a connu un grand essor au début des années 2000. Les dépôts de brevet dans ce domaine se sont multipliés et des systèmes applicatifs à grande échelle ont été mis en place, comme le système utilisé lors des élections présidentielles au Venezuela en 2004 ou le système US Visit, devenu opérationnel depuis 2004. Ces systèmes restent basés sur les empreintes digitales, et ceci est dû aux évolutions techniques des algorithmes pour la biométrie basée sur les empreintes digitales.

2.4 Conclusion

Dans ce chapitre, on a passé en revue les principales techniques de reconnaissance de visages, puis nous avons énoncé quelques systèmes biométrique commercialisés sur le marché. Cet engouement pour les systèmes de reconnaissance des visages est justifié par les nombreux avantages de cette approche. En effet cette technologie est peu couteuse, peu encombrante, elle est de surcroit peu contraignant pour les usagers. Ce chapitre n'a pas pour finalité de décrire tous les algorithmes de reconnaissance faciale, mais nous nous sommes contentés de présenter les algorithmes qui ont introduit une nette évolution dans le domaine de biométrie, permettant ainsi une véritable amélioration des performances.

CHAPITRE : 3

Apprentissage automatique et réseaux de neurones convolutifs

3.1	Introduction
3.2	Apprentissage automatique
3.2.1	En quoi consiste l'apprentissage automatique ?
3.2.2	Définition
3.2.3	Intelligence artificielle
3.2.4	Pourquoi l'apprentissage automatique ?
3.2.5	Les données d'apprentissage
3.2.6	L'histoire des systèmes d'apprentissage automatique
3.3	Les types d'apprentissage
3.3.1	Apprentissage supervisé
3.3.2	Apprentissage non supervisé
3.3.3	Apprentissage semi-supervisé
3.3.4	Apprentissage par renforcement
3.3.5	Apprentissage par transfert
3.4	Les algorithmes d'apprentissage
3.5	Apprentissage Profond (Deep Learning)
3.6	Réseaux de neurones artificiels : les bases
3.6.1	Neurone formel et perceptron
3.6.2	Les fonctions d'activation
3.6.3	Le perceptron multicouche
3.6.4	L'algorithme de la rétro-propagation du gradient (back-propagation)
3.6.5	Les différents types de réseaux de neurones
3.6.6	Classification et reconnaissance des visages par réseaux de neurones
3.7	Les réseaux de neurones profonds
3.8	Réseaux de neurones convolutifs
3.8.1	Différents modules d'un réseau de neurones convolutif
3.8.2	Outils d'optimisation des réseaux convolutifs
3.9	Utilisation de modèles pré-entraînés pour l'apprentissage par transfert
3.10	Les architectures neuronales convolutifs
3.11	Conclusion

3.1 Introduction

Après avoir présenté un état de l'art des différentes techniques utilisées particulièrement en reconnaissance de visage, cette partie de la thèse a pour objectif de présenter les réseaux de neurones convolutifs profonds (Deep Convolutional Neural Networks CNN) utilisés tout au long de cette thèse, en commençant par une introduction générale à l'apprentissage automatique qui a connu un intérêt exceptionnel ces dernières années, ainsi que les concepts de base qui peuvent aider à mieux comprendre le fonctionnement de l'apprentissage en profondeur.

Ce chapitre décrit quelques concepts de base qui peuvent aider à mieux comprendre le fonctionnement de l'apprentissage en profondeur. Tout d'abord, nous rappellerons rapidement les objectifs de l'apprentissage automatique. Après on commence à détailler la structure de base des neurones artificiels et l'intérêt des réseaux de neurones par rapport aux approches classiques d'apprentissage. Ensuite, les réseaux de neurones profonds sont introduits ainsi que leurs diverses architectures. Enfin, une description des réseaux de neurones de convolutifs est fournie avec ses différents modules. Ce dernier est largement utilisé dans partie expérimentales de cette thèse.

3.2 Apprentissage automatique

3.2.1 En quoi consiste l'apprentissage automatique ?

De manière générale, un programme informatique tente de résoudre un problème pour lequel nous avons la solution. Par exemple : (calculer la moyenne générale des étudiants, classer les étudiants selon leur moyenne...). Pour certains problèmes, nous ne connaissons pas de solution exacte et donc nous ne pouvons pas écrire de programme informatique. Par exemple : reconnaître automatiquement des chiffres écrits à la main à partir d'une image scannée, déterminer automatiquement une typologie des clients d'une banque, jouer automatiquement aux échecs contre un humain ou un autre programme. En revanche, pour ces problèmes il est facile d'avoir une base de données regroupant de nombreuses instances du problème considéré. L'apprentissage automatique consiste alors à programmer des algorithmes permettant d'apprendre automatiquement à partir des données et d'expériences passées, un algorithme cherchant à résoudre au mieux un problème considéré [53]. Pour cette raison, l'apprentissage automatique facilite l'utilisation des ordinateurs dans la construction de modèles à partir de données d'échantillonnage afin d'automatiser les processus de prise de décision en fonction des données saisies [54].

3.2.2 Définition

L'apprentissage automatique (Machine Learning en anglais) que l'on appelle aussi l'apprentissage machine est une méthode utilisée en intelligence artificielle. Il s'agit d'algorithmes (procédures traduites en langages informatiques) qui analysent un ensemble de données afin de déduire des règles qui constituent de nouvelles connaissances permettant d'analyser de nouvelles situations [53]. L'apprentissage est un concept pour définir l'acquisition de connaissance et réutiliser ces nouvelles connaissances. Pour nous les humains, notre apprentissage se fait tout au long de notre vie. Nous apprenons à partir de perception de l'environnement avec les cinq sens, les expériences de la vie, de répétition des événements avec la mémoire, et de notre jugement (libre arbitre ou intelligence). Pour les machines, comme elles ne sont pas dotées de sens ni de jugement dynamique, elles obéissent aux instructions dans un programme avec des données d'entrée et une donnée sortie ou réponse du programme.

Dans la pratique, le mot entraînement (Training, en anglais) est souvent synonyme de : apprentissage. Ainsi, en science cognitive, l'apprentissage est défini comme étant la capacité à améliorer les performances au fur et à mesure de l'exercice d'une activité. Par exemple, un joueur de jeu d'échec, assimile (par expérience, s'entraîne) et raisonne (ceci lui procure une certaine intelligence ou puissance de raisonnement pour qu'il puisse progresser) c'est le cas pour un algorithme intelligent. Dans le contexte du système d'apprentissage automatique, il y a toujours des termes connexes qui doivent être compris afin de mieux appréhender le principe du Machine Learning [53].

3.2.3 Intelligence artificielle

Qu'est-ce que l'intelligence ? Un élève « intelligent » à l'école, c'est celui qui a de bonnes notes. C'est celui qui a bien appris sa leçon et qui est capable de répondre à des questions comme par exemple " l'année de l'indépendance de l'Algérie ? ". L'élève ayant appris sa leçon répondrait en 1962, c'est bien mais si on creuse en lui demandant quelle armée était opposée à quelle armée et les événements qui menaient à cette indépendance ? Il n'aurait pas la réponse car il aura appris les dates historiques par cœur sans chercher à savoir ce qu'il y a derrière chaque date. Ce cas présent est le défaut de l'intelligence comme on le définit souvent c'est apprendre sans pour autant comprendre ce qui est appris [55]. Finalement, on confond connaissance et intelligence alors que la connaissance est une toute petite partie de l'intelligence. Lorsque l'intelligence artificielle fit son apparition, les ingénieurs qui cherchaient à créer des « machines qui pensent » procédaient comme le cas de l'école vu

précédemment : On essaye de leur faire apprendre tout ce qu'on peut. On appelle cette méthode l'apprentissage automatique ou le machine learning [54].

La recherche sur l'intelligence artificielle (IA) tente de créer des machines capables d'agir comme des êtres humains, en effet les ordinateurs et les robots sont censés analyser leur environnement et ainsi prendre la meilleure décision possible. Ils doivent donc se comporter de manière intelligente selon nos normes. Mais cela ouvre une problématique, quels critères devons-nous utiliser pour juger de notre propre intelligence ? [55]. Aujourd'hui, l'IA ne peut simuler l'être humain en entier (notamment l'intelligence émotionnelle). Au lieu de cela, des aspects partiels sont isolés afin de faire face à des tâches spécifiques, précises. C'est ce que l'on nomme communément l'intelligence artificielle faible (IA faible) [55].

3.2.4 Pourquoi l'apprentissage automatique ?

L'apprentissage automatique consiste à utiliser des ordinateurs pour optimiser un modèle de traitement de l'information selon certains critères de performance à partir d'observations, que ce soit des données des exemples ou des expériences passées. Lorsque l'on connaît le bon modèle de traitement à utiliser, on n'aura pas besoin de faire de l'apprentissage [56]. Il est peut-être utile lorsque on n'a pas d'expertise sur le problème (ex. robot navigant sur Mars), on a une expertise, mais on ne sait pas comment l'expliquer (ex. reconnaissance de visages), ou lorsque les solutions au problème changent dans le temps (ex. routage de paquets). L'apprentissage automatique est à la croisée de plusieurs disciplines [54] [56] :

- *Les statistiques* : pour l'inférence de modèles à partir d'échantillons.
- *Les probabilités* : pour modéliser l'aspect aléatoire inhérent aux données et au problème d'apprentissage.
- *L'intelligence artificielle* : pour étudier les tâches simples de reconnaissance de formes que font les humains (comme la reconnaissance de chiffres par exemple).
- *L'optimisation* : pour optimiser un critère de performance, soit d'estimer des paramètres d'un modèle, ou de déterminer la meilleure décision à prendre.
- *L'informatique* : pour construire des algorithmes et des représentations efficaces pour construire et évaluer les modèles.
- *Ingénierie* : résoudre des problèmes sans devoir spécifier ou spécialiser manuellement les modèles.

3.2.5 Les données d'apprentissage

Les données d'apprentissage sont souvent réparties en 3 catégories [59] :

- *L'ensemble d'apprentissage* ou population d'entraînement : constitue l'ensemble des candidats ou exemples (images, attributs, vecteurs, ...) utilisés pour générer le modèle d'apprentissage.
- *L'ensemble de test* : est constitué des candidats sur lesquels sera appliqué le modèle d'apprentissage (pour tester et corriger l'algorithme).
- *L'ensemble de validation* : peut être utilisé lors de l'apprentissage (comme sous population de l'ensemble d'apprentissage) afin de valider (intégrer) le modèle et d'éviter le sur-apprentissage.

Selon les domaines, les connaissances ou données d'apprentissage peuvent être de diverses formes : images, mots, phrases, variables ou attributs, des vecteurs ...etc.

3.2.6 L'histoire des systèmes d'apprentissage automatique

Depuis l'antiquité, le sujet des machines pensantes préoccupe les esprits. Ce concept est la base des pensées pour ce qui deviendra ensuite l'intelligence artificielle, ainsi qu'une de ses sous-branches : l'apprentissage automatique. La relation de l'être humain à la machine pensante a toujours oscillé entre crainte et fascination. Cependant, les réels progrès de l'apprentissage machine ne commencèrent pas avant les années 50, à une époque où les ordinateurs n'en étaient encore qu'à leurs débuts et où l'intelligence artificielle ne pouvait que faire rêver [54]. Au cours des deux siècles précédents, des théoriciens avaient déjà jeté les bases de la recherche, mais il faut attendre les travaux de Turing « L'ordinateur et l'intelligence » en 1950 [57] pour parler concrètement d'apprentissage automatique des machines.

En 1950, Turing a développé une proposition de test d'intelligence artificielle : le test de Turing. Il s'agit d'une sorte de jeu dans lequel un ordinateur prétend être un humain, il imite la conversation humaine. Deux ans plus tard, Arthur Samuel [58] informaticien américain pionnier dans le secteur de l'intelligence artificielle, est le premier à faire usage de l'expression machine learning, a développé un ordinateur qui pouvait jouer aux dames tout en s'améliorant à chaque partie. Le programme avait donc la capacité d'apprendre. Enfin, en 1957, Frank Rosenblatt développa le Perceptron, un premier algorithme d'apprentissage, il s'agit d'un réseau neuronal artificiel. Dès lors, les scientifiques ont commencé à confier à leurs ordinateurs des épreuves de plus en plus complexes, les machines les maîtrisant plus ou moins bien. IBM a développé Watson, un programme informatique qui possède un

immense référentiel de connaissances et qui peut répondre aux questions posées en langage naturel. En 1997 un autre ordinateur d'IBM : le Deep Blue. Est le premier à vaincre le champion mondial d'échecs Garry Kasparov dans une compétition d'échec [55].

Durant les années suivantes, les applications de l'apprentissage automatique médiatisées se succèdent bien plus rapidement qu'auparavant. En 2012, un réseau neuronal développé par Google parvient à reconnaître des visages humains ainsi que des chats dans des vidéos sur YouTube avec un taux de réussite de 97 pourcent [58]. En 2015, une nouvelle étape importante est atteinte lorsque l'ordinateur « AlphaGo » de Google gagne contre un des meilleurs joueurs au jeu de Go, jeu de plateau considéré comme le plus dur du monde. En 2016, un système d'intelligence artificielle à base d'apprentissage automatique nommé LipNet [59] parvient à lire sur les lèvres avec un grand taux de succès.

3.3 Les types d'apprentissage

Une fois les données acquises du terrain ou du réel (images, vidéos, fichiers, base de données) sous forme de données brutes, il faut les analyser et en extraire de l'information sous forme de nouvelles connaissances, de descripteurs ou de vecteur de caractéristiques. Lorsque les données sont conséquentes leur traitement peut se faire par différentes méthodes [59].

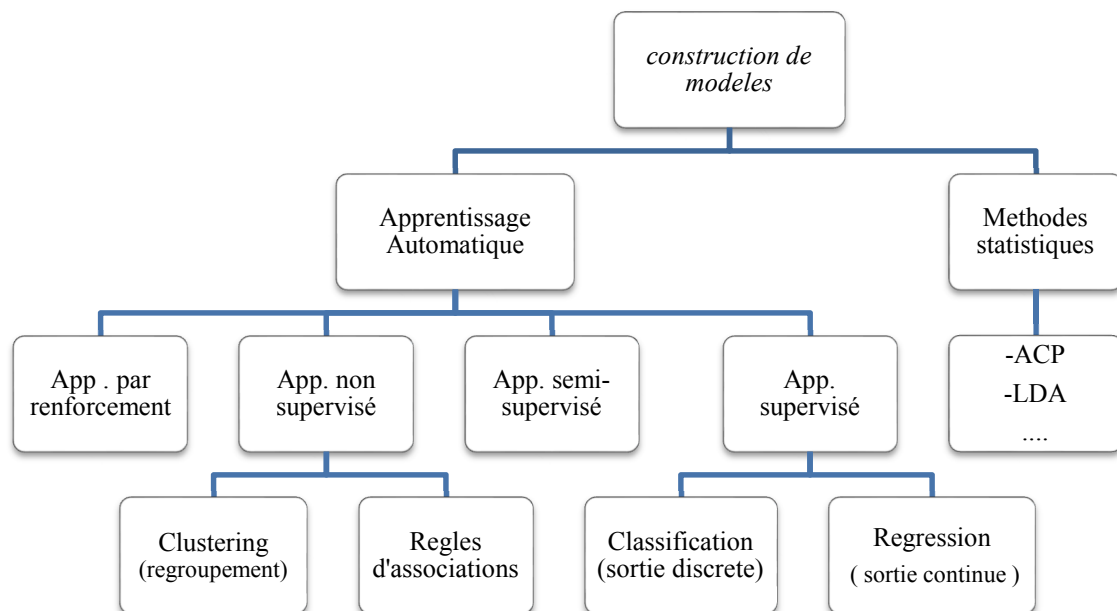


Fig.3.1- Les différentes techniques d'apprentissage automatique [59].

Les deux principales composantes de la construction de modèles d'apprentissage et de classification sont des techniques issues de la statistique ou bien des techniques d'apprentissage automatique. Nous nous intéressons plus spécialement aux techniques d'apprentissage automatique.

Dans l'apprentissage automatique, les tâches sont généralement classées en plusieurs catégories. Ces derniers sont basés sur la façon dont l'apprentissage est reçu ou comment le feedback sur l'apprentissage est donné au système développé. Deux des méthodes d'apprentissage automatique les plus largement adoptées sont l'apprentissage supervisé qui forme des algorithmes basés sur des données d'entrée et de sortie étiquetées par l'homme et l'apprentissage non supervisé qui ne fournit pas à l'algorithme des données étiquetées pour lui permettre de trouver une structure et de découvrir une logique sur les données d'entrées des systèmes d'apprentissage [59].

3.3.1 Apprentissage supervisé

Le type d'apprentissage automatique le plus utilisé est l'apprentissage dit "supervisé" qui permet à la machine d'apprendre ses paramètres en utilisant une base de données annotée. Dans ce cas de figure l'algorithme sera guidé sur la voie d'apprentissage en lui fournissant des exemples après les avoir préalablement étiquetés sur des résultats attendus. L'intelligence artificielle apprend alors de chaque exemple en ajustant ses paramètres de façon à diminuer l'écart entre le résultat obtenu et le résultat attendu. La marge d'erreur se réduit ainsi au fil des entraînements, avec pour but, d'être capable de généraliser son apprentissage à de nouveaux cas. Par exemple, dans le cadre de la classification d'images, on fournit le couple (entrée, sortie) et on modifie les paramètres en fonction de l'erreur entre la sortie désirée et la sortie obtenue. Un modèle entraîné grâce à l'apprentissage supervisé prédit le type d'objet (sa classe) présent dans l'image [60].

3.3.2 Apprentissage non supervisé

Dans ce type d'apprentissage on dispose d'un ensemble d'objets sans aucune valeur cible associée, il faut apprendre un modèle capable d'extraire les régularités présentes au sein des objets pour mieux visualiser ou appréhender la structure de l'ensemble des données [53]. Parfois on l'assimile au clustering puisqu'il ne demande aucun étiquetage préalable des données. Le but est que le modèle réussisse à regrouper les observations disponibles en catégories par lui-même.

3.3.3 Apprentissage semi-supervisé

Pour ce type on dispose d'un petit ensemble d'objets avec pour chacun une valeur cible associée et d'un plus grand ensemble d'objets sans valeur cible, il faut tirer profit à la fois des données avec et sans valeurs cibles pour résoudre des tâches d'apprentissage supervisé ou non supervisé [53].

3.3.4 Apprentissage par renforcement

Les données en entrée sont les mêmes que pour l'apprentissage supervisé, cependant l'apprentissage est guidé par l'environnement sous la forme de récompenses ou de pénalités données en fonction de l'erreur commise lors de l'apprentissage [59].

3.3.5 Apprentissage par transfert

L'apprentissage par transfert peut être vu comme la capacité d'un système à reconnaître et appliquer des connaissances et des compétences, apprises à partir de tâches antérieures, sur de nouvelles tâches ou domaines partageant des similitudes [58].

3.4 Les algorithmes d'apprentissage

Les approches de l'apprentissage automatique sont continuellement développées. Elles abordent plusieurs domaines, notamment les statistiques, la reconnaissance de formes, la biométrie, l'intelligence artificielle, le traitement du signal, le contrôle et l'exploration de données [54]. Suite à cela de nombreuses méthodes liées à ce domaine ont été proposées, le classifieur naïf de Bayes, l'algorithme des k-moyennes (k-means), méthodes multivariées, techniques de réduction de dimensionnalité, clustering, les k plus proches voisins (K-Nearest Neighbors (KNN)), arbres de décision, les réseaux de neurones, les support vecteur machines, modèles de Markov cachés, Les forêts aléatoires (random forest), et bien d'autres algorithmes de comparaison et de classification [55] [56] [59].

Il est à noter l'introduction de l'apprentissage automatique fait ici n'a pas comme objectif d'être parfaitement général. Effectivement, sa simplicité ne permet pas de bien couvrir toutes les notions liées à ce domaine. Cependant, ce choix spécifique permet de traiter plus directement des types d'apprentissage mis en cause dans les travaux de cette thèse, et ainsi devrait faciliter leur compréhension par le lecteur. Pour nos besoins, nous allons passer l'une des approches populaires utilisées dans l'apprentissage automatique, l'apprentissage profond et les réseaux de neurones convolutifs [60]. Ces derniers sont largement utilisés dans les expériences de cette thèse.

3.5 Apprentissage Profond (Deep Learning)

L'apprentissage profond (Deep Learning en anglais) est un type d'intelligence artificielle dérivé de l'apprentissage automatique (machine learning) où la machine est capable d'apprendre par elle-même, contrairement à la programmation où elle se contente d'exécuter à la lettre des règles prédéterminées. Ce type de techniques tentent de modéliser avec un haut niveau d'abstraction des données grâce à des architectures articulées de différentes transformations non linéaires.

Le deep learning s'appuie sur un réseau de neurones artificiels s'inspirant du cerveau humain. Ce réseau est composé de dizaines voire de centaines de couches de neurones, chacune recevant et interprétant les informations de la couche précédente. Introduits pour la première fois en 1943 par McCulloch en 1943 [60], connectés entre eux et modélisant le fonctionnement des neurones biologiques, mais leur étude a stagné à partir de la fin des années 90 jusqu'à une explosion de leurs popularités depuis 2012. Ils ont depuis surpassé beaucoup de méthodes dans les tâches de vision par ordinateur (classification, détection, segmentation...). Cet engouement récent, notamment dans le domaine de la reconnaissance de formes, s'explique de plusieurs manières. Premièrement, d'énormes bases de données publiques annotées sont actuellement disponibles, par exemple la base ImageNet [61] contient environ 14 millions d'images correspondant à 22 000 classes d'objets. La seconde raison ayant relancé l'étude des réseaux de neurones est la capacité des machines modernes à effectuer beaucoup de calculs dans un temps raisonnable notamment grâce à l'utilisation des cartes graphiques (GPU). Cela permet de construire des réseaux de neurones de plus en plus complexes et performants [60]. Enfin, la multiplication des framework de Deep Learning (Torch [62], Caffe [63], Tensorflow [64],...etc.) permet une démocratisation de ces méthodes d'apprentissage automatique.

La tâche de la reconnaissance des traits caractéristiques à partir des visages a été un sujet de recherche répandu avec de nombreuses applications déjà en usage commercial. Presque toutes ces méthodes suivent une approche basée sur la modélisation du visage pour résoudre le problème de reconnaissance faciale. Compte tenu du récent succès d'apprentissage en profondeur, il est intéressant de citer les principaux avantages des approches basées sur l'apprentissage profond par rapport à ceux basées sur la modélisation du visage par d'autres approches classiques [65] :

- Les caractéristiques sont désignées par le réseau de neurones profonds lui-même et ne nécessiteront pas l'extraction manuelle des caractéristiques / points de repère, comme dans une approche basée sur la modélisation du visage.
- Outre les caractéristiques évidentes telles que les lèvres, la position des sourcils, etc, les caractéristiques les plus représentatives, qui peuvent ne pas être présentes dans une modélisation de visage typique basée sur une approche globale, peut être identifiée par le réseau profond.
- Les réseaux profonds sont intrinsèquement parallélisables et ils se comportent bien dans un environnement de calcul parallèle. Par conséquent, leur apprentissage et leur exécution peuvent être beaucoup plus rapides avec l'utilisation de plates-formes telles que les processeurs graphique GPU (Graphics Processing Unit) et les processeurs multitâches.
- Autre différence entre les algorithmes d'apprentissages traditionnels et ceux de deep learning, c'est l'étape de l'extraction de caractéristiques. Dans les approches traditionnelles l'extraction de caractéristiques est faite manuellement, c'est une étape difficile, coûteuse en temps et requiert un spécialiste en la matière, alors qu'en Deep Learning cette étape est exécutée automatiquement par l'algorithme [65].

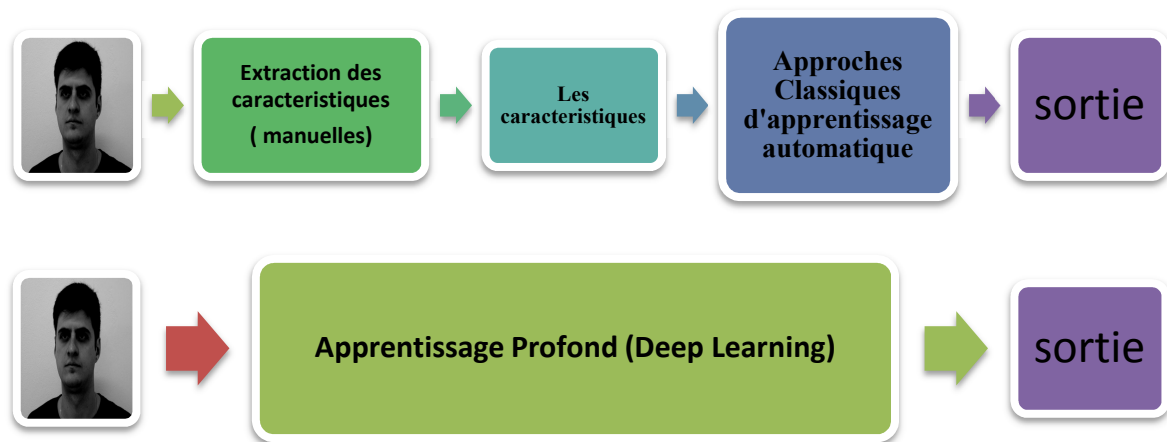


Fig.3.2- Le procédé d'apprentissage classique comparé à celui du Deep Learning.

3.6 Réseaux de neurones artificiels : les bases

3.6.1 Neurone formel et perceptron

Un réseau de neurones artificiel est un modèle de calcul dont la conception est très schématiquement inspirée du fonctionnement des neurones biologiques. Il s'agit d'un des plus puissants classifieurs d'aujourd'hui [66]. Il est modélisé mathématiquement par un réseau dans une architecture, plus ou moins complexe, hiérarchique sous forme de couches

dont les nœuds élémentaires sont appelés neurones. Il est composé généralement de trois types de couches : la couche d'entrée (input layer), couche(s) cachée(s) (hidden layer) et la couche de sortie (output layer). La couche d'entrée est composée de neurones qui correspondent aux caractéristiques des données d'entrée représentées par une grille multidimensionnelle (par exemple la matrice de pixels de l'image ou la forme vectorielle de la donnée). La couche de sortie représente les résultats de la tâche assignée au réseau.

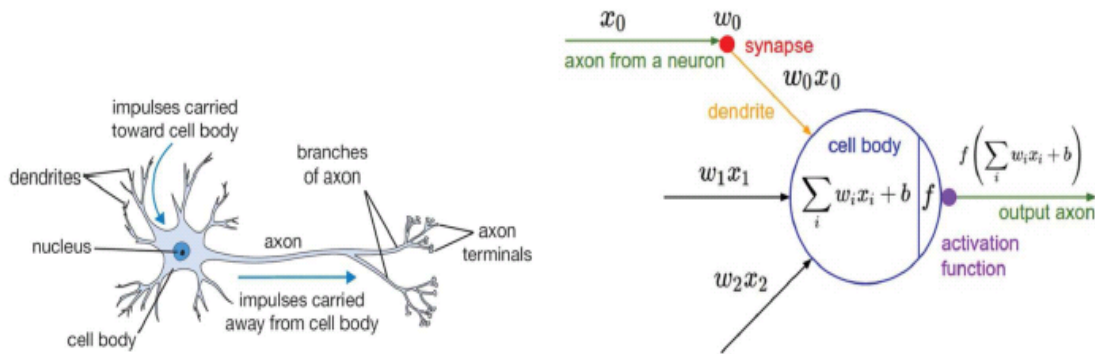


Fig.3.3- Comparaison entre un neurone biologique et le modèle mathématique d'un neurone artificiel (formel) [66].

Le neurone formel (ou artificiel), initialement introduit par McCulloch et Pitts en 1943 [66], est une modélisation mathématique du neurone biologique. Généralement le réseau de neurone reçoit les données X_i sur la couche d'entrée, chacun de ces données à un poids synaptiques W_i mesure l'importance de chaque entrée. Dans la couche cachées ou couche de traitement le réseau applique une fonction appelé fonction d'activation sur la somme pondérée de $W_i X_i$. Le résultat exprime la sortie y de ce réseau (voir fig3.4).

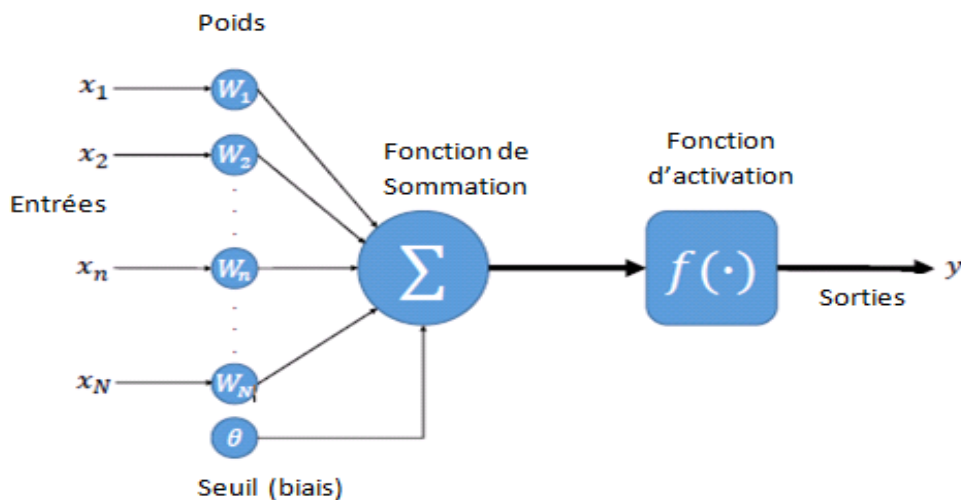


Fig.3.4- Représentation d'un réseau de neurone perceptron [66].

3.6.2 Les fonctions d'activation

Dans le domaine des réseaux de neurones artificiels, la fonction d'activation est une fonction mathématique appliquée à un signal en sortie d'un neurone artificiel. Le terme de "fonction d'activation" vient de l'équivalent biologique "potentiel d'activation", seuil de stimulation qui, une fois atteint entraîne une réponse du neurone. La fonction d'activation est souvent une fonction non-linéaire [60] [67]. L'utilisation de fonctions d'activation non-linéaires est tout simplement indispensable pour la simple et bonne raison que les fonctions linéaires ne fonctionnent qu'avec une seule couche de neurone. Le neurone calcule le produit scalaire entre les signaux d'entrée et le vecteur poids en ajoutant un biais. Le résultat est appliqué dans la fonction d'activation. La fonction d'activation sert à introduire une opération de non-linéarité après l'opération de produit scalaire. Cette non-linéarité permet d'avoir différentes variations de l'état interne sur un objet de la même classe [66]. Mais le choix d'une fonction d'activation se révèle être un élément constitutif important des réseaux de neurones. Cette fonction peut profondément influencer sur la performance du réseau. Il est donc important de bien choisir de type de fonction d'activation dans un réseau de neurones. Au fil des ans, diverses fonctions ont été utilisées. Dans cette partie seules les principales ont été citées. Il existe plusieurs fonctions d'activation à choisir. La fonction sigmoïde (logistique), la tangente hyperbolique (Tanh) et la fonction d'activation unité de rectification linéaire (ReLU) sont les plus couramment utilisées [66].

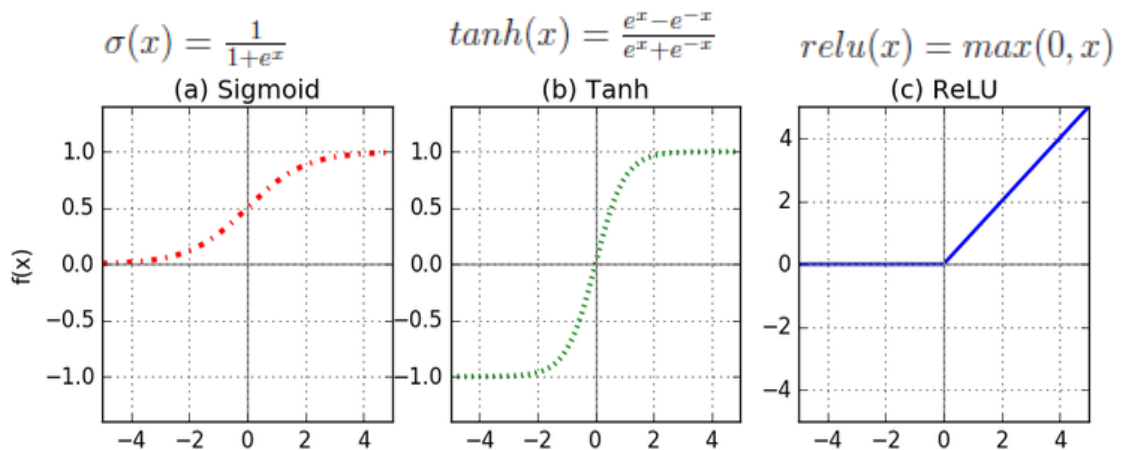


Fig.3.5- Fonctions d'activation couramment utilisées.

Les modèles Tanh, sigmoïde, ReLU, sont bien adaptés aux algorithmes d'optimisation comme la descente stochastique de gradient car elles sont différentiables. Mais actuellement c'est le ReLU qui est le plus utilisé dans les réseaux de neurones profonds [66]. À cause d'un problème que l'on appelle la disparition du gradient "vanishing gradient". Comme les

autres fonctions d'activation (sigmoïde, tanh) produisent des valeurs entre 0 et 1, le gradient va se tendre vers 0 au fur et à mesure que l'on a des réseaux profonds (plusieurs couches). Le ReLU permet d'amortir le "vanishing gradient" car sa valeur de sortie est invariante. De plus le calcul des exponentielles présent dans les fonctions sigmoïde et tanh est lent par rapport au calcul du max dans ReLU qui est donc préféré [66] [67].

Pour le choix d'une fonction d'activation à la couche de sortie, pour un problème de classification, il vaut mieux utiliser des fonctions d'activation non linéaires ('logistique' ou 'softmax') à la couche de sortie [60] [66] [67]. La fonction d'activation de softmax est habituellement utilisée dans la dernière couche du réseau. Il s'agit d'une généralisation de la fonction logistique et n'est qu'un autre nom pour un modèle de classification multinomiale quand on suppose qu'il n'existe aucune hiérarchie parmi les classes. La fonction softmax est agréable car elle donne une approximation de la probabilité qu'une classe soit correcte. L'approche la plus simple consiste à simplement choisir la classe avec la probabilité la plus élevée et ignorer le reste. Mais étant donné qu'il s'agit d'une fonction probabiliste, On peut ainsi interpréter les sorties du réseau comme des probabilités que l'entrée appartienne à chacune des classes.

3.6.3 Le perceptron multicouche

En 1958, Rosenblatt [66] introduit un algorithme d'apprentissage automatique basé sur le neurone artificiel appelé perceptron, c'est le premier réseau neuronal bien que constitué d'un seul neurone. Le perceptron est un modèle simple et adapté pour des problèmes linéairement séparables. Pour résoudre des problématiques plus complexes comme par exemple le problème XOR (ou exclusif), le perceptron simple ne suffit plus. De ce fait, un réseau de neurones plus complexe a été introduit, le perceptron multicouche (PMC).

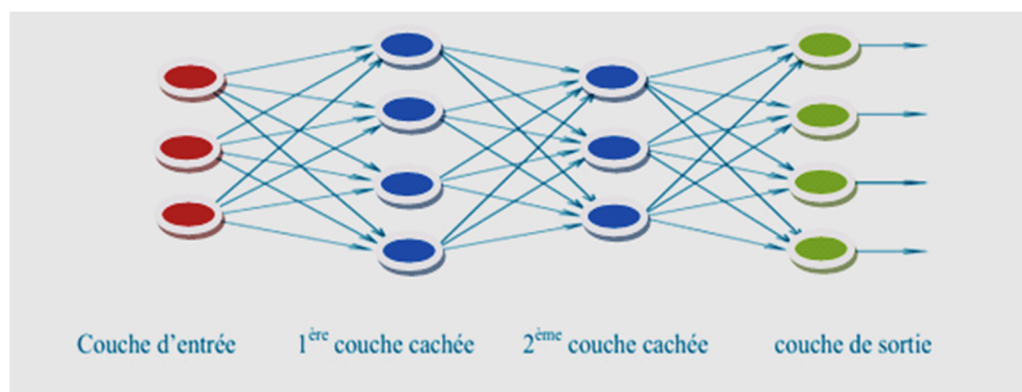


Fig.3.6- Architecture d'un perceptron multicouche. (Constitué de deux couches cachées. Chaque cercle représente un neurone formel).

Le perceptron multicouche (en anglais MLP : Multi Layer Perceptron) est une amélioration du perceptron comprenant une ou plusieurs couches cachées qui font du réseau PMC un outil robuste pour des tâches complexes [67]. Il est largement utilisé pour la décision (classification) dans le domaine de reconnaissance faciale. Ces réseaux sont généralement des réseaux entièrement connectés. Les neurones de la première couche reçoivent le vecteur d'entrée, ils calculent leurs sorties qui sont transmises aux neurones de la seconde couche qui calculent eux même leurs sorties et ainsi de suite de couche en couche jusqu'à celle de décision [66] [67]. Dans un réseau PMC il n'y a aucune connexion entre les cellules d'une même couche. Les perceptrons multicouches sont utilisés avec apprentissage supervisé et aussi avec une technique basée sur l'algorithme de rétro-propagation (back-propagation) pour la correction de l'erreur.

3.6.4 L'algorithme de la rétro-propagation du gradient (back-propagation)

Les réseaux de neurones sont des outils fortement non-linéaires qui utilisent généralement des techniques itératives pour l'apprentissage [68]. IL existe différents algorithmes, mais les techniques les plus conseillées pour l'apprentissage des réseaux de neurones sont les algorithmes de gradient descendant avec taux d'apprentissage variable, Rétro-propagation résiliente, L'algorithme de Fletcher-Reeves, algorithme de Quasi-Newton et l'algorithme de Levenberg-Marquardt [68] [69]. Ces méthodes fournissent généralement de meilleurs résultats que les algorithmes plus classiques comme la descente de gradient mais ils sont également plus gourmands en termes de mémoire et nécessitent davantage de calculs. L'apprentissage se conduit avec une méthodologie précise. La première, est la plus simple, connue sous le nom de règle de Hebb [70], indique que lorsque deux neurones sont excités en même temps, il faut modifier les coefficients synaptiques pour renforcer cette excitation simultanée. La règle de Widrow-Hoff [69] [70] établie en 1960, indique que la modification de ses coefficients est proportionnelle à l'erreur entre le résultat souhaité, le résultat réel et aux valeurs d'entrée. L'algorithme de rétro-propagation du gradient suit la démarche de Widrow-Hoff pour définir une notion d'erreur sur un exemple puis calcule la contribution de cette erreur de chacun des poids synaptiques.

L'objectif de la fonction rétro-propagation est de minimiser l'erreur quadratique moyenne et de chercher à corriger les valeurs de poids du réseau. Le principe de la rétro-propagation consiste à présenter au réseau un vecteur d'entrées, de procéder au calcul de la sortie par propagation à travers les couches, de la couche d'entrée vers la couche de sortie en passant par les couches cachées. Cette sortie obtenue est comparée à la sortie désirée, une erreur est

alors obtenue. A partir de cette erreur, est calculé le gradient qui est à son tour, propagé de la couche de sortie vers la couche d'entrée, d'où le terme de rétro-propagation. Cela permet la modification des poids du réseau et donc l'apprentissage. L'opération est répétée pour chaque vecteur d'entrée et cela jusqu'à ce que le critère d'arrêt soit vérifié [69].

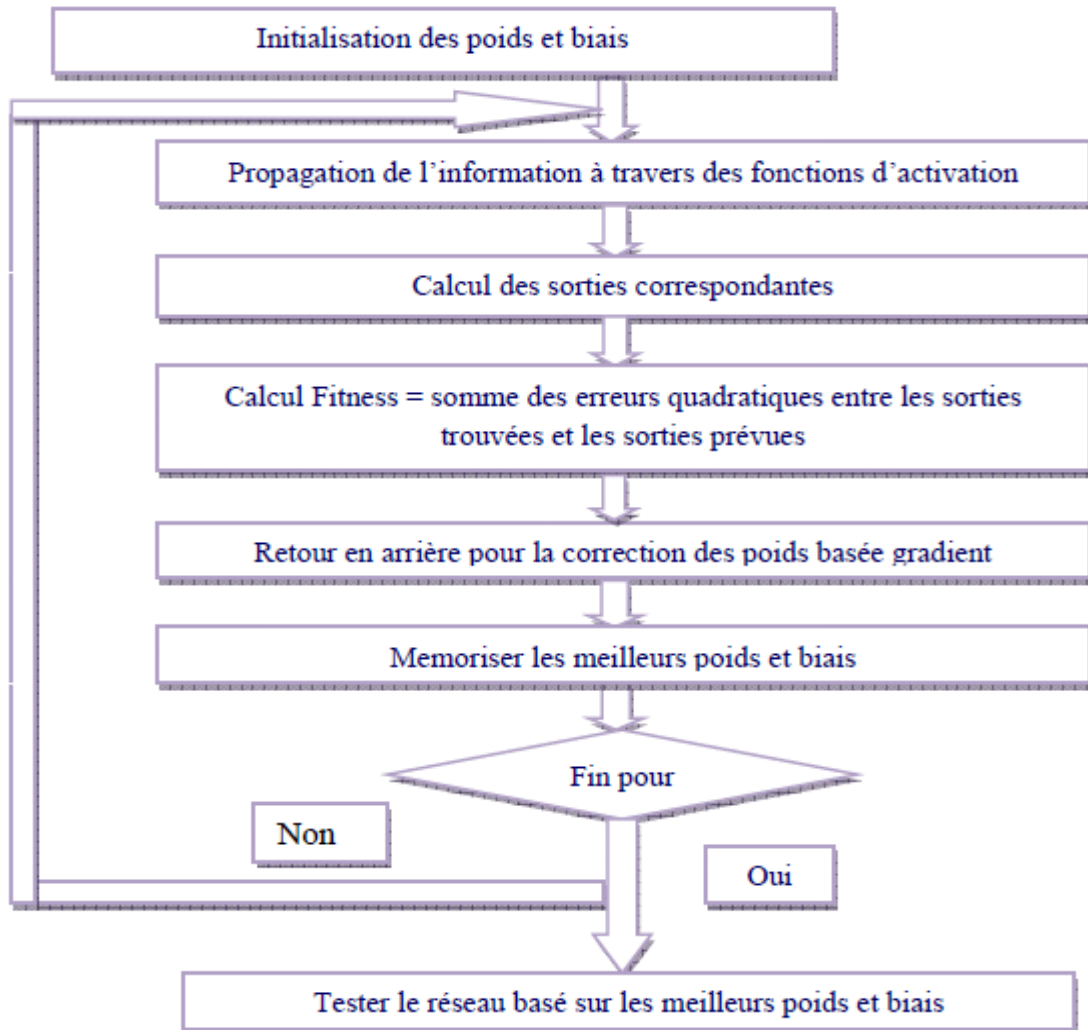


Fig.3.7- Le principe général de l'algorithme de la rétro-propagation de gradient.

3.6.5 Les différents types de réseaux de neurones

Il n'est pas possible d'énumérer l'ensemble des types de réseaux de neurones disponibles à ce jour. Cependant, à titre illustratif sont présentés quelques modèles les plus populaires. L'un des plus anciens réseaux qui a été introduit par Bernard Widrow au début des années 1960 [70] [71], l'Adaline (ADaptive LInear NEuron, ou ADaptive LInear Element), est très proche du modèle perceptron, à l'instar des autres neurones artificiels, l'adaline comprend un circuit qui calcule le niveau d'activation du neurone, suivi d'une fonction de sortie. Le premier étage est un filtre adaptatif qui ajuste des coefficients en fonction d'un

signal d'erreur entre la sortie et une valeur de référence, la fonction de sortie permet de transformer le filtre en classificateur binaire. Le madaline (Many ADALINES) qui l'a suivi peut-être considéré comme les premiers représentants de réseaux neuronaux d'inspiration non biologique. Ces réseaux sont souvent employés en traitement de signaux, notamment pour la réduction de bruit. Contrairement aux réseaux où les neurones sont modélisés de la façon la plus simple possible. Les réseaux de Kohonen recherche un modèle de neurone plus proche de la réalité, ils sont inspirés des observations biologiques du fonctionnement des systèmes nerveux de perception des mammifères. Ces réseaux décrivent en fait trois familles de réseaux de neurones [72], le Vector Quantization (VQ), introduit par Grossberg (1976) [71], la quantification vectorielle est une méthode d'apprentissage non supervisé. Elle permet de retrouver des groupes sur un ensemble de données, de façon relativement similaire à un k-means [73], que l'on préférera d'ailleurs généralement à un VQ si la simplicité d'implémentation n'est pas un élément majeur de la résolution du problème. Les réseaux Self Organizing Map (SOM) ou les cartes auto organisatrices sont issus des travaux de Fausett et Kohonen (1995) [72]. Ces réseaux sont très utilisés pour l'analyse de données. Ils permettent de cartographier en deux dimensions et de distinguer des groupes dans des ensembles de données. Les SOM sont encore largement utilisés mais les scientifiques préfèrent maintenant les Learning Vector Quantization (LVQ), ces réseaux reposent sur un apprentissage supervisé, ils ont été proposés par Kohonen (1988) [72]. Des trois types de réseaux présentés ici, le LVQ est la seule méthode qui soit réellement adaptée à la classification de données par recherche du plus proche voisin. On peut citer aussi les Réseaux de Hopfield [71], proposée par J. Hopfield en 1982. Ces réseaux sont des réseaux récursifs, un peu plus complexes que les perceptrons multicouches. Chaque cellule est connectée à toutes les autres et les changements de valeurs de cellules s'enchainent en cascade jusqu'à un état stable. Ces réseaux sont bien adaptés à la reconnaissance de formes. Un autre réseau très largement utilisé est la fonction de base radiale RBF (Radial Basis Function), Il s'agit d'un cas particulier d'un PMC [74]. Un RBF est constitué uniquement de trois couches, une couche d'entrée qui retransmet les entrées sans distorsion, une seule couche cachée et une couche de sortie qui contient une fonction linéaire. Chaque neurone élémentaire de la couche cachée, calcule la distance entre l'entrée et son centre qu'il passe ensuite dans un non linéarité réalisée par une fonction d'activation qui est généralement de type gaussien [67] [68] [74]. Il existe de diverses topologies possibles pour un nombre spécifique de neurones. Certains sont des représentations réalistes des réseaux de neurones

biologiques, et les autres ont des avantages pour résoudre des problèmes spécifiques en informatique [74] [75].

Les réseaux Feed-forward sont simples, n'ayant pas de retour (connexion) en arrière comme le réseau de neurones probabiliste (PNN) [75], alors que les réseaux récurrents possèdent un comportement plus complexe, avec des boucles ou connexions bouclées dont on peut citer, le réseau de neurone récurrent (RNN) [76] et le réseau de neurones à délais temporels (TDNN) [76].

3.6.6 Classification et reconnaissance des visages par réseaux de neurones

Classer un ensemble d'objets, c'est attribuer à chacun une classe (ou catégorie) parmi plusieurs classes définies à l'avance. Cette tâche est appelée classification ou discrimination. Un algorithme qui réalise automatiquement une classification est appelé classifieur [77]. Les statisticiens appellent aussi classification, la tâche qui consiste à regrouper des données qui se ressemblent dans des classes qui ne sont pas définies à l'avance. Les réseaux de neurones, comme la plupart des outils discriminatifs, peuvent également s'utiliser pour résoudre des problématiques de classification. Contrairement aux problématiques de régression, la classification par les réseaux de neurones va affecter l'appartenance des individus aux classes [68]. Les applications des classifieurs sont très nombreuses, dans le domaine de la reconnaissance des formes (biométries, chiffres et caractères manuscrits ou imprimés, images, parole...), mais également dans bien d'autres domaines (économie, finance, sociologie, traitement du langage...).

Historiquement, les premiers réseaux de neurones utilisés pour la classification étaient conçus pour fournir une décision. Grâce aux progrès accomplis dans la compréhension des propriétés fondamentales des réseaux de neurones, on sait à présent qu'ils peuvent donner une information beaucoup plus riche et fine qu'une simple décision binaire, ils peuvent estimer la probabilité d'appartenance de l'objet inconnu à chacune des classes, ce qui leur permet notamment de s'intégrer dans des systèmes de reconnaissance complexes qui utilisent plusieurs systèmes de classification différents, chacun d'eux fournissant une estimation de l'appartenance de l'objet inconnu à chacune des classes. La décision finale est prise au vu de ces estimations, et en fonction, des résultats des classifieurs [74] [77].

Dans cette partie, nous décrivons l'application des réseaux de neurones artificiels dans le domaine de la reconnaissance de visages. La reconnaissance est le noyau de ce système, c'est la comparaison du vecteur code du visage de test avec ceux de la base de données, et

partons du fait que nous voulons modéliser une fonction du cerveau humain et qu'on a un problème de classification. L'utilisation des réseaux de neurones dans le cadre d'une tâche de reconnaissance de visage, ont été employés dans tous les modules de traitement liés à ce domaine, détection de visage, extraction des caractéristiques et pour la classification. L'application de ces méthodes en reconnaissance automatique des personnes est relativement récente, et reste en évolution continue vu les taux de reconnaissance obtenus qui ne sont pas encore satisfaisants [74].

Le réseau de neurone qu'on va décrire dans cette partie pour la reconnaissance de visages, est un réseau multicouche avec une seule couche cachée et le nombre d'entrée dépend de la taille de l'image. Durant la phase d'enrôlement, un apprentissage permet d'adapter les paramètres du réseau afin qu'il puisse par la suite discriminer les signatures présentées sur ses entrées. Pour ce faire, on procède de façon itérative, on sélectionne un visage de la base d'apprentissage dont on calcule la signature qu'on place en entrée du réseau. On calcule les sorties du réseau pour cette valeur de ses entrées, puis on évalue une fonction d'erreur en considérant que les sorties idéales sont toutes nulles sauf celle qui correspond à la classe dont fait partie l'individu. Pour ajuster les paramètres du réseau, on utilise la technique standard de rétro-propagation de l'erreur, cette procédure est itérée en présentant à l'entrée, plusieurs fois, toutes les signatures extraites des images de la base. Durant la phase d'identification, on calcule la signature du visage cible et on l'applique en entrée du réseau. La sortie obtenue qui prend la valeur maximale indique à quelle classe appartient le visage cible. L'introduction d'un seuil sur la valeur de la sortie maximale permet de déterminer si l'image cible ne correspond à aucun individu connu de la base [74].

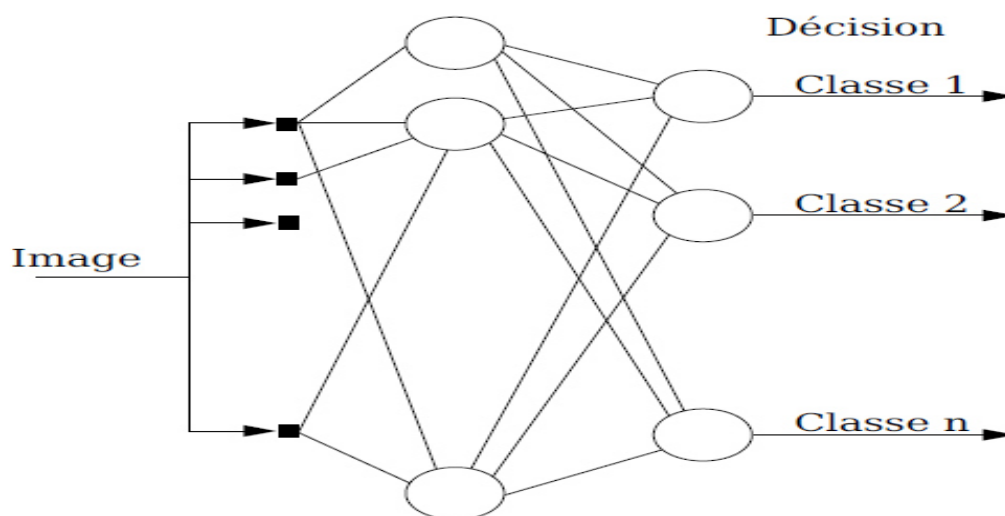


Fig.3.8- Réseau de Neurones Artificiels discriminant pour la reconnaissance de visages.

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

3.7 Les réseaux de neurones profonds

A présent que nous avons précisé comment fonctionnent les réseaux de neurones de manière générale, nous allons aborder le domaine des réseaux de neurones profonds. Cette famille d'algorithmes a permis de faire des progrès importants dans les domaines de la classification des images et de reconnaissance de forme. Ces modèles s'appuient sur un réseau de neurones artificiels s'inspirant du cerveau humain. Ce réseau est composé de dizaines voire de centaines de couches de neurones, chacune recevant et interprétant les informations de la couche précédente [78]. Il convient de souligner que les différentes couches intermédiaires sont plus nombreuses. Chacune des couches intermédiaires va être subdivisée en sous partie, traitant un sous problème, plus simple et fournissant le résultat à la couche suivante, et ainsi de suite jusqu'à la couche de sortie (décision). Le système apprendra par exemple à reconnaître les lettres avant de s'attaquer aux mots dans un texte, ou détermine s'il y a un visage sur une photo avant de découvrir de quelle personne il s'agit. Une définition générale d'un réseau de neurones profonds peut être exprimée en tant qu'un réseau de neurone artificiel avec plusieurs couches cachées entre les couches d'entrée et de sortie du réseau (il n'existe pas de définition absolue de la « profondeur » d'un réseau de neurones). Les réseaux multicouches avaient été considérés comme les premiers réseaux profonds, En effet, plus l'on ajoute des couches cachées, plus la profondeur du réseau augmente [78].

Les réseaux de neurones profonds surpassent les autres techniques utilisées dans le traitement d'images. Les méthodes phares des années passées (extraction de caractéristiques puis apprentissage) sont progressivement remplacées par différents types de réseaux de neurones avec un apprentissage en profondeur [79]. Cependant, la formation de réseaux de neurones profonds posait un problème fondamental de ressources de calcul élevées, qui jusqu'à récemment n'était pas disponible. Ces types de réseaux ne sont pas un simple empilement de couches de neurones. En effet, pour résoudre des problèmes de plus en plus complexes, il n'est pas suffisant d'ajouter toujours plus de couches. Les deux grosses

problématiques des réseaux de neurones, qui sont la difficulté d'apprentissage et la complexité calculatoire croissante avec le nombre de couches, sont en effet toujours présentes. Les solutions récentes apportées par les recherches permettent de limiter ces problèmes mais pas de les régler complètement [80]. La disponibilité récente de puissants processeurs graphiques (GPU) a été un facteur important dans l'avancement la recherche dans ce domaine. Avec les GPU, le temps d'apprentissage pour de si grands réseaux a été réduit de plusieurs ordres de grandeur. Le domaine des réseaux de neurones profonds est un domaine à croissance rapide, et de nouvelles architectures, variantes ou algorithmes apparaissent toutes les années. Il existe un grand nombre de variantes d'architectures profondes. La plupart d'entre elles sont dérivées de certaines architectures neuronale classique. Il n'est pas toujours possible de comparer les performances de toutes les architectures, car elles ne sont pas toutes évaluées sur les mêmes ensembles de données.

Dans cette section, certaines architectures de réseau de neurones profonds sont expliquées brièvement, avec leurs principales caractéristiques, leurs modèles d'apprentissage automatique, leurs architectures et leurs propriétés de profondeur. Il existe en pratique différents types de réseaux de neurones profonds qui visent à résoudre différents problèmes. La citation de la totalité des modèles neuronale ne peut être complète. On s'est limité à décrire la plupart des grandes architectures et la majorité des réseaux profonds qui ont apporté une vraie originalité et une avancée sensible dans le domaine.

Les avancées récentes en apprentissage profond ont montré d'excellentes performances dans des domaines applicatifs comme la reconnaissance de forme [81] [82], le traitement d'image et de la parole. Dans ce cadre, les auto-encodeurs sont souvent utilisés pour obtenir des représentations latentes capables de capturer suffisamment d'informations pour reconstruire les données d'origine. Un auto-encodeur (AE) [83], est un réseau de neurones artificiels, son objectif est d'apprendre une représentation (encodage) d'un ensemble de données, généralement dans le but de réduire la dimension de cet ensemble. Les auto-encodeurs sont des réseaux de neurones simples composés de trois couches. La première couche et la couche cachée forment l'encodeur, la couche cachée ainsi que la dernière couche formant le décodeur comme décrit dans la figure 3.9. Il faut noter qu'à la différence d'un grand nombre de réseaux de neurones, les auto-encodeurs peuvent être entraînés de manière non-supervisée, ce qui permet d'appliquer ces méthodes à des jeux de données non annotés [82] [83].

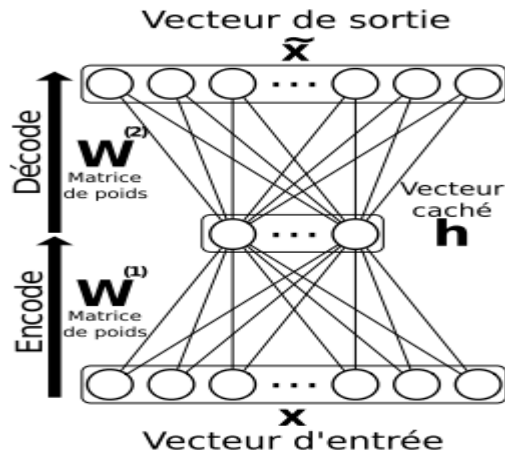


Fig.3.9- Architecture d'un auto-encodeur (AE) [84].

En 2006, le chercheur Geoffrey Hinton propose une nouvelle méthode d'entraînement des réseaux de neurones profonds qui va faire son petit effet dans le domaine. Cette percée a été faite par Hinton et al. [85] [86] qui ont défini une nouvelle architecture réseaux de neurones avec de nombreuses couches cachées basée sur un apprentissage non supervisé. Plus précisément, les auteurs ont proposé un algorithme itératif pour l'apprentissage des réseaux de croyances profonds (en anglais Deep Belief Network : DBN) [85], qui est un modèle multicouche non supervisé composé de plusieurs machines de Boltzmann restreintes (RBM) à couche unique [86]. Les machines de Boltzmann restreintes (RBM) sont des modèles particuliers de réseaux de neurones qui permettent de modéliser des distributions de probabilité sur des espaces discrets. Cet algorithme [87], est caractérisé par des connexions limitées à un sous ensemble strict de toutes les connexions possibles par rapport aux machines de Boltzmann standards (c'est à dire non restreintes), Dans sa formulation la plus simple, une RBM est une machine de Boltzmann sur un graphe bipartite, les deux parties de ce graphe sont respectivement appelées couche visible (v) qui représente les données connectée à une couche cachée(h) censée extraire et expliquer les caractéristiques statistiques de ces données [86].

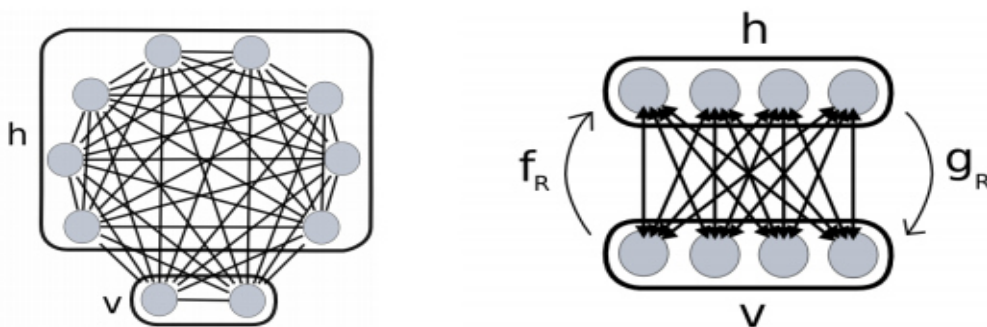


Fig.3.10- À gauche : Topologie d'une machine de Boltzmann classique. À droite : Topologie d'une machine de Boltzmann restreinte (RBM) [88].

L'empilement des différentes RBM constituent un réseaux multicouche appelé Deep Belief Network (DBN). On applique ensuite aux DBN une étape d'apprentissage supervisée. Cette dernière étape est appelée fine-tuning, et se révèle en général beaucoup plus efficace après l'étape de pré-entraînement, car on ne part plus d'une solution initiale aléatoire [88].

Contrairement à un réseau de neurones acyclique (Feed-Forward Neural Network : FFNN), Les réseaux de neurones récurrents (Recurrent Neural Network : RNN) ont été introduit en machine learning afin de pouvoir traiter des données séquentielles [88] [89]. Pour modéliser ces données et les éventuelles dépendances temporelles qui y apparaissent, les RNNs conservent un état caché qui résume leur historique. Un réseau de neurones récurrent est un réseau de neurones dont le graphe de connexion contient au moins un cycle, de nombreux types de RNN ont été développés au cours des ces dernières années tels que les réseaux d'Elman, les réseaux de Jordan et les Echo State Networks [89]. Récemment un type de réseaux RNN a émerger grâce à ses excellentes performances sur des tâches aussi nombreuses que variées ,les réseaux de neurones à base de cellules (Long Short-Term Memory LSTM) ou plus explicitement réseau de neurones récurrents à mémoire court-terme et long terme [89], cette architecture de réseau de neurones récurrents est la plus utilisée, en pratique elle était proposé par Sepp Hochreiter et Jürgen Schmidhuber en 1997 [90] pour résoudre le problème de disparition de gradient. Les RNN ont connu un grand succès dans de nombreuses tâches à titre d'exemple le traitement du langage naturel, traduction automatique, la reconnaissance vocale, et la description des images [89] [90]. En 2012, les travaux de Krizhevsky et al. [91] ont révolutionné les domaines de reconnaissance de formes et vision par ordinateur, par le développement des réseaux de neurones convolutifs (CNN) ont devenues un élément indispensable de la quasi-totalité des approches de pointe du domaine de reconnaissance et classification d'images. Nous explorons ce type de réseaux profonds plus en détail dans la suite de ce chapitre.

3.8 Réseaux de neurones convolutifs

Un réseau de neurones convolutif ou réseau de neurones à convolution (CNN ou ConvNet en anglais pour Convolutional Neural Networks) sont une forme particulière de réseau neuronal multicouches dont l'architecture des connexions est inspirée de celle du cortex visuel des mammifères [92]. Ces réseaux sont capables de catégoriser les informations des plus simples aux plus complexes. Ils consistent en un empilage multicouche de neurones et fonctions mathématiques à plusieurs paramètres ajustables, qui prétraitent de petites quantités d'informations.

Le premier réseau de neurones convolutif (CNN) a été introduit à la fin des années 80 par LeCun [93] [60]. C'est le premier réseau de neurones pour la reconnaissance d'images, ce réseau permettait la reconnaissance des chiffres manuscrits. L'idée est de passer l'image dans une succession de filtres convolutifs apportant une description réduite et pertinente de l'image. Ces caractéristiques sont, par la suite, envoyées à un perceptron multicouches composé de couches cachées et d'une couche de sortie complètement connectées permettant la classification du chiffre présent dans l'image. Par la suite, en 1998, Yann LeCun et al [94] ont proposé une structure de réseaux de convolutions LeNet-5 représentée dans la figure 3.11, associée à un algorithme d'apprentissage pour la classification des chiffres manuscrits.

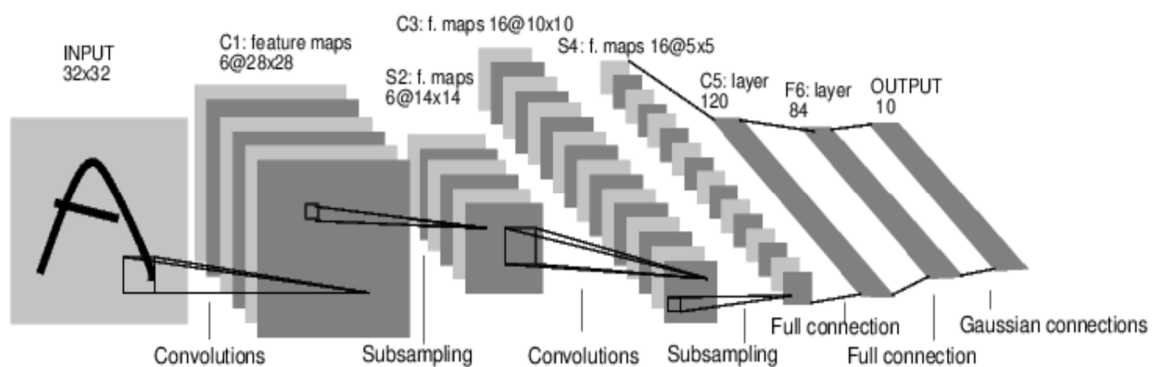


Fig.3.11- Architecture d'un réseau de neurones convolutifs LeNet-5 [94].

Les réseaux CNN sont particulièrement utiles pour identifier des modèles sur des images afin de reconnaître des objets, des visages et des scènes. Ils apprennent directement à partir de données d'entrées (images) et utilisent des modèles pour les classer, éliminant ainsi la nécessité d'effectuer une extraction manuelle des caractéristiques [94]. Le terme convolution vient de l'opération de convolution des matrices utilisées dans le traitement de signal. Dans les convNet, deux nouveaux types de couche ont été ajoutés dans le réseau par rapport aux réseaux de neurones conventionnels : la couche convolution (convolutional layer) et la couche de mise en commun ou de sous-échantillonnage (pooling layer) [66], nous les décrivons dans les parties suivantes.

Les CNN sont un type particulier de réseaux de neurones, applicables facilement sur des images pour capter spatialement de l'information [60]. De plus, de par leur structure convolutive, ils permettent de prendre en entrée des données de grande dimension ce qui est une limite du perceptron multicouches. Une image à trois canaux (RGB) de taille 224×224 pixels représente un vecteur d'entrée de taille (150528) pour un perceptron multicouche.

Cela implique (150528) poids à apprendre pour chaque neurone de la couche cachée connectée aux entrées, ce qui est énormes et compliqué à apprendre. Les CNN peuvent être vus comme un assemblage de modules en série permettant l'extraction de caractéristiques de manière hiérarchique à partir des pixels d'une image [60].

Au final, les CNN sont composés de couches non rebouclées, chaque couche est composée de neurones qui sont connectés à la couche précédente. Dans l'architecture standards d'un réseaux ConvNet (la figure 3.12), l'image d'entrée (input image) est convoluée avec un filtre a noyaux (kernels) du réseau convolutif pour l'obtention de la carte des caractéristiques de l'image (feature maps) [94]. Ces caractéristiques sont ensuite sous-échantillonnées à l'étape de pooling, ce processus est répété à plusieurs reprise en alternance (convolution et pooling). Pour la dernière phase celle de sortie, les résultats sont obtenus en utilisant un perceptron multicouche pour aboutir à la classification finale [60] [95].

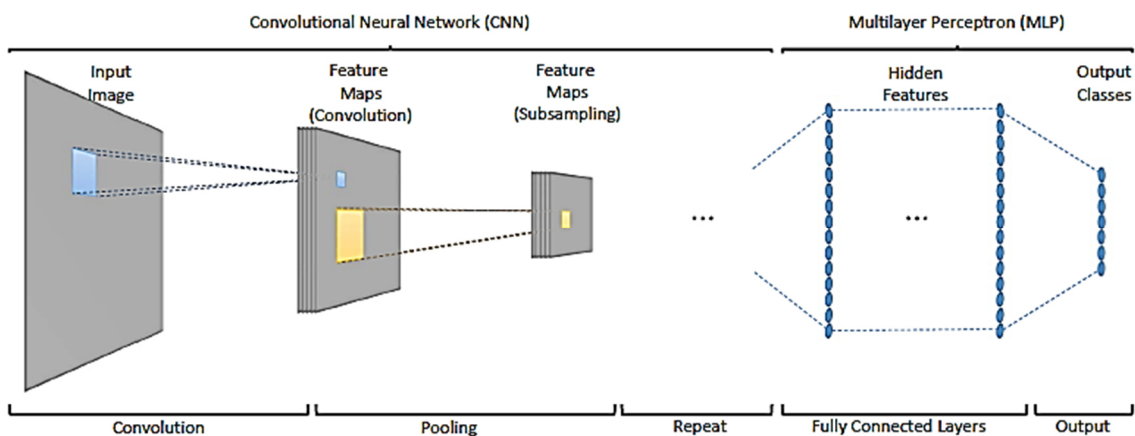


Fig.3.12- Les différentes couches d'un réseau de neurones convolutif standard [96].

3.8.1 Différents modules d'un réseau de neurones convolutif

Au vu de cette structure globale, nous allons donc nous intéresser maintenant aux différents couches (module) classiquement utilisés dans les réseaux CNN.

- **La convolution** : La pièce maîtresse d'un réseau CNN est la couche de convolution. La sortie en résultant est appelée carte de caractéristiques [60]. Une couche convolutive est constituée de plusieurs filtres (ou noyaux) de convolution appliquée sur une matrice d'entrée (une image ou une carte de caractéristique précédente). La couche de convolution fait passer les images d'entrée par un ensemble de filtres convolutifs, chacun de ces filtres activant certaines caractéristiques des images.

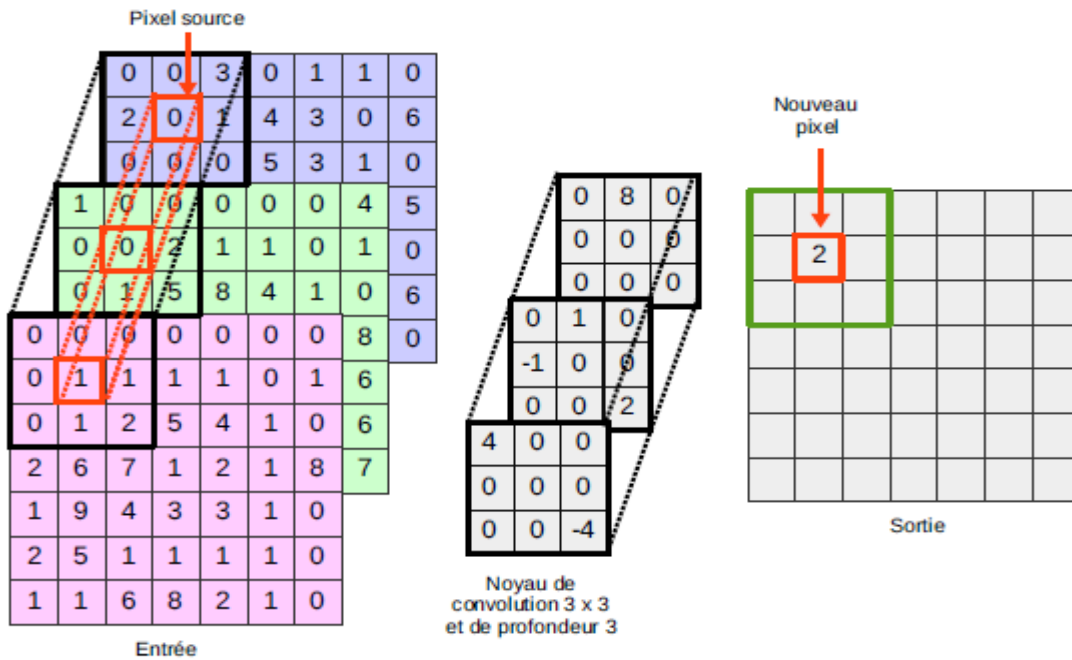


Fig.3.13- Illustration de la convolution [66].

Étant donné en entrée, un filtre de convolution (ou noyau de convolution) est appliqué pour chaque position de la matrice d'image en question. La profondeur du noyau dépend de de l'image d'entrée sur laquelle il est appliqué. Dans l'exemple illustré dans la figure 3.13 l'image d'entrée possède trois canaux (image couleur RGB) donc la profondeur du noyau est de trois. Le résultat pour une position donnée correspond à la somme de la multiplication des éléments du noyau par ceux de l'entrée. Dans l'exemple précédant $(2 \times -4) + (5 \times 2) = 2$ représente le résultat de la multiplication élément par élément de la matrice par le noyau de convolution. Dans le cadre des CNN, la sortie d'une étape de convolution est appelée carte de caractéristiques. Le nombre de cartes de caractéristiques dépend du nombre de filtres appliqués sur l'image d'entrée [60] [66].

Un filtre dans l'étape de convolution sert à faire ressortir certaines caractéristiques d'une image donnée (couleur, contour, luminosité, netteté, etc.). Un filtre d'une couche convolutive est appliqué à toutes les positions de la matrice d'entrée. Ce filtre va être déplacé par pas successifs sur l'ensemble d'images. Pour chaque position du filtre, les valeurs des deux matrices en superposition (filtre et image à traiter) sont multipliées puis projetée dans une nouvelle matrice. Pour cela il faut utiliser un paramètre de la convolution appelé Stride ou pas de convolution. Il représente de combien de pixels dans l'image d'entrée on va se décaler pour réappliquer la convolution [97]. Cependant, il peut arriver que la convolution dépasse l'image si on choisit mal le kernel et le stride. Par exemple, pour

une image de 4×4 (pixels), un kernel de 2×2 et un stride de 3, on va avoir un problème après la première convolution. De plus, le résultat de la convolution est une image plus petite que l'image d'entrée. A ces deux problèmes, il y a un paramètre qu'on appelle le Padding qui est utilisé dans ce cas de figure en ajoutant tout simplement des 0 autour de l'image d'entrée, pour éviter les dépassements [60].

-**Le pooling** : Le but de la couche pooling ou sous-échantillonnage (subsampling) est de réduire la taille des images après le procédé de la convolution, elle permet de rajouter de l'invariance spatiale lors de l'extraction de caractéristiques tout en réduisant la dimension des entrées. Elle peut être de différentes natures mais les types de pooling les plus utilisés sont le Max Pooling (illustré dans la Figure 3.14). Le Max Pooling renvoie l'élément maximum sur une fenêtre de calcul (la matrice issue de la convolution) [60].

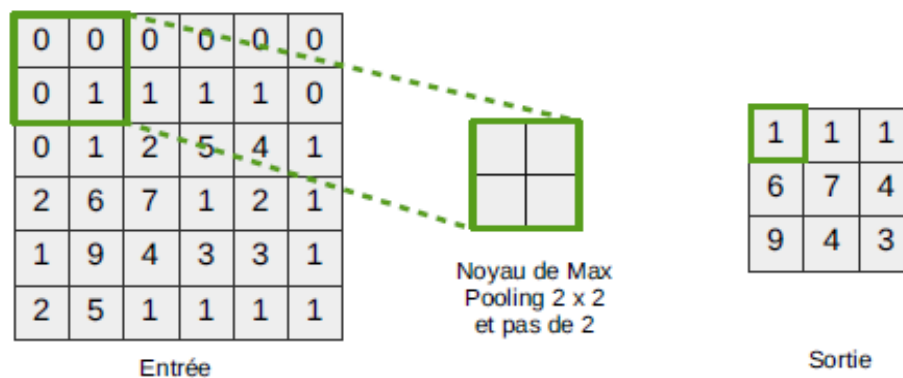


Fig.3.14- Illustration du Max Pooling [60].

En extrayant les valeurs importantes des pixels, cette étape permet de réduire une image tout en conservant les caractéristiques pertinentes. Dans cet exemple (de la figure 3.14), le noyau de pooling est de taille 2×2 et est appliqué tous les deux pixels (stride = 2). Le maximum des quatre éléments sur une fenêtre de la matrice d'entrée est gardé.

Il existe plusieurs types de pooling [97]. Le max-pooling qui revient à prendre la valeur maximale de la sélection. C'est le type le plus utilisé car il est rapide à calculer (immédiat), et permet de simplifier efficacement l'image. Le mean-pooling (ou average-pooling), soit la moyenne des pixels de la fenêtre sélectionné, on calcule la somme de toutes les valeurs et on divise par le nombre de valeurs. On obtient ainsi une valeur intermédiaire pour représenter ce lot de pixels. Le sum-pooling, représente la moyenne sans avoir divisé par le nombre de valeurs (on ne calcule que leur somme) [60].

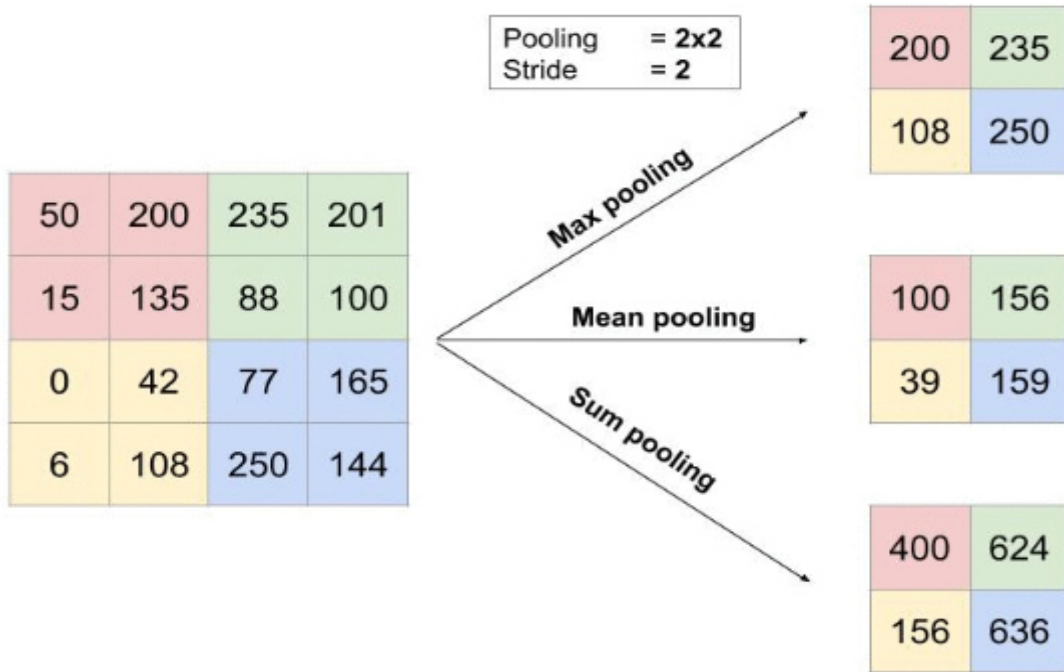


Fig.3.15- les différents types de Pooling avec un filtre 2x2 et un pas de 2.

-Les fonctions d'activation : Il est possible d'améliorer l'efficacité du réseau CNN, en intercalant entre les couches de traitement une fonction qui va opérer comme une couche de correction sur les signaux de sortie. Cela permet d'introduire des non-linéarités pour permettre au réseau d'apprendre des systèmes complexes non linéaires. Il existe différentes fonctions d'activation permettant la non-linéarité dans les différentes couches des CNN. Parmi les plus connues : le sigmoïde (logistique), la tangente hyperbolique et la fonction d'unité de rectification linéaire (Rectified Linear Unit) ReLU [60] [66].

La fonction ReLU (abréviation de Unités Rectifié linéaires : $F(x)=\max(0, x)$), est une fonction qui vient briser (une partie de) la linéarité en supprimant une partie des valeurs (toutes celles négatives). En supprimant une partie des données, ReLU permet également d'accélérer les calculs. En ne modifiant pas les données positives, ReLU n'impacte pas les caractéristiques mises en évidence par la convolution, au contraire elle les met davantage en évidence en creusant l'écart (valeurs négatives) entre deux caractéristiques. Celle-ci est certainement la fonction la plus utilisée dans les réseaux CNN profonds [60], car elle permet une optimisation plus facile. Elle a pour avantage de fournir des réponses parcimonieuses, ne sature pas et permet de réduire les problèmes de disparition de gradient. La fonction ReLU renvoie un gradient constant pour une entrée, permettant ainsi d'apprendre plus rapidement (en particulier les réseaux d'une certaine profondeur) [97].

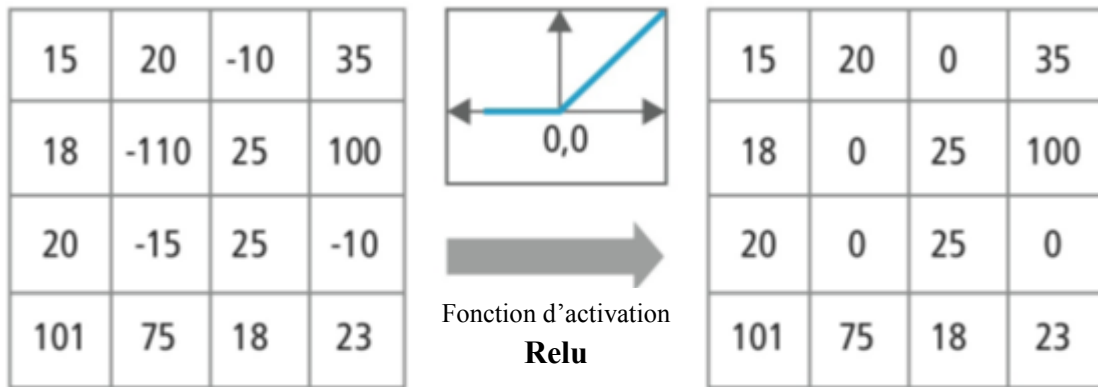


Fig.3.16- Exemple de fonctionnement d'une couche Relu. (Tous les nombres négatifs valorisés dans la case de gauche ont été mis à zéro après la fonction de d'activation appliquée, toutes les autres valeurs restent inchangées).

-Couche entièrement connectée (Fully Connected Layer (FC)) : Après plusieurs couches de convolution et de max-pooling, La couche fully-connected constitue toujours la dernière couche d'un réseau de neurones convolutif. Les neurones dans une couche entièrement connectée ont des connexions vers toutes les sorties de la couche précédente (comme on le voit régulièrement dans les réseaux conventionnels de neurones perceptron multicouches) [67]. En outre, la couche entièrement connectée est la dernière couche où la classification est effectuée. Ici, nos images filtrées et réduites passent par une étape dite de "Flattening" (ou aplatissement). Cette opération consiste à mettre toutes les données dans un seul vecteur. Ce vecteur permettra la création d'une première couche de neurones entièrement connectée. C'est à dire que chacune des valeurs de ce vecteur sera connectée aux neurones de la première couche du réseau permettant la classification de l'image.

La taille de cette couche dépendra de la taille de la sortie des couches précédentes. Par exemple, si nous avons besoin de dix étiquettes de classe différentes, la taille de cette couche serait de 10. La probabilité pour chacune des classes est calculée sur cette couche. Selon les besoins, les couches entièrement connectées peuvent être modélisées pour la régression ou la classification [60].

La forme la plus commune d'une architecture de réseau de neurones convolutifs empile quelques couches Convolution-ReLU, les suit avec des couches Pooling, et répète ce schéma jusqu'à ce que l'entrée soit réduite dans un espace d'une taille suffisamment petite. À ce moment, il est fréquent de placer des couches entièrement connectées (FC) [66] [67]. La dernière couche entièrement connectée est reliée vers la sortie (la figure 3.17 présente l'architectures communément utilisés dans un modèle de réseau de neurones convolutifs).

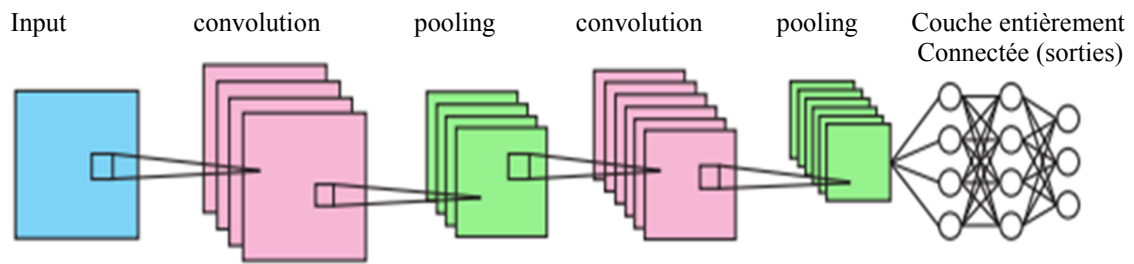


Fig.3.17- Architecture standard d'une réseaux convolutifs.

3.8.2 Outils d'optimisation des réseaux convolutifs

Les réseaux de neurones convolutifs utilisent différents paramètres d'optimisation à la différence des réseaux de neurones standard multicouche. Dans cette section, nous nous attachons à décrire des méthodes permettant une meilleure optimisation associée au réseau CNN. La plupart de ces méthodes ont été mises en œuvre lors de la thèse pour le développement et l'apprentissage des réseaux de neurones convolutifs [60].

-La batch normalization : Dans le cadre d'un apprentissage automatique, l'ordre de présentation des échantillons est très important puisque les paramètres du réseau sont mis à jour après chaque passage d'un échantillon d'apprentissage. Cependant, il n'y a pas de méthodes efficaces pour déterminer à priori quel est l'échantillon qui apportera le plus d'information, sauf peut-être la recherche exhaustive (et donc coûteuse) [98]. Une méthode simple consiste donc à mélanger l'ordre de passage des échantillons après chaque passage de l'ensemble d'apprentissage. Ainsi, les échantillons successifs n'appartiennent pas à la même classe, et contiennent donc potentiellement plus d'information [66].

Une technique a été introduite récemment en 2015 [98] afin de rendre l'apprentissage des CNN plus rapide et efficace. Elle part de l'observation suivante, pendant l'apprentissage, la distribution des entrées des différentes couches du réseau change à chaque itération. Cela induit une adaptation permanente des paramètres du CNN à ces différentes distributions, ce qui augmente le temps d'apprentissage. L'idée de la batch normalization (normalisation par lots) [60], est de normaliser les entrées de chaque couche afin que les distributions de celles-ci soient de moyenne nulle et de variance unitaire. Durant l'apprentissage, les couches de batch normalization utilisent des paramètres (un facteur d'échelle et un biais) permettant d'ajuster cette normalisation, ces paramètres permettent d'appliquer une transformation sur la distribution normalisée. Autrement dit, durant l'apprentissage, si le réseau considère que la distribution normalisée n'est pas adaptée pour une couche donnée, il apprend les paramètres permettant de l'ajuster.

-Les fonctions de perte : La fonction (ou couche) de perte spécifie comment l'entraînement du réseau pénalise l'écart entre le signal prévu et réel. Il existe plusieurs fonctions de perte (ou fonctions d'objectif) utilisables pour l'apprentissage des réseaux de neurones. Ces fonctions sont dépendantes de la tâche que le réseau doit effectuer (classification, régression...). Nous listons ici, les fonctions de perte les plus utilisées [60].

- La fonction de perte Softmax : généralement utilisée pour l'optimisation de réseau de classification d'images. Elle permet la maximisation de la probabilité qu'à une entrée d'appartenir à une classe plutôt qu'à une autre.
- La fonction de perte par entropie croisée (sigmoïde) : permettant une régression sur des probabilités.
- La perte quadratique (squared loss) : également appelée perte L_2 . Cette fonction calcule les carrés de la différence entre la valeur prédite d'un modèle pour un exemple étiqueté et la valeur réelle de l'étiquette. Une autre fonction associée avec ce type de perte, est l'erreur quadratique moyenne (Mean Squared Error (MSE)) qui est calculée en divisant la perte quadratique par le nombre d'exemples.

-Méthodes de régularisation : Le nombre d'entrées et de sorties dans un réseaux est généralement déterminé par les données d'apprentissage, mais le nombre total de neurones des couches intermédiaires et le nombre de ces couches est un paramètre qui doit être ajusté pour éviter à la fois le sur-apprentissage (overfitting) et le sous apprentissage (underfitting) [98]. En apprentissage automatique, la régularisation est un procédé visant à améliorer les performances de généralisation d'un algorithme d'apprentissage, autrement dit à diminuer l'erreur sur les nouvelles données si elles suivent les mêmes lois que celles d'apprentissage. La régularisation est introduite pour réduire l'erreur de généralisation sans perturber l'erreur d'apprentissage. Différents types de régularisation peuvent être envisagées [98] [99] :

- *Le dropout* : Pour éviter le sur-apprentissage (overfitting), la couche de dropout a été introduite [99]. Cette couche est utilisée pendant l'apprentissage. Elle permet de désactiver aléatoirement des neurones durant les différentes itérations de l'apprentissage. Cette manière de faire permet d'apprendre des paramètres plus génériques qui ne se focalisent pas sur des détails de la base d'apprentissage. Une fois l'apprentissage terminé, tous les neurones sont réactivés [60].
- *Early stopping* : consiste à entraîner le réseaux en utilisant à la fois une base d'entraînement et une base de test, et à stopper l'entraînement lorsque l'erreur de test se met à ré-augmenter

- *Augmentation de données* : consiste à augmenter la taille de la base d'apprentissage en lui ajoutant des données obtenues par transformations (ajout de bruit, transformations géométriques, etc.) des données de la base de départ.
- *La régularisation par norme 1* : la spécificité de cette régulation est de diminuer le poids des entrées aléatoires et faibles et d'augmenter le poids des entrées « importantes ». Le système devient moins sensible au bruit [66].
- *La régularisation par norme 2* : (norme euclidienne) La spécificité de cette régulation est de diminuer le poids des entrées fortes, et de forcer le neurone à plus prendre en compte les entrées de poids faible [60] [66].

3.9 Utilisation de modèles pré-entraînés pour l'apprentissage par transfert

Un réseau de neurones à convolution est entraîné sur des centaines, des milliers, voire des millions d'images. Lorsque vous travaillez avec de grandes quantités de données et des architectures de réseau complexes, les GPU peuvent réduire considérablement le temps de traitement nécessaire pour entraîner un modèle. L'apprentissage d'un réseau à partir du début, nécessite l'adaptation du nombre de couches, de filtres, ainsi que d'autres paramètres réglables. Entraîner un modèle précis à partir de zéro nécessite également de grandes quantités de données, parfois des millions d'échantillons, ce qui peut s'avérer extrêmement chronophage [95] [96]. Plutôt que d'entraîner un CNN à partir du début, il est courant d'utiliser un modèle pré-entraîné pour extraire automatiquement des caractéristiques à partir d'un nouveau jeu de données. Cette méthode, appelée apprentissage par transfert (Transfer Learning en anglais), est une solution pratique pour appliquer un apprentissage profond sans avoir à recourir à un très grand jeu de données ni à procéder à des calculs et un apprentissage très longs [95].

Pour créer un réseau, on doit déterminer la configuration de ce dernier. Cette approche est celle qui vous donne le plus grand contrôle sur le réseau, mais elle implique de comprendre la structure d'un réseau de neurones et les nombreuses options disponibles en matière de types et de configurations des couches. Si les résultats peuvent parfois dépasser ceux de l'apprentissage par transfert, cette méthode nécessite généralement davantage d'images pour l'apprentissage, car le nouveau réseau a besoin de nombreux exemples de l'objet pour comprendre la variation des caractéristiques. La durée de l'apprentissage est souvent plus longue, et les combinaisons possibles des couches du réseau sont si nombreuses qu'il peut s'avérer fastidieux de le configurer [100].

Ajuster un réseau pré-entraîné à l'aide de l'apprentissage par transfert est généralement beaucoup plus rapide et plus simple que d'entraîner un réseau depuis le début du processus. Cette méthode est celle qui nécessite le moins de données et de ressources de calcul. L'apprentissage par transfert exploite les connaissances acquises en résolvant un problème pour résoudre des problèmes similaires [95]. L'un des avantages de l'apprentissage par transfert, c'est que le réseau pré-entraîné a déjà appris un riche ensemble de caractéristiques. Ces caractéristiques peuvent ensuite être appliquées à de nombreuses autres tâches similaires. Par exemple, vous pouvez prendre un réseau entraîné sur des millions d'images et le ré-entraîner pour une nouvelle classification d'images en utilisant seulement une centaine d'images. Pour des usages pratiques, il est possible d'exploiter la puissance des CNN sans être un expert du domaine, avec du matériel accessible et une quantité raisonnable de données annotées. Toute la complexité de création de CNN peut être évitée en adaptant des réseaux pré-entraînés disponibles publiquement. Ces techniques exploitent la connaissance acquise sur un problème de classification général pour l'appliquer de nouveau à un problème particulier.

L'utilisation de réseaux CNN pré-entraînés sur d'immenses bases de données pour ensuite les adapter à une tâche spécifique. C'est ce que l'on appelle le fine-tuning [60]. Cette pratique permet d'apprendre des réseaux profonds en utilisant une initialisation des poids et des biais déjà très pertinente et générique. L'adaptation de ces paramètres est ensuite réalisée durant la phase d'apprentissage de la tâche spécifique que l'on souhaite réaliser. Cela a pour incidence une vitesse d'apprentissage bien plus rapide et une convergence quasiment garantie [101].

3.10 Les architectures neuronales convolutifs

Nous présentons ici les architectures de réseaux convolutifs profonds utilisées couramment dans les travaux de recherche en vision par ordinateur, reconnaissance de formes et notamment dans des tâches d'identification faciale. Il est important de remarquer que les architectures proposées dans la littérature ont une forte tendance à devenir de plus en plus profonde avec les années. Autrement dit, il semble que plus le réseau est profond, plus les performances sont meilleures. Néanmoins, cette profondeur implique de faire face à certaines difficultés notamment en termes de temps de calcul et d'optimisation durant l'apprentissage. C'est pourquoi la communauté reste très active sur la problématique de conception d'architectures neuronales [60] [66]. Plusieurs architectures CNN ont déjà été proposées, certaines seront décrites dans cette partie.

-LeNet (1998) : LeCun [94] a été le premier à mettre en œuvre avec succès une application d'un réseaux CNN, utilisé pour la reconnaissance de l'écriture manuscrite. Il se compose de sept couches, sans compter la couche d'entrée. Les images d'entrée utilisées étaient de taille 32×32 . La première couche est composée de six filtres 5×5 , qui après la convolution amènent la taille à 28×28 . Après la convolution vient une couche de sous-échantillonnage max-pooling, puis seize autres filtres 5×5 pour la deuxième couche de convolution, suivie de la dernière couche de sous-échantillonnage avant d'être projetés dans une couche entièrement connectée [94] [101].

-AlexNet (2012) : Cette architecture proposée par Krizhevsky et al [82]. C'est l'un des premiers travaux à avoir popularisé les réseaux convolutifs en informatique, grâce à la victoire lors de la compétition de classification d'images ImageNet. Cette architecture utilise cinq couches de convolution et trois couches de pooling. La taille des noyaux de convolution est variable (11×11 , 5×5 , 3×3) en fonction de la couche considérée. La fonction d'activation utilisée entre chaque couche est la fonction ReLU. Après le passage de l'image dans les couches de convolution, de pooling et d'activation, une carte de caractéristiques est obtenue. Celle-ci est envoyée dans un perceptron multicouches composé de deux couches cachées et d'une couche de sortie [60].

-VGGNet (2014) : Ce CNN a été introduit par Simonyan et al [101]. Au lieu d'utiliser une seule convolution par niveau de profondeur comme AlexNet, cette architecture utilise des séquences de convolution. De plus, les filtres convolutifs sont de plus petite taille que dans AlexNet (noyau de taille 3×3) [60].

-GoogLeNet (2015) : Cette architecture de CNN a été créée par Szegedy et al [102], elle permet une réduction du temps de calcul par rapport à l'architecture VGG présentée précédemment. Pour cela, le GoogLeNet est composé de plusieurs couches appelées couches d'inception. Elles sont composées de plusieurs modules de convolution exécutés en parallèle sur la carte de caractéristiques résultant de la couche précédente d'autres modules d'inception ont par la suite été proposés notamment Inception V2 et V3 [60].

-ResNet (2016) : Ce type de réseau CNN (réseau résiduel) permet l'apprentissage de réseaux très profonds (plus de 150 couches) [60]. L'idée développée dans ResNet est l'utilisation de connexions résiduelles permettant une meilleure optimisation des réseaux très profonds. Une connexion résiduelle permet de passer l'entrée dans deux filtres de convolution mais également de passer directement cette entrée aux couches suivantes.

3.11 Conclusion

Tout au long de cette partie de la thèse, nous avons introduit les concepts de base de l'apprentissage automatique ainsi que ceux de l'apprentissage profond. Dans ce contexte, nous avons introduit les réseaux de neurones et plus particulièrement les réseaux neuronaux convolutifs profonds (CNN) qui est la principale technique utilisée dans la thèse. Depuis ses origines dans le domaine de la biologie cérébrale, nous retraçons l'état actuel de la technologie des réseaux de neurones profond appliqués au problème de reconnaissance de forme. Ces modèles permettent d'apprendre, grâce à l'apprentissage supervisé, les caractéristiques visuelles discriminantes à extraire sur les images afin de résoudre des problématiques de classification et/ou de régression. Les CNN sont au centre de l'attention, en particulier en vision par ordinateur et reconnaissance de forme, vu les résultats impressionnants qu'ils produisent sur ces différentes tâches.

Dans ce chapitre nous avons passé en revue les fondements essentiels des réseaux de neurones convolutifs. Nous avons défini l'architecture détaillée de ce type de réseau ainsi que les différentes couches qui le constitue. Une partie a été consacrée aux propriétés et paramètres du CNN. Cette étude détaillée du modèle CNN nous permet de mieux comprendre son processus et faire une conception pour une tâche de reconnaissance du visage qui fera objet du chapitre suivant.

CHAPITRE :4

Expérimentations, Résultats et discussions

4.1	Introduction
4.2	Les données biométriques
4.2.1	ORL
4.2.2	FERET
4.2.3	CASIA-WebFace DATABASE
4.2.4	Séparation des bases de données et matériel utilisé
4.3	Les différentes étapes de notre approche de reconnaissance de visage
4.3.1	Extraction des caractéristiques faciales
4.3.1.1	Convolution et pooling
4.3.1.2	La structure du réseau de neurones convolutifs CNN proposée
4.3.1.3	Le pré-entraînement du réseau CNN
4.3.2	Réduction des dimensions par analyse en composantes principales
4.3.3	Classification par support vecteur machine SVM
4.4	Expérimentations
4.4.1	Données utilisées
4.4.2	Méthodologies des tests, résultats et discussions
4.5	Conclusion

4.1 Introduction

Nous proposons dans ce chapitre une nouvelle approche d'identification de visage sur des bases de données biométriques. Dans cette partie, nous aborderons l'expérimentale de la thèse, dans un premier temps, nous discuterons les bases de données sur lesquelles nous avons réalisées les différents tests, ainsi qu'une description de leurs contenus. Ensuite, nous décrirons les méthodes sélectionnées dans ce travail, leurs paramètres et les différentes expérimentations. Enfin, nous évaluons les résultats d'implémentation de ces approches, qui seront principalement discutées.

Nous étudions dans ce travail plusieurs voies de recherches relatives à l'apprentissage automatique pour l'extraction des signatures biométriques faciales, la taille de ces vecteurs extraits, et les résultats de classification. L'apprentissage profond est un candidat idéal pour résoudre un problème de reconnaissance de visage. Cette partie donne un aperçu sur la construction et l'application d'une nouvelle approche de préapprentissage des réseaux (CNN) pour la reconnaissance faciale. De plus, il décrit les résultats substantiels obtenus par ces méthodes neuronales combinés avec l'analyse en composante principale et le support vecteur machine sur des images de visages.

4.2 Les données biométriques

Les données biométriques à l'entrée d'un système d'identification doivent être acquises sans contraintes imposées aux utilisateurs et aux personnes à identifier. Alors que les systèmes d'authentification conçus essentiellement pour des applications de contrôle d'accès, peuvent utiliser une large panoplie de modalités biométriques, les systèmes d'identification biométrique visent principalement des applications de surveillance et d'investigation criminelle et nécessitent donc des modalités biométriques faciles à enregistrer avec une coopération réduite de la personne. La reconnaissance de visages est l'une des techniques privilégiées dans ce contexte, parce qu'elle est non intrusive et à cause de la relative facilité d'acquisition de photographies, pouvant même être effectuée à l'insu de la personne concernée. Cette facilité d'acquisition des visages rend la constitution d'une base de données assez aisée pour les applications d'identification.

La plupart des approches de reconnaissance de visage qui se fondent sur des techniques d'apprentissage automatique, nécessitent des données d'entraînement volumineuses. Il est de plus, crucial de pouvoir comparer les algorithmes les uns avec les autres, ce qui suppose là encore de pouvoir disposer de données annotées (images pour lesquels les visages sont

localisés et identifiés). En conséquence, de nombreuses bases de données d'images faciales ont été construites et rendues publiques afin que la communauté scientifique puisse les utiliser pour entraîner différents algorithmes et comparer leurs résultats.

En ce qui concerne les images faciales, de nombreuses bases de données existent pour les différents aspects étudiés dans la reconnaissance des visages, ce qui implique l'importance du choix des données utilisées. Une base de données complète doit contenir des images avec de multiples sources de variations possibles telles que l'éclairage, la pose, position de la caméra. Compte tenu de ces contraintes, nous avons tout d'abord choisi trois grandes bases de visages populaires qui sont publiquement disponibles, qui contiennent plusieurs images faciales acquises sous multiples variations d'illumination, de poses et d'expression faciale. Le choix est également motivé par le fait que la plupart des méthodes de reconnaissance utilisent ces bases pour évaluer leurs performances.

4.2.1 ORL

Cette base ORL (Olivetti Research Laboratory) a été collectée entre avril 1992 et avril 1994 par le laboratoire AT&T de l'université de Cambridge [103]. La base contient 40 personnes, chacune étant enregistrée sous 10 vues différentes (figure 4.1). Les images sont de taille 112×92 pixels en format Pgm (portable format de gris). Pour quelques sujets, les images ont été collectées à des dates différentes, avec des variations dans les conditions d'éclairage, les expressions faciales (expression neutre, sourire et yeux fermés...) et des occultations partielles par les lunettes. Toutes les images ont été collectées sur un fond foncé, les poses de la tête présentent quelques variations en profondeur par rapport à la pose frontale.



Fig.4.1- Extrait de la base ORL. Pour chacune des 40 personnes enregistrées, on dispose de 10 vues avec des changements de pose, d'expression et d'éclairage.

4.2.2 FERET

Le programme de technologie d'identification de visage FERET (The Face Recognition Technology) est une base de données volumineuse pour l'évaluation du système de reconnaissance faciale [104]. FERET est géré par la Defense Advanced Research Projects Agency (DARPA) et le National Institute of Standards and Technology (NIST). Elle a été collectée en 15 sessions entre août 1993 et juillet 1996. Feret contient cinq galeries de données : Fa (1196 images), Fb (1195 images), Fc (194 images), Dup1 (722 images) et Dup2 (234 images), chacune contenant 7 images par personnes sous différentes conditions d'expression, de lumière et de posture. Dans ce travail la galerie Fa été utiliser pour l'apprentissage des différentes approches, et les autres jeux de données sont utilisés pour les différents tests [105].



Fig.4.2- Extraits d'images de visage de la base de données FERET.

4.2.3 CASIA-WebFace DATABASE

La base de données CASIA-WebFace [106] est utilisée pour la recherche scientifique sur la reconnaissance faciale sans contrainte. Les images de visage dans la base de données sont extraites d'internet, par l'institut d'automatique de l'académie chinoise des sciences (CASIA). Cette base est annotée avec 10 575 personnes avec plusieurs prises, soit 494 414 images au total. Il s'agit du deuxième plus grand ensemble de données public disponible pour les problèmes de vérification de visage et de reconnaissance [107]. Les images du jeu de données CASIA WebFace incluent des variations aléatoires de poses, illuminations, expressions faciales et résolutions d'image.



Fig.4.3- Extraits d'images faciale de la base de données de CASIA-WebFace.

4.2.4 Séparation des bases de données et matériel utilisé

Les tâches principales de la partie expérimentale consistaient à mettre en œuvre un système de reconnaissance de visages pour tester l'efficacité de notre méthode d'apprentissage en profondeur, particulièrement un réseau de neurones convolutifs. Cette tâche ainsi que l'environnement de mise en œuvre pratique ont déterminé les critères de sélection des approches et des bases de données utilisés pour les expériences. Lors des expérimentations menées dans ce travail, nous n'avions pas accès à une batterie de serveurs ni à un GPU haut de gamme utilisé à des fins de recherche. Pour la réalisation des expériences, un ordinateur portable a été utilisé, doté d'un processeur Intel i7-4700MQ 3.40 GHz, une mémoire RAM de 8Go, une carte graphique dédiée GeForce GTX 840M. Outre la puissance de calcul, les réseaux convolutifs profonds nécessitent une grande quantité de données d'apprentissage. Étant donné que la collecte et l'annotation d'une base de données d'une taille suffisante est laborieuse, pour ses raisons, nous avons exclu la collecte de données nous-mêmes durant les travaux de cette thèse.

La première composante de la création d'un réseau d'apprentissage profond, consiste à rassembler notre ensemble de données initial. Nous avons besoin d'images de visage humain ainsi que des étiquettes associées à chaque image. Ces étiquettes doivent provenir d'un ensemble de classe, le nombre d'images pour chaque classe devrait être approximativement uniforme (c'est-à-dire le même nombre d'exemples par catégorie), Sinon notre classifieur deviendra naturellement plus disposé à privilégier certaines catégories fortement représentées. Le déséquilibre des classes est un problème courant dans l'apprentissage automatique. Afin de développer une application de reconnaissance de visage, il est nécessaire de disposer de deux bases de données : une base pour effectuer l'apprentissage (training) et une autre pour tester (testing) les techniques utilisées et déterminer leurs performances, mais il n'y a pas de règle pour déterminer ce partage de manière quantitative. Il résulte souvent d'un compromis tenant compte du nombre de données dont on dispose et du temps d'apprentissage du modèle.

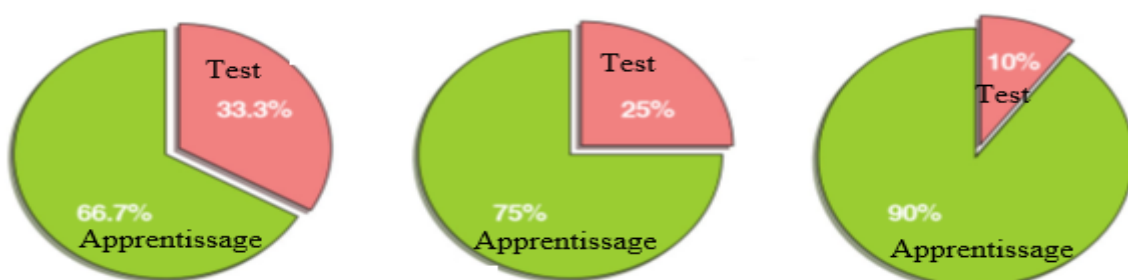


Fig.4.4-Exemples de séparation de données d'apprentissage et de test couramment utilisées.

4.3 Les différentes étapes de notre approche de reconnaissance de visage

L'architecture générale de notre approche avec ces différents processus proposés pour élaborer un système de reconnaissance faciale est illustrée dans la figure 4.5. Un réseau de neurones convolutif (CNN) est d'abord adopté pour l'extraction des caractéristiques des images faciales d'entrée au réseau. Pour obtenir une représentation des visages plus compacte, l'analyse en composante principale (ACP) est utilisée pour réduire le nombre de composantes tout en conservant les informations caractérisant les visages à analyser. Enfin, la matrice des caractéristiques est présentée à l'entrée d'un classifieur support vecteur machine (SVM), pour réaliser la classification des images d'entrées, afin de déterminer la classe d'appartenance des visages. Néanmoins, un réglage des paramètres est primordial pour avoir des résultats plus au moins satisfaisants. La figure suivante représente le schéma global du système de reconnaissance de visage proposé :

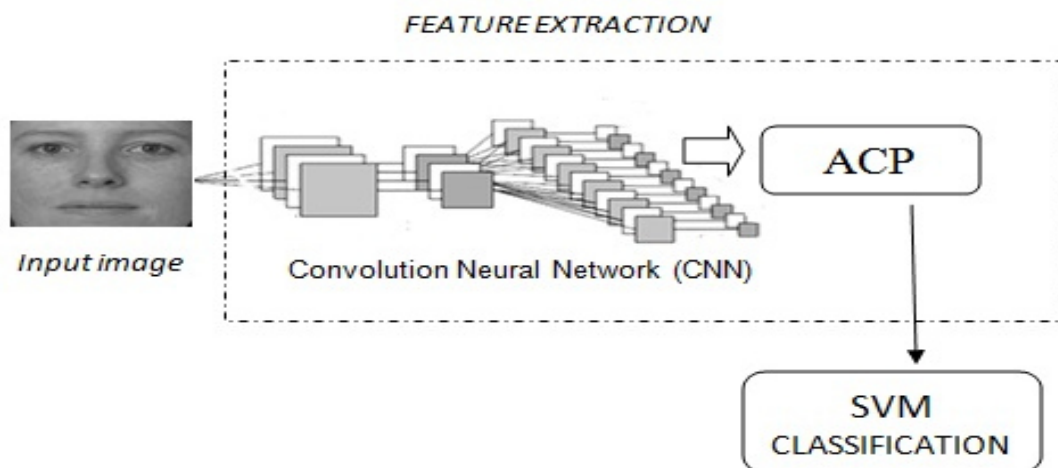


Fig.4.5- L'architecture générale de notre approche de reconnaissance faciale.

De nombreuses variantes d'architectures de réseaux CNN, notamment le nombre de couches, les cartes de caractéristiques et le schéma de connexion, peuvent être adaptées à la tâche de classification de visage. Les figures 4.6 et 4.7, illustrent respectivement le processus d'apprentissage et de test de notre modèle de reconnaissance faciale proposées. Dans notre architecture de réseau de neurones convolutifs, dans un premier temps tous les échantillons sont redimensionnés à 32×32 pixels. Nous utilisons ensuite un modèle de réseau convolutif pré-entraîné pour la reconnaissance faciale, entraîné sur une base de données pour la reconnaissance faciale autre que celle utilisée pour la tâche de reconnaissance faciale. Dans ce travail, nous utilisons une partie du jeu de données CASIA Webface [108] pour former notre CNN et obtenir des poids qui représentent les traits des visages d'apprentissage.

Des études récentes [109] [110] [111] ont montré qu'un modèle pré-entraîné peut-être appliqué à des problèmes de classification similaires. Cette technique nous permet d'éviter le sous-apprentissage du un modèle neuronale. Cela consiste à prendre un réseau déjà entraîné sur un ensemble de données semblables mais plus grand que celui que nous avons à disposition (CASIA Webface). Ce réseau aura donc appris à détecter des formes et des objets dans notre ensemble de données, après il suffit d'adapter la dernière couche du réseau de neurones convolutifs afin que celui-ci classifie correctement les éléments qui nous intéressent. Pendant la phase d'apprentissage, tous les poids restent gelés, c'est-à-dire, ils ne sont pas adaptés à chaque itération. Seuls les poids de la dernière couche (sortie) s'adaptent [111].

Une étape est en fait insérée entre l'extraction des caractéristiques par le CNN et la classification par le SVM. Au cours de cette phase, nous appliquons des techniques de réduction de dimension par l'introduction de l'analyse en composante principale (ACP) pour réduire la taille des vecteurs des caractéristiques et avoir une meilleure représentation des données. Pour la base de données de test, nous utilisons le CNN pré-entraîné pour extraire les vecteurs de caractéristiques faciale et utiliser ces fonctionnalités pour reconnaître toutes les visages par le classifieur SVM.

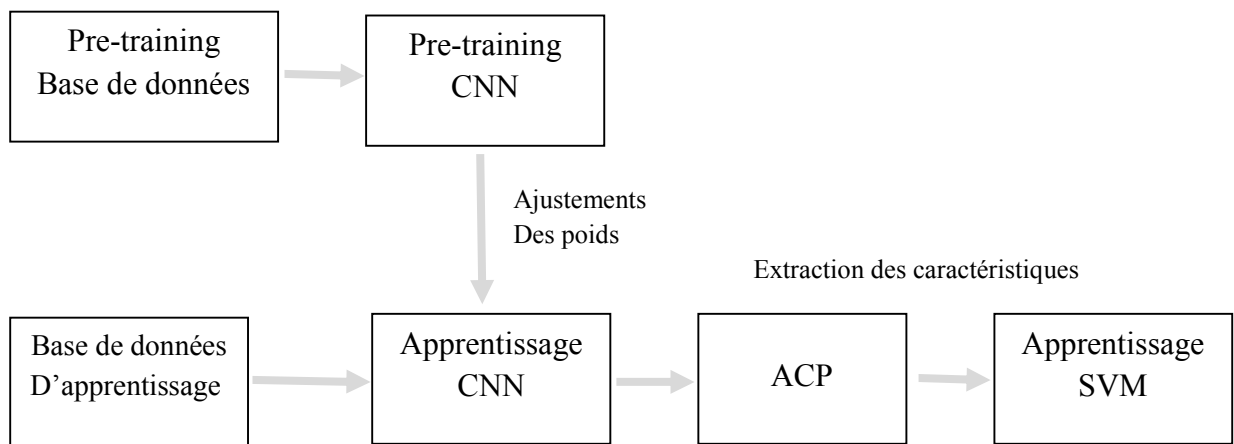


Fig.4.6 - L'ensemble du système d'apprentissage.

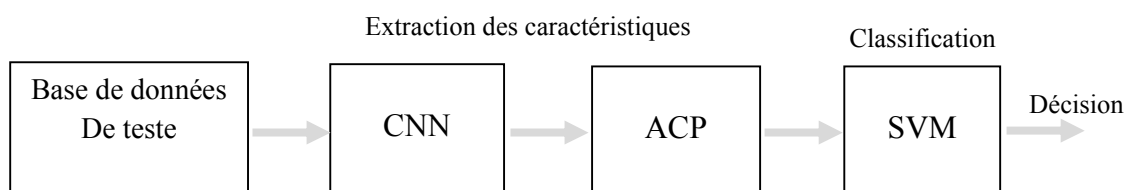


Fig.4.7- Phase de reconnaissance (décision).

4.3.1 Extraction des caractéristiques faciales

4.3.1.1 Convolution et pooling

L'architecture d'un réseaux CNN est constitué d'une succession de blocs de traitement, tels que des couches de convolution, des couches de pooling (sous-échantillonnage) et des couches entièrement connectées. Un réseau neuronal profond (CNN) prend l'image en entrée sans extraction préalable des caractéristiques, ce qui évite grandement la reconstruction massive de données et l'extraction complexe de caractéristiques dans les algorithmes de reconnaissance traditionnels. Le CNN peut effectuer à la fois une extraction et une classification des caractéristiques. La couche d'entrée reçoit des images normalisées de mêmes tailles, ensuite, l'image est convertie sous plusieurs noyaux de filtre à l'aide de poids partagés. Un CNN est composé d'une pile de couches convolutives, qui est paramétrée par le nombre de cartes. On applique un certain nombre de filtres aux données (pixels brutes) d'une image pour extraire les caractéristiques, que le modèle peut ensuite utiliser pour la classification. La carte de caractéristiques de chaque matrice est calculée comme suit :

$$C_n = f(x * W + b) \dots\dots\dots 4.1$$

Où, "*" indique l'opération de convolution, n est le pixel de la carte de caractéristiques, x est la valeur de pixel, W et b sont respectivement le poids et le biais de la couche, f est le noyau de la convolution.

Après cette couche de convolution, nous appliquons immédiatement une couche non linéaire (couche d'activation). Dans ce travail la fonction ReLU, unité de rectification linéaire (Rectified Linear Unit) est appliquée à toutes les valeurs du volume d'entrée pour introduire des non-linéarités dans le modèle.

$$F(x) = \max(0, x) \dots\dots\dots 4.2$$

La couche pooling est un autre point clé de CNN, après la fonction ReLU, nous appliquons une couche de pooling afin de réduire le temps de traitement et la taille des données extraites par les couches de convolution. Cette couche du réseau améliore la robustesse aux variations des images, telles que la rotation, le bruit, la distorsion, et réduit également les dimensions de la sortie. Il existe plusieurs méthodes de pooling. Dans ce travail, nous adoptons le pooling maximale (max-pooling), qui extrait les sous-régions de la carte de caractéristiques à sa valeur maximale. Les couches de convolution et pooling composent la partie extraction des caractéristiques, ensuite les couches entièrement connectées effectuent la classification des entités extraites.

4.3.1.2 La structure du réseau de neurones convolutifs CNN proposée

Ces dernières années, les améliorations rapides des performances du matériel informatique ont bien répondu à la forte demande de traitement de données volumineuses et de réseaux neuronaux profonds. De nombreux chercheurs ont tendance à concevoir des réseaux de neurones convolutifs plus profonds. Les auteurs dans [112] ont conçu un CNN très profond avec 37 couches pour gérer une base de données particulièrement volumineuse, ce qui leur a pris plusieurs jours pour se former. Le CNN conçu dans ce travail ne comporte que neuf couches comme le montre la figure 4.8, comprenant une couche d'entrée, trois couches de convolution, trois couches pooling, une couche entièrement connectée, et une couche de sortie. Les couches Conv1, Conv2 et Conv3 sont des couches de convolution qui consistent respectivement en 30, 60 et 80 cartes de caractéristiques issues de ces différents traitements. Dans les trois couches de convolution, chaque neurone des cartes de caractéristiques est connecté à un champ de réception local 5×5 de la couche précédente. Les couches S1, S2 et S3 représentent les couches de sous-échantillonnage (pooling) dont le nombre de cartes de caractéristiques est égal au nombre de cartes de la couche de convolution précédente. Dans les trois couches de pooling, chaque neurone des cartes de caractéristiques est connecté à un champ récepteur local 2×2 de la couche précédente. Fc représente la couche entièrement connectée avec 512 neurones en sortie, chacun d'entre eux étant connecté à toutes les cartes de S3. La couche de sortie est également une couche entièrement connectée, chaque neurone se connectant à la première couche entièrement connectée.

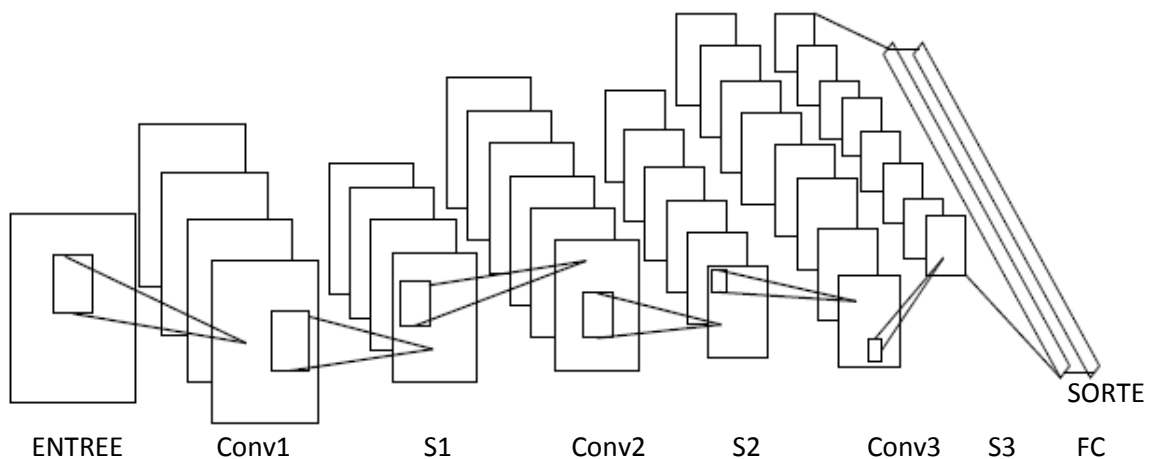


Fig.4.8- Le modèle de réseaux CNN utilisée dans ce travail.

L'architecture de notre réseau CNN, comme illustré dans la figure 4.8, pour la tâche de reconnaissance des visages comprend plusieurs couches. L'entrée du CNN est une image de

visage en niveau de gris de taille 32 x 32 pixels, celle-ci est attribuée à trois couches de convolution qui composent le réseau avec Conv1 constitué de 30 cartes de caractéristiques, Conv2 constitué de 60 cartes de caractéristiques et Conv3 avec 80 cartes de caractéristiques, puis trois couches de pooling sont utilisées, les couches S1, S2 et S3 sont des couches de sous-échantillonnage. Après les couches de prétraitement, on obtient une couches complètement connectées Fc avec 512 neurones, chacun se connecte à toutes les cartes de la dernière couche de pooling suivie par la couche de sortie.

Tab.4.1- Architecture détaillée de notre CNN utilisé pour l'extraction des caractéristiques.

Couches	Type	Taille de sortie
Entrée	Données d'entrée	1×32 ×32
Conv1	Convolution + ReLU	30 5×5 filtres
S1	Max Pooling	2×2, stride 2
Conv2	Convolution + ReLU	60 5×5 filtres
S2	Max Pooling	2×2, stride 2
Conv3	Convolution + ReLU	80 5×5 filtres
S3	Max Pooling	2×2, stride 2
Fc	Fully Connected + dropout	512 units1×1
Sortie	Fully Connected	

Certains ajustements sont utilisés pour optimiser notre réseau CNN. Chaque couche de convolution est suivie d'une fonction d'activation unité linéaire rectifiée (ReLU), étant donné que ReLU peut accélérer l'apprentissage dans des réseaux de neurones profonds et améliorer les performances [111] [112]. Nous appliquons également la technique de dropout pour la régularisation des paramètres d'apprentissage du réseau afin de lutter contre le problème de sur-apprentissage des données d'entraînement. Le dropout améliore les performances de réseaux de taille réduite, en plus les couches entièrement connectées comportent de nombreuses unités et doivent être régularisées. En utilisant cette technique d'optimisation, nous obtenons un meilleur CNN avec des performances stables.

4.3.1.3 Le pré-entraînement du réseau CNN

Afin d'améliorer la capacité de généralisation du réseau, nous introduisons des données auxiliaires pour pré-entraîner notre réseau convolutifs CNN. Le pré-apprentissage ou pré-entraînement (the pre-training) parfois aussi appelé l'apprentissage par transfert d'un réseau de neurones convolutifs est une technique de régularisation pour améliorer la précision de généralisation de notre modèle. Étant donné que le réseau est exposé à une grande quantité de données, les paramètres de pondération sont transférés dans un espace qui est plus susceptible de représenter la distribution des données dans son ensemble plutôt que de surcharger un sous-ensemble spécifique de données. Les CNN partagent le même sort que d'autres réseaux de neurones, il y a trop de paramètres à régler, taille de données d'entrée, nombre de couches cachées, nombre de cartes de caractéristiques par couche, tailles de pooling, taux d'apprentissage, etc. Ainsi, le problème de la sélection du modèle est relativement plus difficile par rapport aux autres techniques d'apprentissage automatique.

Dans ce travail, nous avons sélectionné l'architecture d'un réseau de neurones profond pour la reconnaissance des visages en tant que tâche de pré-entraînement ensuite nous avons examiné la capacité du réseau à extraire et classer les caractéristiques. Le nombre d'échantillons dans une base de données est très important pour l'apprentissage d'un réseau de neurones convolutifs, ces derniers sont souvent entraînés sur de grandes quantités de données pour bien extraire les caractéristiques [110] [113] [114]. Avec l'apprentissage par transfert, au lieu de commencer le processus d'apprentissage à partir de zéro, on part des modèles acquis lors de la résolution d'un problème différent. Nous explorons cette idée en introduisant la base de données CASIA-Webface [107] pour le pré-apprentissage de notre CNN. Nous utilisons des images d'entrée mises à l'échelle 32×32 pour l'apprentissage de notre réseau CNN avec une architecture détaillée dans le tableau 4.1.

Nous introduisons la base de données Casia-Webfaces comme notre base de données auxiliaire pour le pré-entraînement du réseau de neurone, cette base de données dispose de 500 000 images avec 10 000 échantillons dans différentes conditions d'expression, de pose, de lumière, d'âge et autres. Nous sélectionnons une partie des échantillons comme notre jeu de données de pré-apprentissage contenant 10 000 images de visage de Casia-Webfaces. L'ensemble de données de pré-apprentissage est injecté à l'entrée du réseau de neurones convolutif (décrit dans la sous-section 4.3.1.2) et les pondérations du réseau seront mises à jour après un certain nombre d'itérations. Le réseau CNN commence à extraire certaines entités linéaires de la première couche de convolution, puis d'autres caractéristiques de

contour du visage dans la dernière couche de convolution. Comme le montre la figure 4.9. L'idée de l'apprentissage par transfert est que les connaissances acquises à partir d'un problème peuvent être appliquées à des problèmes de classification similaires.

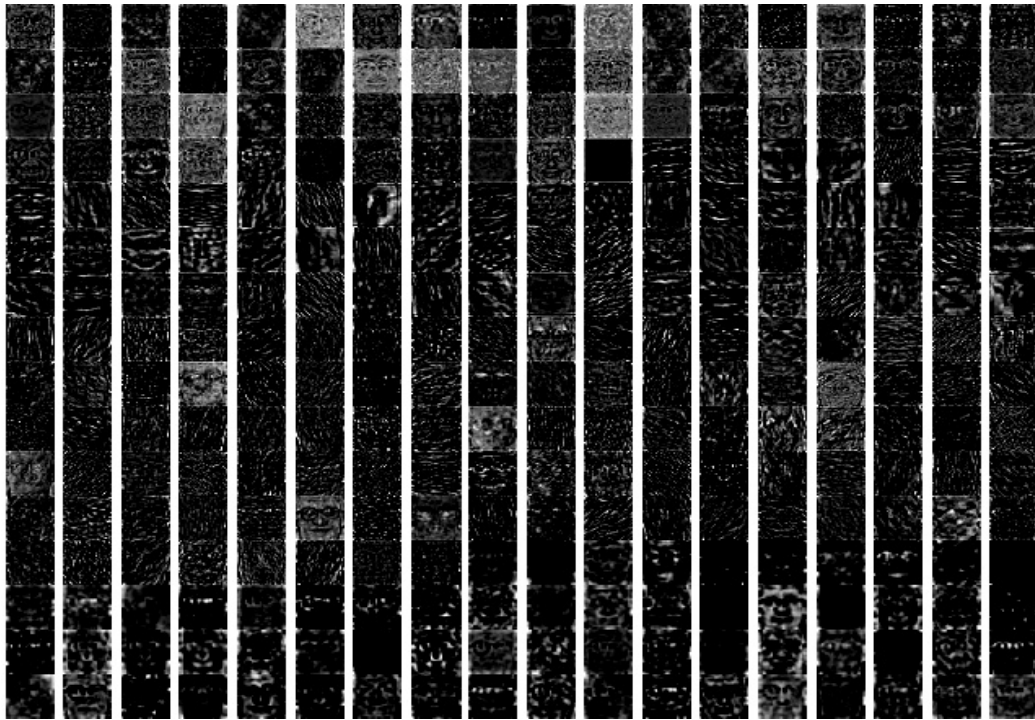


Fig.4.9- Extraction des caractéristiques faciale par pré-apprentissage d'un réseau CNN.

4.3.2 Réduction des dimensions par analyse en composantes principales

L'analyse en composantes principales (Principal Component Analysis (PCA) en anglais) est une transformation linéaire orthogonale qui projette les données dans un espace de dimensions inférieures ou égales d'attributs non corrélées appelées composantes principales, d'où le nom : Analyse en composantes principales. Elle utilise la variance comme une mesure de l'information pertinente et dérive les nouvelles données de manière à maintenir le maximum d'information possible.

Après extraction des caractéristiques par la méthode CNN, les caractéristiques peuvent contenir des dimensions volumineuses conduisant à une redondance des calculs et des informations non discriminantes. Par conséquent, nous adoptons une analyse en composantes principales [110] afin de réduire la redondance, le bruit, et la dimension de vecteurs de sorties avant la phase de reconnaissance faciale. L'ACP est une méthode statistique permettant de rechercher des corrélations entre les caractéristiques et de réduire les dimensions des données. Lorsqu'elle est utilisée sur de visages, les images résultantes sont appelées Eigenfaces.

L'idée de l'ACP est de projeter les vecteurs caractéristiques issue du CNN dans un sous-espace de dimension inférieure. En mappant les données vers un espace de dimension inférieure, cela représente une propriété très intéressante pour l'apprentissage du CNN, car une faible dimensionnalité peut réduire considérablement la mémoire et les calculs.

La méthode des Eigen Faces repose sur une utilisation des premiers vecteurs propres comme visages propres, d'où le terme Eigen Faces. La base formée par ces vecteurs constitue un espace qu'on utilise pour représenter les images des visages. Les personnes se voient donc attribuer un vecteur d'appartenance pour chacune de leur image. Cela étant dit, la reconnaissance est réalisée en comparant les coefficients de projection d'un visage test avec ceux appartenant aux visages d'entraînement. Après avoir rassemblé les visages dans une unique matrice, On calcule ensuite le visage moyen de toutes ses images collectées, qui peut être vue comme le centre de gravité de toutes les matrices. On ajuste ensuite les données par rapport à la moyenne. L'image moyenne est alors soustraite de chaque image de la base de données (Fig. 4.10).

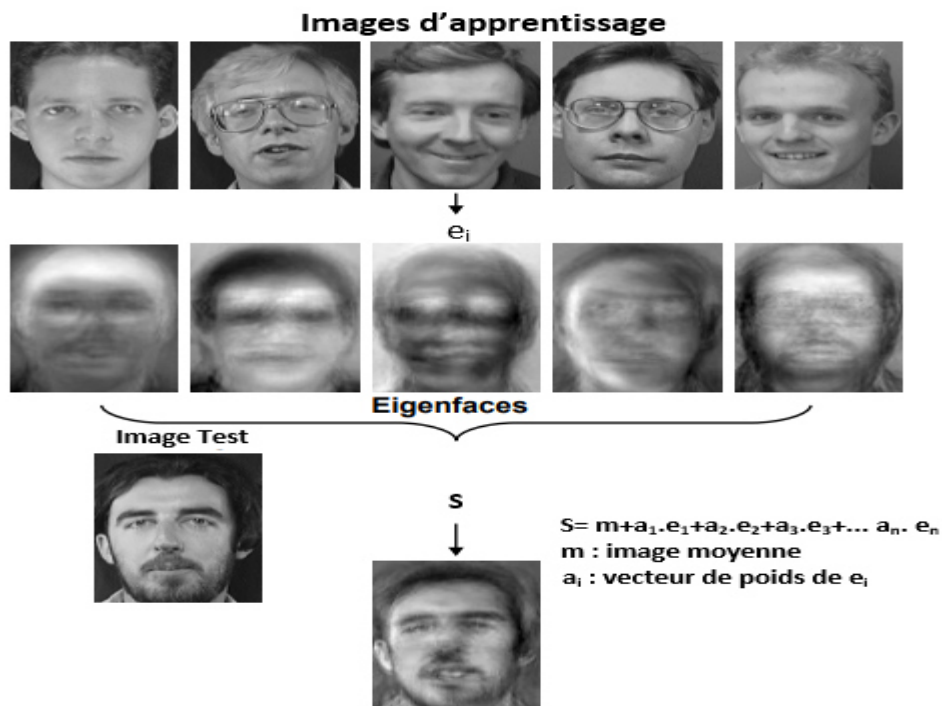


Fig.4.10- Schéma de base d'un système basé sur l'ACP.

Le nombre possible d'eigenfaces est égal au nombre d'images de visage de l'ensemble de la base d'apprentissage. Néanmoins les visages peuvent être approximés, en utilisant seulement les meilleurs Eigenfaces (ayant les plus larges valeurs propres qui représentent en fait, la plus grande variance dans l'ensemble d'images de visage), permettant ainsi de diminuer les calculs.

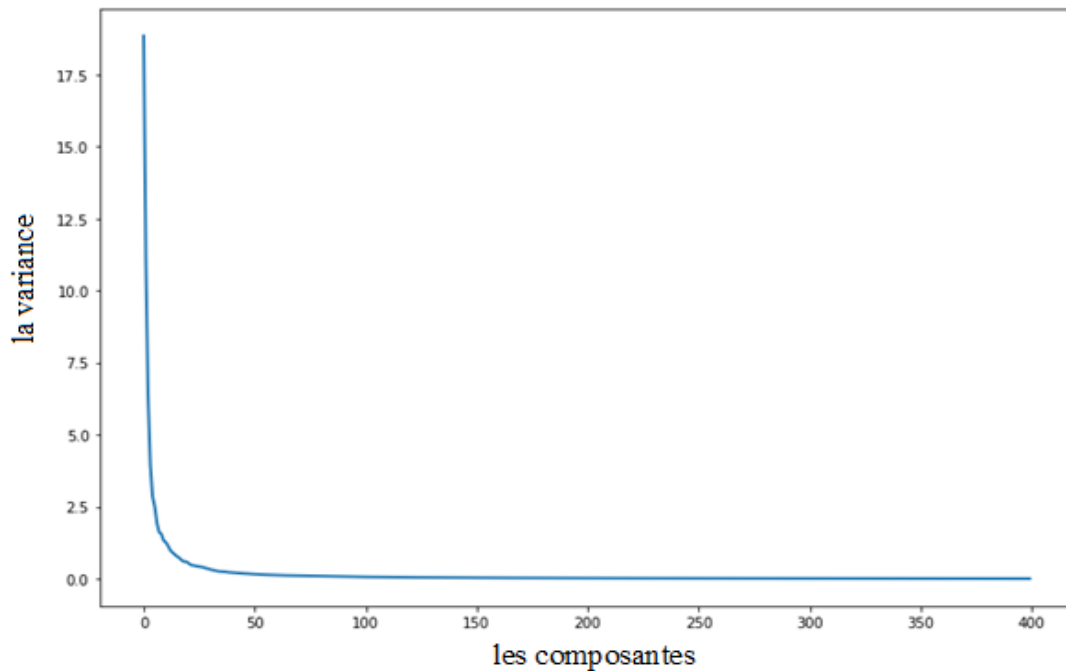


Fig.4.11- Le nombre optimal de composantes principales.

La figure ci-dessus indique le nombre optimal de composantes principales, lorsque nous interprétons le diagramme de variance, il est nécessaire d'obtenir le nombre de composants correspondant à une variance nulle. La figure ci-dessus montre qu'avec juste 100 composants ACP, on peut représenter la totalité des données de la base de visages ORL. Cela correspond aux vecteurs propres ayant les plus grandes valeurs propres.

Algorithme : Calcul de l'ACP

Entrées : matrice X

Sorties : vecteur moyen \bar{x} , base de vecteurs propres U , valeurs propres associées λ_i

Calcul du vecteur moyen :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Normalisation des images d'entrées :

$$\hat{x} = x_i - \bar{x}$$

$$\hat{x} = [\hat{x}_1; \dots; \hat{x}_n]$$

si Données de grande dimension alors

$$M = \frac{1}{n-1} \hat{x} \hat{x}^T$$

Calcul des éléments propres de M :

$$E = [e_1; \dots; e_n]$$

$$\delta = [\delta_1; \dots; \delta_n]$$

Calcul des éléments finaux :

$$u_i = X e_i \quad U = [u_1; \dots; u_n]$$

$$\lambda_i = \delta_i \quad \lambda = [\lambda_1; \dots; \lambda_n]$$

sinon

$$C = \frac{1}{n-1} \hat{x} \hat{x}^T$$

Calcul des éléments propres de C :

$$U = [u_1; \dots; u_n]$$

$$\lambda = [\lambda_1; \dots; \lambda_n]$$

retourner \bar{x}, U, λ

4.3.3 Classification par support vecteur machine SVM

Après avoir extrait les caractéristiques faciales via le réseau CNN, nous utilisons ensuite une technique basée sur les machines à vecteurs de support (SVM : *Support Vector Machines*) comme classifieur final pour la reconnaissance des visages. Les SVMs sont une classe de méthodes d'apprentissage statistique basées sur le principe de la maximisation de la marge (séparation des classes) [115].

Dans ce travail, nous adoptons la fonction SVC (Support Vector Clustering) [116] pour l'étape de classification des vecteurs issus du réseau CNN. L'objectif de cette technique est de partitionner un ensemble de données en plusieurs groupes (classes) afin de les organiser de manière plus significative. SVC [117] peut mapper les points de données sur un espace de grandes dimensions en utilisant un noyau de type gaussienne, qui pourrait trouver une plus petite sphère pouvant entourer tous les points de données. La sphère est renvoyée à l'espace de données, ce qui crée un ensemble de lignes de contour de points de données fermés. Ces points de données fermés par la ligne de contour appartiennent à la même classe (cluster).

Le schéma de combinaison pour définir l'ensemble des sous-problèmes binaires à résoudre passe par la définition d'un principe qui permet de sélectionner la classe la plus vraisemblable à partir des prédictions réalisées par l'ensemble des fonctions de décision binaire. Un sous-problème de discrimination binaire issu d'un problème multi-classe est défini à partir d'un vecteur d'entrée. Ce vecteur représente pour chaque classe de la base initiale sa correspondance avec des exemples considérés comme appartenant à la classe positive (+1), à la classe négative (-1) ou non présent (0) [117]. La décomposition d'un problème multi-classe en N problèmes binaires est la première étape d'un schéma de combinaison binaire. Elle correspond à définir les vecteurs à partir desquelles une matrice de codage du problème multi-classe est définie.

Parmi les méthodes de décomposition, on peut citer les méthodes : un-contre-tous [118], un-contre-un [119], DDAG (Decision Directed Acyclic Graph) [120] et la méthode moitié contre moitié [121]. Dans ce travail, nous utilisons le schéma de combinaison un contre tous. Le principe de cet algorithme est d'utiliser autant de classifieurs qu'il y a de classes. Pour l'apprentissage de chaque classifieur, toutes les observations de la classe considérée sont étiquetées positives (+1), alors que toutes les observations n'appartenant pas à la classe sont considérées négatives (-1). A la fin de cette phase, on dispose de k hyperplans (k égale au nombre de classes).

La méthode *un-contre-tous* (*one against all*) produit autant de problèmes binaires que de classes, chaque problème binaire correspond à la discrimination des exemples d'une classe avec ceux des autres classes, ce qui permet de créer peu de problèmes binaires. L'approche la plus naturelle est d'utiliser cette méthode de discrimination et d'apprendre N fonctions de décision ($\{f_k\} k = 1 \dots N$), permettant de faire la discrimination entre chaque classe de toutes les autres (chaque classe est opposée à toutes les autres), il faut donc poser N problèmes binaires, le k -ième classifieur est destiné à distinguer la catégorie d'indice k de toutes les autres. Pour classifier un visage, on le présente donc à N classifieurs, la décision s'obtient par l'application du principe *winner-takes-all*, l'étiquette retenue est celle associée au classifieur ayant renvoyé la valeur la plus élevée. Chaque classificateur renvoie 1 si la forme à reconnaître appartient à la classe correspondante, -1 sinon. Alors pour reconnaître une forme il faut la soumettre à tous les autres classificateurs, pour prendre une décision de classification, on garde la classe qui a eu la valeur maximale de toutes les fonctions de décision. Donc l'affectation d'un nouveau visage x à une classe C_i se fait par la relation :

$$i = \arg \max_{k=1 \dots N} f_k(x) \dots\dots\dots 4.3$$

Avec cette décomposition, chaque problème binaire est de taille plus faible que le problème initial, ce qui réduit les temps d'entraînement liés à chaque problème de classification. Néanmoins ce gain peut être atténué par l'augmentation du nombre de problèmes binaires à traiter. La discrimination sera plus facile, car la frontière de décision ne concerne que deux classes à chaque fois. Cette approche, aussi simple soit-elle, lorsqu'elle est mise en œuvre avec des SVM correctement paramétrées, permet d'obtenir des performances qui ne sont pas significativement inférieures à celles des autres méthodes de décomposition.

Dans notre système, l'entrée de SVC correspond aux caractéristiques faciales de la couche de sortie dans le réseau CNN après réduction par l'ACP. Le jeu de données d'apprentissage en entrée et la base de données d'essai (test) de SVM sont les entités en sortie de l'ensemble de données d'apprentissage et de l'ensemble de données de test de CNN, respectivement. L'étiquette (label) d'apprentissage et l'étiquette de test de SVM sont respectivement identiques à l'étiquette d'apprentissage et à celle de test de CNN. Nous adoptons la fonction de base radiale (RBF) comme fonction du noyau. Pour la reconnaissance finale, nous utilisons la méthode un contre tous pour concevoir un SVM entre deux classes d'échantillons du jeu de données d'apprentissage et déterminer la classe à laquelle appartient vraisemblablement l'échantillon dans chaque classification. La technique SVC fonctionne mieux pour les données de faible dimension, ainsi une étape de prétraitement à l'aide de l'analyse en composantes principales produit des données qui se regroupent bien.

4.4 Expérimentations

Cette section décrit les expériences réalisées et les résultats expérimentaux de notre système de reconnaissance de visage. Ce travail met l'accent sur la fusion des trois différents approches CNN-ACP-SVC afin de réduire les erreurs de classification et augmenter la précision d'identification faciale, cette précision est matérialisée par le taux de reconnaissance obtenu sur différentes bases de données de visage.

4.4.1 Données utilisées

Les expérimentations ont été réalisées principalement sur la base de données FERET (voir la section 4.2.2). La galerie Fa est utilisée comme base d'apprentissage alors que pour les différents tests, nous utilisons une partie de la galerie Fb qui contient des images prises le même jour que les images de Fa, en utilisant le même appareil photo et dans les mêmes conditions d'éclairage. Dans nos expériences, FERET nécessite une étape de prétraitement par extraction de région d'intérêt (ROI) et un redimensionnement de l'image à 128×128 pixels.

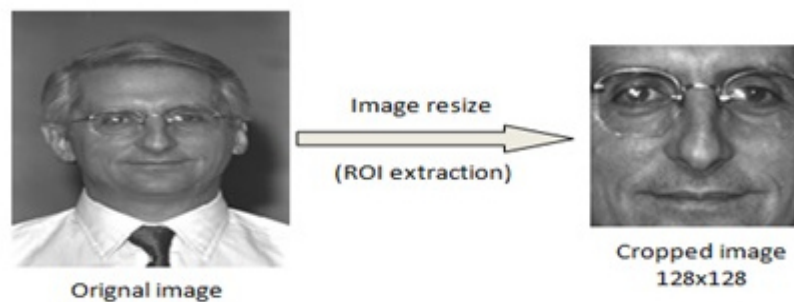


Fig-4.12 Prétraitement et redimensionnement des image de la base FERET.

Nous présentons aussi certaines expérimentations réalisées en utilisant un sous-ensemble de la base ORL. Cette base contient 40 personnes, chacune étant enregistrée sous 10 vues différentes (figure 4.1). Les images sont de taille 112×92 pixels. Dans les séries de test que nous avons effectué la base a été scindé de la façon suivante :

- **Images apprentissages** : Les 6 premières images servent pour la phase d'apprentissage, ce qui fait un sous ensemble de 240 images dédiées à cette phase.
- **Images Tests** : Les 4 dernières images de chaque individu nous ont servies pour la réalisation des différents tests, ce qui nous fait un sous ensemble de 160 images dédiées à la phase test.

4.4.2 Méthodologies des tests, résultats et discussions

Nous avons réalisé une série d'expériences et de tests sur les différentes bases de données utilisées dans ce travail, les résultats obtenus seront présentés, suivis par une discussion les concernant.

-Expérience 1 :

La première expérience, illustre l'influence du nombre d'exemples d'apprentissage sur les performances de la reconnaissance de notre modèle CNN pré-entraîné (décrit dans la section 4.2.2), nous comparons les résultats sur la base de données FERET, afin d'évaluer notre réseau CNN. L'apprentissage d'un CNN consiste à optimiser les coefficients du réseau, à partir d'une initialisation aléatoire, pour minimiser l'erreur de classification en sortie. En pratique, les coefficients du réseau sont modifiés de façon à corriger les erreurs de classification rencontrées, selon une méthode de descente de gradient. Ces gradients sont rétro-propagés dans le réseau depuis la couche de sortie, d'où le nom rétro-propagation du gradient (back propagation) donné aux algorithmes d'entraînement des réseaux de neurones.

Pendant la phase d'apprentissage, chaque cycle complet d'une phase d'entraînement s'appelle « Epoch », l'ensemble de validation est testé à chaque fin d'epoch. Pour s'assurer que le réseau est bien entraîné, il faut tout de même avoir un ensemble de test qui n'est testé qu'une fois que tous les cycles d'entraînement ont été finis. On présente dans le tableau 4.2 les paramètres tels que le nombre d'itérations, fonction de performance, le temps d'apprentissage et l'erreur d'entraînement, sont indiqués pour différents architectures de réseaux testés. L'objet cette partie est de proposer divers paramètres pouvant permettre d'accomplir plus aisément cette phase d'apprentissage.

Tab.4.2- Paramètres d'apprentissage du CNN.

Paramètre	Valeur
Fonction d'activation	Relu
Fonction de perte (erreur)	MSE (0.001)
Taux d'apprentissage (dropout)	0.5
Batch size	100
Nombre d'itérations (epoch) du CNN	50

Nous avons varié la taille des données d'entraînement pour voir l'effet du pré-apprentissage par rapport aux données d'apprentissage du réseau. Des fractionnements de 10 000 et 5 000 exemples ont été sélectionnés de manière aléatoire du jeu de données CASIA webface.

Nous utilisons ensuite un sous ensemble de la base FERET comme données cibles pour la phase de reconnaissance faciale avec un réseau CNN. Compte tenu de la similarité des données cibles et des données auxiliaires, nous pouvons initialiser les poids de CNN, à l'exception de sa dernière couche, avec les poids obtenus sur la base de données Casia-Webfaces. Le réseau neuronal de convolution finale aurait une capacité de généralisation plus forte et un taux de convergence plus rapide. Nous évaluons les modèles pré-entraînés pour une tâche de reconnaissance de visage sur 50 itérations et rapportons le taux de classification sur un corpus de 100 personnes de la base FERET.

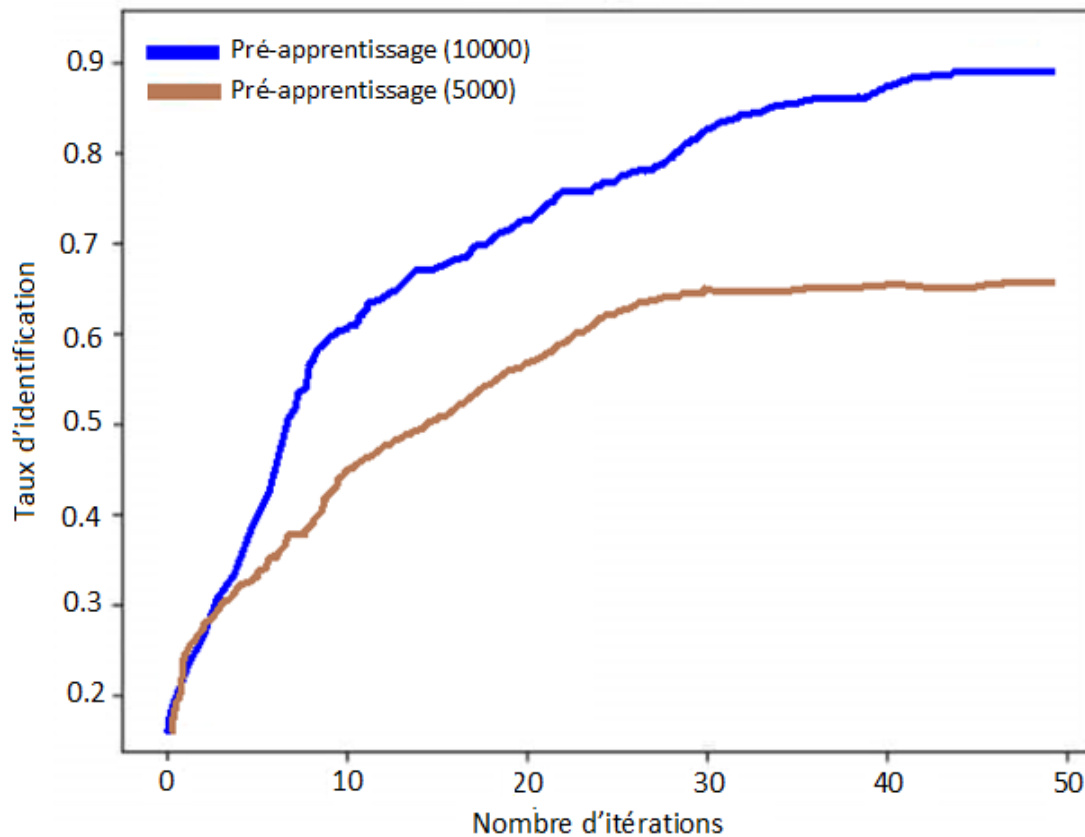


Fig.4.13- Taux d'identification avec les modèles pré-entraînés.

Nous constatons (figure 4.13) que le résultat de l'utilisation du pré-entraînement avec une grande quantité de données dans la phase d'apprentissage (10000 échantillons) est meilleur que lorsque le nombre d'échantillons est limité à (5000 échantillons) en termes de taux de reconnaissance et la vitesse de convergence. Le pré-apprentissage du premier modèle avec 10000 échantillons améliore le taux de reconnaissance de 25% et réduit le temps d'entraînement du jeu de données FERET en comparaison avec le deuxième modèle, qui atteint un taux d'identification de 64% par rapport à 89% pour le premier, ce qui montre que les performances avec un grand ensemble de données sont meilleures.

Les réseaux de neurones convolutifs profonds ont montré des performances significatives dans le domaine de la reconnaissance de forme, y compris le problème de la reconnaissance faciale. Bien que l'apprentissage des modèles CNN montre des performances meilleures avec un ensemble de données volumineux, ce qui confirme qu'ils ne sont pas adaptés à l'apprentissage à partir d'ensembles de données avec peu d'échantillons.

-Expérience 2 :

Dans la deuxième expérience, nous allons utiliser différentes architectures de notre réseau CNN (features-map) sans avoir recours à un préapprentissage, afin d'avoir des performances optimales pour une tâche de reconnaissance de visage. L'architecture du réseau CNN proposée pour le cas de notre étude est composée trois couches, mais le nombre de cartes de caractéristiques sur chaque couche diffère pour chaque domaine et problématique, il dépend également du niveau de complexité de la base de données appliquée. Le nombre de cartes de caractéristiques de chaque couche utilisée dans cette expérimentation, est déterminé de manière optimale à l'aide d'une technique de validation croisée (cross-validation) par 5. La technique de validation croisée est une approche statistique populaire qui permet de concevoir un grand réseau de neurones avec pour objectif une bonne généralisation [122]. Cette technique guide le concepteur dans la sélection du meilleur nombre de neurones ou de cartes caractéristiques sur chaque couche, et du moment opportun pour arrêter l'entraînement.

Par le biais d'une méthode de sélection de modèle par validation croisée, le total des échantillons sera divisé en portions de 20% et 80%. La portion de 20% formera l'ensemble de données de test, tandis que les échantillons restants à 80% sont utilisés pour l'apprentissage. Nous adaptons les recommandations des travaux de [122] [123], dans lesquels ils ont appliqué cette approche au problème de reconnaissance de visage.

Le nombre de sujets utilisés pour l'évaluation de la conception est de 650 sujets de la base de données FERET. La division des échantillons d'entraînement et de test suit un ratio de 80/20, c'est-à-dire 520 et 130 échantillons respectivement. Dans ce travail, la classification est basée sur la règle du gagnant-prend-tout. Les neurones avec la valeur « +1 » correspondent au neurone « gagnant » (qui correspond à la classe correcte) tandis que « -1 » correspond à la mauvaise classe. Dans ce mécanisme, chaque neurone de sortie représente une classe unique du modèle défini. Cette approche est beaucoup plus simple et efficace du point de vue du calcul puisque seule l'opération consistant à une valeur maximale de sortie parmi ces neurones de sortie est requise afin de déterminer la classe de l'image du visage de test.

L'expérience menée dans cette section est mise en œuvre pour trouver le meilleur modèle parmi six architectures différentes en utilisant la technique de validation croisée. Les architectures à tester sont : 15-45-130, 15-46-130, 15-47-130, 15-48-130, 16-45-130, et 16-45-130 respectivement. Ces nombres représentent le nombre de cartes de caractéristiques des couches C1, C2 et C3. La figures 4.14 et le tableau 4.3 montre les résultats de classification et l'erreur de validation des différents modèles. Dans ce travail la fonction Softmax est utilisé à la dernière couche du réseau et l'erreur quadratique moyenne (MSE) est appliquée comme fonction d'erreur.

$$E(k) = \frac{1}{p*m} \sum_{i=1}^p \sum_{j=1}^m (y_i^j - t_i^j)^2, k=1, \dots, s_k \dots\dots\dots 4.4$$

Où, p fait référence au nombre d'échantillons d'apprentissage et m celles de neurones de sortie, s_k représente le nombre d'itération et y_i^j, t_i^j sont respectivement l'actuel et l'erreur désiré des neurones de sorties.

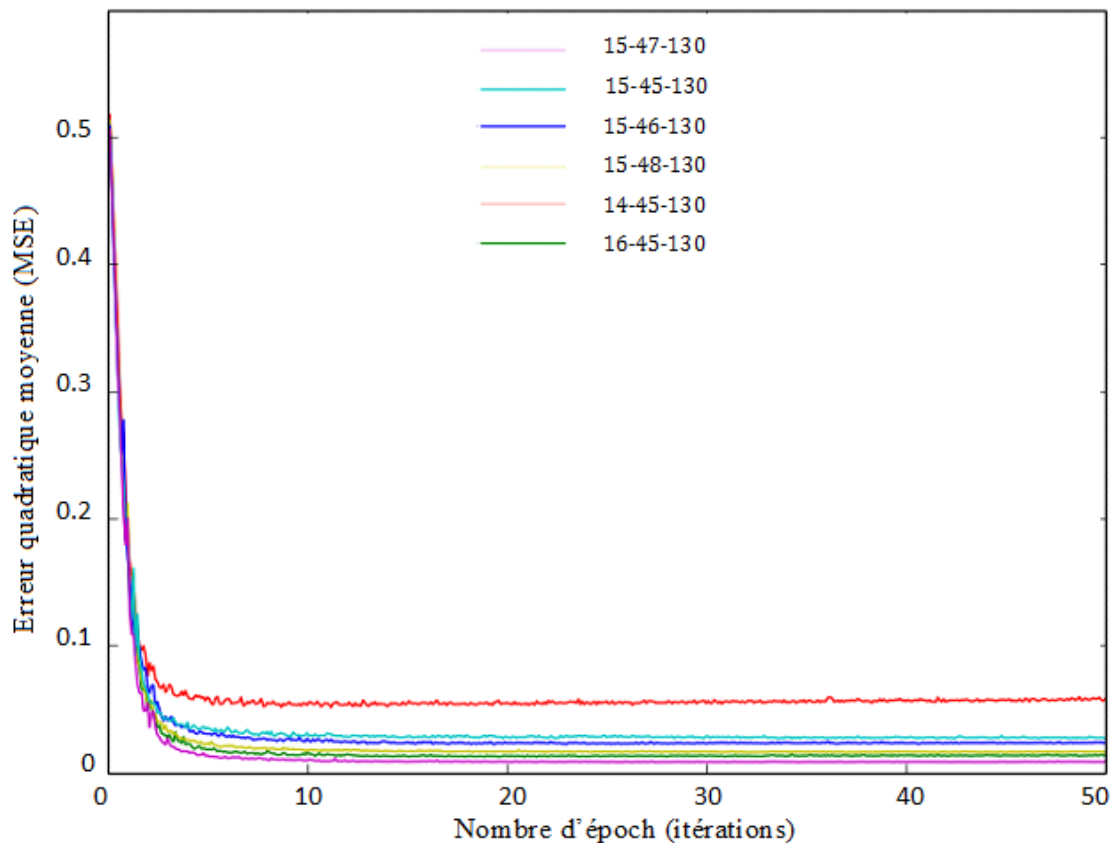


Fig.4.14- Erreur de validation moyenne pour chaque modèle du réseaux CNN (FERET).

Comme illustré sur la figures 4.14, l'erreur de validation se stabilise à partir des itérations 5 à 10. Par conséquent, nous évaluons l'erreur de validation au point 10 epoch. L'erreur de validation la plus faible est indiquée par le modèle 15-47-130 pour la base de données FERET.

Tab.4.3- Performances de classification des modèles testé sur la base de données FERET.

Le modèle	TAUX DE RECONNAISSANCE
14 - 45 - 130	67.2 %
15 - 45 - 130	72.5 %
15 – 46 - 130	78.3 %
15 – 47 - 130	85.5 %
15 – 48 - 130	75.6 %
16 – 45 - 130	66.7 %

La mesure de performance de cette capacité de généralisation est basée sur le pourcentage de classification correcte des échantillons de test. Le tableau 4.3 récapitule les performances des différents modèles de carte de caractéristiques testés sur la base de données FERET. La meilleure précision de 85,50% est obtenue avec le modèle 15-47-130.

Tab.4.4- Configuration expérimentale et performance pour la base de données ORL.

Paramètre du CNN	Valeur et résultat
Architecture (C1-C2-C3 carte de caractéristiques)	5-14-80
Fonction d'erreur	MSE (0.001)
Taille d'image d'entrée	56*46
Nombre d'échantillons d'apprentissage	320
Nombre d'échantillons de test	80
Taux de classification	100 %

Les résultats obtenus par les deux bases de données prouvent que l'architecture CNN à trois couches proposées pourrait être ajustée en fonction du niveau de complexité de la base de données employée. L'exploration nécessaire consiste à ajuster le nombre de cartes de caractéristiques sur chaque couche et déterminé un schéma de connexion optimal pour les deux premières couches de CNN. Nous constatons que les taux de reconnaissance de visage de la base FERET ont considérablement chuté par rapport au résultat de la base ORL (tableau 4.4), ceci peut être expliqué par de l'augmentation du nombre de classes (qui passe de 80 à 130) d'une part, d'autre part cela s'explique aussi par la complexité des échantillons de données.

Les modèles CNNs qui ont été développés pour la reconnaissance faciale présente une même architecture, c'est-à-dire que le nombre de couches est identique, seul le nombre de cartes de caractéristiques de chaque couche est différents d'un modèle à un autres. Parallèlement à la complexité réduite du modèle, l'approche gagnant-prend-tout et le nombre optimisé de cartes de caractéristiques à chaque couche ont contribué à améliorer le temps et le taux de reconnaissance des échantillons de données. Les précisions obtenues de 85,50% (modèle 15-47-130) et 100% (modèle 5-14-80) sur les bases FERET et ORL respectivement, reste dans la même gamme que les autres travaux existants [122] [123], avec une base de données et un nombre de sujets similaires.

-Expérience 3 :

Au cours de cette expérimentation, un apprentissage par transfert de notre réseaux CNN basé sur une fusion avec des techniques ACP-SVM est proposé pour résoudre le problème d'identification de visage et améliorer les taux de reconnaissance. Nous adoptons l'analyse en composante principale pour minimiser les écarts de distribution entre le jeu de données d'apprentissage (CASIA-WebFace) et la base de données de test (FERET), en projetant les entités issues du réseau CNN vers un nouveaux sous-espace de faible dimension. Cette projection des vecteurs de caractéristiques par l'ACP réduit non seulement la redondance et le bruit, mais réduit également la complexité de calcul. Ensuite, la technique machine à vecteur de support (SVM) est utilisé pour la tâche d'identification faciale et donné la classe d'appartenance de chaque visage de la base de données.

Pour valider l'approche proposé, l'ensemble de données d'image de la galerie Fa de la base FERET est utilisé pour l'apprentissage et la galerie Fb pour les différents tests. Après l'extraction des caractéristiques par le modèle pré-entraînés du réseau CNN, dans un premier temps on applique l'analyse en composante principale à toutes les vecteurs issus du réseaux CNN pour la réduction de dimensionnalités de ces derniers. Un classificateur est ensuite nécessaire pour rechercher l'étiquette (label) correspondante pour chaque image test. Différents classificateurs de type machine à vecteur support peuvent être utilisés pour cette tâche d'identification. Dans ce travail, une technique basée sur le support vector clustering (SVC) a été employer, qui est l'une des techniques les plus puissantes pour la classification des images. Dans cette section, l'approche combiner CNN-ACP-SVC est évalué et comparer avec d'autres modèles, les résultats sont résumés dans le tableau 4.5.

Tab.4.5- Taux de reconnaissance des différents approches utilisés dans ce travail (FERET).

Modèles	Taux de reconnaissance
CNN	83,2 %
CNN- linear SVM L1	87.1 %
CNN- linear SVM L2	86.4
CNN-SVC	93.4 %
CNN-PCA-SVC	95.2 %

À la lumière des résultats expérimentaux, nous constatons que le taux de reconnaissance du réseau de neurones convolutif introduit dans ce travail, atteint une performance de classification de 83,2%. Afin d'améliorer ce résultat, nous avons ensuite utilisé différentes méthodes basées sur une classification par machine à vecteur de support pour remplacer la dernière couche de notre CNN. Comme illustré sur le tableau 4.5, un modèle SVC jumelé avec un réseaux CNN permet une performance bien améliorée par rapport à SVM avec une régularisation L1 ou une régularisation L2. Afin d'obtenir une bonne qualité de représentation des visages, et garantir qu'une reconnaissance efficace puisse être menée, une fois que les caractéristiques ont été extraites en utilisant notre réseau CNN, ces dernières sont transformées en une meilleure représentation en introduisant l'analyse en composante principale, les composantes les plus importantes sont ensuite utilisées comme nouvelles caractéristiques. Durant la phase de classification les vecteurs résultants seront finalement passés au classifieur SVC. Les résultats montrent que le taux de reconnaissance de l'approche CNN - ACP - SVC peut atteindre 95,2% en termes de taux de classification ce qui valide les performances de cette technique par rapport aux autres méthodes (tab 4.5). Dans ce travail, la fusion des approches réseau de neurones convolutif, l'analyse en composante principale et le classificateur SVC pour résoudre le problème de reconnaissance de visage permet d'atteindre les meilleurs résultats. Tout d'abord, l'ACP réduit non seulement la redondance, mais l'écart de distribution des données. L'introduction de la technique SVC augmente la distance inter-classes et réduit la distance intra-classes. Enfin, la structure CNN-ACP-SVC avec une technique de pré-apprentissage permet d'obtenir une précision de validation supérieure à celle des algorithmes discutés dans ce travail sur la base de données FERET.

Une étude comparative de la performance de notre approche proposée par rapport à d'autres travaux de reconnaissance de visage antérieurs sur la base de données FERET.

Tab.4.6- Comparaison des performances de certaines méthodes citées dans l'états de l'arts.

Référence	Année	Approche	Taux de classification (FERET)
A. R. Syafeeza et al. [122]	2014	CNN	81.25 %
S. Guo et al. [124]	2016	CNN+SVM	87.29%
Y. Zhang et al. [125]	2016	ACNN	91.67%
CE TRAVAIL	2017	CNN+PCA+SVC	95.5%

À la lumière des résultats comparatifs ci-dessus (tab .4.6), notre approche proposée permet d'obtenir les taux de reconnaissance et de précision les plus élevés, ce qui valide l'architecture proposée de l'ACP-SVC basée sur un réseaux convolutif CNN pour la reconnaissance faciale. On constate que notre approche se positionne correctement par rapport aux autres techniques cités dans l'état de l'art.

4.5 Conclusion

Dans ce chapitre, nous avons proposé une approche neuronale convolutif profonde visant à développer une solution pour la reconnaissance de visage sur différentes bases de données faciales. Dans le but de concevoir un modèle performant pour la reconnaissance de visage, une combinaison de plusieurs approches a été adoptée, dans un premier temps nous avons utilisé une technique pré-apprentissage de notre réseau convolutifs CNN avec certaines données auxiliaires pour augmenter la capacité de généralisation de ce dernier, ce qui améliore considérablement les performances d'extraction des caractéristiques faciales de l'ensemble de données cible. Ces vecteurs caractéristiques passent ensuite par une étape de réduction de dimensionnalités par le biais de la technique ACP, enfin la méthode SVC a été utilisée en tant que classifieur pour effectuer l'identification faciale dans la phase de décision.

Le modèle CNN combiné avec l'ACP et le SVC permet d'obtenir de meilleurs résultats avec un taux de reconnaissance élevé, il atteint d'excellentes performances pour les tâches de classification et de réduction de dimensions. Les expériences effectuées sur les bases de données FERET et ORL confirment l'avantage de notre système. Cette performance est due à la fois à la projection hybride combinant le pouvoir représentatif des réseaux CNN et l'ACP et le pouvoir discriminant du SVC.

Conclusion Générale

Bilan et contributions

Ces dernières années la biométrie a connue un intérêt grandissant et de nombreuses approches ont été développées et proposées par la communauté scientifique, basées sur les empreintes digitales, l'iris, la voix ou le visage. La modalité faciale se distingue des autres, par la simplicité des systèmes d'identification, ne nécessite aucun contact physique ni la coopération de l'individu. Tout au long de la réalisation de ce projet, on a approché un domaine de biométrie très intéressant et d'avenir, avec tous les domaines qui lui sont attachés comme le traitement d'images, vision par ordinateur, reconnaissance de formes et l'analyse de données. Jusqu'à maintenant il n'existe pas de système de reconnaissance de visages performant à 100%, ceci est due à la multitude de contraintes à qui l'image du visage est confrontée, ce qui altère l'information donnée aux techniques de reconnaissances et donc non stabilité des performances.

Cette thèse présente nos travaux sur la reconnaissance de visage faiblement résolus sur des images de base de données volumineuse. Après avoir implanté et évalué des méthodes classiques de reconnaissance faciale, nous avons proposé différentes techniques visant à prendre en compte les spécificités de notre problématique. Les travaux réalisés dans cette thèse s'inscrivent dans le contexte d'identification automatique de visage. Nous avons introduit une approche hybride de reconnaissance faciale mettant en œuvre une classification neuronale basée sur un apprentissage profond, l'analyse en composante principale pour réduire la taille des entités avec lesquelles sont appariés les visages requêtes et une classification par support vecteur machine. L'utilisation conjointe de ces différentes approches permet d'améliorer les performances, tant au niveau de la précision qu'au niveau de la vitesse d'exécution. Dans cette section, nous rappelons dans un premier temps les résultats de ce travail, avant de discuter certaines perspectives.

Les résultats principaux de cette thèse se résument en plusieurs contributions qui sont présentées ci-dessous :

- La thèse décrit et présente un système biométrique de reconnaissance de visage basée sur un apprentissage profond, pour y parvenir, nous avons proposé de nouvelles approches. Les solutions proposées se basent sur l'utilisation des réseaux de neurones convolutifs profonds comme modèle de base d'apprentissage automatique. Ce choix est motivé par leur capacité à

résoudre des problèmes de vision et reconnaissance de forme très complexes grâce à l'extraction de caractéristiques discriminantes de haut niveau. Cependant, ces réseaux ont besoin d'un grand nombre de données d'apprentissage pour être performants.

- Nous avons également étudié l'impact de l'augmentation de la taille de l'ensemble d'apprentissage et l'intégration d'une architecture de réseau par pré-apprentissage (transfert learning) pour améliorer les performances de reconnaissance faciale. Nous avons obtenu une amélioration majeure par rapport à l'utilisation d'une architecture de réseau conventionnelle.

- Nos travaux ont démontré que l'efficacité des caractéristiques extraites par les CNNs dépendent de la complexité du problème cible. Plus particulièrement, les descripteurs des réseaux CNNs se sont avérés beaucoup plus efficaces que les descripteurs conçus par des méthodes classiques (ACP, MLP...). Néanmoins, les CNNs sont souvent entraînés sur de grandes quantités de données pour bien extraire des caractéristiques discriminantes.

- Les résultats de cette étude indiquent que l'utilisation l'analyse en composante principale en vue d'en extraire une information discriminante, permet de réduire la dimension de représentation des données et minimise l'effets de sur-apprentissage. Cette réduction est nécessaire pour rendre les images de visages exploitables par le module de classification, alors que l'extraction de caractéristiques discriminantes permet de présenter les visages à ce dernier de manière plus pertinente.

- Dans un dernier temps, la classification binaire et multi-classes ont été examinées. A cette fin, la méthode utilisée dans ce travail est les machines à vecteurs supports (SVM), en se basant sur l'apprentissage supervisé. Elles sont principalement conçues pour une classification binaire, néanmoins peuvent être adaptées pour une classification multi-classes selon une stratégie de décompositions employer dans ce travail "un-contre-tous". Cette méthode a été appliquée pour la discrimination des visages pour distinguer les personnes sur les différentes bases de données utilisées. De plus, la classification, en utilisant une variantes SVM par clustering SVC, a démontré son efficacité pour une classification meilleur et plus performante.

- La majeure partie de ce travail a porté sur le développement de méthodes performantes d'extraction de caractéristiques d'une part, et la mise en œuvre des réseaux de neurones convolutifs, l'analyse en composante principale et les SVMs pour la classification d'autre part. L'association de ces techniques a permis d'obtenir des taux de reconnaissance de 100% et 95,2%, respectivement pour les bases de données ORL et FERET, ses performances s'avèrent supérieures à celles discutées dans l'état de l'art.

Perspectives de recherche

Les travaux réalisés dans cette thèse, laissent place à de nombreuses perspectives :

- Dans un premier temps une extension de ce travail peut être envisager par l'étude et la réalisation d'un système de détection et de localisation du visage avec des performances assez hautes, une autre consiste à appliquer ce système sur d'autres bases de visages présentant de fortes variations d'éclairage et de pose, ainsi que d'envisager la possibilité d'employer une approche basée sur les éléments locaux du visage.
- Dans un travail futur, nous planifions de généraliser les conclusions de cette thèse a d'autre challenge de la reconnaissance de visage (comme l'expression faciale, la prédiction de genre, l'estimation de l'âge, etc.).
- Un des plus grands challenges dans ce domaine, serait de pouvoir mieux maîtriser les variations d'environnement, qui perturbent encore trop les systèmes de reconnaissance, Les dernières avancées de la technologie de capture 3D des images de visage ont permis de mettre en place des systèmes de reconnaissance assez robustes par rapport à la 2D pour y remédier à ce genre de problèmes.

Si la biométrie est un enjeu important au niveau économique, la recherche, en particulier dans le domaine de la reconnaissance des visages offre encore un champ d'investigations très ouvert.

Bibliographie

- [1]. D. John, Jr. Woodward, C.Horn, J.Gatune, and A.Thomas, Biometrics: A Look at Facial Recognition, documented briefing by RAND Public Safety and Justice for the Virginia State Crime Commission, 2003.
- [2]. S. G. ABABSA. Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D, 2008.
- [3]. S. Prabhakar, S. Pankanti, and A. K. Jain, Biometric Recognition: Security and Privacy concerns, IEEE Security Privacy Mag., vol. 1, no. 2, pp.33-42, 2003.
- [4]. K. Jain, A. Ross, and S. Prabhakar, An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 1, pp.4-19, January 2004.
- [5]. B. Miller, Vital signs of identity. IEEE Spectrum, vol. 31, no. 2, pp 22-30, Feb 1994.
- [6]. P. Verlinde, A Contribution to Multi-Modal Identity Verification Using Decision Fusion. Department of Signal and image Processing, U- Paris, 17 Septembre 1999.
- [7]. J. Daugman. How Iris Recognition Works, IEEE transactions on circuits and systems for video technology, vol.14, no.1, January 2004.
- [8]. B. E. Koenig "Spectrographic voice identification: A forensic survey", FBI, Engineering section, Technical Service Division, 1986.
- [9]. A. Kumar and Y. Zhou. Human identification using knucleocodes. In Biometrics: Theory, Applications and Systems, 2009.
- [10]. Technique de Contrôle d'Accès par Biométrie, <http://www.clusif.fr>, dossier technique, CLUSIF, consulté le 20/01/2019.
- [11]. S. Akrouf, Une Approche Multimodale pour l'Identification du Locuteur, université Ferhat Abbas Sétif, 2011.
- [12]. M. Belahcen, Authentification et Identification en Biométrie. Université Mohamed Khider Biskra ,2013.
- [13]. M. T. Laskri and D.Chefour, Who_Is : système d'identification des visages humains, A R I M A ,Volume 1 ,pages 39 à 61,2002.
- [14]. W. Zhao,R. Chellappa, P. J. Phillips and A. Rosenfeld, Face Recognition: A Literature Survey, ACM Computing Surveys, Vol. 35, No. 4, pp.399–458, December 2003.
- [15]. X. xu, Image Based Face Recognition using Global Features, Research Centre for Integrated Microsystems Electrical and Computer Engineering University of Windsor, May13, 2005.
- [16]. A.Lemieux Systèmes d'identification de personnes par vision numérique ,Université Laval, Québec, Décembre 2003.

- [17]. D. Blackburn, M. Bone, P. J Phillips. Face recognition vendor test 2000, 2001.
- [18]. A. Mellakh, Reconnaissance des visages en conditions dégradées, l'Institut National des Télécommunications de l'université d'Evry-Val d'Essonne, Avril 2009.
- [19]. H. Abdi, L.J. Williams .Principal component analysis; Wiley Interdisciplinary Reviews: Computational Statistics, pp 433-459, 2010.
- [20]. K. Kyungnam, Face Recognition using Principle Component Analysis, Computer Vision and Pattern Recognition, University of Maryland, USA, 1996.
- [21]. M.Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-12(1) :103–108, January 1990.
- [22]. A. M. Martínez and A. C. Kak. PCA versus LDA. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol (2), pp228–233, 2001.
- [23]. B. Scholkopf, A. Smola, and KR Muller. Nonlinear component analysis as a kernel eigenvalue problem. Technical Report No 44, December 1996.
- [24]. J. Lu, K. N. Plataniotis, and A.N. Venetsanopoulos, Face Recognition Using LDA-Based Algorithms, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 14, NO. 1, 2003.
- [25]. H. Yu and J. Yang. A direct lda algorithm for high-dimensional data - with application to face recognition. Pattern Recognition, 34(10) :2067 2070, 2001.
- [26]. J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition. PRL, 26(2) :181191, January 2005.
- [27]. S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.R. Mullers. Fisher discriminant analysis with kernels. Neural Networks for Signal Processing IX, Proceedings of the IEEE Signal Processing Society Workshop, pages 41-48, Aug 1999.
- [28]. S. Mika, G. Rätsch, J. Weston, B. Schölkopf, A. Smola, and K.R. Müller. Invariant feature extraction and classification in kernel spaces, 2000.
- [29]. K. Delac, M. Grgic, S. Grgic, "Independent Comparative Study of PCA, ICA, and LDA on the FERET Data Set", Technical Report, University of Zagreb, FER, 2004.
- [30]. M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. Transactions on Neural Networks, August 15 2002.
- [31]. I. Buciu and I. Pitas. Application of non-negative and local non negative matrix factorization to facial expression recognition. In International Conference on Pattern Recognition, pages: 288–291, 2004.
- [32]. H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana. Discriminative common vectors for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol(1) :pp 4–13, 2005.

- [33]. M. Visani, C. Garcia, and J. M. Jolion. Normalized radial basis function networks and bilinear discriminant analysis for face recognition. In IEEE Conference on Advanced Video and Signal Based Surveillance, pages 342–347, 2005.
- [34]. M.S. Bartlett, H.M. Lades, T. SEJNOWSKI, "Independent component representation for face recognition", Proceedings SPIE Symposium on Electronic Imaging: Science and Technology, vol. 3, pp. 528-539, 1998.
- [35]. M.Benkiniouar, M. Benmohamed. Méthodes d'identification et de reconnaissance de visages en temps réel basées sur AdaBoost ,2005.
- [36]. J.T. Chien; C.P. Liao, Maximum Confidence Hidden Markov Modeling for Face Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.30, n° 4, pp.606-616, 2008.
- [37]. W.Gao, and S. Shan, Face Verification for Access Control, "Biometrics Solutions for Authentication in an E-World", Edited by David Zhang, Kluwer Academic Publishers, Chapter 13, pp339-376, 2002.
- [38]. W.Zhao, R.Chellappa, P.J.Phillips, ACM Computing Surveys , Vol. 35, No. 4, pp. 399–458, December 2003.
- [39]. G. Guo, S.Z. Li, K. Chan, Face Recognition by Support Vector Machines, Proc. Of the IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, pp. 196-201,26-30 March 2000.
- [40]. K. Ouahab, H. Mouss, M.D.Mouss, A. Abdelhadi Reconnaissance des Formes par SVM pour le Diagnostic du Système de Pasteurisation d'une Usine de Lait. Article in Journal of Science and Technology. January 2013.
- [41]. D. Conte, P. Foggia, C. Sansone, M. Vento, "Thirty years of graph matching in pattern recognition", Int. Journal of Pattern Recognition and Artificial Intelligence, vol. 18, n°3, pp.265-298, 2004.
- [42]. A. Albiol, D. Monzo, A. Martin, J. Sastre, A. Albiol, "Face recognition using HOG-EBGM", Pattern Recognition Letters, vol. 29, no. 10, pp. 1537-1543, July 2008.
- [43]. T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. PAMI,24(7) :971_987, 2002.
- [44]. T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(12) :2037–2041, December 2006.
- [45]. X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition Under difficult lighting conditions. In Analysis and Modelling of Faces and Gestures, volume 4778 of LNCS, pages 168 ,182. Springer, 2007.
- [46]. P. S Penev and J. J Atick. Local feature analysis : a general statistical theory for object representation. Network : Computation in Neural Systems, vol(3) :477– 500, August 1996.
- [47]. V. Perlibakas. Face recognition using principal component analysis and log-gabor filters. CoRR, 2006.

- [48]. T. F. Cootes and C. J. Taylor. Constrained active appearance models. In International Conference on Computer Vision, pages 748–754, 2001.
- [49]. A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic face identification system using flexible appearance models. *Image and Vision Computing*, 13(5) :393–401, June 1995.
- [50]. G. J. Edwards and C. J. Taylor and T. F. Cootes. Interpreting face images using active appearance models. In *Automatic Face and Gesture Recognition*, pages 300–305, 1998.
- [51]. E. H. Facci, *Australia Locks in High Technology*, Sydney SECURITY, 2000.
- [52]. M. Kieron, K. Josef, Face authentication competition on the BANCA database, 2004.
- [53]. J. Ah-Pine. Apprentissage automatique. Université -Lyon 2. pp (4-10), 2018.
- [54]. E. Alpaydin. *Introduction to Machine Learning*. MIT Press, 2010.
- [55]. Digital guide. Qu'est-ce-que-l' apprentissage-automatique, <https://www.ionos.fr>, (consulté le 04.02.2019).
- [56]. C. Gagné. Apprentissage et reconnaissance. Université Laval, septembre 2016.
- [57]. <https://www.csee.umbc.edu/courses/471/papers/turing.pdf>, (consulté le 12.12.2018).
- [58]. <https://www.history.computer.org>, Arthur Lee Samuel (consulté 12.12.2018).
- [59]. M. Taffar. *Initiation à l'apprentissage automatique*. Université de Jijel, 2017.
- [60]. F. Chabot. Analyse fine 2D/3D de véhicules par réseaux de neurones profonds, université Clermont auvergne, France, 2018.
- [61]. J. Deng, W. Dong, R. Socher, L. Li, K. Li et L. Fei-fei. Imagenet : A large-scale hierarchical image database. *CVPR*, 2009.
- [62]. R. Collobert, S. Bengio et J. Marthoz. *Torch: A Modular Machine Learning Software Library*, 2002.
- [63]. M. Abadi, A. Agarwal et P. Barham. *Tensor Flow: Large-Scale Machine Learning on Heterogeneous Systems*, 2016.
- [64]. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and T. Darrell. *Caffe: Convolutional Architecture for Fast Feature Embedding*. 2014.
- [65]. A. Gudi. *Recognizing Semantic Features in Faces using Deep Learning*, University of Amsterdam, September 2014.
- [66]. R. M. Ignace. *Détection de concepts et annotation automatique d'images médicales par apprentissage profond*, Université d'Antananarivo, 2018.
- [67]. M.K. Benkaddour, S. Beggari, K. Khamra. *Système de reconnaissance de visage par un réseau de neurone convolutionnel*, Université kasdi merbah de ouargla, Algérie, 2017.

- [68]. Réseaux de Neurones .Concepts Fondamentaux en Statistique. <http://www.statsoft.fr>. (Consulté le 15/02/2019).
- [69]. B.Widrow, and M.A. Lehr. 30 years of adaptive neural networks: Perceptron, adaline, and back-propagation. Proceedings of the Institute of Electrical and Electronics Engineers, Vol. 78, pp. 1415–1442, 1990.
- [70]. D. O. Hebb. The organization of behaviour. Wiley, New-York, 1949.
- [71]. S.Barthélémy,Introduction aux Réseaux de Neurones ,Juin 2000 , www.sylbarth.com. Consulté le 16/02/2019.
- [72]. T. Kohonen, Self-Organizing Maps, vol. 30, Springer Verlag, 1995.
- [73]. D. Arthur, B. Manthey and H. Röglin, Smoothed Analysis of the k-Means Method, Journal of the ACM, vol. 58, n° 5, 2011.
- [74]. A.Chaari. Nouvelle approche d'identification dans les bases de données biométriques basée sur une classification non supervisée, université de Lille 1 France, 2009
- [75]. S.Nebti , Reconnaissance de Caractères Manuscrits par Intelligence Collective, Université Ferhat Abbas Sétif,2013
- [76]. M. Bodn, A guide to recurrent neural networks and backpropagation, the NUTEK-supported project Technical Report, Sweden. 2002.
- [77]. G.Dreyfus, J-M. Martinez, M. Samuelides, MB. Gordon, F. Badran, SThiria , Apprentissage statistique, Eyrolles.2008
- [78]. www.futurasciences.com/tech/definitions/intelligenceartificielle-deep-learning. Consulté le 20/02/2019.
- [79]. R. Giot. Visualisation de réseaux de neurones Profonds, Université de Bordeaux,2016.
- [80]. S. Collet. qu'est-ce-que-le-deep-learning , disponible sur : <https://www.saagie.com>. Consulté le 20/02/2019.
- [81]. Y. LeCun , Y. Bengio and G.Hinton. Deep learning. Nature , vol521, pages 436–444 , May 2015.
- [82]. A.Krizhevsky , I.Sutskever, and G.E. Hinton .ImageNet classification with deep convolutional neural networks. In Proc. Advances in Neural Information Processing Systems ,2012.
- [83]. G.E. Hinton, et al. Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine 29, 82–97 ,2012.
- [84]. K. Janod et al, Auto-encodeurs pour la compréhension de documents parlés. Actes de la conférence conjointe JEP-TALN-RECITAL ,2016.
- [85]. G. E. Hinton, O.Simon, and T.Yee-Why. A fast learning algorithm for deep belief nets. Neural Computation, vol. 18, No. 7, pp. 1527–1554. 2006.

- [86]. G. E. Hinton, and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, Vol. 313, No. 5786, pp. 504–507. 2006.
- [87]. P. Smolensky, Information processing in dynamical systems: Foundations of harmony theory. In *Parallel Distributed Processing*, MIT Press. Vol. 1, No. 6, pp. 194–281. 1986.
- [88]. L. Arnaold, H. Paugam-Moisy et M. sebag, optimisation de la topologie pour les réseaux de neurones profonds, 17e congrès francophone AFRIF-AFIA reconnaissance de forme et intelligence artificiel, Caen,France, 2010.
- [89]. M. Gregory Gelly ,Réseaux de neurones récurrents pour le traitement automatique de la parole , Paris-Saclay -Orsay, septembre 2017.
- [90]. S.Hochreiter,J.Schmidhuber,LongShort-TermMemory , *Neural Computation*, , Vol 9, 1997.
- [91]. A. Krizhevsky, I. Sutskever and G. E. Hinton. “Imagenet classification with deep convolutional neural networks”. *Magazine communications of the ACM*.volume 60, issues 6,pages 84-90,june 2017.
- [92]. Reseade neurones convolutifs. Document disponible sur : <https://dataanalyticspost.com>, Consulté 24/02/2019.
- [93]. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard et L. D. Jackel. Handwritten digit recognition with a backpropagation network. *NIPS*, 1989.
- [94]. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [95]. Neurones à convolution. Article disponible sur : <https://fr.mathworks.com/solutions/deep-learning/convolutional-neural-network.html>. Consulté le 24/02/2019.
- [96]. M. Zaltzhendler. A Deep-Learning Convolutional Neural Network Framework for Multiple Sclerosis Lesion Detection and Segmentation in Patient Brain Images, McGill University, Montreal, Canada, November 2015.
- [97]. R.Lambert .Focus : Le Réseau de Neurones Convolutifs. Document disponible sur : <http://penseeartificielle.fr>. Consulté le 25/02/2019.
- [98]. S. Ioffe et C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ICML*, 2015.
- [99]. J. Delon. Introduction aux réseaux de neurones et à l'apprentissage profond. Université paris Descartes, août 2018.
- [100]. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever et R. Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *JMLR*, 2014.
- [101]. K. Simonyan et A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ICLR*, 2014.
- [102]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich. Going Deeper with Convolutions. *CVPR*, 2015.

- [103]. ORL database: [Http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html](http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html) .
- [104]. FERET Database: <https://www.nist.gov/programs-projects/face-recognition-technology-feret>
- [105]. J. Phillips, H. Moon, S. Rizvi, P. Rauss. The FERET evaluation methodology for face recognition algorithms IEEE Trans. Pattern Anal. Mach. Intell., Vol. 22, pp. 1090-1104, 2000.
- [106]. Casia Webface Database: <http://www.cbsr.ia.ac.cn/english/CASIA-WebFace-Database.html>
- [107]. W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld .Face recognition: A literature survey. ACM Computing Surveys (CSUR), Vol. 35, No. 4, pp. 399-458, 2003.
- [108]. Y. Dong, L. Zhen, L. Shengcai , Z. L. Stan Learning face representation from scratch. arXiv preprint arXiv: 1411.7923,2014.
- [109]. A.R. Chowdhury, T. Y. Lin, S. Maji, E. Learned-Miller. One-to-many face recognition with bilinear CNNs. Applications of Computer Vision (WACV), pp. 1-9. ,2016.
- [110]. M.K. Benkaddour, A. Bounoua. Feature extraction and classification using deep convolutional neural networks, PCA and SVC for face recognition, Traitement du Signal, Vol. 34, No. 1-2, pp. 77-91, 2017.
- [111]. J. Yosinski, J. Clune, Y. Bengio, H. Lipson. How transferable are features in deep neural networks? Advances in Neural Information Processing Systems, pp. 3320-3328,2014.
- [112]. O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” Proceedings of the British Machine Vision, vol. 1, no. 3, pp. 6, 2015.
- [113]. F. Schroff, D. Kalenichenko, J. Philbin. Facenet: A unified embedding for face recognition and clustering. In CVPR,2015.
- [114]. Y. Sun, X. Wang, X. Tang. Deeply learned face representations are sparse, selective, and robust. arXiv preprint arXiv: 1412.1265,2015.
- [115]. E. Osuna, R. Freund, and F. Girosi, Support Vector Machines: Training and Applications, A.I. Memo No. 1602, Artificial Intelligence Laboratory, MIT, 1997.
- [116]. A. Ben-Hur, D. Horn, H.T. Siegelmann, V. Vapnik, “Support Vector Clustering”, Journal of Machine Learning Research 2, pp. 125-137, 2001.
- [117]. J. Saketha Nath, S.K. Shevade, “An Efficient Clustering Scheme Using Support Vector Methods”, Pattern Recognition, 1473-1480, 2006.
- [118]. V. N. Vapnik. « Statistical Learning Theory ». New York, wiley édition,1998.
- [119]. U. H.G. Krebel. Pairwise classification and support vector machines. Advances in kernel methods: support vector learning, pages 255–268, 1999.
- [120]. J. C. Platt, N. Cristianini & J. Shawe-taylor. Large Margin DAGs for Multiclass Classification. NIPS, pages 547–553, 1999.
- [121]. H. Lei & V. Govindaraju. Half-Against-Half Multi-class Support Vector Machines. Multiple Classifier Systems, pages 156–164, 2005.

- [122]. A. R. Syafeeza, M. Khalil-Hani, S. S. Liew, R. Bakhteri. Convolutional Neural Network for Face Recognition with Pose and Illumination Variation, International Journal of Engineering and Technology (IJET), Vol 6 No 1, Feb-Mar 2014.
- [123]. A. R. Syafeeza, M. Khalil-Hani, H. Imran, M. M. Ibrahim and Y. C. Wong. Generalizing convolutional neural networks for pattern recognition tasks. ARPN Journal of Engineering and Applied Sciences. VOL. 10, NO. 12, JULY 2015.
- [124]. S. Guo, S. Chen, Y. Li. Face recognition based on convolutional neural network and support vector machine. Proceedings of the IEEE International Conference on Information and Automation Ningbo, China,2016.
- [125]. Y. Zhang, D. Zhao, J. Sun. Adaptive convolutional neural network and its application in face recognition. Neural Processing Letters, Vol. 43, pp. 389-399,2016.