

Ministère de l'enseignement Supérieur et de la recherche Scientifique

وزارة التعليم العالي والبحث العلمي

Badji Mokhtar Annaba University
Université Badji Mokhtar – Annaba
Faculté des Technologie



جامعة باجي مختار – عنابة

كلية تكنولوجيا

قسم الإعلام الآلي

Département d'informatique

Thèse

Présentée pour obtenir le diplôme de

Doctorat En-Sciences

Spécialité : Informatique

Par :

DJELLAB ISSAM

Thème :

Un multi-classifieur distribué et coopératif pour la reconnaissance des visages humains.

N°	Nom et prénom	Grade	Etablissement	Qualité
01	Ghoualmi Zine Nassira	Prof	Université Badji Mokhtar - Annaba	Président
02	Redjimi Mohammed	Prof	Université 20 Août 1955- Skikda	Rapporteur
03	Azizi Nabiha	Prof	Université Badji Mokhtar - Annaba	Examineur
04	Bekhouche Abdelaali	MCA	Université Abbés Laghrour - Khenchela	Examineur
05	Hemam Mounir	Prof	Université Abbés Laghrour - Khenchela	Examineur
06	Benmohammed Mohamed	Prof	Université Abdelhamid Mehri Constantine 2	Examineur

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

صَدَقَ اللَّهُ الْعَظِيمُ

DEDICACES

A ceux que j'aime ...et ceux qui m'aiment



REMERCIEMENTS

La valeur d'un homme tient dans
sa capacité à donner et non dans
sa capacité à recevoir.

ALBERT EINSTEIN - PHYSICIEN

Ce travail n'aurait pu voir le jour sans l'aide et le soutien de nombreuses personnes.

Je voudrais tout d'abord exprimer ma profonde gratitude à Pr Redjimi Mohammed, Professeur l'Université 20 Aout 1955-Skikda et mon directeur de thèse, dont les conseils éclairés m'ont guidé tout au long de cette recherche. Je remercie également Dr Laimeche Lakhdar, pour son soutien constant et sa disponibilité.

Mes remerciements vont également à Pr Ghoualmi Zine Nassira, présidente du jury, ainsi qu'aux membres du jury Pr Azizi Nabiha, Pr Hemam Mounir, Dr Bekhouche Abdelaali et Pr Benmohammed Mohamed, pour l'intérêt qu'ils ont porté à mon travail en acceptant la lourde tâche d'examineur.

Je voudrais exprimer ma plus profonde gratitude à ma femme, qui a su supporter avec patience mes longues heures de travail et mes moments de doute. Ses encouragements constants et son amour inconditionnel ont été indispensables à la réussite de ce projet. Je remercie également mes enfants, Chihab Eddine et Rassim, dont les sourires ont illuminé mes journées et m'ont donné la force d'avancer. Enfin, je dédie ce travail à mes parents, qui ont toujours cru en moi et m'ont offert tout leur soutien. C'est grâce à eux que j'ai pu réaliser ce rêve.

ISSAM

Résumé

Ces dernières années, les chercheurs ont examiné diverses approches de combinaison de données pour la reconnaissance faciale, ouvrant ainsi une nouvelle voie d'exploration visant à améliorer la fiabilité de la reconnaissance en capitalisant sur la synergie inhérente à diverses sources de données. Ce travail met en œuvre une comparaison complète entre deux méthodes de combinaison basées sur le niveau de score et le niveau de caractéristique, afin de déterminer quelle méthode améliore considérablement les performances globales du système. Dans la méthode initiale appelée "Combinaison de Classifieurs basée sur l'apprentissage profond", nous introduisons une nouvelle règle de fusion basée sur la combinaison au niveau du score. Ce modèle novateur comprend trois Classifieurs, chacun formé à l'aide de techniques bien établies d'extraction de caractéristiques : les motifs binaires locaux (LBP), l'histogramme des gradients orientés (HOG) et les descripteurs faciaux binaires compacts (CBFD). Au lieu de suivre des règles de combinaison conventionnelles, telles que le vote majoritaire ou les scores maximaux, les scores dérivés de chaque Classifieur sont fusionnés, puis formés à l'aide d'un Classifieur Perceptron Multicouche (MLP) pour parvenir à la décision finale. Dans la méthode suivante, appelée "Reconnaissance Faciale basée sur l'Apprentissage Profond avec CNN", nous extrayons des caractéristiques de haut niveau à partir de plusieurs régions d'images considérées comme des données séquentielles en utilisant un ensemble de Réseaux de Neurones Convolutifs (CNN). Dans ce schéma, les couches entièrement connectées de chaque région d'images basée sur CNN sont combinées et alimentées dans un Réseau de Neurones Profonds (DNN) conçu pour la reconnaissance faciale. Les résultats expérimentaux obtenus à partir de bases de données faciales bien connues, notamment Labeled Faces in the Wild (LFW), Olivetti Research Laboratory (ORL) et IARPA Janus Benchmark-C (IJB-C), mettent en évidence les performances compétitives du modèle de combinaison multi-Classifieur proposé et du modèle d'apprentissage profond par rapport aux méthodes de pointe.

Mots-clés: Combinaison de classifieurs, Apprentissage profond, Ensemble de CNN, Reconnaissance faciale, Apprentissage automatique (Machine learning).

Abstract

In recent years, researchers have investigated various approaches to data combination for face recognition, opening up a novel path of exploration aimed at enhancing recognition reliability by capitalizing on the synergy inherent in diverse data sources. This work implements a comprehensive comparison between two combination methods based on the score-level and feature-level combination, to determine which method highly improves the overall system performance. In the initial method called Fusion-based Classifier Combination, we introduce a new fusion rule based on a score-level combination. This novel model comprises three classifiers; each trained utilizing well-established feature extraction techniques: Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG) and Compact Binary Facial Descriptors (CBFD). Instead of adhering to conventional combination rules, such as majority vote or maximum scores, the derived scores from each classifier are merged and then trained using a Multi-Layer Perceptron (MLP) classifier to reach the final decision. In the subsequent method, named Sequential CNN deep learning-based face recognition, we extract high-level features from multiple image regions considered as sequential data, employing an ensemble of Convolutional Neural Networks (CNNs). In this scheme, the fully connected layers of each CNN-based image region are combined and fed into a Deep Neural Network (DNN) tailored for facial recognition. The experimental results obtained from well-known face datasets, including Labeled Faces in the Wild (LFW), Olivetti Research Laboratory (ORL) and IARPA Janus Benchmark-C (IJB-C) highlight the competitive performance of both the proposed multi-classifier combination model and the deep-learning model when compared to state-of-the-art methods.

Keywords: *Classifier Combination, Deep Learning, Ensemble CNN, Face Recognition, Machine Learning.*

الملخص

في السنوات الأخيرة، قام الباحثون بالتحقيق في مختلف المناهج لدمج البيانات في مجال التعرف على الوجوه، مما فتح الباب أمام مسار استكشافي جديد يهدف إلى تعزيز موثوقية التعرف على الوجوه عن طريق الاستفادة من التآزر الطبيعي المتواجد في مصادر البيانات المتنوعة، نقوم في هذه الأطروحة بتنفيذ مقارنة شاملة بين طريقتين للدمج تعتمدان على مستوى النقاط ومستوى السمات، لتحديد أي منهن تعزز بشكل كبير أداء النظام بشكل عام. في الطريقة الأولى المسماة "مزج الصفوف الأساسي للمصنف"، نقدم قاعدة جديدة للمزج تعتمد على مستوى النقاط. يتضمن هذا النموذج الجديد ثلاثة مصنفات، تم تدريب كل منها باستخدام تقنيات استخراج السمات المعترف بها جيداً: أنماط البت المحلية (LBP)، و هيستوغرام التدرجات الموجهة (HOG)، ووصفات الوجه الثنائية المضغوطة (CBFD). بدلاً من اتباع قواعد الدمج التقليدية مثل تصويت الأغلبية أو العلامات القصوى، يتم دمج العلامات المشتقة من كل مصنف ثم يتم تدريبها باستخدام مصنف برسيبترون متعدد الطبقات (MLP) للوصول إلى القرار النهائي. في الطريقة الثانية، والمسماة "التعرف على الوجه بناءً على التعلم العميق باستخدام الشبكات العصبية التكرارية التسلسلية"، نقوم باستخراج سمات عالية المستوى من مناطق الصور المتعددة التي تعتبر بيانات تسلسلية، باستخدام مجموعة من الشبكات العصبية التكرارية. في هذا النهج، يتم مزج الطبقات المتصلة بالكامل من كل منطقة صورة تعتمد على شبكة عصبية تكرارية وتغذى شبكة عصبية عميقة (DNN) مصممة خصيصاً للتعرف على الوجوه. تُظهر النتائج التجريبية المطبقة على قواعد بيانات الوجوه المعروفة بشكل جيد، بما في ذلك ("Labeled Faces in the Wild" و "LFW") و ("Olivetti Research Laboratory ORL") و ("IARPA Janus Benchmark-C")، أداءً تنافسيًا لكلا النموذجين المقترحين للدمج متعدد الصفوف ونموذج للتعلم العميق مقارنة بالأساليب الحديثة في مجال التعرف على الوجوه.

الكلمات المفتاحية : دمج المصنفات، التعلم العميق، الشبكات العصبية التلافيفية (CNN)، التعرف على الوجوه، التعلم

الآلي.

TABLE DES MATIERES

Résumé	IV
Liste des figures	XI
Liste des tableaux	XIII
Abréviations	XIV

Introduction générale	1
1. Contexte général	1
2. Objectifs de la recherche	3
3. Organisation du manuscrit	4
Chapitre 1. Biométrie Et Systèmes De Reconnaissance Faciale	5
1.1. Introduction	5
1.2. Définition de la biométrie	5
1.3. Modalités biométriques	6
1.3.1. Biométrie physique (morphologique)	7
1.3.2. La biométrie comportementale	10
1.3.3. Biométries biologiques	13
1.4. Propriétés d'une modalité biométrique	15
1.5. Applications de la biométrie	16
1.6. Structure d'un système biométrique	16
1.6.1. Phase d'enrôlement	17
1.6.2. Phase de reconnaissance	18
1.7. Principaux Modules d'un système biométrique	20
1.7.1. Module d'acquisition (Capteur)	21
1.7.2. Module d'extraction de caractéristiques	21
1.7.3. Module de comparaison /similarité (Matching)	21
1.7.4. Module de décision	21
1.8. Systèmes biométriques uni-modaux et multimodaux	21
1.8.1. Système biométrique uni-modal	22
1.8.2. Système biométrique multimodale	22
1.9. Fusion des données	23
1.10. Stratégies de fusion	24
1.10.1. Systèmes multi-capteurs	24
1.10.2. Systèmes multi-algorithmes	25
1.10.3. Systèmes multi-instances ou multi-unités	26

1.10.4.	Systèmes multi-échantillons	26
1.10.5.	Systèmes multi-biométries	27
1.10.6.	Systèmes hybrides	27
1.11.	Niveaux de fusion	28
1.11.1.	Fusion au niveau des capteurs	28
1.11.2.	Fusion au niveau des caractéristiques	29
1.11.3.	Fusion au niveau des scores (Score level)	29
1.11.4.	Fusion au niveau des décisions (Decision level)	30
1.12.	Conclusion	32
Chapitre 2. Système de Reconnaissance Faciale		33
2.1.	Introduction	33
2.2.	Evolution de la reconnaissance faciale	33
2.3.	Pourquoi utiliser le visage pour la reconnaissance ?	36
2.4.	Domaines d'application de la reconnaissance faciale	36
2.5.	Système de reconnaissance faciale	37
2.5.1.	Détection de visages	38
2.5.2.	Prétraitement des images	38
2.5.3.	Extraction de caractéristiques	39
2.5.4.	Comparaison et correspondance	40
2.5.5.	Décision	40
2.6.	Défis de la reconnaissance faciale	41
2.6.1.	Variabilité des visages	41
2.6.2.	Diversité des bases de données	42
2.6.3.	Complexité des algorithmes de la reconnaissance faciale	42
2.6.4.	Confidentialité et éthique	43
2.7.	Travaux connexes	43
2.7.1.	Approches globales	43
2.7.1.1.	Techniques linéaires	44
2.7.1.2.	Techniques non linéaires	46
2.7.2.	Approches locales	47
2.7.2.1.	Méthodes de reconnaissance faciale basées sur les points d'intérêt	48
2.7.2.2.	Méthodes de reconnaissance faciale basées sur l'apparence locale	50
2.7.3.	Approches hybrides	54
2.8.	Bases de données de référence pour la reconnaissance faciale	60
2.8.1.	Base de données Yale	60
2.8.2.	Base de données LFW	60
2.8.3.	Base de données FERET	61
2.8.4.	Base de visages de chirurgie esthétique (Plastic Surgery Face Dataset)	62
2.8.5.	Base de données AT&T/ORL	63

2.8.6.	Base de données IJB-C	64
2.8.7.	Base de données MS-Celeb-1M	65
2.8.8.	Base de données CASIA-WebFace	66
2.8.9.	Base de données CAS-PEAL	66
2.9.	Conclusion	67
Chapitre 3. Détection de visage et représentation des caractéristiques		68
3.1.	Introduction	68
3.2.	Détection de visage	68
3.3.	Algorithmes de détection Landmark	70
3.3.1.	Méthodes holistiques	70
3.3.1.1.	Modèle d'apparence active	70
3.3.1.2.	Algorithmes d'ajustement	71
3.3.1.3.	Autres extensions	72
3.3.2.	Méthodes locales contraintes	73
3.3.3.	Méthodes basées sur la régression	74
3.3.3.1.	Méthodes de régression directe	74
3.3.3.2.	Méthodes de régression en cascade	74
3.3.3.3.	Méthodes basées sur l'apprentissage profond	75
3.4.	Annotations des points de repère	77
3.5.	Extraction de caractéristiques faciales	79
3.5.1.	Techniques d'extraction de caractéristiques faciales	79
3.5.1.1.	Histogramme des gradients orientés (HOG)	79
3.5.1.2.	Modèles binaires locaux (LBP)	83
3.5.1.3.	CBFD (Descripteurs faciaux binaires compacts)	84
3.5.1.4.	Speeded up robust features (SURF)	86
3.5.1.5.	Scale invariant feature transform (SIFT)	87
3.5.1.6.	Réseau de neurones convolutionnel (CNN)	88
3.5.1.7.	Filtre de gabor	88
3.5.1.8.	Visages propres (Eigenfaces)	89
3.5.2.	Analyse comparative	90
3.6.	Conclusion	92
Chapitre 4. Contributions		93
4.1.	Introduction	93
4.2.	Reconnaissance faciale basée sur l'apprentissage automatique	94
4.2.1.	Étape de prétraitement	95
4.2.2.	Extraction des caractéristiques	97
4.2.3.	Modèle de combinaison des scores	98

4.3.	Approche de reconnaissance faciale basée deep learning	99
4.3.1.	Concept de reconnaissance séquentielle basée sur les régions	99
4.3.2.	Exploitation des relations géométriques	99
4.3.3.	Mécanisme de coopération des classifieurs	100
4.4.	Expérimentations et résultats	102
4.4.1.	Description du protocole du test	102
4.4.2.	Métriques d'évaluation	103
4.4.3.	Évaluation des performances	104
4.4.3.1.	Évaluation de l'approche utilisant l'apprentissage automatique	104
4.4.3.2.	Performances de la reconnaissance faciale basée sur le deep learning	110
4.5.	Évaluation du modèle proposé basée sur les fonctions de perte	114
4.5.1.	Détails de l'implémentation	114
4.6.	Étude Comparative	117
4.7.	Conclusion	119
	Conclusion et perspectives	120
	Bibliographie	122

LISTE DES FIGURES

Figure 1.1	Exemple des caractéristiques biométriques	6
Figure 1.2	Exemple de caractéristiques biométriques du visage.	7
Figure 1.3	Empreinte digitale.	8
Figure 1.4	Modalité biométrique : Iris	9
Figure 1.5	Modalité biométrique : Géométrie de la main.	10
Figure 1.6	Modalité biométrique : Voix.	11
Figure 1.7	Modalité biométrique : Analyse de la démarche.	12
Figure 1.8	Modalité biométrique : Signature.	12
Figure 1.9	Dynamique de Frappe sur le clavier.	13
Figure 1.10	Analyse de l'ADN.	14
Figure 1.11	Modalité biométrique : Veines de la main.	15
Figure 1.12	Système de reconnaissance biométrique.	17
Figure 1.13	Schéma de l'étape enrôlement dans un système biométrique.	18
Figure 1.14	Schéma d'un système biométrique fonctionnant en mode d'authentification.	19
Figure 1.15	Schéma d'un système biométrique fonctionnant en mode d'identification.	20
Figure 1.16	Différentes sources des systèmes biométriques multimodaux.	24
Figure 1.17	Fusion au niveau des capteurs (Sensor level).	28
Figure 1.18	Fusion au niveau des caractéristiques (Features level).	29
Figure 1.19	Fusion au niveau des scores (Score level).	30
Figure 1.20	Fusion au niveau des décisions.	31
Figure 2.1	Architecture générique d'un système de reconnaissance faciale.	38
Figure 2.2	Exemple d'images de la base de données Yale	60
Figure 2.3	Exemple d'images de la base de données LFW	61
Figure 2.4	Exemple d'images de la base de données FERET.	62

Figure 2.5	Exemple d'images de la Base de données Plastic Surgery Face.	63
Figure 2.6	Exemple d'images de la base de données AT&T/ORL.	64
Figure 2.7	Exemple d'images de la Base de données IJB-C	65
Figure 2.8	Exemple d'images de la base de données MS-Celeb-1M .	65
Figure 2.9	Exemple d'images de la base de données CASIA-WebFace [113]	66
Figure 2.10	Exemple d'images de la base de données CAS-PEAL.	67
Figure 3.1	Exemple des points de repère faciaux "landmark".	69
Figure 3.2	Évolution des modèles CLM lors de la recherche d'une image statique. (a) Points de départ, (b) modèles de départ, (c) après l'itération 1, (d) après l'itération 2, (e) après l'itération 3 et (f) points finaux.	73
Figure 3.3	Illustration de détection faciale (points de repère) avec la Méthodes de régression en cascade.	75
Figure 3.4	Structure du modèle CNN (Convolutional Neural Network).	76
Figure 3.5	La numérotation des points de repère (Landmark).	76
Figure 3.6	Annotations des points de repère (20 points)	78
Figure 3.7	Annotations des points de repère (68 points)	78
Figure 3.8	Annotations des points de repère (194 points).	79
Figure 3.9	Gradients calculés pour une section de l'image.	80
Figure 3.10	Division de la fenêtre de détection en cellules. Histogramme généré pour une seule cellule.	81
Figure 3.11	Points d'intérêt détectés avec SURF.	87
Figure 4.1	Combinaison de Classifieurs Basée sur la Fusion des Scores.	95
Figure 4.2	Forme du visage avec 68 points de repère.	96
Figure 4.3	Architecture de reconnaissance faciale basée sur le Deep Learning	100

Liste des figures

Figure 4.4	Les sept régions faciales utilisées dans la méthode.	110
Figure 4.5	Taux de reconnaissance de la méthode proposée, fondée sur l'apprentissage profond, en fonction du nombre de neurones dans le DNN.	112

LISTE DES TABLEAUX

Tableau 3.1	Les numéros des points de repère (Landmark) principaux et secondaires avec définition.	77
Tableau 3.2	Avantages et limitations des techniques d'extraction de caractéristiques.	91
Tableau 4.1	Performances du système proposé avec les configurations optimales de descripteur LBP.	105
Tableau 4.2	Performances du système proposé avec les configurations optimales de l'algorithme HOG.	106
Tableau 4.3	Performances du système proposé avec les configurations optimales de l'algorithme Cbfd.	108
Tableau 4.4	Performances du système proposé basé sur la fusion des scores.	109
Tableau 4.5	Performances des CNNs pour la reconnaissance des régions faciales.	111
Tableau 4.6	Analyse des performances de la méthode proposée basée sur la fonction de perte CosFace et ArcFace.	116
Tableau 4.7	Analyse comparative des méthodes proposées avec les techniques de l'état de l'art.	118

ABREVIATIONS

ANN	Artificial Neural Network
AFIS	Système Automatisé d'Identification des Empreintes Digitales
ASM	Active Shape Model
ADCDA	Active Discriminative Cross-Domain Alignment
ADN	Acide DésoxyriboNucléique
AAM	Modèle d'Apparence Active
B-ICA	Block ICA
BPNN	Réseau Neuronal à Rétro-Propagation
CTLBP	Curvelets and Local Binary Pattern Features
CBFD	Descripteurs Faciaux Binaires Compacts
CNN	Convolutional Neural Networks
CLM	Modèle Local Contraint
DFD	Direction Field Descriptor
DNN	Deep Neural Network
DCT	Discrete Cosine Transform
DLF	Discriminant Linéaire de Fisher
D-LDA	Analyse Discriminante Linéaire Direct
DWT	Transformée en Ondelettes Discrètes
DOG	Différence De Gaussiennes
EBGM	Elastic Bunch Graph Matching
F-LDA	Analyse Discriminante Linéaire Fractionnaire
GWT	Gabor Wavelet Transform
HOG	Histogramme des Gradients Orientés
HMM	Hidden Markov Model
HMM4	Modèle de Markov Caché à quatre états
IEEE	Institute of Electrical and Electronics Engineers

IJB-C	IARPA Janus Benchmark-C
iMORF	Iterative Multi-Output Random Forests
ICA	Independent Component Analysis
LSH	Locality-Sensitive Hashing
LDA	Analyse Discriminante Linéaire
LPP	Projections Préservant la Localité
LFW	Labeled Faces in the Wild
LDP	Motifs Directionnels Locaux
LGBP	Motifs Binaires Gabor Locaux
LSPBPS	Local Steerable Pyramid Binary Pattern Sequence
LMCL	Large Margin Cosine Loss
LBP-WKPCA	Local Binary Pattern et Wavelet Kernel PCA
LSTM	Long short-term memory
MLP	Multi-Layer Perceptron
MFM	Max Feature Map
MTCNN	Multi-Task Convolutional Neural Network
NN	Neural Network
ORL	Olivetti Research Laboratory
PCA	Analyse en Composantes Principales
PIE	P ose, I llumination et E xpression
PHMM	Pseudo Hidden Markov Model
PDV	Vecteurs de Différence de Pixels
RON	Region Offering Network
RFID	Radio Fréquence Identification Distance
RF	forêt aléatoire
RBF	Radial Basis Function
RKPCA	Robust face recognition based on a new Kernel-PCA
RRQR-PCA	Rank-Revealing QR PCA
RoIs	Regions of Interest
SVM	Machines à Vecteurs de Support

SSD	Single Shot Multibox Detector
SURF	Speeded Up Robust Features
SIFT	Scale Invariant Feature Transform
SVD	Décomposition en Valeurs Singulières
WLD	Descripteurs Locaux de Weber
YOLO	You Only Look Once

INTRODUCTION GENERALE

Soignez le commencement, pensez à la fin, la fin viendra sans fatigue. Si vous oubliez le but ,vous succomberez avant la fin.

CHOU KING - *Philosophe chinois*

1. Contexte Général

Au cours des dernières années, des progrès significatifs ont été réalisés dans le développement de systèmes biométriques pour la reconnaissance des personnes, qui utilisent des modalités biométriques physiologiques ou comportementales. Parmi ces modalités, la technologie faciale a gagné en popularité dans les systèmes biométriques en raison de sa capacité à trouver un équilibre entre l'acceptabilité par les utilisateurs et la précision du système. La reconnaissance faciale offre des informations précieuses pour l'identification des individus, notamment leur identité, leur genre, leur origine ethnique, leur âge et leurs expressions émotionnelles.

Dans notre vie quotidienne, nous avons une capacité naturelle à reconnaître rapidement et facilement les individus en fonction de leurs caractéristiques faciales. Cette capacité remarquable s'étend à la reconnaissance des personnes à partir de photographies, car elle reste robuste face aux variations des caractéristiques faciales, des angles de vue, des conditions d'éclairage et des postures. Les humains peuvent facilement gérer des défis tels que les occlusions, les changements d'expressions faciales, les coiffures ou les effets du vieillissement. Cependant, former un système informatique à effectuer la reconnaissance faciale, une tâche dans laquelle les humains excellent, présente un grand défi. Les chercheurs ont mené des études pour examiner la perception humaine de la reconnaissance faciale à partir d'images faciales, en tirant des enseignements des découvertes psychologiques. De plus, l'existence de relations biologiques et de similitudes observables dans les traits au sein de la même famille a encouragé les chercheurs à exploiter ce phénomène et à développer des systèmes de reconnaissance faciale [1].

Un système de reconnaissance faciale est un système biométrique composé de deux processus principaux : l'enregistrement et le test (la reconnaissance), qui impliquent la vérification ou l'identification. Au cours du processus d'enregistrement, il est essentiel de stocker les caractéristiques biométriques faciales d'une personne extraites à partir d'échantillons fiables dans une base de données. Ces caractéristiques stockées sont ensuite comparées aux caractéristiques extraites des traits de la personne dont les caractéristiques biométriques faciales doivent être vérifiées ou identifiées.

Le processus de test comprend deux modes : la vérification et l'identification. Dans le mode de vérification, le système vérifie les caractéristiques biométriques faciales d'une personne en comparant les caractéristiques biométriques faciales capturées avec le modèle biométrique stocké dans la base de données du système. Ce mode implique une comparaison un-à-un pour déterminer l'authenticité de la relation biométrique faciale. D'autre part, dans le mode d'identification, le système vise à reconnaître un utilisateur légitime en recherchant des modèles correspondants fournis par l'utilisateur dans la base de données. Le système effectue des comparaisons un-à-plusieurs pour identifier une entité individuelle ou ne parvient pas à l'identifier si la personne n'est pas enregistrée dans la base de données du système [2].

2. Objectifs de la recherche

Dans la littérature existante, diverses techniques de reconnaissance faciale ont été proposées, qui peuvent être largement classées en deux catégories : les méthodes basées sur des caractéristiques artisanales [3, 4] et les méthodes de classification basées sur l'apprentissage profond [5, 6]. Ces techniques ont obtenu un succès notable en termes d'identification ou de vérification faciale. Cependant, malgré leurs réalisations, il existe encore plusieurs défis à relever dans le domaine de la reconnaissance faciale. Ces défis peuvent être classés en deux types. Le premier type de défis est directement lié aux relations entre les visages. Ces défis incluent la gestion des variations de pose, d'éclairage, d'expressions faciales et d'occlusions. La capacité à capturer et représenter précisément ces relations est cruciale pour des systèmes de reconnaissance faciale robustes. Le deuxième type de défis est lié à l'environnement de la base de données, et comprend des problèmes tels que la disponibilité limitée de données, des bases de données déséquilibrées et des préoccupations liées à la vie privée. Il est essentiel de

relever ces défis pour développer des systèmes de reconnaissance faciale fiables et efficaces.

Bien que les méthodes précédentes, telles que l'extraction de caractéristiques artisanales et l'apprentissage profond, aient montré un potentiel pour surmonter les défis de la reconnaissance faciale, des recherches supplémentaires et des innovations sont nécessaires pour améliorer les performances et la fiabilité de ces systèmes [7]. Dans ce travail, nous proposons deux nouveaux systèmes de reconnaissance faciale, l'un basé sur l'apprentissage automatique et l'autre sur l'apprentissage profond, afin de contribuer à relever ces défis et atteindre nos objectifs.

Le système basé sur l'apprentissage automatique comprend trois composants essentiels. Tout d'abord, il utilise une combinaison de techniques d'extraction de caractéristiques artisanales, y compris les descripteurs de caractéristiques LBP [8] et HOG [9], ainsi qu'une technique d'extraction de caractéristiques apprise appelée Cbfd [10]. Cette combinaison permet d'extraire des caractéristiques faciales pertinentes. Deuxièmement, un ensemble de machines à vecteurs de support (SVM) est utilisé pour générer des scores de reconnaissance faciale pour chaque représentation de caractéristiques. Enfin, un Classifieur Perceptron Multicouche (MLP) sert de modèle de combinaison, intégrant les scores SVM pour déterminer la fusion optimale de l'information.

Le deuxième système, basé sur l'apprentissage profond se compose de trois composants clés. Tout d'abord, il se concentre sur les régions faciales et les traite comme des données séquentielles en utilisant une série de CNN. Cela signifie que la reconnaissance d'une région faciale prend en compte non seulement l'entrée actuelle, mais aussi les connaissances acquises à partir des régions faciales précédemment traitées. Deuxièmement, les couches entièrement connectées les plus efficaces obtenues à partir de la reconnaissance des régions faciales par chaque CNN sont combinées. Enfin, les caractéristiques extraites sont introduites dans un réseau de neurones profond (DNN) à des fins de reconnaissance faciale.

3. Organisation du manuscrit

Outre l'introduction générale dans laquelle nous mettons en relief le contexte et les objectifs traités dans cette thèse, le présent document est structuré en quatre chapitres.

Le premier chapitre présente un état de l'art sur la biométrie ainsi que sur les systèmes biométriques existants dans la littérature, en détaillant le principe de fonctionnement de chaque système biométrique. Les principaux avantages et inconvénients de ces systèmes sont également présentés. Enfin, dans la dernière section de ce chapitre, nous donnons une présentation de la fusion des données et des stratégies de fusion. Nous introduisons les systèmes multimodaux ainsi que les principes de la fusion des données, les stratégies de fusion et les différents niveaux de fusion.

Puisque nous avons choisi le visage comme modalité biométrique pour élaborer un nouveau système de reconnaissance faciale, le deuxième chapitre est donc consacré à un tour d'horizon de la littérature (état de l'art) portant sur la reconnaissance faciale.

Le troisième chapitre traite de la détection de visage avec les algorithmes de détection de points de repère (Landmark). Les techniques d'extraction des caractéristiques utilisées dans la reconnaissance faciale sont abordées en détail dans ce chapitre.

L'évaluation des systèmes de reconnaissance faciale que nous avons proposés fait l'objet du quatrième chapitre de cette thèse. Dans ce chapitre, nous fournissons une explication détaillée des modèles proposés et présentons ainsi qu'analysons les résultats expérimentaux obtenus à partir de nos modèles. En outre, dans le reste de ce chapitre, nous menons une étude comparative entre nos modèles proposés et des méthodes similaires.

Enfin, une conclusion générale, accompagnée des perspectives envisagées, est donnée à la fin de cette thèse.

CHAPITRE

1

La sécurité est le fondement de la tranquillité.

Voltaire. *Écrivain et philosophe français*

BIOMETRIE ET SYSTEMES DE RECONNAISSANCE FACIALE

1.1. Introduction

La biométrie est une technologie d'identification et de vérification qui convertit une caractéristique biologique, morphologique ou comportementale en une empreinte numérique [11]. Actuellement, ces technologies biométriques sont mises en œuvre dans des secteurs variés tels que : sécurité, contrôle d'accès, transactions bancaires, santé, applications mobiles etc...

Les systèmes biométriques utilisent les caractéristiques physiologiques et comportementaux tels que l'empreinte digitale, le visage, l'empreinte de la paume, l'iris ou les caractéristiques comportementales tels que la voix, le style d'écriture et la démarche pour l'identification et la vérification d'une personne. Ces modalités biométriques sont caractérisées par l'universalité, la distinctivité, la collectivité et la performance [12].

1.2. Définition de la biométrie

Le mot "biométrie" provient de la fusion de deux mots grecs anciens : « bios », signifiant « vie », et « metron », qui se traduit par « mesure ». La biométrie englobe tous les processus visant à identifier une personne en mesurant une ou plusieurs de ses caractéristiques physiques, physiologiques ou comportementales telles qu'illustrées dans la Figure.1.1. Le terme « biométrie » faisait référence à l'utilisation de mesures mathématiques en biologie. Aujourd'hui, il désigne un ensemble de techniques,

dispositifs et systèmes qui permettent aux machines de reconnaître des individus et de confirmer ou d'authentifier leur identité [13].

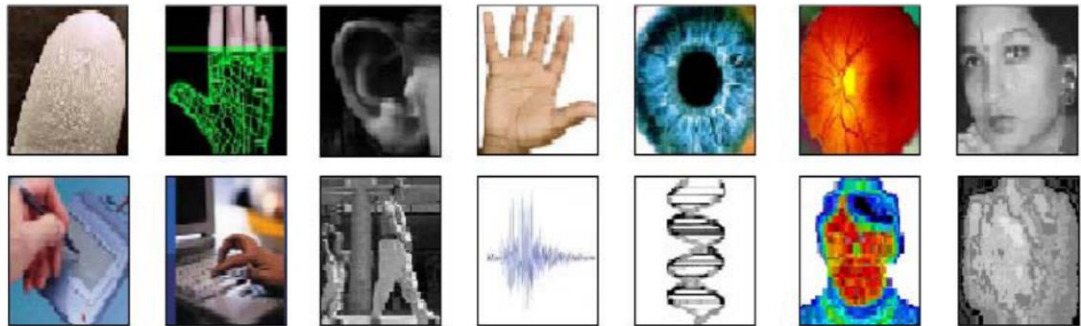


Figure.1.1. Exemples des caractéristiques biométriques

La biométrie présente plusieurs avantages majeurs dans le domaine de l'identification et de l'authentification des individus. Tout d'abord, elle offre un niveau de sécurité élevé grâce à l'utilisation de caractéristiques biologiques et comportementales uniques, ce qui rend la falsification ou la contrefaçon difficile. De plus, la biométrie élimine le besoin de se souvenir de mots de passe ou de porter des cartes d'identification, ce qui réduit le risque de fraude ou d'usurpation d'identité. En outre, elle offre une commodité accrue, car les individus peuvent être identifiés rapidement et facilement sans avoir besoin de porter des dispositifs supplémentaires [14].

1.3. Modalités biométriques

Les modalités biométriques sont généralement classées en deux grandes catégories : les biométries morphologiques et les biométries comportementales. Les premières utilisent des caractéristiques physiques telles que les empreintes digitales ou l'iris, tandis que les secondes exploitent des traits comportementaux comme la signature ou la manière de marcher. Il existe également des biométries biologiques, qui englobent des caractéristiques telles que la thermographie faciale, la structure des veines de la main et l'ADN. Pour être efficaces dans le domaine de la biométrie, les caractéristiques utilisées doivent posséder certaines qualités essentielles pour garantir le développement de systèmes biométriques fiables et robustes [14]. Ces qualités indispensables incluent l'universalité (présence chez tous les individus), l'unicité (capacité à différencier deux

individus), la permanence (stabilité dans le temps pour chaque individu) et la mesurabilité (possibilité de collecte et d'analyse des échantillons).

1.3.1. Biométrie physique (morphologique)

Les techniques biométriques morphologiques ou physiques reposent sur l'identification de caractéristiques morphologiques spécifiques extraites de différentes parties du corps humain, comme l'œil (pour l'iris et la rétine), la main (incluant les empreintes digitales, les lignes palmaires, les articulations des doigts et la géométrie de la main) ou le visage. Cette liste peut être étendue à d'autres éléments tels que la forme de l'oreille ou les vaisseaux sanguins de la main.

1) Visage

Le visage constitue une modalité biométrique largement utilisée dans de nombreuses applications commerciales liées à la sécurité publique [14]. Cela s'explique par son acceptation généralisée par le public, sa facilité d'utilisation, l'absence de participation active de l'utilisateur, son coût modéré, ainsi que sa compatibilité avec les dispositifs existants de capture d'images [15]. Cependant, les images capturées peuvent être bruitées, floues ou de mauvaise qualité. Pour améliorer ces images, des techniques de prétraitement sont appliquées, suivies d'un ensemble de méthodes permettant d'extraire des caractéristiques illustrées dans la Figure.1.2 et de générer des vecteurs pour le processus de reconnaissance [16]. Enfin, cette modalité est vulnérable aux modifications physiques de l'individu, telles que les changements dus à l'âge, aux blessures ou à la chirurgie esthétique, qui peuvent altérer la fiabilité des systèmes de reconnaissance.



Figure 1.2. Exemple de caractéristiques biométriques du visage.

La reconnaissance faciale est utilisée dans divers domaines, notamment pour le contrôle d'accès, qu'il soit physique ou logique. Elle est couramment employée par les

enquêteurs pour identifier les criminels et les corps non identifiés. On la retrouve également dans des lieux tels que les aéroports et certains grands magasins, où elle contribue à renforcer la sécurité. De plus, elle peut être combinée avec d'autres technologies afin d'améliorer son efficacité et sa précision [14].

2) Empreintes digitales

La technique biométrique d'empreintes digitales est largement acceptée car son coût est faible et son efficacité élevée. Un capteur d'empreintes digitales capture l'image de l'empreinte digitale de la personne (voir Figure 1.3). Un traitement d'amélioration de l'image est effectué, suivi de l'extraction des caractéristiques et de la protection du modèle [15]. Les crêtes (points minutieux), les vallées, les îlots et les lacs sont les caractéristiques extraites de l'image d'empreinte digitale. Des vecteurs caractéristiques sont sélectionnés à partir de l'image acquise. De nombreux capteurs tels que les capteurs capacitifs, thermiques et à ultrasons sont utilisés pour acquérir l'image d'empreinte digitale [17].



Figure 1.3. Modalité biométrique : Empreinte digitale.

Cette technologie est très précise pour la distinction des utilisateurs, ce qui en fait un outil hautement discriminant. Elle est compacte, facilitant son intégration dans la plupart des applications telles que les téléphones portables, les ordinateurs, et les systèmes de contrôle d'accès, qu'ils soient physiques (locaux, machines) ou logiques (systèmes d'information). Éprouvée et largement reconnue par le grand public, elle est simple à mettre en place, tout en restant abordable et robuste. Sa précision élevée et son adaptabilité à divers environnements en font une solution performante. Toutefois, certains inconvénients persistent, notamment la nécessité d'un environnement propre et la coopération de l'utilisateur pour un placement correct du doigt sur le lecteur. De plus, l'enregistrement exige un contact physique, ce qui peut soulever des réticences d'ordre psychologique ou hygiénique.

3) Iris

Le système biométrique d'iris est reconnu comme le plus précis parmi les technologies biométriques. L'image de l'iris est capturée, puis des filtres sont appliqués pour en extraire les caractéristiques pertinentes [15]. Des éléments comme les paupières et les cils sont également pris en compte (voir Figure 1.4). L'acquisition de l'image se fait par des éclairages infrarouges, et les étapes du traitement incluent la segmentation, la normalisation, l'extraction des caractéristiques, et enfin la classification [18].



Figure 1.4. Modalité biométrique : Iris

Ce système présente plusieurs avantages : la structure de l'iris demeure inchangée tout au long de la vie, sa texture reste stable, il permet de distinguer les vrais jumeaux avec précision, et offre un potentiel de précision extrêmement élevé. Toutefois, certains inconvénients existent, notamment le coût élevé du matériel, qui nécessite un éclairage spécifique. De plus, l'acquisition des images demande une formation spécifique et de la pratique. De plus, la fiabilité de la reconnaissance diminue à mesure que la distance entre l'œil et la caméra augmente. L'enregistrement peut également poser problème, car il exige que l'utilisateur reste immobile face à la caméra pendant quelques secondes, ce qui peut être dissuasif pour certains. Le processus peut également être inconfortable, limitant parfois l'enrôlement. Ce système est utilisé dans des applications telles que le contrôle d'accès, qu'il soit physique ou logique, ainsi que dans les distributeurs automatiques de billets.

4) Géométrie de la main

La biométrie basée sur la géométrie de la main repose sur plusieurs caractéristiques, telles que la longueur, la largeur, l'épaisseur et la surface de la main ou des doigts de l'utilisateur (voir Figure 1.5) [15]. Cette méthode présente plusieurs avantages par rapport à d'autres techniques biométriques, notamment la petite taille des caractéristiques et des exigences de calcul réduites grâce à l'utilisation d'images à faible

résolution. Bien qu'elle soit largement utilisée aujourd'hui, la géométrie de la main présente certains inconvénients, tels que des coûts élevés et une précision limitée. Des techniques de détection de contours et des méthodes basées sur la texture sont utilisées pour extraire les caractéristiques [19].



Figure 1.5. Modalité biométrique : Géométrie de la main.

Parmi les avantages de ce système, on note son utilisation intuitive, la faible quantité d'espace de stockage nécessaire, son insensibilité à l'humidité ou à la propreté des doigts, ainsi qu'une grande acceptation des utilisateurs. De plus, il offre une haute fiabilité et une rapidité de traitement. Cependant, cette méthode présente également des inconvénients, notamment une faible capacité discriminatoire pouvant entraîner des erreurs chez les jumeaux ou les membres d'une même famille. Elle est également sensible aux changements physiques de la main, dus à des accidents, au vieillissement ou à des maladies comme l'arthrose, ce qui limite la précision et complique l'utilisation chez les personnes atteintes d'arthrite. La biométrie de la main est souvent utilisée pour le contrôle d'accès dans divers lieux, tels que les zones de visite ou les bâtiments privés moins sensibles, y compris les entreprises et différents établissements.

1.3.2. La biométrie comportementale

Les techniques biométriques comportementales se fondent sur l'analyse de certains comportements spécifiques d'une personne, tels que les caractéristiques de sa voix, le style de sa signature, sa manière de marcher [20], ainsi que sa façon d'utiliser une souris ou de taper sur un clavier. En somme, La biométrie comportementale identifie les schémas uniques dans les aspects comportementaux humains [21].

1). Voix

L'empreinte vocale, unique à chaque individu, combine des éléments physiologiques et comportementaux pour l'identifier et le vérifier (voir Figure 1.6) [21]. La

reconnaissance vocale se base sur des caractéristiques telles que la fréquence, la tonalité et la modulation des sons. Bien que la voix puisse distinguer les genres, elle est fortement influencée par la qualité de l'enregistrement et le contenu verbal. L'analyse porte sur divers aspects, comme l'intensité et la forme des ondes sonores, mais la voix peut varier en fonction de l'âge, de la santé ou des émotions de l'individu [20].



Figure 1.6. Modalité biométrique : Voix.

Malgré ces variations, cette forme de biométrie est bien acceptée pour sa simplicité et la possibilité d'utiliser les infrastructures téléphoniques existantes. Cependant, elle reste sensible à l'état physique et émotionnel de la personne, au bruit ambiant, et peut être falsifiée avec un enregistrement vocal [15]. Les applications incluent les forces de l'ordre, les agences de renseignement, les services d'immigration, ainsi que les hôpitaux et systèmes de téléphonie.

2). Analyse de la démarche

La biométrie de la démarche capture et analyse la manière dont une personne marche, en enregistrant les mouvements cycliques de ses jambes, bras et coudes (voir Figure 1.7) [21]. Grâce à une caméra vidéo, ces mouvements sont transformés en données mathématiques, permettant de créer une signature biométrique basée sur la vitesse et l'accélération. Bien que la démarche ne soit pas parfaitement unique, elle reste suffisamment caractéristique pour être utilisée dans des contextes de sécurité faible, notamment pour la surveillance. L'un des principaux avantages de cette méthode est qu'elle peut être détectée à distance avec une caméra à basse résolution et sous différents angles [15, 22].



Figure 1.7. Modalité biométrique : Analyse de la démarche.

Cependant, la démarche peut varier en fonction du poids, des vêtements, des chaussures ou encore des changements physiques, ce qui réduit sa constance et sa fiabilité à long terme. Elle est principalement utilisée dans la surveillance et les installations militaires.

3). Signature

La biométrie de la signature capture les signatures à l'aide de tablettes à stylet (voir Figure 1.8). La vérification peut se faire en ligne, avec une analyse en temps réel, ou hors ligne, à partir d'images pré-traitées [15]. Ces images sont extraites pour obtenir des caractéristiques telles que les coordonnées X et Y, la pression, et le type d'événement [23]. Les avantages incluent la possibilité d'archiver la signature sur des documents et une acceptation élevée par les utilisateurs.



Figure 1.8. Modalité biométrique : Signature.

Cependant, la technologie présente des limites telles que la stabilité considérée comme moyenne à faible, la possibilité de changements dans la signature avec le temps, et des difficultés pour les utilisateurs à s'adapter à la signature sur une tablette graphique. De plus, elle est moins adaptée pour le contrôle d'accès. Cette technologie est principalement utilisée dans les banques, les hôpitaux et les compagnies d'assurance pour l'authentification de documents électroniques [20].

4). Dynamique de frappe

La biométrie de la dynamique de frappe analyse le modèle de frappe d'un individu en capturant des données telles que la vitesse et les erreurs de frappe stilet (voir Figure 1.9). Ces caractéristiques sont traitées à l'aide des méthodes de l'intelligence artificielle pour créer et affiner un modèle de référence, basé sur plusieurs saisies du mot de passe de l'utilisateur. Cette technologie est utilisée pour authentifier les individus lors des tests en ligne [24].



Figure 1.9. Dynamique de Frappe sur le clavier.

Ses avantages incluent une bonne acceptation par les utilisateurs, une amélioration de la sécurité et l'utilisation de gestes naturels avec le matériel existant. Cependant, elle est sensible à l'état physique de l'utilisateur comme l'âge, les émotions et la fatigue [20]. Elle est principalement appliquée dans les cours en ligne, ainsi que dans le secteur bancaire, les compagnies d'assurance, et le traitement de documents administratifs.

1.3.3. Biométries biologiques

Dans le domaine de la biométrie, l'étude des traces biologiques représente une catégorie importante de la reconnaissance automatique d'individus. Elle se concentre sur l'analyse et l'interprétation de caractéristiques biologiques uniques à chaque individu, telles que l'odeur, le sang, la salive, les cheveux, l'ADN et les empreintes veineuses des mains, afin d'identifier ou de vérifier l'identité d'une personne.

1). Analyse de l'ADN

Les avancées en génétique ont permis de mettre en lumière l'ADN comme la caractéristique la plus stable et fiable pour identifier un individu de manière unique et précise. En effet, l'ADN, présent dans chaque cellule du corps humain, conserve sa structure inchangée tout au long de la vie. Les systèmes biométriques ADN impliquent l'identification d'une personne en utilisant son ADN comme caractéristique (voir figure

1.10) [25]. Le code ADN est composé de quatre bases : adénine (A), guanine (G), cytosine (C) et thymine (T). L'ADN peut être extrait de la salive, du sang et des cheveux. Le coût d'investissement pour la mise en place d'un laboratoire de biométrie ADN est élevé. Le système biométrique ADN aide dans les cas criminels pour retrouver la victime et dans le diagnostic médical.



Figure 1.10. Analyse de l'ADN.

L'analyse ADN offre une identification précise et fiable des individus, ce qui facilite l'identification des coupables dans le cadre d'enquêtes criminelles. Cependant, elle présente des inconvénients, notamment un coût élevé en raison des réactifs, des équipements, et de l'expertise nécessaire, ainsi qu'un délai d'analyse pouvant s'étendre sur plusieurs jours, voire semaines. Les applications de cette technologie sont variées, allant de l'identification des auteurs de crimes et des tests de paternité à la reconnaissance des victimes et au diagnostic médical.

2) Veines de la main

Le réseau veineux de la main présente une caractéristique remarquable, il est unique à chaque individu et demeure stable tout au long de la vie. Cette singularité s'explique par la complexité et la variabilité des embranchements des veines, formant un motif propre à chaque personne. La technologie d'identification par les veines de la main repose sur la capture et l'analyse de ce motif unique. L'utilisateur place sa main dans un capteur de lumière infrarouge ou un gabarit de lecture, permettant à un système de capturer une image détaillée des veines (voir Figure 1.11). Cette image est ensuite enregistrée pour une comparaison ultérieure. Les caractéristiques extraites sont des lignes noires. Les veines à l'intérieur des mains sont internes au corps et il est difficile de les falsifier [26, 27].



Figure 1.11. Modalité biométrique : Veines de la main.

Les systèmes biométriques basés sur les Veines de la main offrent une sécurité renforcée contre la fraude, car ils sont difficiles à falsifier, et présentent un taux de fiabilité élevé. Cependant, leur mise en place peut être coûteuse. Ils sont largement utilisés pour assurer la sécurité des transactions financières, le contrôle d'accès sécurisé, et l'authentification des utilisateurs.

1.4. Propriétés d'une modalité biométrique

Les propriétés d'une modalité biométrique se réfèrent aux caractéristiques intrinsèques de la mesure utilisée pour identifier ou authentifier les individus. Ces propriétés sont essentielles pour évaluer l'efficacité et la fiabilité d'une modalité biométrique dans différents contextes d'application. Voici quelques-unes des propriétés clés d'une modalité biométrique :

- 1) Universalité :** Il s'agit de la capacité de la modalité biométrique à être présente chez chaque individu de manière unique et distinctive. Une modalité biométrique universelle est présente chez tous les individus, ce qui la rend potentiellement applicable à grande échelle.
- 2) Distinctivité :** La distinctivité d'une modalité biométrique se réfère à sa capacité à produire des caractéristiques uniques et discriminantes pour chaque individu. Une modalité biométrique distincte permet de différencier efficacement entre les individus, réduisant ainsi les risques de confusion ou d'erreur lors de l'identification.
- 3) Permanence :** La permanence d'une modalité biométrique fait référence à sa stabilité au fil du temps. Une modalité biométrique permanente maintient ses caractéristiques essentielles relativement constantes tout au long de la vie d'un individu, ce qui garantit une fiabilité à long terme dans les systèmes biométriques.
- 4) Mesurabilité :** La mesurabilité d'une modalité biométrique se réfère à sa capacité à être facilement acquise et mesurée de manière fiable à l'aide de technologies appropriées.

Une modalité biométrique mesurable permet une acquisition précise et efficace des données biométriques, facilitant ainsi leur utilisation dans les systèmes d'identification.

- 5) **Acceptabilité** : L'acceptabilité d'une modalité biométrique est liée à la volonté des individus à coopérer avec le processus de collecte des données biométriques. Une modalité biométrique acceptable est perçue comme étant non intrusive, respectueuse de la vie privée et socialement acceptable par les utilisateurs potentiels.

En évaluant ces propriétés, les concepteurs de systèmes biométriques peuvent choisir la modalité biométrique la mieux adaptée à leurs besoins spécifiques tout en assurant des performances optimales en termes d'identification et d'authentification des individus.

1.5. Applications de la biométrie

La biométrie est utilisée dans une grande variété d'applications, notamment:

- *Contrôle d'accès*: pour limiter l'accès aux bâtiments, aux zones sécurisées et aux appareils électroniques.
- *Authentification des utilisateurs*: pour s'assurer que les personnes qui accèdent aux systèmes et aux données sont bien celles qu'elles prétendent être.
- *Suivi du temps et de la présence*: pour enregistrer les heures d'arrivée et de départ des employés.
- *Transactions financières*: pour autoriser les paiements et les transferts d'argent.
- *Enquêtes criminelles*: pour identifier les suspects et les victimes.

1.6. Structure d'un système biométrique

Les systèmes biométriques s'imposent de plus en plus en tant qu'outils incontournables pour l'identification et la vérification des individus. Ils reposent sur l'analyse de caractéristiques biologiques ou comportementales uniques à chaque personne pour confirmer son identité. La structure d'un système biométrique typique se décompose en deux phases distinctes: la phase d'enrôlement et la phase de reconnaissance, comme illustré dans la figure 1.12. Un élément essentiel dans la conception de ces systèmes réside dans la méthodologie d'identification de l'individu. Le système peut fonctionner en mode de vérification ou en mode d'identification [14]. Le processus initial implique la collecte des données biométriques de l'individu, suivi de l'analyse pour extraire des

caractéristiques spécifiques de ces données, puis la comparaison de ces caractéristiques avec les modèles préexistants dans la base de données.

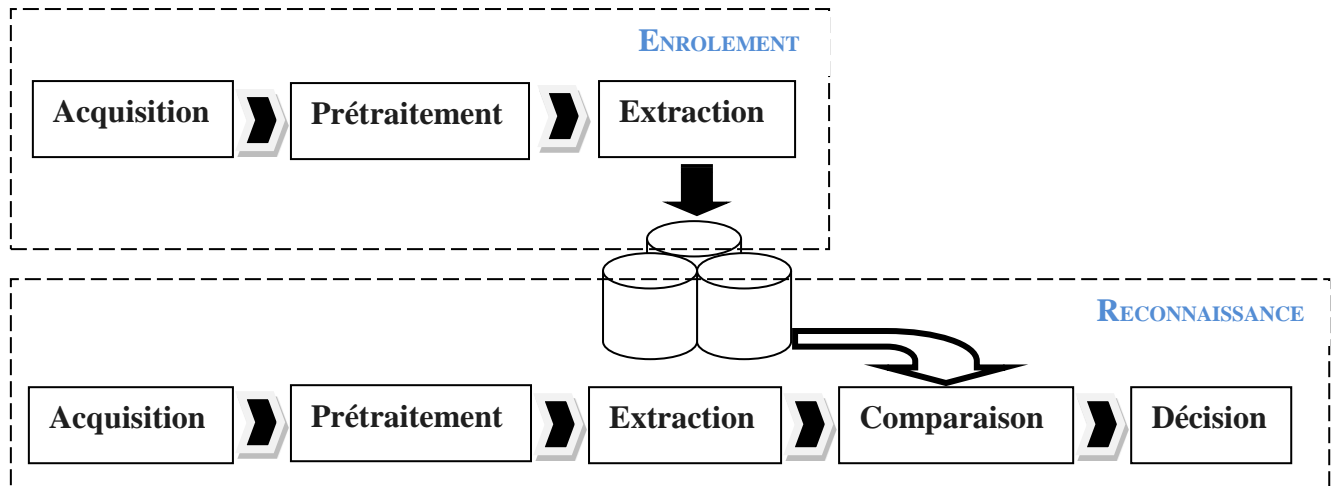


Figure 1.12. Système de reconnaissance biométrique.

1.6.1. Phase d'enrôlement

L'enrôlement, ou enregistrement, constitue l'étape fondamentale de tout système biométrique. Il s'agit de la phase cruciale durant laquelle les données biométriques uniques de chaque individu sont capturées et intégrées pour la première fois dans la base de données du système. L'objectif principal de la phase d'enrôlement est d'inscrire les individus et leurs caractéristiques biométriques (voir Figure 1.13), appelées gabarits biométriques, dans la base de données du système. Ces gabarits servent ensuite de points de comparaison lors des phases ultérieures de reconnaissance ou de vérification d'identité [14]. La phase d'enrôlement peut s'accompagner de l'enregistrement d'informations biographiques supplémentaires dans la base de données. Ces informations complémentaires, telles que le nom, le prénom, la date de naissance et le sexe, permettent d'enrichir le profil de chaque individu et de faciliter son identification en cas de correspondance biométrique.

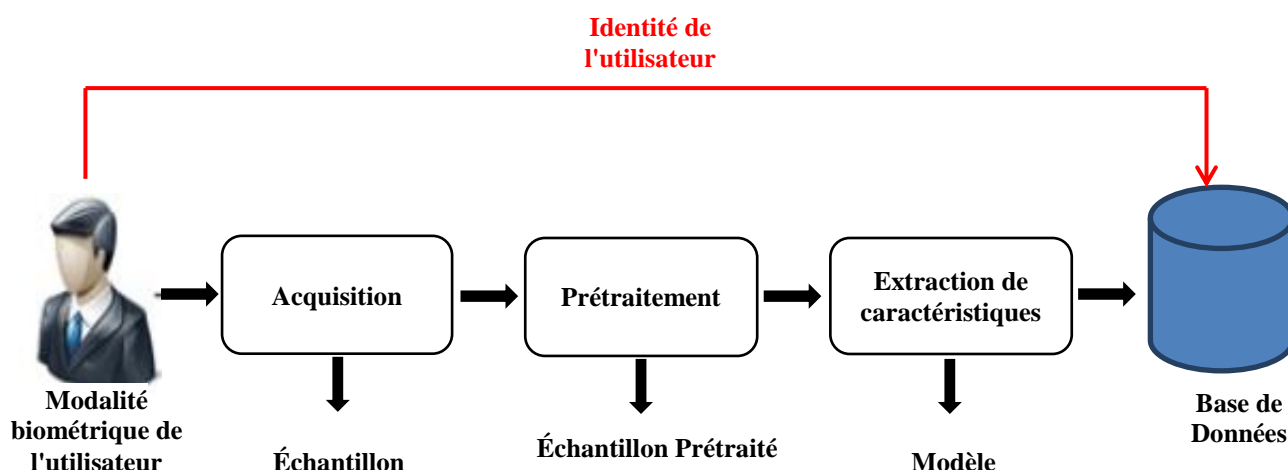


Figure 1.13. Schéma de l'étape Enrôlement dans un système biométrique

1.6.2. Phase de reconnaissance

La phase de reconnaissance constitue le cœur de fonctionnement d'un système biométrique. Son objectif principal est de vérifier ou d'identifier l'identité de la personne souhaitant accéder au système. Au cours de cette phase, la caractéristique biométrique de l'individu est mesurée à l'aide d'un capteur dédié. Cette mesure permet d'extraire un ensemble de paramètres distinctifs, formant un gabarit biométrique ou un vecteur de caractéristiques, similaire à celui créé lors de la phase d'enrôlement. Pour garantir une comparaison optimale, il est crucial que le capteur utilisé durant la reconnaissance présente des caractéristiques aussi proches que possible de celui utilisé lors de l'enrôlement. En effet, des différences de qualité ou de précision entre les capteurs peuvent altérer la performance de la reconnaissance [14]. La suite de la phase de reconnaissance dépend du mode opératoire du système, Les systèmes biométriques peuvent fonctionner selon deux modalités : l'authentification et l'identification.

➤ Mode d'authentification:

En mode d'authentification, l'individu déclare son identité (par exemple, en présentant une carte d'identité) et présente sa caractéristique biométrique au système biométrique qui régule l'accès (par exemple, en plaçant son empreinte digitale sur le capteur), qui capture l'échantillon correspondant. Ensuite, un modèle est calculé à partir de

l'échantillon capturé, et un algorithme de correspondance d'identité est utilisé pour comparer le modèle avec le modèle précédemment stocké dans la base de données pour l'identité présentée (voir Figure 1.14). Le système évalue une correspondance 1:1 entre les identités déclarée et présentée et accorde l'accès en fonction d'une opération de seuillage sur le score de correspondance résultant. Un système biométrique fonctionnant en mode d'authentification peut être inclus dans n'importe quel système qui régule l'accès à une ressource sensible, telle qu'un distributeur automatique de billets, un appareil personnel ou une zone restreinte [14].

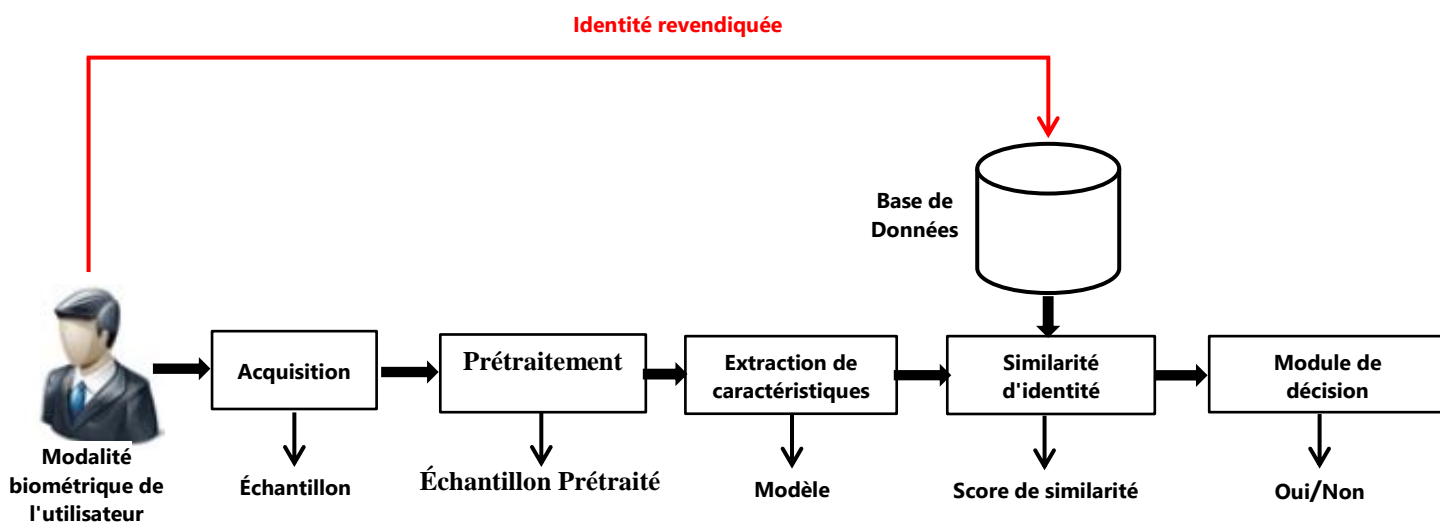


Figure 1.14. Schéma d'un système biométrique fonctionnant en mode d'authentification.

➤ **Mode d'identification:**

En mode d'identification, l'individu ne déclare pas son identité, et le système effectue une correspondance 1:N en comparant le modèle calculé à partir de la caractéristique biométrique présentée par l'individu avec tous les modèles stockés dans la base de données [14]. Ensuite, l'identité de l'individu correspondant au modèle le plus similaire, sur la base des scores de correspondance résultants, est sélectionné (voir Figure 1.15). La modalité d'identification est typique des systèmes biométriques dans les installations de maintien de l'ordre, telles que le Système Automatisé d'Identification des Empreintes Digitales (AFIS). Dans l'AFIS, une empreinte digitale extraite d'une scène de crime doit être comparée à chaque empreinte digitale présente dans la base de données pour déterminer l'identité de l'individu associé.

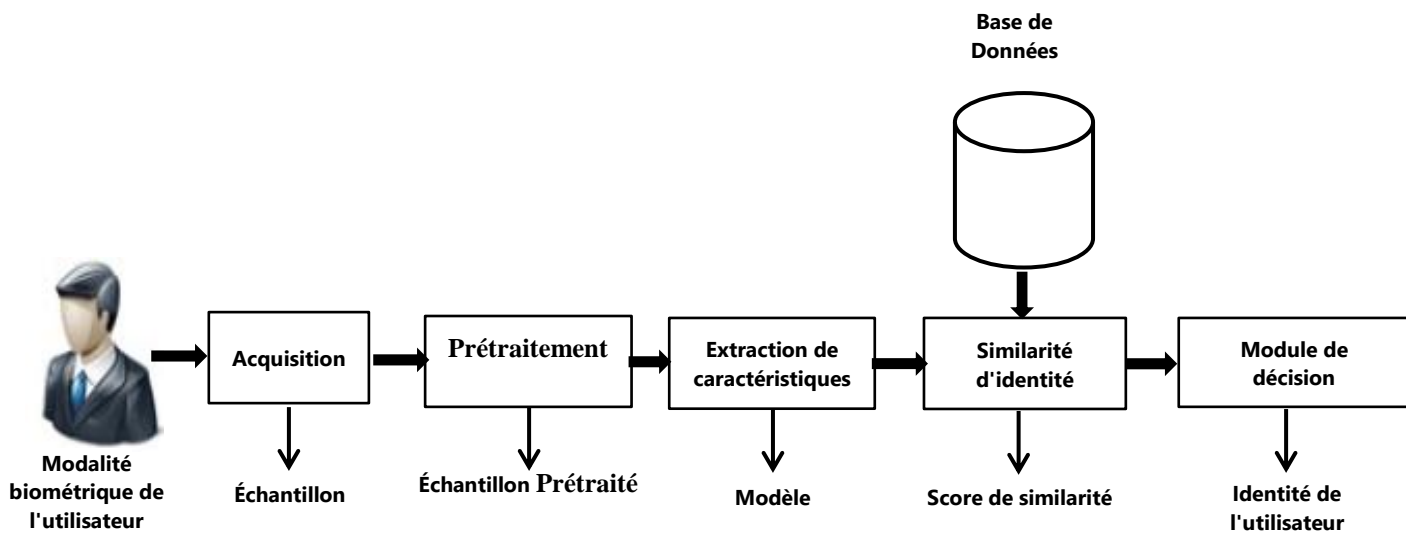


Figure 1.15. Schéma d'un système biométrique fonctionnant en mode d'identification.

Dans la plupart des cas, le terme général "reconnaissance" est utilisé pour désigner le processus de correspondance biométrique lorsqu'il n'y a pas besoin de faire une distinction entre les modalités d'authentification et d'identification. Les systèmes biométriques fonctionnant selon les modalités d'authentification et d'identification nécessitent une étape d'inscription, similaire à une procédure d'inscription, lors de laquelle un modèle est calculé à partir de la caractéristique biométrique de l'utilisateur et stocké dans la base de données du système, ainsi que l'identité associée [14]. Une illustration de l'étape d'inscription est présentée dans la figure 1.13. Un système biométrique fonctionnant en mode d'authentification est montré dans la figure 1.14, tandis qu'un système fonctionnant en mode d'identification est montré dans la figure 1.15.

1.7. Principaux Modules d'un système biométrique

Un système biométrique est un système de sécurité qui utilise les caractéristiques biologiques uniques et mesurables d'un individu pour l'identification ou l'authentification [14]. Il fonctionne en capturant l'empreinte biométrique de l'utilisateur, en l'extrayant et en la comparant à une base de données de modèles stockés. Les principaux modules d'un système biométriques sont :

1.7.1. Module d'acquisition (Capteur)

Dans ce module, les données brutes de l'utilisateur sont mesurées à l'aide du capteur biométrique ou du scanner. Ces données biométriques brutes sont enregistrées, puis transférées au module suivant pour l'extraction des caractéristiques [28]. La conception du module capteur du système biométrique est influencée par divers facteurs tels que le coût et la taille.

1.7.2. Module d'extraction de caractéristiques

Dans ce module, les données brutes transférées depuis le module capteur génèrent une représentation numérique synoptique mais indicative des traits ou modalités sous-jacents. Après l'extraction des caractéristiques, elles sont fournies en entrée au module de comparaison pour une comparaison ultérieure.

1.7.3. Module de comparaison /similarité (Matching)

Les caractéristiques extraites, lorsqu'elles sont comparées aux modèles de la base de données, génèrent un score de correspondance. Ce score de correspondance peut être influencé par la qualité des données biométriques fournies. Le module de comparaison inclut également un module de prise de décision dans lequel le score de correspondance généré est utilisé pour valider l'identité revendiquée.

1.7.4. Module de décision

Le module de prise de décision identifie si l'utilisateur est un utilisateur légitime ou un imposteur en se basant sur les scores de correspondance. Ces scores sont utilisés soit pour valider l'identité d'une personne, soit pour fournir un classement des identités enregistrées afin d'identifier un individu.

1.8. Systèmes biométriques uni-modaux et multimodaux

Il existe deux principales catégories de systèmes biométriques : uni-modaux et multimodaux. Les systèmes uni-modaux, qui se basent sur un seul trait biométrique pour identifier une personne [21]. Ces systèmes peuvent présenter des difficultés telles que des variations dans les données biométriques, un manque de caractères distinctifs, une précision de reconnaissance limitée et une vulnérabilité aux attaques par simulation

[14]. Pour surmonter ces défis, on utilise des systèmes biométriques multimodaux, qui combinent plusieurs traits biométriques pour améliorer la fiabilité et la sécurité.

1.8.1. Système biométrique uni-modal

Le système biométrique uni-modal repose sur un seul trait biométrique. Un unique trait biométrique est utilisé pour l'identification ou la vérification d'une personne [14]. Ce système est déployé dans diverses applications, notamment à des fins de sécurité. Bien qu'il présente une large gamme d'applications, ce système comporte les inconvénients suivants [29].

- Données bruitées : En raison de données bruitées, la correspondance est imprécise, entraînant ainsi des rejets erronés.
- Variation intra-classe : La variation intra-classe accroît le taux de rejet erroné. Cela est dû à la disparité entre les données biométriques acquises et celles utilisées pour créer le modèle.
- Similitudes inter-classes : Les similitudes inter-classes résultent du chevauchement de l'espace des caractéristiques en raison de plusieurs individus. Cela conduit à une augmentation du taux d'acceptation erronée.
- Non-universalité : En raison de maladies ou de handicaps, certaines personnes ne peuvent pas fournir les données biométriques requises.
- Sensibilité aux attaques (spoof attacks): Les données peuvent être falsifiées en raison de tentatives de contrefaçon (spoofing). Des recherches ont démontré la possibilité de voler une empreinte digitale et de la reproduire en utilisant des matériaux tels que la silicone. Cela est également applicable aux modalités de la signature et de la voix, qui se révèlent être des aspects relativement faciles à contrefaire.

Pour éviter ces inconvénients, le système biométrique multimodal est mis en œuvre.

1.8.2. Système biométrique multimodale

Dans notre quotidien, nous nous identifions naturellement les uns aux autres en combinant diverses informations biométriques, qu'elles soient physiques ou comportementales. La biométrie multimodale s'inspire de cette approche intuitive en fusionnant les données provenant de plusieurs modalités biométriques pour une identification plus précise et fiable [14].

En effet, chaque modalité biométrique prise isolément peut présenter des faiblesses limitant son efficacité. Par exemple, les empreintes digitales peuvent s'user ou se déformer avec le temps, tandis que la reconnaissance faciale peut être sujette à des variations d'éclairage ou d'expressions faciales. En combinant les forces de diverses modalités, telles que les empreintes digitales, la reconnaissance faciale et l'iris, les systèmes biométriques multimodaux surmontent ces limitations. Cette approche permet d'améliorer la précision de la reconnaissance. En tirant parti des points forts de chaque modalité, les systèmes multimodaux réduisent les erreurs d'identification et augmentent la robustesse face aux variations individuelles et aux environnements changeants [14] [21]. La combinaison de plusieurs caractéristiques rend plus difficile la contrefaçon simultanée de toutes les données biométriques requises, augmentant ainsi la sécurité du système. Les systèmes multimodaux peuvent s'adapter à des contextes où l'utilisation d'une seule modalité est compromise, comme en cas de blessure ou de mauvaise qualité d'image.

1.9. Fusion des données

Le système biométrique multimodal présente différentes approches de conception basées sur l'intégration d'informations provenant de multiples modalités, instances, capteurs, échantillons [30], ou toute combinaison des cinq sources de preuves telles que Multi-échantillon, Multi-capteur, Multi-algorithme, Multi-instance, Multimodal [21].

De ses origines militaires à ses applications actuelles, la fusion de données s'avère un outil puissant pour extraire des informations plus précises et pertinentes. Née dans le contexte militaire pour des missions de localisation et d'imagerie radar, la fusion de données s'est ensuite déployée dans de nombreux domaines. Exploitant des techniques issues du traitement du signal, de l'intelligence artificielle, de la reconnaissance des formes et de la classification, la fusion de données intègre plusieurs sources d'information pour en créer une nouvelle plus riche et représentative.

Aujourd'hui, elle s'avère un atout précieux pour les scientifiques, leur permettant d'extraire des données plus précises et pertinentes. Initialement dédiée à l'amélioration des réponses aux problématiques militaires, la fusion de données touche désormais des

domaines aussi variés que la télédétection, la météorologie, la biométrie multimodale, la médecine et la robotique [14].

1.10. Stratégies de fusion

Le défi principal de la fusion réside dans la synthèse des informations provenant de différentes sources. Dans le domaine des systèmes biométriques multimodaux, la fusion d'informations provenant de sources diverses constitue un enjeu crucial [21]. Cette symphonie de données, telle que représentée dans la figure 1.16, vise à composer une reconnaissance de l'identité précise et fiable. La diversité des sources d'information, comme l'illustre la figure 1.16, soulève une multitude de scénarios de fusion potentiels. Chacun d'entre eux présente des défis et des opportunités uniques, exigeant une approche adaptée pour harmoniser les données et produire une identification robuste.

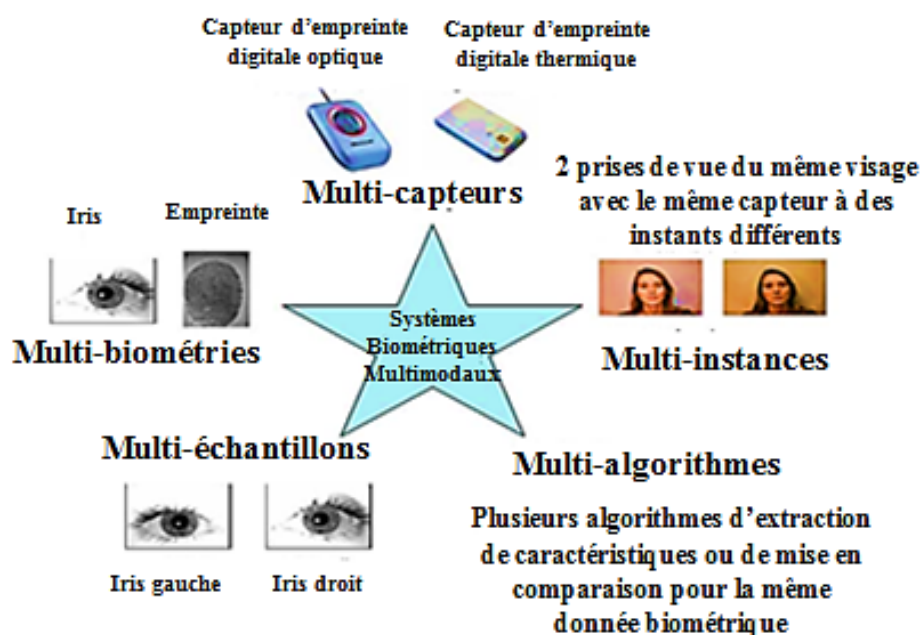


Figure 1.16. Différentes sources des systèmes biométriques multimodaux [31]

1.10.1. Systèmes multi-capteurs

Dans ces systèmes, un seul trait biométrique est imagé à l'aide de plusieurs capteurs afin d'extraire des informations diverses. Dans les systèmes biométriques unimodaux,

l'analyse d'une même modalité biométrique s'enrichit grâce à l'utilisation de capteurs multiples [21]. Ces capteurs, tels des yeux scrutateurs, extraient une multitude d'informations précieuses à partir un seul trait biométrique [28]. Prenons l'exemple de la reconnaissance faciale. Un système peut combiner les images capturées par une caméra CCD, révélant la texture 2D du visage, avec les données de profondeur 3D obtenues par un autre capteur. Cette fusion d'informations permet une identification plus précise et robuste, prenant en compte à la fois les détails visuels et la structure tridimensionnelle du visage. De la même manière, dans le domaine de l'empreinte digitale, l'association d'un capteur optique et d'un capteur thermique peut enrichir l'analyse des empreintes. Le capteur optique capture les détails visuels des crêtes et des sillons, tandis que le capteur thermique détecte la chaleur résiduelle sur la surface du doigt, révélant des informations supplémentaires sur la texture et la composition de l'empreinte [32].

1.10.2. Systèmes multi-algorithmes

Dans ces systèmes, les mêmes données biométriques sont traitées à l'aide de plusieurs algorithmes [28]. Cette multi-modalité algorithmique peut se manifester à deux niveaux :

- **Au niveau de l'extraction de caractéristiques:** Plusieurs algorithmes peuvent être combinés pour extraire des ensembles de caractéristiques distincts à partir des mêmes données biométriques. Par exemple, dans l'analyse d'empreintes digitales, des algorithmes d'analyse de texture et de minutie peuvent collaborer pour identifier des détails visuels et des points de repère caractéristiques [32].
- **Au niveau de la comparaison:** Divers algorithmes de comparaison peuvent être employés pour comparer les caractéristiques extraites aux modèles enregistrés. En combinant les résultats de ces algorithmes, le système peut prendre des décisions d'identification plus robustes et fiables.

Prenons l'exemple de la reconnaissance faciale. Un système peut utiliser des algorithmes de détection de visages pour identifier la présence d'un visage dans une image, puis employer des algorithmes d'extraction de caractéristiques (comme LBP et HOG) ainsi que des algorithmes de reconnaissance faciale pour comparer les traits du visage aux modèles enregistrés [21]. Cette combinaison d'algorithmes permet une identification précise même dans des conditions d'éclairage ou de pose faciales variées.

1.10.3. Systèmes multi-instances ou multi-unités

Ces systèmes utilisent plusieurs instances du même trait corporel et sont également appelés systèmes multi-unités dans la littérature. Par exemple, les doigts d'index gauche et droit, ou les iris gauche et droit d'un individu peuvent être utilisés pour vérifier l'identité d'une personne. Ces systèmes ne nécessitent généralement pas l'introduction de nouveaux capteurs, ni le développement de nouveaux algorithmes d'extraction et de correspondance des caractéristiques, ce qui les rend rentables [32]. Cependant, dans certains cas, un nouvel agencement de capteurs peut être nécessaire pour faciliter la capture simultanée des différentes unités/instances. Les systèmes d'identification automatique des empreintes digitales (AFIS), qui obtiennent des informations sur les dix empreintes digitales d'un sujet, peuvent bénéficier de capteurs capables d'acquérir rapidement des empreintes de tous les doigts. Les systèmes multi-instances sont particulièrement bénéfiques pour les utilisateurs dont les traits biométriques ne peuvent pas être capturés de manière fiable en raison de problèmes inévitables [28]. Par exemple, un seul doigt peut ne pas être un discriminant suffisant pour une personne ayant la peau sèche. Cependant, l'intégration des preuves provenant de plusieurs doigts peut servir de bon discriminant dans ce cas. De même, un système d'iris peut ne pas être en mesure d'imager des parties significatives de l'iris d'une personne en raison de paupières tombantes. La prise en compte des deux iris permettra d'obtenir plus d'informations de texture pouvant être utilisées pour établir l'identité de l'individu de manière plus fiable. Les systèmes multi-instances sont souvent nécessaires dans les applications où la taille de la base de données du système (c'est-à-dire le nombre d'individus inscrits) est très grande (la base de données du FBI compte actuellement environ 50 millions d'images de dix empreintes digitales et plusieurs doigts et fournissent des informations discriminantes supplémentaires).

1.10.4. Systèmes multi-échantillons

Un capteur unique peut être utilisé pour acquérir plusieurs échantillons du même trait biométrique afin de tenir compte des variations pouvant survenir dans le trait, ou pour obtenir une représentation plus complète du trait sous-jacent [32]. Par exemple, un système de reconnaissance faciale peut capturer (et stocker) le profil frontal du visage d'une personne ainsi que les profils gauche et droit afin de prendre en compte les variations de la pose faciale. De même, un système d'empreintes digitales équipé d'un

capteur de petite taille peut acquérir plusieurs empreintes de la même personne afin d'obtenir des images de différentes régions de l'empreinte digitale [21]. Un schéma de mosaïque peut ensuite être utilisé pour assembler les différentes empreintes et créer une image composite. L'un des principaux problèmes dans un système multi-échantillons est de déterminer le nombre d'échantillons qui doivent être acquis auprès d'un individu. Il est important que les échantillons obtenus représentent à la fois la variabilité et la typicité des données biométriques de l'individu. À cette fin, la relation souhaitée entre les échantillons doit être établie au préalable afin d'optimiser les avantages de la stratégie d'intégration. Par exemple, un système de reconnaissance faciale utilisant à la fois des images de face et de profil latéral d'un individu peut stipuler que l'image de profil latéral doit être une vue de trois-quarts du visage.

1.10.5. Systèmes multi-biométries

Dans ces systèmes, différentes modalités biométriques sont combinées afin d'établir l'identité d'une personne. Par exemple, on pourrait associer les caractéristiques de l'empreinte palmaire à celles de l'empreinte digitale [32]. Cette approche de fusion vise à tirer parti des atouts de chaque système biométrique tout en palliant leurs faiblesses. En effet, les systèmes qui intègrent plusieurs informations issues de la même biométrie améliorent la précision de reconnaissance en minimisant les variations intra-classe. Toutefois, ils ne résolvent pas entièrement les défis rencontrés par les systèmes unimodaux [21]. C'est pourquoi les systèmes multi-biométriques suscitent un vif intérêt au sein de la communauté scientifique

1.10.6. Systèmes hybrides

Ce type de systèmes intègre diverses stratégies de fusion mentionnées précédemment dans cette section (telles que la fusion multi-biométrique, multi-échantillon, multi-algorithmes...), dans le but de maximiser les avantages de chaque stratégie. Les systèmes hybrides tentent d'extraire autant d'informations que possible des différentes modalités biométriques. Par exemple, dans un système de fusion hybride, il est envisageable d'exploiter simultanément des images faciales et des empreintes digitales, en recourant à de multiples algorithmes d'extraction de caractéristiques et divers algorithmes de reconnaissance [28].

Outre le scénario mentionné ci-dessus, il est également possible d'utiliser des caractéristiques biométriques en conjonction avec des jetons d'identité non biométriques afin d'améliorer les performances d'authentification [32, 21]. Par exemple, il est possible d'utiliser un authentificateur à double facteur qui combine un nombre pseudo-aléatoire (généralisé par un jeton) avec un ensemble de caractéristiques faciales pour produire un ensemble de codes compacts spécifiques à l'utilisateur, appelés BioCode.

1.11. Niveaux de fusion

Dans le domaine de la reconnaissance biométrique, les systèmes multimodaux ont révolutionné l'identification en exploitant simultanément plusieurs modalités biométriques, telles que les empreintes digitales, la reconnaissance faciale et l'analyse de l'iris. Pour tirer pleinement parti de ces systèmes puissants, la fusion des données biométriques joue un rôle crucial [32]. Cette fusion peut être réalisée à différents niveaux, chacun présentant ses propres avantages et défis. L'information du système multimodal peut être fusionnée à l'un des quatre modules :

1.11.1. Fusion au niveau des capteurs

Cette approche consiste à fusionner les données brutes provenant de différents capteurs (voir Figure 1.17) [32]. On peut utiliser soit des échantillons du même trait biométrique obtenus à partir de plusieurs capteurs compatibles, soit de multiples instances du même trait biométrique obtenues à l'aide d'un seul capteur. Les données sont fusionnées à un stade très précoce, fournissant ainsi une grande quantité d'informations par rapport aux autres niveaux de fusion [28].

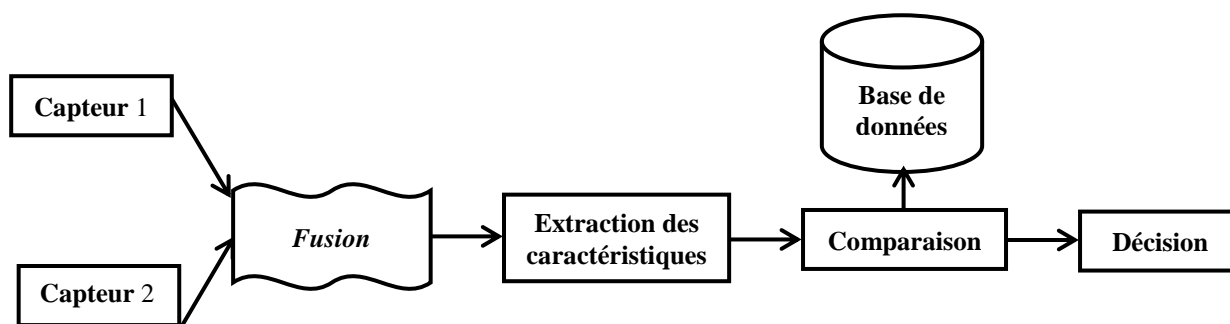


Figure 1.17 : Fusion au niveau des capteurs (Sensor level)

1.11.2. Fusion au niveau des caractéristiques

Les données ou l'ensemble de caractéristiques provenant de plusieurs capteurs ou sources sont fusionnées ensemble. Les caractéristiques extraites de chaque capteur forment un vecteur de caractéristiques [32]. Ces vecteurs de caractéristiques sont ensuite concaténés pour former un nouveau vecteur unique [21]. Dans la fusion au niveau des caractéristiques, le même algorithme d'extraction de caractéristiques ou un autre algorithme d'extraction de caractéristiques sur plusieurs modalités peut être utilisé, car les caractéristiques doivent être fusionnées [28]. La fusion au niveau des caractéristiques est difficile car la relation entre les caractéristiques n'est pas connue. Les caractéristiques structurellement incompatibles sont courantes. En raison de ces difficultés, peu de travaux ont été rapportés sur la fusion au niveau des caractéristiques dans les systèmes biométriques multimodaux. La fusion au niveau des caractéristiques est illustrée dans la figure 1.18.

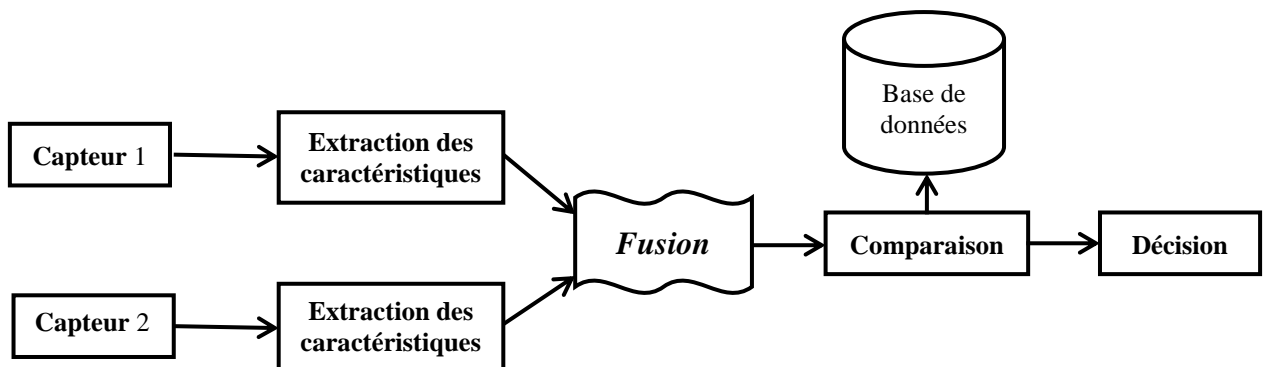


Figure 1.18. Fusion au niveau des caractéristiques (Features level)

1.11.3. Fusion au niveau des scores (Score level)

Chaque système fournit un score de correspondance indiquant la proximité du vecteur de caractéristiques avec le vecteur du modèle. Ces scores peuvent être combinés pour affirmer la véracité de l'identité revendiquée [32]. Les scores obtenus à partir de différents comparateurs (matchers) ne sont pas identiques, une technique de normalisation des scores est adaptée pour mapper les scores obtenus à partir de différents comparateurs (matchers) sur une même plage [21, 28]. Ces scores contiennent des informations riches sur l'entrée. De plus, il est assez facile de combiner

les scores de différentes biométries, c'est pourquoi beaucoup de travaux ont été réalisés dans ce domaine. La fusion au niveau du score de correspondance est illustrée dans la figure 1.19.

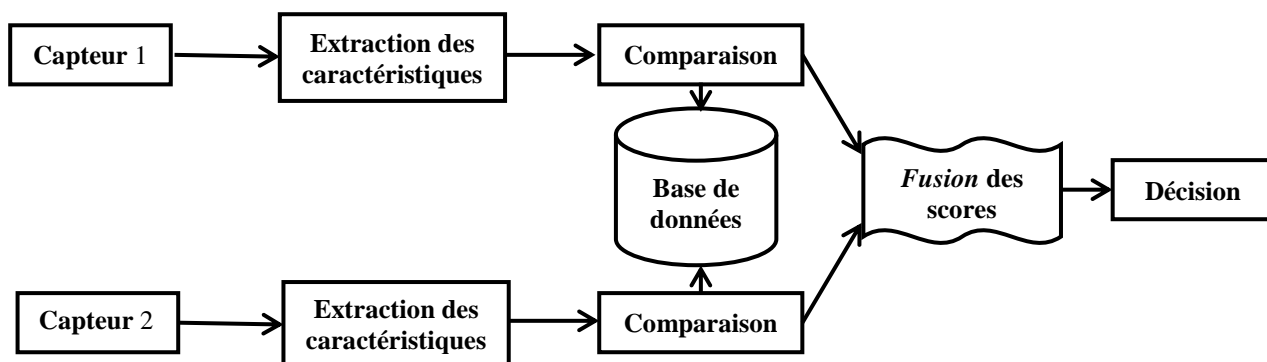


Figure 1.19. Fusion au niveau des scores (Score level)

1.11.4. Fusion au niveau des décisions (Decision level)

Les sorties finales des multiples classifieurs sont combinées. Un schéma de vote à la majorité peut être utilisé pour prendre la décision finale [28]. La fusion au niveau de la décision comprend un niveau d'information très abstrait. De ce fait, elle est moins préférée dans la conception des systèmes biométriques multimodaux. Les systèmes biométriques qui intègrent des informations aux premiers stades sont plus efficaces que ceux dans lesquels l'intégration est réalisée à des stades ultérieurs. Ainsi, la fusion au niveau des caractéristiques est censée donner de meilleurs résultats de reconnaissance, mais il est difficile de réaliser cette intégration à ce niveau car les ensembles de caractéristiques des différents systèmes peuvent ne pas être compatibles. De plus, la plupart des systèmes biométriques commerciaux ne fournissent pas l'accès aux ensembles de caractéristiques qu'ils utilisent dans leurs produits. La fusion au niveau du score du comparateur est généralement préférée car il est relativement facile d'accéder et de combiner les scores présentés par différentes modalités [21, 32]. La fusion au niveau de la décision est illustrée dans la figure 1.20.

Pour arriver à la décision finale plusieurs méthodes peuvent être utilisées :

- 1) **Vote majoritaire** : [33] le vote majoritaire est souvent utilisé pour prendre une décision lorsque plusieurs algorithmes ou modèles fournissent une réponse. les méthodes de vote consistent à interpréter chaque sortie d'un classifieur comme un vote pour l'une des classes possibles. La classe ayant un nombre de votes majoritaire ou supérieur à un seuil préfixe est retenue comme décision finale. Les votes des classifieurs ne sont pas pondérés et chaque classe reçoit autant de votes qu'il y a de classifieurs à combiner.
- 2) **Aléatoire** : La fusion aléatoire est une technique de fusion biométrique qui consiste à choisir aléatoirement une des modalités biométriques disponibles pour prendre une décision d'authentification ou d'identification. Cette technique est utilisée pour lutter contre l'usurpation d'identité, car elle rend plus difficile pour un imposteur de prédire quelle modalité biométrique sera utilisée [21].
- 3) **Les règles « ET et OU »** : En fusion biométrique, les règles ET et OU sont deux techniques utilisées pour prendre une décision finale en fonction des décisions des sous-systèmes biométriques.
 - **Règle (ET)** : La règle ET est une technique de fusion biométrique qui consiste à prendre une décision finale d'acceptation ou de rejet uniquement si toutes les décisions des sous-systèmes sont d'acceptation [21].
 - **Règle (OU)** : La règle OU est une technique de fusion biométrique qui consiste à prendre une décision finale d'acceptation ou de rejet si au moins une des décisions des sous-systèmes est d'acceptation.

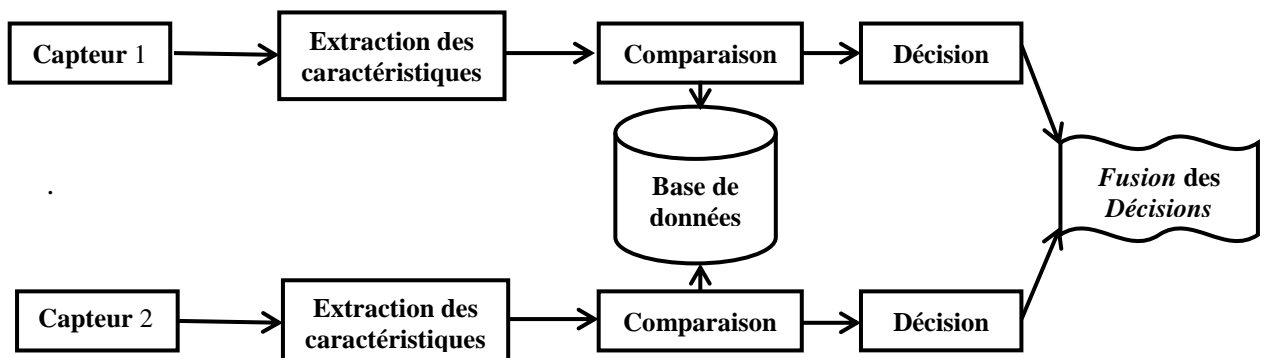


Figure 1.20. Fusion au niveau des décisions

1.12. Conclusion

En conclusion, la biométrie s'impose aujourd'hui comme l'un des piliers de la sécurité, offrant un niveau de protection sans équivalent dans divers domaines. Son adoption croissante dans le monde réel témoigne de ses nombreux avantages, notamment sa fiabilité et sa robustesse face aux tentatives d'intrusion. Ce chapitre a présenté un panorama assez large des technologies biométriques disponibles, soulignant leur diversité et leurs applications variées. Chaque technologie possède ses propres forces et faiblesses, mettant en lumière le fait qu'elles ne présentent pas toutes le même niveau d'efficacité. Enfin, nous avons introduit le concept d'un système biométrique typique, en décrivant sa structure générale et les différentes étapes impliquées dans le processus d'identification ou de vérification d'un individu.

CHAPITRE

2

SYSTEME DE RECONNAISSANCE FACIALE

2.1. Introduction

La reconnaissance faciale constitue le processus automatisé permettant d'identifier un individu à partir d'images ou de vidéos capturées [34]. Ce domaine de recherche revêt aujourd'hui une importance significative, avec des applications de plus en plus cruciales dans divers secteurs tels que l'identification criminelle, la visioconférence, les distributeurs automatiques de billets, la vérification de permis de conduire et de passeports, ainsi que la surveillance des espaces publics. La complexité de la reconnaissance faciale est accentuée par des conditions d'imagerie variables, comme l'occultation de certaines caractéristiques, la similitude interpersonnelle et les variations au niveau des visages [35]. Au cours des dernières décennies, de nombreux algorithmes ont été proposés pour la reconnaissance faciale, avec des niveaux de succès variables. Ces algorithmes analysent les images et extraient des informations telles que la forme, la taille et l'emplacement des caractéristiques faciales. Ainsi, les algorithmes les plus précis exigent généralement des ressources computationnelles intensives (des ressources informatiques qui nécessitent une puissance de calcul importante) [34].

Ce chapitre présente un aperçu historique de l'évolution de la reconnaissance faciale tout en énumérant brièvement ses principales applications. De même qu'il décrit, aussi, l'architecture fondamentale et le fonctionnement des systèmes de reconnaissance faciale. Enfin, les techniques actuelles de détection et de reconnaissance faciale y sont présentées de la façon la plus complète que possible.

2.2. Evolution de la reconnaissance faciale

La reconnaissance faciale peut être considérée comme étant l'une des applications les plus pertinentes de l'analyse d'image. Développer un système automatisé capable

d'égaliser la capacité humaine à reconnaître les visages constitue un défi majeur. Bien que les êtres humains soient compétents pour identifier des visages familiers, leur aptitude diminue considérablement lorsqu'ils sont confrontés à un grand nombre de visages inconnus. Les ordinateurs, dotés de mémoires de capacités théoriquement illimitées et de vitesses de calcul élevées, sont censés surmonter ces limitations humaines lorsqu'ils sont dotés d'algorithmes puissants et adaptés.

La reconnaissance faciale demeure un problème non résolu et une technologie très demandée, La richesse des résultats en 2023 se manifeste à travers un total de 42 694 occurrences lors de la recherche de l'expression "face recognition" dans la bibliothèque numérique "IEEE", 5499 articles ont été recensés. Divers secteurs industriels s'intéressent aux applications potentielles de cette technologie, parmi lesquels la vidéosurveillance, l'interaction homme-machine, la photographie, la réalité virtuelle. Cet intérêt multidisciplinaire stimule la recherche et attire l'attention de diverses disciplines, dépassant ainsi le cadre de la seule vision par ordinateur. La reconnaissance faciale revêt une importance particulière dans des domaines tels que la reconnaissance de formes, les réseaux de neurones, le traitement d'image et la psychologie.

Les premiers travaux, jugés pertinents, sur ce sujet remontent aux années 1960, se situant initialement dans le domaine de la psychologie et s'articulant autour de questions liées à l'expression faciale, l'interprétation des émotions et la perception des gestes. Dans les années 1990 et au début des années 2000, la reconnaissance faciale a connu un développement rapide, avec des méthodologies dominées par des approches holistiques (par exemple, l'espace linéaire [36], l'apprentissage de variétés [37, 38]), qui extraient des caractéristiques de basse dimension en prenant l'ensemble du visage comme entrée. Plus tard, dans les années 2000 et au début des années 2010, des descripteurs locaux (comme Gabor [39], Local Binary Pattern (LBP) [40], Direction Field Descriptor (DFD), etc.) et le Support Vector Machine (SVM) [41] ont été appliqués à la reconnaissance faciale, améliorant davantage les performances de reconnaissance. Cependant, ces méthodes traditionnelles [36, 37, 38, 39, 40, 41] souffrent de conceptions élaborées et de représentations peu profondes, et ont du mal à obtenir des performances de reconnaissance robustes face à des variations complexes de poses, d'occlusions, d'éclairages, d'expressions, etc.

Au cours des dernières années, la reconnaissance faciale a connu une évolution rapide. En 2014, des algorithmes d'intelligence artificielle et particulièrement de l'apprentissage profond (The deep learning) [42] ont été appliqués au problème de la reconnaissance faciale et lui ont permis de connaître des avancées remarquables. Par exemple, DeepFace [43, 44] et DeepID [45], construits sur plusieurs couches de Réseaux de Neurones Convolutifs (CNN) [46], ont atteint des performances remarquables (précision de 99% dans la base de données LFW [47]) En comparaison avec les méthodes traditionnelles, les méthodes basées sur le deep learning présentent un avantage écrasant en termes de précision de reconnaissance [42]. Inspirés par ces résultats, de nombreux chercheurs se sont activement impliqués dans la recherche sur la reconnaissance faciale profonde et ont développé une série d'algorithmes pour repousser les limites des performances.

Le succès du deep learning [46] dans la reconnaissance faciale provient de plusieurs aspects. Tout d'abord, la structure appropriée des couches empilées des réseaux CNN permet d'apprendre des caractéristiques discriminatives avec une forte invariance face à des variations relativement importantes de la pose, des occlusions, de l'éclairage, etc. Ensuite, l'entraînement des méthodes basées sur le deep learning avec des paramètres d'apprentissage massif sur des bases de données à grande échelle (par exemple, CASIA-WebFace [48] et MS-Celeb-1M [49]) permet l'extraction de représentations discriminatives de manière efficace. Enfin, le développement des technologies informatiques et du matériel (GPU) fournit un fort soutien pour l'entraînement à grande échelle.

Puis, et suite aux développements technologiques qu'ont connus le matériel et le logiciel informatiques, la précision des systèmes de reconnaissance faciale s'est considérablement améliorée. En 2023, les systèmes de reconnaissance faciale les plus avancés peuvent identifier des visages familiers, même dans des conditions difficiles. Cette amélioration de la précision est due à plusieurs facteurs, notamment :

- Le développement de nouveaux algorithmes de reconnaissance faciale.
- Le développement de nouvelles Méthodes d'extraction de caractéristiques.
- L'augmentation de la taille des bases de données de visage.
- L'amélioration des performances des processeurs (Hardware).

A tout cela, on peut rajouter les niveaux de maîtrise et le volume important de recherche investi dans le domaine.

2.3. Pourquoi utiliser le visage pour la reconnaissance ?

Il existe de nombreuses façons d'authentifier une personne, tandis que les techniques basées sur la biométrie sont apparues comme l'option la plus prometteuse pour la reconnaissance des individus ces dernières années. D'autres approches telles que les jetons, les clés, les mots de passe et les cartes intelligentes nécessitent l'implication des individus de quelque manière que ce soit et présentent des risques d'être égarées, volées, oubliées ou contrefaites. Cependant, les traits biologiques d'un individu ne peuvent pas être égarés ou oubliés. Les systèmes de reconnaissance basés sur la biométrie incluent les caractéristiques individuelles telles que le visage, la paume, la rétine, la voix, la géométrie des doigts, et ainsi de suite. Parmi toutes ces modalités biométriques, le visage offre de nombreux avantages [50], c'est pourquoi elle est préférée. Néanmoins, toutes ces techniques nécessitent une action volontaire de l'utilisateur, et la reconnaissance faciale peut être réalisée passivement sans aucun effort explicite, car les images faciales peuvent être capturées à distance par une caméra. Cela est utile à des fins d'identification, de sécurité et de surveillance. De même, certaines méthodes dépendent des mains et des doigts, qui peuvent être rendues inutiles si le tissu épidermique est endommagé de quelque manière que ce soit. L'identification de la rétine nécessite un équipement coûteux, et elle est beaucoup trop sensible aux mouvements du corps. De même, la reconnaissance vocale est sujette aux bruits de fond dans les espaces publics et aux fluctuations auditives sur une ligne téléphonique ou un enregistrement magnétique. En revanche, les images faciales peuvent être obtenues facilement avec une caméra abordable. De bons algorithmes de reconnaissance faciale et un prétraitement d'image approprié peuvent compenser et corriger le bruit et les légères variations d'orientation, d'échelle et d'éclairage.

2.4. Domaines D'application de la reconnaissance faciale

La reconnaissance faciale est une technologie qui permet d'identifier une personne à partir de son visage. Elle est utilisée dans de nombreux domaines [51, 52], tels que la sécurité, le commerce, la santé et les loisirs [53].

- **Dans le domaine de la sécurité** [52], la reconnaissance faciale est utilisée pour identifier les personnes, contrôler l'accès à des zones sensibles ou surveiller des zones publiques. Par exemple, elle peut être utilisée pour identifier les personnes qui entrent dans un bâtiment gouvernemental, pour contrôler l'accès à un port ou pour surveiller une rue.

- **Dans le domaine du commerce**, la reconnaissance faciale est utilisée pour authentifier les clients, permettre le paiement sans contact ou cibler la publicité. Par exemple, elle peut être utilisée pour authentifier les clients d'une banque en ligne [51], pour permettre le paiement sans contact dans un magasin ou pour cibler la publicité en fonction de l'âge ou du sexe d'une personne.

- **Dans le domaine de la santé**, la reconnaissance faciale est utilisée pour identifier les patients, suivre les patients ou mener des recherches médicales [52]. Par exemple, elle peut être utilisée pour identifier les patients dans un hôpital, pour suivre les progrès d'un patient en chimiothérapie ou pour identifier les personnes à risque de développer une maladie.

- **Dans le domaine des loisirs**, la reconnaissance faciale est utilisée pour authentifier les utilisateurs [51], contrôler la parentalité ou jouer à des jeux vidéo. Par exemple, elle peut être utilisée pour authentifier les utilisateurs d'une console de jeu, pour contrôler l'accès à un site Web pour enfants ou pour jouer à un jeu vidéo qui reconnaît les expressions faciales du joueur.

2.5. Système de reconnaissance faciale

L'architecture générique et simplifiée d'un système de reconnaissance faciale s'articule autour de plusieurs modules interconnectés qui travaillent ensemble pour détecter, traiter, extraire des caractéristiques, comparer et décider [52, 35] (voir Figure 2.1).

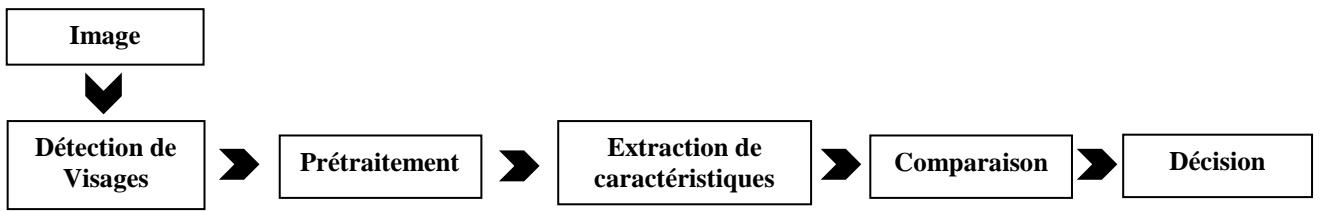


Figure.2.1 : Architecture générique d'un système de reconnaissance faciale. [35]

2.5.1. Détection de visages

La première étape dans un système de reconnaissance de visage consiste à localiser le ou les visages présents dans une image ou une vidéo [35, 51]. Cette tâche est généralement accomplie à l'aide de méthodes de détection qui reposent sur des techniques comme :

- **Détection de caractéristiques faciales** : Les algorithmes recherchent des traits caractéristiques du visage comme les yeux, le nez, la bouche ou les contours du visage. Une méthode classique est celle du **classifieur en cascade de Haar** [51], tandis que les techniques modernes utilisent des algorithmes de **Deep Learning** [46], tels que **MTCNN (Multi-Task Convolutional Neural Network)**, qui sont particulièrement efficaces pour repérer les visages dans différentes conditions d'éclairage, d'angles ou d'expressions faciales.
- **Modèles de réseaux neuronaux** : Des réseaux comme **YOLO (You Only Look Once)** [54] et **SSD (Single Shot Multibox Detector)** [55] permettent également une détection en temps réel, ce qui est crucial dans les applications de surveillance.

La détection de visage est une étape critique, car elle détermine les zones à analyser pour l'identification ou la vérification et toute erreur à ce stade peut rendre l'ensemble du processus inefficace [52].

2.5.2. Prétraitement des images

Une fois le visage détecté, il est important de normaliser l'image pour améliorer la précision de la reconnaissance [52, 56]. Le prétraitement comprend plusieurs étapes :

- **Redimensionnement** : Les visages peuvent être redimensionnés pour obtenir une taille uniforme, ce qui facilite l'analyse et réduit la complexité computationnelle.
- **Correction de l'orientation** : Les images peuvent être tournées ou ajustées pour aligner les visages de façon standard (par exemple, en redressant les visages inclinés) [57].
- **Correction d'éclairage** : Des algorithmes sont appliqués pour compenser les variations d'éclairage afin que les caractéristiques faciales soient visibles et distinctes même dans des conditions de faible lumière [56].
- **Normalisation des couleurs** : Les images peuvent être converties en niveaux de gris pour se concentrer uniquement sur les traits structuraux sans être influencé par les variations de couleur de peau [57].

Le but du prétraitement est de rendre les images plus cohérentes afin d'améliorer la qualité des caractéristiques qui seront extraites par la suite [56].

2.5.3. Extraction de caractéristiques

C'est l'étape clé qui permet de traduire le visage en une forme que le système peut comprendre et comparer. Les **caractéristiques faciales uniques** sont extraites sous forme de vecteurs numériques [51]. Il existe plusieurs méthodes pour cette extraction :

- **Points d'intérêt** : Le système identifie des points spécifiques du visage (par exemple, la distance entre les yeux, la largeur du nez, la forme de la mâchoire) [58]. Ces caractéristiques sont souvent appelées **points nodaux** et servent à créer une sorte de carte du visage.
- **Descripteurs locaux** : Les descripteurs tels que **HOG (Histogram of Oriented Gradients)** ou **LBP (Local Binary Patterns)** [59] analysent les textures et les formes présentes dans des sous-régions du visage.
- **Réseaux de neurones convolutifs (CNN)** : Les méthodes basées sur les CNNs sont aujourd'hui largement dominantes [46]. Des architectures comme **FaceNet**, **DeepFace**, ou **VGG-Face** sont capables d'extraire des représentations de haut niveau du visage sous forme de vecteurs numériques en se basant sur des millions de paramètres optimisés pour capturer les différences subtiles entre les visages [43].

Ces caractéristiques sont ensuite transformées en une représentation numérique appelée **vecteur de caractéristiques** [51] qui permet de différencier un visage d'un autre.

2.5.4. Comparaison et correspondance

Une fois que le visage est représenté sous forme d'un vecteur **de caractéristiques** [52], ce vecteur est comparé à ceux enregistrés dans une base de données pour déterminer une correspondance. Les étapes de la comparaison incluent :

- 1) **Mesure de similarité** : Les systèmes utilisent souvent des mesures de distance telles que la **distance euclidienne** ou la **distance cosinus** pour comparer deux vecteurs [60]. Si la distance entre les vecteurs est en dessous d'un certain seuil, cela signifie qu'il y a une forte probabilité que les deux visages appartiennent à la même personne.
- 2) **Recherche dans la base de données** : Dans les systèmes à grande échelle, des techniques d'indexation rapides comme **KD-Trees** ou **LSH (Locality-Sensitive Hashing)** sont utilisées pour accélérer le processus de recherche parmi des milliers, voire des millions de vecteurs faciaux.

La comparaison peut être utilisée soit pour :

- **Identification** (1, N) où le système cherche à identifier une personne parmi un grand base de visages [60].
- **Vérification** (1:1), où le système compare un visage à une seule entrée dans la base de données pour valider l'identité (par exemple, dans le cas d'un contrôle d'accès) [60].

2.5.5. Décision

Après la comparaison, le système prend une décision [35] :

- **Correspondance trouvée** : Si le score de similarité dépasse un certain seuil, le système conclut que le visage correspond à une entrée existante.
- **Aucune correspondance** : Si le score est trop faible, le système conclut que le visage ne correspond à aucune entrée connue.

La précision de cette décision dépend fortement de plusieurs facteurs comme la qualité de l'image, les conditions d'éclairage, ou encore la performance du modèle de reconnaissance utilisé [35]. Des seuils ajustables sont souvent définis pour équilibrer les faux positifs et les faux négatifs en fonction de l'application.

2.6. Défis de la reconnaissance faciale

La détection et la reconnaissance faciales sont deux tâches connexes dans le traitement d'images numériques qui ont de nombreuses applications en matière de sécurité, de biométrie, de médias sociaux et de divertissement. Cependant, malgré les avancées en matière d'apprentissage profond et de vision par ordinateur, ces tâches restent difficiles et limitées par plusieurs facteurs [61]. Dans cette partie, nous explorerons certains des problèmes et obstacles actuels auxquels la recherche en détection et reconnaissance faciales est confrontée et comment ils affectent l'exactitude, la robustesse et l'efficacité des algorithmes. Voici quelques-uns des défis auxquels les chercheurs sont confrontés.

2.6.1. Variabilité des visages

L'un des principaux défis de la détection et de la reconnaissance des visages est la grande variabilité des visages humains en termes de forme, de taille, de pose, d'expression, d'éclairage, d'occlusion [61] et de maquillage. Ces facteurs peuvent rendre difficile la généralisation des algorithmes et leur adaptation à différents scénarios et conditions [51, 61]. Par exemple, la détection des visages peut échouer lorsque le visage est partiellement caché par un masque, un chapeau ou une main, ou lorsque le visage est incliné, tourné ou déformé. La reconnaissance des visages peut également souffrir d'une faible performance lorsque le visage n'est pas aligné, recadré ou normalisé, ou lorsque le visage change au fil du temps en raison du vieillissement, de la pilosité faciale ou des cosmétiques.

➤ **Les variations d'éclairage** : La reconnaissance faciale est sensible aux variations d'éclairage [51]. Par exemple, un visage qui est mal éclairé ou qui est éclairé par une lumière artificielle peut être difficile à reconnaître.

- **La rotation de la tête** : La reconnaissance faciale est également sensible à la rotation de la tête. Par exemple, un visage qui est incliné ou tourné sur le côté peut être difficile à reconnaître.
- **L'expression faciale** [61] : L'expression faciale peut également affecter la précision de la reconnaissance faciale [51]. Par exemple, un visage qui sourit ou qui fronce les sourcils peut être difficile à reconnaître.
- **Le vieillissement** : Le vieillissement peut également affecter la précision de la reconnaissance faciale. Par exemple, un visage qui a vieilli peut être difficile à reconnaître par un système de reconnaissance faciale qui a été formé sur des images de visages jeunes.
- **La discrimination** : La reconnaissance faciale peut être utilisée pour la discrimination. Par exemple, un système de reconnaissance faciale peut être biaisé contre certaines races ou ethnies.

2.6.2. *Diversité des bases de données*

Un autre défi de la reconnaissance faciale réside dans la diversité et la qualité de bases de données utilisées pour former et évaluer les algorithmes. Différents bases de données peuvent présenter différentes caractéristiques, telles que la résolution, le format, la distribution [61]. Par exemple, certaines bases de données peuvent comporter davantage d'images de certaines ethnies, de certains genres, de certaines tranches d'âge ou de certains milieux que d'autres, ce qui peut influencer l'équité et la représentativité des algorithmes. De plus, certaines bases de données peuvent présenter des étiquettes bruyantes, incomplètes ou inexactes, ce qui peut compromettre la fiabilité et la validité des algorithmes.

2.6.3. *Complexité des algorithmes de la reconnaissance faciale*

Un troisième défi de la reconnaissance faciale réside dans la complexité et le compromis des algorithmes utilisés pour mettre en œuvre ces tâches. D'une part, les algorithmes doivent être suffisamment précis et robustes pour gérer la variabilité et la diversité des visages et des bases de données. D'autre part, les algorithmes doivent être suffisamment efficaces pour fonctionner sur différents appareils et plates-formes, tels

que les téléphones mobiles, les caméras ou les serveurs cloud. Par conséquent, les algorithmes doivent trouver un équilibre entre vitesse, mémoire, puissance et précision, ce qui peut être difficile à réaliser et à optimiser.

2.6.4. Confidentialité et éthique

Un quatrième défi de la reconnaissance faciale réside dans la confidentialité et l'éthique des applications et des utilisateurs des tâches. La détection et la reconnaissance faciale peuvent avoir des utilisations positives et bénéfiques, telles que le déverrouillage des appareils, la vérification des identités, l'étiquetage des photos ou l'amélioration du divertissement. Cependant, elles peuvent également avoir des utilisations négatives et nuisibles, telles que l'invasion de la vie privée, le vol de données, l'usurpation d'identité ou la surveillance. Par conséquent, les chercheurs et les développeurs des tâches doivent prendre en compte les implications éthiques et légales de leur travail et veiller au consentement, à la sécurité et aux droits des utilisateurs.

2.7. Travaux connexes

La reconnaissance faciale a attiré l'attention des chercheurs depuis plusieurs décennies et demeure un domaine de recherche attrayant et en évolution constante. Les domaines de la reconnaissance de formes, du traitement d'images et des statistiques ont été mobilisés pour le développement des méthodes de reconnaissance faciale. De plus, les progrès croissants des technologies informatiques et l'existence de vastes bases de données ont permis d'élaborer des algorithmes et des approches de plus en plus complexes, améliorant ainsi les performances de reconnaissance. Dans cette section, nous présentons les travaux connexes les plus récents en se basant sur des caractéristiques globales et locales. De plus, nous discutons également de certains modèles d'apprentissage profond récents pour la reconnaissance faciale.

2.7.1. Approches globales

Dans ces approches, également appelées méthodes basées sur l'apparence, les images faciales sont traitées globalement, c'est-à-dire qu'il n'est pas nécessaire d'extraire des points caractéristiques ou des régions faciales (bouche, yeux, etc.) [53]. Ainsi, une image faciale est représentée par une matrice de pixels, et cette matrice est souvent

transformée en vecteurs de pixels pour en faciliter la manipulation. Bien que ces approches soient faciles à mettre en œuvre, elles sont sensibles aux variations (poses, éclairage, expressions faciales et orientation). En effet, tout changement dans l'image faciale entraîne un changement des valeurs des pixels. Les approches globales peuvent être subdivisées en méthodes linéaires et non linéaires.

2.7.1.1. Techniques linéaires

Ces approches utilisent une projection linéaire des données d'image d'entrée à partir d'un grand espace dans une zone d'un espace de taille relativement plus petite (le sous-espace du visage). Cependant, une telle projection présente deux inconvénients majeurs. Premièrement, les variations de visage, qui nous permettent de distinguer différents individus, ne peuvent pas être préservées. Ainsi, pour comparer les vecteurs des pixels d'un sous-espace linéaire, les distances euclidiennes utilisées ne sont pas très efficaces pour classer les formes de visage/non-visage et les individus. Par conséquent, le taux de détection et de reconnaissance de ces méthodes est généralement insatisfaisant. Plusieurs techniques peuvent être classées comme des techniques linéaires:

Shamrat et al. [62] proposent une méthode de reconnaissance faciale qui repose sur les eigenfaces (ensemble de vecteurs propres utilisés dans le domaine de la vision artificielle), et l'Analyse en Composantes Principales (**PCA**) [51]. Cette approche explore l'espace des caractéristiques faciales pour identifier un sous-ensemble pertinent de composantes abstraites. La méthode SURF (Speeded Up Robust Features) est utilisée pour extraire des points d'intérêt et leurs descripteurs correspondants des images faciales. Ensuite, l'algorithme PCA (Analyse en Composantes Principales) qui est un algorithme de machine learning non supervisé mis en place lorsque le nombre de dimensions est très grand est appliqué pour réduire la dimensionnalité des descripteurs extraits, en conservant les composantes les plus informatives. Finalement, les eigenfaces, dérivées du PCA, sont utilisées pour représenter les visages stockés dans la base de données. La reconnaissance d'un nouveau visage s'effectue en mesurant sa similarité avec les eigenfaces et en identifiant la correspondance la plus proche. Les résultats expérimentaux sur la base de données d'images faciales ORL [63] démontrent que cette approche atteint une précision et un taux de reconnaissance élevés.

S. Q. Nur Septi et al. [64] ont présenté un système basé sur l'Analyse en Composantes Principales (**PCA**) et les "Fisherfaces", utilisant les Machines à Vecteurs de Support (SVM) [51] pour classifier un ensemble de données contenant des visages masqués. Cette approche de reconnaissance faciale s'appuie sur deux techniques principales. La première, PCA, réduit la dimensionnalité des images faciales tout en conservant les informations essentielles. La deuxième, le Discriminant Linéaire de Fisher (DLF), optimise la séparation entre les classes dans l'espace des caractéristiques réduit par PCA. Cette méthode atteint un taux de précision moyen de 99,76% sur l'ensemble des données de visages masqués.

Rasmikanta Pati et al. [65] présentent un système basé sur une variante d'ICA (Independent Component Analysis) appelée Block ICA (**B-ICA**). Contrairement à l'ICA traditionnelle qui traite l'image entière comme un unique vecteur, la B-ICA divise l'image du visage en blocs et utilise ces blocs comme vecteurs d'entraînement. Cela permet de réduire la dimensionnalité des données de manière plus efficace. Les auteurs comparent leur méthode à la B-ICA classique sur trois bases de données d'images et montrent que leur méthode atteint un taux de reconnaissance faciale supérieur pour différentes tailles de blocs. Ils obtiennent de meilleurs résultats avec 70%, 80% et 90% des données utilisées pour l'entraînement du modèle.

Belhumeur et al. [36] ont travaillé avec d'autres techniques également basées sur la décomposition linéaire, telles que l'Analyse Discriminante Linéaire (**LDA**) [51]. Contrairement à la PCA, qui construit un sous-espace pour représenter de manière optimale uniquement l'objet "visage", la LDA crée un sous-espace discriminant pour distinguer de manière optimale les visages de différentes personnes. La LDA, également connue sous le nom d'analyse discriminante linéaire de Fisher, est l'une des approches les plus largement utilisées pour la reconnaissance faciale. Elle se fonde sur le critère de séparabilité des données par classe pour réduire la dimensionnalité. La LDA comprend deux étapes : la réduction de l'espace d'origine par la PCA et la projection dans l'espace final, utilisant des vecteurs appelés "Fisher faces". Ces derniers sont calculés sur la base du critère de séparabilité des classes, mais dans l'espace réduit. Cette réduction est nécessaire en raison du critère de singularité de la matrice de dispersion totale dans l'approche LDA. Des études comparatives montrent que les méthodes basées sur la

LDA offrent généralement de meilleurs résultats que celles fondées sur la PCA [66, 67, 68].

Pentland, A. et al. [69] ont amélioré la PCA pour traiter la variation de pose en introduisant un sous-espace probabiliste. Ce dernier fournit une mesure de similarité plus pertinente dans un contexte probabiliste. Par ailleurs, Lu, J.; Plataniotis, K.N. et al. [70] ont combiné le **D-LDA** (LDA Direct) et le **F-LDA** (LDA Fractionnaire), une variante de la LDA où des fonctions pondérées sont utilisées pour minimiser les erreurs de classification dues à la proximité des catégories. Les améliorations apportées dans ces travaux ont donné de bons résultats.

Magesh Kumar et al. [71] proposent une approche basée sur les caractéristiques de Gabor, **LDA** et un classifieur ANN (Artificial Neural Networks) pour renforcer la fiabilité de la reconnaissance faciale. Cette approche emploie deux techniques de modélisation statistique populaires : **PCA** et **LDA**, pour extraire les caractéristiques distinctives des visages. Un prétraitement des images faciales est effectué à l'aide d'ondelettes de Gabor pour réduire les variations dues à la pose et à l'éclairage. Les techniques PCA et LDA génèrent des vecteurs de caractéristiques discriminantes de faible dimension, qui sont ensuite utilisés pour la classification, employant un Réseau Neuronal à Rétro-Propagation (BPNN) comme classifieur. La méthode proposée a été testée avec succès sur la base de données faciale ORL. En utilisant les méthodes d'extraction de caractéristiques Gabor et PCA, combinées avec la classification ANN, le système a atteint des taux de reconnaissance significatifs. Par exemple, avec 199 caractéristiques, le taux de reconnaissance a atteint 98,6% en utilisant l'ANN, contre 97,8% avec la méthode de mesure de distance euclidienne.

2.7.1.2. Techniques non linéaires

Les techniques non linéaires offrent une alternative aux approches linéaires traditionnelles pour la reconnaissance faciale. Elles permettent de capturer les complexités des visages et d'améliorer les performances de reconnaissance dans des conditions d'éclairage et de pose variées [60]. Contrairement aux techniques linéaires qui projettent les données dans un sous-espace de dimension réduite, les techniques non linéaires transforment les données dans un espace de dimension plus élevée pour capturer les relations non linéaires entre les caractéristiques faciales. Parmi les

techniques non linéaires, on trouve **Kernel-PCA** et **Kernel-LDA**. L'utilisation de ces techniques offre de meilleures performances dans des conditions d'éclairage et de pose variées, et permet d'utiliser des représentations faciales plus robustes. Dans ce contexte, plusieurs approches non linéaires ont été proposées :

Maafiri, Ayyad et al. [72] ont proposés un nouvel algorithme appelé **RRQR-PCA** (Rank-Revealing QR PCA) afin d'améliorer les performances de **QR-PCA** en exploitant la factorisation QR Révélatrice de Rang (RRQR). Ils ont également amélioré le taux de reconnaissance de RRQR-PCA en développant une extension non linéaire. De plus, un nouveau noyau robuste basé sur la RBF (Radial Basis Function) L_p -norm kernel est proposé afin de réduire l'effet des valeurs aberrantes et du bruit. Des expériences approfondies sur deux bases de données de visages standard bien connues, ORL et FERET, prouvent que l'algorithme proposé est plus robuste que les PCA conventionnels, 2DPCA, PCA-L1, WTPCA-L1, LDA et 2DLDA en termes de précision de reconnaissance faciale.

Dans [73], les auteurs ont introduit une nouvelle approche utilisant **PCANet** comme base, combinée à des classifieurs SVM linéaires et un Réseau de Neurone (NN Neural Network). Le modèle PCANet, tel que décrit dans cette étude, se compose de deux étapes d'extraction de caractéristiques et d'une seule étape de sortie non linéaire. Les caractéristiques extraites sont ensuite utilisées séparément dans les Classifieurs SVM linéaires et un Réseau de Neurone (NN). Pour évaluer la méthode proposée, les auteurs ont mené des expériences en comparant ses résultats à des techniques d'extraction de caractéristiques bien établies, telles que LBP, Gabor et Hierarchical Multiscale LBP. Cette évaluation est réalisée à l'aide de plusieurs bases de données, notamment XM2VTS ORL, AR, Extended Yale-B et LFW. Les résultats des tests montrent que PCANet présente une résistance supérieure aux variations causées par l'occultation, l'éclairage, la pose, le bruit et l'expression. Par conséquent, cette méthode offre une promesse significative pour améliorer les applications de reconnaissance faciale.

2.7.2. Approches locales

Les approches locales traitent uniquement certaines caractéristiques faciales qui sont ensuite classées selon des statistiques bien définies [53]. Les méthodes locales,

également appelées méthodes basées sur les caractéristiques, peuvent être classées en deux catégories :

- Méthodes de reconnaissance faciale basées sur les points d'intérêt : dans ces approches, les points d'intérêt sont d'abord identifiés. Par la suite, les caractéristiques présentes sur ces points sont extraites.
- Méthodes de reconnaissance faciale basées sur l'apparence locale : le visage est divisé en petites régions (ou patches) à partir desquelles des caractéristiques locales sont directement extraites.

2.7.2.1. Méthodes de reconnaissance faciale basées sur les points d'intérêt

Dans ces méthodes, les caractéristiques géométriques spécifiques, telles que la largeur de la tête, la distance entre les yeux, etc. sont extraites. Ensuite, ces données seront utilisées dans l'étape de classification afin de reconnaître les individus. Ces méthodes peuvent être divisées en deux classes selon le point d'intérêt. La première catégorie se concentre sur la performance des détecteurs des points caractéristiques du visage tandis que la deuxième classe traite des représentations plus élaborées des informations véhiculées par les points caractéristiques du visage, plutôt que seulement les caractéristiques géométriques.

Xianming Chen et al. [74] ont réalisé un travail présentant une méthode parallèle pour la reconnaissance faciale basée sur la mise en correspondance de graphes élastiques (Elastic Bunch Graph Matching EBGM) [51]. Comparée à d'autres méthodes telles que l'analyse en composantes principales (PCA) et l'analyse discriminante linéaire (LDA), l'EBGM offre l'avantage d'une meilleure précision, mais au prix d'un temps de calcul et d'une utilisation de la mémoire plus importants, ce qui la rend moins pratique. Ils proposent une méthode parallèle pour l'EBGM en équilibrant l'unité des images. Le processus d'apprentissage est distribué et les images de test sont réparties uniformément sur tous les processeurs. Ensuite, le processus de reconnaissance est effectué simultanément sur tous les processeurs grâce à une communication mutuelle.

Laurenz Wiskott et al. [75] ont présenté un système de reconnaissance de visages humains à partir d'images uniques extraites d'une vaste base de données. Ce système est basé sur l'EBGM [51], une extension de LDA dans laquelle les nœuds des graphes sont

situés sur un certain nombre de points sélectionnés du visage. Les visages sont représentés par des graphes étiquetés, basés sur une transformation en ondelettes de Gabor. Les graphes d'images de nouveaux visages sont extraits par un processus de correspondance de graphes élastique et peuvent être comparés par une simple fonction de similarité. Le système se distingue des travaux précédents à trois égards. [76, 77] L'information de phase est utilisée pour un positionnement précis des nœuds. Des graphes adaptés à l'objet sont utilisés pour gérer les grandes rotations en profondeur. L'extraction du graphe d'image est basée sur une nouvelle structure de données, le graphe de paquets, qui est construit à partir d'un petit ensemble de graphes d'images échantillons.

M. Vasanthi et al dans [78] ont introduit une nouvelle approche pour la reconnaissance d'images faciales basée sur la biométrie, utilisant une analyse de corrélation multivariée qui extrait des points de caractéristiques géométriques et des caractéristiques visuelles de bas niveau. Les caractéristiques visuelles de bas niveau, telles que la couleur et la texture, sont extraites localement des régions saillantes sélectionnées de l'image faciale. Les caractéristiques géométriques sont capturées à l'aide du modèle de forme active (Active Shape Model ASM) [79]. Les caractéristiques de couleur sont extraites du modèle de couleur YCbCr (Y: la Luminance, Cb : la Chrominance Bleue, Cr : la Chrominance Rouge), et la méthode d'autocorrélation est utilisée pour extraire les caractéristiques de texture. Les caractéristiques extraites sont formées en une matrice tensorielle de caractéristiques. La matrice de caractéristiques de l'image faciale clé est comparée aux matrices de caractéristiques de l'image faciale cible stockées dans la base de données de vecteurs de caractéristiques en utilisant la méthode de corrélation canonique. La corrélation entre les matrices de caractéristiques clé et cible est testée pour voir si elle est hautement significative ou non. Si une corrélation significative est trouvée, cela signifie que les images faciales clé et cible sont identiques. Les bases de données d'images faciales de référence, tels que GT, LFW, ont été considérés pour les expériences ; en plus de ces bases de données, une base de données d'images faciales a été construite avec des célébrités. La méthode proposée a abouti à un score de précision moyen de 95,27 %, 94,20 %, 96,19 % et 96,05 % pour les bases de données GT, LFW, Pointing'04 et la base de données construite par les chercheur, respectivement. De plus, le F-score a été calculé, soit 96,78 %, 95,15 %, 97,08 % et 96,96 % pour les mêmes

bases de données, respectivement. Les résultats obtenus par la méthode proposée sont comparables à ceux des méthodes existantes.

D. Zheng et al. [80] ont proposé une nouvelle technique d'alignement discriminant actif inter-domaines ADCDA (Active Discriminative Cross-Domain Alignment) pour la reconnaissance faciale basse résolution. Cette technique explore conjointement et de manière unique les propriétés géométriques et statistiques du domaine source et du domaine cible. Plus précisément, la méthode ADCDA proposée s'appuie sur trois éléments clés : 1) elle réduit simultanément le décalage de domaine à la fois dans la distribution marginale et la distribution conditionnelle entre le domaine source et le domaine cible, 2) elle aligne les données des deux domaines dans un sous-espace latent commun grâce à un alignement discriminant local, 3) elle sélectionne des échantillons représentatifs et variés à l'aide d'une stratégie d'apprentissage actif afin d'améliorer encore les performances de classification. Des expériences approfondies menées sur six bases de données de référence (YALE-B, UMIST, ORL, FERET, CMU-PIE, et AR) confirment que la méthode proposée surpasse significativement les approches précédentes à la pointe de la technologie.

Pour conclure, de nombreuses méthodes, basées sur l'extraction de points caractéristiques, ont été proposées. Elles peuvent être utilisées efficacement pour la reconnaissance faciale lorsque seule une image de référence est disponible. Cependant, leur performance dépend de nombreux algorithmes efficaces pour localiser les points caractéristiques du visage. En pratique, la tâche de détection de points caractéristiques précis n'est pas facile et n'a pas été complètement résolue, en particulier dans les cas où la forme ou l'apparence d'une image faciale peut varier largement.

2.7.2.2. Méthodes de reconnaissance faciale basées sur l'apparence locale

Les méthodes de reconnaissance faciale basées sur l'apparence locale exploitent des caractéristiques locales du visage pour identifier ou vérifier l'identité d'un individu. Contrairement aux approches qui utilisent des représentations globales du visage, celles basées sur l'apparence locale se concentrent sur des régions spécifiques ou des motifs caractéristiques, tels que les yeux, le nez et la bouche, ainsi que les textures et les contours qui les entourent.

Muqet et Holambe dans [81] ont introduit une nouvelle approche pour extraire des caractéristiques faciales robustes aux variations d'expressions et de poses. La méthode utilise des caractéristiques d'histogramme **LBP** basées sur la Transformée en Ondelettes Directionnelles (DIWT, Directional Wavelet Transform) et utilise un schéma de partitionnement en quadtree efficace pour mettre en œuvre la DIWT. En utilisant la DIWT, l'approche permet une sélection directionnelle adaptative basée sur les caractéristiques de l'image et représente les ensembles de bords de l'image. La combinaison de caractéristiques d'histogramme LBP multi-régions des sous-bandes de niveau supérieur {LL, HL, LH (**L**:"Low", **H**:"High.")} forme un ensemble de caractéristiques très efficace. Pour évaluer la méthode proposée, divers bases de données d'images faciales sont utilisés et les résultats démontrent sa capacité de discrimination supérieure. Comparé à d'autres méthodes, l'approche proposée obtient les meilleurs résultats de reconnaissance au rang un. Les résultats expérimentaux indiquent que ce travail surpasse les approches holistiques, telles que la technique LDA de caractéristiques de texture et les Projections Préservant la Localité (LPP), ainsi que les descripteurs locaux, tels que LBP, les Motifs Directionnels Locaux (LDP) et les Descripteurs Locaux de Weber (WLD), lorsqu'il s'agit d'images faciales présentant différents niveaux d'expressions et de variations de poses. De plus, ce travail présente de meilleures performances par rapport aux méthodes d'analyse multi-résolution LBP non adaptatives, telles que les Motifs Binaires Gabor Locaux (LGBP), LSPBPS (Local Steerable Pyramid Binary Pattern Sequence) et CTLBP (Curvelets and Local Binary Pattern Features) [82].

Dans [4], une nouvelle approche pour la reconnaissance faciale en trois dimensions est présentée, combinant les descripteurs de caractéristiques LBP et un classifieur SVM. La méthode proposée comprend deux étapes principales. Tout d'abord, l'algorithme LBP est utilisé pour extraire des informations de caractéristiques pertinentes à partir de l'image de profondeur du visage en trois dimensions. Ensuite, l'algorithme SVM est utilisé pour classer ces caractéristiques extraites. Pour évaluer l'efficacité de la méthode proposée, des échantillons sont sélectionnés à partir du Texas Three-dimensional Face Recognition et d'une base de données de profondeur personnalisée. Les résultats expérimentaux montrent que l'algorithme atteint un taux de reconnaissance plus élevé tout en réduisant le temps de calcul nécessaire pour la reconnaissance. Cette approche offre des perspectives prometteuses pour la reconnaissance faciale en trois dimensions.

Dans [83], l'utilisation du LBP pour l'extraction de caractéristiques de texture, ainsi que l'utilisation de ces caractéristiques comme données d'entraînement pour le CNN parallèle, améliore efficacement la précision de la reconnaissance faciale en réduisant les effets négatifs des variations d'éclairage sur les caractéristiques faciales. L'architecture du CNN intègre plusieurs composants cruciaux pour améliorer ses performances. Le module inception est employé pour élargir le réseau et améliorer sa capacité à représenter des caractéristiques complexes. La normalisation par lots accélère le processus d'entraînement et améliore la convergence. De plus, des connexions de saut sont intégrées pour faciliter la circulation de l'information entre différentes couches, augmentant ainsi la précision de la reconnaissance. La stratégie d'apprentissage en ensemble parallèle transforme la structure du réseau d'un seul réseau en un ensemble, augmentant considérablement la précision et les capacités de généralisation de l'approche proposée. Pour évaluer les performances de cette méthode, des expériences complètes ont été menées, la comparant avec trois autres méthodes : l'Analyse en Composantes Principales (PCA), HOG-CNN et CNN utilisées de manière indépendante. Les résultats constamment supérieurs démontrent l'efficacité de l'approche proposée dans les tâches de reconnaissance faciale, mettant en évidence sa précision et son efficacité remarquables dans la gestion des défis liés à l'éclairage.

Lakshmi et Ponnusamy dans [84] ont introduit un nouveau descripteur de caractéristiques pour la reconnaissance des expressions faciales. L'approche proposée combine les descripteurs de caractéristiques HOG et LBP modifiés. La méthodologie se compose de plusieurs étapes. Tout d'abord, l'algorithme de détection de visages Viola-Jones est utilisé pour localiser la région faciale. Ensuite, un filtre passe-haut de Butterworth est appliqué pour améliorer la région détectée, permettant l'identification des régions des yeux, du nez et de la bouche grâce à l'approche Viola-Jones. À l'étape suivante, les descripteurs de caractéristiques HOG et LBP modifiés proposés sont utilisés pour extraire des caractéristiques des régions des yeux, du nez et de la bouche détectées. Ces caractéristiques sont ensuite concaténées et leur dimensionnalité est réduite à l'aide de codeurs automatiques empilés en profondeur. Enfin, un classifieur à vecteur de support multi-classes est utilisé pour la classification et la reconnaissance des expressions faciales. Les résultats expérimentaux démontrent l'efficacité des descripteurs de caractéristiques modifiés proposés dans la reconnaissance précise des émotions sur les bases de données CK+ et JAFFE.

Dans [85], les auteurs ont introduit une nouvelle méthode de reconnaissance faciale qui résout efficacement les problèmes liés à l'éclairage et au désalignement. Leur approche combine les descripteurs de caractéristiques **LBP** avec une version améliorée de la méthode d'apprentissage de plusieurs métriques. Initialement, le LBP est utilisé pour extraire des caractéristiques de texture à partir des images de visages. Ensuite, l'Analyse Discriminante Linéaire (LDA) sert à réduire la dimensionnalité de ces caractéristiques. Les caractéristiques de Fisher sont alors divisées en sous-blocs, chaque bloc étant traité comme un vecteur colonne. L'utilisation de l'IPMML (Improved Pairwise-constrained Multiple Metric Learning method) [86] permet de dériver une matrice de Mahalanobis optimale, utilisée pour calculer la distance discriminante nécessaire à la reconnaissance faciale. Enfin, un classifieur du K Plus Proche Voisin (KPPV) est employé pour classer les images de visages. Les résultats expérimentaux ont démontré l'efficacité de cette méthode, qui atteint des taux de reconnaissance élevés et montre une robustesse face aux variations d'éclairage, d'expression faciale et aux images de visages désalignés. Par ailleurs, une autre approche proposée par Tang et al [83]. associe l'apprentissage parallèle d'ensemble des descripteurs de caractéristiques LBP et CNN pour la reconnaissance faciale.

L'étude décrite dans [3] se concentre sur la mise en œuvre pratique d'un algorithme sophistiqué appelé FaceNet. Cet algorithme est utilisé au sein d'un système de contrôle d'accès conçu pour détecter efficacement les visages et les yeux, même dans des conditions d'éclairage difficiles. Cette capacité de détection est obtenue grâce à l'utilisation de techniques de codage du visage. De plus, l'extraction des caractéristiques faciales est effectuée en utilisant l'algorithme **HOG** [9, 59]. Le système de contrôle d'accès comprend une fonction de comparaison de visages qui intègre un Classifieur de machine à vecteurs de support (SVM) pour classer les encodages faciaux et générer la sortie souhaitée. Afin d'améliorer encore la fonctionnalité du système, des capteurs RFID (Radio Fréquence Identification Distance) et des capteurs infrarouges sont intégrés de manière transparente. Dans l'ensemble, l'étude présente une solution complète pour le contrôle d'accès en mettant en œuvre efficacement l'algorithme FaceNet en conjonction avec divers capteurs et une interface conviviale sur une page web.

2.7.3. *Approches hybrides*

Cette troisième catégorie comprend des approches qui utilisent simultanément des caractéristiques globales et locales afin d'exploiter les avantages de ces deux méthodes [51]. Elle inclut, également, les techniques basées sur les modèles statistiques. Ce dernier formalise les relations entre les variables sous forme d'équations mathématiques qui décrivent comment une ou plusieurs variables aléatoires sont liées à une ou plusieurs autres variables aléatoires. Ce modèle est considéré comme statistique lorsque les variables ne sont pas déterministes, mais stochastiquement liées [53].

Kar et al. [87] ont proposé une nouvelle méthodologie combinant la Transformée en Ondelettes de Gabor (Gabor Wavelet Transform **GWT**) et le Pseudo Modèle de Markov Caché (Pseudo Hidden Markov Model **PHMM**) pour la reconnaissance faciale. Contrairement à la méthode traditionnelle de balayage en zigzag, la méthode proposée utilise un balayage en spirale pour l'extraction de caractéristiques. Ce balayage en spirale permet une meilleure sélection des caractéristiques faciales. Le PHMM utilise le concept de structures locales pour identifier les zones faciales importantes et extraire les caractéristiques les plus discriminantes d'une image de visage. Le PHMM permet de réduire significativement le problème de dépendance à long terme inhérent aux HMM traditionnels. La méthode proposée utilise les pixels les plus informatifs au lieu de l'image entière, ce qui la rend plus rapide pour la reconnaissance faciale. Cette méthode a été testée avec succès sur des images de faces frontales provenant des bases de données ORL, FRAV2D et FERET avec une Précision de la reconnaissance faciale de 91.5% [87].

Dans [88], les auteurs proposent une méthode de reconnaissance faciale basée sur un modèle de Markov caché à quatre états (HMM4). Cette méthode vise à réduire la taille des vecteurs de caractéristiques en tronquant leurs coefficients. Les images faciales sont divisées en une séquence de blocs superposés. Une séquence d'observation contenant les coefficients des valeurs propres et des vecteurs propres de ces blocs est utilisée pour entraîner le modèle. Chaque individu est associé à un HMM distinct. Afin de minimiser la complexité de calcul du modèle proposé, une Transformée en Ondelettes Discrètes (DWT) en prétraitement est appliquée. De plus, une Décomposition en Valeurs Singulières (SVD) est appliquée aux images faciales. Un seuil empirique sur les valeurs singulières permet de rejeter ou d'accepter les images test. L'extraction des

caractéristiques est réalisée à l'aide de l'Analyse en Composantes Principales (PCA), ensuite, la classification des images test acceptées se base sur un vote majoritaire en utilisant différentes séquences d'observation des caractéristiques de l'image. Les expérimentations menées sur les bases de données Yale et ORL, en présence de bruit de type sel et poivre et en environnement sans bruit, démontrent que la méthode proposée atteint une précision de reconnaissance comparable aux techniques existantes, tout en offrant un coût de calcul réduit.

Dans leur étude présentée dans [89], les auteurs ont introduit une fonction de perte innovante nommée "Large Margin Cosine Loss" (LMCL). Cette fonction de perte redéfinit la perte Softmax en une perte cosinus, obtenue par une normalisation L2 des vecteurs de caractéristiques et des vecteurs de poids afin d'éliminer les variations. De plus, un terme de marge cosinus est intégré pour améliorer la marge de décision dans l'espace angulaire. En conséquence, cette approche permet de minimiser la variance intra-classe tout en maximisant la variance inter-classes, grâce à la normalisation et à l'amélioration de la marge de décision basée sur le cosinus.

Dans [90], Wanling et Shijun ont proposé une approche efficace de détection de la contrefaçon de visage utilisant une combinaison de la Transformée en Ondelettes Discrètes (Discrete Wavelet Transform DWT), de la Transformée en Cosinus Discrète (Discrete Cosine Transform DCT) et de l'histogramme des motifs locaux (LBP), ainsi qu'un classifieur SVM. Cette stratégie se déroule en plusieurs étapes. Tout d'abord, des caractéristiques DWT sont générées par la décomposition d'images sélectionnées en différentes composantes de fréquence au sein de blocs multi-résolutions de 8x8. Ensuite, des caractéristiques DWT-LBP sont construites pour capturer l'information spatiale de ces blocs en reliant horizontalement les histogrammes LBP des blocs DWT correspondants dans chaque image. Par la suite, des caractéristiques DWT-LBP-DCT sont obtenues en appliquant verticalement des opérations DCT sur les caractéristiques DWT-LBP, incorporant ainsi des informations temporelles issues du fichier vidéo. Ce processus permet aux caractéristiques extraites DWT-LBP-DCT de représenter efficacement les caractéristiques fréquentielles-spatio-temporelles de la vidéo. Enfin, un classifieur SVM avec un noyau à fonction radiale (RBF) est formé pour la détection de la contrefaçon de visage. Les évaluations expérimentales menées sur deux bases de données de référence, à savoir REPLAY-ATTACK et CASIA-FASD, ont révélé que

l'approche proposée atteint des performances de détection élevées par rapport aux méthodes existantes.

Dans [5], un nouvel algorithme basé sur la pyramide Laplacienne pour la reconnaissance faciale 3D profonde, avec des applications pratiques dans des environnements publics, est proposé. Cet algorithme intègre des techniques de fusion multi-mode, d'alignement 3D dense, et de fusion résiduelle à plusieurs échelles. L'approche commence par utiliser une méthode de représentation de structure de 2D à 3D pour capturer efficacement les informations des points de repère faciaux clés et effectuer une modélisation d'alignement dense. Ensuite, un réseau de profondeur Laplacienne à cinq couches est construit en utilisant le modèle de points de repère faciaux 3D. Au cours du processus d'entraînement, un poids résiduel multi-échelle est intégré dans la fonction de perte pour améliorer les performances du réseau. Pour garantir des performances en temps réel, le réseau proposé est conçu comme une cascade de bout en bout. Cette conception permet à la fois une identification précise et un dépistage efficace du personnel, en particulier dans le contexte de mesures de contrôle épidémique. L'algorithme permet une reconnaissance faciale rapide et de haute précision, facilitant la création d'une base de données de visages 3D. Il démontre son adaptabilité et sa robustesse dans des environnements difficiles caractérisés par une faible luminosité et du bruit, tout en étant capable de gérer diverses couleurs de peau et postures pour la reconstruction et la reconnaissance faciale.

Dans [91], une nouvelle approche de reconnaissance faciale utilisant la méthode YOLO-Face pour la détection des visages a été proposée. Pour l'étape de classification, les auteurs ont exploré l'idée de remplacer la couche entièrement connectée d'un réseau de neurones convolutionnel (CNN) par une machine à vecteurs de support (SVM). Ils ont également analysé l'utilisation d'une forêt aléatoire (RF) et de la méthode des k Plus Proches Voisins (KNN). Leurs résultats expérimentaux ont montré que le modèle FaceNet+SVM atteignait un taux de précision élevé de 99,7 % sur la base de données LFW. De plus, les modèles FaceNet+KNN et FaceNet+RF ont atteint des précisions respectives de 99,5 % et 89,1 % sur cette même base de données.

Mamieva et al. [92] ont introduit une nouvelle technique de détection de visages basée sur l'apprentissage profond. Cette technique comprend deux composants : un Réseau de Proposition de Régions (Region Offering Network "RON") et un réseau de prédiction.

Le RON génère une liste de propositions de zones susceptibles de contenir des visages ou des Régions d'Intérêt (Regions of Interest "RoIs"). Le réseau de prédiction est chargé de classifier ces zones et d'affiner les cadres autour des visages détectés. Les deux composants partagent des paramètres communs avec les couches de convolution d'extraction de caractéristiques, permettant ainsi à l'architecture d'obtenir des performances compétitives dans les tâches de détection de visages. Pour former le modèle, les auteurs ont utilisé la base de données WIDER FACE. Les résultats expérimentaux montrent que leur méthode excelle dans les tâches d'identification de visage, en obtenant une précision plus élevée malgré une taille de modèle réduite et une efficacité de calcul accrue.

Dans [93], les auteurs ont introduit un algorithme innovant appelé "Iterative Multi-Output Random Forests" (iMORF) pour améliorer les performances dans plusieurs tâches d'analyse faciale. Cet algorithme modélise explicitement les relations entre ces tâches et les exploite de manière itérative pour améliorer les performances globales. L'algorithme iMORF adopte une approche hiérarchique de l'analyse faciale, où une forêt de premier niveau est dédiée à la classification de la pose et de l'expression, tandis qu'une forêt de niveau inférieur se concentre sur la régression des positions des points de repère. En estimant la pose et l'expression, l'algorithme intègre une contrainte forte qui limite la variation des positions des points de repère. Les positions estimées de ces points offre, ainsi, des caractéristiques liées à la forme la plus discriminative, améliorant les prédictions de la pose et de l'expression. Cette exploitation itérative des interconnexions entre les tâches d'analyse faciale se poursuit à travers des forêts hiérarchiques en cascade jusqu'à atteindre la convergence. Les expériences menées sur des bases de données d'images faciales du monde réel disponibles publiquement ont démontré que l'algorithme iMORF améliore significativement les performances de chaque tâche individuelle impliquée dans l'analyse faciale.

Le travail proposé dans [94] a introduit plusieurs modifications pour améliorer les performances du modèle de réseau pour la reconnaissance faciale. Ces modifications incluent le remplacement de la couche de convolution traditionnelle par une couche de convolution MLP pour améliorer l'extraction des caractéristiques. De plus, la fonction d'activation MFM (Max Feature Map) est incorporée pour séparer efficacement les signaux de bruit des signaux d'information, améliorant ainsi la reconnaissance.

L'inclusion de la fonction de perte Center Loss réduit la distance entre les éléments et améliore la généralisation des caractéristiques apprises. Grâce à des expériences approfondies, le modèle de réseau présente des résultats prometteurs. Dans des expériences de classification de prédiction faciale à grande échelle, le modèle atteint un taux de reconnaissance de 82,3 %. De plus, dans des expériences de vérification faciale réalisées sur la base de données de visages LFW, le modèle atteint un taux de précision de 84,5 %, ce qui indique de hautes performances de reconnaissance. Les expériences menées sur des images faciales capturées dans différentes conditions mettent en évidence la robustesse du modèle de réseau, à l'exception d'une légère réduction de la précision dans la vérification faciale avec des visages de côté. Globalement, le modèle de réseau présente une reconnaissance efficace.

Dans [6], un cadre complet nommé 3DPalsyNet a été introduit pour détecter le mouvement de la bouche et évaluer la paralysie faciale. Ce cadre emploie une architecture CNN 3D modifiée, avec une épine dorsale ResNet, conçue pour capturer les actions dynamiques des vidéos. Les performances de l'architecture proposée ont été évaluées sur deux bases de données (The Extended Cohn-Kanade (CK+) et Facial Palsy dataset [95]), révélant un score F1 de 82 % pour la détection des mouvements de la bouche et un impressionnant score F1 de 88 % pour l'évaluation de la paralysie faciale.

Dans [96], l'étude présente un réseau d'incrustation de caractéristiques basé sur ResNet-100 associé à des fonctions de perte de pointe, notamment Center Loss, Marginal Loss, Angular Softmax Loss, Large Margin Cosine Loss et Additive Angular Margin Loss. Ils effectuent une évaluation complète impliquant des tâches de vérification et d'identification de visage, en utilisant les bases de données IJB-B et IJB-C pour évaluer les performances en matière de variations de **P**ose, d'**I**llumination) et d'**E**xpression (PIE), la base de données FG-Net pour l'analyse liée à l'âge et SCface (Surveillance Cameras Face database [97]) pour des scénarios d'images à faible résolution. La base de données MS-1MV2 est utilisée comme principale base de données d'entraînement pour le développement du système. Suite à cela, l'étude évalue les performances du réseau avec la fonction de perte la plus adaptée pour la reconnaissance de visages masqués synthétiques sur la base de données de visages masqués réels, la base de données RMFRD nettoyée (c-RMFRD).

Dans [98], Zhou et Feng ont présenté une nouvelle technique d'ensemble d'arbres de décision novatrice appelée gcForest (forêt en cascade multi-échelonnée). Cette méthode construit un ensemble de forêts profondes avec une structure en cascade, permettant un apprentissage de représentation efficace. Grâce à la détermination adaptative des niveaux de cascade, gcForest peut ajuster automatiquement la complexité du modèle, ce qui donne d'excellents résultats même avec des données limitées. Notamment, gcForest présente une réduction substantielle du nombre de paramètres hyperparasites par rapport aux réseaux de neurones profonds. Les conclusions expérimentales de leur travail montrent que gcForest atteint des performances très compétitives au niveau des réseaux de neurones profonds.

Maafiri Ayyad et al. [99] ont présenté une nouvelle méthode d'extraction de caractéristiques pour une reconnaissance faciale robuste appelée LBP-WKPCA (Local Binary Pattern et Wavelet Kernel PCA). La méthode proposée vise à extraire les informations discriminantes et robustes afin de minimiser les erreurs de reconnaissance. Ceci est obtenu tout d'abord par l'utilisation d'un algorithme de projection non linéaire performant appelé RKPCA (Robust face recognition based on a new Kernel-PCA). Ensuite, ils ont adapté l'algorithme pour réduire la dimensionnalité des caractéristiques extraites en utilisant la transformation proposée du Motif Local Binaire Couleur et du descripteur ondelettes. L'idée générale de ce descripteur est de trouver la meilleure représentation d'une image de visage dans une structure vectorielle discriminante grâce à une stratégie de groupement de caractéristiques originale générée par la décomposition à trois niveaux de la transformée en ondelettes discrètes bidimensionnelle (2D-DWT) et du motif local binaire (LBP). Des expériences approfondies sur quatre bases de données de visages bien connus, à savoir ORL, GT, LFW et YouTube Celebrities, montrent que la méthode proposée atteint un taux de reconnaissance de 100 % pour ORL, de 96,84 % pour GT, de 99,34 % pour LFW et de 95,63 % pour YouTube Celebrities.

2.8. Bases de données de référence pour la reconnaissance faciale

Cette section présente les principales bases de données utilisées pour la reconnaissance faciale. Ces bases comprennent des contraintes spécifiques reflétant des situations réelles, permettant aux chercheurs d'évaluer leurs techniques et approches de reconnaissance faciale. Un résumé de ces bases de données est fourni, avec leurs caractéristiques et leur pertinence.

2.8.1. Base de données Yale

La base de données de visages de Yale [100] (voir Figure 2.2) se compose d'une collection de 165 images de 15 individus, chacun contribuant avec 11 images distinctes. Ces images sont prises dans diverses conditions, notamment des variations de conditions d'éclairage (lumière droite, lumière centrale et lumière gauche), d'expressions faciales (normale, triste, heureuse, surprise, endormie et clin d'œil), ainsi qu'avec ou sans lunettes [101]. De légers changements dans l'expression faciale et la position de la tête sont perceptibles, car toutes les 64 images d'un visage pour une pose particulière ont été capturées en environ 2 secondes. La base de données de visages de Yale a été subdivisée en quatre sous-bases différents selon l'angle entre la source de lumière et l'axe de la caméra (120, 250, 500, 770 degrés). Les emplacements étiquetés à la main des deux yeux et du centre de la bouche sont inclus avec la base de données.

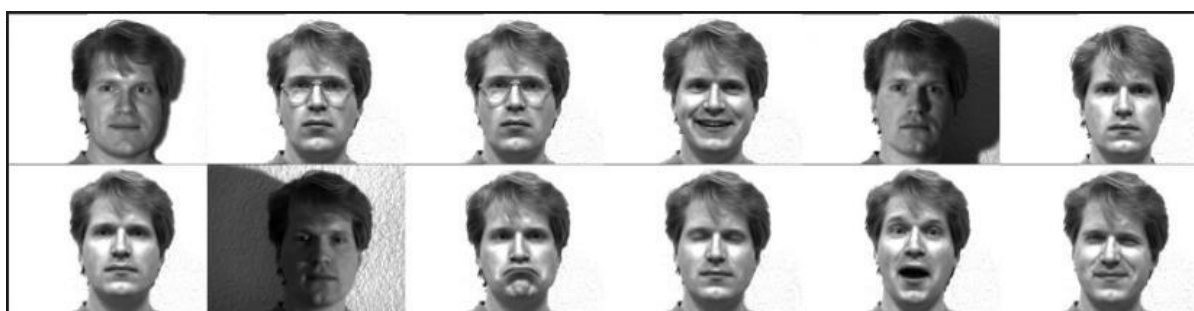


Figure 2.2: Exemple d'images de la Base de données Yale [101].

2.8.2. Base de données LFW

La base de données Labelled Faces in the Wild (LFW) [47] (voir Figure 2.3) est une collection de 13 233 images de visages collectées sur le Web et étiquetées avec le nom

de chaque personne. Il se compose de 5 749 individus différents, dont 1 680 ont au moins deux images distinctes dans le jeu de données [102]. Les 4 069 restants ne contiennent qu'une seule image. La plupart des images du jeu de données LFW sont en couleur et disponibles en JPEG 250 × 250 pixels avec seulement quelques images en niveaux de gris [102].



Figure 2.3: Exemple d'images de la base de données LFW [102, 103].

2.8.3. Base de données FERET

La base de données de technologie de reconnaissance faciale (FERET) [104] (voir Figure 2.4) est une collection de 24 catégories d'images faciales, obtenues à partir de l'Université George Mason et du laboratoire de recherche de l'armée américaine dans le cadre du programme FERET parrainé par le programme de développement et de technologie de lutte contre la drogue du ministère de la Défense des États-Unis [105]. FERET a été largement utilisé dans les travaux passés. Les images faciales ont été enregistrées entre août 1993 et juillet 1996 sur 15 sessions. Un appareil photo 35 mm a été utilisé pour l'enregistrement d'images avant de les numériser et de les convertir en images en niveaux de gris 8 bits de 256 X 384 pixels. La base de données FERET contient peu de variations entre les sessions d'enregistrement en raison de la nécessité de réassembler l'équipement d'enregistrement pour chaque session [87].

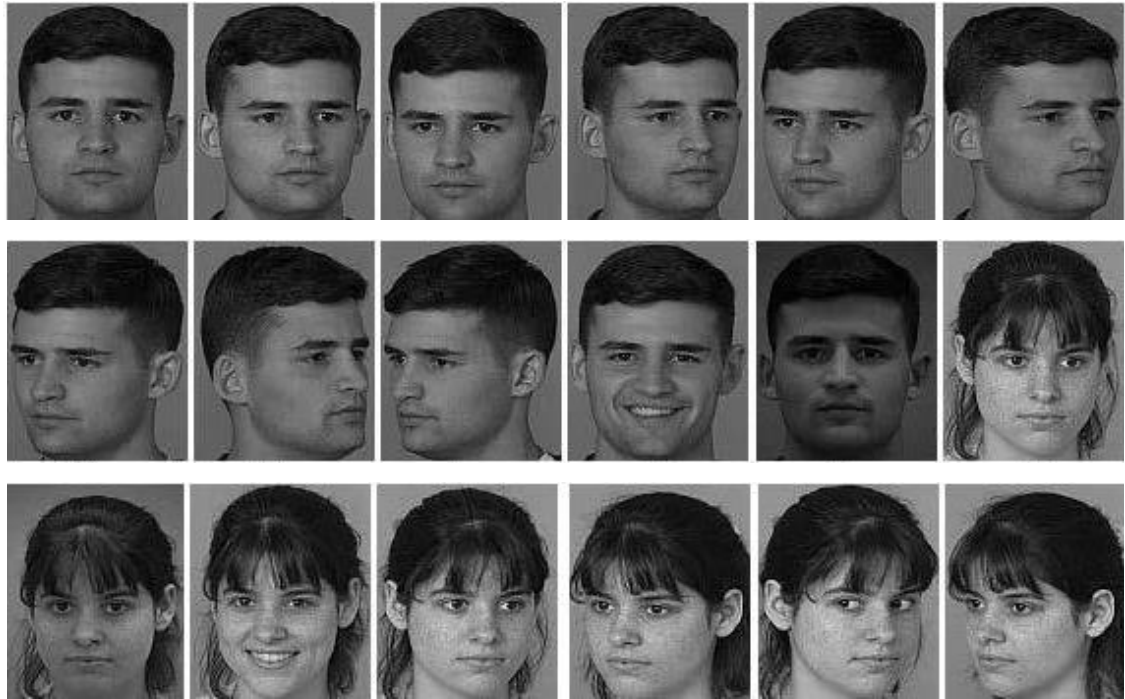


Figure 2.4: Exemple d'images de la base de données FERET. [87]

2.8.4. *Base de visages de chirurgie esthétique (Plastic Surgery Face Dataset)*

La base de visages de chirurgie esthétique [106] (voir Figure 2.5) est une collection de visages illustrant un scénario réel de l'impact de la chirurgie sur les systèmes de reconnaissance faciale. La base contient 1800 images de visages avant et après la chirurgie, appartenant à 900 individus différents. 519 de ces individus représentent des situations de chirurgies localisées, c'est-à-dire des changements sur une partie du visage, et 381 individus représentent des cas de chirurgies globales, c'est-à-dire un changement total de l'image du visage. La base de données contient différents types de chirurgies esthétiques, notamment la rhytidectomie (lifting facial), la rhinoplastie (chirurgie du nez) et la blépharoplastie (chirurgie des paupières) [107].



Figure 2.5: Exemple d'images de la Base de données Plastic Surgery Face. [107]

2.8.5. Base de données AT&T/ORL

La base de données de visages AT&T/ORL [63] (voir Figure 2.6) est une base de données classique utilisée pour évaluer les performances des algorithmes de reconnaissance faciale. Elle contient 40 individus, chacun représenté par 10 images distinctes. Ces images ont été prises entre 1992 et 1994 par le groupe de Vision et Robotique du laboratoire du département d'ingénierie de l'Université de Cambridge [60]. Elles présentent une grande variété d'expressions faciales, d'éclairages et d'accessoires. Malgré cette diversité, les images de la base de données AT&T/ORL présentent une composition cohérente. Les individus sont photographiés de face, contre un fond uniforme. Cette configuration permet aux chercheurs de se concentrer sur les subtilités des traits du visage, sans être perturbés par des éléments distrayants [60].



Figure 2.6: Exemple d'images de la base de données AT&T/ORL. [71]

2.8.6. Base de données IJB-C

La base de données IJB-C (IARPA Janus Benchmark-C [108]) (voir Figure 2.7) est largement utilisée dans le domaine de la reconnaissance faciale [109]. Il est conçu pour évaluer et améliorer les performances des algorithmes de reconnaissance faciale dans des conditions réelles difficiles. La base de données contient un total de 31 334 images fixes, dont 21 294 images comportent des visages humains et 10 040 images contiennent du contenu non facial. En moyenne, il y a environ 6 images disponibles pour chaque sujet dans la base de données. Ces images capturent diverses expressions faciales, poses et conditions d'éclairage, ce qui en fait un ensemble de données diversifié et stimulant pour les tâches de reconnaissance faciale. En plus des images fixes, IJB-C comprend 117 542 images extraites de 11 779 vidéos en mouvement complet. Chaque vidéo contient généralement plusieurs images des mêmes sujets [103], contribuant à une évaluation plus complète des algorithmes de reconnaissance faciale. IJB-C est accompagné d'un protocole d'évaluation bien défini qui spécifie comment diviser la base de données en bases d'entraînement et de test, ainsi que les métriques de performance utilisées pour évaluer les algorithmes de reconnaissance faciale [103].



Figure 2.7: Exemple d'images de la Base de données IJB-C [103].

2.8.7. Base de données MS-Celeb-1M

La base de données Microsoft Celeb [110] (voir Figure 2.8) est constituée d'environ 8,2 millions d'images faciales récupérées sur Internet dans le but de développer des technologies de reconnaissance faciale. Elle représente l'une des plus vastes collections publiques d'images faciales, regroupant environ 100 000 individus. La plupart des personnes présentes dans cette base de données sont des acteurs américains et britanniques [49], ainsi que des célébrités populaires sur Internet. Étant donné que les images sources n'ont pas été enregistrées délibérément, la base de données offre un soutien très limité pour l'illumination et les invariants de pose [49].

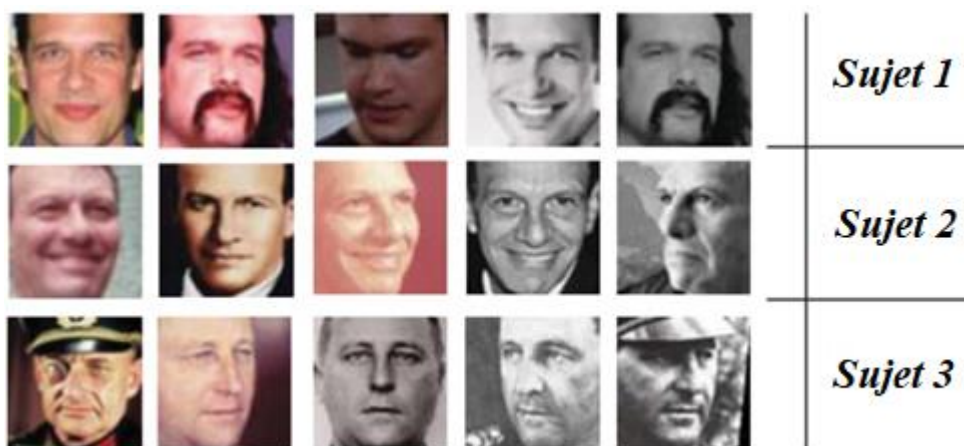


Figure 2.8: Exemple d'images de la base de données MS-Celeb-1M. [111]

2.8.8. Base de données CASIA-WebFace

La base de données CASIA-WebFace (CW [48]) (voir Figure 2.9) est une vaste collection d'images de visages de célébrités, constituant une ressource précieuse pour la recherche en reconnaissance faciale [112]. Cette base de données (CW) contient 494 414 images de visages de 10 575 identités réelles collectées sur le web [112]. Une méthode d'annotation automatique et semi-automatique a été mise en place pour associer chaque visage à l'identité correspondante. Le résultat est une base de données à grande échelle et hautement varié, qui surpasse en taille et en complexité de nombreuses autres bases de données existantes. CW est particulièrement utile pour entraîner et évaluer des modèles de reconnaissance faciale, notamment dans des conditions réelles où les images peuvent être de mauvaise qualité ou contenir plusieurs visages [112].



Figure 2.9: Exemple d'images de la base de données CASIA-WebFace [113].

2.8.9. Base de données CAS-PEAL

La base de données faciales Chinese Academy of Sciences Pose, Expression, Accessory, and Lighting (CAS-PEAL [114]) (voir Figure 2.10) a été mise en place en tant que projet clé du Programme national de haute technologie de l'Académie chinoise des sciences. L'objectif était de constituer un référentiel central d'images de visages provenant de différents groupes ethniques chinois. Cette base de données offre un solide support pour les invariants de pose, d'expression et d'éclairage, mais elle présente des limitations face aux défis liés au vieillissement et à l'occlusion. Il est important de noter que seuls les visages chinois sont disponibles, ce qui restreint l'évaluation des algorithmes [115].

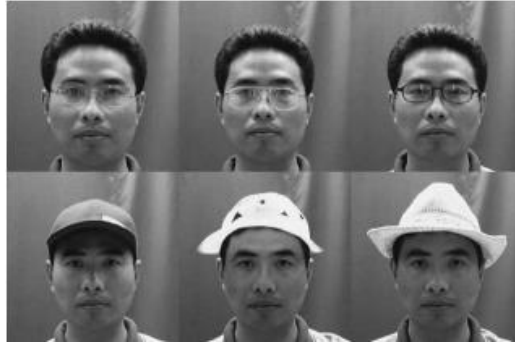


Figure 2.10: Exemple d'images de la base de données CAS-PEAL. [115]

2.9. Conclusion

Les systèmes biométriques basés sur la reconnaissance faciale ont prouvé leur supériorité par rapport aux systèmes utilisant d'autres modalités biométriques. Cette supériorité découle de leur praticité et de l'acceptation facile par les utilisateurs en raison de leur ergonomie. Les progrès technologiques ont permis de développer des systèmes biométriques performants et fiables, contribuant ainsi à améliorer la sécurité et la commodité dans de nombreux domaines d'application. Toutefois, il reste encore plusieurs défis à relever, notamment en matière de protection de la vie privée et de précision des systèmes. Malgré ces défis, le potentiel de la reconnaissance faciale pour faciliter la vie quotidienne et renforcer la sécurité demeure indéniable.

CHAPITRE

3

La vérité est une caractéristique
immuable de la science.

Albert Einstein- *physicien*

DETECTION DE VISAGE ET REPRESENTATION DES CARACTERISTIQUES

3.1. Introduction

Dans un système de reconnaissance biométrique, une fois la modalité acquise, on procède à la détection de la biométrie et l'extraction des caractéristiques nécessaires pour l'identification ou la vérification [116]. De nombreuses techniques ont déjà été proposées pour la détection de la biométrie et l'extraction des caractéristiques biométriques, notamment pour le visage. Chaque méthode de la détection de la biométrie et d'extraction biométrique présente ses avantages et ses inconvénients. Dans ce chapitre, nous allons passer en revue les principales techniques utilisées pour la détection biométrique et l'extraction des caractéristiques, avec un focus particulier sur la reconnaissance faciale.

3.2. Détection de visage

Le visage joue un rôle important dans la communication visuelle. En regardant un visage, les humains peuvent automatiquement extraire de nombreux messages non verbaux, tels que l'identité, l'intention et l'émotion des personnes. En vision par ordinateur, pour extraire automatiquement ces informations faciales, la localisation des points clés du visage est généralement une étape clé et de nombreuses méthodes d'analyse faciale reposent sur la détection précise de ces points de repère. Par exemple, la reconnaissance des expressions faciales et les algorithmes d'estimation de la pose de la tête peuvent fortement dépendre des informations de forme du visage fournies par les emplacements des points de repère. Les points de repère autour des yeux peuvent fournir une estimation initiale des positions du centre des pupilles pour la détection des yeux et le suivi du regard.

Les algorithmes de détection des points de repère faciaux visent à identifier automatiquement les emplacements des points de repère clés sur des images ou des

vidéos faciales. Ces points clés sont soit des points dominants décrivant l'emplacement unique d'un composant facial (par exemple, les coins de l'œil) ou un point interpolé reliant ces points dominants autour des composants faciaux et du contour du visage (voir Table 3.1). Formellement, étant donné une image faciale notée I , un algorithme de détection de points de repère prédit les emplacements de N points de repère $X = \{(x_1, y_1); (x_2, y_2); \dots; (x_n, y_n)\}$, où (x_i, y_i) représentent les coordonnées des points de repère faciaux sur l'image (voir Figure 3.1). La détection des points de repère faciaux est difficile pour plusieurs raisons. Tout d'abord, l'apparence faciale change considérablement entre les personnes sous différentes expressions faciales et poses de tête. Ensuite, les conditions environnementales, telles que l'illumination, affectent l'apparence des visages sur les images faciales.

Au cours des dernières décennies, il y a eu des développements significatifs des algorithmes de détection des points de repère faciaux. Les premiers travaux se concentrent sur les images faciales moins difficiles sans les variations faciales mentionnées précédemment. Plus tard, les algorithmes de détection de points de repère faciaux visent à gérer plusieurs variations au sein de certaines catégories, et les images faciales sont généralement collectées dans des conditions « contrôlées ». Par exemple, dans des conditions « contrôlées », les poses et expressions faciales peuvent seulement être dans certaines catégories. Plus récemment, la recherche se concentre sur les conditions difficiles « en situation réelle », où les images faciales peuvent subir des expressions faciales arbitraires, des poses de tête, des illuminations, des occlusions faciales, etc. En général, il manque encore une méthode robuste capable de gérer toutes ces variations.

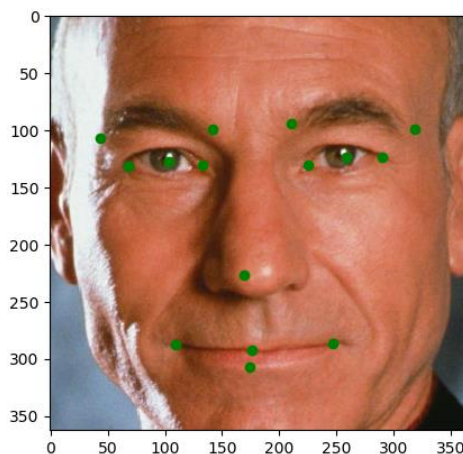


Figure 3.1: Exemple des points de repère faciaux "landmark".

3.3. Algorithmes de détection Landmark

Les algorithmes de détection Landmark peuvent être classés en fonction de la manière dont ils modélisent l'apparence faciale et les motifs de forme faciale en trois catégories principales : les méthodes holistiques, les méthodes de Modèle Local Contraint (CLM) et les méthodes basées sur la régression. L'apparence faciale fait référence aux motifs distinctifs d'intensité de pixel autour des points de repère faciaux ou dans toute la région du visage, tandis que les motifs de forme faciale font référence aux motifs des formes du visage tels que définis par les emplacements des points de repère et leurs relations spatiales. Les méthodes holistiques modélisent explicitement l'apparence faciale holistique et les motifs de forme faciale globale. Les CLM s'appuient sur l'apparence faciale locale explicite et les motifs de forme faciale globale explicite. Les méthodes basées sur la régression utilisent des informations d'apparence holistique ou locale et peuvent intégrer implicitement les motifs de forme faciale globale pour la détection conjointe des points de repère. En général, les méthodes basées sur la régression montrent des performances meilleures récemment.

3.3.1. Méthodes holistiques

Les méthodes holistiques exploitent explicitement les informations relatives à l'apparence faciale holistique ainsi que les motifs de forme faciale globale pour la détection des points de repère faciaux.

3.3.1.1. Modèle d'apparence active

Le Modèle d'Apparence Active (AAM) a été introduit par Taylor et al. [117]. Il s'agit d'un modèle statistique qui ajuste les images faciales à l'aide d'un petit nombre de coefficients, contrôlant à la fois les variations d'apparence et de forme du visage. Durant la construction du modèle, l'AAM construit le modèle de forme faciale globale et le modèle d'apparence faciale holistique séquentiellement, en se basant sur l'Analyse en Composantes Principales (PCA). Lors de la détection, il identifie les emplacements des points de repère.

3.3.1.2. Algorithmes d'ajustement

La plupart des méthodes holistiques se concentrent sur l'amélioration des algorithmes d'ajustement). Elles peuvent être classées en méthodes d'ajustement analytiques et méthodes d'ajustement basées sur l'apprentissage.

1. Méthodes d'ajustement analytiques

Les méthodes d'ajustement analytiques formulent le problème d'ajustement du Modèle d'Apparence Active (AAM) comme un problème d'optimisation non linéaire et le résolvent analytiquement. En particulier, l'algorithme recherche le meilleur ensemble de coefficients de forme et d'apparence p , λ qui minimise la différence entre l'image reconstruite et l'image testée avec une formulation des moindres carrés non linéaire.

2. Méthodes d'ajustement basées sur l'apprentissage

Au lieu de résoudre directement le problème d'ajustement analytiquement, les méthodes d'ajustement basées sur l'apprentissage apprennent à prédire les coefficients de forme et d'apparence à partir des apparences d'image. Elles peuvent être classées en méthodes d'ajustement de régression linéaire, méthodes d'ajustement de régression non linéaire et autres méthodes d'ajustement basées sur l'apprentissage.

Les méthodes d'ajustement de régression linéaire supposent qu'il existe une relation linéaire entre les mises à jour des coefficients du modèle et l'image d'erreur ou les caractéristiques de. Elles apprennent une fonction de régression linéaire pour la prédiction, qui suit l'AAM conventionnel.

Méthodes d'ajustement de régression non linéaire, Les méthodes d'ajustement de régression linéaire supposent que la relation entre les caractéristiques et l'image d'erreur autour de la vraie solution des coefficients du modèle est proche du quadratique, ce qui garantit qu'une procédure itérative avec des mises à jour linéaires et des tailles de pas adaptatives conduirait à la convergence. Cependant, cette hypothèse linéaire n'est vraie que lorsque l'initialisation est autour de la vraie solution, ce qui rend les méthodes d'ajustement de régression linéaire sensibles à l'initialisation. Pour résoudre ce problème, les méthodes de régression non linéaire utilisent des modèles non linéaires pour apprendre la relation entre les caractéristiques de l'image et les mises à jour des coefficients du modèle.

3.3.1.3. Autres extensions

1) Représentation des caractéristiques

Il existe d'autres extensions des méthodes AAM conventionnelles. Une direction particulière consiste à améliorer les représentations des caractéristiques. Il est bien connu que le modèle AAM a une capacité de généralisation limitée et qu'il a du mal à s'ajuster aux variations de visage non vues auparavant (par exemple, entre différents sujets, sous différentes illuminations, avec une occlusion partielle, etc.) [118]. Cette limitation est en partie due à l'utilisation de l'intensité brute des pixels comme caractéristiques. Pour résoudre ce problème, certains algorithmes utilisent des caractéristiques d'image plus robustes. Par exemple, dans [119], au lieu d'utiliser l'intensité brute des pixels, les caractéristiques des ondelettes sont utilisées pour modéliser l'apparence faciale. De plus, seule l'information d'apparence locale est utilisée pour améliorer la robustesse face à l'occlusion partielle et à l'illumination. Dans [120], l'ondelette de Gabor avec le modèle de mélange gaussien est utilisée pour modéliser l'apparence locale de l'image, ce qui permet une recherche rapide des points locaux. Les deux méthodes améliorent les performances de la méthode AAM conventionnelle.

2) Ensemble de modèles AAM

Un seul modèle AAM (Active Appearance Model) suppose intrinsèquement une linéarité dans la variation de la forme et de l'apparence du visage. Conscients de cette limitation, certaines méthodes utilisent des modèles en ensemble pour améliorer les performances. Par exemple, dans le travail [121] un modèle AAM de régression séquentielle est proposé, qui entraîne une série de modèles AAM pour un ajustement séquentiel dans une approche en cascade. Le modèle AAM au début prend en compte les grandes variations (par exemple, la pose), tandis que ceux des étapes ultérieures s'ajustent aux petites variations. Dans ce travail, à la fois des ensembles indépendants de modèles AAM et des modèles séquentiels couplés sont utilisés. Les ensembles indépendants de modèles AAM utilisent des coefficients de modèle de manière indépendante avec différents réglages, tandis que les modèles couplés appliquent le modèle de prédiction appris dans les premiers niveaux pour générer les données d'entraînement dans les niveaux suivants. La régression par boosting et la régression par

forêt aléatoire sont toutes deux utilisées pour prédire les mises à jour du modèle. De manière similaire aux ensembles couplés de modèles AAM dans les travaux de [121] et [122], le problème d'ajustement du modèle AAM est formulé comme un problème d'optimisation avec une solution par descente de gradient stochastique. Cela conduit à un algorithme approximé qui apprend de manière itérative le modèle de prédiction linéaire à partir des données d'entraînement avec des coefficients de modèle lors de la première itération, puis met à jour les coefficients du modèle pour les données d'entraînement à utiliser lors de l'itération suivante suivant une approche en cascade.

3.3.2. Méthodes locales contraintes

Comme illustré dans la Figure 3.2, Les Méthodes Locales Contraintes CLM [123] déduisent les emplacements des points de repère x en se basant sur les motifs globaux de la forme du visage ainsi que sur les informations d'apparence locale indépendante autour de chaque point de repère. Ces informations locales sont plus faciles à capturer et plus robustes face à l'illumination et à l'occlusion, en comparaison avec l'apparence globale.

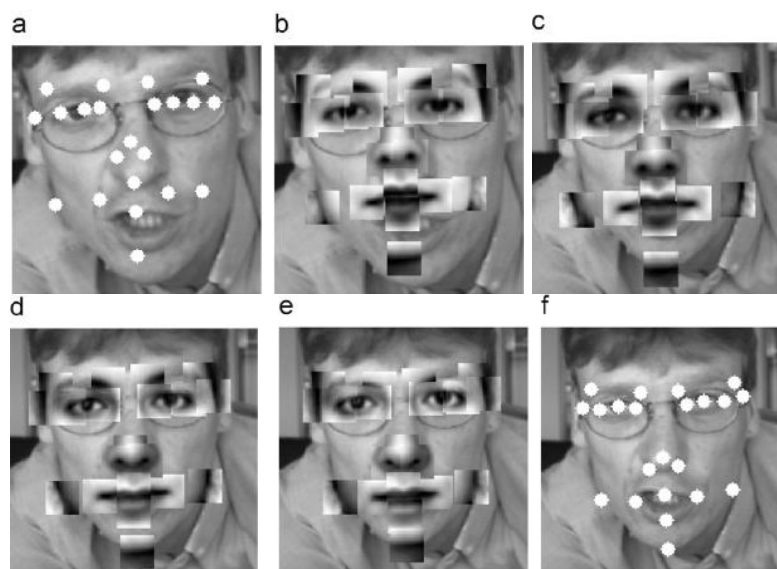


Figure 3.2: Évolution des modèles CLM lors de la recherche d'une image statique. (a) Points de départ, (b) modèles de départ, (c) après l'itération 1, (d) après l'itération 2, (e) après l'itération 3 et (f) points finaux. [124]

3.3.3. Méthodes basées sur la régression

Les méthodes basées sur la régression apprennent directement la correspondance entre l'apparence de l'image et les emplacements des points de repère. Contrairement aux méthodes holistiques et aux méthodes de modèle local contraint, elles ne construisent généralement pas explicitement de modèle global de forme du visage. Au lieu de cela, les contraintes de forme du visage peuvent être implicitement intégrées. En général, les méthodes basées sur la régression peuvent être classées en méthodes de régression directe, méthodes de régression en cascade, et méthodes de régression basées sur l'apprentissage profond. Les méthodes de régression directe prédisent les points de repère en une seule itération sans aucune initialisation, tandis que les méthodes de régression en cascade effectuent une prédiction en cascade et nécessitent généralement des emplacements initiaux des points de repère. Les méthodes basées sur l'apprentissage profond suivent soit la régression directe, soit la régression en cascade. Étant donné qu'elles utilisent des méthodes d'apprentissage profond uniques, nous les discutons séparément.

3.3.3.1. Méthodes de régression directe

Les méthodes de régression directe apprennent la correspondance directe entre l'apparence de l'image et les emplacements des points de repère du visage sans aucune initialisation des emplacements de ces points. Elles sont généralement effectuées en une seule étape [125]. Ces méthodes peuvent être classées en approches locales et approches globales. Les approches locales utilisent des patches d'image, tandis que les approches globales utilisent l'apparence holistique du visage.

3.3.3.2. Méthodes de régression en cascade

Contrairement aux méthodes de régression directe qui réalisent une prédiction en une seule étape, les méthodes de régression en cascade commencent par une estimation initiale des emplacements des points de repère du visage (par exemple, le visage moyen)[126], puis mettent à jour progressivement ces emplacements à travers différentes étapes en utilisant des fonctions de régression différentes apprises pour chaque étape (voir Figure 3.3). Plus précisément, lors de l'entraînement, à chaque étape des modèles de régression sont appliqués pour apprendre la correspondance entre les

apparences d'image indexées par la forme (par exemple, l'apparence locale extraite en fonction des emplacements des points de repère estimés actuellement) et les mises à jour de la forme. Le modèle appris lors de la première étape sera utilisé pour mettre à jour les données d'entraînement pour l'étape suivante. Lors du test, les modèles de régression appris sont appliqués de manière séquentielle pour mettre à jour les formes à travers les itérations Comme illustré dans la Figure3.3 [127].



Figure3.3: Illustration de détection faciale (points de repère) avec la Méthodes de Régression en Cascade.

3.3.3.3. Méthodes basées sur l'apprentissage profond

Récemment, les techniques d'apprentissage profond ont gagné en popularité en tant qu'outils très puissants pour résoudre les problèmes de vision par ordinateur. Pour la détection et le suivi des points de repère faciaux, il y a une tendance à passer des méthodes traditionnelles aux méthodes basées sur l'apprentissage profond. Dans les premiers travaux [128], le modèle de machine de Boltzmann profonde, qui est un modèle profond probabiliste, a été utilisé pour capturer les variations de forme du visage dues aux poses et aux expressions pour la détection et le suivi des points de repère faciaux. Plus récemment, les modèles de Réseaux de Neurones Convolutifs (CNN) (voir Figure 3.4) sont devenus les modèles d'apprentissage profond dominants pour la détection des points de repère faciaux, et la plupart d'entre eux suivent le cadre de régression directe globale ou le cadre de régression en cascade. Ces méthodes peuvent être largement classées en méthodes purement basées sur l'apprentissage et en méthodes hybrides. Les méthodes purement basées sur l'apprentissage prédisent directement les emplacements des points de repère faciaux, tandis que les méthodes hybrides combinent des méthodes d'apprentissage profond avec un modèle de projection en vision par ordinateur

pour la prédiction. Les méthodes de cette catégorie utilisent les puissants modèles CNN pour prédire directement les emplacements des points de repère à partir d'images faciales. Le travail précurseur de Sun et al. [129] prédit cinq points clés du visage de manière en cascade. Au premier niveau, un modèle CNN avec quatre couches de convolution est appliqué pour prédire les emplacements des points de repère à partir de l'image faciale. Ensuite, Plusieurs réseaux profonds affinent localement chaque point séparément.

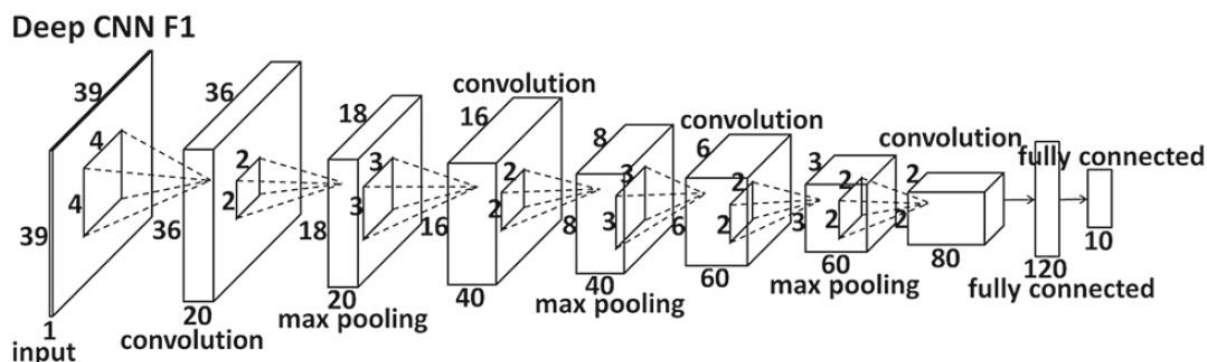


Figure 3.4. Structure du modèle CNN (Convolutional Neural Network).

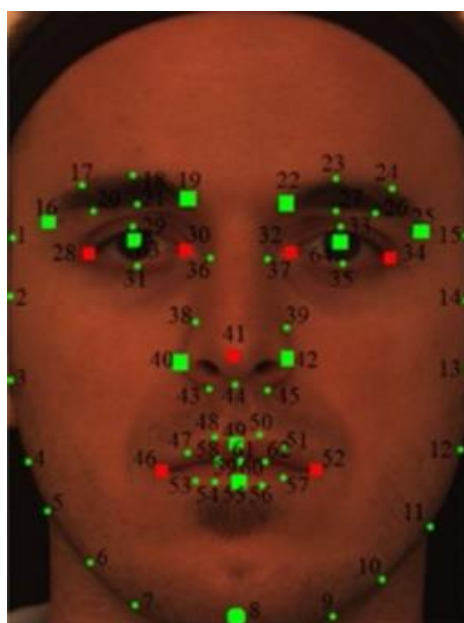


Figure 3.5: La numérotation des points de repère (Landmark).

Tableau 3.1: Les numéros des points de repère (Landmark) principaux et secondaires avec définition.

Points de repère principaux		Points de repère secondaires	
Numéro	Définition	Numéro	Définition
16	Coin extérieur du sourcil gauche	1	Tempe gauche
19	Coin intérieur du sourcil gauche	2-7, 9-14	Contours des joues
22	Coin intérieur du sourcil droit	8	Pointe du menton
25	Coin intérieur du sourcil droit	15	Tempe droite
28	Coin extérieur de l'œil gauche	16-19	Contours du sourcil gauche
30	Coin intérieur de l'œil gauche	22-25	Coins du sourcil droit
32	Coin intérieur de l'œil droit	29, 33	Centres de la paupière supérieure
34	Coin extérieur de l'œil droit	31, 35	Centres de la paupière inférieure
41	Pointe du nez	36, 37	Arêtes du nez
46	Coin gauche de la bouche	40, 42	Sommets du nez (narines)
52	Coin droit de la bouche	38-40, 42-45	Contours du nez
63, 64	Centres des yeux	47-51, 53-62	Contours de la bouche

3.4. Annotations des points de repère

Il existe généralement deux types de points de repère faciaux, les points clés du visage et les points de repère interpolés (voir Figure 3.5 et Table 3.1). Les points clés du visage sont les points de repère dominants sur le visage, comme les coins des yeux, la pointe du nez, les coins de la bouche, etc. Ils possèdent des motifs d'apparence/forme locale uniques. Les points de repère interpolés décrivent soit le contour du visage, soit connectent les points clés (voir Les Figure 3.8). Dans les premières recherches, seuls des points clés de repère clairsemés sont annotés et détectés. Récemment, davantage de points sont annotés dans les nouvelles bases de données. Par exemple, dans BioID [130], 20 points de repère sont annotés, tandis que 68 et 194 points de repère sont annotés dans les bases de données ibug et Helen [131, 132, 133] (voir Les Figures 3.6 ; 3.7 ; 3.8).

Il existe certains problèmes avec les annotations de points de repère existantes. Tout d'abord, les annotations de points de repère sont intrinsèquement biaisées et sont incohérentes entre les bases de données. Par conséquent, il est difficile de combiner plusieurs bases de données pour les évaluations. L'incohérence des annotations existe également pour l'annotation individuelle des points de repère. Par exemple, pour l'annotation des coins des yeux, certaines bases de données ont tendance à fournir une annotation à l'intérieur de la région de l'œil, tandis que d'autres peuvent annoter le point à l'extérieur de la région de l'œil. Pour résoudre ce problème, dans [134], une méthode est proposée pour combiner des bases de données avec différentes annotations de points de repère faciaux. Elle génère une union des annotations de points de repère en transférant les annotations de points de repère de la base de données source vers la base de données cible.



Figure3.6: Annotations des points de repère (20 points) [130].

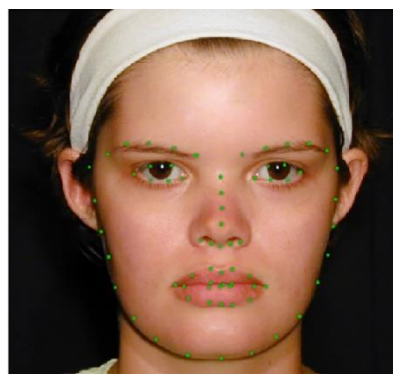


Figure3.7: Annotations des points de repère (68 points). [132]



Figure3.8: Annotations des points de repère (194 points). [133]

3.5. Extraction de caractéristiques faciales

L'extraction de caractéristiques est une tâche essentielle pour collecter l'ensemble des caractéristiques à partir d'une image. L'extraction de caractéristiques inclut la réduction de la taille des données qui représentent un grand ensemble de données. L'extraction de caractéristiques est le processus de récupération de caractéristiques précieuses à partir d'une image. Elle génère le vecteur de caractéristiques de toutes les caractéristiques extraites d'une image [135]. L'extraction de caractéristiques faciales peut être caractérisée comme un processus pour localiser des caractéristiques particulières dans une image faciale. C'est un processus pour extraire les caractéristiques faciales. Il existe diverses techniques d'extraction de caractéristiques telles que LBP, HOG, Cbfd, SURF, SIFT, le réseau de neurones convolutionnel (CNN), la transformée de Gabor, les Eigenfaces,...

3.5.1. Techniques d'extraction de caractéristiques faciales

Pour traiter les images, nous devons extraire les caractéristiques et les représenter sous forme de vecteurs de caractéristiques. Il y a plusieurs techniques d'extraction de caractéristiques telles que SURF, SIFT, le Réseau de Neurones Convolutionnel (CNN), la transformée de Gabor, les Eigenfaces et LBP.

3.5.1.1. Histogramme des gradients orientés (HOG)

L'Histogramme des Gradients Orientés (HOG) est une technique d'extraction de caractéristiques utilisée en vision par ordinateur pour décrire la texture d'une image. Il fonctionne en calculant un histogramme de la distribution des orientations des gradients

dans des régions locales d'une image. Le HOG est une technique puissante et robuste qui a été utilisée avec succès dans de nombreuses applications de vision par ordinateur, notamment la reconnaissance d'objets, la reconnaissance des visages, la classification des images, la segmentation d'images et la détection de défauts.

1) Principe du HOG

Le HOG est basé sur l'idée que la texture d'une image peut être décrite par la distribution des orientations des gradients dans l'image. Pour calculer le HOG, on commence par calculer les gradients de l'image. Les gradients sont des mesures de la direction et de l'intensité du changement de luminosité entre deux pixels adjacents, Comme indiqué dans la Figure 3.9.

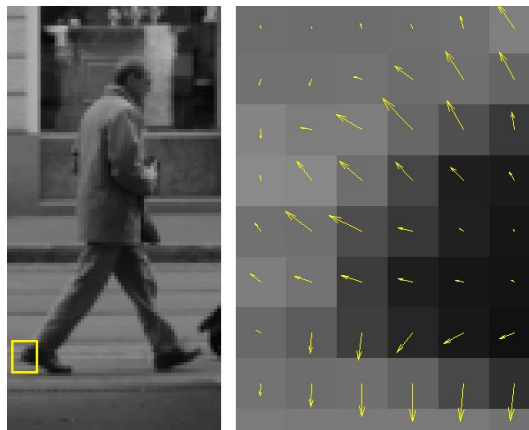


Figure3.9: Gradients calculés pour une section de l'image. [59]

Ensuite, on divise l'image en petites régions locales, appelées cellules. Pour chaque cellule, on calcule un histogramme de la distribution des orientations des gradients dans la cellule. L'histogramme est une liste des fréquences d'occurrence des différentes orientations de gradient [59], Comme indiqué dans la Figure 3.10.

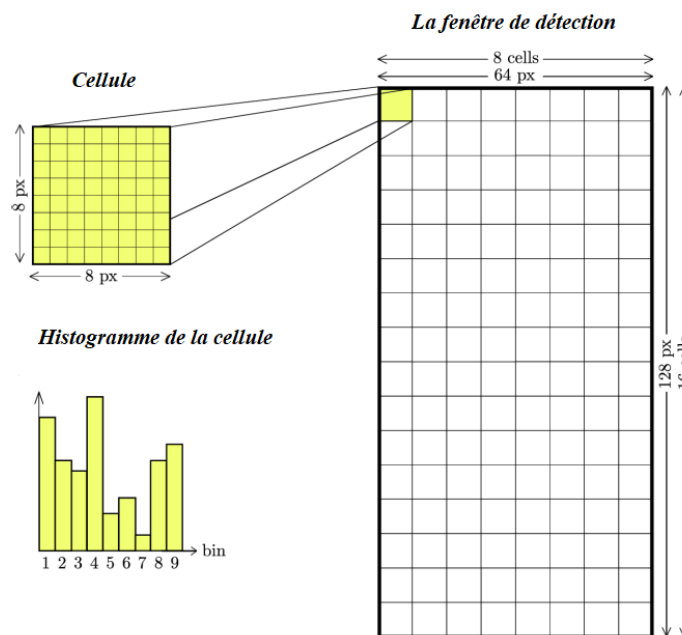


Figure3.10: Division de la fenêtre de détection en cellules. Histogramme généré pour une seule cellule.

2) Calcul du gradient

La première étape pour produire le descripteur HOG consiste à calculer les dérivés ponctuelles unidimensionnelles G_x et G_y en direction x et y en convoluant les masques de gradient M_x et M_y avec l'image brute I.

$$G_x = M_x * I \quad M_x = [-1 \ 0 \ 1] \quad (3.1)$$

$$G_y = M_y * I \quad M_y = [-1 \ 0 \ 1] \quad (3.2)$$

Sur la base des dérivées G_x et G_y , Il s'ensuit une estimation du module du gradient $|G(x,y)|$ et l'angle de direction $\phi(x,y)$ pour chaque pixel. Le module du gradient reflète l'intensité du gradient à un pixel spécifique [59].

$$|G(x,y)| = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \quad (3.3)$$

L'angle d'orientation du gradient peut être calculé directement à partir de :

$$\tan(\phi(x, y)) = \frac{G_y(x, y)}{G_x(x, y)} \quad (3.4)$$

3) Les avantages de la méthode HOG

L'HOG présente de nombreux avantages, notamment :

- **Robustesse aux bruits et aux variations d'éclairage** : Le HOG est relativement insensible aux bruits et aux variations d'éclairage, ce qui le rend utile pour l'analyse d'images dans des conditions difficiles.
- **Capacité à décrire une large gamme de textures** : Le HOG peut être utilisé pour décrire une large gamme de textures, y compris les textures régulières, irrégulières et complexes.
- **Fonctionnalités invariantes à l'échelle** : Le HOG est invariant à l'échelle, ce qui signifie que les caractéristiques HOG d'un objet restent les mêmes lorsqu'il est vu à des échelles différentes.

4) Applications de la méthode HOG

Le HOG a été utilisé avec succès dans de nombreuses applications de vision par ordinateur, notamment :

- *Reconnaissance d'objets* : identifier des objets dans des images, tels que des personnes, des voitures, des animaux ou des produits manufacturés.
- *Reconnaissance des visages*: identifier des personnes en fonction de leur texture faciale.
- *Classification des images* : classer des images en fonction de leur texture.
- *Segmentation d'images* : segmenter des images en fonction de leur texture.
- *Détection de défauts* : détecter des défauts dans des images, tels que des défauts de surface sur des produits manufacturés.

Le HOG est une technique puissante et polyvalente qui peut être utilisée dans de nombreuses applications de vision par ordinateur. Elle est robuste aux bruits et aux variations d'éclairage, et peut être utilisée pour décrire une large gamme de textures [59], c'est exactement ce dont nous avons besoin pour un système de reconnaissance faciale automatique.

3.5.1.2. Modèles Binaires Locaux (LBP)

Le LBP encode les informations de texture d'une image en niveaux de gris en comparant la différence d'intensité de chaque pixel avec ses pixels voisins. Initialement, l'image est convertie en niveaux de gris. Ensuite, l'image est divisée en cellules rectangulaires $[k \times k]$. Chaque pixel i dans la cellule est comparé (en termes d'intensité) aux pixels voisins dans un cercle centré sur le pixel i et de rayon r . Dans [135], les pixels voisins sont au nombre de 8 et le rayon est de 1. En définissant la valeur du pixel central comme seuil (variant de 0 à 255), les pixels adjacents obtiennent des valeurs binaires : ceux ayant des valeurs égales ou supérieures reçoivent une valeur de 1, et ceux ayant des valeurs inférieures reçoivent une valeur de 0. Ces valeurs binaires sont converties en décimal en les multipliant avec des puissances de 2 (en gardant la même direction) et en les additionnant. Le processus est répété avec chaque pixel appartenant à 9 cellules différentes de 3×3 , de sorte qu'il puisse prendre 2^8 valeurs différentes. L'histogramme à 256 bacs de la fréquence des valeurs prises par chaque pixel constitue le vecteur caractéristique de 256 dimensions.

Le fait que certains des motifs binaires apparaissent plus souvent que d'autres a conduit à une version avancée de cet algorithme, appelée la version de motif uniforme, qui est invariante par rotation. Un motif est dit uniforme lorsqu'il comporte au plus deux transitions 0!1 ou 1!0. L'histogramme, dans ce cas, comporte un bac pour chaque motif uniforme et un bac pour tous les motifs non uniformes. Le nombre de bacs est égal à $P(P - 1) + 3$, où P est le nombre de pixels voisins. Pour 8 pixels voisins, l'histogramme de dimension 256 est transformé en un histogramme de dimension 59, ce qui représente une réduction de 77% de la taille de la caractéristique. Pour une image de taille $[M \times N]$, la taille de la caractéristique est donnée par l'équation (3.5) :

$$\text{Taille de la Caractéristique LBP} = \lfloor M/k \rfloor \lfloor N/k \rfloor [P(P - 1) + 3] \quad (3.5)$$

La structure locale des niveaux de gris est résumée à l'aide de l'opérateur des motifs binaires locaux (LBP). Le LBP est utile car il est résistant aux effets d'éclairage et

efficace pour l'analyse de texture et la classification selon [136]. L'image faciale est divisée en plusieurs blocs et les caractéristiques sont extraites sur ces blocs par le LBP.

$$LBP_{p,r} \sum_{p=0}^{p-1} s(g_i - g_c) 2^i, s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{Sinon} \end{cases} \quad (3.6)$$

Où g_c est le niveau de gris du pixel central. g_i ($i=0,1,..7$) est le niveau de gris de chaque pixel environnant. Si g_i est inférieur à g_c , le résultat binaire du pixel est défini à 0, sinon à 1. Tous les résultats sont combinés pour obtenir une valeur de 8 bits. $LBP_{p,r}$ représente la caractéristique LBP des voisinages circulaires d'un pixel, où r est le rayon et p est le nombre de points du voisinage sur le cercle [136].

3.5.1.3. Cbfd (Descripteurs faciaux binaires compacts)

Jiwen et al. [10] introduisent une nouvelle méthode d'extraction de caractéristiques appelée Cbfd (Compact Binary Face Descriptor), qui vise à améliorer les performances des codes binaires grâce à une phase d'apprentissage. Cette approche intègre l'intelligence artificielle pour pallier les limitations et améliorer l'efficacité. La phase d'apprentissage et l'extraction des caractéristiques de l'image peuvent être résumées comme suit :

1) La Phase d'apprentissage

Cette méthode se concentre sur la robustesse des codes binaires par rapport aux variations locales de la texture de l'image. Les codes binaires compacts sont appris directement à partir des pixels bruts pour représenter les images. Il est important de noter que pour de meilleurs résultats de classification, les caractéristiques Cbfd doivent être construites à l'aide d'un ensemble d'échantillons d'images fournis dans le même contexte. Pendant l'apprentissage des caractéristiques Cbfd, les vecteurs d'entraînement sont générés en tenant compte de la relation entre chaque pixel et son voisinage. Plus précisément, une analyse est effectuée sur l'image à l'aide d'une fenêtre rectangulaire de taille 3 x 3. Cette fenêtre est centrée sur chaque pixel, ce qui permet d'extraire des informations pertinentes du contexte local entourant ce pixel.

Soit l'ensemble des vecteurs d'apprentissage $X = [x_1, x_2, \dots, x_n]$, appelés Vecteurs de Différence de Pixels (PDV). Ces PDV sont obtenus en mesurant la différence entre le pixel central et ses pixels voisins dans une fenêtre prédéfinie. La taille de chaque vecteur est $(2R+1) \times (2R+1) - 1$, à l'exclusion du PDV entre le pixel central et lui-même (PDV₀=0).

L'objectif de l'extraction de caractéristiques Cbfd est d'apprendre K fonctions $(w_k)_{k=1 \dots n}$ de hachage qui quantifient chaque vecteur $x_n, (n = 1, \dots, N)$ en un vecteur binaire $b_n = [b_{n1}, b_{n2}, \dots, b_{nk}]$.

Cette quantification est réalisée grâce à la formule (3.7) suivante :

$$b_{nk} = 0.5 \times (\text{sgn}(w_k^T \times x_n) + 1) \quad (3.7)$$

Où, $\text{sgn}(v) = 1$ if $v > \tau$ and -1 , désigne τ le seuil utilisé pour la conversion binaire des caractéristiques.

Afin de construire la matrice de projection W qui englobe toutes les fonctions de hachage w_k , nous l'initialisons avec les K premiers vecteurs propres de la matrice de covariance ($C = XX^T$). Par la suite, une procédure d'optimisation est menée pour minimiser la fonction objective, $J(w_k)$ définie par l'équation (3.8) comme suit :

$$\min J(w_k) = J_1(w_k) + \lambda_1 \times J_2(w_k) + \lambda_2 \times J_3(w_k) \quad (3.8)$$

$$V_b = 0.5 \times (\text{sgn}(w^T \times X) + 1) \quad (3.9)$$

Les paramètres λ_1 et λ_2 sont prédéfinis et utilisés pour équilibrer les effets de différents termes. Les termes J_1, J_2 , et J_3 sont sélectionnés pour garantir que :

- (3.7) La variance des codes binaires appris est maximisée.
- (3.8) La perte de quantification entre la caractéristique originale et les codes binaires encodés est minimisée.

- (3.9) Les bacs de caractéristiques dans les codes binaires appris sont répartis le plus uniformément possible.

2) Apprentissage du codebook

Le but du codebook est de réduire le nombre de vecteurs binaires associés à chaque image. Les vecteurs d'entraînement (PDVs) sont projetés sur la matrice w , puis l'algorithme de clustering k-means est appliqué pour obtenir les centroïdes des vecteurs binaires résultants. Ces centroïdes forment le codebook, qui représente les classes.

3) Phase d'extraction des caractéristiques de l'image

Le processus d'extraction des caractéristiques repose sur la matrice de projection w (caractéristique Cbfd) et le codebook obtenu lors de la phase d'apprentissage. Après avoir obtenu tous les vecteurs PDV $X = [x_1, x_2, \dots, x_n]$ pour l'image, leurs homologues binaires sont déterminés en les projetant sur la matrice.

Chaque vecteur binaire est ensuite remplacé par la coordonnée vectorielle la plus proche dans le codebook (bin). Par la suite, un histogramme est construit en utilisant les différentes coordonnées, représentant l'ensemble des caractéristiques de l'image. Pour extraire des vecteurs de caractéristiques discriminants, l'image brute est segmentée en plusieurs régions, traitant chaque région comme une image individuelle avec ses propres caractéristiques Cbfd (w) et son codebook. Pour chaque région, un histogramme (H_s) est créé. Enfin, la concaténation de tous les histogrammes donne un vecteur complet (v) donnée par l'équation (3.10) qui représente l'image entière :

$$v = [H_1, H_2, \dots, H_M] \quad (3.10)$$

3.5.1.4. *Speeded up robust features (SURF)*

SURF est un détecteur et descripteur de caractéristiques locales. La version standard de SURF est plusieurs fois plus rapide et donne des résultats robustes contre différentes transformations d'image par rapport à SIFT. Les détecteurs SURF découvrent les points d'intérêt dans une image. Pour détecter les points d'intérêt, SURF utilise une approximation entière du déterminant du détecteur de blobs Hessian [137], qui peut être

calculée avec 3 opérations entières à l'aide d'une image intégrale pré-calculée. Son descripteur de caractéristiques est basé sur la somme de la réponse de l'ondelette de Haar autour du point d'intérêt. Cela peut être calculé à l'aide de l'image intégrale. SURF est invariant à l'échelle et aux rotations dans le plan des caractéristiques. Les vecteurs de caractéristiques par SURF sont formés par des motifs locaux autour des points clés qui sont détectés à l'aide d'un filtre agrandi.

Les principales étapes effectuées dans SURF sont le détecteur de points d'intérêt (voir Figure3.11) et le descripteur de points d'intérêt [137].

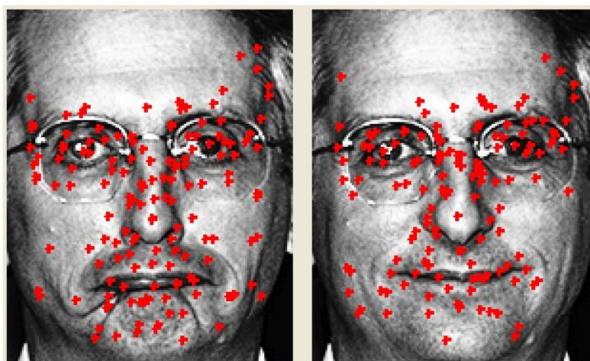


Figure3.11.Points d'intérêt détectés avec SURF. [137]

3.5.1.5. Scale invariant feature transform (SIFT)

Scale invariant feature transform (SIFT) est une technique d'extraction de caractéristiques conventionnelle et largement utilisée. Elle utilise le concept de pyramide d'images et la Différence De Gaussiennes (DOG). L'image est convoluée par un filtre gaussien à différentes échelles. SIFT est invariant à l'échelle, à la rotation et offre des performances optimales sous les changements d'éclairage, les changements de pose [138]. Avec la technique SIFT l'image de test est appariée en comparant de manière autonome chaque caractéristique de l'image de test à la base de données actuelle. Le meilleur appariement entre les caractéristiques extraites est basé sur la distance euclidienne de leurs vecteurs de caractéristiques [138]. SIFT comprend quatre étapes majeures de calcul. Les quatre étapes comprennent la détection des extrema dans l'espace des échelles, la localisation des points clés, l'attribution d'orientation et le descripteur des points clés. Bien que SIFT soit simple et performant pour l'extraction de

caractéristiques, elle accuse un retard en termes de vitesse de calcul et de complexité computationnelle [138].

3.5.1.6. Réseau de neurones convolutionnel (CNN)

Le réseau de neurones convolutionnel (CNN) est une méthode d'apprentissage profond largement utilisé dans la reconnaissance d'images et l'analyse de la parole. Les CNN sont utiles dans l'étude de la reconnaissance de la variabilité des formes 2D. Il combine des champs de caractéristiques locaux avec des poids partagés. Il utilise également le sous-échantillonnage spatial pour prendre en compte le niveau de décalage, la déformation et l'invariance à l'échelle [139]. Le réseau de convolution contient des ensembles de couches composées de nombreux plans de neurones. Chaque unité dans le plan est associée à la couche précédente en utilisant un voisinage local. Dans les CNN, chaque couche possède de nombreuses cartes de caractéristiques. La sortie d'une couche est l'entrée de la couche suivante [139]. L'unité se compose d'un détecteur de caractéristiques locales et sa caractéristique d'activation est obtenue dans la phase d'apprentissage. Les sorties de l'ensemble des unités se combinent pour former une carte de caractéristiques. La même opération est appliquée sur chaque partie de l'image ou les anciennes cartes de caractéristiques pour toutes les unités dans une carte de caractéristiques afin d'extraire toutes les caractéristiques de la même image [139]. Après avoir balayé l'image d'entrée avec une seule unité avec des poids pour former un champ réceptif local et stocker les sorties de la première unité dans ses emplacements associés dans la carte de caractéristiques séquencée. Les cartes de caractéristiques convolutionnelles ajoutent un niveau d'invariance pour les translations et la distorsion. Elles incluent des caractéristiques ayant une résolution spatiale décroissante à mesure que la complexité augmente ainsi que la détection globale des unités dans les couches successives.

3.5.1.7. Filtre de gabor

Les représentations de fréquence et d'orientation de l'ondelette de Gabor sont basées sur le système visuel humain. L'ondelette de Gabor est adaptée à la représentation des informations de texture. L'ondelette de Gabor est utilisée pour extraire les caractéristiques de l'image qui sont alignées à des orientations particulières. L'ondelette

de Gabor donne une analyse des images en considérant différentes échelles ou résolutions et orientations. Elle prend en compte à la fois les rotations et les variations de fréquence simultanément [140]. L'ondelette de Gabor est une sinusoïde modulée utilisant une enveloppe gaussienne comme décrit dans [141]. Le noyau du filtre de Gabor 2D est donné par :

$$f(x, y, \theta_k, \lambda) = \exp \left[-\frac{1}{2} \left\{ \frac{R_1^2}{\sigma_x^2} + \frac{R_2^2}{\sigma_y^2} \right\} \right] \exp \left\{ i \frac{2\pi R_i}{\lambda} \right\} \quad (3.11)$$

$$R_1 = x \cos \theta_k + y \sin \theta_k \quad , \quad R_2 = -x \sin \theta_k + y \cos \theta_k$$

Où σ_x et σ_y sont définis comme les écarts-types de l'enveloppe gaussienne dans les dimensions x et y , λ est la longueur d'onde et θ est l'orientation du plan sinusoïdal respectivement. Le spectre de l'enveloppe gaussienne basé sur la longueur d'onde λ est défini comme :

$$\theta_k = \frac{\pi(k-1)}{n}, k = 1, 2, \dots, n. \quad (3.12)$$

où n est le nombre d'orientations à prendre en compte. Le processus d'extraction de caractéristiques en utilisant la transformée d'ondelettes de Gabor consiste d'abord à convertir l'image couleur en image en niveaux de gris. Ensuite, construire le filtre de Gabor en utilisant des fréquences et différents angles de phase. Enfin, la moyenne et l'écart-type sont calculés et stockés dans un vecteur de caractéristiques.

3.5.1.8. Visages propres (Eigenfaces)

Les visages propres décomposent les images de visages en un petit ensemble d'images de caractéristiques distinctives. La méthode des visages propres est basée sur l'Analyse en Composantes Principales (ACP). La reconnaissance est effectuée en projetant une autre image dans le sous-espace engendré par les visages propres, puis en classifiant le visage en comparant sa position dans l'espace des visages avec les emplacements des individus connus [141]. Les visages propres sont utilisés pour trouver l'espace de dimension inférieure. Ils réduisent la dimensionnalité de l'espace de données original. Les visages propres sont les vecteurs propres qui représentent chacune des dimensions de l'espace des visages et ils peuvent être considérés comme de multiples caractéristiques de visages. Les vecteurs propres sont dérivés de la matrice de

covariance de l'ensemble des images de visages, où une image de $N \times N$ pixels est considérée comme un point (ou vecteur) dans un espace dimensionnel N^2 . Cette approche offre simplicité, rapidité et capacité d'apprentissage.

3.5.2. Analyse comparative

Cette section présente l'analyse comparative des techniques d'extraction de caractéristiques. Le tableau 1 montre les différentes techniques d'extraction de caractéristiques faciales ainsi que leurs avantages et limitations.

Tableau 3.2: Avantages et limitations des techniques d'extraction de caractéristiques.

Technique	Avantages	Limitations
SURF [137]	+ Détecteur et descripteur invariants à l'échelle et à la rotation. + Reconnaissance des images dans différentes conditions telles que faible luminosité, images floues.	- Mauvaise approximation des orientations de points clés. - Ne prend en compte que les images en niveaux de gris.
SIFT [138]	+ Invariant à l'échelle et à la rotation. + Reconnaît les images sous différentes conditions d'éclairage. + Reconnaît les visages avec variations d'échelle, de pose et d'expressions.	- Haute dimensionnalité lors de l'appariement. - Plus lent que SURF (SURF offre des performances similaires tout en étant plus rapide).
CNN [139]	+ Utilisation de poids partagés dans les couches convolutionnelles, ce qui signifie que le même filtre (banque de poids) est utilisé pour chaque pixel de la couche. + Réduit l'empreinte mémoire et améliore les performances. + Gère les images faciales avec différentes poses, expressions faciales et illuminations. + Invariant aux distorsions.	- Charge de calcul élevée.
FILTRE GABOR [140]	+ Extrait des composants importants des visages tels que la bouche, les yeux, le nez, etc. + Invariant à l'illumination, la rotation, l'échelle, la transformation. + Multi-orientation et robuste.	- La dimension du vecteur de caractéristiques est énorme.
EIGEN FACES [141]	+ Utilise spécifiquement les données d'intensité brute pour l'apprentissage et la reconnaissance sans traitement significatif de bas ou moyen niveau. + Simple et efficace.	- Sensible à l'échelle. Par conséquent, un prétraitement de bas niveau reste nécessaire pour la normalisation de l'échelle.
LBP [136]	+ Invariant à la rotation. + Facile à mettre en œuvre.	- Charge computationnelle importante.
HOG[59]	+Robustesse aux bruits et aux variations d'éclairage .+ Simple et efficace.+ Capacité à décrire une large gamme de textures. + Fonctionnalités invariantes à l'échelle.	- Moins performant sur les textures fines.
CBFD[10]	+Rapidité de calcul. +Robustesse au bruit.+Efficacité en termes de stockage	-Sensibilité aux variations d'éclairage et de pose

3.6. Conclusion

Dans ce chapitre, nous avons présenté diverses méthodes d'extraction des caractéristiques et de détection de visage. Nous avons détaillé en particulier les méthodes que nous avons utilisées, notamment : les algorithmes de détection de points de repère (landmark detection), ainsi que les techniques d'extraction de caractéristiques HOG, LBP et CBF, permettant une meilleure modélisation des visages. Dans le chapitre suivant, nous présenterons les résultats obtenus sur les trois bases de données (ORL, LFW et IJB-C) afin de démontrer, d'une part, l'efficacité de nos méthodes d'extraction de caractéristiques et, d'autre part, l'impact positif de la fusion d'informations sur les performances des systèmes ainsi développés.

Un problème sans solution est un
Problème mal posé.

ALBERT EINSTEIN - *Physicien*

CONTRIBUTIONS

4.1. Introduction

Les méthodes de combinaison sont des techniques utilisées pour fusionner les sorties de plusieurs modèles, classifieurs ou sources d'information, dans le but d'améliorer les performances globales, la robustesse ou de fournir des prédictions plus fiables. Ces méthodes trouvent des applications dans divers domaines tels que l'apprentissage automatique, la reconnaissance de formes et la fusion de données. Afin d'améliorer la performance, la robustesse et la fiabilité des systèmes de reconnaissance faciale, la présente étude implémente deux méthodes de combinaison basées sur la fusion au niveau des scores et au niveau des caractéristiques. Ces méthodes sont employées pour déterminer laquelle améliore significativement les performances globales du système.

Les contributions dans ce manuscrit sont résumées comme suit :

☞ Dans notre proposition initiale, contrairement aux techniques de combinaison traditionnelles telles que celles basées sur le niveau des scores, des caractéristiques et des images, nous introduisons une règle de fusion inventive basée sur un classifieur MLP (Multi-Layer Perceptron). Opérant au niveau des scores, cette méthodologie consiste à concaténer les scores dérivés des modèles individuels, puis à entraîner le classifieur MLP pour calculer le score adéquat.

☞ La reconnaissance faciale ne repose pas uniformément uniquement sur la structure faciale complète ; au contraire, elle peut dépendre de composants faciaux spécifiques dans certaines conditions. Dans cette optique, Nous explorons un nouveau modèle basé sur l'apprentissage profond et l'analyse de régions spécifiques du visage. Le concept fondamental de notre proposition repose sur le lien entre la reconnaissance d'une région faciale donnée et la reconnaissance de la région faciale précédente. Ce principe s'inspire des paradigmes de reconnaissance de données séquentielles, tels que ceux utilisés pour la génération de texte [142, 143]. En d'autres termes, la couche entièrement connectée du CNN qui réalise une reconnaissance optimale pour la première région faciale est combinée avec les caractéristiques de la région faciale suivante, et cette séquence continue. Finalement, les couches entièrement connectées des CNN composites sont fusionnées et introduites dans un classifieur DNN (Deep Neural Network) pour évaluer la performance globale du système [144].

4.2. Reconnaissance faciale basée sur l'apprentissage automatique

L'approche de reconnaissance faciale proposée [145], basée sur l'apprentissage automatique, illustrée dans la Figure 4.1 comprend les étapes suivantes : 1) Prétraitement de l'image : la première étape comprend la détection des visages et le recadrage de l'image pour séparer le visage dans l'image d'entrée. 2) Extraction des caractéristiques : dans la deuxième étape, une variété de caractéristiques sont extraites des images faciales prétraitées. Trois schémas distincts, à savoir LBP [8], Cbfd [10] et HOG [9], sont utilisés pour extraire différents vecteurs de caractéristiques. 3) Modèle de combinaison des scores: cette étape utilise un modèle de combinaison qui intègre les sorties de trois classifieurs. Les scores obtenus à partir des trois classifieurs sont combinés et fournis en entrée du classificateur MLP pour déterminer la combinaison optimale pour la reconnaissance faciale.

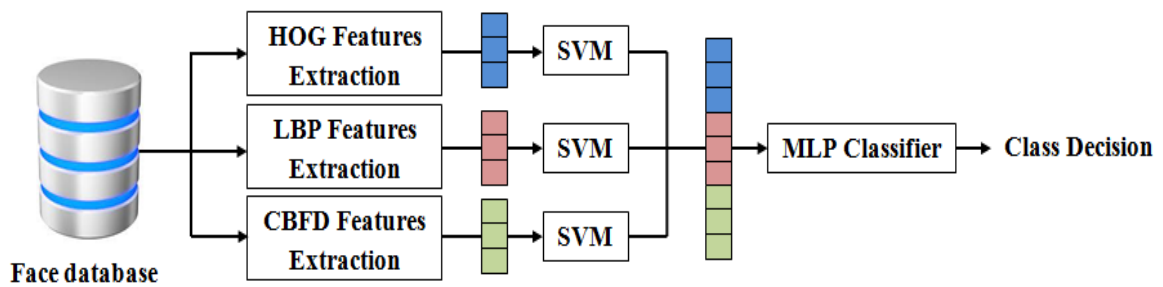


Figure 4.1. Combinaison de Classifieurs Basée sur la Fusion des Scores.

4.2.1. *Étape de prétraitement*

L'étape de prétraitement joue un rôle crucial dans l'extraction d'informations précieuses à partir d'images faciales numériques, menant à une amélioration significative de la précision de nos schémas de reconnaissance faciale. Dans nos schémas proposés, cette étape de prétraitement comprend deux phases principales : détection du visage, recadrage et redimensionnement.

☞ **Détection du visage** : Dans les applications liées au visage, la détection du visage joue un rôle fondamental. Elle implique l'utilisation d'algorithmes conçus pour détecter et localiser précisément des points de repère sur un visage, appelés "landmarks". L'objectif principal de cette étape est d'identifier avec précision la région faciale afin d'extraire les caractéristiques exclusivement des zones pertinentes de l'image d'entrée.

Pour la détection de la région faciale dans chaque image, nous avons utilisé un algorithme de détection des points de repère-68 connu par Landmarks-68 qui identifie automatiquement les points de repère faciaux [146].

Le principe de cet algorithme est comme suit :

L'algorithme Landmark-68 est conçu pour identifier 68 points spécifiques correspondant à des caractéristiques clés du visage. Ces points sont soigneusement sélectionnés pour couvrir les zones anatomiques importantes, permettant ainsi une représentation détaillée et précise de la géométrie faciale (voir Figure 4.2).

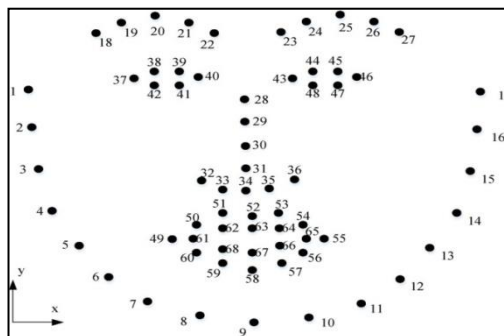


Figure 4.2:Forme du visage avec 68 points de repère.

La localisation de ces points de repère s'effectue généralement à l'aide de modèles statistiques, qui offrent une grande précision dans l'identification des points. Le modèle utilisé pour cette tâche est le Modèle de Forme Active (ASM). Ce dernier emploie une approche itérative visant à minimiser la différence entre la forme du modèle et les caractéristiques détectées sur l'image, en ajustant les paramètres du modèle pour correspondre aux contours des traits faciaux.

La répartition des principaux groupes de points de repère est la suivante :

☞ Coins des yeux

- Points identifiés : 36 points pour les deux yeux (18 par œil).
- Utilité : Ces points permettent de déterminer l'orientation et la forme des yeux, essentiels pour l'expression faciale et l'identification.

☞ Sourcils

- Points identifiés : 8 points (4 par sourcil).
- Utilité : Ils sont cruciaux pour détecter les expressions et les émotions, telles que la surprise ou la colère.

☞ Nez

- Points identifiés : 6 points.

- Utilité : Ces points aident à définir la largeur, la longueur et la position du nez, des éléments essentiels à la reconnaissance faciale.

☞ Coins de la bouche

- Points identifiés : 12 points (6 par coin).
- Utilité : Ces points jouent un rôle clé dans l'expression des émotions, comme le sourire ou la tristesse.

☞ Ligne de la mâchoire

- Points identifiés : 17 points le long du contour de la mâchoire.
- Utilité : Ces points contribuent à définir la forme globale du visage et sont utiles dans des applications comme la modélisation 3D du visage.

☞ **Étape de recadrage et de redimensionnement** : Dans notre étude, nous avons utilisé trois points de repère spécifiques situés le long de la ligne de la mâchoire (points 1, 9 et 17) ainsi que deux points situés sur le sourcil (points 20 et 25) pour déterminer la largeur et la hauteur d'un rectangle. Pour garantir une localisation précise de la région faciale, nous avons défini des rectangles légèrement plus grands qui se chevauchent, couvrant ainsi une zone plus étendue du visage. Une fois la région faciale identifiée, l'image d'entrée est recadrée pour inclure uniquement la portion délimitée par les rectangles définis par ces points de repère.

4.2.2. *Extraction des caractéristiques*

L'extraction des caractéristiques est une étape cruciale dans les applications de reconnaissance de formes, car les performances de classification dépendent fortement de la distinctivité et de la variabilité des caractéristiques extraites. Dans nos méthodes proposées, nous avons utilisé trois techniques d'extraction de caractéristiques : LBP (Local Binary Pattern), Cbfd (Compact Binary Facial Descriptors), et HOG (Histogram of Oriented Gradients), détaillées dans le chapitre 3.

4.2.3. *Modèle de combinaison des scores*

Le modèle de combinaison proposé repose sur la combinaison des scores de sortie de trois classifieurs (SVM) pour prendre une décision finale, exploitant ainsi leurs connaissances et capacités collectives. Ce modèle de combinaison peut être résumé de la manière suivante :

☞ Tout d'abord, nous disposons d'un ensemble d'échantillons d'entraînement constitué de paires x_i, y_i où i varie de 1 à n . Chaque échantillon x_i , obtenu à partir d'une méthode d'extraction de caractéristique, est représenté par un vecteur de caractéristiques de dimensions d , tandis que y_i correspond à l'étiquette de catégorie associée, prenant des valeurs dans l'ensemble $\{1, n\}$.

☞ Ensuite, chaque classifieur traite un ensemble de données d'entrée et génère des prédictions pour chaque échantillon x_i , produisant un vecteur de scores de taille n , représentant les probabilités attribuées à chaque classe.

$$S_j^1, S_j^2, \dots, S_j^n \text{ avec } S_j^i \in [0,1] \quad (4.1)$$

L'étape suivante consiste à utiliser un réseau de perceptron multicouche (MLP). Ce MLP comprend une couche d'entrée qui reçoit la combinaison des scores S_j^n avec $j \in \{1,2,3\}$, une seule couche cachée, et utilise la fonction d'activation sigmoïde. La décision pour la classe i est prise par le modèle de combinaison en se basant sur la couche de sortie des classifieurs SVM, selon la formule suivante :

$$C = \sum_{i=1}^n w_{ij} \times S_j^i + b_j \quad (4.2)$$

Dont, w_{ij} représente les poids et b_j est la valeur du biais du classifieur j .

☞ Enfin, pour mettre à jour les poids w_{ij} et la valeur de biais b_j , il est nécessaire de calculer l'erreur de prédiction du modèle. Cela peut être fait en utilisant la formule (4.3) pour calculer l'erreur du $n^{\text{ème}}$ nœud dans la couche de sortie. L'erreur de prédiction dans la couche de sortie est donnée par :

$$\varepsilon = v_i(1 - v_i) - (d_i - v_i) \quad (4.3)$$

Où, d_i représente la valeur de sortie souhaitée du modèle et v_i la valeur de sortie prédite.

4.3. Approche de reconnaissance faciale basée deep learning

Dans le domaine de la reconnaissance faciale, les humains se concentrent instinctivement sur des traits spécifiques pour identifier et différencier les visages. Les yeux, le nez, la bouche et les sourcils jouent un rôle fondamental, non seulement en raison de leur unicité, mais également grâce aux relations géométriques qui les relient. Par exemple, la distance entre les centres des yeux peut servir de ligne de base, à partir de laquelle les distances verticales des autres traits, comme le nez ou la bouche, apportent des informations discriminantes essentielles. Ces observations inspirent une approche où chaque région faciale est étudiée de manière ciblée et où les relations spatiales sont exploitées pour améliorer la précision.

4.3.1. Concept de reconnaissance séquentielle basée sur les régions

La méthode proposée s'inspire des architectures de réseaux neuronaux telles que les RNN et LSTM (Long Short-Term Memory) [142,143], utilisées pour la génération de mots, où la reconnaissance d'un mot dépend du contexte fourni par les mots précédents. De manière analogue, cette approche traite une image faciale comme une séquence de données. Plutôt que d'analyser le visage dans son intégralité en une seule étape, l'image est segmentée en régions spécifiques (yeux, nez, bouche, etc.), qui sont étudiées individuellement. Cette approche repose sur des techniques de traitement séquentiel couramment appliquées à des tâches comme la reconnaissance de texte ou de vidéo, où la compréhension découle de l'analyse progressive des séquences. Chaque région faciale est ainsi considérée comme une entité distincte, permettant une exploration détaillée avant que les informations ne soient intégrées pour produire une décision finale cohérente et précise, tout en exploitant les capacités des Réseaux de Neurones Convolutionnels (CNN).

4.3.2. Exploitation des relations géométriques

Une caractéristique distinctive de cette méthode est l'intégration explicite des relations géométriques entre les différentes régions faciales. Après avoir extrait les

caractéristiques de chaque région faciale à l'aide des CNN, les distances et positions relatives entre ces régions sont calculées et incorporées dans le modèle. Par exemple, la distance verticale du nez ou de la bouche à la ligne de base des yeux peut être utilisée pour renforcer la distinction entre les visages. Ce traitement contextuel des relations spatiales garantit que l'analyse ne repose pas uniquement sur des traits isolés, mais également sur leur agencement global.

4.3.3. Mécanisme de coopération des classifieurs

La Figure 4.3 illustre le fonctionnement de l'approche coopérative basée sur des CNN agissant comme des classifieurs locaux. Chaque CNN est spécifiquement entraîné pour analyser une région faciale donnée, comme les yeux, le nez ou la bouche, et produit une prédiction individuelle pour cette région. Ces prédictions reflètent les caractéristiques uniques capturées par chaque classifieur local, fournissant des informations complémentaires sur le visage analysé.

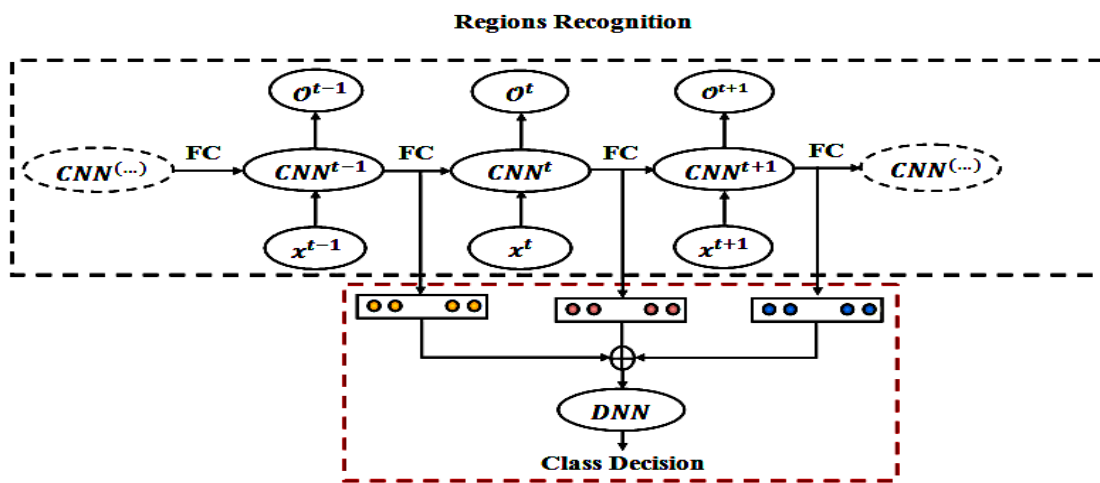


Figure 4.3. Architecture de reconnaissance faciale basée sur le Deep Learning

Considérons une image faciale comme une séquence composée de plusieurs régions faciales, notée x . À chaque étape temporelle discrète t , une région spécifique de l'image, notée x^t , est sélectionnée et utilisée comme entrée pour le CNN correspondant. Lors de chaque étape temporelle t , un état caché, noté $FC^{(t)}$, est calculé. Cet état joue un rôle clé en tant que « mémoire » du réseau, stockant les informations extraites

jusque-là. Il est déterminé en combinant l'entrée actuelle x^t avec l'état caché de l'étape précédente $FC^{(t-1)}$. Cette combinaison, réalisée par concaténation, est exprimée mathématiquement dans l'Équation (4.4).

$$CNN^{(t)} = x^t \oplus FC^{(t-1)} \quad (4.4)$$

Le résultat de ce processus, représenté par $O^{(t)}$, encapsule les caractéristiques spécifiques que le CNN extrait de la région faciale analysée x^t . Cela permet de conserver une représentation riche et structurée des informations pour chaque région faciale.

Pour arriver à une décision globale, les sorties des différents CNN, qui agissent comme des classifieurs locaux, sont combinées dans un cadre coopératif. Cette fusion des prédictions peut être réalisée en utilisant diverses approches, notamment :

1. La moyenne pondérée, où chaque classifieur contribue proportionnellement à un poids reflétant sa fiabilité sur une région spécifique.
2. Le modèle de fusion dense, qui repose sur une couche dense (FullyConnected Layer) pour apprendre une intégration optimisée des prédictions locales.
3. Le mécanisme séquentiel (LSTM ou transformateur), qui exploite les relations contextuelles entre les différentes régions faciales en traitant les prédictions comme des étapes d'une séquence.

Dans notre travail, nous avons opté pour le modèle de fusion dense. Concrètement, une couche dense, aussi appelée réseau entièrement connecté, a été utilisée pour combiner les sorties des classifieurs locaux. Cette couche dense agit comme un intégrateur, apprenant à analyser et à fusionner les informations extraites par les différents CNN dédiés à chaque région faciale. L'objectif est de produire une prédiction finale qui soit cohérente et globale, en tenant compte des caractéristiques uniques de chaque région tout en exploitant les corrélations entre elles.

Enfin, ce vecteur combiné est fourni en entrée à un classifieur basé sur un Réseau de Neurones Profonds (DNN). Ce dernier exploite la richesse des caractéristiques

encapsulées dans le vecteur pour effectuer une reconnaissance faciale précise et performante, garantissant une classification robuste et efficace.

Ce choix se justifie par la capacité du modèle de fusion dense à capturer les interactions complexes entre les régions faciales et à s'adapter à diverses variations telles que les expressions faciales ou les angles de vue.

Notre approche présente plusieurs avantages clés. Elle permet une analyse approfondie des traits faciaux locaux tout en exploitant les relations globales, offrant ainsi une précision accrue. En outre, le traitement séquentiel des régions faciales permet de gérer les variations dues aux expressions, aux orientations ou aux conditions d'éclairage. Les applications potentielles incluent les systèmes de sécurité biométrique, le diagnostic médical basé sur les anomalies faciales, et des systèmes interactifs de reconnaissance dans des domaines comme la réalité augmentée ou la personnalisation marketing.

4.4. Expérimentations et résultats

Pour valider l'efficacité des méthodes proposées, une série d'expériences a été réalisée. Chaque expérience a porté sur l'évaluation et la comparaison des performances des méthodes en utilisant trois bases de données de référence bien établis : LFW [47], ORL [63] et IJB-C [108]. Ces bases ont été sélectionnées en raison de leur diversité et de leur représentativité dans le domaine de la reconnaissance faciale, offrant un éventail varié d'images faciales. L'objectif principal des expériences était de mesurer les performances des méthodes sur ces bases de données, fournissant ainsi des preuves empiriques convaincantes de leur efficacité dans des contextes pratiques.

4.4.1. Description du protocole du test

Dans nos expérimentations, les images faciales du jeu de données ORL ont été séparées en deux ensembles distincts. L'ensemble d'entraînement, destiné à alimenter les systèmes de reconnaissance faciale, est composé de 240 images, soit 6 images par individu. Les 160 images restantes sont utilisées pour les tests. Il est à noter que la base de données ORL comprend 40 individus, chacun étant représenté par 10 images uniques.

Concernant le jeu de données LFW, 3 300 images faciales (soit 80 % du jeu de données) ont été utilisées pour l'entraînement, tandis que les 769 images restantes ont été affectées aux tests. Il convient de préciser que la base de données LFW contient 4 096 individus, chacun étant représenté par une image unique.

Pour la base IJB-C, comme indiqué dans le chapitre 2, contient un total de 31 334 images. Elle est accompagnée de 4 protocoles permettant d'évaluer la détection faciale, la vérification, la reconnaissance et le regroupement (clustering) à différentes échelles, à savoir :

- **Le protocole de vérification mixte de base 1:1** évalue la capacité d'un algorithme à vérifier l'identité des personnes dans des scénarios de vérification.
- **Le protocole de vérification des covariables 1:1** utilise des templates composés d'une seule image pour analyser en détail les performances de l'algorithme en fonction de différentes covariables.
- **Le protocole de reconnaissance mixte de base 1:N** teste les capacités de l'algorithme dans des scénarios d'identification.
- **Le protocole de clustering**, comprenant 4 sous-protocoles couvrant de 32 à 3 531 personnes, évalue la capacité de l'algorithme à regrouper des visages en clusters à différentes échelles.

Dans notre travail, nous avons spécifiquement adopté le protocole de reconnaissance mixte 1:N, conçu pour évaluer la capacité des algorithmes dans des scénarios d'identification. Ce protocole utilise deux galeries distinctes, Galerie 1 (G1) et Galerie 2 (G2). Chaque galerie contient un modèle par personne, généré en sélectionnant aléatoirement la moitié des images d'une personne. Les images restantes sont attribuées à la base de tests. G1 comprend 1772 personnes accompagnées de 5 588 images, tandis que G2 se compose de 1 759 personnes et de 6 011 images. Il est important de noter que ces galeries sont entièrement distinctes l'une de l'autre, facilitant ainsi les scénarios d'identification à base ouvert.

4.4.2. Métriques d'évaluation

Afin d'évaluer l'efficacité de notre nouveau système d'identification faciale, nous avons réalisé des évaluations approfondies sur une série de jeux de données, incluant ORL,

LFW et IJB-C. Ces évaluations ont été effectuées en utilisant la métrique d'évaluation de la précision (Accuracy). La précision est une mesure fondamentale et largement utilisée dans les systèmes de classification, en particulier en reconnaissance faciale. Elle détermine la précision globale du modèle en quantifiant le ratio des prédictions correctes par rapport au total des prédictions effectuées. Mathématiquement, l'exactitude est définie comme suit :

$$\text{Accuracy} = \frac{\text{Number of Corrected Prediction}}{\text{Total of Number of Prediction}} \times 100\% \quad (4.5)$$

Dans le contexte des systèmes de reconnaissance faciale, une prédiction est considérée comme exacte si le système parvient à identifier ou vérifier l'individu sur l'image. À l'inverse, une prédiction est considérée comme incorrecte lorsque le système échoue à reconnaître l'individu ou l'identifie à tort comme une autre personne. La précision est une métrique facilement compréhensible qui offre une évaluation globale de l'efficacité du système.

4.4.3. Évaluation des performances

Les résultats expérimentaux présentés dans cette section sont organisés en deux parties principales. La première partie porte sur les résultats de l'approche proposée utilisant l'apprentissage automatique, tandis que la deuxième partie se concentre sur les résultats de l'approche basée sur l'apprentissage profond. Une comparaison entre les systèmes de pointe est ensuite réalisée afin d'identifier celui offrant les meilleures performances.

4.4.3.1. Évaluation de l'approche utilisant l'apprentissage automatique

Une étape de configuration des paramètres est effectuée pour les algorithmes d'extraction de caractéristiques utilisés dans l'approche d'apprentissage automatique, notamment LBP, Cbfd et HOG, afin d'optimiser leurs performances avant l'évaluation.

1. Configuration des paramètres du LBP

Afin d'optimiser les performances des descripteurs de caractéristiques basés sur LBP (Local Binary Patterns), deux paramètres essentiels ont été étudiés : le rayon r

définissant le motif autour du pixel central, et le nombre de points p le long de ce rayon. Une série d'expériences a été menée pour identifier les valeurs optimales de ces paramètres. L'algorithme LBP a été évalué sur diverses combinaisons de p et r , telles que (12,2), (12, 2), (12,2), (12,4), (12, 4), (12,4), (12,6), (12, 6), (12,6), (12,8), (12, 8), (12,8), (16,2), (16, 2), (16,2), (16,4), (16, 4), (16,4), (16,6), (16, 6), (16,6), et (16,8), (16, 8), (16,8).

Le Tableau 4.1 présente les performances du système proposé en fonction des meilleurs paramètres du LBP. Sur le jeu de données ORL, le descripteur LBP atteint une précision de **85,1 %** avec 16 échantillons et un rayon de 4, démontrant ainsi son efficacité exceptionnelle pour la reconnaissance faciale. Pour le jeu de données LFW, bien que la précision soit légèrement inférieure à **76,8 %**, le descripteur reste performant pour cette tâche. Enfin, sur le jeu de données IJB-C, une légère baisse de précision est observée, atteignant **73,4 %** avec une configuration de 12 échantillons et un rayon de 2. Cette diminution peut être attribuée à la qualité moindre des images incluses dans le protocole de validation de l'IJB-C. Ces résultats confirment la robustesse des descripteurs LBP, tout en soulignant l'importance du choix des paramètres pour maximiser leurs performances selon les jeux de données.

Tableau 4.1. Performances du système proposé avec les configurations optimales de descripteur LBP.

<i>Base de données</i>	<i>Paramètres du LBP (p, r)</i>	<i>Précisions [%]</i>
ORL	(16, 4)	85.1
LFW	(12, 8)	76.8
IJB-C	(12, 2)	73.4

2. Configuration des paramètres du HOG

De manière similaire, une phase d'ajustement des paramètres a été réalisée pour l'algorithme HOG. Cette étape se concentre sur deux aspects clés : la taille des blocs et le pourcentage de chevauchement entre les blocs adjacents. En particulier, l'objectif est de déterminer la taille optimale des blocs, tandis que les recherches indiquent qu'un chevauchement de 50 % entre les blocs est généralement suffisant pour garantir des performances efficaces de l'algorithme [9]. Pour identifier les meilleurs paramètres, des expérimentations ont été menées en testant différentes tailles de blocs, telles que 10×10, 12×12, jusqu'à 32×32, tout en maintenant un pourcentage de chevauchement constant. Cette exploration des tailles de blocs a permis d'évaluer leurs impacts sur la précision et de sélectionner la configuration optimale offrant les meilleurs résultats.

Le Tableau 2 présente les résultats obtenus par le système proposé sur les bases de données ORL, LFW et IJB-C en utilisant les tailles de blocs optimales. Il met en évidence l'impact des variations de ces tailles sur les performances de l'algorithme HOG, offrant ainsi des informations précieuses sur l'efficacité des différentes configurations pour la reconnaissance faciale.

Tableau 4.2. Performances du système proposé avec les configurations optimales de l'algorithme HOG

<i>Base de données</i>	<i>Paramètres du HOG ($w \times w$)</i>	<i>Précisions [%]</i>
ORL	16×16	87.0
LFW	20×20	79.8
IJB-C	12×12	76.2

Les résultats présentés dans le Tableau 4.2 montrent que la reconnaissance faciale atteint des niveaux de précision plus élevés avec le jeu de données ORL. Plus précisément, en utilisant une taille de bloc de 16×16, l'algorithme SVM, combiné à la méthode HOG, atteint une précision remarquable de **87,0 %**, mettant en évidence l'efficacité de HOG pour l'identification de visage. Pour le jeu de données LFW, les performances de reconnaissance restent compétitives par rapport à celles obtenues avec ORL. Une précision de **79,8 %** est atteinte avec une taille de bloc de 20×20, illustrant la robustesse de l'extraction de caractéristiques via HOG pour capturer les relations au

sein de LFW. Ces résultats s'alignent étroitement sur ceux observés avec le jeu de données ORL. En revanche, pour le jeu de données IJB-C, une légère diminution de la précision est constatée, atteignant **76,2 %** avec une taille de bloc de 12×12 . Cette baisse peut s'expliquer par la diversité du jeu de données, qui inclut des variations importantes telles que la pilosité faciale, la couleur de peau, et des changements significatifs de pose, rendant la tâche de reconnaissance faciale plus complexe.

3. Configuration des paramètres du Cbfd

La méthode Cbfd (Compressed Binary Facial Descriptor) repose sur une base de paramètres prédéfinis, dont certains nécessitent des ajustements et des tests pour optimiser les performances du système de reconnaissance faciale. À travers une série d'expériences, nous avons cherché à identifier les configurations les plus appropriées pour maximiser l'efficacité de cette technique.

L'algorithme Cbfd s'appuie sur plusieurs paramètres clés, notamment la taille de la fenêtre, le seuil binaire, la méthode de quantification et la dimension de la matrice de projection. Nos expérimentations se sont concentrées principalement sur l'impact de la taille de la fenêtre, avec pour objectif de déterminer la configuration optimale selon les métriques de performance. Pour cela, nous avons testé différentes tailles de régions, telles que (3, 3), (5, 5), (7, 7), (9, 9), (11, 11), (13, 13), (15, 15), et (17, 17). Ce paramètre a été systématiquement varié afin d'identifier la meilleure configuration en termes de performance. Concernant la quantification, nous avons adopté une méthode adaptative avec un seuil fixé à 0,9, permettant de discrétiser efficacement les caractéristiques à valeurs continues générées par Cbfd. Pour normaliser les caractéristiques, nous avons utilisé la normalisation Z-score, une technique garantissant l'invariance des données face aux variations d'apparence ou d'éclairage des images. Par ailleurs, pour prévenir le sur-apprentissage et améliorer la généralisation de l'algorithme, nous avons employé une régularisation L1 avec un paramètre lambda de 0,01. Cette régularisation contribue à renforcer la robustesse et l'efficacité de la méthode Cbfd dans des scénarios variés.

Tableau 4.3. Performances du système proposé avec les configurations optimales de l'algorithme Cbfd.

<i>Base de données</i>	<i>Taille de fenêtre(n ,n)</i>	<i>Précision [%]</i>
ORL	7×7	88,9
LFW	13×13	81,6
IJB-C	17×17	79.54

Le Tableau 4.3 présente les résultats obtenus à partir d'expériences de reconnaissance faciale menées avec la technique Cbfd, où différents paramètres de taille de fenêtre ont été utilisés. Les résultats montrent que lorsqu'une taille de fenêtre de 7×7 est utilisée, une précision impressionnante de 88,9 % est atteinte lorsqu'elle est appliquée au jeu de données ORL. De même, avec le jeu de données LFW, l'utilisation d'une taille de fenêtre de 13×13 permet d'obtenir un taux de reconnaissance d'environ 81,6 %. Cependant, en examinant le jeu de données IJB-C, une légère baisse de la précision de la reconnaissance faciale est observée, atteignant spécifiquement 79,54 % lorsqu'une taille de fenêtre de 17×17 est utilisée.

4. Performances du modèle proposé basé sur la combinaison des scores

Dans cette étape, les performances du modèle proposé s'appuient sur les paramètres optimaux des techniques d'extraction de caractéristiques afin de générer des scores de reconnaissance. Ces scores, représentant les performances respectives des techniques utilisées, sont ensuite fusionnés par concaténation pour former un vecteur global. Ce vecteur sert comme un vecteur d'entrée à un perceptron multicouche (MLP), qui agit comme un modèle de combinaison, chargé d'intégrer les résultats des différentes méthodes pour améliorer la reconnaissance faciale.

Dans un premier temps, des classifieurs basés sur les machines à vecteurs de support (SVM) sont entraînés en appliquant deux noyaux distincts : un noyau polynomial et un noyau RBF (Radial Basis Function). Les jeux de données LFW, ORL et IJB-C sont utilisés pour ces entraînements. Les paramètres adoptés sont les suivants : $n = 3$ pour le noyau polynomial, $\sigma = 0,125$ pour le noyau RBF, et $C = 0,2$ pour le paramètre de

régularisation des SVM. Ces configurations assurent une performance optimale pour chaque méthode d'extraction de caractéristiques.

Ensuite, le perceptron multicouche (MLP) intervient comme modèle de combinaison. Il est utilisé pour capturer les relations complexes entre les scores concaténés des SVM grâce à une architecture non linéaire. La fonction d'activation ReLU est utilisée dans les couches cachées, introduisant de la non-linéarité et permettant au modèle d'apprendre des représentations plus expressives. Enfin, la couche de sortie utilise une fonction d'activation Softmax pour produire des probabilités normalisées, facilitant ainsi la prise de décision dans les tâches de classification. Cette approche intégrée permet de tirer parti des forces de chaque technique d'extraction tout en améliorant la robustesse et la précision du système global de reconnaissance faciale.

Tableau 4.4. Performances du système proposé basé sur la fusion des scores

<i>Bases de données</i>	<i>Méthodes à noyaux</i>	<i>Précision</i>
LFW	Polynomial kernel	78.67%
	RBF kernel	82.43%
ORL	Polynomial kernel	93.72%
	RBF kernel	98.48%
IJB-C	Polynomial kernel	75.45%
	RBF kernel	81.84%

Le Tableau 4.4 présente une analyse détaillée des performances de la méthode de fusion basée sur un MLP appliquée à trois bases de données distinctes : LFW, ORL et IJB-C, en utilisant différentes méthodes de noyau. Parmi celles-ci, le noyau RBF se distingue comme le plus performant, affichant les niveaux de précision les plus élevés sur l'ensemble des bases de données. En détail, avec le noyau RBF, des taux de précision remarquables de **98,48 %** sont obtenus sur la base de données ORL, **82,43 %** sur LFW et **81,84 %** sur IJB-C. Ces résultats mettent en évidence l'efficacité du noyau RBF, qui s'impose comme l'option privilégiée pour maximiser la précision de reconnaissance dans les trois contextes étudiés.

Par ailleurs, le noyau RBF présente plusieurs avantages pratiques : il nécessite moins de paramètres et est moins sujet aux problèmes numériques que le noyau polynomial. Ces

qualités renforcent son attrait, en particulier pour les applications réelles, où sa robustesse et sa simplicité en font un choix de premier ordre pour les systèmes de fusion.

4.4.3.2. Performances de la reconnaissance faciale basée sur le deep learning

L'architecture de notre deuxième contribution pour la reconnaissance faciale comprend sept réseaux de neurones convolutifs (CNN), chacun dédié à la reconnaissance de régions faciales spécifiques : les yeux, le nez, la bouche, ainsi que les coins supérieurs gauche et droit, coins inférieur gauche et droit du visage, comme illustré dans la **Figure 4.4**.

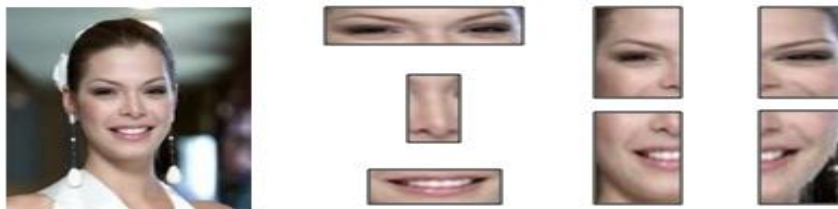


Figure 4.4. Les sept régions faciales utilisées dans la méthode proposée.

Pour obtenir une précision optimale en reconnaissance faciale, une série d'expériences a été menée pour chaque CNN, en ciblant individuellement les différentes régions faciales. Ces expériences avaient pour objectif de déterminer les paramètres idéaux des CNN. Les paramètres explorés incluaient les tailles des filtres (3×3, 5×5, 7×7, 11×11) et le nombre de filtres dans la première couche (8, 16, 32, 64, 128). Ce processus de sélection a été répété pour les couches suivantes de l'architecture.

Étant donné le grand nombre de combinaisons possibles (7 régions faciales × 4 tailles de filtres × 4 nombres de filtres = 112 configurations), le Tableau 4.5 présente une synthèse des meilleurs résultats obtenus. Il met en avant les précisions maximales atteintes sur les bases de données LFW, ORL et IJB-C, en soulignant les tailles et les nombres de filtres ayant joué un rôle déterminant dans ces performances.

Tableau 4.5. Performances des CNNs pour la reconnaissance des régions faciales.

		<i>Base de données</i>		
		LFW	ORL	IJB-C
Facial regions	Yeux	73.74 %	94.23 %	82.85%
	Nez	71.80 %	92.33%	79.54%
	Bouche	72.10%	96.37%	77.60%
	CoinSupérieur Gauche	74.10 %	95.28%	75.76%
	Coin Supérieur Droit	83.90%	97.95%	74.88%
	Coin Inférieur Gauche	84.20 %	96.10%	77.11%
	Coin Inférieur Droit	85.60 %	97.49%	79.19%

Les résultats présentés dans le Tableau 4.5 mettent en évidence l'efficacité de notre méthode de reconnaissance faciale pour l'identification des différentes régions faciales, à travers divers bases de données.

Pour le jeu de données **LFW**, notre approche affiche des performances globalement satisfaisantes. Parmi les régions analysées, le coin inférieur droit se distingue particulièrement, atteignant une précision notable de **85,60 %**, soulignant sa contribution significative à l'amélioration globale de la reconnaissance faciale.

En appliquant notre méthode au jeu de données **ORL**, les performances atteignent des niveaux encore plus élevés. Notamment, la région du coin supérieur droit se révèle être la plus performante, avec une précision exceptionnelle de **97,95 %**. Ce résultat met en avant la robustesse de notre méthode face aux variations contenues dans ce jeu de données.

Concernant le jeu de données **IJB-C**, notre approche montre des améliorations substantielles par rapport à l'approche basée sur l'apprentissage automatique. La précision obtenue atteint un impressionnant **82,85 %**, témoignant de la capacité de notre méthode à gérer efficacement les variations complexes, telles que les changements de pose, d'expression et de qualité d'image, présentes dans ce jeu de données.

Pour garantir une reconnaissance faciale précise et robuste face aux variations, les couches entièrement connectées (FullyConnected Layer) de chaque CNN sont concaténées afin de former un vecteur de caractéristiques global. Ce vecteur synthétise les informations discriminantes extraites de chaque région faciale. Il est ensuite utilisé comme entrée pour un classifieur basé sur un réseau de neurones profonds (DNN). Ce classifieur exploite les relations complexes entre les régions pour fournir des résultats de reconnaissance faciale robustes et précis. Cette stratégie combinée démontre ainsi l'efficacité de notre méthode dans divers contextes et bases de données.

Nos expériences incluent également l'optimisation du nombre de neurones dans la couche cachée, comme illustré dans la Figure 4.5. Nous avons mené une série de tests en utilisant un classifieur MLP avec un nombre variable de neurones, allant de 10 à 100, tout en maintenant un nombre constant d'itérations maximales (entre 2000 et 5000) et en utilisant l'erreur quadratique moyenne (MSE) pour l'entraînement. Les fonctions de transfert adoptées étaient de type sigmoïde, permettant d'obtenir une configuration optimale pour la reconnaissance faciale.

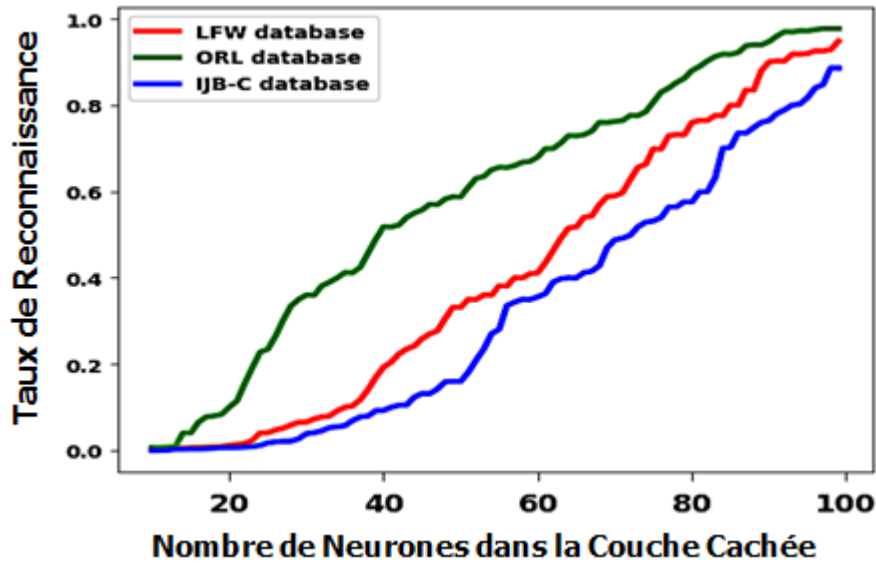


Figure 4.5. Taux de reconnaissance de la méthode proposée, fondée sur l'apprentissage profond, en fonction du nombre de neurones dans le DNN.

La Figure 4.5 présente une analyse détaillée des performances du système de reconnaissance faciale proposé. Les principales observations tirées de cette évaluation peuvent être résumées comme suit :

- ☞ **Amélioration significative avec plus de 90 neurones** : L'analyse révèle une nette augmentation des taux de reconnaissance lorsque la couche cachée du classifieur DNN contient plus de 90 neurones. Dans cette configuration, les performances atteignent des niveaux impressionnants, avec des taux variant entre 95,54 % et 97,75 %. Ces résultats soulignent l'efficacité de l'architecture DNN combinée aux CNN pour la reconnaissance faciale.
- ☞ **Influence des bases de données** : La performance varie selon la base de données utilisée. En particulier, la base de données ORL surpasse les bases de données LFW et IJB-C avec un taux de reconnaissance maximal de 97,75 %, contre 92,90 % pour LFW et 88,59 % pour IJB-C. Cela montre l'importance du choix des bases de données sur les résultats obtenus.
- ☞ **Limites avec 20 à 60 neurones** : Un nombre limité de neurones (entre 20 et 60) dans la couche cachée pose des défis. Dans cette plage, le classifieur MLP, intégré au DNN, éprouve des difficultés de convergence, notamment pour les bases de

données LFW et IJB-C. Ce problème peut être attribué à une capacité insuffisante de la couche cachée pour capturer efficacement les caractéristiques des données.

☞ **Défis spécifiques aux bases de données** : Les bases LFW et IJB-C posent des difficultés uniques. Les variations marquées d'orientation et d'expressions faciales dans LFW compliquent la reconnaissance, tandis que les faibles résolutions et les variations de qualité des images dans IJB-C augmentent la complexité du processus.

Ces observations mettent en lumière la nécessité d'adapter l'architecture du modèle et les paramètres en fonction des caractéristiques des données pour optimiser les performances.

4.5. Évaluation du modèle proposé basée sur les fonctions de perte

L'objectif principal de la reconnaissance faciale, englobant la vérification et l'identification, repose sur la différenciation efficace des caractéristiques faciales. Toutefois, la fonction de perte Softmax classique utilisée dans les réseaux de neurones convolutifs profonds (CNN) montre souvent des limites en termes de capacité discriminante. Pour surmonter cette insuffisance, plusieurs fonctions de perte innovantes ont été développées ces dernières années, notamment la perte à grande marge cosinus (CosFace) [147], la perte d'angle additive (ArcFace) [148] et la perte SphereFace [149].

Ces fonctions de perte avancées visent à renforcer le pouvoir discriminant des embeddings de caractéristiques en établissant une relation spécifique entre les vecteurs de caractéristiques et les centroïdes des classes.

Dans cette section, nous évaluons les performances du modèle basé sur l'apprentissage profond proposé en intégrant, de manière indépendante, deux de ces fonctions de perte : CosFace et ArcFace.

4.5.1. Détails de l'implémentation

Les étapes clés pour intégrer chaque fonction de perte (CosFace et ArcFace) dans un CNN de base sont les suivantes :

- A.** Dans la dernière couche entièrement connectée de chaque CNN du modèle, une couche supplémentaire a été ajoutée spécifiquement pour le calcul de la fonction de perte intégrée. Cette couche prend en entrée les vecteurs de caractéristiques générés par les couches précédentes et applique la fonction de perte sélectionnée.

Les fonctions de perte utilisées s'expriment comme suit :

☞ **Fonction de perte CosFace**

$$L_A = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{(\cos(\theta_{y_i})-m)}}{e^{(\cos(\theta_{y_i})-m)} + \sum_{j=1, j \neq y_i}^N e^{\cos \theta_j}} \quad (4.6)$$

☞ **Fonction de perte ArcFace**

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{\cos(\theta_{y_i}+m)}}{e^{\cos(\theta_{y_i}+m)} + \sum_{j=1, j \neq y_i}^N e^{\cos \theta_j}} \quad (4.7)$$

Les principaux paramètres utilisés dans ces fonctions de perte sont : 1) le paramètre θ qui contrôle l'échelle des scores de similarité cosinus. Il détermine à quel point nous voulons amplifier ou réduire la marge angulaire appliquée aux valeurs de similarité cosinus, et 2) le paramètre m qui spécifie la marge angulaire ajoutée à la similarité cosinus entre les caractéristiques et les vecteurs de poids associés aux classes correctes.

- B.** La sortie générée par la couche de perte spécifiée a été utilisée comme résultat final de notre modèle CNN. Cette sortie a ensuite servi pour les phases d'entraînement, d'évaluation et de test dans le cadre de nos expériences.
- C.** Enfin, nous avons évalué notre modèle entraîné sur des bases de données de test, en prenant en compte la fonction de perte spécifiée pour l'encodage des caractéristiques et la classification.

Dans notre configuration expérimentale, nous avons sélectionné les valeurs de marge $m = 0,35$ pour CosFace et $m = 0,50$ pour ArcFace. Ces valeurs ont été choisies en raison de leur efficacité démontrée pour obtenir des performances constantes, en

particulier sur des bases de données de petite taille, comme le soulignent des travaux antérieurs [148].

Tableau 4.6. Analyse des performances de la méthode proposée basée sur la fonction de perte CosFace et ArcFace.

<i>Fonction de Perte</i>	<i>ORL</i>	<i>LFW</i>	<i>IJB-C</i>
Méthode proposée basée sur Softmax	97.75%	92.20%	88.59%
Méthode proposée basée sur CosFace	97.96%	93.35%	91.60%
Méthode proposée basée sur ArcFace	98.95%	96.80%	92.25%

Comme le montre le Tableau 4.6, l'évaluation des performances du modèle proposé sur les bases de données ORL, LFW et IJB-C révèle que l'utilisation de la fonction de perte ArcFace permet d'obtenir les meilleurs résultats. Le modèle proposé utilisant ArcFace a surpassé les variantes basées sur Softmax et CosFace, atteignant des taux de reconnaissance de **98,95 %** sur la base de données ORL, **96,80 %** sur LFW et **92,25 %** sur IJB-C.

La variante du modèle utilisant la perte CosFace a également démontré des performances solides, supérieures à celles obtenues avec Softmax. Les taux de reconnaissance enregistrés étaient de **97,96 %** pour ORL, **93,35 %** pour LFW et **91,60 %** pour IJB-C.

En revanche, la méthode proposée utilisant la fonction de perte Softmax a affiché les performances les plus faibles parmi les trois approches, avec des taux de reconnaissance de **97,75 %** pour ORL, **92,20 %** pour LFW et **88,59 %** pour IJB-C.

Il est également important de noter que les bases de données ORL et LFW ont systématiquement conduit à des taux de reconnaissance plus élevés par rapport à la base de données IJB-C, et ce, pour chacune des trois fonctions de perte évaluées. Ces résultats soulignent non seulement l'efficacité du modèle proposé associé à des fonctions de perte avancées, mais aussi les différences inhérentes aux caractéristiques des bases de données, qui influencent la performance globale du système.

4.6. Étude comparative

Dans cette section, nous procédons à une comparaison détaillée des performances de notre méthode proposée avec celles des techniques récentes de pointe [6, 73, 83, 85, 91, 94], y compris celles reposant sur diverses fonctions de perte [89, 96]. Les résultats de cette comparaison sont présentés dans le Tableau 4.7, qui récapitule les précisions obtenues par ces méthodes, couvrant à la fois les approches d'apprentissage automatique et d'apprentissage profond.

Pour notre méthode basée sur l'apprentissage automatique, l'évaluation démontre que la fusion des classifieurs atteint des performances impressionnantes, avec des précisions de **98,48 %** sur la base de données ORL, **82,43 %** sur LFW et **81,84 %** sur IJB-C. Il est à noter que notre approche dépasse celle de Muqet et al. [85], qui a obtenu une précision de **97,00 %**. Par ailleurs, en comparaison avec les méthodes basées sur l'apprentissage profond, notre approche surpasse également les résultats obtenus par Kong et al. [73], utilisant PCANet + KNN et PCANet + SVM, avec des précisions respectives de **91,50 %** et **97,50 %**.

Le Tableau 4.7 montre également une comparaison détaillée des performances de notre méthode d'apprentissage profond par rapport aux approches récentes dans ce domaine. Sur la base de données ORL, notre méthode atteint une précision de **97,75 %**, surpassant ainsi les résultats de Kong et al. [73], avec des précisions de **91,50 %** et **97,50 %**. Pour la base de données LFW, notre approche obtient une précision de **92,20 %**, dépassant la combinaison FaceNet + RF [91] (précision de **89,10 %**) et la combinaison MLP + MFM avec CNN [94] (précision de **84,5 %**). En outre, notre méthode se trouve en concurrence avec celle de Storey et al. [6], qui a obtenu une précision de **93,60 %**.

Enfin, nous comparons notre méthode proposé avec les méthodes de fonction de perte [96, 89], comme détaillé dans le Tableau 4.7. Sur les bases de données LFW et IJB-C, notre modèle surpasse l'approche présentée dans [89], atteignant des précisions de **93,35 %** et **96,80 %** pour les fonctions de perte CosFace et ArcFace respectivement. Par rapport à la méthode de [96], notre approche montre des résultats compétitifs, obtenant des précisions de **91,60 %** pour CosFace et **92,25 %** pour ArcFace, contre **92,20 %** pour CosFace et **95,20 %** pour ArcFace dans [96].

Tableau 4.7. Analyse comparative des méthodes proposées avec les techniques de l'état de l'art.

Apprentissage automatique.

Méthode	Base de données		
	ORL	LFW	IJB-C
KNN (DWT+LBP) [85]	97.00%	-	-
Méthode proposée	98.48%	82.43%	81.84%

Apprentissage profond.

FaceNet + RF [91]	-	89.10%	-
LBP + Base CNN [83]	100%		-
MLP + MFM in CNN [94]	-	84.50%	-
3D-CNN+ResNe [6]	-	93.60%	-
PCANet + KNN [73]	91.50%	-	-
PCANet + SVM[73]	-	97.50%	-
ResNet-100 _{CosFace} [96]	-	-	92.20%
ResNet-100 _{ArcFace} [96]	-	-	95.20%
LMCL _{CosFace} [89]	-	92.69%	-
LMCL _{ArcFace} [89]	-	93.30%	-
Comparaison basée sur les fonctions de perte			
Méthode proposée basée sur Softmax	97.75%	92.20%	88.59%
Méthode proposée basée sur CosFace	97.96%	93.35%	91.60%

Méthode proposée basée sur ArcFace	98.95%	96.80%	92.25%
------------------------------------	--------	--------	--------

4.7. Conclusion

Au cours de ce chapitre, nous avons proposé une approche de reconnaissance faciale utilisant l'apprentissage automatique, structurée en plusieurs étapes. Le prétraitement de l'image, incluant la détection du visage et le recadrage, est crucial pour extraire les informations pertinentes. L'algorithme Landmark-68 permet une localisation précise des points de repère faciaux, optimisant ainsi cette étape.

Ensuite, trois techniques d'extraction de caractéristiques (LBP, Cbfd, et HOG) sont utilisées pour capturer une variété de données discriminantes. Ces caractéristiques sont combinées à l'aide d'un modèle de fusion des scores basé sur un perceptron multicouche (MLP), permettant une décision finale optimale. Les résultats expérimentaux montrent des performances supérieures à celles de plusieurs méthodes de l'état de l'art, avec des taux de reconnaissance élevés.

Enfin, une approche basée sur l'apprentissage profond traite chaque région faciale indépendamment avant de fusionner les résultats dans un modèle coopératif, renforçant ainsi la précision de la reconnaissance, même avec des variations d'expressions ou d'angles. Cette méthode offre un modèle robuste et précis, adapté à des applications telles que la sécurité biométrique, le diagnostic médical et la réalité augmentée.

CONCLUSION ET PERSPECTIVES

La bataille contre l'ignorance se gagne tous les jours, et elle finit par ouvrir sur des perspectives insoupçonnées.

Dalai Lama. Leader Tibétain

Ce travail de recherche aborde les défis de la reconnaissance faciale à travers l'introduction de deux approches innovantes : l'une basée sur l'apprentissage automatique et l'autre sur l'apprentissage profond. L'efficacité de trois techniques, à savoir LBP, HOG et Cbfd, est évaluée pour surmonter ces défis. La solution proposée repose sur l'utilisation d'un nouveau modèle de combinaison de classifieurs multi-classes et d'une méthode unique pour extraire des caractéristiques de haut niveau à partir de plusieurs régions d'image traitées comme des données séquentielles à l'aide d'un ensemble de CNN, suivie d'un classifieur DNN pour la reconnaissance faciale.

Les résultats expérimentaux obtenus à partir de bases de données faciales renommées, telles que LFW, ORL et IJB-C, révèlent la performance compétitive du modèle de combinaison de classifieurs multi-classes proposé ainsi que du modèle de deep learning, comparés aux méthodes de pointe. De plus, nous avons évalué l'efficacité du modèle basé sur l'apprentissage profond proposé en parallèle avec des fonctions de perte de pointe, telles que CosFace et ArcFace. En nous basant sur les résultats obtenus lors de nos expériences, nous pouvons mettre en lumière les forces et faiblesses spécifiques de notre approche :

- Les résultats expérimentaux montrent que la méthode de l'apprentissage automatique, basée sur une combinaison au niveau du score de correspondance, est susceptible de fournir de meilleures performances de reconnaissance car elle contient plus d'informations pertinentes, ce qui est à la fois faisable et pratique.
- Cette thèse illustre comment utiliser le CNN comme un modèle séquentiel, et nous pensons qu'il pourrait ouvrir la voie à une alternative aux réseaux neuronaux profonds

pour de nombreuses tâches. Le CNN traditionnel traite une entrée et passe à la suivante en ignorant sa séquence. Dans le modèle basé sur l'apprentissage profond proposé, une image est considérée comme une série de régions faciales séquentielles qui doivent être suivies pour être comprises. En d'autres termes, le premier CNN reçoit une région d'image et la transmet sous forme de vecteur de caractéristiques au CNN suivant pour prédire la région faciale suivante en fonction de la région précédente, et ainsi de suite.

- De plus, nous tenons à mentionner qu'il est possible que le modèle basé sur l'apprentissage profond proposé puisse être utilisé dans d'autres applications telles que l'estimation de l'âge, la prédiction du sexe ou la reconnaissance des émotions faciales.
- Cependant, les méthodes proposées doivent être suffisamment précises et robustes pour gérer la variabilité et la diversité des visages et des ensembles de données.

Dans les futures recherches, nous envisageons d'explorer l'application de mécanismes d'attention (attention mechanisms) [150] pour identifier automatiquement les régions faciales distinctives tout en minimisant efficacement l'impact des zones bruyantes. Le bruit, qui peut provenir de maquillage, lunettes, ou angles de prise de vue, peut fausser les résultats. Les mécanismes d'attention aideraient à minimiser l'effet de ces perturbations en leur attribuant un faible poids dans le processus de décision.

- [1] Bendib, I., Meraoumia, A., Haouam, M. Y., & Laimeche, L. (2022). A new cancelable deep biometric feature using chaotic maps. *Pattern Recognition and Image Analysis*, 32(1), 109-128.
- [2] Haouam, M. Y., Meraoumia, A., Laimeche, L., & Bendib, I. (2021). S-DCTNet: Security-oriented biometric feature extraction technique: An effective pathway to secure and reliable biometric systems. *Multimedia Tools and Applications*, 80(28), 36059-36091.
- [3] Rameswari, R., Kumar, S. N., Aananth, M. A., & Deepak, C. (2021). Automated access control system using face recognition. *Materials Today: Proceedings*, 45, 1251-1256.
- [4] Shi, L., Wang, X., & Shen, Y. (2020). Research on 3D face recognition method based on LBP and SVM. *Optik*, 220, 165157.
- [5] Kong, W., You, Z., & Lv, X. (2023). 3D face recognition algorithm based on deep Laplacian pyramid under the normalization of epidemic control. *Computer Communications*, 199, 30-41.
- [6] Storey, G., Jiang, R., Keogh, S., Bouridane, A., & Li, C. T. (2019). 3DPalsyNet: A facial palsy grading and motion recognition framework using fully 3D convolutional neural networks. *IEEE access*, 7, 121655-121664.
- [7] Ramgopal, M., Roopesh, M. S., Chowdary, M. V., Madhav, M., & Shanmuga, K. (2022). Masked Facial Recognition in Security Systems Using Transfer Learning. *SN Computer Science*, 4(1), 27.
- [8] Samai, D., Meraoumia, A., Bendjenna, H., & Laimeche, L. (2017, December). Oriented Local Binary Pattern (LBP θ): A new scheme for an efficient feature extraction technique. In *2017 International Conference on Mathematics and Information Technology (ICMIT)* (pp. 155-161). IEEE.
- [9] Başa, B. (2015). Implementation of Hog Edge Detection Algorithm Onfpga's. *Procedia-Social and Behavioral Sciences*, 174, 1567-1575.
- [10] Lu, J., Liong, V. E., Zhou, X., & Zhou, J. (2015). Learning compact binary face descriptor for face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(10), 2041-2056.
- [11] Abdulrahman, S. A., & Alhayani, B. (2023). A comprehensive survey on the biometric systems based on physiological and behavioural characteristics. *Materials Today: Proceedings*, 80, 2642-2646.
- [12] Jain, A. K., Ross, A., & Prabhakar, S. (2004). An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1), 4-20.
- [13] International Biométrie Group-IBG, En ligne : <http://www.biometricgroup.com/reports/publicmarketreport.html> accessed on 28/11/2024
- [14] Dargan, S., & Kumar, M. (2020). A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities. *Expert Systems with Applications*, 143, 113114.
- [15] Adeoye, O. S. (2010). A survey of emerging biometric technologies. *international journal of computer applications*, 9(10), 1-5.
- [16] Liu, X., Pedersen, M., Charrier, C., & Bours, P. (2018, September). Can image quality enhancement methods improve the performance of biometric systems for degraded face images?. In *2018 Colour and Visual Computing Symposium (CVCS)* (pp. 1-5). IEEE.
- [17] Joshi, M., Mazumdar, B., & Dey, S. (2018). Security vulnerabilities against fingerprint biometric system. arXiv preprint arXiv:1805.07116.

- [18] Ammour, B., Bouden, T., & Boubchir, L. (2018). Face-iris multi-modal biometric system using multi-resolution Log-Gabor filter with spectral regression kernel discriminant analysis. *IET biometrics*, 7(5), 482-489.
- [19] Brockly, M., Guest, R., Elliott, S., & Scott, J. (2011, October). Dynamic signature verification and the human biometric sensor interaction model. In *2011 Carnahan Conference on Security Technology* (pp. 1-6). IEEE.
- [20] Alsaadi, I. M. (2021). Study on most popular behavioral biometrics, advantages, disadvantages and recent applications: A review. *Int. J. Sci. Technol. Res*, 10(1).
- [21] Modak, S. K. S., & Jha, V. K. (2019). Multibiometric fusion strategy and its applications: A review. *Information Fusion*, 49, 174-204.
- [22] Boyd, J. E., & Little, J. J. (2005). Biometric gait recognition. In *Advanced Studies in Biometrics: Summer School on Biometrics, Alghero, Italy, June 2-6, 2003. Revised Selected Lectures and Papers* (pp. 19-42). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [23] Querini, M., Gattelli, M., Gentile, V., & Italiano, G. F. (2015). A new system for secure handwritten signing of documents. *INTERNATIONAL JOURNAL OF COMPUTER SCIENCE & APPLICATIONS*, 12(2), 37-56.
- [24] Singh, B., Sonawane, S., Shah, Y., & Singh, V. (2017). Literature survey on keystroke dynamics for user authentication. *International Journal on Recent and Innovation Trends in Computing and Communication*, 5(5), 280-282.
- [25] Butler, J. M., McCord, B. R., & Shen, Y. (2003). The development of reduced size STR amplicons as tools for analysis of degraded DNA. *Journal of forensic sciences*, 48(5), JFS2003043.
- [26] Sarkar, I., Alisherov, F., Bhattacharyya, D., & Kim, T. H. (2010). Palm vein authentication system: a review. *International Journal of Control and Automation*, 3(1), 27-34.
- [27] Mulyono, D., & Jinn, H. S. (2008, April). A study of finger vein biometric for personal identification. In *2008 International Symposium on Biometrics and Security Technologies* (pp. 1-8). IEEE.
- [28] Li, S., Fei, L., Zhang, B., Ning, X., & Wu, L. (2024). Hand-based multimodal biometric fusion: A review. *Information Fusion*, 102418.
- [29] Jain, A. K., & Ross, A. (2004). Multibiometric systems. *Communications of the ACM*, 47(1), 34-40.
- [30] Telgad, R. L., Deshmukh, P. D., & Siddiqui, A. M. (2014, May). Combination approach to score level fusion for Multimodal Biometric system by using face and fingerprint. In *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)* (pp. 1-8). IEEE.
- [31] Cherrat, E., Alaoui, R., & Bouzahir, H. (2019). Système d'Identification Biométrique par Fusion Multimodale.
- [32] Singh, M., Singh, R., & Ross, A. (2019). A comprehensive overview of biometric fusion. *Information Fusion*, 52, 187-205.
- [33] Muhammad, A. R. I. F. (2005). Fusion de Données: Ultime Etape de Reconnaissance de Formes, Applications à l'Identification et à l'Authentification. *these de doctorat, Universite de Tours*.

- [34] Soyata, T., Muraleedharan, R., Funai, C., Kwon, M., & Heinzelman, W. (2012, July). Cloud-vision: Real-time face recognition using a mobile-cloudlet-cloud acceleration architecture. In *2012 IEEE symposium on computers and communications (ISCC)* (pp. 000059-000066). IEEE.
- [35] Kortli, Y., Jridi, M., Al Falou, A., & Atri, M. (2020). Face recognition systems: A survey. *Sensors*, *20*(2), 342.
- [36] Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, *19*(7), 711-720.
- [37] He, X., Yan, S., Hu, Y., Niyogi, P., & Zhang, H. J. (2005). Face recognition using laplacianfaces. *IEEE transactions on pattern analysis and machine intelligence*, *27*(3), 328-340.
- [38] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2008). Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, *31*(2), 210-227.
- [39] Lei, Z., Li, S. Z., Chu, R., & Zhu, X. (2007). Face recognition with local gabor textons. In *Advances in Biometrics: International Conference, ICB 2007, Seoul, Korea, August 27-29, 2007. Proceedings* (pp. 49-57). Springer Berlin Heidelberg.
- [40] Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, *28*(12), 2037-2041.
- [41] Guo, G., Li, S. Z., & Chan, K. (2000, March). Face recognition by support vector machines. In *Proceedings fourth IEEE international conference on automatic face and gesture recognition (cat. no. PR00580)* (pp. 196-201). IEEE.
- [42] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [43] Raji, I. D., & Fried, G. (2021). About face: A survey of facial recognition evaluation. *arXiv preprint arXiv:2102.00813*.
- [44] Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1701-1708).
- [45] Sun, Y., Wang, X., & Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1891-1898).
- [46] Wang, M., & Deng, W. (2021). Deep face recognition: A survey. *Neurocomputing*, *429*, 215-244.
- [47] <https://vis-www.cs.umass.edu/lfw/> accessed on 28/11/2024
- [48] <http://biometrics.idealtest.org/findTotalDbByMode.do?mode=Face#/datasetDetail/9> accessed on 28/11/2024
- [49] Guo, Y., Zhang, L., Hu, Y., He, X., & Gao, J. (2016). Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14* (pp. 87-102). Springer International Publishing.

- [50] Masupha, L., Zuva, T., Ngwira, S., & Esan, O. (2015, December). Face recognition techniques, their advantages, disadvantages and performance evaluation. In *2015 International Conference on Computing, Communication and Security (ICCCS)* (pp. 1-5). IEEE.
- [51] Rath, S. K., & Rautaray, S. S. (2014). A survey on face detection and recognition techniques in different application domain. *International Journal of Modern Education and Computer Science*, 6(8), 34.
- [52] Wójcik, W., Gromaszek, K., & Junisbekov, M. (2016). Face recognition: Issues, methods and alternative applications. *Face Recognition-Semisupervised Classification, Subspace Projection and Evaluation Methods*, 7-28.
- [53] Parmar, D. N., & Mehta, B. B. (2014). Face recognition methods & applications. *arXiv preprint arXiv:1403.0485*.
- [54] Cheng, R. (2020). A survey: Comparison between Convolutional Neural Network and YOLO in image identification. In *Journal of Physics: Conference Series* (Vol. 1453, No. 1, p. 012139). IOP Publishing.
- [55] Jawad, T., Maisha, B., & Nazmus, S. (2019). Targeted face recognition and alarm generation for security surveillance using single shot multibox detector (SSD). *Int. J. Comput. Appl*, 177, 8-13.
- [56] Han, H., Shan, S., Chen, X., & Gao, W. (2013). A comparative study on illumination preprocessing in face recognition. *Pattern Recognition*, 46(6), 1691-1699.
- [57] Banerjee, S., Brogan, J., Krizaj, J., Bharati, A., Webster, B. R., Struc, V., ... & Scheirer, W. J. (2018, March). To frontalize or not to frontalize: Do we really need elaborate pre-processing to improve face recognition?. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 20-29). IEEE.
- [58] Taskiran, M., Kahraman, N., & Erdem, C. E. (2020). Face recognition: Past, present and future (a review). *Digital Signal Processing*, 106, 102809.
- [59] Bauer, S., Brunsmann, U., & Schlotterbeck-Macht, S. (2009, July). FPGA implementation of a HOG-based pedestrian recognition system. In *Proc. MPC-Workshop* (pp. 49-58).
- [60] Barnouti, N. H., Al-Dabbagh, S. S. M., & Matti, W. E. (2016). Face recognition: A literature review. *International Journal of Applied Information Systems*, 11(4), 21-31.
- [61] Singh, S., & Prasad, S. V. A. V. (2018). Techniques and challenges of face recognition: A critical review. *Procedia computer science*, 143, 536-543.
- [62] Javed Mehedi Shamrat, F. M., Ghosh, P., Tasnim, Z., Khan, A. A., Uddin, M. S., & Chowdhury, T. R. (2022). Human Face recognition using eigenface, SURF method. In *Pervasive Computing and Social Networking: Proceedings of ICPCSN 2021* (pp. 73-88). Springer Singapore.
- [63] The Olivetti Research Laboratory (ORL) face dataset :<https://www.kaggle.com/datasets/tavarez/the-orl-database-for-training-and-testing>
- [64] Septi, S. Q. N., Yulita, I. N., & Napitupulu, H. (2021, October). Face Recognition Using Fisherface and Support Vector Machine Method. In *2021 International Conference on Artificial Intelligence and Big Data Analytics* (pp. 50-55). IEEE.
- [65] Pati, R., Pujari, A. K., & Gahan, P. (2021). Face recognition using particle swarm optimization based block ICA. *Multimedia Tools and Applications*, 80(28), 35685-35695.

- [66] Jiang, Y., & Guo, P. (2007, October). Comparative studies of feature extraction methods with application to face recognition. In *2007 IEEE International Conference on Systems, Man and Cybernetics* (pp. 3627-3632). IEEE.
- [67] Vyas, R. A., & Shah, S. M. (2017). Comparison of PCA and LDA techniques for face recognition feature based extraction with accuracy enhancement. *International Research Journal of Engineering and Technology (IRJET)*, 4(6), 3332-3336.
- [68] Bouzalmat, A., Kharroubi, J., & Zarghili, A. (2014). Comparative study of PCA, ICA, LDA using SVM classifier. *Journal of Emerging Technologies in Web Intelligence*, 6(1), 64-68.
- [69] Pentland, Moghaddam, & Starner. (1994, June). View-based and modular eigenspaces for face recognition. In *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (pp. 84-91). IEEE.
- [70] Lu, J., Plataniotis, K. N., & Venetsanopoulos, A. N. (2003). Face recognition using LDA-based algorithms. *IEEE Transactions on Neural networks*, 14(1), 195-200.
- [71] MageshKumar, C., Thiyagarajan, R., Natarajan, S. P., Arulselvi, S., & Sainarayanan, G. (2011, March). Gabor features and LDA based face recognition with ANN classifier. In *2011 International Conference on Emerging Trends in Electrical and Computer Technology* (pp. 831-836). IEEE.
- [72] Maafiri, A., & Chougali, K. (2021). Robust face recognition based on a new Kernel-PCA using RRQR factorization. *Intelligent data analysis*, 25(5), 1233-1245.
- [73] Kong, J., Chen, M., Jiang, M., Sun, J., & Hou, J. (2018). Face recognition based on CSGF (2D) 2 PCANet. *IEEE Access*, 6, 45153-45165.
- [74] Chen, X., Zhang, C., Dong, F., & Zhou, Z. (2013, July). Parallelization of elastic bunch graph matching (EBGM) algorithm for fast face recognition. In *2013 IEEE China Summit and International Conference on Signal and Information Processing* (pp. 201-205). IEEE.
- [75] Wiskott, L., Fellous, J. M., Kruger, N., & Malsburg, C. V. D. (1996). Face recognition by elastic bunch graph matching. *TR96-08, Institut für Neuroinformatik, Ruhr-Universität Bochum*.
- [76] Krüger, N., Pöttsch, M., & von der Malsburg, C. (1997). Determination of face position and pose with a learned representation based on labelled graphs. *Image and vision computing*, 15(8), 665-673.
- [77] Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., Von Der Malsburg, C., Wurtz, R. P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on computers*, 42(3), 300-311.
- [78] Vasanthi, M., & Seetharaman, K. (2022). Facial image recognition for biometric authentication systems using a combination of geometrical feature points and low-level visual features. *Journal of King Saud University-Computer and Information Sciences*, 34(7), 4109-4121.
- [79] Iqtait, M., Mohamad, F. S., & Mamat, M. (2018, March). Feature extraction for face recognition via active shape model (ASM) and active appearance model (AAM). In *IOP Conference Series: Materials Science and Engineering* (Vol. 332, No. 1, p. 012032). IOP Publishing.
- [80] Zheng, D., Zhang, K., Lu, J., Jing, J., & Xiong, Z. (2020). Active discriminative cross-domain alignment for low-resolution face recognition. *IEEE Access*, 8, 97503-97515.
- [81] Muqheet, M. A., & Holambe, R. S. (2019). Local binary patterns based on directional wavelet transform for expression and pose-invariant face recognition. *Applied Computing and Informatics*, 15(2), 163-171.
- [82] Zhou, L., Liu, W., Lu, Z. M., & Nie, T. (2014). Face recognition based on curvelets and local binary pattern features via using local property preservation. *Journal of Systems and Software*, 95, 209-216.

- [83] Tang, J., Su, Q., Su, B., Fong, S., Cao, W., & Gong, X. (2020). Parallel ensemble learning of convolutional neural networks and local binary patterns for face recognition. *Computer Methods and Programs in Biomedicine*, 197, 105622.
- [84] Singhal, N., Ganganwar, V., Yadav, M., Chauhan, A., Jakhar, M., & Sharma, K. (2021). Comparative study of machine learning and deep learning algorithm for face recognition. *Jordanian Journal of Computers and Information Technology*, 7(3).
- [85] Zhou, L., Wang, H., Lin, S., Hao, S., & Lu, Z. M. (2020). Face recognition based on local binary pattern and improved pairwise-constrained multiple metric learning. *Multimedia Tools and Applications*, 79, 675-691.
- [86] Zhou, L., Wang, H., Lu, Z. M., Nie, T., & Zhao, K. (2016). Face recognition based on LDA and improved pairwise-constrained multiple metric learning method. *Journal of Information Hiding and Multimedia Signal Processing*, 7(5), 1092.
- [87] Kar, A., Bhattacharjee, D., Nasipuri, M., Basu, D. K., & Kundu, M. (2013). High performance human face recognition using gabor based pseudo hidden Markov model. *International Journal of Applied Evolutionary Computation (IJAEC)*, 4(1), 81-102.
- [88] Ali, D., Touqir, I., Siddiqui, A. M., Malik, J., & Imran, M. (2022). Face recognition system based on four state hidden Markov model. *IEEE Access*, 10, 74436-74448.
- [89] Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4690-4699).
- [90] Zhang, W., & Xiang, S. (2020). Face anti-spoofing detection based on DWT-LBP-DCT features. *Signal Processing: Image Communication*, 89, 115990.
- [91] Sanchez-Moreno, A. S., Olivares-Mercado, J., Hernandez-Suarez, A., Toscano-Medina, K., Sanchez-Perez, G., & Benitez-Garcia, G. (2021). Efficient face recognition system for operating in unconstrained environments. *Journal of Imaging*, 7(9), 161.
- [92] Mamieva, D., Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Improved face detection method via learning small faces on hard images based on a deep learning approach. *Sensors*, 23(1), 502.
- [93] Z. Mortezaie and H. Hassanpour, "A Survey on Age-invariant Face Recognition Methods," *Jordanian Journal of Computers and Information Technology (JJCIT)*, vol. 05, no. 02, pp. 87 - 96, August 2019.
- [94] Yu, H., Zhao, J., & Zhu, Y. (2019, October). Research on face recognition method based on deep learning. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 1-5). IEEE.
- [95] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010, June). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops* (pp. 94-101). IEEE.
- [96] Hsu, G. S. J., Wu, H. Y., Tsai, C. H., Yanushkevich, S., & Gavrilova, M. L. (2022). Masked face recognition from synthesis to reality. *IEEE Access*, 10, 37938-37952.
- [97] Grgic, M., Delac, K., & Grgic, S. (2011). SCface—surveillance cameras face database. *Multimedia tools and applications*, 51, 863-879.
- [98] Zhou, Z. H., & Feng, J. (2019). Deep forest. *National science review*, 6(1), 74-86.

- [99] Maafiri, A., Bir-Jmel, A., Elharrouss, O., Khelifi, F., & Chougali, K. (2022). Lwkpca: A new robust method for face recognition under adverse conditions. *IEEE Access*, 10, 64819-64831.
- [100] <https://vision.ucsd.edu/datasets/yale-face-database> accessed on 28/11/2024
- [101] Gupta, P., Saxena, N., Sharma, M., & Tripathi, J. (2018). Deep neural network for human face recognition. *International Journal of Engineering and Manufacturing (IJEM)*, 8(1), 63-71.
- [102] Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition*.
- [103] Maze, B., Adams, J., Duncan, J. A., Kalka, N., Miller, T., Otto, C., ... & Grother, P. (2018, February). Iarpa janus benchmark-c: Face dataset and protocol. In *2018 international conference on biometrics (ICB)* (pp. 158-165). IEEE.
- [104] <https://www.nist.gov/itl/products-and-services/color-feret-database> accessed on 28/11/2024
- [105] Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, 16(5), 295-306.
- [106] <https://www.iab-rubric.org/resources/biometric-datasets/face> accessed on 28/11/2024
- [107] Nappi, M., Ricciardi, S., & Tistarelli, M. (2016). Deceiving faces: When plastic surgery challenges face recognition. *Image and Vision Computing*, 54, 71-82.
- [108] <https://github.com/DanJun6737/TransFace?tab=readme-ov-file> accessed on 28/11/2024
- [109] Shi, Y., & Jain, A. K. (2021). Boosting unconstrained face recognition with auxiliary unlabeled data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2795-2804).
- [110] <https://academictorrents.com/details/9e67eb7cc23c9417f39778a8e06cca5e26196a97/tech&hit=1&filelist=1> accessed on 28/11/2024
- [111] Li, Y., Zeng, J., Zhang, J., Dai, A., Kan, M., Shan, S., & Chen, X. (2017, October). Kinnet: Fine-to-coarse deep metric learning for kinship verification. In *Proceedings of the 2017 workshop on recognizing families in the wild* (pp. 13-20).
- [112] Yi, D., Lei, Z., Liao, S., & Li, S. Z. (2014). Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*.
- [113] Boutros, F., Huber, M., Siebke, P., Rieber, T., & Damer, N. (2022, October). Sface: Privacy-friendly and accurate face recognition using synthetic data. In *2022 IEEE International Joint Conference on Biometrics (IJCB)* (pp. 1-11). IEEE.
- [114] <https://github.com/bioidiap/bob.db.caspeal> accessed on 28/11/2024
- [115] Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., & Zhao, D. (2007). The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(1), 149-161.

- [116] Pflug, A., & Busch, C. (2012). Ear biometrics: a survey of detection, feature extraction and recognition methods. *IET biometrics*, 1(2), 114-129.
- [117] Edwards, G. J., Taylor, C. J., & Cootes, T. F. (1998, April). Interpreting face images using active appearance models. In *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 300-305). IEEE.
- [118] Gross, R., Matthews, I., & Baker, S. (2004). Appearance-based face recognition and light-fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(4), 449-465.
- [119] Gross, R., Matthews, I., & Baker, S. (2005). Generic vs. person specific active appearance models. *Image and Vision Computing*, 23(12), 1080-1093.
- [120] Jiao, F., Li, S., Shum, H. Y., & Schuurmans, D. (2003, June). Face alignment using statistical models and wavelet features. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.* (Vol. 1, pp. I-I). IEEE.
- [121] Sauer, P., Cootes, T. F., & Taylor, C. J. (2011, August). Accurate Regression Procedures for Active Appearance Models. In *Bmvc* (pp. 1-11).
- [122] Saragih, J., & Göcke, R. (2009). Learning AAM fitting through simulation. *Pattern Recognition*, 42(11), 2628-2636.
- [123] Cristinacce, D., & Cootes, T. F. (2006, September). Feature detection and tracking with constrained local models. In *Bmvc* (Vol. 1, No. 2, p. 3).
- [124] Cristinacce, D., & Cootes, T. (2008). Automatic feature localisation with constrained local models. *Pattern Recognition*, 41(10), 3054-3067.
- [125] Yang, H., & Patras, I. (2013, April). Privileged information-based conditional regression forest for facial feature detection. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (pp. 1-6). IEEE.
- [126] Cao, X., Wei, Y., Wen, F., & Sun, J. (2014). Face alignment by explicit shape regression. *International journal of computer vision*, 107, 177-190.
- [127] Chen, D., Ren, S., Wei, Y., Cao, X., & Sun, J. (2014). Joint cascade face detection and alignment. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13* (pp. 109-122). Springer International Publishing.
- [128] Wu, Y., Wang, Z., & Ji, Q. (2013). Facial feature tracking under varying facial expressions and face poses based on restricted boltzmann machines. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3452-3459).
- [129] Sun, Y., Wang, X., & Tang, X. (2013). Deep convolutional network cascade for facial point detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3476-3483).
- [130] Milborrow, S., & Nicolls, F. (2008). Locating facial features with an extended active shape model. In *Computer Vision—ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part IV 10* (pp. 504-513). Springer Berlin Heidelberg.
- [131] Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2013). 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 397-403).
- [132] McKone, E., Robbins, R. A., He, X., & Barnes, N. (2018). Caricaturing faces to improve identity recognition in low vision simulations: How effective is current-generation automatic assignment of landmark points?. *PloS one*, 13(10), e0204361.

- [133] Zhu, S., Li, C., Loy, C. C., & Tang, X. (2014). Transferring landmark annotations for cross-dataset face alignment. *arXiv preprint arXiv:1409.0602*.
- [134] Smith, B. M., & Zhang, L. (2014). Collaborative facial landmark localization for transferring annotations across datasets. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13* (pp. 78-93). Springer International Publishing.
- [135] Lajevardi, S. M., & Hussain, Z. M. (2012). Automatic facial expression recognition: feature extraction and selection. *Signal, Image and video processing*, 6, 159-169.
- [136] Meena, K., & Suruliandi, A. (2011, June). Local binary patterns and its variants for face recognition. In *2011 International Conference on Recent Trends in Information Technology (ICRTIT)* (pp. 782-786). IEEE.
- [137] Paul, E., & AS, A. B. (2015, November). Mining images for image annotation using SURF detection technique. In *2015 International Conference on Control Communication & Computing India (ICCC)* (pp. 724-728). IEEE.
- [138] Purandare, V., & Talele, K. T. (2014, April). Efficient heterogeneous face recognition using scale invariant feature transform. In *2014 International conference on circuits, systems, communication and information technology applications (CSCITA)* (pp. 305-310). IEEE.
- [139] Yan, C., Lang, C., Wang, T., Du, X., & Zhang, C. (2014). Age estimation based on convolutional neural network. In *Advances in Multimedia Information Processing—PCM 2014: 15th Pacific-Rim Conference on Multimedia, Kuching, Malaysia, December 1-4, 2014, Proceedings 15* (pp. 211-220). Springer International Publishing.
- [140] Kamarainen, J. K. (2012, October). Gabor features in image analysis. In *2012 3rd international conference on image processing theory, tools and applications (IPTA)* (pp. 13-14). IEEE.
- [141] Murala, S., Gonde, A. B., & Maheshwari, R. P. (2009, March). Color and texture features for image indexing and retrieval. In *2009 IEEE International Advance Computing Conference* (pp. 1411-1416). IEEE.
- [142] Kaur, M., & Mohta, A. (2019, November). A review of deep learning with recurrent neural network. In *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 460-465). IEEE.
- [143] Hochreiter, S. (1997). Long Short-term Memory. *Neural Computation MIT-Press*.
- [144] Yi, H., Shiyu, S., Xiusheng, D., & Zhigang, C. (2016, October). A study on deep neural networks framework. In *2016 IEEE advanced information management, communicates, electronic and automation control conference (IMCEC)* (pp. 1519-1522). IEEE.
- [145] Issam, D., Lakhdar, L., & Mohamed, R. (2023). CAN THE COMBINATION OF FACIAL FEATURES ENHANCE THE PERFORMANCE OF FACE RECOGNITION?. *Jordanian Journal of Computers and Information Technology*, 9(4).
- [146] Devries, T., Biswaranjan, K., & Taylor, G. W. (2014). Age estimation and face verification across aging using landmarks. In *Canadian Conference on Computer and Robot Vision, IEEE Montreal* (pp. 98-103).
- [147] Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., ... & Liu, W. (2018). Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5265-5274).

- [148] Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 212-220).
- [149] Kim, M., Jain, A. K., & Liu, X. (2022). Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 18750-18759).
- [150] Guo, M. H., Xu, T. X., Liu, J. J., Liu, Z. N., Jiang, P. T., Mu, T. J., ... & Hu, S. M. (2022). Attention mechanisms in computer vision: A survey. *Computational visual media*, 8(3), 331-368.