

Université Mohamed Khider – Biskra
Faculté des Sciences et de la technologie
Département : Génie Electrique
Ref :.....



جامعة محمد خيضر بسكرة
كلية العلوم و التكنولوجيا
قسم: الهندسة الكهربائية
المرجع:

Thèse présentée en vue de l'obtention
Du diplôme de
Doctorat en sciences : Electronique

Spécialité (Option) : Electronique

Intitulé

Optimisation des Systèmes Multimodaux pour l'Identification dans l'Imagerie

Présentée par :

Elaggoune Hocine

Soutenue publiquement le **22/12/2022**

Devant le jury composé de :

Pr. DEBILOU AbdErazak	Professeur	Président	Université de BISKRA
Pr. BELAHCENE Mebarka	Professeur	Rapporteur	Université de BISKRA
Pr. CHERIF Fella	Professeur	Examinatrice	Université de OUARGLA
Dr. SEMIRA Hichem	Maitre de Conférences 'A'	Examineur	Université de CONSTANTINE

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

وَالْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ

وَالصَّلَاةُ وَالسَّلَامُ عَلَى مَنْ بُعِثَ رَحْمَةً لِّلْعَالَمِينَ وَعَلَى آلِهِ وَصَحْبِهِ أَجْمَعِينَ.

Dédicace -----

Je ne saurais trouver les mots qu'il faut pour exprimer ma gratitude, mon amour, ma reconnaissance... C'est tout simplement que je dédie ce travail : À tous ceux qui m'ont soutenu pour que je puisse arriver à ce stade particulièrement : la mémoire de mes chers parents, ma femme, mes enfants, mes sœurs et leurs enfants, à toute ma famille, à tous mes amis, à tous les enseignants et enseignantes qui ont participé de près ou de loin à ma constitution.

Hocine ELAGGOUNE

Remercîment -----

Louange à ALLAH le tout Puissant, qui m'a permis d'achever cette thèse.

*Je tiens d'abord à remercier ma Directrice de thèse " Pr. **BELAHCENE Mebarka** ", Professeur à l'université Mohamed Khider, Biskra, qui m'a encouragée et m'a poussée à terminer ce travail dans les délais. Je la remercie d'avoir été mon support et soutien durant toutes ces années d'études et de m'avoir dirigé avec patience.*

J'adresse mes sincères remerciements aux président et membres du jury qui m'ont fait l'honneur d'accepter d'évaluer ce travail, chacun par son nom :

Pr. DEBILOU Abderezzak : Professeur à l'université Mohamed Khider, Biskra.

Dr. CHARIF Fella : Maître de Conférence A à l'université Kasdi Merbah, Ouargla.

Dr. SEMIRA Hichem : Maître de Conférence A à l'Université Emir AbdElkader, Constantine.

Je tiens à remercier aussi mes Co-auteurs :

M. BOURENNANE Salah – Professeur à école centrale-Marseille et Directeur du groupe GSM à l'Institut Fresnel Marseille, France.

M. MESSAADIA Mourad - Professeur associé au CESI Engineering School (Rouen-France).

M. BENATIA Mohamed Amin - Professeur associé au CESI Engineering School (Rouen-France).

*Tous mes remerciements et mes respects au Recteur de l'université Mohamed Khider, Biskra le Professeur **M. BOUTARFAIA Ahmed**.*

*Sans oublier le **Dr. OUAMANNE Abdelmalik** : Maître de Conférence A à l'université Mohamed Khider, Biskra pour son aide et son soutien.*

Mes remerciements à ma femme et mes enfants de m'avoir supporté toute ces années, et je demande leur pardon pour le temps que j'ai passé loin d'eux.

Mes remerciements aux travailleurs du domaine de l'éducation et de l'enseignement qui font leur devoir sans attendre aucune récompense sauf d'Allah.

En fin, mes remerciements vont à tous ceux qui m'ont soutenu ou qui, d'une manière ou d'une autre, ont contribué à l'élaboration de ce travail.

Hocine ELAGGOUNE

Publications & communications scientifiques-----

A. Publications internationales

- [1] Elaggoune H, Belahcene M, & Bourennane S (2022) Hybrid descriptor and optimized CNN with transfer learning for face recognition. *Multimedia Tools and Applications*, 81(7), 9403-9427. <https://doi.org/10.1007/s11042-021-11849-1>

B. Communications internationales

- [1] Elaggoune H, Belahcene M, Bourennane S (2020) Hybrid Descriptor Optimization for Face Recognition. International Conference in Optimization and Learning (OLA2020), Cadiz, Spain, 17-19 Feb 2020. https://ola2020.sciencesconf.org/data/pages/Program_OLA_2025.pdf.
- [2] Elaggoune H, Belahcene M, Messaadia M, & Benatia MA (2021, May) Hybrid Descriptor and Patches Optimization for Face Recognition. In *2021 1st International Conference On Cyber Management And Engineering (CyMaEn)* (pp. 1-6). IEEE. <https://doi.org/10.1109/CyMaEn50288.2021.9497271>

ملخص:

تعد الصورة من بين الوسائط الأكثر شعبية، والتي احتلت مكاناً أساسياً لتطوير: أنظمة التعرف على القياسات الحيوية بصفة عامة والتعرف على الوجه بصفة خاصة.

أحد الاستخدامات الأكثر شيوعاً للصور هو تحديد الهوية / التحقق في القياسات الحيوية ، والتي شهدت اهتماماً متزايداً في السنوات الأخيرة. ترتبط فعالية تقنيات تحديد الهوية في التصوير اليوم ارتباطاً وثيقاً بالقبود القوية المفروضة على المستخدم. لذلك يتحول خط البحث الحالي إلى إدارة المواقف التي يكون فيها اكتساب البيانات أقل تقييداً. أخيراً ، غالباً ما يكون استخدام طريقة واحدة محدوداً من حيث الأداء أو صعوبات الاستخدام ، ولهذا السبب يبدو من المثير للاهتمام تقييم مساهمة الطريقة المتعددة في هذا السياق.

فالهدف من الأطروحة هو إجراء بحث موجه نحو تقنيات التحسين مع الاعتماد على الواصفات الهجينة والبقع واشراك تقنيات الاندماج الخاصة بهم من ناحية، ومن ناحية أخرى الاعتماد على التقنيات التعلم العميق (نقل التعلم). نحن مهتمون بشكل خاص بصورة الوجه ويتم التحقق من صحة مناهجنا باستغلال عدد من قواعد البيانات الشاملة المحدثه في بنات مقيدة وغير مقيدة.

الكلمات المفتاح: خصائص الوجه، إندماج ، تعدد البصمات، الواصفات، التحسين، نقل التعلم، التعرف على الوجه.

Abstract :

Among the most popular media that have taken an essential place for the development of biometric recognition systems in general and face recognition systems in particular, we find Image.

One of the most common uses of images is identification/verification in biometrics, which has seen growing interest in recent years. The effectiveness of identification techniques in imaging is today very strongly linked to strong constraints imposed on the user. A current line of research therefore turns to the management of situations where data acquisition is less constrained. Finally, the use of a single modality is often limited in terms of performance or difficulties of use, why it seems interesting to evaluate the contribution of multi-modality in this context.

The objective of the thesis is to carry out a work to pursue a research directed toward the techniques of optimization based on the one hand on the hybrid descriptors and the patches as well as their techniques of fusions, and on the other hand on the Deep Learning (Transfer Learning). We are particularly interested in the image of faces and our approaches are validated on several universal databases to tackle all the hazards of acquisition and uncontrolled environments.

Keywords: *Patches; Fusion; Multimodality; Descriptors; Optimisation, Transfer Learning; Face recognition.*

Résumé :

Parmi les médias les plus populaires qui ont pris une place incontournable pour le développement des systèmes de reconnaissances biométriques en général et les systèmes de la reconnaissance de visage en particulier on trouve l'Image.

L'une des utilisations les plus courantes des images est l'identification/vérification en biométrie qui connaît un intérêt grandissant depuis quelques années. L'efficacité des techniques d'identification en imagerie est aujourd'hui très fortement liée à des contraintes fortes imposées à l'utilisateur. Une voie de recherche actuelle se tourne donc vers la gestion de situations où l'acquisition des données est moins contrainte. Finalement, l'usage d'une seule modalité est souvent limité en termes de performance ou de difficultés d'usage, c'est pourquoi il apparaît intéressant d'évaluer l'apport de la multi-modalité dans ce contexte.

L'objectif de la thèse est de mener un travail pour poursuivre une recherche tournée à la fois vers les techniques d'optimisation basées d'une part sur les descripteurs hybrides et les patches ainsi que leurs techniques de fusions, et d'autre part sur le Deep Learning (*Transfer Learning*). Nous nous intéressons plus particulièrement à l'image du visage et nos approches sont validées sur plusieurs bases de données universelles pour défier tous les aléas d'acquisition et d'environnements non contrôlés.

Mots clés : *Patches ; Fusion ; Multimodalité ; Descripteurs ; Optimisation, Transfer Learning ; Reconnaissance de visage.*

TABLE DES MATIERES

<i>LISTE DES FIGURES</i>	IX
<i>LISTE DES TABLEAUX</i>	XI
<i>ACRONYMES</i>	XIII
1. Introduction générale	1
1.1 Contexte	1
1.2 Problématique	2
1.3 Positionnement	3
1.4 Objectifs	3
1.5 Contributions	4
1.6 Matériel et logiciels mise en œuvre pour la réalisation de nos expériences	5
1.7 Organisation de la thèse	5
2. Les systèmes de reconnaissances biométriques et fusion multimodal	6
2.1 Introduction	6
2.2 La biométrie	6
2.3 Structure d'un système biométrique	8
2.4 Fonctionnalités biométriques	9
<i>1) Vérification (Authentication)</i>	9
<i>2) Identification</i>	10
<i>3) Surveillance</i>	10
2.5 Systèmes de reconnaissances biométriques multimodale	11
2.5.1 Types de fusion [7]	12
<i>1) Systèmes multi algorithmes :</i>	12
<i>2) Systèmes multi échantillons :</i>	12
<i>3) Systèmes multi capteurs :</i>	12
<i>4) Systèmes multi instances :</i>	13

5) <i>Systèmes multi caractères</i> :	13
2.5.2 Les approches de fusion [7]	13
2.5.3 Normalisation de scores [7]	14
1) <i>Normalisation par la méthode Min-Max</i> :	15
2) <i>Normalisation par la méthode Z-Score</i> :	15
3) <i>Normalisation par la médiane et l'écart absolu médian (MAD) [8]:</i>	15
4) <i>Normalisation par la méthode tangente hyperbolique "Tanh"[8] :</i>	16
2.6 Evaluation des performances d'un système biométrique « SB »	16
2.7 Les méthodes de la reconnaissance du visage	20
2.7.1 Méthodes globales	21
2.7.2 Méthodes locales	22
2.7.3 Méthodes hybrides	23
2.7.4 Apprentissage profond (Deep Learning)	23
2.8 Conclusion	27
3. État de L'art	28
3.1 Introduction	28
3.2 Les descripteurs de visage	29
03.3 Les méthodes d'optimisation de la RV	30
3.4 Apprentissage profond (Deep Learning) pour la RV	31
3.4.1 Aperçu sur les méthodes récentes DL pour la RV	33
1) <i>Augmentation de données (Data Augmentation)</i>	34
2) <i>Extraction des caractéristiques profond (Deep Feature Extraction)</i>	34
3) <i>Perte d'entraînement (Training Loss) pour l'appariement du visage</i>	37
3.4.2 Bases de données visage utilisées dans les modèles DL	39

3.5 Conclusion	41
4. Optimisation des descripteurs hybrides et Patchs pour la RV	43
4.1 Introduction	43
4.2 Méthodologie	45
4.2.1 Extraction des caractéristiques	45
1) <i>Filtre de Gabor</i>	45
2) <i>Histogramme des Gradients orientés</i>	46
3) <i>Optimisation à Essaim de Particules</i>	48
4) <i>Quantification de Phase Locale</i>	51
5) <i>Analyse en composantes principales</i>	52
4.2.2 Apprentissage et Modélisation	52
1) <i>Analyse Discriminante Linéaire</i>	52
2) <i>Apprentissage du modèle LDA</i>	53
3) <i>Test du modèle LDA</i>	53
4.3 Bases de Données	54
4.3.1 Base de données de visages Yale étendue B	54
4.3.2 Base de données LFW “Labeled Faces in the Wild”	54
4.4 Résultats des descripteurs hybrides optimisés proposés	55
4.4.1 Résultats sans fusion multimodales	55
4.4.2 Résultats avec fusion multimodales	56
4.4.3 Comparaison de nos résultats avec ceux existants en littérature	58
4.5 Conclusion	61
5. Optimisation des caractéristiques de visages en utilisation des descripteurs hybrides et CNN en apprentissage par transfert	62
5.1 Introduction	62
5.2 Approches proposées	65
5.2.1 Descripteurs Hybrides Optimisés	65

5.2.2 Apprentissage par Transfert Optimisée	67
<i>1) Réseau de neurones convolutionnels</i>	68
<i>2) Apprentissage par transfert</i>	68
5.3 Bases de données	69
5.3.1 Base de données ORL	69
5.3.2 Base de données AR	70
5.3.3 Base de données LFW	71
5.3.4 Base de données "Thermal Tufts Face (TTF)"	71
5.4 Résultats expérimentaux et discussion	71
5.4.1 Descripteurs Hybrides Optimisés	75
<i>1) Comparaison avec l'état de l'art</i>	76
5.4.2 Apprentissage par Transfert Optimisée	86
5.4.3 Comparaison des résultats des deux approches	87
5.5 Conclusion	87
6. Conclusion et perspectives	88
BIBLIOGRAPHIE	91

LISTE DES FIGURES

Fig 2. 1. Différentes modalités biométriques.	7
Fig 2. 2. Structure générale d'un système biométrique.	8
Fig 2. 3. Modèle de vérification [3].	9
Fig 2. 4. Modèle d'Identification [3].	10
Fig 2. 5. Détermination de l'image de la requête sur la liste restreint d'individus.	10
Fig 2. 6. Différentes possibilités de la fusion d'informations [6].	11
Fig 2. 7. Sources de différents types de fusion de traits biométriques.	12
Fig 2. 8. Schéma de la fusion de scores [7].	13
Fig 2. 9. Matrice de confusion d'hypothèses de décision.	17
Fig 2. 10. Illustration de FAR, FRR et de EER en fonction du seuil de décision [9], [10].	18
Fig 2. 11. a) Courbe ROC ; b) Courbe DET [10][11].	19
Fig 2. 12. Courbe CMC.	19
Fig 2. 13. Classification des algorithmes principaux utilisés en reconnaissance faciale [8].	20
Fig 2. 14. Réseau de neurones d'apprentissage profond "Deep Learning Neural Network" [23].	24
Fig 3. 1. Évolution de l'architecture	35
Fig 3. 2. Architectures [64]: a) AlexNet, b) VGG-Net, c) GoogleNet, d) ResNet, e) SeNet.	36
Fig 4. 1. Schéma de descripteurs hybrides proposé et de l'optimisation de patchs pour le système RV.	45
Fig 4. 2. HOG extraction des caractéristiques [111].	47

Fig 4. 3. Exemples d'images de la base de données Yale B et leur localisation des yeux et de la bouche	54
Fig 4. 4. (a) Image du visage (Yale B), (b) Histogramme du visage, (c) visage après égalisation,	54
Fig 4. 5. Exemples d'images de la base de données LFW avec leurs localisation des yeux et de la bouche.	55
Fig 4. 6. Exemple d'alignement en sélectionnant cinq points du visage [121].	55
Fig 4. 7. Courbes CMC avec et sans l'algorithme PSO de YALE.	59
Fig 4. 8. Courbes ROC avec et sans l'algorithme PSO de LFW.	60
Fig 5. 1. Schéma de principe de systèmes hybrides proposé d'identification/vérification optimisé.	66
Fig 5. 2. Schéma de principe de l'optimisation combinée avec AlexNet et Transfer Learning.	67
Fig 5. 3. Exemples d'images des bases de données : (a) ORL, (b) AR, (c) LFW et (d) TTF.	71
Fig 5. 4. Courbes CMC avec et sans l'algorithme PSO de ORL, AR & TTF.	73
Fig 5. 5. Courbes ROC avec et sans l'algorithme PSO de LFW.	73
Fig 5. 6. TR avec variation du taux de test par AlexNet_TL sans PSO de ORL, AR & TTF.	74
Fig 5. 7. Performances du TR sur la base de données AR, ORL, LFW, et TTF.	81
Fig 5. 8. Temps de traitement consommé par PSO_AlexNet_TL.	82
Fig 5. 9. Comparaison Temps de test entre l'approche Alexnet_TL_PSO (750 Caractères, 20 itérations) et Alexnet_TL.	86

LISTE DES TABLEAUX

Tableau 2. 1	Paramètres d'AlexNet	25
Tableau 3. 1	Résultats de quelques travaux liés aux descripteurs RV	30
Tableau 3. 2	Résultats de quelques travaux liés aux méthodes d'optimisation de RV	31
Tableau 3. 3	Résultats de quelques travaux liés aux méthodes DL avec optimisation	41
Tableau 4. 1	TR et Temps Test pour la base de données YALE-B	56
Tableau 4. 2	TR et Temps Test pour la base de données LFW	56
Tableau 4. 3	TR et Temps Test obtenu en Fusion des modalités Yeux & Bouche (BD : YALE-B)	57
Tableau 4. 4	TR et Temps Test obtenu en Fusion des modalités Yeux & Bouche (BD : LFW)	57
Tableau 4. 5	TR de visages avec YALE-B, comparaison avec d'autres méthodes existantes dans la littérature.	58
Tableau 4. 6	TR de visages (%) Rang 1 avec LFW, comparaison avec d'autres méthodes existantes dans la littérature	58
Tableau 5. 1	Rang 1- Taux de reconnaissance (TR) pour les bases de données ORL, AR et TTF	72
Tableau 5. 2	Comparaison de nos meilleurs résultats du TR avec d'autres méthodes existantes dans la littérature obtenus pour La base de données ORL.	75
Tableau 5. 3	Comparaison de nos meilleurs résultats du TR avec d'autres méthodes existantes dans la littérature obtenus pour La base de données AR.	75
Tableau 5. 4	Taux de reconnaissance de visages, Rang 1 avec LFW et comparaison avec d'autres méthodes existantes dans la littérature	76
Tableau 5. 5	Taux de reconnaissance pour le modèle CNN avec TTF : comparaison avec d'autres méthodes existantes dans la littérature.	77

Tableau 5. 6	Taux de reconnaissance pour le modèle CNN avec ORL : comparaison avec d'autres méthodes existantes dans la littérature.	77
Tableau 5. 7	Taux de reconnaissance pour le modèle CNN avec AR : comparaison avec d'autres méthodes existantes dans la littérature.	77
Tableau 5. 8	Taux de reconnaissance pour le modèle CNN avec LFW : comparaison avec d'autres méthodes existantes dans la littérature.	78
Tableau 5. 9	Variation de TR (%) en fonction des Caractéristiques et des Itérations.	84
a)	ORL	84
b)	AR	84
c)	LFW	85
d)	Thermal Tufts Face	85
Tableau 5. 10	Comparaison du TR et du temps de test des deux approches proposées pour ORL, AR, LFW, et TTF.	86

ACRONYMES

ADN	Acide Désoxyribose Nucléique
ANN	Artificial Neural Network
BD	Data Base
BSIF	Binarized Statistical Image Features
CAO	Conception Assistée par Ordinateur
CD-EJSR	Customized Dictionary- Extended Joint Sparse Representation
CL	Contextual Loss
CMC	Cumulative Match Characteristic
CMCGSO	Cloud Model Chaos Glowworm Swarm Optimization
CMS	Cumulative Match Score
CNN	Convolutional Neural Networks
CRN	Cascaded Refinement Network
CS	Cuckoo Search
CT	Contourlet Transform
DBN	Deep Belief Network
DCA	Discriminant Correlation Analysis
DCT	Discrete Cosine Transform
DET	Detection Error Tradeoff
DFT	Discrete Fourier Transform
DLSR	Discriminative Locality-constrained Sparse Representation
DWT	Discrete Wavelet Transform
DWT E-CLAHE	Discrete Wavelet Transform Enhanced Contrast Limited Adaptive Histogram Equalization
ECG	Electrocardiography
EER	Equal Error Rate
ESPCN	Efficient Sub-Pixel Convolutional Neural network

FAR	False Accept Rate
FKP	Finger-Knuckle-Print
FRCNN	Fast Region-based Convolutional Network
FRS	Face Recognition System
FRR	False Reject Rate
FSRC	Fast Sparse Representation-based
GA	Genetic Algorithm
GAR	Genuine Accept Rate
GFC	Gabor Fisher Classifier
GPU	Graphics Processing Unit
GSR-FR	Group Sparse Representation-Face Recognition
HMM	Hidden Model Markov
HOG	Histogram of Oriented Gradient
HTER	Half Total Error Rate
HVS	Human Vision System
ICA	Independent Component Analysis
IKLDA	Improved Kernel Linear Discriminate Analysis
KLDA	Kernel Linear Discriminate Analysis
K-NN	K Nearest Neighbours
KNN:	K-nearest neighbors
KPCA	Kernel Principal Component Analysis
KSVD-IJSR	K Singular Value Decomposition-Improved Joint Sparse Representation
LARK	Locally Adaptive Regression Kernel
LBP	Local Binary Patterns
LBP-HF	Local Binary Pattern Histogram Fourier Features
LDA	Linear Discriminate Analysis
LPQ	Local Phase Quantization
LTP	Local Ternary Pattern

MMD	Mean of Minimum Distances
MPSO	Modified Particle Swarm Optimization
Multi-AP-GAN	Multi Attribute Preserved Generative Adversarial Network
PAO	Publication Assistée par Ordinateur
PCA	Principal Component Analysis
PCASVD	Principal Component Analysis Singular Value Decomposition
PHOG	Pyramid of. Histograms of Orientation Gradients
PNN	Probabilistic Neural Network
PSO	Particle Swarm Optimization
QO	Quadratic Optimisation
ReLU	Rectified Linear Unit
RNN	Rough Neural Network
ROC	Receiver Operating Characteristics
RV	Reconnaissance de Visages
SFD-CR	Space Frequency Domain-Chirp Rate
SRC-EnNPSO	Sparse Representation-based pattern Classification - Elman Neural Network weights are optimized with PSO
SVM	Support Vector Machine
TER	Total Error Rate
TL	Transfer Learning
TPLBP	Three patch Local Binary Patterns
TTF	Thermal Tufts Face
TR	Taux de Reconnaissance
VLC	VanderLugt Correlator
WEB	World Wide Web

1. Introduction générale

1.1. Contexte

Au fil des jours, le besoin de développer des techniques d'acquisition, d'analyse et de représentation des données devient plus qu'une nécessité. Parmi les médias les plus populaires utilisés par ces techniques, on trouve l'image, le son, et la vidéo, qui ont vu leur importance s'accroître au fil du temps. L'usage de ces nouveaux médias en tant qu'information est devenu une nécessité dans plusieurs domaines : la médecine, la géographie, la Conception Assistée par Ordinateur (CAO) en mécanique, la Publication Assistée par Ordinateur (PAO), l'histoire, l'art, l'archéologie, la bureautique, l'éducation, le domaine de l'information et de l'audiovisuel, et la surveillance [1] etc.

La biométrie est la science qui porte sur l'analyse des caractéristiques biologiques, morphologiques ou comportementales strictement propres à chaque individu et permettant la reconnaissance et l'identification de son identité. Elle peut également être définie comme un ensemble de caractéristiques mesurables, robustes, distinctives et unique à chaque individu.

Nous pouvons distinguer quatre grands types d'applications de la biométrie : le contrôle d'accès (Access control), l'authentification des transactions (Transaction authentication), la répression (Law-enforcement) et la personnalisation des appareils utilisés au quotidien (Personalization).

La reconnaissance de visages s'est avérée être l'une des technologies les plus réussies et a eu un impact sur des domaines hétérogènes. Dans cette thèse, certains systèmes de reconnaissance de visage les plus efficaces ont été étudiés. Premièrement, on donne un aperçu d'un système général de reconnaissance de visage. Deuxièmement, l'étude couvre diverses architectures de réseau et des pertes de formation qui ont eu un impact substantiel. Finalement, toutes les expériences sont validées sur diverses bases de données publiques pour évaluer les capacités du système de reconnaissance de visages (FRS).

Dans ce travail, nous nous intéressons plus particulièrement à l'image et son rôle incontournable dans les systèmes de reconnaissances biométriques (en particulier la reconnaissance de visage). L'évolution réelle de l'image en matière de création, de stockage et de manipulation en informatique, date de 1963 avec le projet sketchpad d'Ivan Sutherland [2]. Le début des années 90 ont connues la création du WEB qui a fait un saut quantique dans le monde entier en offrant la possibilité à toute la planète d'avoir un accès varié à un ensemble de média (texte, image, vidéo, son, etc.), et a également contribué à l'exploitation des images.

1.2. Problématique

Bien que les êtres humains puissent détecter et identifier des visages dans une scène sans beaucoup de peine, construire un système automatique qui accomplit de telles tâches représente un sérieux défi. Ce défi est d'autant plus grand lorsque les conditions d'acquisition des images sont très variables. Il existe deux types de variations associées aux images de visages : inter et intra sujet. La variation inter-sujets est limitée à cause de la ressemblance physique entre les individus. Par contre la variation intra-sujet est plus vaste. Elle peut être attribuée à plusieurs facteurs : changement d'éclairage, expressions, jumeaux, poses, présence ou absence des composants structurels (lunettes, moustache...).

Parmi les utilisations les plus courantes des images est l'identification en biométrie qui est un domaine de la vision artificielle et qui a connu un intérêt grandissant depuis quelques années. Plusieurs approches intéressantes ont été développées dans le spectre visible, multi spectral et hyper spectral. Les systèmes opérant dans ces spectres souffrent de plusieurs limitations dues aux changements dans l'éclairage, le mouvement et dans les expressions du visage. Des images multidimensionnelles et des stratégies de fusion intra et inter-spectres opérant dans l'espace des textures et utilisant des approches de réduction de dimensionnalités linéaires et non linéaires sont des pistes à explorer et à développer.

L'efficacité des techniques d'identification en imagerie est aujourd'hui très fortement liée à des contraintes fortes imposées à l'utilisateur, une voie de recherche actuelle se tourne donc vers la gestion de situations où l'acquisition des données est moins contrainte. Finalement, l'usage d'une seule modalité est souvent limité en termes de performance ou de

difficultés d'usage, c'est pourquoi il apparaît intéressant d'évaluer l'apport de la multi modalité dans ce contexte.

1.3. Positionnement

Notre étude, menée au Laboratoire d'Identification, Commande, Contrôle et Communication (LI3C) de l'université Mohamed KHIDER BISKRA, s'inscrit dans une des problématiques étudiées par l'équipe RB_IAIM « *Reconnaissance Biométrique & Identification d'Anomalie sur Imagerie Médicale* ». Ce travail contribue aux travaux menés au sens plus large par l'équipe sur des recherches liées à deux axes en parallèles :

- a. Identification/Authentification et reconnaissance de la parenté biométrique ;
- b. Identification d'anomalies dans les images médicales.

1.4. Objectifs

L'objectif principal de la thèse est de contribuer à un travail pour poursuivre une recherche tournée à la fois vers la technique et les usages.

L'objectif global des travaux est donc de permettre une meilleure compréhension et amélioration des performances des systèmes multimodaux pour l'identification, et par la suite une optimisation sur l'aspect spectrale et spatiale ainsi qu'un apprentissage automatique des images rendant notre approche plus souple, rapide et efficace.

L'étude et la recherche relative à cette thèse porte sur trois axes, en vue de la réalisation d'un système de reconnaissance robuste, efficace et performant :

- Une première partie concerne l'étude et la conception des descripteurs locaux et globaux récents ainsi que les descripteurs hybrides de pointe ;
- La deuxième partie concerne l'étude des méthodes d'optimisation et la mise en œuvre de la technique retenue. Le long de ce travail la multi modalité est prise en compte ;

- D'autre part, l'apprentissage en profondeur s'est avéré être le plus efficace dans les tâches de vision informatique en raison de son architecture basée sur la convolution. Depuis l'avènement de l'apprentissage profond, les technologies de reconnaissance de visage n'ont cessé d'accroître et ont eues une augmentation substantielle de précision. Ce qui a motivé notre recherche à optimiser un système de reconnaissance de visage basé sur le Deep Learning.

La validation des résultats est menée sur diverses bases de données.

1.5. Contributions

Dans ce travail de thèse, la première contribution est consacrée pour l'élaboration de systèmes d'identification biométrique multimodal optimisé, basé essentiellement sur la modalité du visage et ses patches.

Pourquoi le choix du visage ? pour la simple raison, qu'elle est naturelle, non intrusive et moins coûteuse.

Nos apports dans cette contribution sont d'abord la recherche du meilleur descripteur hybride qui caractérise le visage. Puis la recherche de la meilleure modalité caractérisant au mieux le visage, d'où la proposition de l'approche repose sur l'élaboration de plusieurs systèmes hybrides et multimodaux de reconnaissance visages, et leurs applications sur plusieurs bases de données.

La deuxième contribution est élue pour le développement d'une structure d'optimisation de fonctionnalités basée sur une structure de Deep Learning et Transfer Learning passant par l'algorithme d'optimisation pour la reconnaissance de visages. Pour cela, des structures basées sur l'apprentissage par transfert « Transfer Learning » sont utilisées en s'appuyant sur les réseaux de neurones convolutifs (Convolutional Neural Network (CNN)) et ce afin d'extraire les caractéristiques du visage les plus pertinentes. Dans cette contribution la validation des résultats est menée sur diverses bases de données en milieux contrôlés, non contrôlés et thermiques.

1.6. Matériel et logiciels mise en œuvre pour la réalisation de nos expériences

Nos travaux ont été exécuté avec un ordinateur de Processeur : Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz 2.90 GHz, avec RAM de 8,00 Go. Le logiciel de programmation utilisé est le MATLAB R2017a.

1.7. Organisation de la thèse

Le manuscrit de thèse s'articule autour de cinq chapitres : après avoir introduit la reconnaissance de visage et les techniques utilisées, le chapitre 2 est consacré à la présentation des systèmes de reconnaissances biométriques et la fusion multimodale. Le chapitre 3 est dédié à l'état de l'art de la recherche scientifique dans le domaine où les méthodes et techniques sur la caractérisation paramétrique de l'image, l'optimisation des caractéristiques ainsi que l'utilisation de l'apprentissage automatique et la classification sont tous mis en revue. Le chapitre 4 présente la conception et l'implémentation de la première approche proposée basée sur la multi modalité, l'optimisation et tous les travaux menés sur la modalité visage et les différentes modalités patches. Le dernier chapitre 5 : la recherche est poussée vers l'apprentissage profond et l'optimisation des caractéristiques extraites d'un modèle CNN (Convolutional Neural Networks) utilisant l'apprentissage par transfert (Transfer learning). Dans tous les chapitres les résultats sont discutés et améliorés le long du travail pour optimiser les taux de performances et aussi optimiser le temps de test. Enfin, une conclusion générale et des perspectives clôturent la thèse.

2. Les systèmes de reconnaissances biométriques et fusion multimodale

2.1. Introduction

Un système de reconnaissance biométrique est un type de système informatique qui vise à identifier, authentifier un individu sur la base d'une ou plusieurs caractéristiques biologiques et / ou comportementales (biométrie) en les comparant avec des données précédemment acquises (enregistrées dans la base de données du système). Ces données sont acquises et enregistrées à travers divers outils électroniques innovants : Caractéristiques du visage, couleur et la taille de l'iris, empreintes digitales, empreinte vocale, etc.

L'évolution des techniques logicielles et matérielles informatiques ainsi que l'apparition des nouvelles sources d'images numériques de haute définition, relancent depuis quelques années ce domaine qui est en plein extension.

Ce chapitre porte sur une vision générale sur les systèmes de reconnaissances biométrique à Mono-modalité et Multi-modalité. Également, quelques concepts et lexiques très utilisés dans le système biométrique ont été discutés.

2.2. La biométrie

La biométrie est la science qui porte sur l'analyse des caractéristiques biologiques, morphologiques ou comportementales strictement propres à chaque individu et permettant la reconnaissance et l'identification de son identité. Elle peut également être définie comme un ensemble de caractéristiques mesurables, robustes, distinctives et uniques à chaque individu (Fig 2. 1).

- **Biométries morphologiques (physiologiques)**

Elles sont basées sur l'analyse des caractéristiques physiques particulières, qui sont permanents et uniques pour chaque individu. On peut citer (l'iris de l'oeil,

l’empreintes des articulations des doigts (FKP), les empreinte digitales, le visage, les empreintes palmaires, etc.).

- **Biométries comportementales**

Elles reposent sur l’analyse de comportements d’un individu. On peut citer (la démarche, dynamique de frappe au clavier, dynamique de la signature, inclinaison et vitesse de déplacement du stylo, l’empreinte de sa voix, sa démarche et sa façon de taper sur un clavier, etc.). Mais ces biométries dépendent généralement de l’état physique ou moral de l’individu, ce qui peut altérer l’information fournie par ces modalités pour la reconnaissance.

- **Biométries biologiques**

Elles se basent sur l’analyse des données biologiques liées à l’individu (Salive, ADN, l’odeur corporelle, la forme de l’oreille, la forme des lèvres, veines de la main, etc.). Ce type de biométrie est très complexe à mettre en œuvre dans un système usuel de reconnaissance et n’est utilisé que dans un cas d’extrême de nécessité (Exemples : Enquête criminelle, test de paternité...etc.).

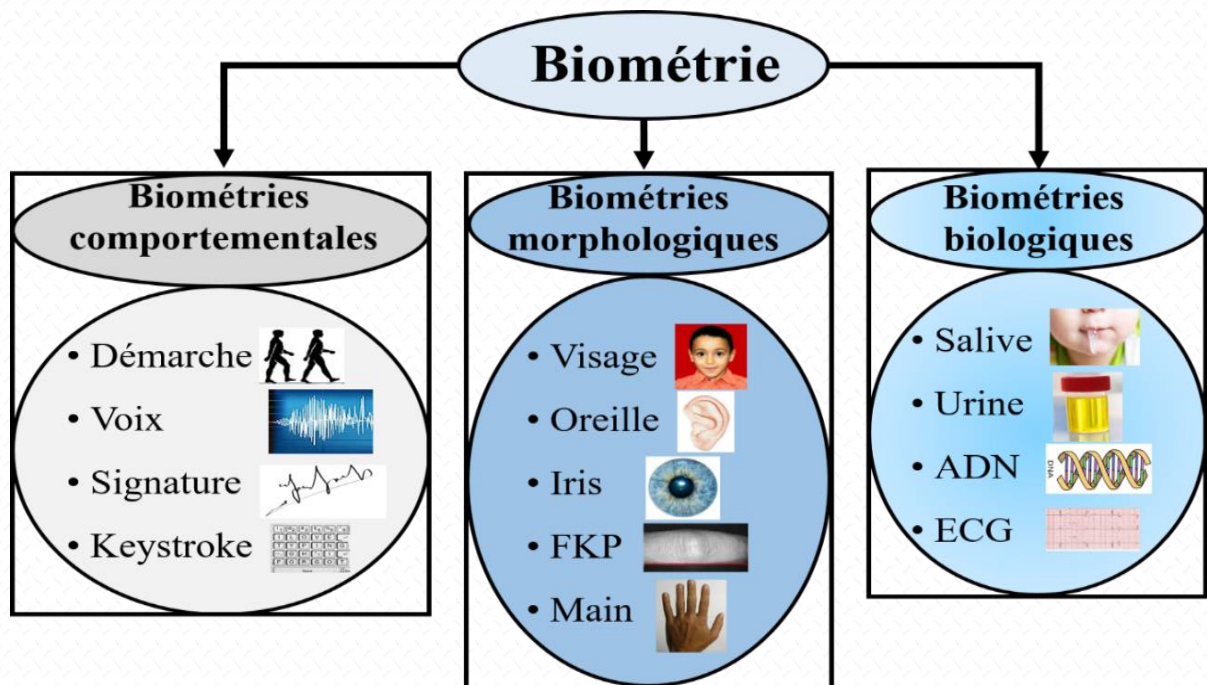


Fig 2. 1. Différentes modalités biométriques.

2.3. Structure d'un système biométrique

Un système biométrique doit avoir au moins deux modules, l'un pour l'apprentissage (on l'appelle aussi enrôlement) et l'autre pour la reconnaissance. Aussi, un autre facultatif pour l'adaptation.

Dans un système biométrique classique, on procède en premier lieu par l'acquisition des données biométriques de l'objet à reconnaître. Une étape d'extraction de ses paramètres est indispensable, ces paramètres doivent ensuite être comparées aux modèles de bases de données (Voir Fig 2. 2).

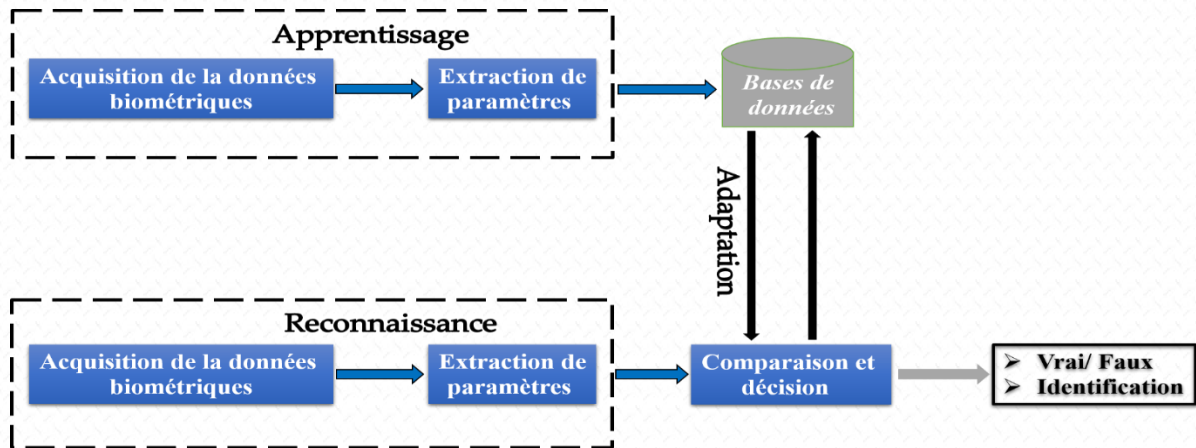


Fig 2. 2. Structure générale d'un système biométrique.

Les trois modules essentiels de tout système biométrique sont définis ci-dessous :

Module d'apprentissage (Enrôlement) : Au cours de ce module, il est question de l'acquisition ou la capture de la caractéristique. Cette capture n'est stockée dans la base de données qu'après des certaines transformations lui appliquées.

En effet, le signal contient de l'information inutile à la reconnaissance et seuls les paramètres pertinents sont extraits. Le modèle ou gabarit est une représentation compacte du signal qui permet de faciliter la phase de reconnaissance, mais aussi de diminuer la quantité de données à stocker. Il est à noter que la qualité du capteur peut grandement influencer les performances du système. Meilleure est la qualité du système d'acquisition, moins il y aura de prétraitements à effectuer pour extraire les paramètres du signal.

- **Module de reconnaissance :** Au cours de la reconnaissance, la caractéristique biométrique est mesurée et un ensemble de paramètres est extrait comme lors de l'apprentissage. Le capteur utilisé doit avoir des propriétés aussi proches que possibles du capteur utilisé durant la phase d'apprentissage. Si les deux capteurs ont des propriétés trop différentes, il faudra en général appliquer une série de prétraitements supplémentaires pour limiter la dégradation des performances.
 - **Module d'adaptation :** Pendant la phase d'apprentissage, le système biométrique ne capture souvent que quelques instances d'un même attribut afin de limiter la gêne pour l'utilisateur. Il est donc difficile de construire un modèle assez général capable de décrire toutes les variations possibles de cet attribut. De plus, les caractéristiques de cette biométrie ainsi que ses conditions d'acquisition peuvent varier. L'adaptation est donc nécessaire pour maintenir ou améliorer la performance d'un système utilisation après utilisation. Elle est quasi indispensable pour les caractéristiques non permanentes comme la voix.

2.4. Fonctionnalités biométriques

Un système de reconnaissance a trois types de tâches à remplir [3]:

1) Vérification (Authentication)

Appelée aussi One-to-One (1 :1), consiste à vérifier que l'identité prétendue est bien la bonne (Fig 2. 3).

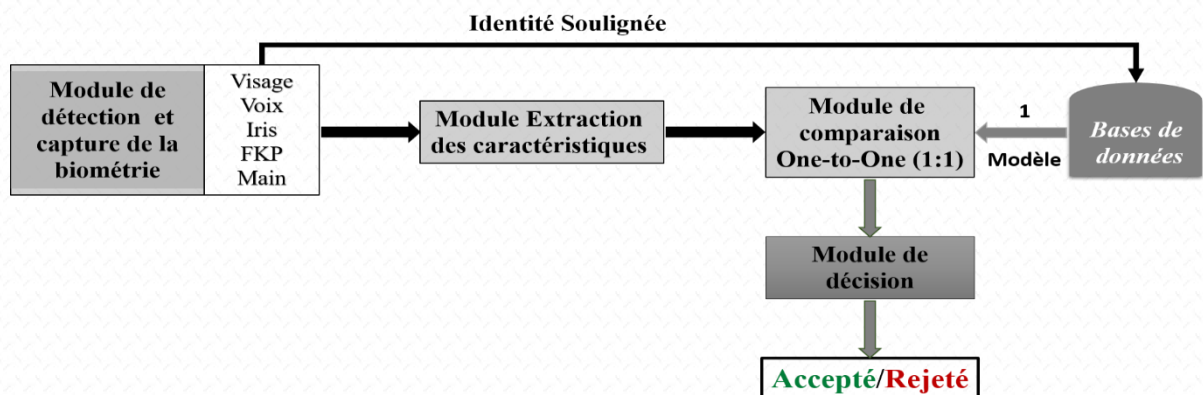


Fig 2. 3. Modèle de vérification [3].

2) Identification

Appelée aussi One-to-Many en anglais (1 : N), consiste à déterminer l'identité d'un individu parmi N identités connues, présentes dans une base (Fig 2. 4).

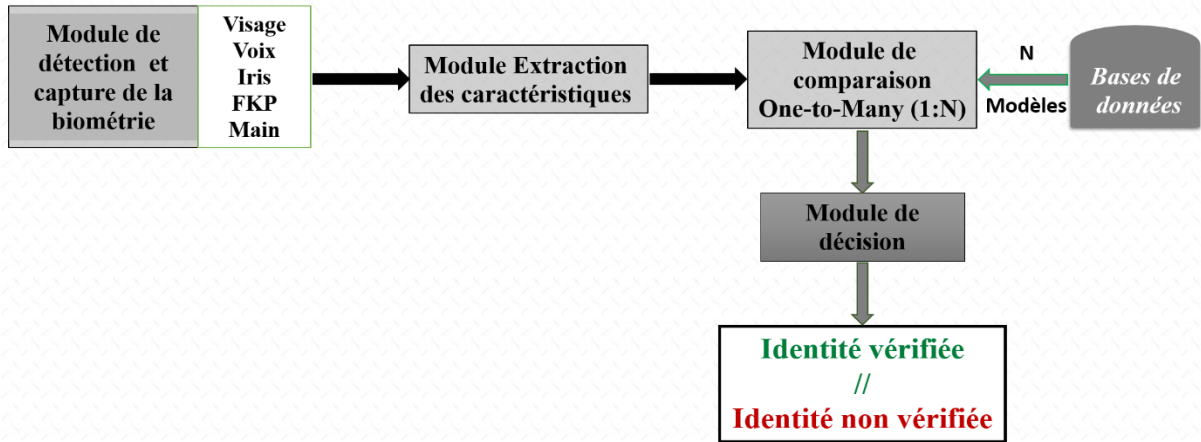


Fig 2. 4. Modèle d'Identification [3].

3) Surveillance

Consiste à déterminer de façon discrète si un échantillon biométrique correspond à un individu sur une liste restreinte d'individus recherchés (Fig 2. 5).

On peut dire que la surveillance est une application des applications de l'identification.

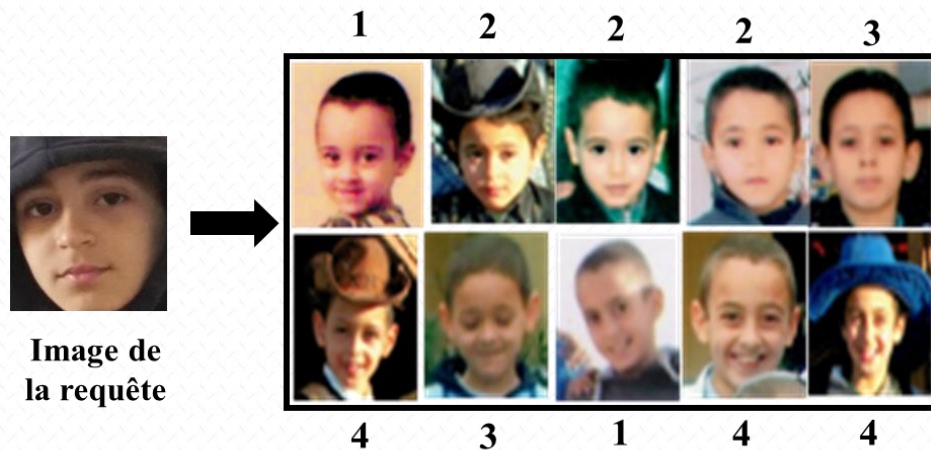


Fig 2. 5. Détermination de l'image de la requête sur la liste restreint d'individus.

2.5. Systèmes de reconnaissances biométriques multimodale

En vue du développement technologique terrible et accéléré et des techniques innovantes de falsification des données biométrique, les systèmes de reconnaissances biométriques multimodales sont devenus une nécessité pour surmonter certaines limites imposées par les systèmes uni-modaux. L'utilisation de différentes modalités dans un système biométrique permet de traiter les problèmes imposés par les système uni-modaux, comme le bruit introduit par le capteur, manque d'individualité due à des facteurs génétiques (deux frères ou père et frère), l'universalité de modalités, manque de représentation invariante, et sensibilité aux attaques. Ainsi, à cause de tous ces problèmes pratiques, le taux de fausse acceptation (FAR) d'un système biométrique augmente, ce qui rend ce type de systèmes uni-modaux inacceptables pour un déploiement d'applications critiques de sécurité [4]-[5].

Un système biométrique multimodal repose sur la fusion de l'information, qu'on peut effectuer à quatre niveaux différents : au niveau du capteur, au niveau de l'extraction de caractéristiques, au niveau des scores issus du module de comparaison et au niveau de la décision (voir Fig 2. 6).

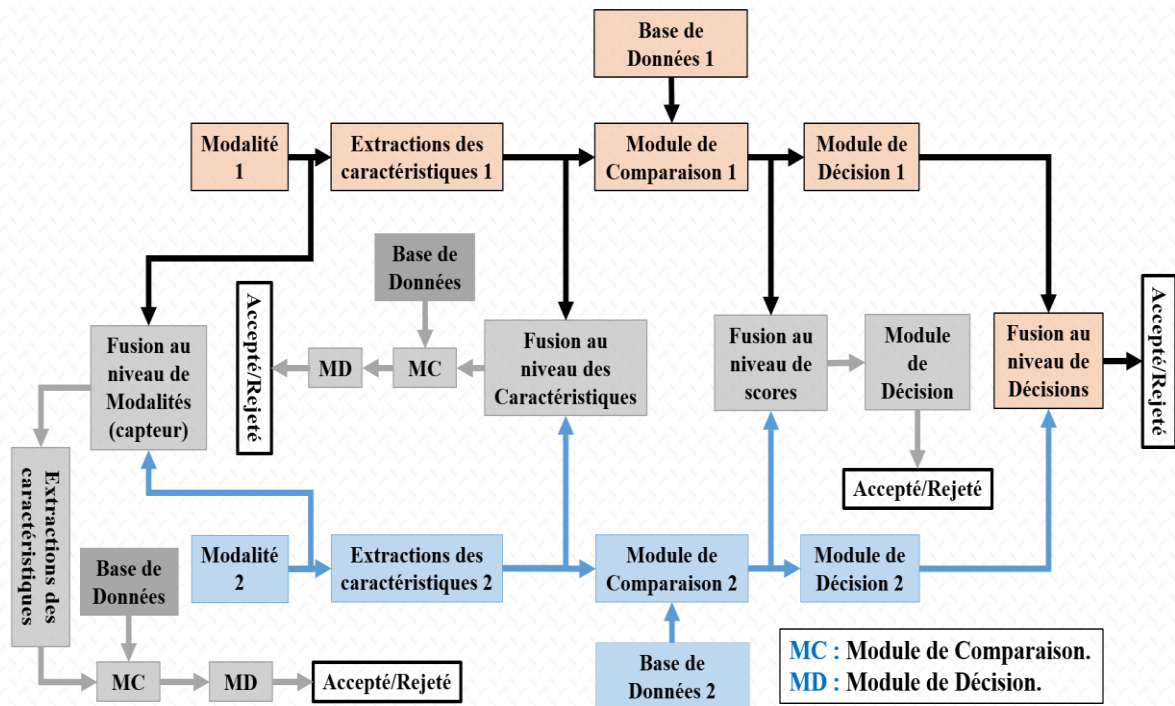


Fig 2. 6. Différentes possibilités de la fusion d'informations [6].

2.5.1. Types de fusion [7]

La fusion d'éléments biométriques peut se référer à de nombreux scénarios différents (Fig 2. 7).

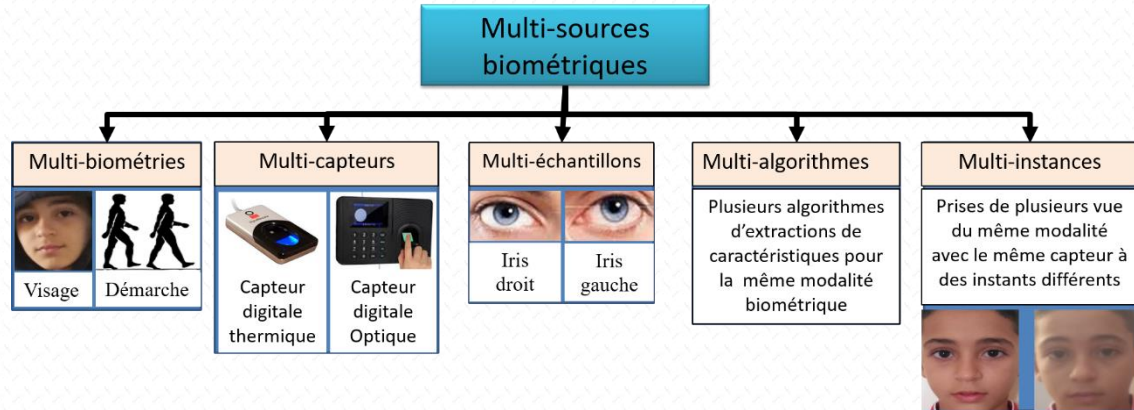


Fig 2. 7. Sources de différents types de fusion de traits biométriques.

1) *Systemes multi algorithmes :*

Il est utilisé par de nombreuses approches. L'extraction des caractéristiques est élaborée via différents algorithmes puis fusionnées. La fusion de caractéristiques extraites via un algorithme analysant les textures et un autre la forme d'un caractère biométrique entre dans ce cadre.

2) *Systemes multi échantillons :*

Concerne la capture de plusieurs instances d'une même modalité biométrique par un capteur unique, et cela dans le but de rendre plus robuste l'extraction des caractéristiques ou d'enrichir le modèle biométrique d'une personne. Par exemple : captures de visage d'une personne sous différents angles en plusieurs reprise. L'utilisation de vidéos entre également dans ce cadre.

3) *Systemes multi capteurs :*

Elle est caractérisée par l'utilisation de plusieurs capteurs utiliser pour la capture d'une modalité biométrique unique sous différents angles. Ce type de système permet notamment la fusion au niveau capteur.

4) Systèmes multi instances :

Ce type de système permet de capturer plusieurs instances du même caractère biométrique. L'exemple typique pour ce type de système est l'acquisition de plusieurs empreintes digitales via un capteur unique. L'avantage de ces systèmes, c'est qu'ils n'entraînent pas de surcoût de capteurs, ni de développer de nouveaux algorithmes. À ne pas confondre avec les systèmes multi échantillons.

5) Systèmes multi caractères :

Ce type de système combine différentes modalités biométriques d'un individu. Ces systèmes nécessitent différents capteurs ainsi que des algorithmes dédiés à chaque modalité biométrique. Ce qui caractérise ce type de système, c'est que les modalités biométriques considérées peuvent être plus décorrélés que pour les systèmes multi capteurs.

2.5.2. Les approches de fusion [7]

Un système de fusion est constitué de deux modules, un module de fusion et un module de décision (voir Fig 2. 8). Le problème devient donc un problème de classification à 2 classes (OUI ou NON, Client ou Imposteur) à partir d'un vecteur de nombre réels dont la dimension est égale au nombre de sous-systèmes du système multi-algorithmes.

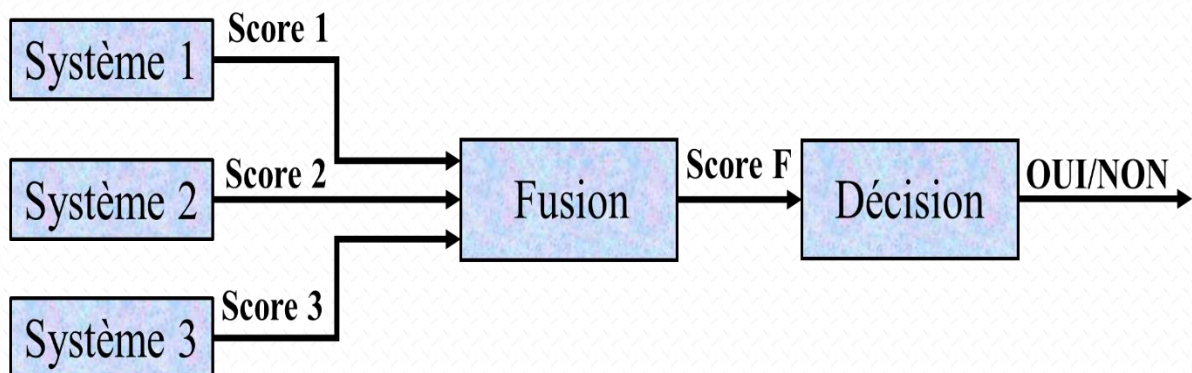


Fig 2. 8. Schéma de la fusion de scores [7].

Il existe deux approches pour combiner les scores obtenus par différents systèmes.

- **Approche par combinaison de scores :** Elle Consiste à traiter le sujet comme un problème de combinaison de scores par des méthodes mathématiques de combinaison, d'une manière de combiner les scores individuels pour avoir un unique score, qui est ensuite utilisé pour prendre la décision finale. Afin de s'assurer que la combinaison de scores provenant de différents systèmes soit cohérente, les scores doivent d'abord être transformés dans un domaine commun : on parle alors de normalisation de score.
- **Approche par classification de scores :** Elle Consiste à traiter le problème des scores individuels comme un problème de classification de scores. Pour cela un vecteur de caractéristiques est construit en utilisant les scores donnés en sortie par les systèmes individuels ; ce vecteur est ensuite attribué à une des deux classes : Client ou Imposteur. Aucun traitement n'est requis avant de les envoyer dans le classifieur. Le classifieur utilisé pour cette opération est capable d'apprendre la frontière de décision sans tenir compte de la manière dont le vecteur de caractéristiques a été généré. Les scores en sortie de différents systèmes peuvent être non-homogènes (mesure de distance ou de similarité, différents intervalles de valeurs prises, etc....) [7].

2.5.3. Normalisation de scores [7]

Les méthodes de normalisation de scores cherchent de rendre les scores issus des systèmes individuellement homogènes avant de les combiner. Certains systèmes produisent des scores de similarité (plus le score est grand, plus la référence ressemble au test, donc l'utilisateur est un Client), d'autres produisent des distances (plus la distance est faible, plus la référence et le test sont proches, plus l'utilisateur est un Client). De plus chaque système peut avoir des intervalles de variations des scores différents, par exemple pour un système les scores varient entre 0 et 1 et pour un autre les scores varient entre 0 et 1000. Cela nous permet de confirmer la nécessité de la normalisation des scores avant de les combiner.

Parmi les techniques de normalisation de scores, on trouve :

1) Normalisation par la méthode Min-Max :

La normalisation **Min – max** remet à l'échelle de manière linéaire toutes les caractéristiques de l'intervalle **[0, 1]**. Le but d'avoir un tel intervalle restreint est de réduire l'espace de variation des valeurs d'une caractéristique et par conséquent réduire l'effet des valeurs aberrantes (**Outliers**).

La transformation se fait grâce à la formule suivante :

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \dots \dots \dots (2.1)$$

Avec :

X_{min} : La plus petite valeur observée pour la caractéristique *X*.

X_{max} : La plus grande valeur observée pour la caractéristique *X*.

X : La valeur de la caractéristique qu'on cherche à normaliser.

2) Normalisation par la méthode Z-Score :

La transformation a pour but de ramener la moyenne arithmétique **μ à 0** et l'écart-type **σ à 1**. Encore une fois, le procédé est simple si on a à notre disposition la moyenne **μ** et l'écart-type **σ** de la variable, la formule pour le faire est la suivante :

$$X_{\text{norm}} = \frac{X - \mu}{\sigma} \dots \dots \dots (2.2)$$

3) Normalisation par la médiane et l'écart absolu médian (MAD) [8]:

Sont insensibles aux valeurs aberrantes et aux points aux extrémités d'une distribution. Cette technique de normalisation ne conserve pas la distribution d'entrée et ne transforme pas les scores dans un intervalle commun.

La transformation se fait grâce à la formule suivante :

$$X_{\text{norm}} = \frac{X - \text{median}}{\text{MAD}} \dots \dots \dots (2.3)$$

Où le : $MAD = \text{median}(\{|X - \text{median}(\{X\})|\})$.

4) Normalisation par la méthode tangente hyperbolique "Tanh"[8] :

Toutes les valeurs sont converties en tangente hyperbolique (**Tanh**).

La normalisation est donnée par la formule suivante :

$$X_{\text{norm}} = \frac{1}{2} \left\{ \tanh \left(0.01 \left(\frac{X - \mu_{GH}}{\sigma_{GH}} \right) \right) + 1 \right\} \dots \dots \dots (2.4)$$

Où μ_{GH} et σ_{GH} sont respectivement les estimateurs de la moyenne et de l'écart-type de la distribution des scores authentiques.

2.6. Evaluation des performances d'un système biométrique « SB »

Un système de reconnaissance vise soit à établir une identité d'une personne déterminée, ou investiguer une identité revendiquée provenant d'une personne dont l'identité ayant été auparavant établie ou non [3]. Ainsi, cela nous permise de dire que chaque système biométrique est en face de deux populations :

- **Les clients** appartenant au système, qui ont l'autorisation de l'utiliser et d'y entrer.
- **Les imposteurs** non autorisés à utiliser le système, mais essayent quand même de pénétrer.

L'évaluation d'un **SB** se fait par plusieurs mesures à effectuer voir [Fig 2. 9](#) :

		Vrais Classes	
		P	N
Classe hypothétique	Oui	Vrais Positifs	Faux Positifs
	Non	Faux Négatifs	Vrais Négatifs

Fig 2. 9. Matrice de confusion d'hypothèses de décision.

Les systèmes biométriques sont sensibles aux erreurs suivantes :

- **FRR (False Rejection Rate) ou TFR (taux de faux rejets) :** par définition est la proportion de fois qu'un individu légitime (abonné ou client) est rejeté par le système. Il est appelé aussi **FNR (False Negative Rate)**. A noter qu'un faux rejet ne signifie pas nécessairement que l'erreur est due au système, par exemple, dans le cas d'un système à base d'empreintes digitales, un mauvais positionnement du doigt sur le capteur ou la saleté peuvent produire des faux rejets.

Le **FRR** est lié à la sensibilité **TPR (True Positive Rate)** : la proportion de fois qu'un individu légitime (client) est accepté par le système.

$$FRR = FNR = 1 - TPR = 1 - \left(\frac{TP}{P}\right) \dots \dots \dots (2.5)$$

- **FAR (False Acceptance Rate) ou TFA (Taux de Fausses Acceptations) :** par définition est la proportion de fois qu'un individu non-légitime (Imposteur) est accepté par le système. Il est appelé aussi **FPR (False Positive Rate)**.

Le **FAR** est lié à la spécificité ou **TNR (True Negative Rate)**: la proportion de fois qu'un individu non-légitime (Imposteur) est rejeté par le système.

$$FAR = FPR = 1 - TNR = 1 - \left(\frac{TN}{N}\right) \dots \dots \dots (2.6)$$

- **TER (Total Error Rate)** : c'est le taux d'erreur totale d'un système biométrique. Cette mesure est calculée par la relation suivante :

$$TER = FRR + FAR \dots \dots \dots (2.7)$$

- **HTER (Half Total Error Rate)** : c'est le taux d'erreur moyenne d'un système biométrique. Cette métrique est donnée par l'équation :

$$HTER = \frac{TER}{2} = \frac{FRR + FAR}{2} \dots \dots \dots (2.8)$$

- **EER (Equal Error Rate)** : c'est le taux d'erreur égale du système, qui correspond au taux d'erreur pour lequel **FAR** est égal à **FRR**.

Un bon système de vérification minimise à la fois les deux taux FAR et FRR en ajustant le seuil de décision. Le plus souvent, les performances du système biométrique sont exprimées en taux d'erreur égale (Equal Error Rate, EER), qui correspond à la valeur où le FRR est égal au FAR. Plus la valeur de l'EER est faible, plus le système est meilleur. L'EER est utilisé pour comparer les systèmes biométriques entre eux. L'EER correspond au point d'intersection des deux courbes FRR et FAR (voir Fig 2. 10).

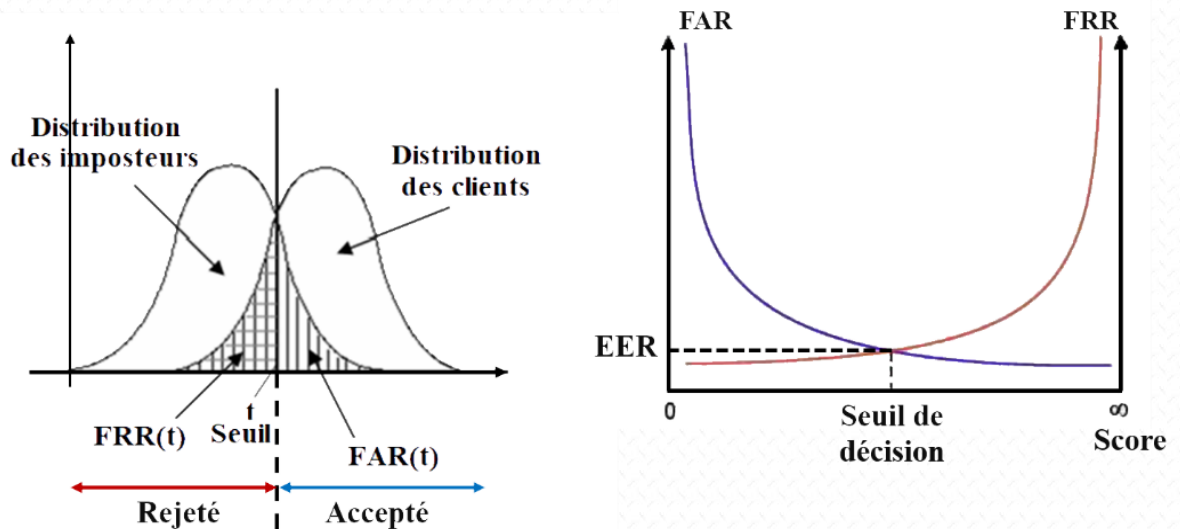


Fig 2. 10. Illustration de FAR, FRR et de EER en fonction du seuil de décision [9], [10].

La relation entre FRR et FAR est illustrée par les courbes appelées **ROC** (Receiver Operating Characteristics) et **DET** (Detection Error Trade-Off). Ces courbes sont obtenues en faisant varier le seuil de décision et en calculant à chaque fois les deux valeurs FRR et FAR. Fig 2. 11 présente un exemple de courbes **ROC** et **DET**.

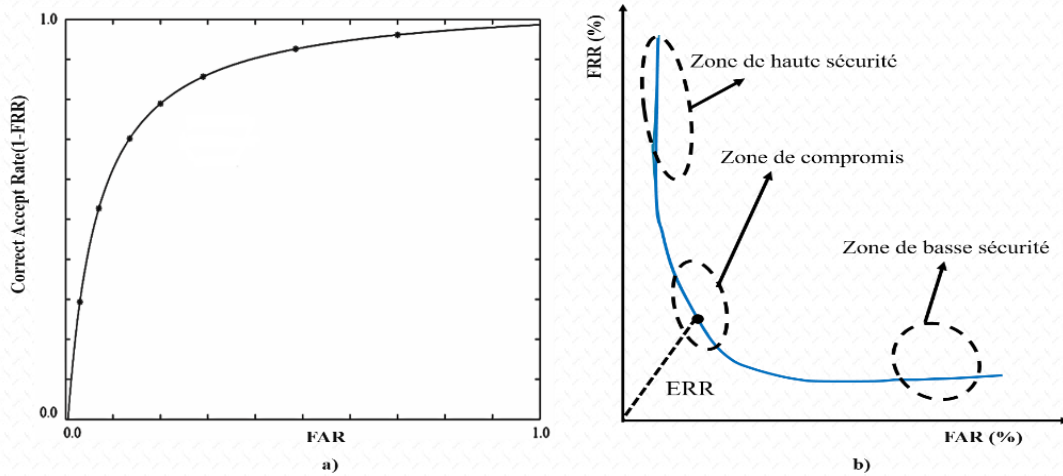


Fig 2. 11. a) Courbe ROC ; b) Courbe DET [10][11].

Une autre courbe très importante est la courbe de scores cumulés **CMC** (Cumulative Match Characteristic). Elle indique la probabilité des bonnes identifications pour un rang n (probabilité que la personne correspond aux n meilleures scores (Fig 2. 12). On dit qu'un système reconnaît au rang 1 lorsqu'il choisit l'image la plus proche suite à la reconnaissance. On dit qu'un système reconnaît au rang 2, lorsqu'il choisit parmi deux images, celle qui correspond le mieux à l'image d'entrée.

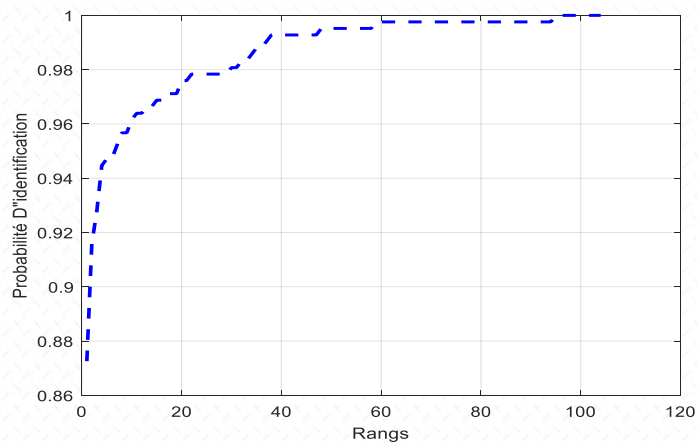


Fig 2. 12. Courbe CMC.

2.7. Les méthodes de la reconnaissance du visage

Un système de reconnaissance de visages à trois types de tâches à remplir, comme nous avons mentionné au paravent : **L'identification**, **La vérification** ou **authentification**, et **La surveillance**. Afin d'améliorer les performances des **SRV**, plusieurs méthodes ont été achevées depuis les premiers travaux menés durant les années 1960-1970 [12]-[14]. Ces méthodes peuvent être répartie en trois catégories : **Méthodes globales** (ou **holistiques**), **Méthodes locales**, et **Méthodes hybrides** [7], [8]. Depuis la fin de l'année 2012 une nouvelle catégorie de méthodes d'apprentissage profond (Deep Learning) a révolutionné le domaine de la reconnaissance de visage et le domaine de la Vision par ordinateur (Computer Vision).

Dans cette section, on va expliquer avec un peu plus de détail ces catégories de méthodes (voir Fig 2. 13).

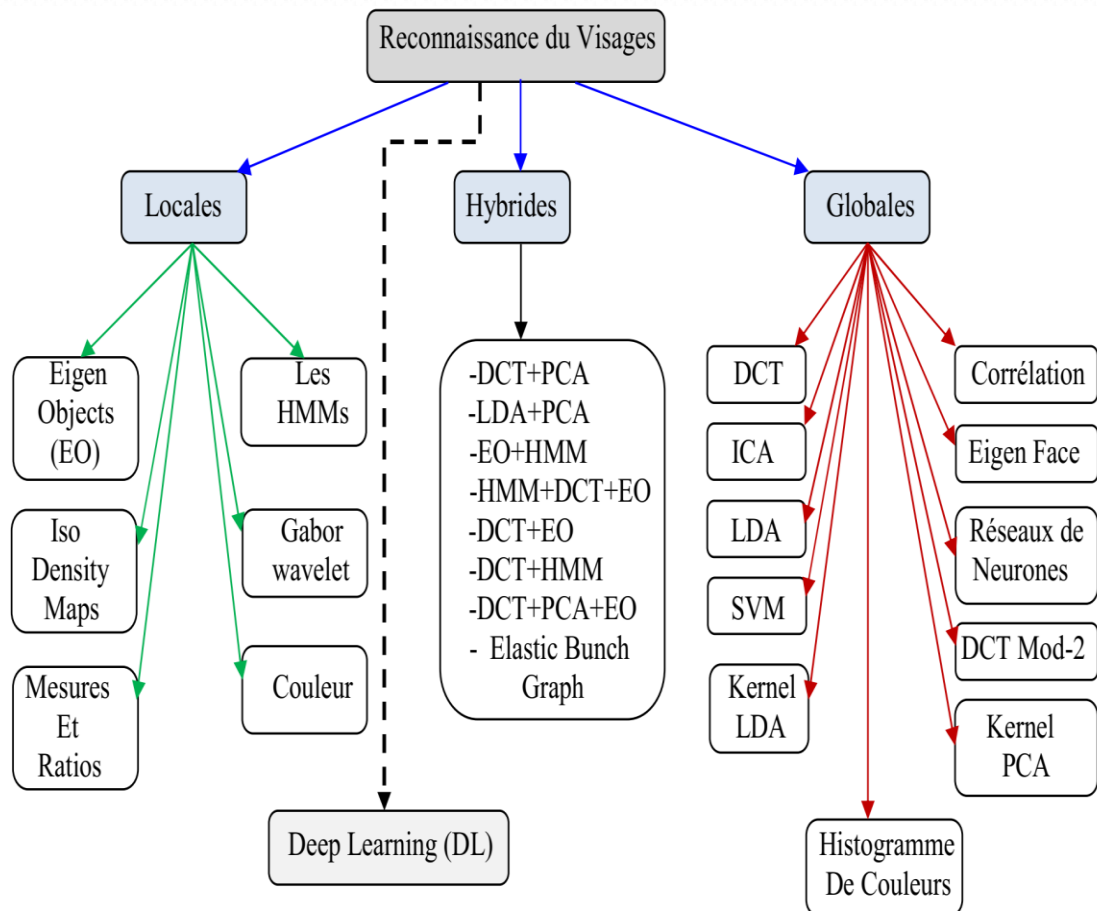


Fig 2. 13. Classification des algorithmes principaux utilisés en reconnaissance faciale [8].

2.7.1. Méthodes globales

Ce sont des méthodes basées sur l'apparence. Cela veut dire qu'il n'est pas nécessaire de localiser certains repères du visage (comme les centres des yeux, les narines, le centre de la bouche, etc.). Dans ces méthodes, l'une des étapes est la transformation des images de visages en matrices de valeurs de pixels qui seront à leur tour transformer en vecteurs faciles à manipuler.

Elles sont relativement rapides à mettre en œuvre, mais l'inconvénient principal, est qu'elles sont très sensibles aux variations d'éclairage, de pose et d'expression faciale. Ce qui nous ramène à établir que la moindre variation des conditions de l'environnement ambiant de l'image influe et entraîne des changements des valeurs de pixels.

Parmi les techniques des méthodes globales les plus exploitées en littérature, on trouve les techniques linéaires et les techniques non linéaires.

Les **techniques linéaires** réalisent une projection linéaire des visages (espace dont la dimension est égale à la dimension des images, donc grande) sur un espace de plus faible dimension. L'inconvénient principal de ces techniques, qu'ils sont incapables de préserver les variations non convexes des variétés de visages afin de différencier des individus. Ainsi, l'utilisation de distances classiques dans l'espace projeté ne permet pas toujours de réaliser une bonne classification entre les classes « visages » et « non visages ». Ce facteur crucial limite le pouvoir des techniques linéaires pour obtenir une détection et une reconnaissance du visage très précises. La plus connue de ces techniques est sans aucun doute l'**Analyse en Composantes Principales (abrégé PCA en anglais)**, également appelée **transformée de Karhunen-Loeve [15]-[16]**. Le PCA fut d'abord utilisé afin de représenter efficacement des images de visages humains. Il existe d'autres techniques également construites à partir de *décompositions linéaires* comme l'analyse discriminante linéaire (abrégé **LDA en anglais**) [17] ou encore l'analyse en composantes indépendantes (abrégé **ICA en anglais**) [18]. Tandis que le PCA construit un sous-espace pour représenter de manière "optimale" (mathématiquement parlant) seulement "**l'objet**" visage, le LDA construit un sous-espace discriminant pour distinguer de façon "optimale" les visages de différentes personnes. Elle

permet donc d'effectuer une véritable séparation de classes. Des études comparatives montrent que les méthodes basées sur le LDA donnent généralement de meilleurs résultats que les méthodes basées sur le PCA. L'algorithme ICA, quant à lui, est une généralisation de l'algorithme PCA avec lequel il coïncide dans le cas de **données gaussiennes**. L'algorithme ICA est basé sur le concept intuitif de **contraste** et permet d'éliminer la redondance statistique des données de départ. Bien que ces méthodes globales linéaires basées sur l'apparence évitent l'instabilité des toutes premières méthodes géométriques qui ont été mises au point, elles ne sont pas assez précises pour décrire les subtilités des variétés (géométriques) présentes dans l'espace de l'image originale. Ceci est dû à leurs limitations à gérer la non-linéarité en reconnaissance de visage : les déformations de variétés non linéaires peuvent être lissées et les concavités peuvent être remplies, causant des conséquences défavorables.

Afin de pouvoir traiter ce problème de non-linéarité en reconnaissance de visage; de telles méthodes linéaires ont été étendues à des **techniques non linéaires** basées sur la notion mathématique de noyau (*kernel*) comme le Kernel PCA [19] et le Kernel LDA [20]. Ici, une projection non linéaire (réduction de dimension) de l'espace de l'image sur l'espace de caractéristiques (*feature space*) est effectuée ; les variétés présentes dans l'espace de caractéristiques résultant deviennent simples, de même que les subtilités des variétés qui sont préservées. Bien que les méthodes basées sur le noyau peuvent atteindre une bonne performance sur les données d'entraînement, il ne peut pas en être de même pour de nouvelles données en raison de leur plus grande flexibilité ; contrairement aux méthodes linéaires.

2.7.2. Méthodes locales

Les méthodes locales [8], basées sur des modèles, utilisent des connaissances a priori que l'on possède sur la morphologie du visage et s'appuient en général sur des points caractéristiques de celui-ci. Kanade [14] présenta un des premiers algorithmes de ce type en détectant certains points ou traits caractéristiques d'un visage puis en le comparant avec des paramètres extraits d'autres visages. Ces méthodes constituent une autre approche pour prendre en compte la non-linéarité en construisant un espace de caractéristiques local et en utilisant des filtres d'images appropriés, de manière à ce que les distributions des visages soient moins

affectées par divers changements. Les approches Bayésiennes, les machines à vecteurs de support, la méthode des modèles actifs d'apparence ou encore la méthode « local binary pattern » ont été utilisées dans ce but. Toutes ces méthodes ont l'avantage de pouvoir modéliser plus facilement les variations de pose, d'éclairage et d'expression par rapport aux méthodes globales. Toutefois, elles sont plus lourdes à utiliser puisqu'il faut souvent placer manuellement un assez grand nombre de points sur le visage alors que les méthodes globales ne nécessitent de connaître que la position des yeux afin de normaliser les images, ce qui peut être fait automatiquement et de manière assez fiable par un algorithme de détection.

2.7.3. Méthodes hybrides

Les méthodes hybrides permettent d'associer les avantages des méthodes globales et locales en combinant la détection de caractéristiques géométriques (ou structurales) avec l'extraction de caractéristiques d'apparence locales. Elles permettent d'augmenter la stabilité de la performance de reconnaissance lors de changements de pose, d'éclairage et d'expressions faciales. L'analyse de caractéristiques locales [21] et les caractéristiques extraites par ondelettes de Gabor, sont des algorithmes hybrides typiques. Plus récemment, l'algorithme LogGabor PCA effectue une convolution avec des ondelettes de Gabor orientées autour de certains points caractéristiques du visage afin de créer des vecteurs contenant la localisation et la valeur d'amplitudes énergétiques locales ; ces vecteurs sont ensuite envoyés dans un algorithme PCA afin de réduire la dimension des données.

La Fig 2. 13 fournit une classification des algorithmes principaux de reconnaissance de visage.

2.7.4. Apprentissage profond (Deep Learning)

Le démarrage de l'apprentissage profond (**Deep Learning**) a eu lieu réellement en 2012 [22]. Par définition le DL est un ensemble de méthodes d'apprentissage automatique non linéaire composé d'un ensemble de "**Réseaux de neurones à plus de deux couches** [23]" (voir Fig 2. 14).

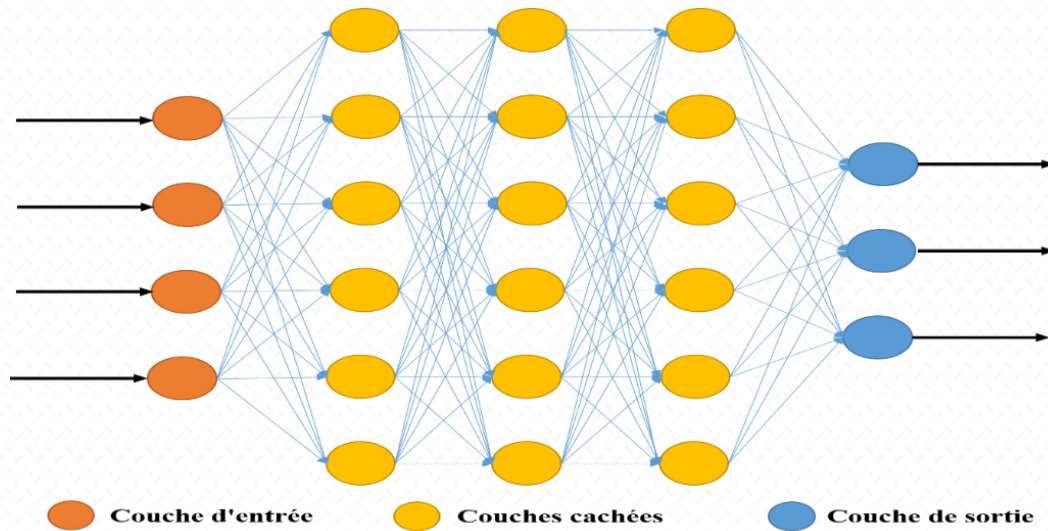


Fig 2. 14. Réseau de neurones d'apprentissage profond "Deep Learning Neural Network" [23].

Autrement dit, l'apprentissage profond (DL) fait partie du domaine du Machine Learning, qui fait lui-même partie du domaine d'Intelligences Artificielles, et l'inverse n'est pas vrai. Les quatre architectures fondamentales de **DL** sont :

- Réseaux pré-entraînés non supervisés (Unsupervised pretrained networks UPN).
- Réseaux de neurones convolutionnels (Convolutional neural networks CNN).
- Réseaux de neurones récurrents (Recurrent neural networks RNN).
- Réseaux de neurones récursifs (Recursive neural networks RNN).

On est intéressé dans cette thèse par la structure CNN, qui est bien adaptée aux problèmes de classification d'images, la détection d'objets. Ces structures peuvent identifier des visages, des individus, des panneaux de signalisation, et de nombreux autres aspects des données visuelles [24]. L'une des structures CNNs qui a pris une place très importante dans la littérature en dépassant un total de 6184 citations, on trouve le réseau **AlexNet**. Il est largement considéré comme l'une des publications les plus influentes dans le domaine de la vision par ordinateur. Il a remporté le concours ImageNet LSVRC-2012 2012 avec une large marge (taux d'erreur de 15,3 % contre 26,2 % (deuxième place)) [22]. Il se compose de 5 couches convolutives et de 3 couches entièrement connectées (Fig 5. 2). Les statistiques détaillées se trouvent dans le [Tableau 2. 1](#).

Tableau 2. 1. Paramètres d'AlexNet

1	'data'	Image Input	227x227x3 Images with 'zerocenter' normalization
2	'conv1'	Convolution	96 11x11x3 Convolutions with stride [4 4] and padding [0 0 0 0]
3	'relu1'	ReLU	ReLU
4	'norm1'	Cross Channel Normalization	Cross channel normalization with 5 channels per element
5	'pool1'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
6	'conv2'	Convolution	256 5x5x48 convolutions with stride [1 1] and padding [2 2 2 2]
7	'relu2'	ReLU	ReLU
8	'norm2'	Cross Channel Normalization	Cross channel normalization with 5 channels per element
9	'pool2'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
10	'conv3'	Convolution	384 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
11	'relu3'	ReLU	ReLU
12	'conv4'	Convolution	384 3x3x192 Convolutions with stride [1 1] and padding [1 1 1 1]
13	'relu4'	ReLU	ReLU
14	'conv5'	Convolution	256 3x3x192 Convolutions with stride [1 1] and padding [1 1 1 1]
15	'relu5'	ReLU	ReLU
16	'pool5'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
17	'fc6'	Fully Connected	4096 Fully connected layer
18	'relu6'	ReLU	ReLU
19	'drop6'	Dropout	50% Dropout
20	'fc7'	Fully Connected	4096 Fully connected layer
21	'relu7'	ReLU	ReLU
22	'drop7'	Dropout	50% Dropout
23	'fc8'	Fully Connected	1000 Fully connected layer
24	'prob'	Softmax	Softmax
25	'output'	Classification Output	Crossentropyex with 'tench' and 999 other classes

Dans certains cas, certains systèmes de reconnaissance ne peuvent pas être entraînés facilement et peuvent prendre trop de temps pour fonctionner correctement. Lorsqu'il n'y a pas assez de temps pour construire un nouveau système (Modèle), ou qu'il y a trop de tâches

à gérer par le système, une solution élaborée, d'adopter un modèle existant similaire pré-entraîné, on parle de l'apprentissage par transfert (Transfer Learning). L'un des objectifs de cette technique est de gagner ainsi du temps pour construction du système, au lieu d'en créer un nouveau.

Le Transfer Learning, ou apprentissage par transfert, désigne les techniques permettant de transférer les connaissances de plusieurs sources vers d'autres problèmes à régler. Il s'agit des compétences apprises à partir de tâches antérieures servant à améliorer de nouvelles tâches similaires. Ce transfert de savoirs global est très efficace dans le cas du Deep Learning.

Il existe différentes catégories de Transfer Learning :

- **L'apprentissage par transfert inductif** : les sources et cibles relèvent du même domaine, mais les problèmes à résoudre sont différents.
- **L'apprentissage par transfert non supervisé** : les sources et cibles sont similaires, et les tâches sont différentes. Par contre, les données ne sont pas labellisées, ce qui permet d'obtenir un plus grand nombre de sources, notamment sur le web.
- **L'apprentissage par transfert transductif** : les tâches et sources sont les mêmes, mais les domaines sont différents.

L'exploitation du TL pour les réseaux de neurones convolutionnels (CNN) sont utilisés avec succès dans les environnements traditionnels d'exploration de données [25].

L'apprentissage par transfert [26] a récemment été introduit dans la RV profonde pour résoudre le problème du changement de domaine. Il apprend les fonctionnalités transférables à l'aide d'un domaine source étiqueté (données d'apprentissage) et d'un domaine cible non étiqueté (données de test), de sorte que l'écart de domaine est réduit et que les modèles formés sur le domaine source fonctionneront également bien sur le domaine cible. De plus, le réglage fin des paramètres CNN à partir d'un modèle pré-appris à l'aide d'un ensemble de données d'apprentissage cible est un type particulier d'apprentissage par transfert et est couramment utilisé par de nombreuses méthodes [27]-[29].

2.8. Conclusion

Dans ce chapitre nous avons introduit les concepts de systèmes biométriques de reconnaissance, leurs structures et leurs applications. Nous avons aussi montré que les performances des systèmes biométriques dépendent de plusieurs facteurs et qu'elles varient d'un système à un autre. Les fusions multimodales ont été discutées, en concluant que c'est l'une des solutions d'amélioration des performances biométriques. Les systèmes de reconnaissances de visages ont pris un espace de discussion dans ce chapitre, ainsi que les catégories des méthodes exploitées dans RV.

3. État de L'art

3.1. Introduction

La reconnaissance de visage s'est avérée être l'une des technologies les plus réussies et a eu un impact sur des domaines hétérogènes.

Une étape préalable et indispensable à toute tentative d'amélioration des résultats souhaitables des systèmes physiques est "la modélisation". La grande majorité des techniques utilisées dans les systèmes de reconnaissance biométrique se basent sur la connaissance d'un modèle formalisé ou non du système à étudier. L'analyse et la synthèse de l'imagerie ont connues beaucoup d'évolutions et sont devenues beaucoup plus précises après l'introduction des techniques de modélisation pour les scènes complexes. L'analyse d'images repose sur un large éventail de techniques de modélisation différentes basées sur la géométrie, la représentation et le mouvement. Modéliser un processus n'est pas une fin en soi ; ce qui est demandé au processus est la réalisation d'une mission cible. Pour que cette mission cible soit garantie en quantité et qualité, il faut que toute anomalie de fonctionnement soit rapidement détectée puis prise en compte dans la stratégie de conduite du système considéré [30]. Ces anomalies ou ces problèmes donnés peuvent souvent être résolus par des processus d'optimisation. Ces processus d'optimisation, d'une ou de plusieurs fonctions coûts (fonctions objectives) dont le but est de minimiser ou de maximiser la "fonction coûts " par rapport à tous les paramètres concernés.

Dans la littérature, on trouve une grande variété de méthodes, conçues pour la résolution des problèmes d'optimisation : en premier lieu on trouve les méthodes classiques qui reposent uniquement sur les mathématiques, mais imposent des conditions d'application strictes. En deuxième lieu il y'a les méthodes numériques que l'on pourrait qualifier d'heuristiques n'essaient pas de trouver une solution idéale mais tentent d'obtenir une solution dans un temps donné disponible pour le calcul. Finalement on trouve les méthodes dites méta-heuristiques,

qui ont émergées dans les années 1980. La méta-heuristique présente de nombreuses similitudes avec la physique, la biologie ou même l'éthologie [31].

Dans ce chapitre état de l'art nous présentons premièrement, un aperçu de descripteurs pour la reconnaissance de visage. Deuxièmement, un résumé des méthodes d'optimisation utilisée pour la RV. Nous présentons dans la dernière partie du chapitre une mise en revue qui prend en charge diverses architectures de reconnaissances de visages qui ont eu un impact substantiel.

3.2. Les descripteurs de visage

Bien que la reconnaissance de visage soit assez facile à traiter pour les humains, il est assez difficile de faire en sorte que les machines comprennent les traits du visage et les calculent pour une reconnaissance future. De nombreuses caractéristiques affectent le processus de reconnaissance de visage, à savoir les conditions d'éclairage, l'angle du visage, la couverture du visage, différentes caractéristiques du visage comme la barbe, la moustache, les lunettes, etc. En tenant compte de ces caractéristiques, il relève en outre le défi de détecter un visage par machine un niveau en avant.

Pour localiser des visages sur des photos ou des vidéos, un détecteur de visage est d'abord utilisé. Les visages sont alignés sur des coordonnées canoniques normalisées à l'aide du détecteur de repère facial. Ces photos de visage alignées sont utilisées pour créer le module de reconnaissance de visage. De plus, avant que les caractéristiques ne soient transmises au module de reconnaissance de visage, on s'assure que le visage est en direct et non usurpé. Pour que le système de reconnaissance de visage fonctionne, il nécessite trois modules, à savoir le prétraitement du visage, l'extraction des caractéristiques et la perte d'entraînement pour l'appariement du visage, qui seront expliqués brièvement ci-dessous.

Les approches conventionnelles d'apprentissage de sous-espace basées sur les orientations de gradient d'image n'emploient que l'information de premier ordre. Pourtant, récent recherches sur le système de vision humain (HVS) a dévoilé qu'un paysage ou les propriétés d'une surface géométrique ne peuvent être capturés par l'information de second

ordre. Les orientations de gradient d'image du second ordre peuvent atténuer l'effet négatif des bruits dans les images de visage et aussi réduire la redondance. Pour cela nous nous intéressons à différents descripteurs. Le [Tableau 3. 1](#) présente quelques résultats de quelques travaux liés aux descripteurs RV.

Tableau 3. 1. Résultats de quelques travaux liés aux descripteurs RV

Références	Base de données	Méthodes	TR(%)
Hussain et al. [33]	FERET	LPQ	99.20%
	LFW		75.30%
Karaaba et al. [34]	FERET	HOG and MMD	68.59%
	LFW		23.49%
Arigbabu et al. [35]	LFW	PHOG and SVM	88.50%
Seo et al. [36]	LFW	LARK + PCA	78.90%
Preeti & Kumar D [37]	ORL	DCT-PCA & LDA	93.00
Huang M et al. [38]	ORL	DLSR	95.63
Malhotra P et al. [39]	ORL	PCA + DCT + CS	96.50
Keinert F et al. [40]	AR	GSR-FR	96.70
Peng Y et al. [41]	AR	SFD-CR	96.94
Shang K et al. [42]	AR	CD-EJSR	99.17
Liu J et al. [43]	AR	KSVD-IJSR	99.17
Ayyavoo T et al. [44]	AR	DWT E-CLAHE	98.78
Elaggoune H et al. [45]	Thermal Tufts Face	Gabor + HOG + PCASVD	87.26

3.3. Les méthodes d'optimisation de la RV

La plupart des algorithmes d'optimisations de la vision par ordinateur sont des méthodes d'optimisation globale. De telles méthodes cherchent à minimiser ou de maximiser la "fonction coûts " par rapport à des paramètres bien définis et les intègre de manière globale, c'est-à-dire sur tout le domaine de l'image. Étant donné que ces méthodes de modélisation sont très transparentes et intuitives, il n'est pas surprenant que de telles méthodes soient devenues des outils populaires et efficaces pour résoudre de nombreux problèmes fondamentaux de la vision par ordinateur tels que la reconnaissance de visage.

Dans le monde numérique, des algorithmes évolutifs tels que l'algorithme génétique (GA), l'optimisation des essaims de particules (PSO) sont largement utilisés pour éliminer les caractéristiques du visage redondants et irrévérrencieux [46]. Le **Tableau 3. 2** présente un résumé des travaux liés au connexes.

Tableau 3. 2. Résultats de quelques travaux liés aux méthodes d'optimisation de RV

Références	Base de données	Méthodes	TR(%)
Elaggoune H et al. [45]	ORL	Gabor + LPQ + PSO	99.50
	AR	Gabor + HOG+PSO+PCASVD	99.32
	Thermal Tufts Face	Gabor + HOG+PSO+PCASVD	87.26
	LFW	Gabor+HOG+PSO	72.35
Elaggoune H & al. [47]	Yale-B	Gabor+HOG+PSO	99.58
	LFW		69.25
Yaji GS et al. [48]	ORL	DWT+PSO	96.8
		DCT+PSO	94.7
Ramadan RM et al. [49]	ORL	DWT+DCT	95.03
Yin H et al. [50]	ORL	PCA+PSO	97.5
Khadhraoui T et al. [51]	YALE-B	DWT+MPSO	98.29
Sasirekha K et al. [52]	ORL	PSO-KNN	98.75
Samma H & al. [53]	LFW	HOG+ LBP+GABOR+PSO	50.00
Zhou G et al. [54]	ORL	SRC-EnNPSO	98.92
Di X et al. [55]	TTF	Multi-AP-GAN	75:86

3.4. Apprentissage profond (Deep Learning) pour la RV

L'apprentissage en profondeur s'est avéré être le plus efficace en informatique tâches de vision en raison de son architecture basée sur la convolution.

Le système de reconnaissance de visage est une technologie que l'on peut utiliser efficacement et avec précision pour détecter et reconnaître un individu utilisant les caractéristiques de son visage. Ce système peut être utilisé pour les détecter dans les images, vidéos ou en temps réel. Les approches traditionnelles, telles que le filtrage réponses, histogrammes de caractéristique codes, dictionnaire de distribution d'atomes (distribution of dictionary atoms), recherchent à distinguer un visage humain n'utilisant que des

représentations à une ou deux couches. Les méthodes d'apprentissage en profondeur, comme CNN, utilisent une cascade de plusieurs couches d'unités de traitement pour l'extraction de caractéristiques et transformations.

DeepFace [56] a atteint un Taux de reconnaissance de pointe sur le célèbre Benchmark LFW [57] en 2014, le Taux de reconnaissance a atteint pour la première fois (DeepFace : 97,35 % via Humain : 97,53 %) pour des images prises sans contrainte, en entraînant un réseau de neurones de 9 couches. Quatre millions de photos de visages ont été utilisées pour créer un modèle profond. La caractéristique de Gabor [58] découverte par des scientifiques avec des années d'expertise est analogue à la première couche du réseau neuronal profond. La deuxième couche est responsable de l'apprentissage des détails de texture plus compliqués. Les caractéristiques de la troisième couche sont des structures rudimentaires plus compliquées et certaines, comme un nez en pont haut et de grands yeux, ont commencé à être plus décrits. Dans la quatrième, la sortie du réseau est suffisante pour décrire les caractéristiques particulières du visage, qui auraient une discrète réaction à des idées abstraites non ambiguës telles que le sourire, le cri ou même l'œil bleu.

Le principal objectif de l'introduction de reconnaissance de visage est le besoin de sécurité des informations ou des biens physiques. Dans le monde avancé d'aujourd'hui, différents types de fraudes prennent lieu sur tout le monde, utilisant différents types d'infractions. Le seul moyen d'authentification était en les clés, les cartes d'identité, mots de passe, etc à déterminer qui est la personne. Avec l'introduction d'un système efficace de reconnaissance de visage, le système devient extrêmement sécurisé et robuste, et élimine le besoin de se rappeler des mots de passe ... Ceci donne une conclusion que les couches inférieures de CNN apprennent automatiquement les caractéristiques, et les plus hauts niveaux sont en relation avec un certain niveau d'abstraction des modèles à apprendre.

Nous présentons dans les Fig 3. 1, Fig 3. 2 les architectures les plus influentes de l'apprentissage profond de visages par ordre chronologique, et dans le Tableau 3. 3 quelques résultats de travaux liés aux méthodes DL avec optimisation.

3.4.1. Aperçu sur les méthodes récentes DL pour la RV

Les systèmes de reconnaissance de visage [32] sont la voie des machines [59] à se rapprocher des humains, rapprochant le monde de l'automatisation. Les modèles DL utilisés pour la reconnaissance de visage sont sous formes de plusieurs modules essentiels se présentant comme suit :

- **Prétraitement du visage :** Le prétraitement est un moyen efficace de préparation des images de visages pour le traitement, l'analyse et la classification : comme mentionné précédemment, la reconnaissance de visage est affectée par différentes conditions telles que l'éclairage, les occlusions, les poses, etc. Pour résoudre ce problème, le traitement du visage est une étape nécessaire dans ce système.
- **Extraction des caractéristiques profond (Deep Feature Extraction) :** L'architecture du réseau peut être appelée l'épine dorsale de l'ensemble du système de reconnaissance de visage [60]. L'extraction de caractéristiques profondes est essentielle pour la classification d'images multi spectrales et devient un sujet brûlant dans la recherche sur la détection des changements. En raison des effets des variations individuelles et de l'éclairage, certains traits qui sont significativement liés aux changements du visage sont difficiles à extraire. En conséquence, les caractéristiques qui peuvent définir efficacement les changements faciaux sont nécessaires de toute urgence.
- **Perte d'entraînement pour l'appariement des visages :** la perte d'entraînement a joué un rôle essentiel dans l'amélioration de la précision globale des systèmes de reconnaissance de visage. Nous nous limiterons aux deux fonctions de perte notables, la perte euclidienne et la perte cosinus.

1) Augmentation de données (Data Augmentation)

L'augmentation des données, souvent appelée « augmentation des données faciales », est un moyen efficace de compenser les données d'entraînement faciales limitées. C'est une technique pour augmenter la quantité de données d'entraînement ou de test en modifiant des échantillons de visages virtuels réels ou simulés. Les deux méthodes de traitement sont classées comme suit : augmentation un à plusieurs, augmentation plusieurs à un [61].

- **Augmentation de un à plusieurs [62]** : Pour permettre aux réseaux profonds d'apprendre des représentations invariantes de pose, ces approches créent plusieurs patches ou images de variabilité de pose à partir d'une seule image.
- **Augmentation de plusieurs à un [63]** : Ces approches extraient la vue canonique des images de visage d'une ou plusieurs images non frontales, permettant à la RV d'être effectuée dans des paramètres contrôlés.

Les types de transformation sont élaborés en deux catégories pour générer des échantillons améliorés :

- **Transformation générique** : Dans ce type de transformation, l'image est entièrement altérée/transférée et le contenu de haut niveau est ignoré. Il a deux types Géométrique et Photométrique.
- **Transformation spécifique du visage** : cette approche se concentre principalement sur les composants et les attributs du visage, leur permettant de modifier l'âge, les cosmétiques, la posture, la coiffure et les émotions faciales.

2) Extraction des caractéristiques profond (Deep Feature Extraction)

L'extraction de caractéristiques profondes est effectuée en faisant passer l'image d'entrée à travers l'architecture du réseau, et un vecteur de caractéristiques d'intégration correspondant est obtenu. L'idée d'augmenter la précision d'un système de reconnaissance de visage repose soit sur l'amélioration des données de haute qualité, soit sur l'amélioration de

l'architecture du réseau et de la perte de formation. L'amélioration de l'ensemble de données est un défi de tâche et nécessite beaucoup de ressources. Certains géants comme Facebook et Google ont construit des ensembles de données de reconnaissance de visage qui incluent environ 10^6 - 10^7 visages uniques [32] ; [56]. Cependant, les universitaires ne disposent pas des ressources de calcul nécessaires et comptent généralement sur l'amélioration de l'architecture du réseau ou sur la perte de formation pour améliorer encore la précision du système de reconnaissance de visage. Par exemple, la précision de LFW l'indice de référence s'est amélioré de 97 % à 99.7 % en raison de changements dans l'architecture ou ajustements dans la perte de formation.

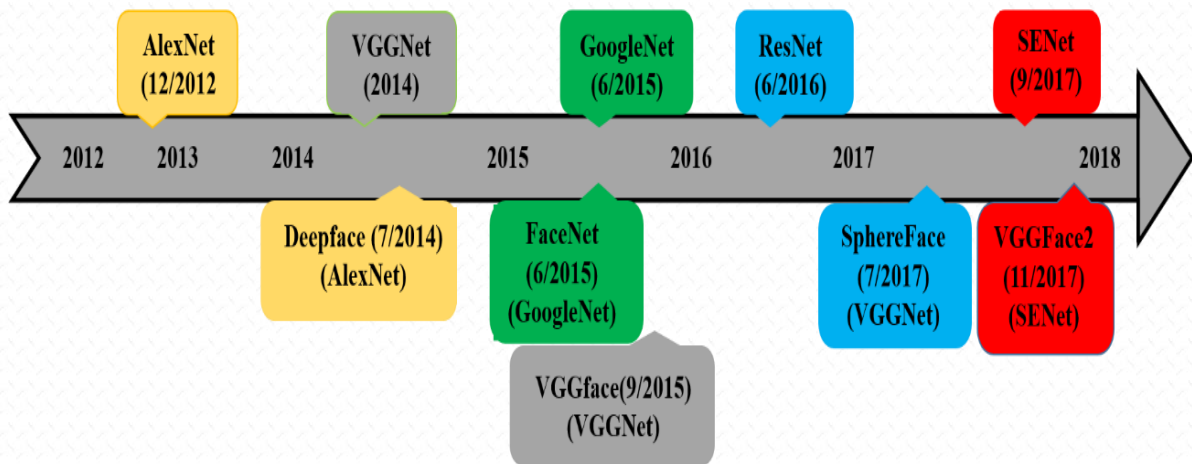


Fig 3. 1. Évolution de l'architecture La rangée du haut présente les architectures de réseaux typiques dans la classification des objets, et la rangée du bas décrit les algorithmes de RV qui utilisent les architectures typiques [26].

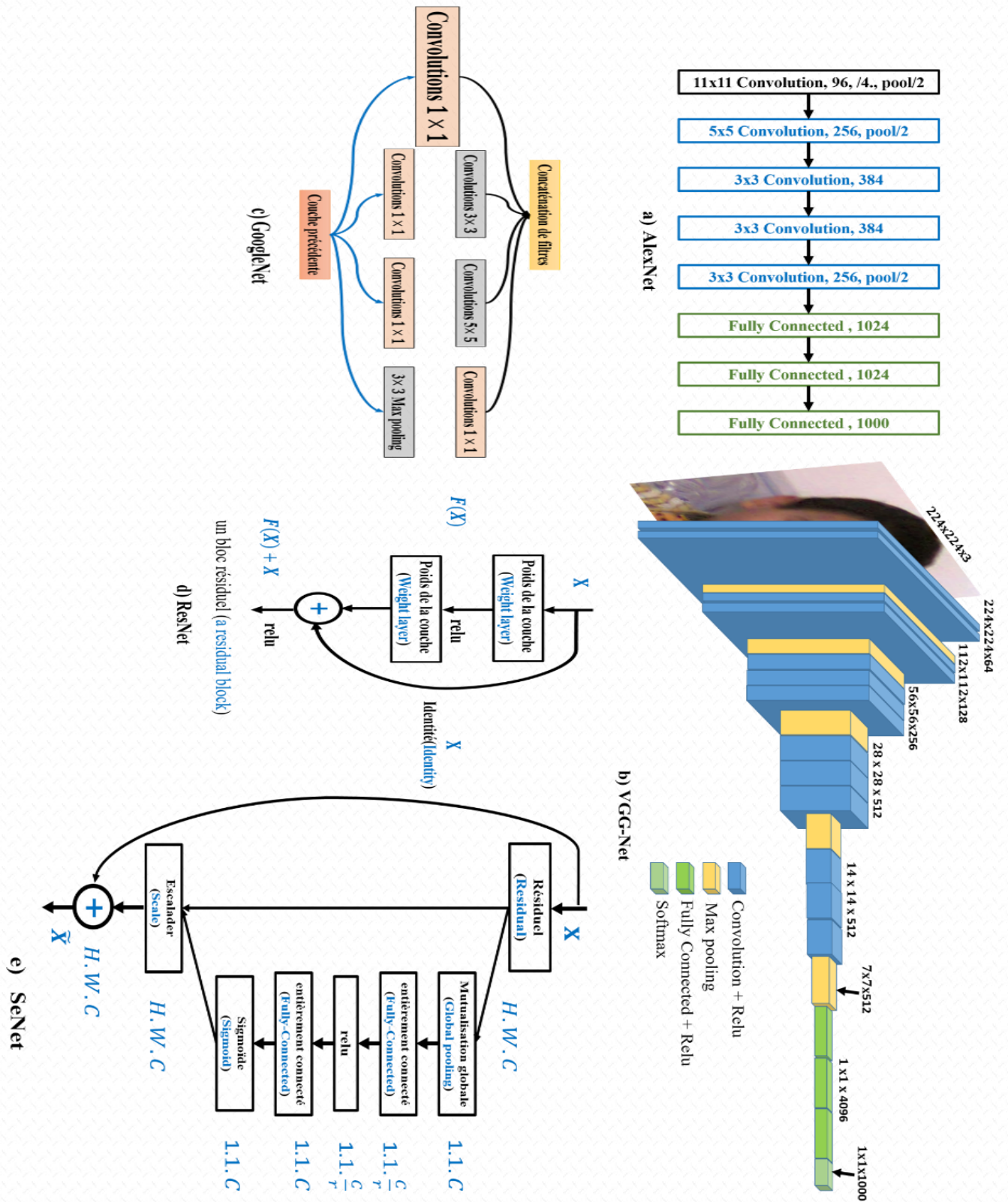


Fig 3. 2. Architectures [64]: a) AlexNet, b) VGG-Net, c) GoogleNet, d) ResNet, e) SeNet.

L'objectif principal d'une architecture de réseau Dorsale Internet "**Backbone network**" en anglais est d'extraire des caractéristiques robustes à partir d'une image faciale d'entrée. Les principales architectures d'extraction de caractéristiques ont toujours été celles qui ont extrêmement bien fonctionné sur les problèmes généraux de classification d'objets. AlexNet [22] a atteint une précision de reconnaissance de pointe lors du concours de reconnaissance visuelle à grande échelle ImageNet (ILSVRC), 2012 en battant les modèles précis de l'époque par une large marge. AlexNet est composé de cinq couches de convolution et de trois entièrement connectées au ReLU en tant que fonction d'activation, ainsi que de couches Dropout et d'augmentation des données pour augmenter encore les performances sur l'ensemble de test. En 2014, VGGNet [65] a proposé une architecture composée de noyaux de convolution 3x3, et des cartes de caractéristiques supplémentaires ont été utilisées après la couche de regroupement 2x2 max. En 2015, GoogleNet à 22 couches [66] a introduit un module de création (en anglais "Inception Module") avec la concaténation de cartes d'entités hybrides, ainsi que deux signaux intermédiaires supplémentaires supervisés Softmax. Il effectue plusieurs circonvolutions avec différents champs récepteurs (1×1 , 3×3 et 5×5) en parallèle, et concatène toutes les cartes d'entités pour fusionner les informations multi-résolution. Des gains de performance significatifs ont également été observés grâce à l'emploi d'une telle structure. Afin de former des réseaux neuronaux extrêmement profonds, peut-être de l'ordre de 150 couches, les chercheurs ont mis au point le modèle ResNet [67]. L'idée clé était l'utilisation d'une cartographie résiduelle afin de lutter contre les limitations antérieures découlant du problème du gradient de fuite (vanishing gradient problem). SENet [68], le lauréat ILSVRC 2017, a développé un bloc "Squeeze-and-Excitation" (SE) qui s'adapte aux réponses des caractéristiques par canal en modélisant explicitement les interdépendances des canaux. Ces blocs peuvent être utilisés avec architectures actuelles telles que ResNet pour renforcer leur capacité de représentation.

3) Perte d'entraînement (Training Loss) pour l'appariement du visage

Les premières architectures de réseau comme DeepFace [56] et DeepID [69] utilisaient la perte Softmax pour apprendre les fonctionnalités et former le modèle. Les chercheurs ont réalisé que de telles fonctions de perte limitaient intrinsèquement la capacité

d'apprentissage des caractéristiques discriminantes. Par conséquent, davantage d'études ont commencé à se concentrer sur les fonctions de perte pour améliorer la précision du système de reconnaissance de visage. Au cours de cette période, la perte euclidienne et la perte en cosinus ont joué un rôle vital dans la transformation globale.

- **Perte euclidienne** : la perte euclidienne [70] ; [71] est une métrique d'apprentissage qui intègre une image faciale donnée dans l'espace représentatif. L'idée de base derrière la perte euclidienne est de réduire la variance Intra-classe et d'augmenter la variance Inter-classe entre les images faciales de différentes personnes. Les versions les plus courantes de la perte euclidienne sont la perte de triplet et la perte contrastive. Perte de triplet [32] ; [72] - [76] considère la différence relative entre les images faciales.

L'équation de perte du triplet est la suivante :

$$L(A, P, N) = \max(\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha, 0) \dots \dots \dots (3.1)$$

Où A est connue comme l'image d'ancrage, P est l'image positive et appartient à la même classe que A , et N est l'image négative qui appartient à une classe différente de A et N . $f(\bullet)$ est une fonction qui convertit un image d'entrée à son vecteur d'intégration respectif. Dans la perte contrastive [77] - [81], la distance euclidienne des paires d'images positives est inférieure et la distance des paires d'images négatives est supérieure.

L'équation de la perte contrastive est la suivante :

$$L = y_{ij} \max(0, \|f(x_i) - f(x_j)\|^2 - \epsilon^+) + (1 - y_{ij}) \max(\epsilon^- - \|f(x_i) - f(x_j)\|^2, 0) \dots \dots (3.2)$$

Où $y_{ij} = 0$ signifie que x_i et x_j sont des échantillons non concordants et $y_{ij} = 1$ signifie des échantillons concordants. $f(\bullet)$ est une fonction qui convertit une image en son vecteur d'intégration correspondant, ϵ^+ contrôle la marge des paires correspondantes et ϵ^- contrôle la marge des paires non correspondantes.

- **Perte en cosinus** : après le développement des fonctions de perte initiales, les chercheurs ont décidé de se concentrer sur le raffinement des fonctions de perte afin de développer un système de reconnaissance de visage robuste. Par conséquent, le cosinus ou la perte angulaire [82] - [86] est entré en jeu pour une amélioration supplémentaire de la précision.

$$L_s = \frac{1}{N} \sum_{i=1}^N -\log \frac{e^{f_{y_i}}}{\sum_{j=1}^C f_j} \dots \dots \dots (3.3)$$

Où N est le nombre d'échantillons d'apprentissage et C est le nombre de classes. f est noté comme l'activation de la couche entièrement connectée avec le vecteur de poids W .

$$f_j = \|W\| \|x\| \cos \theta_j \dots \dots \dots (3.4)$$

Où θ est l'angle entre W et x . La norme et l'angle des vecteurs ont tous deux une influence sur la probabilité a posteriori.

3.4.2. Bases de données visage utilisées dans les modèles DL

De nombreuses bases de données de visages ont été créées au cours des trois dernières décennies, avec une tendance claire de la petite à la grande échelle, de la source unique à la diversité des sources, et des conditions contrôlées en laboratoire aux conditions réelles sans contraintes. La procédure de génération de la base de données faciale influence fortement le chemin de la recherche en reconnaissance de visage.

Un ensemble de données d'entraînement suffisamment volumineux est requis pour une reconnaissance de visage approfondie réussie. Un ensemble de données de reconnaissance de visage à grande échelle est en cours de création pour la communauté de la recherche afin de combler le fossé des données qui existe entre l'industrie et la communauté de la recherche. MS-Celeb-1M [87] - [88], VGGface [72].

Megaface [89] - [90] et WebFace260M [91] sont quatre bases de données contenant plus d'un million [92] de photos.

La VGGFace2 [72] ; [93] est un nouvel ensemble de données de visage à grande échelle, contient 3,31millions de photographies de 9131 personnes, avec une moyenne de 362,6prises de vue pour chaque sujet. Les photos ont été rassemblées à l'aide de Google Image Search et incluent une variété de poses, d'ethnies, d'âges, d'éclairages et d'emplois (par exemple, acteurs, athlètes, politiciens). L'ensemble de données a été compilé avec trois objectifs en tête : (a) un grand nombre d'identités avec des photographies pour chacune; (b) un large éventail d'âges, de postures et d'ethnies ; et (c) un faible niveau de bruit d'étiquette.

MS-Celeb-1M [87], s'est fixé pour objectif de reconnaître un million de célébrités à partir de leurs photographies faciales et de les identifier à l'aide de clés d'entité uniques dans une base de connaissances. Les informations complètes de la base de connaissances aident à lever les ambiguïtés et améliorent la précision de la reconnaissance, tout en contribuant à une variété d'applications réelles telles que le sous-titrage d'images et analyse de vidéos d'actualités.

WebFace260M [91] est une nouvelle référence de visage à l'échelle d'un million qui comprend des données d'entraînement telles que des identités bruyantes [94] 4M/260M visages (WebFace260M) et des identités 2M nettoyées/42M visages (WebFace42M), ainsi qu'un temps -méthode d'évaluation contrainte. Le prétraitement du visage est réalisé par cinq points de repère utilisés pour reconnaître et aligner les visages qui sont prédits par Retina Face [36]. Les statistiques de WebFace260M sont affichées en tant que date de naissance, nationalité et profession. L'ensemble d'apprentissage WebFace260M est automatiquement nettoyé à l'aide de la méthode CAST et l'ensemble de formation WebFace42M est créé. Les annotations d'attributs de visage dans WebFace42M incluent la posture, la race, l'âge, le chapeau, le verre, le masque et le sexe.

Tableau 3.3. Résultats de quelques travaux liés aux méthodes DL avec optimisation

Références	Base de données	Méthodes	TR(%)
Elaggoune H et al. [45]	ORL	PSO_AlexNet_TL	99.17
	AR		99.86
	Thermal Tufts Face		83.41
	LFW		99.57
Di X et al. [55]	Thermal Tufts Face	Multi-AP-GAN (GT)	77.38
Damer N et al. [95]	Thermal Tufts Face	CRN + CL	74.90
Sasirekha K et al. [96]	ORL	RNN_PSO	94.07
Talab MA et al. [97]	ORL	ESPCN + CNN [97]	92.24
Muhammad HI et al. [98]	ORL	PCA-FRCNN	99.00
Almabdy S et al. [99]	ORI	Alex Net + SVM	99.17
Cheng EJ et al. [100]	AR	Deep Learning + softmax	94.42
Manzuri M et al. [101]	AR	Metaheuristic algorithms (IC)	95.30
Deng J et al. [102]	LFW	ArcFace-CNN	99.50
Yin X et al. [103]	LFW	Transfer Learning-Auto encoder	98.83

3.5. Conclusion

La technologie de reconnaissance de visage s'est avérée être l'une des technologies les plus percutantes de ces dernières décennies. Ce chapitre couvre diverses architectures de reconnaissance de visage et procède en couvrant divers paramètres nécessaires à la formation de tels modèles. Tout d'abord, une introduction générale à la reconnaissance de visage et à ses ramifications est présentée. Deuxièmement, un aperçu général du pipeline de reconnaissance de visage est donné et commence par l'élaboration du premier aspect, le prétraitement et extraction de caractéristiques du visage.

Ensuite un état de l'art sur les méthodes d'optimisation liée à la reconnaissance de visage est mis en revue.

Dans la dernière partie du chapitre nous mettons l'accent sur diverses architectures de réseaux utilisées pour l'extraction de caractéristiques profondes et de différentes fonctions de perte utilisées pour former de tels modèles et également pour l'appariement des visages. En outre, il donne des informations sur diverses bases de données d'identification de visage utilisées par la communauté universitaire pour comparer la métrique de précision par rapport à d'autres résultats proposés par d'autres études contemporaines.

4. Optimisation des descripteurs hybrides et Patches pour la RV

4.1. Introduction

Diverses techniques sont en cours de développement, y compris des approches locales et hybrides, qui fournissent une description de l'image du visage en utilisant des parties (patches) de l'image du visage ou l'image complètement [7].

Les premiers travaux orientés vers la reconnaissance de visage ont été achevés dans les années 1950 en psychologie [12]. Ils sont venus afin de résoudre les problèmes liés à l'expression du visage, à l'interprétation de l'émotion ou à la perception des gestes.

Au cours de ces dernières décennies, l'intérêt pour le développement des algorithmes de reconnaissance de visage a augmenté rapidement [104]. La vidéosurveillance, l'identification criminelle, le contrôle d'accès aux bâtiments et les véhicules sans pilote et autonomes ne sont que quelques exemples d'applications concrètes qui gagnent en popularité auprès des industries [105].

Les patches du visage ont de nombreuses applications en vidéosurveillance. Bien que l'extraction et la sélection des caractéristiques aient été bien étudiées, la spécificité de chaque variation d'expression n'est pas entièrement explorée dans les travaux de pointe.

Inspirées par les résultats de l'analyse du visage selon lesquels le visage humain est une classe d'objets hautement structurés, et que par conséquent la position joue un rôle de plus en plus important dans sa reconstruction, certaines méthodes de reconnaissance de visage basées sur des patches ont été proposées. Le problème clé de ces méthodes est de savoir comment représenter le patch d'image et obtenir le vecteur de poids optimal [106]. Par exemple, **Ma X et al.** [107] utilisent des patches à la même position de toutes les images de visage d'entraînement pour halluciner le patch d'image en résolvant un problème des moindres carrés.

Dans ce chapitre, nous présentons des algorithmes de reconnaissance de visage, utilisant une approche de sélection de caractéristiques basée sur la méthode d'optimisation PSO.

Notre approche utilise l'application de descripteurs hybrides composée d'un mélange de descripteurs sélectionnés (filtre de Gabor, HOG, LPQ, PCA) exploités pour l'extraction des caractéristiques de visages. Le vecteur des caractéristiques obtenu sera ensuite traité par un algorithme d'optimisation appelé (PSO)" pour une amélioration significative.

La classification est effectuée par la méthode (projection linéaire de sous-espace) utilisant la mesure de distance cosinus avec la normalisation des données. Nos expérimentations sont réalisées sur un ensemble de données du monde réel : LFW, ainsi que la bases de données Yale B étendue.

La principale valeur ajoutée de ce travail est :

- Elaboration d'un nouvel algorithme de sélection de fonctionnalités pour la reconnaissance de visage basé sur le descripteur hybride associé à l'algorithme PSO. Cet algorithme s'applique à Gabor avec des vecteurs de caractéristiques HOG (LPQ ou PCA) pour rechercher le sous-ensemble de caractéristiques optimal pour augmenter le taux de reconnaissance et la séparation des classes.
- Proposition d'optimisation de l'approche basée sur la représentation des patches pour détecter la région active d'intérêt de l'échantillon de test.
- Évaluation de l'algorithme proposé à l'aide des bases de données de visages LFW et Yale B étendue et comparaison de ses performances avec un algorithme de sélection de caractéristiques basé sur PSO et divers algorithmes RV dans la littérature.

Le reste de ce chapitre est organisé comme suit :

La section 2 présente la méthodologie suivie dans la conception de notre système proposé. Un aperçu des bases de données ciblées à étudier se trouve dans la section 3. La section 4 présente les résultats expérimentaux et leurs discussions. Finalement, une conclusion générale est présentée dans la section 5 de ce chapitre.

4.2. Méthodologie

Le prétraitement des images est nécessaire pour déterminer ou détecter que l'image représente réellement un visage et son emplacement (Fig 4. 1). La deuxième étape est l'extraction des caractéristiques, qui comprend la réduction de la taille, l'extraction de la saillance et le nettoyage du bruit. Ensuite, l'image sera transformée en un vecteur de dimension basé sur le nombre de caractéristiques extraites [108], [109]. L'étape de classification concerne l'utilisation de classificateurs pour l'identification et la vérification.

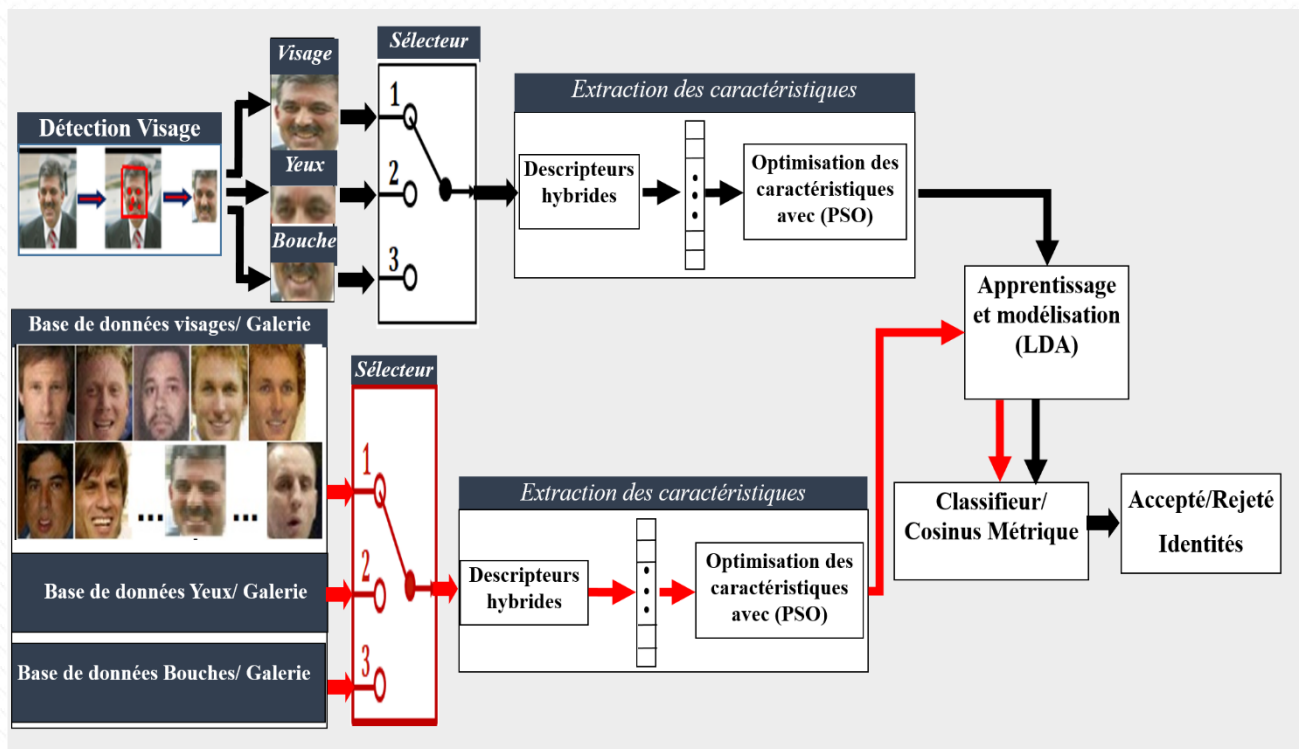


Fig 4. 1. Schéma de descripteurs hybrides proposé et de l'optimisation de patches pour le système RV.

4.2.1. Extraction des caractéristiques

1) Filtre de Gabor

Un filtre de Gabor est capable de générer des informations à partir d'une image de texture à différentes échelles et orientations. Généralement, le filtre de Gabor utilisé en

reconnaissance de visage est un filtre linéaire dont la réponse impulsionnelle est une sinusoïde modulée par une fonction gaussienne (aussi appelée ondelette de Gabor) :

$$\Psi_{\theta,\lambda}(z) = \frac{\|k_{\theta,\lambda}\|}{\sigma^2} e^{-\frac{\|k_{\theta,\lambda}\|^2 \|z\|^2}{2\sigma^2}} \left[e^{ik_{\theta,\lambda}z} - e^{-\frac{\sigma^2}{2}} \right] \quad (4.1)$$

Où $z = (x, y)$ est le point de coordonnées (x, y) dans le plan image. Les paramètres θ et λ définissent l'orientation et la fréquence des noyaux de Gabor. $\| \cdot \|$ est l'opérateur norme et σ est l'écart type de l'enveloppe gaussienne. Le paramètre d'orientation θ (*ou o*) fait tourner l'ondelette autour de son centre. Il détermine l'angle des contours ou lignes de l'image auquel l'ondelette sera sensible. Le paramètre λ (*ou r*) précise la longueur d'onde du cosinus ou inversement la fréquence centrale de l'ondelette. Les ondelettes de grande longueur d'onde seront sensibles aux changements graduels d'intensité dans une image. Les ondelettes de petite longueur d'onde seront sensibles aux contours et aux lignes abruptes (4.1). La convolution de l'image I et d'un noyau de Gabor $\Psi_{\theta,\lambda}(z)$ est définie par :

$$G_{\theta,\lambda}(z) = I(z) * \Psi_{\theta,\lambda}(z) \quad (4.2)$$

Comme on peut le remarquer dans l'équation (4. 1), les filtres de Gabor ont une forme complexe, il est important d'utiliser les informations données par la partie réelle et la partie imaginaire des coefficients de Gabor. Deux choix triviaux s'offrent à nous l'étude de l'**amplitude** et l'étude de la **phase** de Gabor. L'amplitude et la phase seront notées respectivement par :

$$\begin{cases} M(IG_{r,o}(x, y)) = \sqrt{I_m(IG_{r,o}(x, y))^2 + Real(IG_{r,o}(x, y))^2} \\ P(IG_{r,o}(x, y)) = \arctan\left(\frac{I_m(IG_{r,o}(x, y))}{Real(IG_{r,o}(x, y))}\right) \end{cases} \dots \dots \dots (4.3)$$

2) *Histogramme des Gradients orientés*

En anglais "Histogram of Oriented Gradients (HOG)", ont été proposés par Navneet Dalal et Bill Triggs, chercheurs à l'INRIA (Institut national de recherche en informatique et en automatique) de Grenoble, à la conférence CVPR de juin 2005 [110].

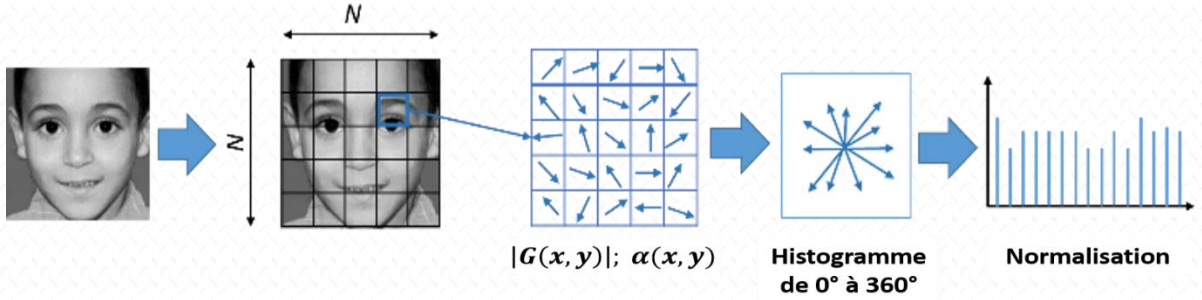


Fig 4.2. HOG extraction des caractéristiques [111].

Il a été développé et étendu à de nombreuses applications biométriques et pour la vision par ordinateur [111]. Le HOG décrit l'aspect et la forme locale de l'objet dans une image en utilisant la distribution de gradients. Pour obtenir Le vecteur caractéristique d'une image par la technique HOG $I(x, y)$ en va suivre les procédures suivante :

- La division de l'image $I(x, y)$ en N cellules régulières et des blocs. Les valeurs de gradient sont calculées pour chaque pixel en utilisant un filtre dérivatif 1-D centré, dans les directions horizontales et verticales. Pour cela on utilise les masques :

$$D_x = [-1 \ 0 \ 1]; \quad D_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad (4.5)$$

$$\begin{cases} G_x(x, y) = I(x, y) * D_x \\ G_y(x, y) = I(x, y) * D_y \end{cases} \quad (4.6)$$

- Le calcul de l'amplitude et l'orientation de gradient à chaque pixel (x, y) , s'effectue par les équations. (4.7) et (4.8), avec $G_x(x, y)$, et $G_y(x, y)$ sont le gradient horizontale et le gradient verticale, du pixel (x, y) , respectivement :

$$|G(x, y)| = \sqrt{(G_x(x, y))^2 + (G_y(x, y))^2} \quad (4.7)$$

$$\alpha(x, y) = \arctan\left(\frac{G_y(x, y)}{G_x(x, y)}\right) \quad (4.8)$$

- L'histogramme de l'orientation du gradient à l'intérieur de chaque cellule est calculé par quantification Gradients non signés à chaque pixel en 12 canaux (bins) d'orientations. Les histogrammes sont uniformes de 0 à 180° (cas non signé) ou de 0 à 360° (cas signé).
- La normalisation du vecteur caractéristique de chaque cellule s'effectue en utilisant des vecteurs dans leurs blocs respectifs. Exemple de la norme L2-norme, son facteur de normalisation est calculé par l'équation :

$$v_n = \frac{v}{\sqrt{\|v\|_2^2 + \epsilon^2}} \quad (4.9)$$

v est le vecteur non normalisé contenant tous les histogrammes dans un bloc, $\|v\|_2$ est la L2-norme du vecteur descripteur, et ϵ est une constante.

- Le vecteur caractéristique HOG sera formée de la concaténation de tous les vecteurs caractéristiques de tous les blocs pour une image donnée.

3) Optimisation à Essaim de Particules

L'optimisation à essaim de particules en anglais "Particle Swarm Optimisation"(PSO) a été développée par James Kennedy (psychosocial), Bureau des statistiques, Département américain du Travail et Russell Eberhart (ingénieur électricien), Université Purdue, en 1995 [112].

L'optimisation à essaim de particules (PSO) est une méthode qui a trouvé son application dans le domaine de l'optimisation continue sans fonction linéaire qui simule des comportements dits sociaux. C'est un algorithme extrêmement efficace et simple pour optimiser un large éventail de fonctions [113], [114]. Il convient de noter que l'algorithme PSO peut être utilisé dans n'importe quel espace dimensionnel, 2D et 3D ... etc. Le but de l'algorithme est de trouver la meilleure évaluation d'une fonction de fitness donnée. Puisque chaque particule représente une position dans l'espace de dimension D, l'objectif est alors d'ajuster sa position selon :

- La meilleure position de la particule trouvée jusqu'à présent, et la meilleure position au voisinage de cette particule. Pour compléter les déclarations précédentes, chaque particule stocke ces valeurs :
- X_i : position actuelle, V_i : sa vitesse actuelle et P_i : sa meilleure position trouvée jusqu'ici.

La formule originale du PSO définit chaque particule comme une solution potentielle à un problème dans l'espace de $D - dimension$. La position de la particule i est représentée comme suit :

$$X_i = (x_{i1}, x_{i2}, \dots, x_{iD}) \quad (4.10)$$

Chaque particule retient tout au long de son déplacement la position de la meilleure solution $lbest$, représentée par :

$$P_i = (p_{i1}, p_{i2}, \dots, p_{iD}) \quad (4.11)$$

La meilleure position est mise à jour à chaque itération définit par l'équation :

$$P_i(t+1) = \begin{cases} P_i(t) & \text{si } f(X_i(t+1)) \geq f(P_i(t)) \\ X_i(t+1) & \text{si } f(X_i(t+1)) < f(P_i(t)) \end{cases} \quad (4.12)$$

Où le symbole f désigne la fonction de fitness; $P_i(t)$ représente la meilleure valeur de fitness et la coordonnée où la valeur a été calculée, t désigne l'étape de génération. Une particule dans un essaim est en mouvement ; par conséquent, il y a une vitesse, qui peut être représentée comme suit :

$$v_i = (v_{i1}, v_{i2}, \dots, v_{iD}) \quad (4.13)$$

Pour le modèle $gbest$, la meilleure particule est déterminée à partir de l'essaim entier. Si la position de la meilleure particule est dénotée par le vecteur P_g alors:

$$P_g \in \{P_1, P_2, \dots, P_N\} \\ f(P_g(t)) = \min\{f(P_1(t)), f(P_2(t)), \dots, f(P_N(t))\} \quad (4.14)$$

Où N est le nombre total de particules dans l'essaim.

A chaque itération, les particules modifient leur vecteur de déplacement en direction de la position de $lbest$ (P_i) et de la position de $gbest$ (P_g) selon les équations suivantes:

$$V_i(t + 1) = wV_i(t) + c_1r_1(P_i - X_i(t)) + c_2r_2(P_g - X_i(t)) \quad (4.15)$$

$$X_i(t + 1) = X_i(t) + V_i(t + 1) \quad (4.16)$$

Où w est le facteur d'inertie, c_1 et c_2 sont les constantes d'accélération positive, avec $c_1 + c_2 \leq 4$, r_1 et r_2 sont deux variables aléatoires uniformément distribuées dans l'intervalle $[0,1]$. L'algorithme *PSO* s'exécute jusqu'à un nombre d'itérations fixe ou une variation de vitesse proche de zéro.

L'aperçu du PSO utilisé pour sélectionner les meilleures caractéristiques du visage est donné dans l'algorithme (Algorithme 4.1).

Algorithme 4.1. L'optimisation par essaim de particules

Début

Initialiser une population de particules aléatoirement en positions et vitesses de l'espace de recherche D-dimension.

Tant que la condition d'arrêt n'est pas atteinte **faire**

Pour chaque particule i des particules générées.

Adapter la vitesse de la particule en utilisant l'équation (4.15);

Ajuster la position de la particule en utilisant l'équation (4.12) ;

Evaluer la fitness $f(X_i)$;

Mettre à jour la meilleure particule locale P_i et globale P_g ;

Fin pour

Fin tant que

Retourner et visualiser la meilleure solution

Fin

4) Quantification de Phase Locale

En anglais " Local Phase Quantization (LPQ) ", est un descripteur local qui extrait des caractéristiques de la transformée de Fourier discrète (DFT) sur une image [115].

Dans LPQ, la phase est examinée dans les voisinages locaux N_x à chaque position de pixel x de l'image $f(x)$. Ces spectres locaux sont calculés en utilisant une transformée de Fourier à court terme définie par [116]-[118] :

$$F(u, x) = \sum_{y \in N_x} f(y) w_R(y - x) e^{-j2\pi u^T y} \quad (4.17)$$

Où $x \in \{x_1, x_2, \dots, x_N\}$ définit chaque pixel contiennent en une simple convolution 1-D pour les lignes et les colonnes successivement.

Les coefficients de Fourier locaux sont calculés à quatre points de fréquence $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u_3 = [a, a]^T$, et $u_4 = [a, -a]^T$, où a est un scalaire suffisamment petit pour satisfaire $H(u_i) > 0$ pour chaque position de pixel, il en résulte un vecteur :

$$F(x) = [F(u_1, x), F(u_2, x), F(u_3, x), F(u_4, x)] \quad (4.18)$$

En observant les signes des parties réelles et imaginaires de chaque composant de $F(x)$, les informations de phase peuvent être comptées à l'aide d'une simple quantification scalaire:

$$q_1 = \begin{cases} 1 & \text{pour } g_j \geq 0 \\ 0 & \text{pour } g_j \leq 0 \end{cases} \quad (4.19)$$

Où g_j est la $j^{\text{ième}}$ composante du vecteur :

$$G(x) = [Re\{F(x)\}, Im\{F(x)\}] \quad (4.20)$$

Alors, l'étiquette de l'image $f_{LPQ}(x)$ est représentée par:

$$f_{LPQ}(x) = \sum_{j=1}^8 q_j(x) 2^{j-1} \quad (4.21)$$

5) Analyse en composantes principales

En anglais "principal component analysis (PCA)", aussi connue sous le nom de transformée de Karhunen–Loève est une méthode très utilisée en statistique. Introduite par Pearson puis plus tard par Hotelling [7], c'est une technique descriptive permettant d'étudier les relations entre variables quantitatives, sans prendre en compte, a priori, aucune structure, variables ou individus. L'ACP est une approche statistique qui utilise une transformation orthogonale pour convertir un ensemble de variables corrélées en ensembles de variables non corrélées appelées composantes principales. Ainsi, l'ACP permet une réduction dimensionnelle des variables. Depuis les années 1990 jusqu'à aujourd'hui, la méthode PCA a pris sa place dans le développement des algorithmes de reconnaissance de visage [119], [120].

4.2.2. Apprentissage et Modélisation

1) Analyse Discriminante Linéaire

En anglais "linear discriminant analysis (LDA)", L'algorithme LDA est né des travaux de Belhumeur et al. De Yale University (USA), en 1997 [17]. Il est aussi connu sous le nom de "*Fisherfaces*".

LDA est une approche probabiliste qui fait partie des techniques d'analyse discriminante prédictive. Elle consiste à regrouper les données où une donnée sera considérée proche de l'autre si elle provient de la même densité (loi) de probabilité plutôt qu'en effectuant un simple calcul de distance déterministe. Chaque classe de données pourra donc être résumée par les paramètres de sa loi de probabilité. Chaque classe est supposée être Gaussienne [7].

Considérons par exemple une image couleur, chaque image contient n pixels, chaque pixel contient d descripteurs **RGB**. On représente le i ème pixel ($i = 1, \dots, n$) par un vecteur \mathbf{x}_i de dimension d : $\mathbf{x}_i = (x_{i1}, \dots, x_{ij}, \dots, x_{id})^T \in \{0, 1, \dots, 255\}^d$.

La loi Gaussienne est simplement décrite par sa moyenne μ_k et sa matrice de covariance et Σ est donnée par :

$$f_{ik}(\mathbf{x}_i; \mu_k; \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{\left(-\frac{1}{2}(\mathbf{x}_i - \mu_k)^T \Sigma^{-1} (\mathbf{x}_i - \mu_k)\right)} \quad (4.22)$$

2) Apprentissage du modèle LDA

L'apprentissage du modèle LDA consiste en l'estimation des paramètres μ_k et Σ à partir des données d'apprentissage. Ces estimations sont données par :

$$\pi_k = \frac{\sum_{i|z_i=k} 1}{n} = \frac{n_k}{n} \quad (4.23)$$

Où n_k est le cardinal de la classe k , z_i est l'étiquette de classe de l'exemple x_i et $z_i = k$ désigne le fait que l'exemple x_i appartient à la classe k .

L'estimation des paramètres μ_k (les moyennes) et Σ (matrice de covariance) à partir des données d'apprentissage sont données par :

$$\mu_k = \frac{1}{n_k} \sum_{i|z_i=k} x_i \quad (4.24)$$

Et

$$\Sigma = \frac{1}{n - K} \sum_{k=1}^K \sum_{i|z_i=k} (x_i - \mu_k)(x_i - \mu_k)^T \quad (4.25)$$

3) Test du modèle LDA

Une fois les paramètres sont estimés, on peut prédire les classes des données de test par la règle du MAP (maximum a posteriori). Cette règle consiste à maximiser les probabilités a posteriori, c.-à-d. affecter chaque donnée de test x_i à la classe \hat{z}_i ayant la plus grande probabilité a posteriori :

$$\tau_{ik} = \mathbb{P}(z_i = k | x_i; \mu_k, \Sigma) = \frac{\pi_k f_{ik}(x_i; \mu_k, \Sigma)}{\sum_{l=1}^K \pi_l f_{il}(x_i; \mu_l, \Sigma)} \quad (4.26)$$

Et on a donc

$$\hat{z}_i = \arg \max_{k=1}^K \tau_{ik} \quad (i = 1, \dots, n) \quad (4.27)$$

4.3. Bases de Données

4.3.1. Base de données de visages Yale étendue B

Il s'agit d'une version mise à jour de la base de données Yale de visage B. Il s'agit de 38 sujets avec 9 expositions et 64 conditions d'éclairage. Les images sont divisées en cinq sous-ensembles en fonction de l'angle entre la direction de la source lumineuse et l'axe de la caméra centrale (Fig 4. 3 (a)). Nous réduisons la variation d'éclairage avec la technique d'égalisation d'histogramme (voir Fig 4. 4).



Fig 4. 3. Exemples d'images de la base de données Yale B et leur localisation des yeux et de la bouche

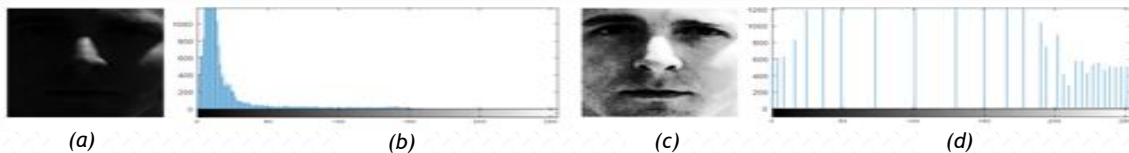


Fig 4. 4. (a) Image du visage (Yale B), (b) Histogramme du visage, (c) visage après égalisation, (d) Histogramme du visage égalisé.

4.3.2. Base de données LFW “Labeled Faces in the Wild”

C'est une grande base de données de photographies de visages étiquetées conçue pour étudier le problème de la reconnaissance des visages dans des environnements sans contraintes. L'ensemble de données contient 13233 images de visages recueillies sur le Web, représentant 5749 personnes, dont 1680 personnes ayant deux images ou plus. Chaque visage a été étiqueté avec le nom de la personne représentée.

La base de données est divisée en deux vues : la vue 1 utilisée pour la sélection du modèle et la vue 2 pour l'évaluation des performances. Elle peut être consultée et téléchargée à l'adresse : <http://vis-www.cs.umass.edu/lfw/>.

Des exemples d'images de visage de cette base de données sont présentés à la Fig 4. 5.



Fig 4. 5. Exemples d'images de la base de données LFW avec leurs localisation des yeux et de la bouche.

4.4. Résultats des descripteurs hybrides optimisés proposés

Dans nos expériences, pour l'algorithme du PSO, nous avons choisis une fonction "fitness" de la forme :

$$fun = mean(data(:)) \quad (4.28)$$

Qui a apporté des améliorations dans la matrice des caractéristiques. Concernant la base de données "LFW", elle a subi un prétraitement d'alignement par la méthode de la sélection de faces et de cinq points faciaux (voir Fig 4. 6).



Fig 4. 6. Exemple d'alignement en sélectionnant cinq points du visage [121].

Nous avons réalisé plusieurs tests, plusieurs descripteurs hybrides ont été élaborer dont les résultats sont présentés dans les figures et les tableaux en ce qui suit avec et sans l'algorithme PSO.

4.4.1. Résultats sans fusion multimodales

Les Tableau 4. 1 et Tableau 4. 2 présentent un aperçu sur les résultats du TR et du temps test obtenus pour les deux bases de données YALE-B et LFW. Ces résultats montrent que l'un des meilleurs résultats en termes de temps et de TR est obtenu par l'utilisation du

descripteur HOG pour l'extraction des caractéristiques. Les résultats montrent que chaque descripteur contribue à sa manière à fournir au système RV des informations efficaces pour une bonne identification. Les descripteurs hybrides Gabor + HOG et Gabor + LPQ restent de bons candidats lorsqu'ils sont optimisés par le PSO.

Tableau 4. 1. TR et Temps Test pour la base de données YALE-B

N°	Méthodes	Base de données YALE-B					
		Visage		YEUX		BOUCHE	
		TR (%)	Temps Test (s)	TR (%)	Temps Test (s)	TR (%)	Temps Test (s)
1	Gabor+HOG+PSO	99.58%	0.082	99.58%	0.059	99.45	0.042
2	Gabor+HOG	99.58	0.113	99.58	0.047	99.45	0.041
3	Gabor +LPQ+ PSO	98.89	0.287	99.17	0.094	98.75	0.093
4	Gabor+LPQ	99.58	0.379	99.72	0.140	99.72	0.134
5	HOG+PSO	99.58	0.077	99.17	0.038	99.17	0.035
6	HOG	99.58	0.156	99.58	0.045	99.58	0.049
7	LPQ+ PSO	99.72	0.402	99.03	0.124	98.61	0.149
8	LPQ	99.72	0.414	99.31	0.146	99.58	0.135
9	Gabor+HOG+ PCASVD	99.58	0.067	99.58	0.130	99.58	0.075
10	Gabor+HOG+ PCASVD +PSO	36.29	0.107	50.00	0.127	37.53	0.128
11	Gabor+HOG+PSO+PCASVD	99.45	0.110	99.58	0.127	99.45	0.130

Tableau 4. 2. TR et Temps Test pour la base de données LFW

N°	Méthodes	Base de données LFW					
		Visage		YEUX		BOUCHE	
		TR (%)	Temps Test (s)	TR (%)	Temps Test (s)	TR (%)	Temps Test (s)
1	Gabor+HOG+PSO	69.25	0.081	70.90	0.063	65.87	0.051
2	Gabor+HOG	69.23	0.150	71.13	0.141	66.20	0.148
3	Gabor +LPQ+ PSO	65.12	0.049	69.30	0.052	64.03	0.050
4	Gabor+LPQ	65.25	0.143	67.08	0.083	64.10	0.160
5	HOG+PSO	70.38	0.056	71.82	0.055	66.92	0.043
6	HOG	70.32	0.150	71.87	0.133	67.05	0.168
7	PSO	65.47	0.177	64.48	0.575	59.98	0.208

4.4.2. Résultats avec fusion multimodales

Selon les résultats enregistrés dans les tableaux : "Tableau 4. 3, et Tableau 4. 4", la remarque principale constatée est la notion de la modalité dominante, c'est-à-dire : la fusion des modalités prend le meilleur résultat des modalités exécuté chacune toute seule. Par exemple : pour l'expérience N° 1 du Tableau 4. 1, pour la base Yale-B avec l'approche "

Gabor+HOG+PSO ", le TR est de **99.58%** pour la modalité **Yeux**, et de **99.45%** pour la modalité **Bouche**, en contre partie pour la même approche dans les mêmes conditions d'exécutions de l'expérience avec les différentes fusions (**Fusion au niveau de Modalités**, **Fusion au niveau des Caractéristiques**, et **Fusion au niveau de l'optimisateur PSO**), le TR est de **99.72%**, **99.58%**, et **99.58%** pour les trois fusions respectivement. Les meilleurs résultats sont obtenus pour la Fusion **au niveau de Modalités**.

Tableau 4. 3. TR et Temps Test obtenu en Fusion des modalités Yeux & Bouche (BD : YALE-B)

N°	Méthodes	Fusion Yeux & Bouche					
		Fusion au niveau de Modalités		Fusion au niveau des Caractéristiques		Fusion au niveau de l'optimisateur PSO	
		TR (%)	Temps Test (s)	TR (%)	Temps Test (s)	TR (%)	Temps Test (s)
1	Gabor+HOG+PSO	99.72	0.0648	99.58	0.0674	99.58	0.1128
2	Gabor+HOG	99.58	0.1109	99.58	0.0572		
3	Gabor +LPQ+ PSO	99.31	0.1168	99.31	0.1025	98.20	0.1209
4	Gabor+LPQ	99.72	0.3284	99.72	0.1807		
5	HOG+PSO	99.45	0.0404	99.45	0.0366	99.17	0.0368
6	HOG	99.58	0.0476	99.58	0.0486		
7	LPQ+ PSO	99.58	0.1472	99.45	0.1728	99.03	0.1005
8	LPQ	99.72	0.1905	99.72	0.2604		
9	Gabor+HOG+ PCASVD	99.58	0.1771	99.58	0.0655		
10	Gabor+HOG+ PCASVD +PSO	37.95	0.0986	49.72	0.0550	11.63	0.2165
11	Gabor+HOG+PSO+PCASVD	99.58	0.0661	99.58	0.0838	99.58	0.1433

Tableau 4. 4. TR et Temps Test obtenu en Fusion des modalités Yeux & Bouche (BD : LFW)

N°	Méthodes	Fusion Yeux & Bouche					
		Fusion au niveau de Modalités		Fusion au niveau des Caractéristiques		Fusion au niveau de l'optimisateur PSO	
		TR (%)	Temps Test (s)	TR (%)	Temps Test (s)	TR (%)	Temps Test (s)
1	Gabor+HOG+PSO	71,53	0,231	72,21	0,166	64,7	0,280
2	Gabor+HOG	72,1	0,106	73,11	0,092		
3	Gabor +LPQ+ PSO	66,6	0,123	69,7	0,255	69,2	0,164
4	Gabor+LPQ	66,61	0,123	69,63	0,083		
5	HOG+PSO	72,48	0,192	72,86	0,195	65,15	0,270
6	HOG	72,96	0,090	73,28	0,117		

4.4.3. Comparaison de nos résultats avec ceux existants en littérature

Les [Tableau 4. 5](#), et [Tableau 4. 6](#) comparent les meilleurs taux obtenus par nos approches proposées à plusieurs méthodes existantes dans la littérature utilisant le PSO. Les résultats présentés dans ces tableaux montrent que le descripteur hybride Gabor+HOG+PSO surpasse toutes les méthodes mentionnées dans ces tableaux.

Tableau 4. 5. TR de visages avec YALE-B, comparaison avec d'autres méthodes existantes dans la littérature.

Méthodes	DWT+MPSO [51]	Complete GFC [122]	Relaxed LTP [123]	LBP and VLC [124]	Nos résultats Gabor+HOG+PSO
TR(%)	98.29	99.18	98.71	95.80	Visage : 99.58 Yeux : 99.58 Bouche : 99.45

Tableau 4. 6. TR de visages (%) Rang 1 avec LFW, comparaison avec d'autres méthodes existantes dans la littérature

		Méthodes						
Samma H & all [53]		HOG	LBP	GABOR	Hybrid		PSO	Modèle Hybride optimisé
TR(%)		43.75	37.50	40.00	50.00		43.75	50.00
Nos résultats		HOG	HOG+PSO	Gabor+HOG	Gabor+HOG+PSO	Gabor+LPQ	PSO	Gabor +LPQ+ PSO
TR(%)	Visage	70.32	70.38	69.23	69.25	65.25	65.47	65.12
	Yeux	71.87	71.82	71.13	70.90	67.08	64.48	69.30
	Bouche	67.05	66.92	66.20	65.87	64.10	59.98	64.03

De plus, d'après les [Fig 4. 7](#) et [Fig 4. 8](#), on peut voir que les résultats les plus précis sont obtenus par le descripteur hybride optimisé (Gabor+HOG+PSO).

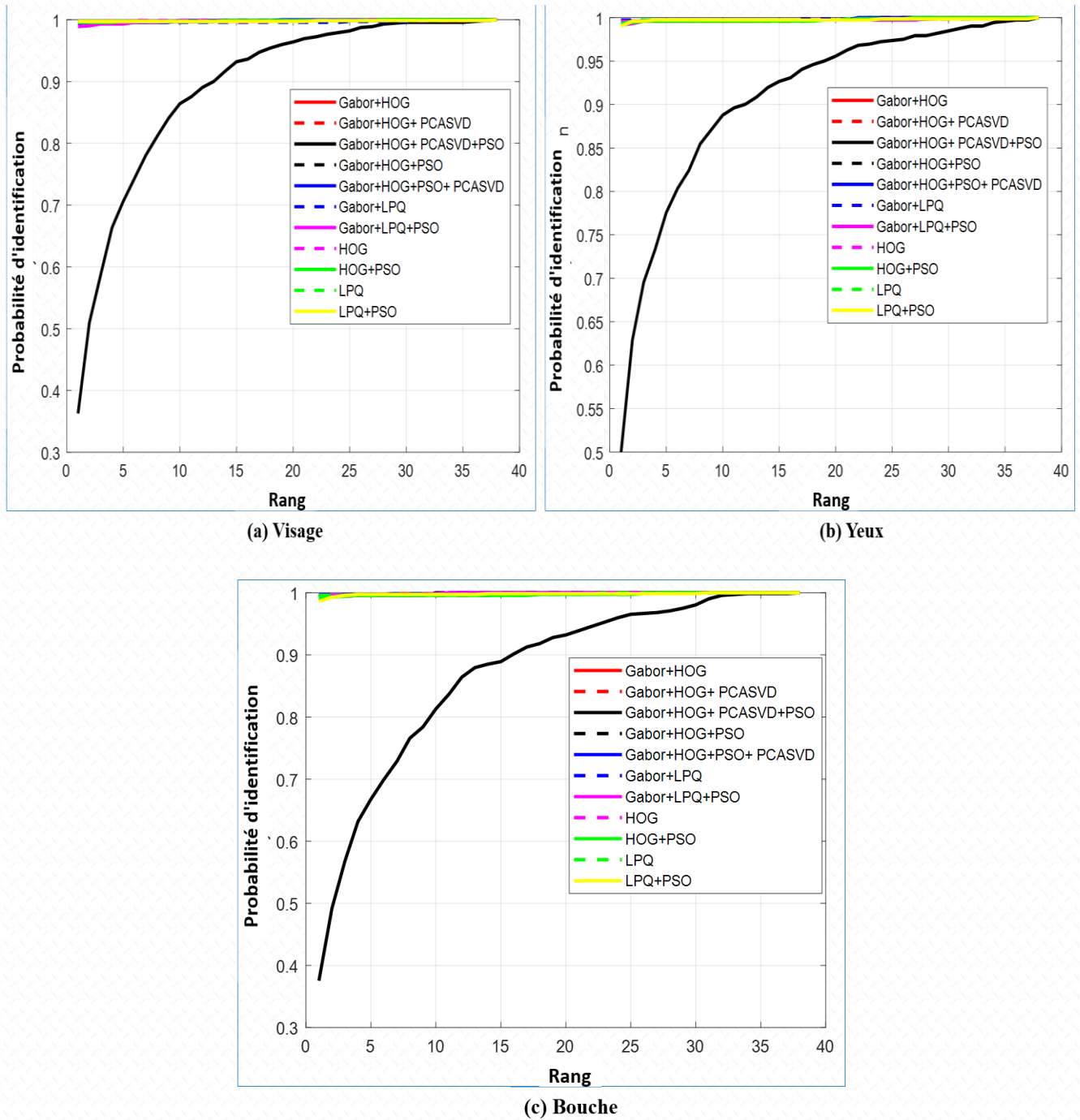


Fig 4. 7. Courbes CMC avec et sans l'algorithme PSO de YALE.

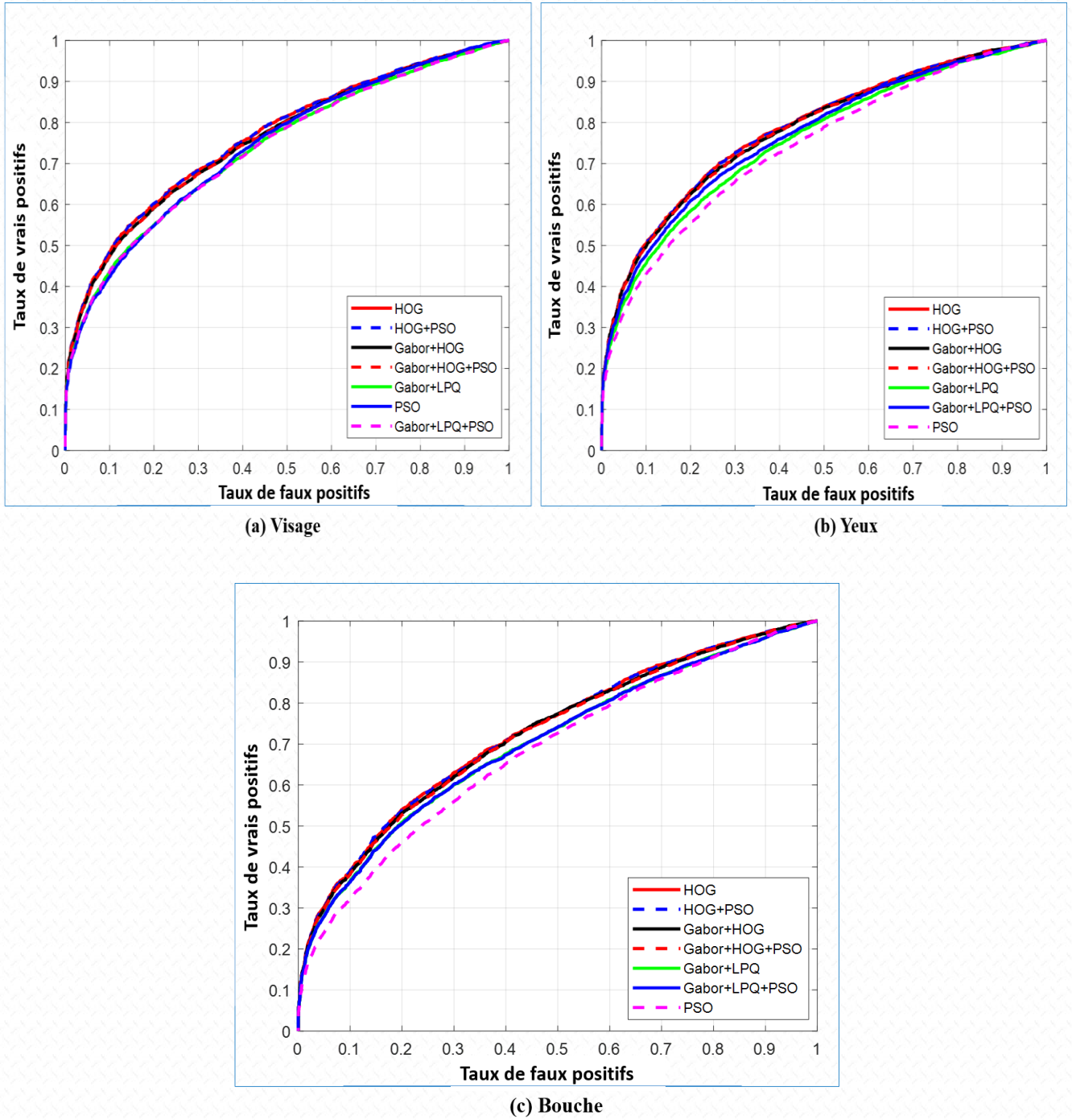


Fig 4. 8. Courbes ROC avec et sans l'algorithme PSO de LFW.

4.5. Conclusion

Dans ce chapitre nous avons démontré l'efficacité des descripteurs hybride dans les systèmes de reconnaissance de visage et ses patches.

Les résultats montrent que le bloc d'extraction des caractéristiques constitué de plusieurs descripteurs en cascade est très efficace. L'algorithme PSO joue un rôle très important dans l'optimisation des systèmes RV, notamment en termes de temps d'exécution, qui est un paramètre à prendre en considération.

La reconnaissance des yeux donne les meilleures TR et temps d'exécution par rapport aux bases de données de la bouche et du visage. Ce résultat qualifie la reconnaissance oculaire comme un bon moyen pour la reconnaissance visage.

5. Optimisation des caractéristiques de visages en utilisation des descripteurs hybrides et CNN en apprentissage par transfert

5.1. Introduction

L'un des systèmes sollicité dans la sécurité et la surveillance à identification biométrique pour accéder aux dispositifs sécurisés, on trouve les systèmes de reconnaissance de visages.

Malgré que le visage offre une accessibilité facile plus importante que la plupart des autres biométries, et qui n'implique pas une collaboration de l'utilisateur, et malgré les bons résultats encourageants obtenu, reste une biométrie encore peu fiable.

Le problème de contrôle optimal d'un système de reconnaissance de visages (lorsque les sources de changements sont multiples : les conditions d'éclairage, de pose, de position, d'expressions faciales, La présence d'attributs comme par exemple la présence de lunettes ou non, de maquillage, barbes, etc.) est un défi à soulever.

Les premiers travaux orientés vers la reconnaissance de visage ont été achevés dans les années 1950 en psychologie [47]. Ils sont venus dans le but de résoudre les problèmes attachés à l'expression du visage, l'interprétation de l'émotion ou la perception des gestes.

En 1964 et 1965, Bledsoe, avec Helen Chan et Charles Bisson, a travaillé sur l'utilisation de l'ordinateur pour reconnaître les visages humains. Comme ces travaux ont été effectués par diverses agences de renseignement mandatées par des contrats liés au Département de la Défense des États-Unis, qui n'a pas permis beaucoup de publicité, peu de travail a été publié [13]. Mais le premier qui a élaboré un système automatisé intégral était bien T. Kenade en 1973 [14].

Les premiers chercheurs qui ont utilisés la méthode basée sur l'Analyse en composantes principale ACP dans le traitement d'image sont L. Sirovich et M. Kirby en 1986. La technique qui deviendra l'approche de fondement de la proposition de nombreux nouveaux algorithmes de reconnaissance de visage dans les années suivantes [16], [125].

Depuis les années 1990, la reconnaissance de visage a connu un développement excessif, de nombreuses approches et algorithmes sont conçus parmi eux l'ACP, l'ICA, la LDA et leurs dérivés [126].

Le problème de l'optimisation de l'imagerie reste un problème d'actualité. Il persiste dans de nombreux travaux de recherche scientifique. En particulier, on peut citer l'algorithme d'optimisation PSO (Particle Swarm Optimization) qui a pris place dans le domaine de l'optimisation de l'imagerie. Il est devenu de plus en plus populaire [127], [128].

En effet, la véritable révolution qui avait bouleversé le monde s'est produite en 2012 avec le Deep Learning (DL). Il a réalisé une performance tout à fait incontournable dans la reconnaissance d'images, qui a marqué les scientifiques de l'époque [22].

L'une des principales différences entre le DL et les algorithmes traditionnels (tels que l'apprentissage automatique "Machine learning") est que leurs performances s'améliorent avec le volume et la richesse des données. De plus, dans les algorithmes traditionnels, l'extraction de caractéristiques se fait manuellement, consomme du temps et nécessite un spécialiste du domaine, alors qu'en DL l'algorithme effectue cette étape automatiquement.

L'année 2013, une grande partie de la recherche s'était concentrée sur l'approche Deep Learning et la machine est devenue surhumaine en matière de reconnaissance d'images.

Actuellement, l'orientation des chercheurs est vers le développement de systèmes de reconnaissance de visage basés sur Cascade CNN (Convolutional neural network) [129], [130], et l'élaboration de systèmes basés sur la structure Deep Learning et l'apprentissage par transfert [131] - [134].

La technique CNN est utilisée dans les problèmes de classification d'images, de reconnaissance de visage et de détection d'objets [135]. Ils ont donné une progression très

importante pour le calcul de la précision des points de référence difficiles par rapport aux jeux de données du monde réel [136].

Nos apports dans ce chapitre sont d'abord la recherche du meilleur descripteur hybride qui caractérise le visage. Puis la proposition de deux approches optimales :

- Descripteurs hybrides Optimisés
- CNN en apprentissage par transfert optimisé

Le principal ajout dans ce travail est le suivant :

- Elaboration d'une nouvelle approche de sélection de caractéristiques de visages. L'approche adoptée représente la fusion de plusieurs descripteurs biométriques en intégrant un algorithme d'optimisation, afin d'améliorer significativement les performances des systèmes basés sur un seul descripteur. Cela se fera en combinant les avantages de chaque descripteur.
- Application d'une nouvelle fonction coût "fitness", qui prend la forme d'une fonction sigmoïde dans l'algorithme d'optimisation PSO, et qui a donné de bons résultats.
- Application de plusieurs systèmes hybrides de reconnaissance de visages sur des bases de données en environnement contrôlé, non contrôlé et visages thermiques pour voir si ces systèmes conservent leurs performances et bénéfices.
- Développement et élaboration d'une structure d'optimisation de fonctionnalités basée sur la structure Deep Learning et Transfer Learning passant par l'algorithme PSO, s'appliquant à l'identification de visages.

Ce document est composé de quatre sections principales. Après avoir introduit la reconnaissance de visage et les techniques utilisées, nous fournissons notre approche proposée en combinant PSO avec l'approche DL. Ensuite, la section 3 présente l'implémentation et les bases de données utilisées. La section 4 présente les résultats obtenus et la discussion. Enfin, nous concluons le chapitre.

5.2. Approches proposées

5.2.1. Descripteurs Hybrides Optimisés

Cette approche utilise un algorithme basé sur la combinaison de plusieurs descripteurs populaires telle que le filtre de Gabor, HOG, LPQ, et PCA pour l'extraction des caractéristiques de visages. Pour sélectionner les paramètres caractéristiques optimaux, nous utilisons l'algorithme PSO avec la fonction coût "fitness" sous forme sigmoïde afin d'améliorer considérablement les performances de reconnaissance de visage du système. La classification pour l'identification/vérification des visages est effectuée par la méthode de projection de sous-espace linéaire en utilisant la métrique de distance cosinus avec la normalisation des données, voir Fig 5. 1.

La méthode du cosinus pour la similarité de deux vecteurs non nuls peut être dérivée de la formule euclidienne :

$$A \cdot B = \|A\| \|B\| \cos\theta \quad (5.1)$$

A et B , la similarité cosinus " $\cos(\theta)$ " est calculée par :

$$\text{Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (5.2)$$

Où A_i et B_i sont respectivement les composantes des vecteurs A et B .

Une étape de prétraitement d'image est nécessaire pour la détermination ou la détection que l'image représente réellement un visage et son emplacement voir Fig 5. 1. La deuxième étape est l'extraction des caractéristiques, qui comprend la réduction de la taille, l'extraction de la saillance et le nettoyage du bruit. Ensuite, l'image est projetée dans un sous-espace orthogonal optimal en fonction du nombre de caractéristiques extraites et sélectionnées. Certaines publications révèlent que l'extraction de caractéristiques entre dans la reconnaissance de visages [137]-[139]. L'étape de classification utilise un classificateur métrique cosinus pour l'identification et la vérification. Pour plus de détails sur l'extraction des caractéristiques et les descripteurs hybrides en question, voir le Chapitre : 4, Section "4.2 Méthodologie".

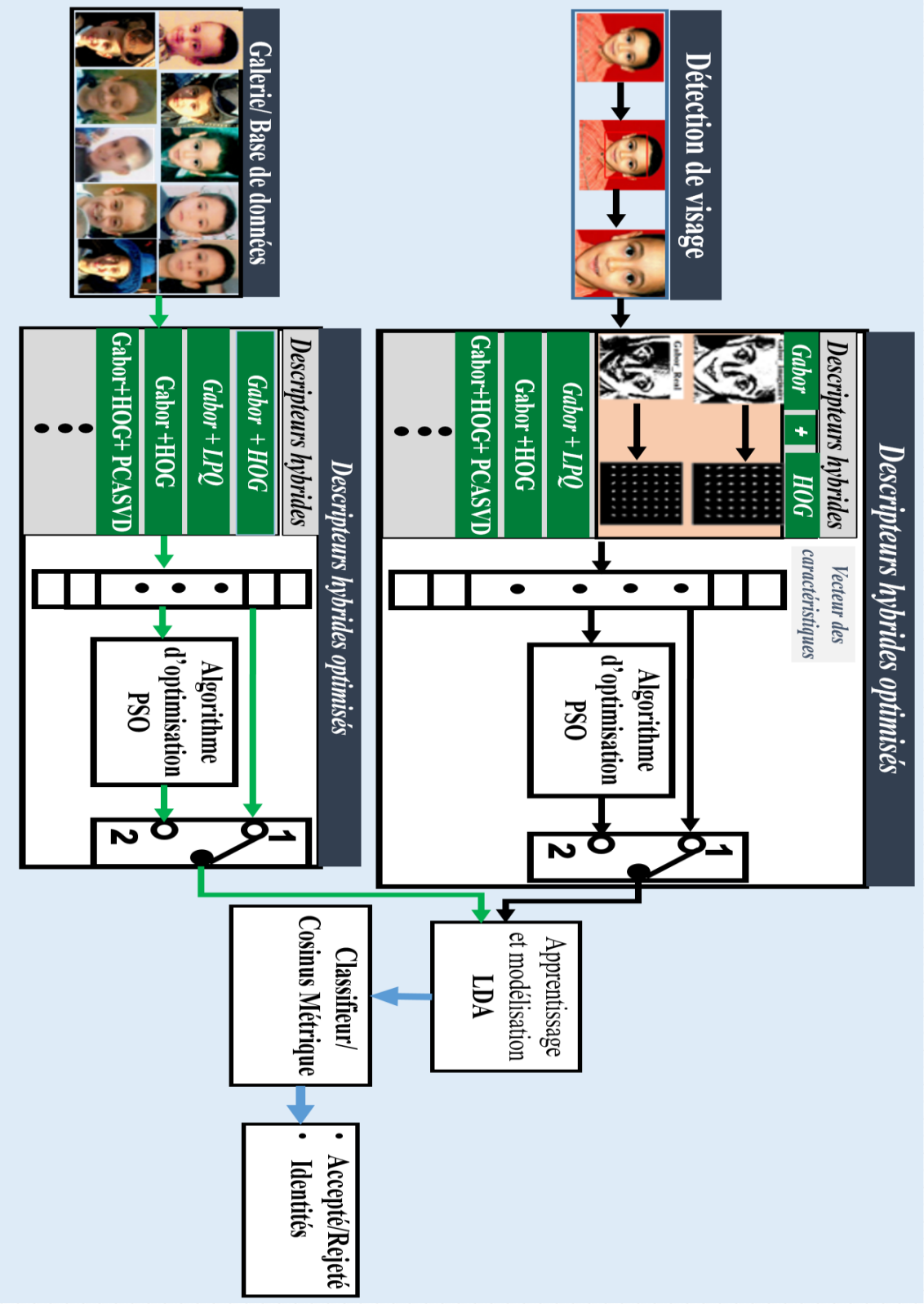


Fig 5. 1. Schéma de principe de systèmes hybrides proposé d'identification/vérification optimisé.

5.2.2. Apprentissage par Transfert Optimisée

Il s'agit d'une structure d'apprentissage approfondi (Transfer Learning) en s'appuyant sur une structure CNN appelé AlexNet dans le but d'extraction des caractéristiques de visages. Les dernières couches de la structure AlexNet sont remplacées par l'association en cascade d'algorithme d'optimisation PSO, suivi d'un algorithme de division de données en deux parties appelé (SplitDataset), une partie pour l'apprentissage et l'autre pour le test, et un classificateur monocouche "Softmax" est ajouté à sa sortie. Les paramètres originaux du modèle AlexNet ont été conservés et utilisés comme initialisation du modèle proposé (Fig 5. 2).

Dans l'algorithme classique, la reconnaissance de visages nécessite l'extraction des caractéristiques du visage avant la classification et la reconnaissance, ce qui affectera le taux de reconnaissance. Pour résoudre ce problème, cette approche utilise la structure CNN-AlexNet pour réaliser un système de reconnaissance de visages.

La motivation et les raisons derrière le choix d'AlexNet comme noyau du modèle proposé sont sa structure profonde, sa structure simple, son temps d'apprentissage rapide et sa moindre occupation d'espace mémoire. On trouve aussi dans la littérature que pour ces raisons mentionnées le modèle AlexNet est adapté à l'analyse d'images falsifiées.

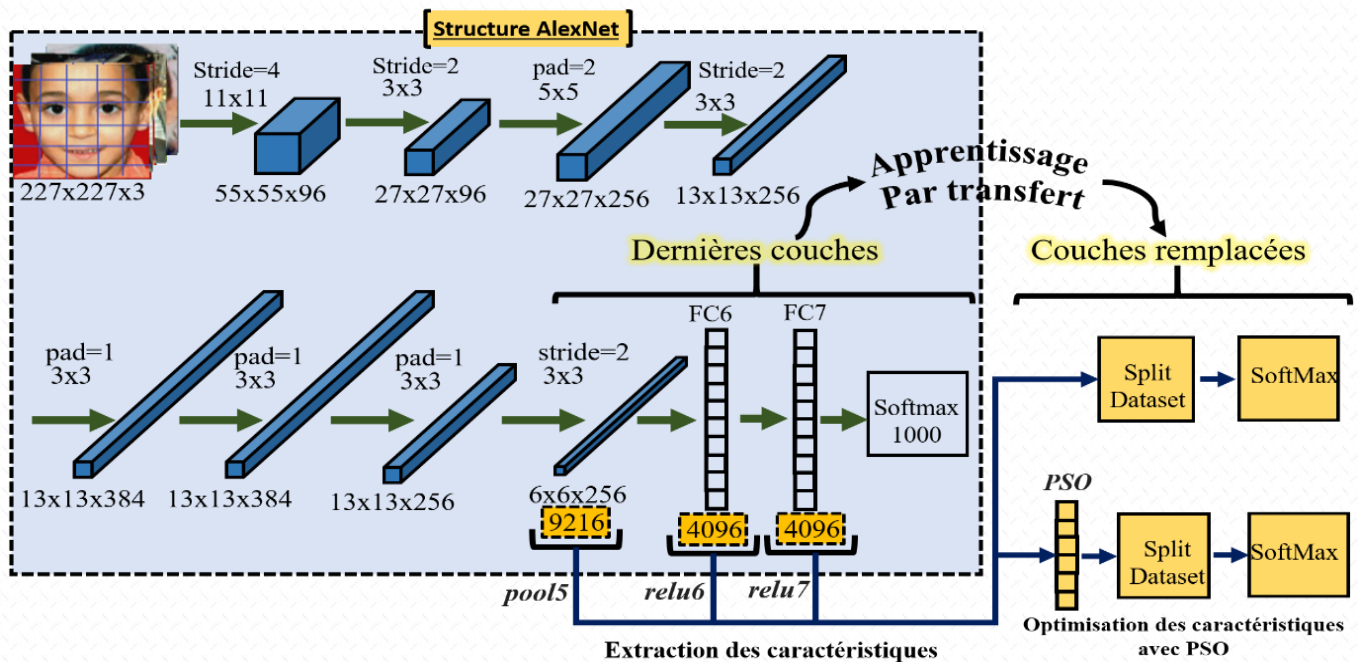


Fig 5. 2. Schéma de principe de l'optimisation combinée avec AlexNet et Transfer Learning.

1) Réseau de neurones convolutionnels

Le terme profond (Deep) fait généralement référence au nombre de couches cachées dans le réseau de neurones. L'une des classes DL est CNN, la plus couramment appliquée à l'analyse de l'imagerie visuelle. Ils ont des applications dans la reconnaissance d'images et de vidéos, les systèmes de recommandation, la classification d'images, l'analyse d'images médicales et le traitement du langage naturel.

Parmi les structures CNN les plus populaires, on retrouve la structure AlexNet [128], [22]. AlexNet était beaucoup plus grand que les CNN précédents utilisés pour les tâches de vision par ordinateur (par exemple, l'article LeNet de Yann LeCun en 1998) [140]. Il possède 60 millions de paramètres et 650 000 neurones. Il a remporté le concours ImageNet LSVRC-2012 2012 avec une large marge (taux d'erreur de 15,3 contre 26,2 (deuxième place)). Il se compose de cinq couches convolutives et de trois couches entièrement connectées.

2) Apprentissage par transfert

L'apprentissage par transfert est une branche de l'apprentissage automatique qui transfère des connaissances utiles du domaine source au domaine cible pour créer un bon modèle de classification [141]. Dans de nombreuses tâches de classification du monde réel, il est difficile de satisfaire l'hypothèse selon laquelle les données d'apprentissage et les données à classer partagent la même distribution et le même espace de caractéristiques, et afin de résoudre les problèmes de classification, nous avons utilisé la stratégie de TL.

La stratégie TL consiste à prendre un modèle préformé, dans notre cas la structure "AlexNet" (les poids et les paramètres d'un réseau qui a été formé sur un grand ensemble de données existant auparavant) et l'ajuster avec de nouveaux propres ensembles de données. Le modèle préformé joue le rôle d'un extracteur de caractéristiques en supprimant les dernières couches du réseau (les trois dernières couches) et en les remplaçant par un nouvel ensemble de modèles de classification.

Pour l'extraction des caractéristiques, nous avons mené plusieurs expérimentations pour choisir la sortie appropriée, qui nous permettra d'extraire les meilleures fonctionnalités du modèle « Alexnet » :

- La sortie " Couche requise = pool5" avec un vecteur de 9216 caractères.
- La sortie " Couche requise = relu6" avec un vecteur de 4096 caractères.
- La sortie " Couche requise = relu7" avec un vecteur de 4096 caractères.

Nous avons constaté que les meilleurs résultats sont obtenus pour le choix de la sortie "**Couche requise = relu6**" qui a été adoptée dans notre structure proposée **d'Apprentissage par transfert**. La sortie "Couche requise = relu6" sera suivie respectivement en cascade d'un module d'optimisation PSO, un module de division de données en deux parties appelé "**SplitDataset**" (une partie pour l'apprentissage et l'autre pour le test), et un module classificateur monocouche "**Softmax**". (voir Fig 5. 2).

5.3. Bases de données

5.3.1. Base de données ORL

La base de données contient une collection d'images de visage prises entre avril 1992 et avril 1994 au Laboratoire d'informatique de l'Université de Cambridge. Cette base de données a fait l'objet d'une étude menée sur la reconnaissance de visage et la vision. La base de données contient 400 images avec dix images par personne. Ces images ont été prises en fonction de l'expression du visage, avec ou sans lunettes et de l'ambiance lumineuse. Toutes les images ont été prises sur un fond sombre et homogène avec des personnes en position verticale et frontale. Des exemples d'images de visage de cette base de données sont présentés dans (Fig 5. 3 .a)). Taille de l'image = 112x92 pixels = 10304 pixels. (Cette base est disponible sur <http://www.camorl.co.uk/facedatabase.html>).

5.3.2. Base de données AR

Cette base de données [142] de visages a été créée en 1998 (Ohio State University, USA) par Aleix Martinez et Robert Benavente dans le centre de vision par ordinateur d'Université autonome de Barcelone. Il apporte plus de 4000 images de 126 personnes. Ces images ont été prises en fonction des expressions faciales, de l'atmosphère lumineuse et des lunettes de soleil, etc. Les images ont été prises avec des conditions contrôlées sans aucune

restriction d'usure et de coiffure. Pour chaque participant, la même photo est prise selon deux séances distinctes.

Des exemples d'images de visage de cette base de données sont présentés par la Fig 5. 3. (b), plus bas dans le document. La taille de l'image = $120 \times 165 \times 3$ pixels = 59400 pixels.

Cette base est accessible au public et peut être obtenue à partir de <http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html>.

5.3.3. Base de données LFW

Cette grande base de données [143], [144] de photographies de visages étiquetées conçue à étudier le problème de reconnaissance de visage dans des environnements non contrôlés. L'ensemble de données contient 13233 images de visages collectés sur le Web, représentant 5749 personnes, dont 1680 personnes avec deux images ou plus. Pour chaque image du visage de la personne, un nom lui a été attribué. La base de données a été divisée en deux vues : la première pour la sélection du modèle et la seconde pour l'évaluation des performances. Elle peut être consultée et téléchargée à l'adresse <http://vis-www.cs.umass.edu/lfw/>.

Pour nos expérimentations, nous avons utilisé l'approche proposée sur la vue 2 en utilisant des images de visage alignées selon un protocole d'image restreint, dans lequel aucune donnée d'entraînement supplémentaire ne peut être impliquée. L'ensemble de données utilisé est divisé en dix sous-ensembles disjoints, où neuf sous-ensembles sont utilisés pour la formation et un pour les tests.

Chaque sous-ensemble contient 300 paires de faces appariées et 300 paires de faces non appariées. Les performances sont rapportées sous la forme d'une précision moyenne des plis et d'une courbe ROC de validation croisée multipliée par 10. Des exemples d'images de visage sur cette base de données sont présentés dans la Fig 5. 3. (c) La taille de l'image d'origine = $250 \times 250 \times 3$ pixels = 187500 pixels.

5.3.4. Base de données "Thermal Tufts Face (TTF)"

TTF est un ensemble de données faciales à grande échelle (plus de 10 000 images, 74 femmes et 39 hommes, de plus de 15 pays avec une tranche d'âge comprise entre 4 et 70 ans). Il contient 6 modalités d'image : visible, proche infrarouge, thermique, croquis informatisé, vidéo enregistrée et images 3D [145]. Dans ce document nous nous intéressons uniquement à la modalité d'images thermiques. La taille de l'image d'origine = 128x128x3 pixels = 49152 pixels (Fig 5. 3. (d)).



Fig 5. 3. Exemples d'images des bases de données : (a) ORL, (b) AR, (c) LFW et (d) TTF.

5.4. Résultats expérimentaux et discussion

Dans nos expérimentations, pour l'algorithme PSO, nous avons choisi une fonction de coût "fitness" de la forme Sigmoïde :

$$fun = \frac{1}{1 + e^{mean(data(:))}} \quad (5.3)$$

Cette dernière a apporté des améliorations significatives à la matrice des caractéristiques. Concernant la base de données « LFW », elle a subi un prétraitement d'alignement utilisant la méthode de sélection des visages et des cinq points faciaux Fig 4. 6.

5.4.1. Descripteurs Hybrides Optimisés

Nous avons implémenté plusieurs méthodes hybrides comme le montrent les figures et tableaux de résultats ci-dessous avec et sans PSO :

Le Tableau 5. 1 montre que le descripteur hybride Gabor+HOG de l'expérience numéro 7 donne le meilleur taux de reconnaissance (TR). Cela nous prouve que chaque descripteur contribue à sa manière à fournir au système FR des informations efficaces pour une bonne

identification. Les descripteurs hybrides Gabor+HOG+PSO et Gabor+HOG+PSO+PCASVD restent de bons candidats lorsqu'ils sont optimisés par le PSO.

On remarque que la méthode PCASVD utilisée dans la phase de réduction des données (expérience 10) n'améliore pas les performances car dans le cas de l'expérience 9 la projection se fait dans le sous-espace LDA qui est plus discriminatoire. Donc l'expérience 10 nous montre que l'optimisation des caractéristiques réduites n'est pas la meilleure solution. Il est préférable de sélectionner les caractéristiques optimales dès le départ puis de réduire. C'est le cas pour tout le reste des expériences réalisées.

Tableau 5. 1. Rang 1- Taux de reconnaissance (TR) pour les bases de données ORL, AR et TTF

N°	Méthodes Hybrides	ORL (%)	AR (%)	Thermal Tufts Face-TTF (%)
1	PSO	87.50	85.11	85.10
2	LPQ	99.00	40.14	29.57
3	Gabor + LPQ	99.50	85.61	31.97
4	Gabor + LPQ + PSO	99.50	86.16	30.05
5	HOG	96.50	95.26	87.26
6	HOG + PSO	93.50	99.37	85.10
7	Gabor + HOG	99.50	99.53	87.26
8	Gabor + HOG + PSO	98.50	99.32	86.06
9	Gabor + HOG + PCASVD	99.00	99.63	87.26
10	Gabor + HOG + PCASVD + PSO	30.50	13.17	5.53
11	Gabor + HOG + PSO + PCASVD	99.00	99.32	87.26

Selon la [Fig 5. 4](#) et [Fig 5. 5](#), on peut voir que les résultats les plus précis sont obtenus par le descripteur hybride optimisé Gabor+HOG+PSO.

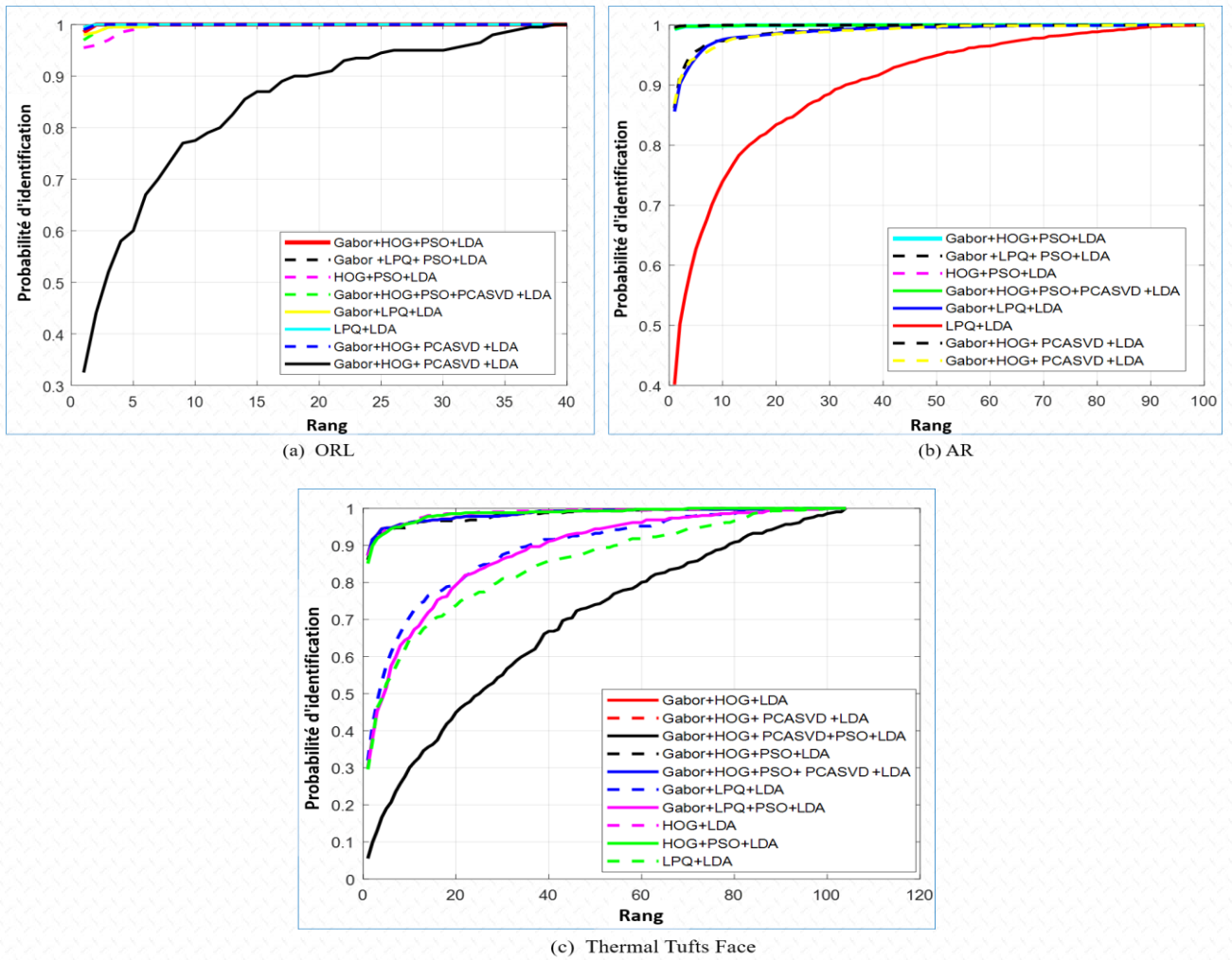


Fig 5. 4. Courbes CMC avec et sans l'algorithme PSO de ORL, AR & TTF.

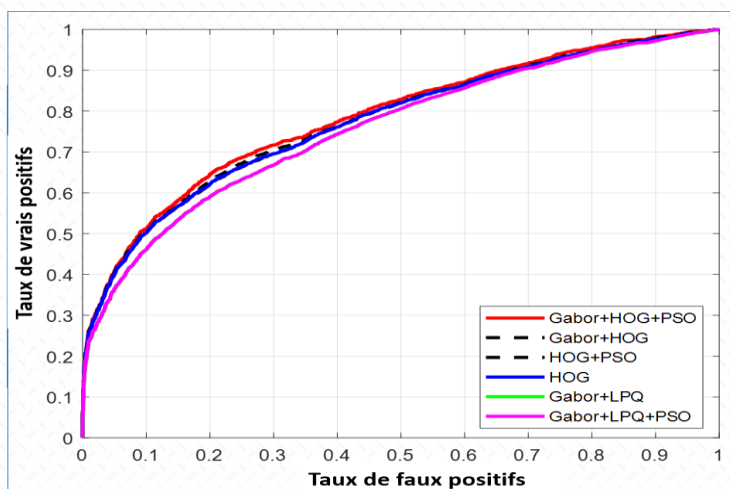


Fig 5. 5. Courbes ROC avec et sans l'algorithme PSO de LFW.

Des expérimentations sur différents échantillons de test (30 à 60) montrent que la précision reste pratiquement stable sur les différentes bases de données ORL, ORL et TTF. Le choix d'une base de données de test est choisi pour les bonnes performances (voir Fig 5. 6).

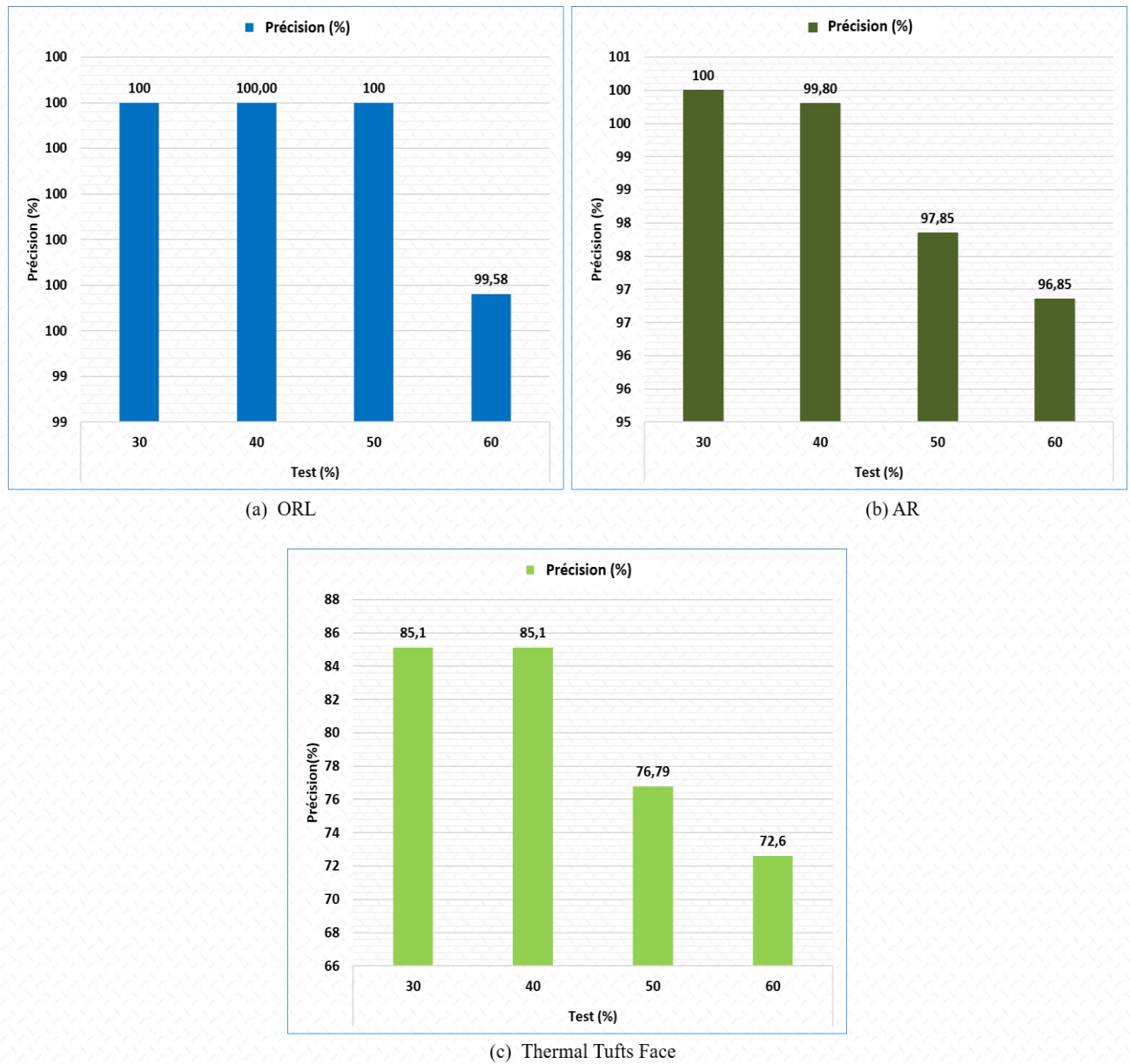


Fig 5. 6. TR avec variation du taux de test par AlexNet_TL sans PSO de ORL, AR & TTF.

1) Comparaison avec l'état de l'art

Les [Tableau 5. 2](#) et [Tableau 5. 3](#) présentent un aperçu des meilleurs résultats d'identification obtenus par notre proposition de méthodes hybrides (Gabor+LPQ+PSO pour le [Tableau 5. 2](#) et Gabor+HOG+PSO pour le [Tableau 5. 3](#)) avec les meilleurs résultats présentés dans la littérature.

Tableau 5. 2. Comparaison de nos meilleurs résultats du TR avec d'autres méthodes existantes dans la littérature obtenus pour La base de données ORL.

N°	Méthodes	TR (%)
1	DCT-PCA and LDA [37]	93.00
2	DLSR [38]	95.63
3	PCA + DCT + CS [39]	96.50
4	PSO-KNN [52]	98.75
5	SRC-EnNPSO [54]	98.92
6	SRC-CMCGSO [54]	99.06
7	IKLDA+PNN [146]	97.22
8	Notre : Gabor + LPQ + PSO	99.50

Tableau 5. 3. Comparaison de nos meilleurs résultats du TR avec d'autres méthodes existantes dans la littérature obtenus pour La base de données AR.

N°	Méthodes	TR (%)
1	GSR-FR [40]	96.70
2	SFD-CR [41]	96.94
3	CD-EJSR [42]	99.17
4	KSVD-IJSR [43]	99.17
5	DWT E-CLAHE [44]	98.78
6	Notre : Gabor + HOG + PSO	99.32

Le [Tableau 5. 4](#) montre que le descripteur hybride Gabor+HOG+PSO donne le meilleur taux de reconnaissance. On remarque que le Gabor+LPQ donne de faibles performances. Il présente un aperçu des résultats obtenus par nos méthodes hybrides proposées par rapport à la littérature utilisant presque les mêmes méthodes hybrides.

Tableau 5. 4. Taux de reconnaissance de visages, Rang 1 avec LFW et comparaison avec d'autres méthodes existantes dans la littérature

	Méthodes					
TR (%)	HOG	LBP	Gabor	Gabor + HOG + LBP	PSO	Optimization models- Hybrid
Samma H & al. [53]	43.75	37.50	40.00	50.00	43.75	50.00
	HOG	HOG+PSO	Gabor+HOG	Gabor+HOG+PSO	Gabor+LPQ	Gabor +LPQ+ PSO
Elaggoune H & al. [47]	70.32	70.38	69.23	69.25	65.25	65.12
Nos Résultats	70.31	70.23	70.73	72.35	67.20	67.20

Les [Tableau 5. 2](#), [Tableau 5. 3](#), [Tableau 5. 4](#), et [Tableau 5. 5](#) comparent les meilleurs taux obtenus par nos approches proposées pour plusieurs méthodes existantes dans la littérature utilisant le PSO. Les résultats présentés dans le [Tableau 5. 2](#) montrent que le descripteur hybride Gabor+LPQ+PSO surpasse toutes les méthodes mentionnées. Le [Tableau 5. 3](#), et le [Tableau 5. 4](#) montrent que le descripteur hybride Gabor+HOG+PSO qui surpasse toutes les méthodes mentionnées. Quant au [Tableau 5. 5](#) c'est le descripteur hybride Gabor+HOG+PSO+PCASVD qui surpasse toutes les méthodes mentionnées.

5.4.2. Apprentissage par Transfert Optimisée

Les [Tableau 5. 5](#), [Tableau 5. 6](#), [Tableau 5. 7](#) et [Tableau 5. 8](#) comparent les meilleurs taux obtenus par nos approches proposées à plusieurs méthodes existantes dans la littérature utilisant L'apprentissage approfondi avec ou sans optimisation.

Tableau 5. 5. Taux de reconnaissance pour le modèle CNN avec TTF : comparaison avec d'autres méthodes existantes dans la littérature.

N°	Méthodes	TR (%)
1	Multi-AP-GAN [55]	75:86
2	Multi-AP-GAN (GT) [55]	77.38
3	CRN + CL [95]	74.90
4	Notre : Gabor + HOG + PSO + PCASVD	87.26
5	Notre : PSO_AlexNet_TL	83.41

Tableau 5. 6. Taux de reconnaissance pour le modèle CNN avec ORL : comparaison avec d'autres méthodes existantes dans la littérature.

N°	Méthodes	TR (%)
1	RNN_PSO [96]	94.07
2	ESPCN + CNN [97]	92.24
3	PCA-FRCNN [98]	99.00
4	Alex Net + SVM [99]	99.17
5	FSRC-QO-CNN [147]	98.45
6	Notre : PSO_AlexNet_TL	99.17

Tableau 5. 7. Taux de reconnaissance pour le modèle CNN avec AR : comparaison avec d'autres méthodes existantes dans la littérature.

N°	Méthodes	TR (%)
1	Deep Learning + softmax [100]	94.42
2	Metaheuristic algorithms (IC) [101]	95.30
3	FSRC-QO-CNN [147]	99.7
4	BDL [148]	92.64
5	CSGT-CRC [149]	97.7
6	Notre : PSO_AlexNet_TL	99.86

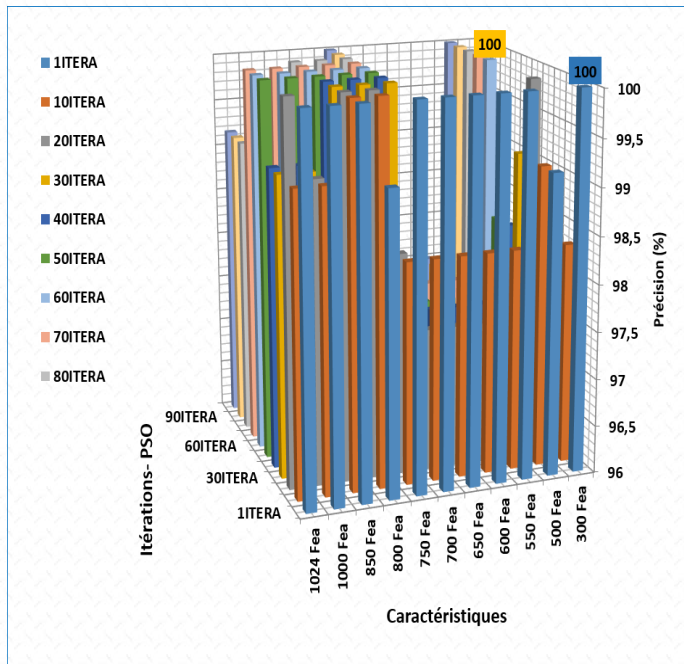
Tableau 5. 8. Taux de reconnaissance pour le modèle CNN avec LFW : comparaison avec d'autres méthodes existantes dans la littérature.

N°	Méthodes	TR (%)
1	AlexNet + SVM [99]	94
2	Transfer learning (AlexNet) [99]	95.63
3	DeepID [100]	97.45
4	ArcFace-CNN [102]	99.50
5	Transfer Learning-Auto encoder [103]	98.83
6	Weighted PCA-EFMNet Deep Learning [150]	95.00
7	Siamese-VGG [151]	95.62
8	Notre : PSO_AlexNet_TL	99.57

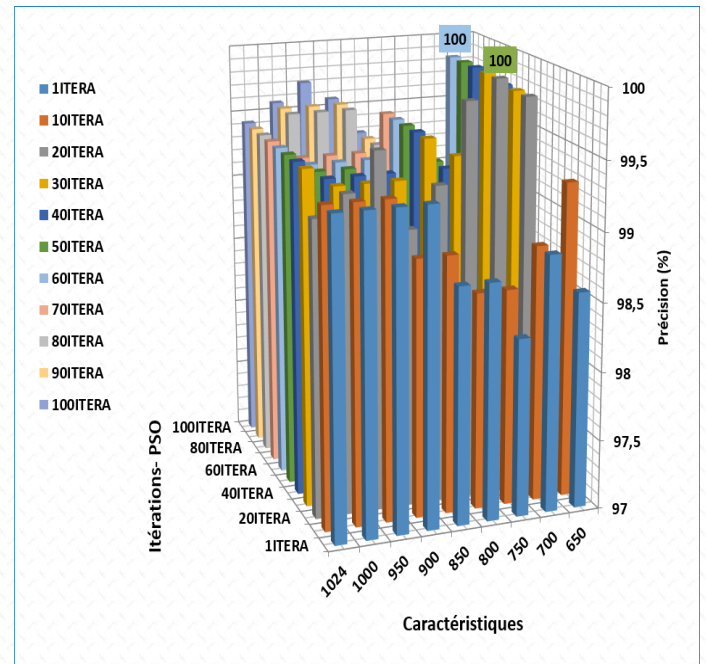
Nous réalisons quelques expériences qui nous permettront de choisir notre échantillon de test et d'établir un protocole de bases de données pour ORL, AR et TTF. Le protocole de la base de données LFW est déjà défini dans la section " [5.3.3 Base de données LFW](#) ". Les résultats de ces expériences sont donnés par la [Fig 5. 7](#).

D'après les résultats obtenus, le choix d'un échantillon de 70% (pour entraînement) et 30 % (pour test) des images d'un candidat de chaque base de données pour entraîner et tester notre système de reconnaissance de visages s'avère être un bon candidat. Ce qui nous amène de le prendre comme choix favorisé dans la suite de nos expérimentations. Plus l'échantillon d'apprentissage est important plus les performances sont meilleures, ce qui est évident.

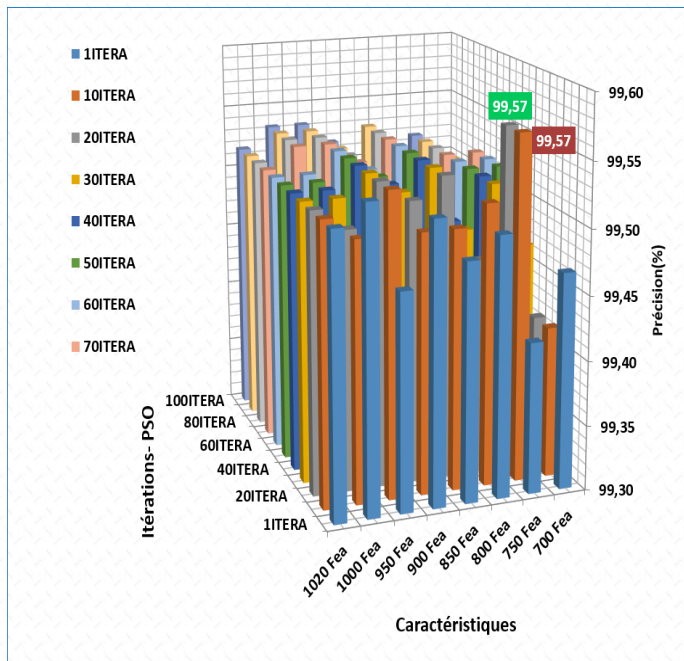
Optimisation des caractéristiques de visages en utilisation des descripteurs hybrides et CNN en apprentissage par transfert



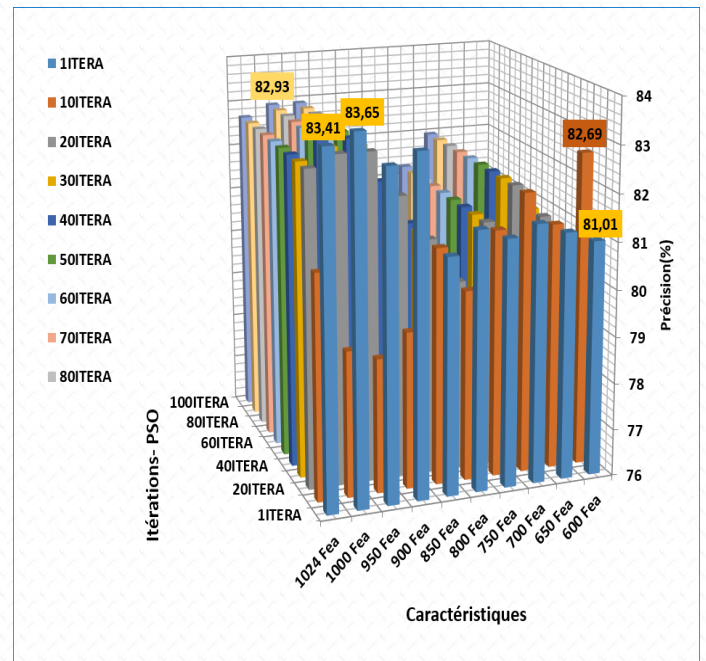
(1) ORL



(2) AR



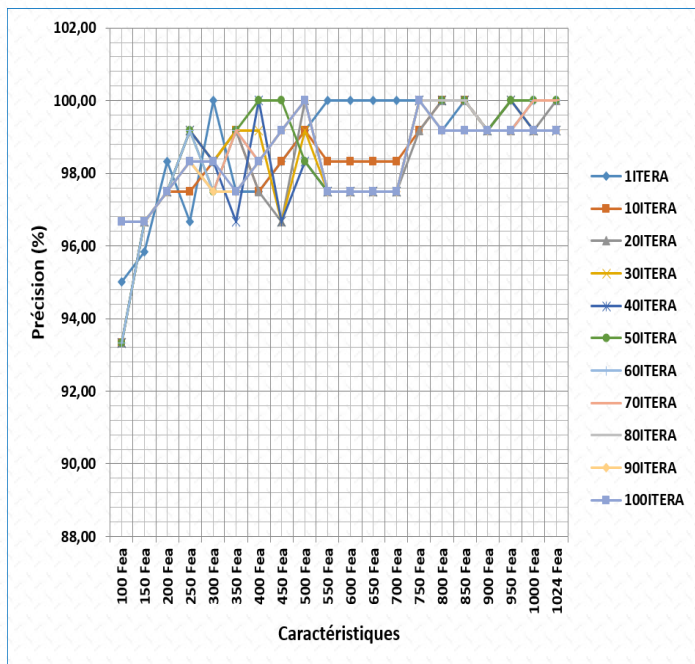
(3) LFW



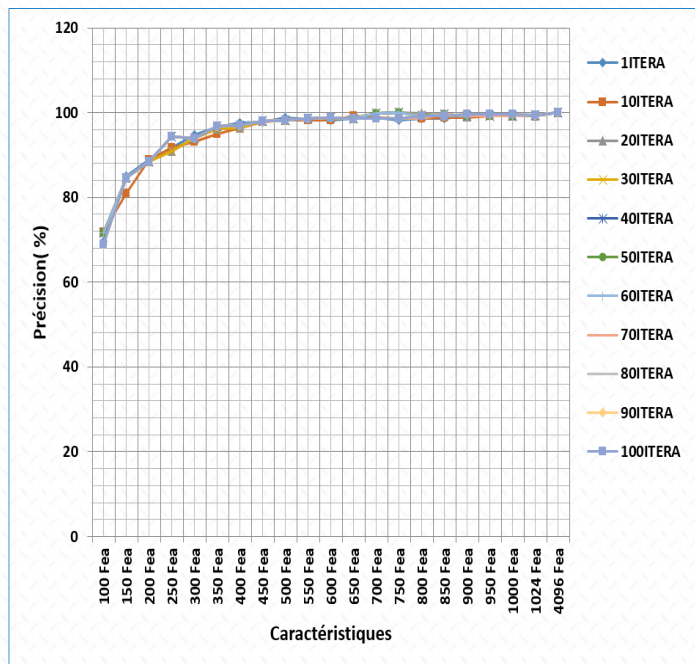
(4) Thermal Tufts Face

a) Histogramme des meilleurs résultats TR en fonctions d'itérations et de Caractéristiques.

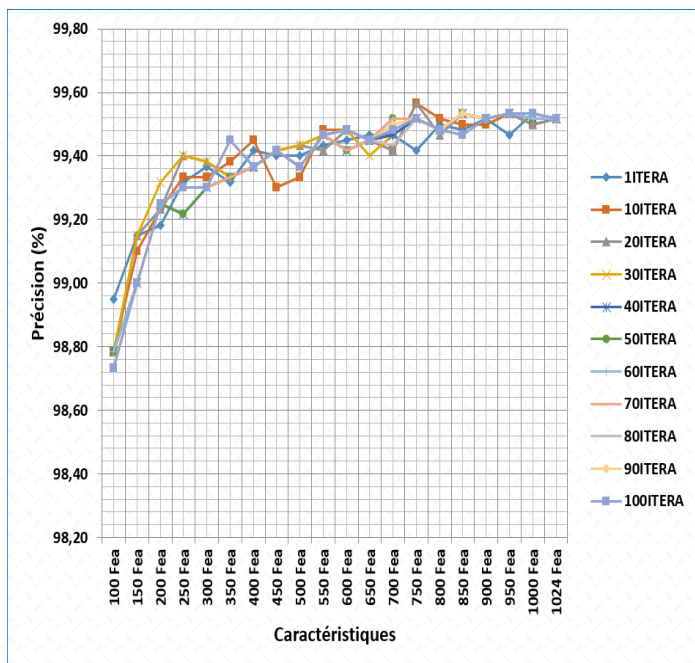
Optimisation des caractéristiques de visages en utilisation des descripteurs hybrides et CNN en apprentissage par transfert



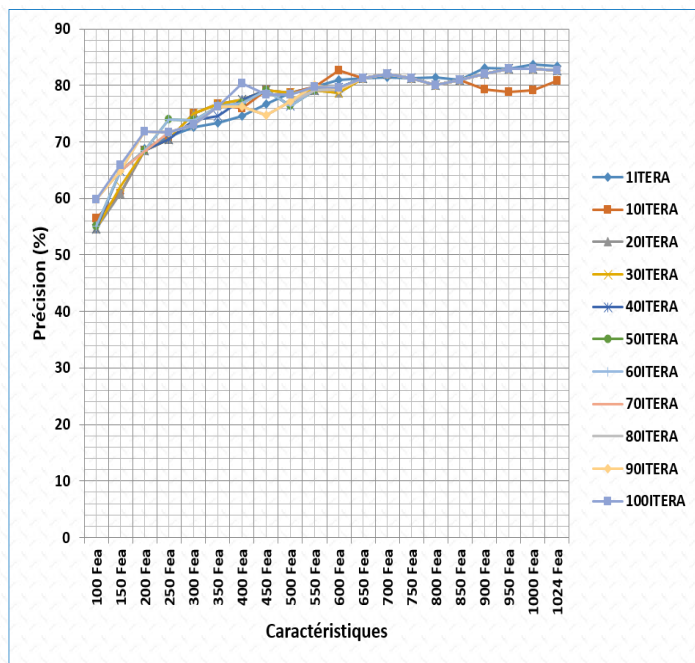
(1) ORL



(2) AR



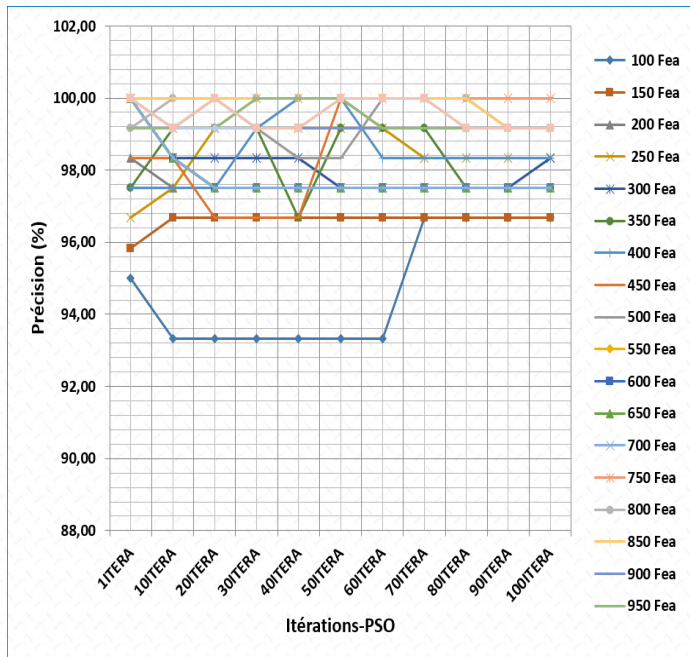
(3) LFW



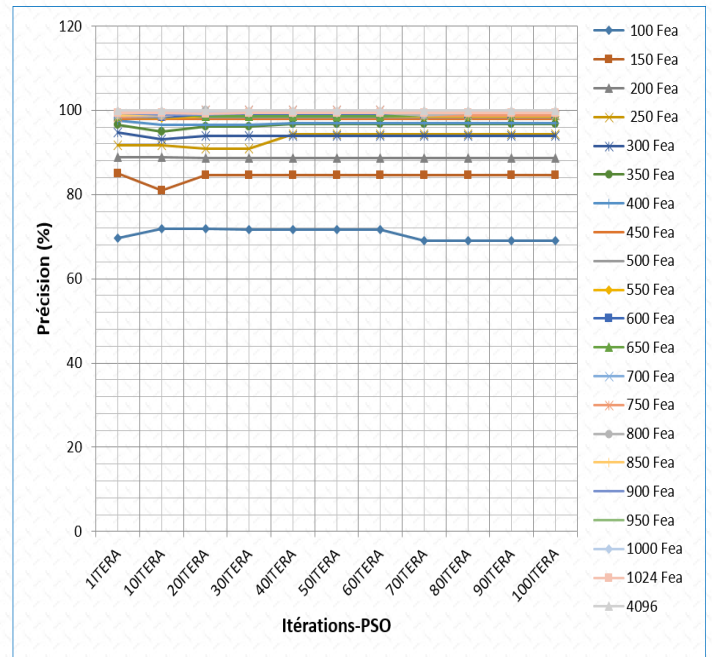
(4) Thermal Tufts Face

b) Variation du TR en fonctions des Caractéristiques avec la fixation d'itérations.

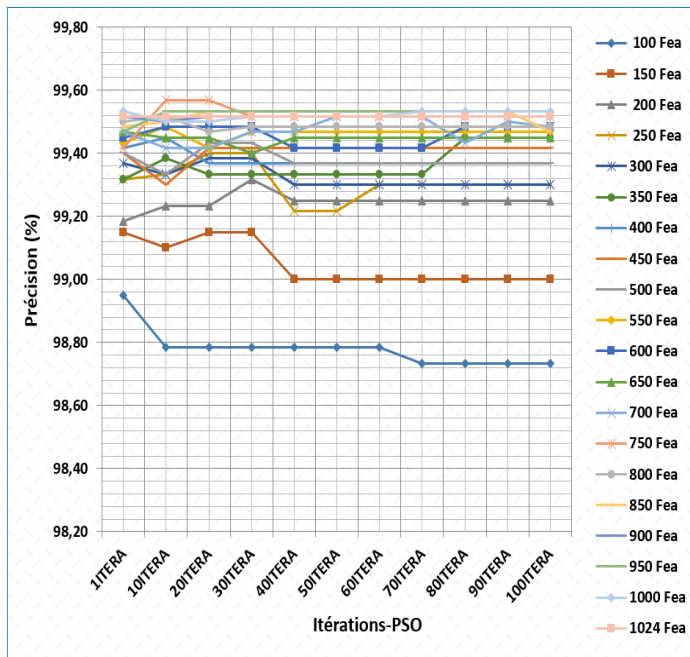
Optimisation des caractéristiques de visages en utilisation des descripteurs hybrides et CNN en apprentissage par transfert



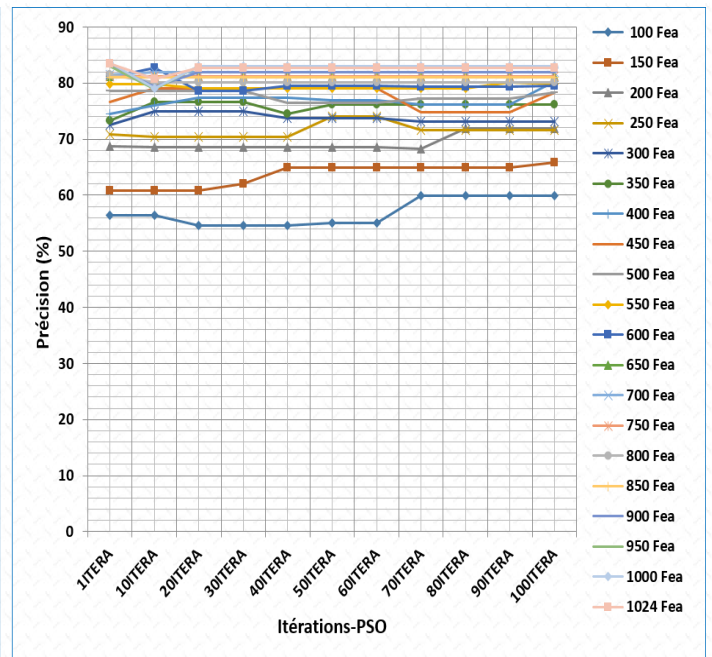
(1) ORL



(2) AR



(3) LFW



(4) Thermal Tufts Face

c) variation du TR en fonctions d'Itérations avec la fixation des Caractéristiques.

Fig 5. 7. Performances du TR sur la base de données AR, ORL, LFW, et TTF.

Une chose importante à retenir à ce stade du travail est le temps de traitement d'une seule image. On constate d'après les résultats obtenus que le temps de traitement global (temps d'apprentissage et de test) pour une image couleur ou infrarouge en environnement contrôlé et non contrôlé reste très proche de celui d'une image en niveaux de gris en milieu contrôlé : 0,55s, 0,58s, 0,62s et 0,6s respectivement pour les bases de données ORL, AR, LFW et TTF (voir Fig 5. 8). Ce résultat est très intéressant puisqu'il nous permet de dire que notre approche est standard et joue son rôle d'apprentissage automatique. On voit que quel que soit la base de données utilisée les performances sont stables et donc le système dans ce cas se caractérise par sa robustesse, ce qui répond à nos objectifs, et c'est encourageant. Les meilleurs résultats sont transportés sur des graphiques explicatifs (voir Fig 5. 8).

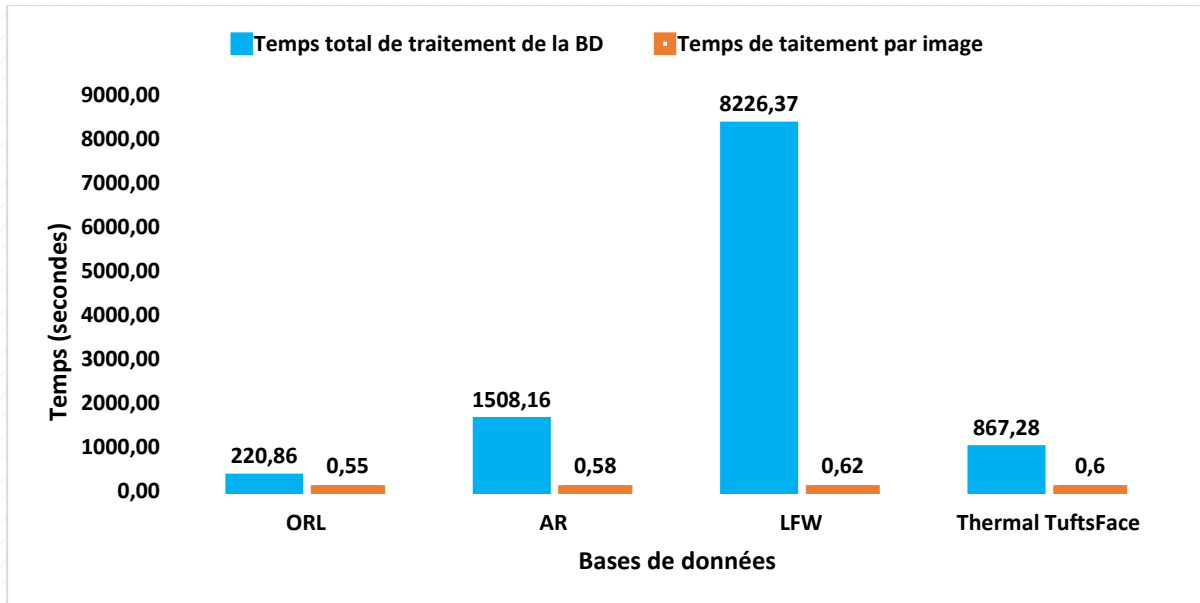


Fig 5. 8. Temps de traitement consommé par PSO_AlexNet_TL.

D'après les résultats graphiques obtenus, nous pouvons constater et déduire que : concernant la base AR, nous obtenons une stabilité à partir de 450 caractéristiques (Fig 5. 7. b) (2)) quel que soit le nombre d'itérations. On note que de bons résultats sont obtenus à partir de 30 itérations voir Tableau 5. 9 et (Fig 5. 7. c) (2)) stabilité et bons résultats quel que soit le nombre de caractéristiques. Le paramètre d'itération est donc très important, ce qui confirme que le PSO optimise les performances du système. Dans ce cas le meilleur résultat est obtenu pour les caractéristiques nombre=1024 et ce quel que soit le nombre d'itérations. Nous

constatons que nous obtenons des performances optimales dès 40 itérations. Nous pouvons également dire que selon les histogrammes obtenus de AR et de LFW, le système nécessite moins d'itérations avec un nombre d'itérations = 10 et un nombre de caractères = 750 pour LFW, un nombre d'itérations = 20 et un nombre de caractères = 800 pour AR, un nombre d'itérations = 60 et nombre de caractères = 500 pour ORL, et un nombre d'itérations = 01 et nombre de caractères = 1024 pour TTF. Par conséquent, le Deep PSO dans ce cas est plus efficace sur des bases de données de grandes tailles avec plus de variantes et dans des environnements non contrôlés.

Nous avons cherché à découvrir la variation de TR en fonction de la variation des caractéristiques et des itérations afin de voir l'influence de l'algorithme d'optimisation PSO sur celle-ci voir [Fig 5. 9](#).

Il est clair que le temps est plus important dans le cas du LFW car il s'agit d'une grande base de données avec beaucoup de variation d'un environnement non contrôlé. Nous obtenons 8226,37s pour les temps d'entraînement et de test avec un temps de test égal à 4,48s pour un TR de 99,57% voir [Fig 5. 9](#), ce qui est très encourageant et montre que le PSO comme méthode d'optimisation est à retenir dans l'amélioration des performances du système ainsi que du temps de traitement.

La [Fig 5. 9](#) montre l'influence du PSO sur la consommation de temps par image. Il augmente avec la taille de la base de données. En retour, le PSO est utilisé pour améliorer et optimiser le temps de test.

Les résultats de la [Fig 5. 9](#) confirment l'importance d'optimiser la durée globale du test. On note une nette amélioration avec tout ce qui est obtenu dans le cas de toutes les expériences réalisées sur le test ; qui est au mieux égal à 3,65s, 28,34s, 4,65s et 16s par rapport aux résultats d'optimisation PSO du test de temps = 2,1s, 6,7s, 4,48s et 4,4s respectivement aux bases de données ORL, AR, LFW et TTF.

Tableau 5.9. Variation de TR (%) en fonction des Caractéristiques et des Itérations.

a) ORL

<i>Caractéristiques</i> \ <i>Itérations</i>	1	10	20	30	40	50	60	70	80	90	100
100	95	93.33	93.33	93.33	93.33	93.33	93.33	96.67	96.67	96.67	96.67
150	95.83	96.67	96.67	96.67	96.67	96.67	96.67	96.67	96.67	96.67	96.67
200	98.33	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5
250	96.67	97.5	99.17	99.17	99.17	99.17	99.17	98.33	98.33	98.33	98.33
300	100	98.33	98.33	98.33	98.33	97.5	97.5	97.5	97.5	97.5	98.33
350	97.5	99.17	99.17	99.17	96.67	99.17	99.17	99.17	97.5	97.5	97.5
400	97.5	97.5	97.5	99.17	100	100	98.33	98.33	98.33	98.33	98.33
450	98.33	98.33	96.67	96.67	96.67	100	99.17	99.17	99.17	99.17	99.17
500	99.17	99.17	100	99.17	98.33	98.33	100	100	100	100	100
550	100	98.33	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5
600	100	98.33	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5
650	100	98.33	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5
700	100	98.33	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5
750	100	99.17	99.17	100	100	100	100	100	100	100	100
800	99.17	100	100	100	100	100	100	100	100	99.17	99.17
850	100	100	100	100	100	100	100	100	100	99.17	99.17
900	99.17	99.17	99.17	99.17	99.17	99.17	99.17	99.17	99.17	99.17	99.17
950	99.17	99.17	99.17	100	100	100	99.17	99.17	99.17	99.17	99.17
1000	100	99.17	99.17	99.17	99.17	100	100	100	99.17	99.17	99.17
1024	100	99.17	100	99.17	99.17	100	100	100	99.17	99.17	99.17

b) AR

<i>Caractéristiques</i> \ <i>Itérations</i>	1	10	20	30	40	50	60	70	80	90	100
100	69.57	71.86	71.86	71.57	71.57	71.57	71.57	69	69	69	69
150	85	81	84.57	84.57	84.57	84.57	84.57	84.57	84.57	84.57	84.57
200	88.86	88.86	88.57	88.57	88.57	88.57	88.57	88.57	88.57	88.57	88.57
250	91.71	91.71	90.86	90.86	94.29	94.29	94.29	94.29	94.29	94.29	94.29
300	94.71	93.14	93.86	93.86	93.86	93.86	93.86	93.86	93.86	93.86	93.86
350	96.57	95	96.14	96.14	96.71	96.71	96.71	96.71	96.71	96.71	96.71
400	97.57	96.43	96.43	96.43	97	97	97	97	97	97	97
450	97.86	97.86	98	98	98	98	98	98	98	98	98
500	98.86	98.29	98.29	98.29	98.29	98.29	98.29	98.29	98.29	98.29	98.29
550	98.43	98.14	98.57	98.57	98.57	98.57	98.57	98.57	98.57	98.57	98.57
600	98.29	98.29	98.86	98.86	98.86	98.86	98.86	98.86	98.86	98.86	98.86
650	98.57	99.29	98.57	98.57	98.57	98.57	98.57	98.71	98.71	98.71	98.71
700	98.86	98.86	99.86	99.86	99.86	99.86	99.86	98.71	98.71	98.71	98.71
750	98.29	98.57	100	100	100	100	100	98.71	98.71	98.71	98.71
800	98.71	98.57	99.86	99.43	99.29	99.29	99.29	98.71	99.29	99.29	99.29
850	98.71	98.86	99.29	99.57	99.57	99.57	99.57	99.57	99.29	99.29	99.29
900	99.29	98.86	99	99.29	99.29	99.29	99.29	99.29	99.57	99.57	99.57
950	99.29	99.29	99.57	99.29	99.29	99.29	99.29	99.29	99.57	99.57	99.71
1000	99.29	99.29	99.29	99.29	99.29	99.29	99.29	99.29	99.57	99.57	99.57
1024	99.29	99.29	99.14	99.43	99.43	99.43	99.43	99.43	99.43	99.43	99.43

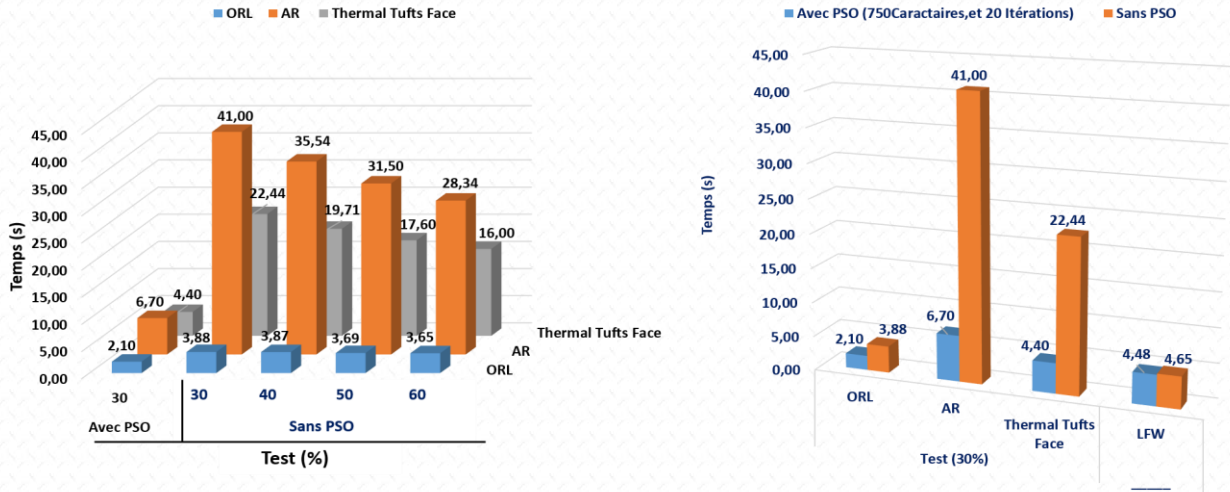
Optimisation des caractéristiques de visages en utilisation des descripteurs hybrides et CNN en apprentissage par transfert

c) LFW

<i>Caractéristiques</i> \ <i>Itérations</i>	1	10	20	30	40	50	60	70	80	90	100
100	98.95	98.78	98.78	98.78	98.78	98.78	98.78	98.73	98.73	98.73	98.73
150	99.15	99.1	99.15	99.15	99	99	99	99	99	99	99
200	99.18	99.23	99.23	99.32	99.25	99.25	99.25	99.25	99.25	99.25	99.25
250	99.32	99.33	99.4	99.4	99.22	99.22	99.3	99.3	99.3	99.3	99.3
300	99.37	99.33	99.38	99.38	99.3	99.3	99.3	99.3	99.3	99.3	99.3
350	99.32	99.38	99.33	99.33	99.33	99.33	99.33	99.33	99.45	99.45	99.45
400	99.42	99.45	99.37	99.37	99.37	99.37	99.37	99.37	99.37	99.37	99.37
450	99.4	99.3	99.42	99.42	99.42	99.42	99.42	99.42	99.42	99.42	99.42
500	99.4	99.33	99.43	99.43	99.37	99.37	99.37	99.37	99.37	99.37	99.37
550	99.43	99.48	99.42	99.47	99.47	99.47	99.47	99.47	99.47	99.47	99.47
600	99.45	99.48	99.48	99.48	99.42	99.42	99.42	99.42	99.48	99.48	99.48
650	99.47	99.45	99.45	99.4	99.45	99.45	99.45	99.45	99.45	99.45	99.45
700	99.47	99.42	99.42	99.47	99.47	99.52	99.52	99.52	99.43	99.5	99.48
750	99.42	99.57	99.57	99.52	99.52	99.52	99.52	99.52	99.52	99.52	99.52
800	99.5	99.52	99.47	99.48	99.48	99.48	99.48	99.48	99.48	99.48	99.48
850	99.48	99.5	99.53	99.53	99.53	99.53	99.53	99.53	99.53	99.53	99.47
900	99.52	99.5	99.52	99.52	99.52	99.52	99.52	99.52	99.52	99.52	99.52
950	99.47	99.53	99.53	99.53	99.53	99.53	99.53	99.53	99.53	99.53	99.53
1000	99.53	99.5	99.5	99.52	99.52	99.52	99.52	99.53	99.53	99.53	99.53
1024	99.52	99.52	99.52	99.52	99.52	99.52	99.52	99.52	99.52	99.52	99.52

d) Thermal Tufts Face

<i>Caractéristiques</i> \ <i>Itérations</i>	1	10	20	30	40	50	60	70	80	90	100
100	56.49	56.49	54.57	54.57	54.57	55.05	55.05	59.86	59.86	59.86	59.86
150	60.82	60.82	60.82	62.02	64.90	64.90	64.90	64.90	64.90	64.90	65.87
200	68.75	68.51	68.51	68.51	68.51	68.51	68.51	68.26	71.88	71.88	71.88
250	70.91	70.43	70.43	70.43	70.43	74.04	74.04	71.63	71.63	71.63	71.63
300	72.59	75.00	75.00	75.00	73.80	73.80	73.80	73.80	73.80	73.80	73.80
350	73.32	76.68	76.68	76.68	74.52	76.20	76.20	76.20	76.20	76.20	76.20
400	74.52	75.96	77.40	77.40	77.40	76.92	76.92	76.20	76.20	76.20	80.30
450	76.68	79.10	79.10	79.10	79.10	79.10	79.10	74.76	74.76	74.76	78.37
500	78.61	78.61	78.61	78.61	76.44	76.44	76.44	77.16	77.16	77.16	78.37
550	79.81	79.81	79.81	79.81	79.81	79.81	79.81	79.81	79.81	80.05	79.81
600	81.01	82.70	78.61	78.61	79.57	79.57	79.57	79.33	79.33	79.33	79.57
650	81.25	81.25	81.25	81.25	81.25	81.25	81.25	81.25	81.25	81.25	81.25
700	81.49	81.97	81.97	81.97	81.97	81.97	81.97	81.97	81.97	81.97	81.97
750	81.25	81.25	81.25	81.25	81.25	81.25	81.25	81.25	81.25	81.25	81.25
800	81.49	80.05	80.05	80.05	80.05	80.05	80.05	80.05	80.05	80.05	80.05
850	81.01	81.01	81.01	81.01	81.01	81.01	81.01	81.01	81.01	81.01	81.01
900	83.17	79.33	81.97	81.97	81.97	81.97	81.97	81.97	81.97	81.97	81.97
950	82.93	78.85	82.93	82.93	82.93	82.93	82.93	82.93	82.93	82.93	82.93
1000	83.65	79.09	82.93	82.93	82.93	82.93	82.93	82.93	82.93	82.93	82.93
1024	83.41	80.77	82.70	82.70	82.70	82.70	82.70	82.70	82.70	82.70	82.70



a) Comparaison Temps de test entre l'approche Alexnet_TL_PSO (750 Caractères, 20 itérations) et Alexnet_TL, avec changement du taux de test.

b) Comparaison Temps de test entre l'approche Alexnet_TL_PSO (750 Caractères, 20 itérations) et Alexnet_TL, avec taux de test 30

Fig 5. 9. Comparaison Temps de test entre l'approche Alexnet_TL_PSO (750 Caractères, 20 itérations) et Alexnet_TL.

5.4.3. Comparaison des résultats des deux approches

Le [Tableau 5. 10](#) présente le temps de test par image de différentes bases de données. D'après ce tableau, on constate généralement que les meilleurs résultats AR sont obtenus pour la première approche proposée quand les bases de données sont de petites tailles, cependant si les bases sont à grande échelle, la deuxième approche proposée Alexnet_TL donne les meilleurs résultats en taux de reconnaissance et en temps de traitement. L'avantage d'utiliser l'optimisation est de minimiser le temps de calculs, et de minimiser l'espace d'occupation pour le traitement de données (voir [Tableau 5. 10](#)).

Tableau 5. 10. Comparaison du TR et du temps de test des deux approches proposées pour ORL, AR, LFW, et TTF.

Bases de données	Méthodes	TR (%)	Temps Test global (s)	Temps Test/Image (ms)
ORL	Gabor + HOG + PSO	98.5	1.70	4.25
	PSO_AlexNet_TL	99.17	2.10	5.25
AR	Gabor +HOG + PSO	99.32	12.93	4.97
	PSO_AlexNet_TL	99.86	6.70	2.57
LFW	Gabor + HOG + PSO	72.35	4.85	0.80
	PSO_AlexNet_TL	99.57	4.48	0.74
Thermal Tufts Face	Gabor + HOG + PSO + PCASVD	87.26	4.47	3.07
	PSO_AlexNet_TL	83.41	4.4	3.02

5.5. Conclusion

Dans ce chapitre, nous avons mis en place deux techniques de reconnaissance de visages. Le premier est basé sur des descripteurs hybrides sélectionnés et le second est basé sur l'apprentissage approfondi en "Transfer Learning" pour l'extraction de caractéristiques. Les descripteurs hybrides sélectionnés reposent sur la combinaison de plusieurs descripteurs classiques tels que le filtre de Gabor, HOG, LPQ et PCA.

Le Transfer Learning s'appuie sur une structure CNN pré-entraînée appelé Alexnet sans ses dernières couches, exploiter dans l'extraction des caractéristiques de visages. Notre contribution à ces deux techniques réside dans l'intégration d'un algorithme d'optimisation PSO. Les résultats obtenus par les deux techniques montrent que l'association de plusieurs descripteurs en cascade et l'utilisation d'une structure CNN AlexNet en Transfer Learning pré-entraînée reste une technique intéressante dans le domaine de la reconnaissance de visages.

Nos modèles proposés sont appliqués sur différentes bases de données ORL, AR, LFW et Thermal-Tufts Face. Elles et comparées aux techniques existantes voir [Tableau 5. 2](#) au [Tableau 5. 8](#). Les résultats montrent que notre modèle proposé atteint un meilleure TR que la plupart des modèles de pointe.

6. Conclusion et perspectives

Dans ce travail, nous nous intéressons plus particulièrement à l'image et son rôle incontournable dans les systèmes de reconnaissances biométriques (en particulier la reconnaissance de visage). Malgré le développement terrible des moyens innovants et approuvés dans divers domaines de la recherche scientifique, l'image reste l'un des principaux piliers dont on ne peut se passer afin d'accéder aux faits scientifiques.

Parmi les utilisations les plus courantes des images est l'identification en biométrie qui est un domaine de la vision artificielle et qui connaît un intérêt grandissant depuis quelques années.

L'efficacité des techniques d'identification en imagerie est aujourd'hui très fortement liée à des contraintes fortes imposées à l'utilisateur, une voie de recherche actuelle se tourne donc vers la gestion de situations où l'acquisition des données est moins contrainte. En revanche, l'usage d'une seule modalité est souvent limité en termes de performance ou de difficultés d'usage, c'est pourquoi il apparaît intéressant d'évaluer l'apport de la multi modalité dans ce contexte. Finalement une stratégie d'optimisation des paramètres reste à développer.

L'objectif global des travaux est donc de permettre une meilleure compréhension et amélioration des performances des systèmes multimodaux pour l'identification, et par la suite une optimisation sur l'aspect spectral et spatial ainsi qu'un apprentissage automatique des images rendant notre approche plus souple, rapide et efficace.

Un des objectifs primordiaux est l'augmentation du taux de reconnaissance en utilisant des systèmes multimodaux, et la diminution de l'espace de stockage de l'information, en utilisant les algorithmes d'optimisations.

Dans ce cadre, et après avoir introduit les concepts généraux en biométrie, nous avons présenté un état de l'art liés au connexes de la recherche scientifique dans le domaine où les méthodes et techniques sur la caractérisation paramétrique de l'image, l'optimisation des

caractéristiques ainsi que l'utilisation de l'apprentissage automatique et la classification sont tous mis en revue.

Nos contributions sont représentées dans les deux derniers chapitres. Le chapitre avant le dernier présente l'élaboration de structures de reconnaissance de visage basées sur des descripteurs hybrides classiques optimisés basé essentiellement sur la modalité du visage et ses patches. La fusion multimodale est pris en compte à deux niveaux, au niveau de modalités(Capteur), et au niveau des caractéristiques. L'algorithme d'optimisation choisi est le PSO qui appartient aux méthodes d'optimisations Méta-heuristiques. La fonction coûts (objective) est de valeur : " $fun = mean(data(:))$ ", c'est-à-dire la valeur moyenne de la matrice des caractéristiques. En revanche, il n'existe pas actuellement de critères universels ni de méthodologie standard permettant de comparer des algorithmes de fusion entre eux. Les résultats obtenus par ces approches montrent que le bloc d'extraction des caractéristiques constitué de plusieurs descripteurs en cascade est très efficace. L'algorithme d'optimisation PSO joue un rôle très important dans l'optimisation du temps d'exécution, qui est un paramètre à prendre en considération. La reconnaissance des yeux donne de meilleures TR et temps d'exécution par rapport aux bases de données de la bouche et du visage. Ce résultat qualifie la reconnaissance oculaire comme un bon moyen pour la reconnaissance visage. Pour la fusion des modalités patches, la remarque principale constatée à partir des résultats obtenus est la notion de la modalité dominante, c'est-à-dire : la fusion des modalités prend le meilleur résultat quand on les exploite indépendamment les unes des autres.

Le dernier chapitre consacré au développement et l'élaboration d'une structure d'optimisation de fonctionnalités basée sur la structure Deep Learning et Transfer Learning passant par l'algorithme d'optimisation pour la reconnaissance de visages. Pour cela, des structures basées sur l'apprentissage par transfert « Transfer Learning » est utilisé en s'appuyant sur les réseaux de neurones convolutifs (Convolutional Neural Network (CNN)) et ce afin d'extraire les caractéristiques du visage les plus pertinentes. Dans cette contribution la validation des résultats est menée sur diverses bases de données en milieux contrôlés, non contrôlés et thermiques. Une nouvelle fonction coûts de valeur : " $fun = \frac{1}{1+e^{mean(data(:))}}$ " est choisie, et qui a donnée des résultats encourageants mieux que ceux obtenus par la fonction

coûts choisie pour notre contribution précédente " $fun = mean(data(:))$ ". Une comparaison est élaborée entre les approches à descripteurs classiques et les approches à base du Transfer Learning en Deep Learning. On a constaté que si les bases de données sont de petite taille "Small Data", les bons résultats en terme du temps en particulier, et en termes de TR en générale, sont obtenus pour les approches à descripteurs classiques. En revanche si les bases de données sont de grande taille "Big Data" on a constaté l'inverse.

Par conséquent, on peut tirer que nos résultats obtenus par les deux techniques montrent que l'association de plusieurs descripteurs en cascade et l'utilisation d'une structure CNN AlexNet Transfer Learning pré-entraînée reste une technique intéressante dans le domaine de la reconnaissance de visage. Nos modèles proposés sont appliqués sur différentes bases de données ORL, Yale-B, AR, LFW et Thermal Tufts Face.

Par ailleurs, certaines perspectives restent à explorer. Nous pouvons citer en particulier les travaux suivants :

- D'explorer d'autres méthodes d'optimisation. Cela nous permettra de développer des modèles basés sur plusieurs nouveaux algorithmes d'optimisation, y compris d'autres bases de données et d'autres structures des réseaux CNN.
- Adapter notre modèle à d'autres domaines d'application (géographie, surveillance, etc.) et à d'autres types de médias tels que la vidéo et le son.
- Tester les méthodes proposées avec d'autres modalités telles que l'Iris et l'empreinte digitale, etc.
- Utiliser d'autres algorithmes pour la classification tels que la SVM "Support Vector Machine", la classification hiérarchique, etc.
- Proposer un autre système multimodal basé sur les images de FKPs et de Palmprints par exemple.
- Utiliser d'autres techniques de fusion telles que la fusion au niveau de la décision ou la fusion au niveau des scores.

BIBLIOGRAPHIE

- [1] Chbeir R (2001) *Modélisation de la description d'images : application au domaine médical* (Doctoral dissertation, Lyon, INSA).
- [2] Sutherland IE (1964) SketchPad: A Man-Machine Graphical Communication System. Proceedings AFIPS Spring Joint Computer Conference, Detroit, MI., Vol. 23, P. 329-346. <https://doi.org/10.1177/003754976400200514>
- [3] Ross AA, Nandakumar K, & Jain AK (2006) *Handbook of multibiometrics* (Vol. 6). Springer Science & Business Media.
- [4] Jain A, Flynn P, & Ross AA (Eds.) (2007) *Handbook of biometrics*. Springer Science & Business Media.
- [5] Sim HM, Asmuni H, Hassan R, & Othman RM (2014) Multimodal biometrics: Weighted score level fusion based on non-ideal iris and face images. *Expert Systems with Applications*, 41(11), 5390-5404. <https://doi.org/10.1016/j.eswa.2014.02.051>
- [6] Ross A, & Poh N (2009) Multibiometric systems: Overview, case studies, and open issues. *Handbook of Remote Biometrics*, 273-292.
- [7] Belahcene M (2013) *Authentification et identification en biométrie* (Doctoral dissertation, Université Mohamed Khider Biskra).
- [8] Morizet N (2009) *Reconnaissance biométrique par fusion multimodale du visage et de l'iris* (Doctoral dissertation, Télécom ParisTech).
- [9] Fagertun J (2005) *Face Recognition* (Master's thesis, Technical University of Denmark, DTU, DK-2800 Kgs. Lyngby, Denmark).
- [10] Tortorella F (2005) A ROC-based reject rule for dichotomizers. *Pattern Recognition Letters*, 26(2), 167-180. <https://doi.org/10.1016/j.patrec.2004.09.004> [Get rights and content](#)
- [11] Perronnin, F., & Dugelay, J. L. (2002). Introduction à la biométrie authentification des individus par traitement audio-vidéo. *TS. Traitement du signal*, 19(4), 253-265.
- [12] Bruner JS (1954) The perception of people. *Handbook of social psychology*, Vol. 2, p. 634-654. Addison-Wesley Publishing. ID (NAID)10024076129.
- [13] Bledsoe WW (1968) Semiautomatic facial recognition. Technical report sri project 6693.
- [14] Kanade T (1973) Picture processing by computer complex and recognition of human faces. Ph. D. Thesis, Kyoto University.
- [15] Sirovich L, Kirby M (1987) Low-dimensional procedure for the characterization of human faces. *Journal of Optical Society of America*, Vol. 4, No. 3, pp. 519-524, 1987.
- [16] Vanderploeg RD, Brown WS, & Marsh JT (1987) Judgements of emotion in words and faces: ERP correlates. *International Journal of Psychophysiology*, 5(3), 193-205. [https://doi.org/10.1016/0167-8760\(87\)90006-7](https://doi.org/10.1016/0167-8760(87)90006-7)

BIBLIOGRAPHIE

- [17] Etemad K, & Chellappa R (1997) Discriminant analysis for recognition of human face images. *Josa a*, 14(8), 1724-1733.
- [18] Bartlett MS, Movellan JR, & Sejnowski TJ (2002) Face recognition by independent component analysis. *IEEE Transactions on neural networks*, 13(6), 1450-1464.
- [19] Schölkopf B, Smola A, & Müller KR (1998) Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5), 1299-1319.
- [20] Mika S, Ratsch G, Weston J, Scholkopf B, & Mullers, KR (1999, August) Fisher discriminant analysis with kernels. In *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)* (pp. 41-48). Ieee.
- [21] Penev P and Atick J (1996) Local features analysis : A general statistical theory for object representation. *Neural Systems*, Vol. 7, No. 3, pp. 477–500, 1996.
- [22] Krizhevsky A, Sutskever I, & Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12)*. Curran Associates Inc., Red Hook, NY, USA, 1097–1105.
- [23] Patterson J, & Gibson A (2017) *Deep learning: A practitioner's approach*. " O'Reilly Media, Inc."
- [24] Masi I, Wu Y, Hassner T, & Natarajan P (2018, October). Deep face recognition: A survey. In *2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)* (pp. 471-478). IEEE. <https://doi.org/10.1109/SIBGRAPI.2018.00067>
- [25] Weiss K, Khoshgoftaar TM, & Wang D (2016) A survey of transfer learning. *Journal of Big data*, 3(1), 1-40. <https://doi.org/10.1186/s40537-016-0043-6>
- [26] Wang M, & Deng W (2021) Deep face recognition: A survey. *Neurocomputing*, 429, 215-244. <https://doi.org/10.1016/j.neucom.2020.10.081>
- [27] AbdAlmageed W, Wu Y, Rawls S, Harel S, Hassner T, Masi I, ... & Medioni G (2016, March) Face recognition using deep multi-pose representations. In *2016 IEEE winter conference on applications of computer vision (WACV)* (pp. 1-9). IEEE. <https://doi.org/10.1109/WACV.2016.7477555>
- [28] Wang D, Otto C, & Jain AK (2016) Face search at scale. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1122-1136. <https://doi.org/10.1109/TPAMI.2016.2582166>
- [29] Chen JC, Patel VM, & Chellappa R (2016, March) Unconstrained face verification using deep cnn features. In *2016 IEEE winter conference on applications of computer vision (WACV)* (pp. 1-9). IEEE. <https://doi.org/10.1109/WACV.2016.7477557>
- [30] Elaggoune H, Benouaret M, & Messaadia M (2015, December) Modeling and fault diagnosis sensor by multi-model approach. In *Xème Conférence Internationale: Conception et Production Intégrées*.
- [31] Nakib A (2007) *Conception de métaheuristiques d'optimisation pour la segmentation d'images. Application aux images biomédicales* (Doctoral dissertation, Université Paris XII Val de Marne).
- [32] Schroff F, Kalenichenko D, & Philbin J (2015) Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815-823). <https://doi.org/10.1109/CVPR.2015.7298682>

- [33] Arashloo SR, & Kittler J (2013, September) Efficient processing of MRFs for unconstrained-pose face recognition. In *2013 IEEE sixth international conference on biometrics: theory, applications and systems (BTAS)* (pp. 1-8). IEEE. <https://doi.org/10.1109/BTAS.2013.6712721>
- [34] Karaaba M, Surinta O, Schomaker L, & Wiering MA (2015, December) Robust face recognition by computing distances from multiple histograms of oriented gradients. In *2015 IEEE Symposium Series on Computational Intelligence* (pp. 203-209). IEEE. <https://doi.org/10.1109/SSCI.2015.39>
- [35] Arigbabu OA, Ahmad SMS, Adnan WAW, Yussof S, & Mahmood S (2017) Soft biometrics: Gender recognition from unconstrained face images using local feature descriptor. *arXiv preprint arXiv:1702.02537*.
- [36] Seo HJ, & Milanfar P (2011) Face verification using the lark representation. *IEEE Transactions on Information Forensics and Security*, 6(4), 1275-1286. <https://doi.org/10.1109/TIFS.2011.2159205>
- [37] Preeti & Kumar D (2018) Performance Evaluation of Face Recognition System using various Distance Classifiers. Second International Conference on Computing Methodologies and Communication (ICCMC), 2018, pp. 322-327. <https://doi.org/10.1109/ICCMC.2018.8487835>
- [38] Huang M, Shao G, Wang K, Liu T, & Lu H (2021) Discriminative locality-constrained sparse representation for robust face recognition. In *Journal of Physics: Conference Series* (Vol. 1780, No. 1, p. 012034). IOP Publishing. <https://doi.org/10.1088/1742-6596/1780/1/012034>
- [39] Malhotra P, & Kumar D (2019) An optimized face recognition system using cuckoo search. *Journal of Intelligent Systems*, 28(2), 321-332. <https://doi.org/10.1515/jisys-2017-0127>
- [40] Keinert F, Lazzaro D, & Morigi S (2019) A robust group-sparse representation variational method with applications to face recognition. *IEEE Transactions on Image Processing*, 28(6), 2785-2798. <https://doi.org/10.1109/tip.2018.2890312>
- [41] Peng Y, Li L, Liu S, & Lei T (2018) Space-frequency domain based joint dictionary learning and collaborative representation for face recognition. *Signal processing*, 147, 101-109. <https://doi.org/10.1016/j.sigpro.2018.01.013>
- [42] Shang K, Huang ZH, Liu W, & Li ZM (2018) A single gallery-based face recognition using extended joint sparse representation. *Applied Mathematics and Computation*, 320, 99-115. <https://doi.org/10.1016/j.amc.2017.07.058>
- [43] Liu J, Liu W, Ma S, Wang M, Li L, & Chen G (2019) Image-set based face recognition using K-SVD dictionary learning. *International Journal of Machine Learning and Cybernetics*, 10(5), 1051-1064. <https://doi.org/10.1007/s13042-017-0782-5>
- [44] Ayyavoo T, & Suseela JJ (2018) Illumination pre-processing method for face recognition using 2D DWT and CLAHE. *Iet Biometrics*, 7(4), 380-390. <https://doi.org/10.1049/iet-bmt.2016.0092>
- [45] Elaggoune H, Belahcene M, & Bourennane S (2022) Hybrid descriptor and optimized CNN with transfer learning for face recognition. *Multimedia Tools and Applications*, 81(7), 9403-9427. <https://doi.org/10.1007/s11042-021-11849-1>
- [46] Thangavel K, Manavalan R (2014) Soft computing models based feature selection for TRUS prostate cancer image classification. *J Soft Comput* 18(6):1165–1176. <https://doi.org/10.1007/s00500-013-1135-2>

- [47] Elaggoune H, Belahcene M, Messaadia M, & Benatia MA (2021, May) Hybrid Descriptor and Patches Optimization for Face Recognition. In *2021 1st International Conference On Cyber Management And Engineering (CyMaEn)* (pp. 1-6). IEEE. <https://doi.org/10.1109/CyMaEn50288.2021.9497271>
- [48] Yaji GS, Sarkar S, Manikantan K, & Ramachandran S (2012) DWT feature extraction based face recognition using intensity mapped unsharp masking and laplacian of gaussian filtering with scalar multiplier. *Procedia Technology*, 6, 475-484.
- [49] Ramadan RM, & Abdel-Kader RF (2009, December) Particle swarm optimization for human face recognition. In *2009 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)* (pp. 579-584). IEEE.
- [50] Yin H, Qiao J, Fu P, & Xia X (2014). Face Feature Selection with Binary Particle Swarm Optimization and Support Vector Machine. *J. Inf. Hiding Multim. Signal Process.*, 5(4), 731-739.
- [51] Khadhraoui T, Ktata S, Benzarti F, & Amiri H (2016, March) Features selection based on modified PSO algorithm for 2D face recognition. In *2016 13th international conference on computer graphics, imaging and visualization (CGiV)* (pp. 99-104). IEEE.
- [52] Sasirekha K, & Thangavel K (2019) Optimization of K-nearest neighbor using particle swarm optimization for face recognition. *Neural Computing and Applications*, 31(11), 7935-7944. <https://doi.org/10.1007/s00521-018-3624-9>
- [53] Samma H, Suandi SA, & Mohamad-Saleh J (2019) Face sketch recognition using a hybrid optimization model. *Neural Computing and Applications*, 31(10), 6493-6508.
- [54] Zhou G, Ouyang A, & Xu Y (2020) Chaos Glowworm Swarm Optimization Algorithm Based on Cloud Model for Face Recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 34(12), 2056009. <https://doi.org/10.1142/S0218001420560091>
- [55] Di X, Riggan BS, Hu S, Short NJ, & Patel VM (2021) Multi-scale thermal to visible face verification via attribute guided synthesis. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(2), 266-280. <https://doi.org/10.1109/TBIOM.2021.3060641>
- [56] Taigman Y, Yang M, Ranzato MA, & 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). <https://doi.ieeecomputersociety.org/10.1109/CVPR.2014.220>
- [57] Huang GB, Mattar M, Berg T, & Learned-Miller E (2008, October) 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*.
- [58] Liu C, & Wechsler H (2002) Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image processing*, 11(4), 467-476. <https://doi.org/10.1109/TIP.2002.999679>
- [59] Mehdipour Ghazi M, & Kemal Ekenel H (2016) A comprehensive analysis of deep learning based representation for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 34-41). <https://doi.org/10.1109/CVPRW.2016.20>
- [60] Hu G, Yang Y, Yi D, Kittler J, Christmas W, Li SZ, & Hospedales T (2015) When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 142-150). <https://doi.org/10.1109/ICCVW.2015.58>

- [61] Wang X, Wang K, & Lian S (2020) A survey on face data augmentation for the training of deep neural networks. *Neural computing and applications*, 32(19), 15503-15531.
- [62] Lv JJ, Shao XH, Huang JS, Zhou XD, & Zhou X (2017) Data augmentation for face recognition. *Neurocomputing*, 230, 184-196.
- [63] Choi Y, Choi M, Kim M, Ha JW, Kim S, & Choo J (2018) Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797). <https://doi.org/10.1109/CVPR.2018.00916>
- [64] Dalvi J, Bafna S, Bagaria D, & Virnodkar S (2022) A Survey on Face Recognition Systems. *arXiv preprint arXiv:2201.02991*. <https://doi.org/10.48550/arXiv.2201.02991>
- [65] Simonyan K, & Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [66] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, ... & Rabinovich A (2015) Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9). <https://doi.org/10.1109/CVPR.2015.7298594>
- [67] 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). <https://doi.org/10.1109/CVPR.2016.90>
- [68] Hu J, Shen L, & Sun G (2018) Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132-7141). <https://doi.org/10.1109/CVPR.2018.00745>
- [69] 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). <https://doi.org/10.1109/CVPR.2014.244>
- [70] Xing E, Jordan M, Russell S J, & Ng A (2002) Distance metric learning with application to clustering with side-information. *Advances in neural information processing systems*, 15.
- [71] Weinberger KQ, & Saul LK (2009) Distance metric learning for large margin nearest neighbor classification. *Journal of machine learning research*, 10(2).
- [72] Parkhi OM, Vedaldi A, & Zisserman A (2015) Deep face recognition.
- [73] Sankaranarayanan S, Alavi A, & Chellappa R (2016) Triplet similarity embedding for face verification. *arXiv preprint arXiv:1602.03418*.
- [74] Sankaranarayanan S, Alavi A, Castillo CD, & Chellappa R (2016, September) Triplet probabilistic embedding for face verification and clustering. In *2016 IEEE 8th international conference on biometrics theory, applications and systems (BTAS)* (pp. 1-8). IEEE. <https://doi.org/10.1109/BTAS.2016.7791205>
- [75] Liu J, Deng Y, Bai T, Wei Z, & Huang C (2015) Targeting ultimate accuracy: Face recognition via deep embedding. *arXiv preprint arXiv:1506.07310*.
- [76] Ding C, & Tao D (2015) Robust face recognition via multimodal deep face representation. *IEEE transactions on Multimedia*, 17(11), 2049-2058. <https://doi.org/10.1109/TMM.2015.2477042>

BIBLIOGRAPHIE

- [77] Sun Y, Wang X, & Tang X (2015) Deeply learned face representations are sparse, selective, and robust. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2892-2900). <https://doi.org/10.1109/CVPR.2015.7298907>
- [78] Sun Y, Chen Y, Wang X, & Tang X (2014) Deep learning face representation by joint identification-verification. *Advances in neural information processing systems*, 27.
- [79] Sun Y, Liang D, Wang X, & Tang X (2015) Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv:1502.00873*.
- [80] Sun Y, Wang X, & Tang X (2016) Sparsifying neural network connections for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4856-4864).
- [81] Yi D, Lei Z, Liao S, & Li SZ (2014) Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*.
- [82] Liu W, Wen Y, Yu Z, Li M, Raj B, & Song L (2017) Spheroface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 212-220). <https://doi.org/10.1109/CVPR.2017.713>
- [83] Liu W, Wen Y, Yu Z, & Yang M (2016, June) Large-margin softmax loss for convolutional neural networks. In *ICML* (Vol. 2, No. 3, p. 7).
- [84] Wang F, Cheng J, Liu W, & Liu H (2018) Additive margin softmax for face verification. *IEEE Signal Processing Letters*, 25(7), 926-930. <https://doi.org/10.1109/LSP.2018.2822810>
- [85] 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). <https://doi.org/10.1109/CVPR.2019.00482>
- [86] Liu W, Zhang, Y M, Li X, Yu Z, Dai B, Zhao T, & Song L (2017) Deep hyperspherical learning. *Advances in neural information processing systems*, 30.
- [87] Guo Y, Zhang L, Hu Y, He X, & Gao J (2016, October) Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European conference on computer vision* (pp. 87-102). Springer, Cham.
- [88] "Ms-celeb-1m challenge 3" <http://trillionpairs.deepglint.com>
- [89] Kemelmacher-Shlizerman I, Seitz SM, Miller D, & Brossard E (2016) The megaface benchmark: 1 million faces for recognition at scale. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4873-4882). <https://doi.org/10.1109/CVPR.2016.527>
- [90] Nech A, & Kemelmacher-Shlizerman I (2017). Level playing field for million scale face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7044-7053). <https://doi.org/10.1109/CVPR.2017.363>
- [91] Zhu Z, Huang G, Deng J, Ye Y, Huang J, Chen X, ... & Zhou, J. (2021). Webface260m: A benchmark unveiling the power of million-scale deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10492-10502). <https://doi.org/10.1109/CVPR46437.2021.01035>

- [92] Masi I, Trần AT, Hassner T, Leksut JT, & Medioni G (2016, October) Do we really need to collect millions of faces for effective face recognition?. In *European conference on computer vision* (pp. 579-596). Springer, Cham. https://doi.org/10.1007/978-3-319-46454-1_35
- [93] Cao Q, Shen L, Xie W, Parkhi OM, & Zisserman A (2018, May) Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)* (pp. 67-74). IEEE. <https://doi.org/10.1109/FG.2018.00020>
- [94] Maze B, Adams J, Duncan JA, 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. <https://doi.org/10.1109/ICB2018.2018.00033>
- [95] Damer N, Boutros F, Mallat K, Kirchbuchner F, Dugelay JL, & Kuijper A (2019). Cascaded generation of high-quality color visible face images from thermal captures. ArXiv, abs/1910.09524.
- [96] Sasirekha K, & Thangavel K (2020) Biometric face classification with the hybridised rough neural network. *International Journal of Biometrics*, 12(2), 193-217. <https://dx.doi.org/10.1504/IJBM.2020.107717>
- [97] Talab MA, Awang S, & Najim SADM (2019, June) Super-low resolution face recognition using integrated efficient sub-pixel convolutional neural network (ESPCN) and convolutional neural network (CNN). In *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, 2019, pp. 331-335. <https://doi.org/10.1109/I2CACIS.2019.8825083>
- [98] Muhammad HI, Musa KI, Abdulrahman ML, Abubakar A, Umar K, & Ishola A (2021) Enhancing Detection Performance of Face Recognition Algorithm Using PCA-Faster R-CNN. *European Journal of Electrical Engineering and Computer Science*, 5(3), 9-16. <https://doi.org/10.24018/ejece.2021.5.3.321>
- [99] Almabdy S, & Elrefaei L (2019) Deep convolutional neural network-based approaches for face recognition. *Applied Sciences*, 9(20), 4397. <https://doi.org/10.3390/app9204397>
- [100] Cheng EJ, Chou KP, Rajora S, Jin BH, Tanveer M, Lin CT, ... & Prasad M (2019) Deep sparse representation classifier for facial recognition and detection system. *Pattern Recognition Letters*, 125, 71-77. <https://doi.org/10.1016/j.patrec.2019.03.006>
- [101] Manzuri M, Pooyan M, & Rikhtegar A (2020) Comparing performance of metaheuristic algorithms for finding the optimum structure of CNN for face recognition. *International Journal of Nonlinear Analysis and Applications*, 11(1), pp. 301-319. <https://doi.org/10.22075/ijnaa.2020.4296>
- [102] 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. 4685-4694). <https://doi.org/10.1109/CVPR.2019.00482>
- [103] Yin X, Yu X, Sohn K, Liu X, & Chandraker M (2018) Feature transfer learning for deep face recognition with under-represented data. ArXiv, abs/1803.09014.
- [104] Kortli Y, Jridi M, Al Falou A, & Atri M (2020) Face recognition systems: A survey. *Sensors*, 20(2), 342.
- [105] Ballantyne M, Boyer RS, & Hines L (1996) Woody Bledsoe: His life and legacy. *AI magazine*, vol. 17, no 1, p. 7-7.

BIBLIOGRAPHIE

- [106] Wong Y, Chen S, Mau S, Sanderson C, & Lovell BC (2011, June) Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In *CVPR 2011 WORKSHOPS* (pp. 74-81). IEEE.
- [107] Ma X, Zhang J, & Qi C (2010) Hallucinating face by position-patch. *Pattern Recognition*, 43(6), 2224-2236.
- [108] Bessaoudi M, Belahcene M, Ouamane A, Chouchane A, & Bourennane S (2018, March) A novel approach based on high order tensor and multi-scale locals features for 3D face recognition. In 2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) (pp. 1-5). IEEE.
- [109] Chouchane A, Belahcene M, Ouamane A, & Bourennane S (2014, October) 3D face recognition based on histograms of local descriptors. In *2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA)* (pp. 1-5). IEEE.
- [110] Cheon M, Lee W, Yoon C, & Park M (2012) Vision-based vehicle detection system with consideration of the detecting location. *IEEE transactions on intelligent transportation systems*, 13(3), 1243-1252.
- [111] Déniz O, Bueno G, Salido J, & De la Torre F (2011) Face recognition using histograms of oriented gradients. *Pattern recognition letters*, 32(12), 1598-1603.
- [112] Dalal N, & Triggs B (2005, June) Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE.
- [113] Kennedy J, & Eberhart R (1995, November) Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks* (Vol. 4, pp. 1942-1948). IEEE.
- [114] Ballardini AL (2018) A tutorial on particle swarm optimization clustering. *arXiv preprint arXiv:1809.01942*.
- [115] Turan C, & Lam KM (2018) Histogram-based local descriptors for facial expression recognition (FER): A comprehensive study. *Journal of visual communication and image representation*, 55, 331-341.
- [116] Turan C, & Lam KM (2018) Histogram-based local descriptors for facial expression recognition (FER): A comprehensive study. *Journal of visual communication and image representation*, 55, 331-341.
- [117] Ameer B, Belahcene M, Masmoudi S, Derbel AG, & Hamida AB (2017, May). A new GLBSIF descriptor for face recognition in the uncontrolled environments. In *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* (pp. 1-6). IEEE.
- [118] Setiowati S, Franita EL, & Ardiyanto I (2017, October) A review of optimization method in face recognition: Comparison deep learning and non-deep learning methods. In *2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE)* (pp. 1-6). IEEE.
- [119] Kirby M, & Sirovich L (1990) Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern analysis and Machine intelligence*, 12(1), 103-108. <https://doi.org/10.1109/34.41390>

- [120] Arora M, & Kumar M (2021) AutoFER: PCA and PSO based automatic facial emotion recognition. *Multimedia Tools and Applications*, 80(2), 3039-3049. <https://doi.org/10.1007/s11042-020-09726-4>
- [121] Zhang K, Zhang Z, Li Z, & Qiao Y (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10), 1499-1503.
- [122] Štruc V, & Pavešić N (2010) The complete gabor-fisher classifier for robust face recognition. *EURASIP Journal on Advances in Signal Processing*, 2010, 1-26.
- [123] Ren J, Jiang X, & Yuan J (2013, September) Relaxed local ternary pattern for face recognition. In *2013 IEEE international conference on image processing* (pp. 3680-3684). IEEE. <https://doi.org/10.1109/ICIP.2013.6738759>
- [124] Napoléon T, & Alfalou A (2014, May) Local binary patterns preprocessing for face identification/verification using the VanderLugt correlator. In *Optical Pattern Recognition XXV* (Vol. 9094, pp. 30-35). SPIE. <https://doi.org/10.1117/12.2051267>
- [125] Forsyth GA, Kushner RI, & Forsyth PD (1981) Human facial expression judgment in a conversational context. *Journal of Nonverbal Behavior*, 6(2), 115-130. <https://doi.org/10.1007/BF00987287>
- [126] Kirby M, & Sirovich L (1990) Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern analysis and Machine intelligence*, 12(1), 103-108. <https://doi.org/10.1109/34.41390>
- [127] Oloyede MO, Hancke GP, & Myburgh HC (2020) A review on face recognition systems: recent approaches and challenges. *Multimedia Tools and Applications*, 79(37), 27891-27922. <https://doi.org/10.1007/s11042-020-09261-2>
- [128] Hermosilla G, Rojas M, Mendoza J, Farias G, Pizarro FT, San Martin C, & Vera E (2018) Particle swarm optimization for the fusion of thermal and visible descriptors in face recognition systems. *IEEE Access*, 6, 42800-42811. <https://doi.org/10.1109/ACCESS.2018.2850281>
- [129] Wu H, Zhang K, & Tian G (2018) Simultaneous face detection and pose estimation using convolutional neural network cascade. *IEEE Access*, 6, 49563-49575. <https://doi.org/10.1109/ACCESS.2018.2869465>
- [130] Li S, Dou Y, Xu J, Yang K, & Li R (2019) GBCNN: A Full GPU-Based Batch Multi-Task Cascaded Convolutional Networks. *IEEE Access*, 7, 20225-20234. <https://doi.org/10.1109/ACCESS.2019.2894589>
- [131] Michelucci, U. (2019). *Advanced applied deep learning: convolutional neural networks and object detection*. Apress.
- [132] Mallouh, A. A., Qawaqneh, Z., & Barkana, B. D. (2019). Utilizing CNNs and transfer learning of pre-trained models for age range classification from unconstrained face images. *Image and Vision Computing*, 88, 41-51. <https://doi.org/10.1016/j.imavis.2019.05.001>
- [133] Du L, Hu H, & Wu Y (2019) Age factor removal network based on transfer learning and adversarial learning for cross-age face recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(9), 2830-2842. <https://doi.org/10.1109/TCSVT.2019.2923262>
- [134] Huang Z, Cao Y, & Wang T (2019, March) Transfer learning with efficient convolutional neural networks for fruit recognition. In *2019 IEEE 3rd Information Technology, Networking, Electronic*

- and Automation Control Conference (ITNEC)* (pp. 358-362). IEEE.
<https://doi.org/10.1109/ITNEC.2019.8729435>
- [135] Adjabi I, Ouahabi A, Benzaoui A, & Taleb-Ahmed A (2020) Past, present, and future of face recognition: A review. *Electronics*, 9(8), 1188. <https://doi.org/10.3390/electronics9081188>
- [136] Wu X, He R, Sun Z, & Tan T (2018). A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*, 13(11), 2884-2896. <https://doi.org/10.1109/TIFS.2018.2833032>
- [137] Li S, Dou Y, Xu J, Yang K, & Li R (2019) GBCNN: A Full GPU-Based Batch Multi-Task Cascaded Convolutional Networks. *IEEE Access*, 7, 20225-20234. <https://doi.org/10.1109/ACCESS.2019.2894589>
- [138] Du L, Hu H, & Wu Y (2019) Age factor removal network based on transfer learning and adversarial learning for cross-age face recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(9), 2830-2842. <https://doi.org/10.1109/ACCESS.2019.2894589>
- [139] Bessaoudi M, Belahcene M, Ouamane A, Chouchane A, Bourennane S (2019) A Novel Hybrid Approach for 3D Face Recognition Based on Higher Order Tensor. In: Demigha O, Djamaa B, Amamra A. (eds) *Advances in Computing Systems and Applications*. CSA 2018. Lecture Notes in Networks and Systems, vol 50. Springer, Cham. https://doi.org/10.1007/978-3-319-98352-3_23
- [140] LeCun Y, Bottou L, Bengio Y, & Haffner P (1998) Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. <https://doi.org/10.1109/5.726791>
- [141] Liu X, Liu Z, Wang G, Cai Z, & Zhang H (2017) Ensemble transfer learning algorithm. *IEEE Access*, 6, 2389-2396. <https://doi.org/10.1109/ACCESS.2017.2782884>
- [142] Martinez A, & Benavente Robert (1998) The AR face database. Tech. Rep. 24 CVC Technical Report.
- [143] Huang GB, Mattar M, Berg T, & Learned-Miller E (2008, October) 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*.
- [144] Huang GB, & Learned-Miller E (2014) Labeled faces in the wild: Updates and new reporting procedures. Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep, 14(003).
- [145] Panetta K, Wan Q, Aгаian S, Rajeev S, Kamath S, Rajendran R, and all (2018) A comprehensive database for benchmarking imaging systems. *IEEE transactions on pattern analysis and machine intelligence*, 42(3), 509-520. <https://doi.org/10.1109/TPAMI.2018.2884458>
- [146] Ouyang A, Liu Y, Pei S, Peng X, He M, & Wang Q (2020) A hybrid improved kernel LDA and PNN algorithm for efficient face recognition. *Neurocomputing*, 393, 214-222. <https://doi.org/10.1016/j.neucom.2019.01.117>
- [147] Song X, Hu G, Luo JH, Feng Z, Yu DJ, & Wu XJ (2019) Fast SRC using quadratic optimisation in downsized coefficient solution subspace. *Signal Processing*, 161, 101-110. <https://doi.org/10.1016/j.sigpro.2019.03.007>

BIBLIOGRAPHIE

- [148] Du Q, & Da F (2021) Block dictionary learning-driven convolutional neural networks for fewshot face recognition. *The visual Computer*, 37(4), 663-672. <https://doi.org/10.1007/s00371-020-01802-y>
- [149] Sun Y and Wang H (2017, March) Face Recognition Based on Circularly Symmetrical Gabor Transforms and Collaborative Representation. In 2nd International Conference on Multimedia and Image Processing (ICMIP), 2017, pp. 103-107. <https://doi.org/10.1109/ICMIP.2017.32>
- [150] Ameer B, Belahcene M, Masmoudi S, Hamida AB (2019) Hybrid descriptors and Weighted PCA-EFMNet for Face Verification in the Wild. *International Journal of Multimedia Information Retrieval*, 8(3), 143-154. <https://doi.org/10.1007/s13735-019-00175-w>
- [151] Heidari M and Fouladi-Ghaleh K (2020, February) Using Siamese Networks with Transfer Learning for Face Recognition on Small-Samples Datasets. *International Conference on Machine Vision and Image Processing (MVIP)*, 2020, pp. 1-4. <https://doi.org/10.1109/MVIP49855.2020.9116915>