



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة أكلي محمد أولحاج - البويرة  
Université Akli Mohand Oulhadj - Bouira



Faculté des Sciences et des Sciences Appliquées  
Département d'Informatique

Laboratoire d'informatique, Mathématique et Physique pour l'Agriculture et  
les Forêts (LIMPAF)

## Thèse

Présentée en vue de l'obtention du diplôme de:

**Doctorat 3<sup>ème</sup> Cycle en Informatique**

Option : Réseaux et Sécurité

Présentée par: **Insaf Adjabi**

Intitulée:

**Reconnaissance Faciale Utilisant des Réseaux de Neurones  
Convolutifs et des Descripteurs de Texture dans des  
Environnements Contraints et Non-Contraints**

### Membres du Jury

Président	Mohamed Lamine BOUCENNA	MCA	Université de Bouira
Directeur de Thèse	Abdeldjalil OUAHABI	Prof	Université de Tours
Examineurs	Abderrahmane AMROUCHE	Prof	Université des Sciences et de la Technologie Houari Boumediène - Alger
	Zahira CHOUIREF	MCA	Université de Bouira
	Bilal SAOUD	MCA	Université de Bouira

Soutenue le 18 Janvier 2024



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



## ***Remerciements***

Tout d'abord, je tiens à remercier Allah le Tout-Puissant de m'avoir donné la force, le courage et la patience nécessaires pour mener à bien ce travail.

Je tiens à exprimer ma sincère gratitude à mon superviseur, **Prof. Abdeldjalil OUAHABI**, pour avoir rendu ce travail possible. Ses conseils et ses orientations m'ont accompagné à chaque étape de la rédaction.

Je souhaite également remercier les membres du jury, le **Dr. Mohamed Lamine BOUCENNA**, le **Prof. Abderrahmane AMROUCHE**, le **Dr. Zahira CHOUIREF** et le **Dr. Bilal SAOUD**, pour l'honneur qu'ils ont accepté de juger ce travail.

Des remerciements spéciaux sont également adressés au **Prof. Amir BENZAOU**I de l'Université de SKIKDA pour son aide, son encouragement et son soutien.

Je souhaite dédier ce travail à ma famille, à mes parents, à mon mari et à mes enfants, Assinat et Anas, pour leur soutien continu et leur compréhension tout au long de mes recherches. Vos prières m'ont soutenu et je suis reconnaissante de votre amour et de votre soutien.

## *Résumé*

La reconnaissance faciale est un domaine de recherche très actif dans la vision par ordinateur. Grâce à l'intelligence artificielle, des avancées significatives ont été réalisées dans ce domaine. Les premières études se concentraient sur des conditions contrôlées dont les approches traditionnelles ont donné de bons résultats. Cependant, ces dernières années, la recherche s'est orientée vers des conditions non contrôlées où l'apprentissage profond, en particulier les réseaux de neurones convolutifs (en anglais, convolutional neural network (CNN)), s'est révélé très efficace pour faire face aux variations telles que les changements d'expression, l'éclairage, et l'occlusion, qui peuvent affecter la reconnaissance faciale. Dans les situations de la vie réelle tels que la surveillance de la sécurité dans les aéroports, les contrôles frontaliers, ou encore les vérifications d'identité en ligne, la reconnaissance faciale à échantillon unique par personne (en anglais, single sample per person (SSPP)) est largement utilisée et considéré comme un défi majeur en reconnaissance faciale, car il n'existe qu'une seule image par personne dans l'ensemble d'apprentissage. Pour relever ce défi, trois contributions sont développées et évaluées dans le cadre de cette thèse. La première contribution propose une méthode efficace appelée multi-block color binarized statistical image features (MB-C-BSIF) qui exploite les informations de texture et de couleur pour obtenir une représentation précise des visages. La deuxième contribution combine une approche traditionnelle avec l'apprentissage profond en utilisant le descripteur MB-C-BSIF et un CNN pour bénéficier les avantages de ces deux méthodes. Enfin, la troisième contribution propose une méthode d'analyse en multi-résolution basée sur la transformée en ondelettes discrètes (en anglais, discrete wavelet transform (DWT)), permettant d'extraire des caractéristiques statistiques à différentes résolutions pour améliorer la reconnaissance des visages dans des conditions variables. Les résultats obtenus pour chacune des contributions de cette thèse ont démontré des performances excellentes par rapport aux autres méthodes existantes: 96.17 % pour la méthode MB-C-BSIF, 97.33 % pour la méthode hybride, et 98.00 % pour la méthode d'analyse en multi-résolution, en ce qui concerne la reconnaissance des variations en expression faciale dans la base de données Alex and Robert (AR).

**Mots clés:** Biométrie, reconnaissance faciale, réseaux de neurones convolutifs (CNN), reconnaissance faciale à échantillon unique.

## *Abstract*

Facial recognition is a highly active research area in computer vision. Thanks to artificial intelligence, significant advancements have been made in this field. Early studies focused on controlled conditions and utilized traditional approaches with good results. However, in recent years, research has shifted towards uncontrolled conditions where deep learning, particularly convolutional neural networks (CNNs), has proven to be highly effective in dealing with variations such as changes in expression, lighting, and occlusion that can affect facial recognition. In real-life situations such as airport security surveillance, border controls, or online identity verification, single sample per person (SSPP) facial recognition is widely used and considered a major challenge. This is due to the fact that there is only one image per person in the training dataset. To address this challenge, three contributions are developed and evaluated in the context of this thesis. The first contribution proposes an efficient method called multi-block color binarized statistical image features (MB-C-BSIF) that leverages texture and color information to obtain an accurate representation of faces. The second contribution combines a traditional approach with deep learning by using the MB-C-BSIF descriptor and a CNN to benefit from the advantages of both methods. Finally, the third contribution proposes a multi-resolution analysis method based on the discrete wavelet transform (DWT), allowing for the extraction of statistical features at different resolutions to improve facial recognition under variable conditions. The results obtained for each contribution in this thesis demonstrated excellent performance compared to other existing methods: 96.17% for the MB-C-BSIF method, 97.33% for the hybrid method, and 98.00% for the multi-resolution analysis method, concerning the recognition of variations in facial expression in the Alex and Robert (AR) dataset.

**Keywords:** Biometrics, facial recognition, convolutional neural networks (CNN), single sample per person facial recognition.

## المُلخَص

إن مجال التعرف على الوجوه هو مجال بحث نشط جدًا في مجال رؤية الحاسوب. بفضل التقدم الكبير الذي تم تحقيقه في هذا المجال بفضل التطورات في ميدان الذكاء الاصطناعي. ركزت الدراسات الأولية على الظروف المراقبة حيث حققت النهج التقليدي نجاحًا جيدًا. ومع ذلك، في السنوات الأخيرة، تحول البحث إلى الظروف غير مسيطر عليها حيث أظهر التعلم العميق، وبشكل خاص شبكات الأعصاب التلافيفية (CNN)، فعالية كبيرة في التعامل مع التغيرات مثل التغيرات في التعابير الوجهية والإضاءة والتظليل، والتي قد تؤثر على التعرف على الوجوه. في السياقات الحياتية الواقعية مثل مراقبة الأمان في المطارات، والتفتيش عند الحدود، أو حتى التحقق من الهوية عبر الإنترنت، يُستخدم التعرف على الوجوه بعينة واحدة لكل شخص (SSPP) بشكل واسع ويُعتبر تحديًا كبيرًا في مجال التعرف على الوجوه، حيث يتوفر صورة واحدة فقط لكل فرد في مجموعة البيانات التعلم. لمواجهة هذا التحدي، تم تطوير وتقييم ثلاثة مساهمات في إطار هذه الأطروحة. تقدم المساهمة الأولى طريقة فعالة تسمى " ميزات الصورة الإحصائية الثنائية الملونة ومتعددة الكتل (MB-C-BSIF) "والتي تستغل معلومات القوام واللون للحصول على تمثيل دقيق للوجوه. المساهمة الثانية تجمع بين نهج تقليدي والتعلم العميق وذلك باستخدام الوصف MB-C-BSIF وشبكة CNN للاستفادة من مزايا كلتا الطريقتين. وأخيرًا، تقدم المساهمة الثالثة طريقة تحليل متعددة الدقة تعتمد على تحويل الموجات المنفصلة (DWT)، مما يسمح باستخراج معلومات إحصائية على مختلف الدقات لتحسين التعرف على الوجوه في ظروف متغيرة. أظهرت النتائج التي تم الحصول عليها لكل من المساهمات في هذه الرسالة أداءً ممتازًا مقارنةً بالطرق الأخرى الموجودة: 96.17% للمساهمة الأولى، 97.33% للمساهمة المختلطة، و 98.00% للمساهمة في التحليل متعدد الدقة، فيما يتعلق بالتعرف على التغيرات في التعابير الوجهية في قاعدة البيانات AR .

**الكلمات الرئيسية:** البيومترية، التعرف على الوجوه، الشبكات العصبية التلافيفية (CNN)، التعرف على الوجوه بعينة واحدة.

## *Liste des abréviations*

<b>AR</b>	Alex and Robert
<b>BDL</b>	Block Dictionary Learning
<b>BSIF</b>	Binarized Statistical Image Features
<b>CDA</b>	Class-Specific Deep Auto-Encoder
<b>CLAHE</b>	Contrast Limited Adaptive Histogram Equalization
<b>CNN</b>	Convolutional Neural Network
<b>DMMA</b>	Discriminant Multi-Manifold Analysis
<b>DNNC</b>	Dissimilarity-Based Nearest Neighbor Classifier
<b>DWT</b>	Discrete Wavelet Transform
<b>EER</b>	Equal Error Rate
<b>ESRC</b>	Extended Sparse Representation Classifier
<b>FAR</b>	False Accept Rate
<b>FC</b>	Fully Connected
<b>FRR</b>	False Rejection Rate
<b>HOG</b>	Histograms Of Oriented Gradients
<b>ICA</b>	Independent Component Analysis
<b>ILSVRC</b>	Imagenet Large Scale Visual Recognition Challenge
<b>ISMMDL</b>	Image Set-Based Multimanifold Discriminant Learning
<b>KNN</b>	K-Nearest Neighbors
<b>KNNMMDL</b>	K-Nearest Neighbors Virtual Image Set-Based Multimanifold Discriminant Learning
<b>KNNVSG</b>	K-Nearest Neighbor-Based Virtual Sample Generating
<b>KPCANet</b>	Kernel Principal Component Analysis Network
<b>LBP</b>	Local Binary Pattern
<b>LFW</b>	Labeled Faces in The Wild
<b>LGR</b>	Local Generic Representation
<b>LPQ</b>	Local Phase Quantization
<b>LRSR</b>	Local Robust Sparse Representation
<b>MB-C-BSIF</b>	Multi-Block Color Binarized Statistical Image Features
<b>MRMR</b>	Maximum Relevance - Minimum Redundancy
<b>NNC</b>	Nearest Neighbor Classifier
<b>PCA</b>	Principal Component Analysis
<b>PCRC</b>	Patch Collaborative Representation-Based Classification
<b>ROC</b>	Courbe Receiver Operator Characteristic
<b>RR</b>	Recognition Rate
<b>SSFR</b>	Single Sample Face Recognition
<b>SSPP</b>	Single Sample Per Person
<b>SVD</b>	Singular Value Decomposition
<b>SVDL</b>	Sparse Variation Dictionary Learning
<b>TAR</b>	True Accept Rate

# Table des matières

Introduction Générale.....	1
----------------------------	---

## **CHAPITRE 1 : Principes de Biométrie et Systèmes Biométriques.**

1.1. Introduction .....	4
1.2. Définition de la biométrie.....	4
1.3. Différentes modalités biométriques .....	5
1.3.1. Reconnaissance faciale .....	6
1.3.2. Reconnaissance de l'iris .....	6
1.3.3. Reconnaissance de la rétine.....	7
1.3.4. Reconnaissance de l'oreille .....	8
1.3.5. Reconnaissance d'empreintes digitales .....	8
1.3.6. Reconnaissance de la géométrie de la main .....	9
1.3.7. Reconnaissance de la signature.....	10
1.3.8. Reconnaissance de la voix.....	10
1.3.9. Reconnaissance de la dynamique des frappes.....	11
1.3.10. Reconnaissance de la démarche.....	11
1.4. Caractéristiques biométriques .....	12
1.5. Architecture fonctionnelle d'un système biométrique .....	13
1.5.1. Vérification vs Identification .....	14
1.5.2. Modules d'un système biométrique.....	14
1.6. Performance et évaluation des systèmes biométriques .....	16
1.6.1. Evaluation de la vérification .....	16
1.6.2. Evaluation de l'identification.....	17
1.7. Applications de la biométrie .....	18
1.8. Motivation : Pourquoi le visage?.....	19
1.9. Conclusion.....	19

## **CHAPITRE 2: La Reconnaissance Faciale.**

2.1. Introduction .....	21
2.2. Le visage en tant que modalité biométrique .....	21



2.3.	Etapes essentielles pour les systèmes de reconnaissance faciale .....	23
2.4.	Défis ouverts de la reconnaissance faciale .....	24
2.4.1.	Reconnaissance faciale et occlusion .....	25
2.4.2.	Reconnaissance de visages hétérogènes .....	25
2.4.3.	Reconnaissance faciale et vieillissement .....	26
2.4.4.	Reconnaissance faciale 3D .....	27
2.4.5.	Reconnaissance faciale dans la vidéosurveillance .....	27
2.4.6.	Reconnaissance faciale à partir d'un seul échantillon .....	28
2.5.	Pourquoi la reconnaissance faciale à partir d'un seul échantillon? .....	29
2.6.	Etat de l'art en reconnaissance faciale à partir d'un seul échantillon .....	29
2.6.1.	Méthodes traditionnelles .....	30
2.6.1.1.	Méthodes de génération d'échantillons virtuels .....	30
2.6.1.2.	Méthodes d'apprentissage générique .....	31
2.6.1.3.	Méthodes de partition d'images .....	31
2.6.2.	Méthodes basées sur l'apprentissage profond .....	32
2.6.2.1.	C'est quoi l'apprentissage profond ? .....	32
2.6.2.2.	Réseau neuronal convolutif ou Convolutional neural network (CNN) .....	33
2.6.2.3.	Architectures populaires des CNNs .....	33
2.6.2.4.	Méthodes de SSFR basées sur l'apprentissage profond .....	36
2.7.	Conclusion .....	36

**CHAPITRE 3: Caractéristiques Statistiques d'Images Binarisées en Multi-Bloc-Couleur pour la Reconnaissance Faciale à Echantillon Unique.**

3.1.	Introduction .....	38
3.2.	Méthode proposée .....	38
3.2.1.	Prétraitement .....	38
3.2.2.	Extraction de caractéristiques basée sur MB-C-BSIF .....	39
3.2.3.	Classification .....	43
3.3.	Analyse expérimentale .....	47
3.3.1.	Expérimentations sur la base de données AR .....	47
3.3.1.1.	Description de la base de données .....	47

3.3.1.2.	Configurations.....	48
3.3.1.3.	Expérimentations #1 (Effets des paramètres du descripteur BSIF) .....	48
3.3.1.4.	Expérimentation #2 (Effets de la distance).....	50
3.3.1.5.	Expérimentation #3 (Effets de la segmentation d'image).....	51
3.3.1.6.	Expérimentation #4 (Effets des informations de la texture couleur).....	53
3.3.1.7.	Comparaison #1 (Protocole I) .....	55
3.3.1.8.	Comparaison #2 (Protocole II) .....	57
3.3.2.	Expérimentations sur la base de données LFW.....	58
3.3.2.1.	Description de la base de données .....	58
3.3.2.2.	Protocole experimental .....	59
3.3.2.3.	Limite des systèmes SSFR .....	59
3.4.	Conclusion.....	61

#### **CHAPITRE 4: Combinaison de caractéristiques de texture et d'apprentissage profond pour la reconnaissance faciale à échantillon unique.**

4.1.	Introduction .....	63
4.2.	Etude comparative entre les caractéristiques basées sur la texture et celles basées sur l'apprentissage profond .....	63
4.2.1.	Méthodes .....	63
4.2.1.1.	Augmentation des données et prétraitement.....	63
4.2.1.2.	Extraction de caractéristiques .....	64
4.2.1.3.	Classification .....	68
4.2.2.	Expérimentations & Etude comparative .....	68
4.2.2.1.	Description de la base de données AR et son protocole d'évaluation.....	69
4.2.2.2.	Expérimentations & Comparaison .....	69
4.3.	Combinaison des caractéristiques de texture et d'apprentissage profond pour la reconnaissance faciale à échantillon unique.....	71
4.3.1.	Méthodes .....	72
4.3.1.1.	Augmentation des données et prétraitement.....	72
4.3.1.2.	Extraction de caractéristiques .....	72
4.3.1.3.	Classification des caractéristiques .....	73
4.3.2.	Système proposé .....	73
4.3.3.	Analyse expérimentale.....	74

4.3.3.1.	Base de données AR.....	75
4.3.3.2.	Expérimentations.....	75
4.3.3.3.	Comparaison #1.....	78
4.3.3.4.	Comparaison #2.....	78
4.4.	Conclusion.....	79

**CHAPITRE 5: La reconnaissance faciale à partir d'un seul échantillon en utilisant la transformée en ondelettes discrète et les caractéristiques de texture.**

5.1.	Introduction.....	81
5.2.	Préliminaires.....	81
5.2.1.	Transformée en ondelettes discrète à 2 dimensions (2D-DWT).....	81
5.2.2.	Binarized statistical image features (BSIF).....	84
5.3.	Approche proposée.....	84
5.3.1.	Prétraitement.....	84
5.3.2.	Extraction des caractéristiques.....	84
5.3.3.	Classification.....	85
5.4.	Analyse expérimentale.....	86
5.4.1.	Base de données AR.....	86
5.4.2.	Expérimentation #1.....	86
5.4.3.	Expérimentation #2.....	87
5.5.	Conclusion.....	88
	Conclusion Générale.....	89

# Liste des Figures

Figure 1.1. Classification générale des schémas d'authentification biométrique.....	5
Figure 1.2. Architecture fonctionnelle d'un système biométrique. Diagramme des processus d'apprentissage, vérification, et identification [8]. .....	15
Figure 1.3. Taux de vraisemblance des utilisateurs légitimes et des imposteurs. ....	17
Figure 2.1. La conception standard d'un système de reconnaissance automatique de visage. ....	23
Figure 2.2. Quelques exemples d'occultation par chapeau, lunettes, masque, main, ombre et auto-occultation.....	25
Figure 2.3. Quelques modalités d'images présentant des visages hétérogènes. (a) Photographies simples, (b) Images de croquis, et (c) Images infrarouges. ....	26
Figure 2.4. Exemple de vieillissement du visage .....	26
Figure 2.5. Exemples de reconnaissance faciale dans la vidéosurveillance.....	28
Figure 2.6. Architecture de l'AlexNet.....	34
Figure 2.7. Architecture du VGGNet.....	35
Figure 2.8. Architecture du ResNet.....	35
Figure 3.1. Illustration sur le fonctionnement du descripteur BSIF avec un filtre de taille $7 \times 7$ et 8 bits.....	40
Figure 3.2. Exemple de décomposition d'une image en multi-blocs (MB) . ....	41
Figure 3.3. Cadre de l'approche proposée pour l'extraction des caractéristiques: MB-C-BSIF.....	42
Figure 3.4. Schéma du système SSFR proposé, basé sur le descripteur MB-C-BSIF. ....	46
Figure 3.5. Les 26 images faciales du premier individu de la base de données AR, accompagnées de leurs descriptions détaillées. ....	47
Figure 3.6. Exemples de deux sujets différents de la base de données Labeled Faces in the Wild (LFW)-a. ....	59
Figure 4.1. Le cadre général du modèle hybride développé.....	74
Figure 5.1. Exemples d'application de la DWT aux premiers et deuxièmes niveaux de décomposition... ..	83
Figure 5.2. Organigramme graphique de l'approche d'extraction des caractéristiques proposée.....	85

# Liste des Tableaux

Tableau 1.1. Comparaison des différentes technologies biométriques (E: Elevée, M: Moyenne, F: Faible). .....	13
Tableau 3.1. Comparaison des résultats obtenus en utilisant six configurations de l'opérateur BSIF avec des changements dans l'expression faciale. ....	49
Tableau 3.2. Comparaison des résultats obtenus avec six configurations de l'opérateur BSIF avec occlusion par lunettes de soleil. ....	49
Tableau 3.3. Comparaison des résultats obtenus avec six configurations de l'opérateur BSIF avec occlusion par écharpes. ....	49
Tableau 3.4. Comparaison des résultats obtenus par différentes mesures de distance avec des changements en expression faciale. ....	50
Tableau 3.5. Comparaison des résultats obtenus par différentes mesures de distance avec occlusion par lunettes de soleil. ....	50
Tableau 3.6. Comparaison des résultats obtenus par différentes mesures de distance avec occlusion par écharpes. ....	51
Tableau 3.7. Comparaison des résultats obtenus à l'aide de différents blocs divisés avec des changements en expression faciale. ....	52
Tableau 3.8. Comparaison des résultats obtenus à l'aide de différents blocs divisés avec occlusion par lunettes de soleil. ....	52
Tableau 3.9. Comparaison des résultats obtenus à l'aide de différents blocs divisés avec occlusion par écharpes. ....	52
Tableau 3.10. Comparaison des résultats obtenus en utilisant différents espaces colorimétriques avec des changements dans l'expression faciale. ....	54
Tableau 3.11. Comparaison des résultats obtenus en utilisant différents espaces colorimétriques avec occlusion par lunettes de soleil. ....	54
Tableau 3.12. Comparaison des résultats obtenus en utilisant différents espaces colorimétriques avec occlusion par écharpes. ....	54
Tableau 3.13. Comparaison avec 18 méthodes de l'état de l'art en utilisant des sous-ensembles de variations en expression faciale. ....	56
Tableau 3.14. Comparaison avec 12 méthodes en utilisant des sous-ensembles d'occlusion et éclairage +occlusion. ....	57
Tableau 3.15. Taux d'identification utilisant la base de données LFW. ....	60
Tableau 4.1. Récapitulatif des configurations utilisées par chaque opérateur de texture. ....	65
Tableau 4.2. Comparaison entre plusieurs modèles CNN utilisés dans notre étude. ....	68

Tableau 4.3. Répartition des personnes et des images entre l'ensemble d'entraînement et les ensembles de test pour chaque sous-ensemble.....	69
Tableau 4.4. Résultats expérimentaux des descripteurs basés sur la texture utilisés dans notre étude..	70
Tableau 4.5. Résultats expérimentaux des modèles basés sur l'apprentissage profond. ....	70
Tableau 4.6. Comparaison des résultats obtenus à l'aide de différentes approches d'extraction/classification de caractéristiques.....	77
Tableau 4.7. Comparaison des performances de l'approche proposée par rapport à d'autres méthodes utilisant le protocole I. ....	78
Tableau 4.8. Comparaison des performances de l'approche proposée avec d'autres méthodes utilisant le protocole II. ....	79
Tableau 5.1. Répartition des personnes et des images entre l'ensemble d'apprentissage et les ensembles de test pour chaque personne. ....	86
Tableau 5.2. Résultats et comparaison de l'expérimentation #1. ....	87
Tableau 5.3. Résultats et comparaison de l'expérimentation #2. ....	88

## ***Introduction Générale***

La reconnaissance faciale est l'un des domaines de recherche les plus actifs dans la vision par ordinateur. Des méthodes, algorithmes, et bases de données ont été proposés ces dernières années pour étudier la reconnaissance faciale dans des environnements contraints et non-contraints. En effet, grâce aux technologies de l'intelligence artificielle, des progrès significatifs ont été réalisés en reconnaissance faciale. Dans les premiers temps, les intérêts de recherche étaient principalement axés sur la reconnaissance faciale dans des conditions contrôlées où des approches classiques et simples ont fourni d'excellentes performances. Ces dernières années, la recherche s'est concentrée sur des conditions non contrôlées dans lesquelles, l'approche basée sur l'apprentissage profond (en anglais, *deep learning*), en particulier les réseaux de neurones convolutifs (CNNs) ont obtenu des succès remarquables en offrant une forte robustesse face aux nombreuses variations telles que les changements d'expression, l'éclairage, et l'occlusion qui peuvent altérer le processus de reconnaissance.

Pour garantir une généralisation précise des systèmes d'apprentissage en reconnaissance faciale, il est crucial de disposer d'ensembles de données d'entraînement considérables. Bien que l'approche basée sur l'apprentissage profond à l'aide de CNN soit considérée comme une technique puissante pour la reconnaissance faciale, il est nécessaire d'avoir une quantité importante de données d'entraînement pour obtenir des performances meilleures. Ainsi, l'importance d'un ensemble de données d'entraînement de haute qualité et important est cruciale pour obtenir des performances précises et généralisables dans la reconnaissance faciale.

Dans les situations de la vie réelle tels que la surveillance de la sécurité dans les aéroports, les contrôles frontaliers, ou encore les vérifications d'identité en ligne, la reconnaissance faciale à échantillon unique par personne (en anglais, *single sample per person* (SSPP)) est un défi majeur en reconnaissance faciale. Contrairement à la reconnaissance faciale traditionnelle, où plusieurs échantillons faciaux sont disponibles pour chaque personne, dans la reconnaissance faciale à échantillon unique (en anglais, *single sample face recognition* (SSFR)), il n'y a qu'un seul échantillon facial par personne dans l'ensemble de données d'apprentissage. Ainsi les images de visages peuvent varier considérablement selon les conditions d'acquisition, comme l'éclairage, l'angle de vue, l'occlusion, l'expression faciale, etc. Cela peut rendre difficile la tâche de reconnaître une personne à partir d'une seule image. Pour résoudre ce problème,

différentes approches ont été proposées dans la littérature et qui peuvent être grossièrement divisées en deux catégories basées sur l'extraction de caractéristiques: les méthodes traditionnelles et méthodes basées sur l'apprentissage profond. Tout d'abord, les méthodes traditionnelles de reconnaissance basées sur des caractéristiques artisanales extraient des caractéristiques à l'aide de descripteurs de texture tels que *binarized statistical image features* (BSIF), *local binary pattern* (LBP) et *local phase quantization* (LPQ). Ensuite les méthodes d'apprentissage profond en particulier les CNNs ont connu un développement rapide et sont devenues une stratégie d'apprentissage axée sur les données pour diverses tâches de vision par ordinateur. Elles intègrent l'extraction de caractéristiques et la classification en effet il utilise une technique de filtrage d'image pour extraire des caractéristiques et un réseau neuronal pour catégoriser les caractéristiques extraites dans les classes appropriées. Cette thèse s'appuie sur trois contributions majeures :

- Dans notre première contribution, nous avons proposé une méthode efficace appelée *multi-block color binarized statistical image features* (MB-C-BSIF), basée sur le descripteur de texture BSIF, afin de résoudre le problème de SSFR. Cette méthode exploite les informations de texture locales, régionales, globales, et de couleur afin d'obtenir une représentation précise et fiable des visages.
- Dans notre deuxième contribution, nous avons proposé une méthode hybride combinant deux approches, l'une traditionnelle et l'autre basée sur l'apprentissage profond. En effet, nous avons concaténé le descripteur de texture couleur MB-C-BSIF et le CNN afin de bénéficier des avantages de ces deux méthodes.
- Dans notre troisième contribution, nous avons proposé une méthode d'analyse en multi-résolution qui utilise la transformée en ondelette discrète (en anglais, *discrete wavelet transform* (DWT)) pour extraire des caractéristiques statistiques binarisées (BSIF) à différentes résolutions. Cette approche a pour objectif de capturer davantage d'informations et d'améliorer la capacité du système à reconnaître les visages dans des conditions variables.

Notre thèse est organisée en cinq chapitres :

- Dans le premier chapitre, nous présentons une vue d'ensemble des biométries et décrivons en détail la méthode de reconnaissance basée sur les caractéristiques



biométriques, tout en examinant les différentes méthodes utilisées pour évaluer les performances des systèmes biométriques.

- Dans le deuxième chapitre nous explorons l'état de l'art de la reconnaissance faciale. Nous examinons les recherches connexes dans ce domaine et les catégorisons en deux types basés sur l'extraction de caractéristiques: traditionnelles ou basés sur l'apprentissage profond.
- Dans le troisième chapitre, nous détaillons le descripteur MB-C-BSIF, qui est notre première contribution. Nous évaluons ses performances et comparons ses résultats avec d'autres méthodes décrites dans la littérature.
- Dans le quatrième chapitre, nous présentons une méthode hybride basée sur la concaténation du descripteur de texture couleur MB-C-BSIF et les CNNs, et évaluons leurs performances en les comparant avec d'autres méthodes de l'état de l'art.
- Dans le cinquième chapitre, nous détaillons notre méthode d'analyse multi-résolution. Nous évaluons ses performances et comparons ses résultats avec d'autres méthodes décrites dans la littérature.

En conclusion, nous présentons une synthèse de nos résultats et discutons des possibilités d'étendre cette recherche à l'avenir.

Les contributions liées à cette thèse sont :

- Adjabi, I., Ouahabi, A., Benzaoui, A., & Taleb-Ahmed, A. (2020). Past, present, and future of face recognition: A review. *Electronics*, 9(8), 1188.
- Adjabi, I., Ouahabi, A., Benzaoui, A., & Jacques, S. (2021). Multi-block color-binarized statistical images for single-sample face recognition. *Sensors*, 21(3), 728.
- Adjabi, I. (2022, May). Combining hand-crafted and deep-learning features for single sample face recognition. In *2022 7th International Conference on Image and Signal Processing and their Applications (ISPA)* (pp. 1-6). IEEE.
- Adjabi, I. (2023, July). Single sample face recognition using multi-resolution analysis of color texture features. In *2023 International Conference on Nonlinear Science and Complexity (ICNSC2023)*. Istanbul, Turkey.

## **CHAPITRE 1: Principes de la Biométrie et Systèmes Biométriques**

### **1.1. Introduction**

L'intérêt pour les technologies de reconnaissance biométrique a connu une croissance significative au cours des dernières décennies. Les avancées technologiques dans ce domaine ont conduit à des recherches approfondies et à des développements visant à créer des systèmes biométriques précis, conviviaux, sans contrainte, et abordables. Par conséquent, les systèmes d'authentification biométrique sont largement utilisés dans divers domaines. Ils trouvent leur utilisation dans le contrôle d'accès physique, la surveillance, la sécurité des auteurs correspondants du réseau, les transactions en ligne, le contrôle de présence, ainsi que dans les technologies d'authentification pour les appareils mobiles et les ordinateurs personnels.

Ce chapitre fournit un aperçu global des concepts et des principes de base de la biométrie. Nous expliquons le fonctionnement des systèmes biométriques, et présentons les modalités biométriques les plus répandues. Ces fondements permettent de poser les bases nécessaires pour comprendre et approfondir les sujets abordés dans les chapitres suivants de cette thèse.

### **1.2. Définition de la biométrie**

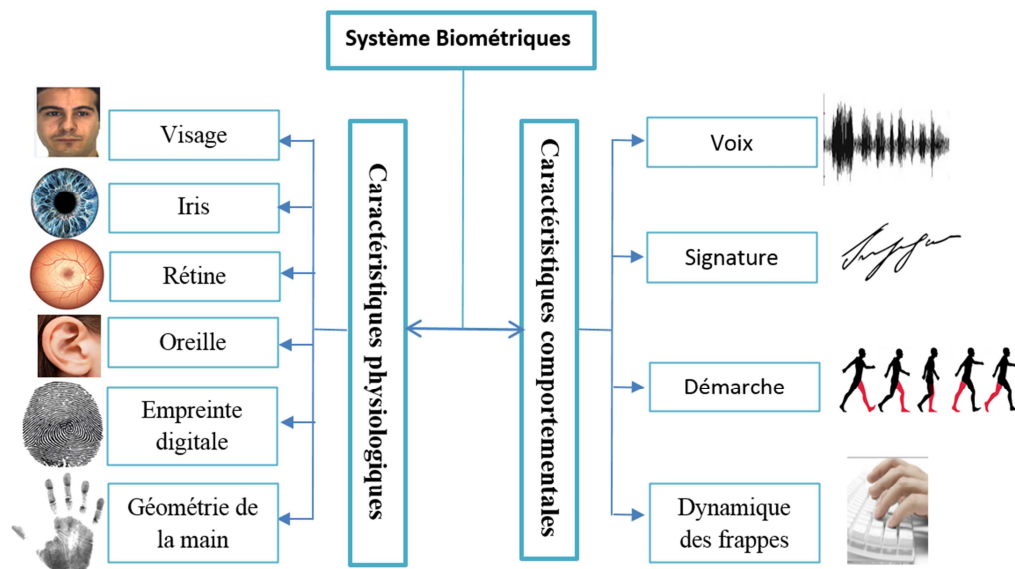
Dans la littérature, il existe de nombreuses définitions pour la biométrie. Dans cette section, nous présenterons certaines des définitions qui existent dans la littérature.

- La biométrie est définie comme la reconnaissance automatisée des individus basée sur leurs caractéristiques comportementales et biologiques [1].
- La biométrie est la science qui mesure les propriétés physiques des êtres vivants [2].
- Le terme biométrie trouve ses origines dans les mots grecs "bio" qui signifie "vie" et "métrique" qui signifie "mesurer". La biométrie fait référence à l'identification ou à la vérification d'une personne en se basant sur ses caractéristiques physiologiques et/ou comportementales [3].
- "Toute caractéristique physique ou trait personnel automatiquement mesurable, robuste, et distinctif qui peut être utilisé pour identifier un individu ou vérifier l'identité revendiquée d'un individu" [4].

Les définitions présentées soulignent que la biométrie repose sur des caractéristiques physiques ou comportementales uniques à chaque individu. Ces caractéristiques sont mesurables et servent à identifier ou vérifier l'identité d'une personne. Ces caractéristiques physiologiques ou comportementales qui permettent d'identifier une personne sont appelées modalités biométriques.

### 1.3. Différentes modalités biométriques

Fondamentalement, les modalités biométriques personnels sont regroupées en deux principales catégories: les caractéristiques physiologiques et les caractéristiques comportementales.



**Figure 1.1.** Classification générale des schémas d'authentification biométrique.

Les caractéristiques physiologiques sont liées aux caractéristiques statiques du corps humain qui ne changent pas avec le temps. Parmi les exemples de traits physiologiques, on retrouve la reconnaissance faciale, la géométrie de la main, l'empreinte palmaire, l'empreinte digitale, la reconnaissance de l'iris, l'ADN, la rétine, et la reconnaissance des veines sanguines. D'un autre côté, l'approche comportementale de la biométrie se concentre sur les traits comportementaux d'un individu, qui sont liés à son comportement personnel. Cela inclut la reconnaissance vocale, la reconnaissance de signature, la démarche, et la dynamique de frappe [5-8]. La Figure 1.1 présente une classification générale des schémas d'authentification biométrique.

### **1.3.1. Reconnaissance faciale**

Les traits du visage tels que les yeux, le nez, les lèvres et le menton jouent un rôle important dans la reconnaissance des individus sur le long terme. Chaque personne a une forme de visage unique, ce qui en fait une méthode populaire dans le domaine de la reconnaissance biométrique. Les ordinateurs ont contribué au développement de la reconnaissance automatique des individus en utilisant les caractéristiques faciales distinctives. Cela a conduit à la popularité croissante des systèmes de reconnaissance faciale. Il existe de nombreux logiciels commerciaux qui ont été programmés pour effectuer une identification précise des traits du visage humain.

#### **- *Avantages***

Les systèmes de reconnaissance faciale offrent plusieurs avantages, tels que la facilité d'utilisation; ils peuvent capturer et analyser les traits du visage sans nécessiter de contact physique ou d'actions spéciales de la part des individus à identifier, la rapidité et l'efficacité; les systèmes de reconnaissance faciale peuvent traiter et comparer les visages rapidement, ce qui permet une identification rapide et efficace, et des coûts de mise en œuvre réduits; ils ne nécessitent pas d'équipements spécialisés coûteux et peuvent utiliser des caméras standard pour capturer les images du visage.

#### **- *Inconvénients***

Plusieurs facteurs peuvent influencer les performances d'un système de reconnaissance faciale biométrique, telle que la qualité ou la résolution des photos collectées pour chaque individu, les variations en éclairage, les angles de rotation du visage, etc. Aussi, différents changements en expressions faciales imposent certains défis à la reconnaissance automatique des identités des personnes, telles que triste expression, heureux, en colère, et autre.

### **1.3.2. Reconnaissance de l'iris**

L'iris est la partie circulaire colorée située au centre de l'œil, et son motif distinctif offre une grande efficacité dans les systèmes de reconnaissance des individus. L'iris possède une composition biologique unique avec un réseau spécifique de tissus qui peuvent être visiblement reconnus.

- *Avantages*

La reconnaissance de l'iris est largement considérée comme la méthode optimale pour l'authentification automatique des individus en raison de plusieurs facteurs. Tout d'abord, les scanners d'iris sont largement utilisés dans divers secteurs de la sécurité. De plus, ces dispositifs offrent une facilité d'utilisation et une flexibilité d'exploitation. L'iris possède des caractéristiques distinctives uniques qui sont difficiles à falsifier, ce qui renforce la sécurité et la précision de la reconnaissance de l'iris.

- *Inconvénients*

Un certain nombre de chercheurs ont abordé différentes questions et facteurs qui contribuent à la diminution de la précision de la reconnaissance de l'iris. Certains de ces facteurs incluent le port de lunettes, de lentilles de contact, etc. Un autre défi de déploiement de la biométrie basée sur l'iris est le coût élevé de mise en œuvre d'une telle technique d'authentification.

### **1.3.3. Reconnaissance de la rétine**

Chaque œil possède une rétine, qui est une couche composée d'un réseau complexe de cellules neurales situé à l'arrière de l'œil. Cette partie de l'œil présente des caractéristiques uniques qui ont été exploitées dans le développement d'un nouveau système d'authentification automatique permettant une distinction rapide entre les individus.

- *Avantages*

La biométrie de la rétine offre des avantages significatifs. Tout d'abord, elle présente une haute unicité et distinction, ce qui signifie qu'elle est capable de différencier même des jumeaux. De plus, les motifs de la rétine restent stables dans le temps, offrant ainsi une identification fiable sur le long terme. De plus, la difficulté de contrefaçon et la protection contre l'environnement font de la reconnaissance de la rétine une technologie sécurisée.

- *Inconvénients*

Cette technique d'authentification biométrique a suscité beaucoup d'attention en raison de son niveau élevé de précision dans la reconnaissance des individus. Cependant, elle présente certains inconvénients. Tout d'abord, elle peut causer un certain inconfort à l'utilisateur en raison des efforts nécessaires pour capturer les vaisseaux de sa rétine. De plus, la biométrie basée sur le

scan de la rétine peut être affectée par certains facteurs médicaux, tels que l'hypertension artérielle, ce qui peut altérer l'exactitude du processus d'authentification.

### **1.3.4. Reconnaissance de l'oreille**

Contrairement à d'autres parties du corps, la forme de l'oreille reste constante tout au long de la vie d'une personne, ce qui permet d'utiliser cette caractéristique comme moyen d'authentification biométrique.

#### *- Avantages*

Il s'agit d'une méthode plus confortable et conviviale en termes de participation de l'utilisateur par rapport à la reconnaissance de l'iris et de la rétine.

#### *- Inconvénients*

Comme d'autres schémas de reconnaissance biométrique, l'authentification basée sur l'oreille présente certains inconvénients. Cette méthode n'a pas encore atteint un niveau de sécurité remarquable. L'un des inconvénients de la reconnaissance de l'oreille est la simplicité des caractéristiques distinctives de l'oreille, qui ne permettent pas d'établir de manière solide l'identité d'un individu [9].

### **1.3.5. Reconnaissance d'empreintes digitales**

Les systèmes d'authentification biométrique basés sur les empreintes digitales sont largement utilisés et considérés comme l'une des méthodes les plus efficaces pour l'identification et la vérification de l'identité. Chaque individu possède des empreintes digitales uniques, même entre des jumeaux identiques. Et encore chaque doigt d'une même personne a également une empreinte différente. Cette unicité des empreintes digitales constitue la base fondamentale des systèmes d'authentification biométrique basés sur les empreintes digitales.

#### *- Avantages*

Les systèmes d'authentification basés sur les empreintes digitales offrent de nombreux avantages, ce qui explique leur déploiement croissant dans différentes institutions. Ils sont faciles à utiliser pour les utilisateurs, nécessitant simplement de placer leur doigt sur le capteur. De plus, ces systèmes utilisent des capteurs optiques abordables, rendant leur mise en œuvre accessible sans nécessiter un investissement financier élevé. Ils ne requièrent pas une consommation

d'énergie importante, ce qui les rend adaptés aux appareils mobiles et contribue à une meilleure gestion de l'énergie. Enfin, cette méthode d'authentification est particulièrement adaptée aux environnements mobiles, déjà intégrée dans de nombreux smartphones, tels que les iPhones, offrant ainsi praticité et commodité aux utilisateurs.

- ***Inconvénients***

Malgré les nombreux avantages des systèmes de reconnaissance d'empreintes digitales, il existe quelques inconvénients associés à cette technologie. L'un des principaux problèmes réside dans la complexité d'obtenir des images de haute qualité des empreintes digitales. Des facteurs tels que la saleté, les coupures, les déchirures, et l'usure peuvent altérer les crêtes et les détails des empreintes digitales.

### **1.3.6. Reconnaissance de la géométrie de la main**

De nombreux systèmes de sécurité et applications dépendent de la reconnaissance de la géométrie de la main en tant que technique automatisée pour l'identification ou la vérification de leurs utilisateurs légitimes. Cette méthode d'authentification biométrique a commencé à gagner en popularité dans divers secteurs de la sécurité.

- ***Avantages***

Cette méthode présente plusieurs avantages par rapport aux autres techniques d'authentification biométrique. Par exemple, elle est facile à utiliser et bénéficie d'une acceptation généralisée de la part des utilisateurs, ce qui en fait un système d'authentification convivial par rapport aux autres systèmes biométriques.

- ***Inconvénients***

Cette méthode nécessite un dispositif matériel spécial et coûteux pour scanner la géométrie de la main. Ce scanner doit être en trois dimensions afin d'obtenir toutes les informations de la paume. De plus, certaines contraintes peuvent affecter l'extraction des informations, telles que le port de bijoux, etc.

### **1.3.7. Reconnaissance de la signature**

Les signatures biométriques offrent une méthode d'identification unique basée sur le comportement d'écriture de chaque individu. Elle permet la capture dynamique de données, telles que la direction, la vitesse, la pression d'écriture, et la forme de la signature.

- *Avantages*

La biométrie de la signature présente des avantages, tels que leur non-intrusivité, leur acceptabilité élevée, leur coût abordable et leur efficacité dans les transactions commerciales. Elle offre une méthode familière d'authentification et est difficile à imiter en raison des modèles comportementaux uniques de chaque individu. Cependant, elle a aussi des inconvénients, tels que leur faible précision.

- *Inconvénients*

Les signatures présentent plusieurs inconvénients. Elles ont une faible précision en raison des variations possibles dans la reproduction de la signature par l'individu. Les signatures peuvent changer avec le temps. Les signatures sont également faciles à falsifier, et différentes personnes peuvent signer de la même manière.

### **1.3.8. Reconnaissance de la voix**

La reconnaissance vocale est une technologie qui permet de convertir les sons, phrases, et mots prononcés par les êtres humains en signaux électriques, puis ces signaux sont convertis en un code de conception. Ce code de conception représente les caractéristiques vocales uniques de chaque individu et est utilisé pour identifier et authentifier une personne.

- *Avantages*

La reconnaissance vocale permet aux utilisateurs de communiquer naturellement en utilisant leurs voix, sans nécessiter de formation préalable. La saisie vocale est rapide, ce qui permet d'augmenter la productivité. Elle facilite l'accès à la technologie pour les personnes ayant des difficultés à utiliser leurs mains. De plus, elle réduit les erreurs de frappe et offre une expérience utilisateur améliorée.



- ***Inconvénients***

Les inconvénients de la reconnaissance vocale incluent les erreurs fréquentes en présence de perturbations, la nécessité d'un microphone proche de l'utilisateur, la vulnérabilité aux enregistrements vocaux préenregistrés, les difficultés avec différentes langues et les coûts élevés [10].

### **1.3.9. Reconnaissance de la dynamique des frappes**

La reconnaissance de la dynamique des frappes est une technologie comportementale basée sur l'analyse de la façon dont une personne tape sur un clavier ou un appareil de saisie. Elle utilise des mesures, telles que le temps de frappe, les intervalles entre les touches, la vitesse, et la pression exercée pour créer un modèle comportemental unique.

- ***Avantages***

La reconnaissance de la dynamique des frappes est une technologie pratique, peu coûteuse et non intrusive. Elle utilise un logiciel simple pour analyser les caractéristiques de la frappe, ce qui réduit la nécessité de changer fréquemment les mots de passe et simplifie la gestion des identifiants.

- ***Inconvénients***

Il est important de noter que la reconnaissance de la dynamique des frappes présente quelques limites et défis. Par exemple, certains facteurs tels que la fatigue, le stress, ou les changements de matériel peuvent affecter la dynamique des frappes d'un individu, ce qui peut entraîner des erreurs de reconnaissance. De plus, il est également possible qu'un utilisateur modifie délibérément leur comportement de frappe pour contourner la technologie.

### **1.3.10. Reconnaissance de la démarche**

La démarche désigne la manière unique dont une personne marche, englobant différents aspects, tels que la longueur des pas, le rythme, la posture, et le tempo. La reconnaissance biométrique de la démarche est une technologie qui vise à vérifier ou à identifier les individus en se basant sur la façon de marcher.

- **Avantages**

La démarche est une caractéristique biométrique attrayante en raison de son caractère non intrusif. Contrairement à d'autres caractéristiques biométriques, elle peut être capturée à distance et sans consentement préalable de la personne observée. De plus, il est difficile de dissimuler, de voler, ou de falsifier la démarche. L'étude des paramètres kinésiologiques qui définissent la démarche humaine peut donc être utilisée comme base pour l'identification.

- **Inconvénients**

La démarche change avec le temps et est influencée par des facteurs tels que les vêtements, les chaussures, la surface de marche, la vitesse, et l'état émotionnel.

#### 1.4. Caractéristiques biométriques

Dans le contexte de l'évaluation des performances d'une modalité biométrique, diverses caractéristiques sont prises en considération. Chaque modalité possède ses propres avantages et inconvénients. La sélection d'une modalité biométrique pour une application spécifique repose sur une série de paramètres liés à la nature de l'application, à ses exigences ainsi qu'aux caractéristiques de la modalité en question. L'étude de Jain et al. [11] a en effet identifié sept paramètres communément employés pour évaluer la pertinence d'une modalité physique ou comportementale dans une application biométrique. Ces paramètres sont les suivants :

- **Universalité**: La modalité biométrique doit être présente chez tous les individus, ce qui signifie que chaque personne doit posséder cette caractéristique.
- **Unicité**: La modalité biométrique doit permettre de distinguer de manière unique chaque individu, c'est-à-dire qu'aucun individu ne devrait partager la même caractéristique biométrique.
- **Permanence**: La modalité biométrique doit être stable et ne pas changer significativement au fil du temps. Cela garantit que la caractéristique reste utilisable sur une longue période.
- **Mesurabilité**: La modalité biométrique doit pouvoir être mesurée de manière objective et précise. Des techniques appropriées doivent être disponibles pour extraire et quantifier la caractéristique biométrique.

- **Acceptabilité**: La modalité biométrique doit être acceptable pour les individus, c'est-à-dire qu'elle doit être facile à acquérir et ne pas susciter de réticence ou d'inconfort chez les utilisateurs.
- **Performance**: La modalité biométrique doit offrir des performances suffisantes en termes de taux de reconnaissance, de taux d'erreur, et de temps de traitement pour répondre aux besoins de l'application.
- **Circumvention**: La modalité biométrique ne doit pas être facilement contournable ou falsifiable. Elle doit être suffisamment robuste pour résister à des tentatives de fraude ou de contrefaçon.

La prise en compte de ces paramètres permet de sélectionner la modalité biométrique la plus appropriée pour une application donnée, en tenant compte des spécificités et des exigences de celle-ci. Le Tableau 1.1 présente une évaluation empirique de différentes technologies biométrique.

**Tableau 1.1.** Comparaison des différentes technologies biométriques (E: Elevée, M: Moyenne, F: Faible).

Modalité	Universalité	Unicité	Permanence	Mesurabilité	Performance	Acceptabilité	Circumvention
Visage	E	F	M	E	F	E	F
Iris	E	E	E	M	E	F	E
Rétine	E	E	M	F	E	F	E
Oreille	M	M	E	M	M	E	M
Empreintes digitales	M	E	E	M	E	M	E
Géométrie de la main	M	M	M	E	M	M	M
Dynamique des frappes	F	F	F	M	F	M	M
Signature	F	F	F	E	F	E	F
Voix	M	F	F	M	F	E	F
Démarche	M	F	F	E	F	E	M

### 1.5. Architecture fonctionnelle d'un système biométrique

Un système biométrique est, généralement, composé de deux principaux modes qui sont: un mode d'apprentissage et un mode de reconnaissance (vérification ou identification). Le mode d'apprentissage ou d'enregistrement est la première étape du système biométrique. Pendant cette phase, les caractéristiques biométriques d'un individu sont collectées, extraites, et converties en

un modèle biométrique unique, également appelé "template" ou "signature biométrique". Ce modèle est ensuite stocké dans une base de données sécurisée pour une utilisation ultérieure.

Le mode de reconnaissance est la deuxième étape du système biométrique. Lorsque l'individu souhaite être identifié ou vérifié, ses caractéristiques biométriques sont capturées et comparées aux modèles biométriques stockés dans la base de données. Si une correspondance est trouvée, l'individu est considéré comme reconnu et vérifié. Cela signifie que ses caractéristiques biométriques correspondent à celles d'un individu enrôlé précédemment. En revanche, si aucune correspondance n'est trouvée, l'individu est considéré comme non reconnu.

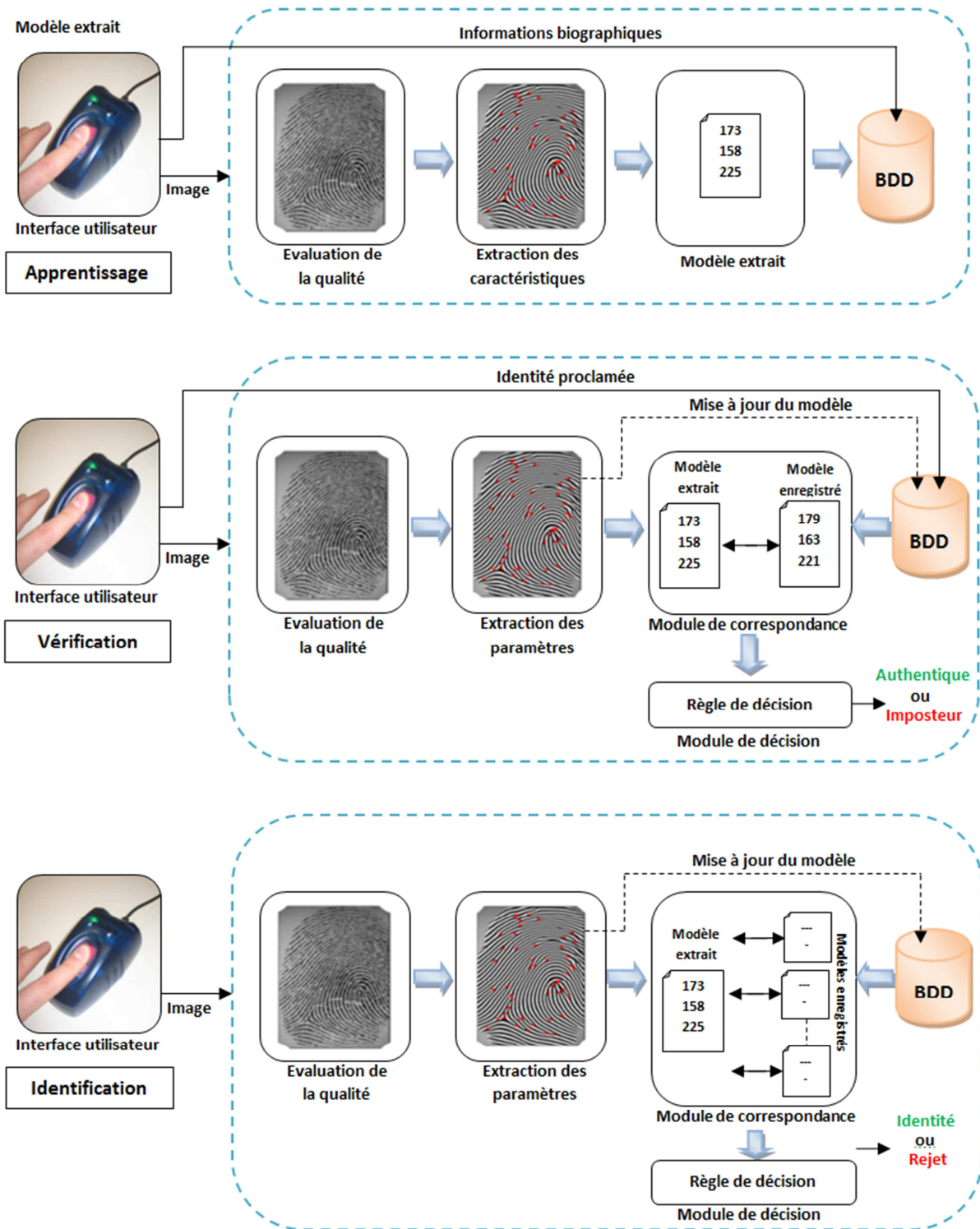
### **1.5.1. Vérification vs Identification**

La reconnaissance biométrique peut prendre deux formes principales: la vérification et l'identification. En mode de vérification [12], le système évalue l'identité d'une personne en comparant son ou ses modèles enregistrés dans la base de données avec les caractéristiques biométriques capturées. Une comparaison un-à-un est effectuée par le système pour décider si l'identité déclarée est vraie ou fausse. En mode d'identification, l'objectif est d'identifier l'individu parmi un grand nombre d'autres individus en comparant ses caractéristiques biométriques avec tous les modèles biométriques stockés dans la base de données.

### **1.5.2. Modules d'un système biométrique**

La plupart des systèmes de technologie biométrique utilisent les mêmes principes de fonctionnement [13]. Le système biométrique peut être représenté par cinq modules qui sont:

- Le module de capture / acquisition d'image permet de capturer les données biométriques brutes de l'individu sous forme de vidéo, audio, image, ou autre signal.
- Le module d'extraction de caractéristiques effectue un processus automatisé pour extraire les caractéristiques biométriques distinctives afin de générer un modèle ou template. Ce processus peut utiliser des techniques d'apprentissage automatique, de vision par ordinateur, et de reconnaissance de motifs.
- Le module de base de données est un répertoire qui contient les informations biométriques enregistrées et enrôlées des utilisateurs. Il stocke différents modèles ou templates d'utilisateurs.



**Figure 1.2.** Architecture classique d'un système biométrique. Diagramme des processus d'apprentissage, vérification, et identification [8].

- Le module de correspondance effectue une comparaison entre les caractéristiques biométriques actuellement extraites et les modèles stockés. Cela permet de générer une valeur de correspondance ou un score de correspondance, qui quantifie la similarité entre les deux échantillons biométriques.
- Le module de prise de décision utilise les scores de correspondance obtenus et les compare à une valeur seuil préétablie. En fonction de cette comparaison, il prend une décision d'acceptation ou de rejet. La Figure 1.2 présente les processus d'apprentissage, vérification, et identification d'un système biométrique.

## 1.6. Performance et évaluation des systèmes biométriques

Un aspect important de la technologie biométrique est d'évaluer sa performance. Lors de l'évaluation de la performance d'un système biométrique, il est essentiel de prendre en compte les différents modes de fonctionnement, à savoir l'identification et la vérification.

### 1.6.1. Evaluation de la vérification

Lorsqu'il s'agit de vérifier un système, notamment un système de reconnaissance ou d'authentification biométrique, il est courant d'utiliser des mesures, telles que le taux de fausses acceptations (en anglais, *false accept rate* (FAR)) et le taux de faux rejets (*false rejection rate* (FRR)).

- **FAR**: mesure la probabilité qu'un individu non autorisé soit accepté par le système. Cela signifie que le système reconnaît incorrectement cette personne comme étant autorisée. Un taux de FAR bas est préférable car il indique une meilleure sécurité.
- **FRR**: mesure la probabilité qu'un individu authentique soit rejeté par le système. Cela signifie que le système ne reconnaît pas correctement cette personne, le considérant à tort comme non autorisée. Un taux de FRR bas est souhaitable car il indique une meilleure convivialité du système, permettant aux utilisateurs légitimes d'être correctement reconnus.
- **Le taux d'acceptation réel (en anglais, *true accept rate* (TAR))**: mesure la proportion d'individus authentiques correctement acceptés par le système. C'est une mesure importante pour évaluer la performance du système en termes de reconnaissance des utilisateurs autorisés.

- **La courbe receiver operator characteristic (ROC)**: est utilisée pour représenter graphiquement les performances du système en fonction des différents seuils de décision. La courbe ROC affiche le FRR sur l'axe des ordonnées et le FAR sur l'axe des abscisses. Elle permet d'observer le compromis entre le FRR et le FAR et de choisir le seuil optimal en fonction des besoins spécifiques du système.
- **Le taux d'erreur égal (equal error rate (EER))**: c'est un point important de la courbe ROC (point d'équilibre). L'EER correspond au point où le taux de faux rejets est égal au taux de fausses acceptations, représentant ainsi le seuil où les deux types d'erreurs sont équilibrés. L'EER est souvent utilisé pour évaluer la performance d'un système biométrique, car il fournit une mesure globale de sa précision. La Figure 1.3 représente la distribution théorique des taux de vraisemblance des utilisateurs légitimes et des imposteurs [14].

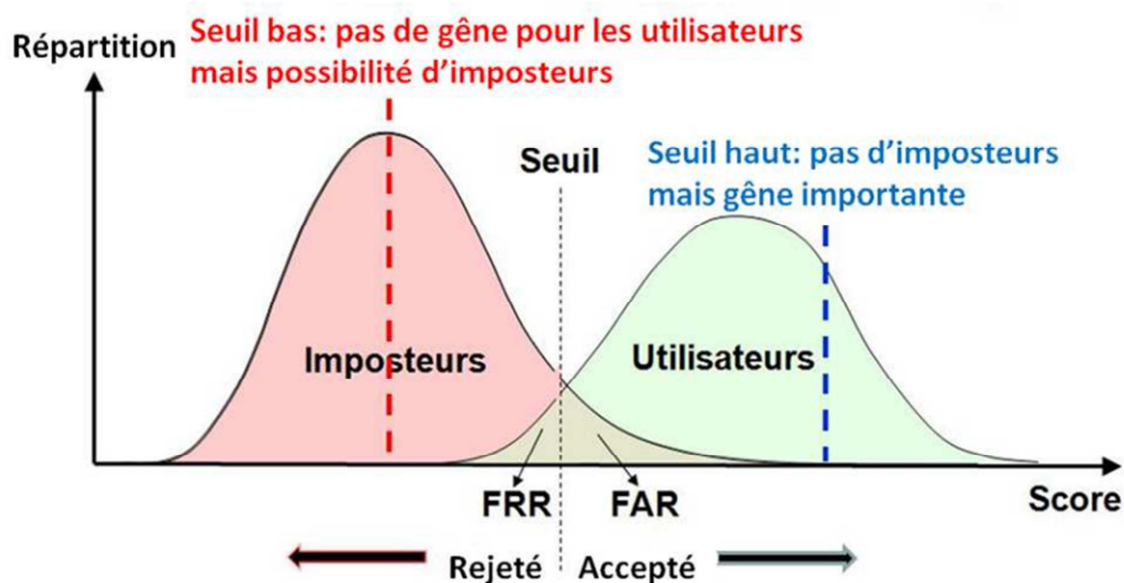


Figure 1.3. Taux de vraisemblance des utilisateurs légitimes et des imposteurs.

### 1.6.2. Evaluation de l'identification

Le taux de reconnaissance (en anglais, recognition rate (RR)) est une mesure couramment utilisée pour évaluer la performance d'un système d'identification. Il mesure la proportion de participants préalablement enregistrés dont les identités ont été correctement reconnues par le

système. Le RR est calculé en divisant le nombre d'images correctement reconnues par le nombre total d'images. Mathématiquement, l'équation pour calculer le RR est la suivante [15]:

$$RR = \frac{\text{Nombre d'images correctement reconnues}}{\text{Nombre total d'images}} \times 100 \quad (1.1)$$

### 1.7. Applications de la biométrie

Les applications de la biométrie sont diverses et couvrent plusieurs domaines. Dans ce qui suit, nous présentons quelques-unes des applications courantes de la biométrie:

- **Authentification bancaire:** Les systèmes de biométrie peuvent être utilisés pour renforcer la sécurité des transactions bancaires en s'assurant que seules les personnes autorisées peuvent accéder à leurs comptes ou effectuer des transactions.
- **Sécurité aéroportuaire:** Les aéroports utilisent la biométrie pour renforcer la sécurité en identifiant les passagers et le personnel autorisé, en vérifiant leur identité et en facilitant le contrôle des frontières.
- **Identification nationale:** Certains pays utilisent des systèmes de biométrie pour créer des cartes d'identité nationales sécurisées, basées sur des caractéristiques biométriques telles que les empreintes digitales ou les empreintes faciales.
- **Forces de l'ordre:** La biométrie est utilisée par les forces de l'ordre pour l'identification des criminels, l'enregistrement des délinquants sexuels, et la gestion des bases de données criminelles.
- **Contrôle d'accès et authentification unique:** Les systèmes de biométrie sont utilisés pour contrôler l'accès aux bâtiments, aux zones restreintes, et aux systèmes informatiques, offrant ainsi une méthode sécurisée d'authentification des utilisateurs.
- **Temps et présence:** La biométrie est utilisée pour enregistrer les heures de travail des employés de manière précise et fiable, remplaçant les anciens systèmes de pointage.

La biométrie offre des avantages, tels que la sécurité renforcée, la commodité pour les utilisateurs, et la responsabilité accrue dans les transactions. Cependant, il est important de prendre en compte les questions de confidentialité et d'éthique liées à la collecte et à l'utilisation des données biométriques [16].



## **1.8. Motivation : Pourquoi le visage?**

Le visage humain a été choisi comme modalité biométrique principale dans cette thèse pour plusieurs raisons. Tout d'abord, le visage est considéré comme la modalité la plus naturelle pour identifier les individus, car il est largement utilisé dans la reconnaissance des personnes par le système neurologique humain. Il est non intrusif, ce qui signifie qu'il peut être capturé à distance sans nécessiter de contact physique avec un capteur, ce qui le rend acceptable pour le public. De plus, l'acquisition des images faciales peut être réalisée de manière secrète et discrète à l'aide d'une simple caméra reliée à un ordinateur.

Le visage est également une caractéristique biométrique largement utilisée dans de nombreux scénarios de sécurité et de surveillance, tels que l'identification des criminels dans les aéroports ou la vidéosurveillance dans les lieux publics. Cependant, malgré ses avantages, l'identification par visage présente également des défis. Le visage humain est une entité dynamique qui peut varier en termes de pose, d'expressions faciales, d'éclairage, d'occlusion, et à d'autres facteurs, ce qui rend l'identification difficile dans certaines situations.

De plus, les systèmes biométriques basés sur le visage sont vulnérables aux attaques de contrefaçon (spoofing), où une personne tente de falsifier les données faciales pour tromper le système. Ces défis ont motivé la recherche dans le domaine de l'évaluation des descripteurs de textures locaux et de l'apprentissage profond pour améliorer la précision et la robustesse de l'identification par visage.

## **1.9. Conclusion**

Dans ce chapitre, nous avons abordé le sujet de la biométrie de manière générale. Nous avons commencé par présenter des définitions et des caractéristiques clés de la biométrie. Ensuite, nous avons examiné différentes modalités biométriques, en mettant en évidence leurs avantages et leurs inconvénients respectifs. Nous avons également exploré la structure des systèmes biométriques et les différents modes de fonctionnement qu'ils utilisent.

Par la suite, nous nous sommes penchés sur l'évaluation des systèmes biométriques, en examinant les aspects spécifiques de l'identification et de la vérification biométrique. Nous avons également illustré quelques exemples concrets d'applications biométriques dans divers domaines.

Enfin, nous avons mis l'accent sur notre choix particulier de la modalité biométrique du visage. Nous avons expliqué les raisons qui motivent ce choix, notamment la naturalité de la reconnaissance faciale, sa non-intrusivité, et son acceptation par le public. Nous avons également souligné les défis auxquels est confrontée l'identification par visage, tels que les variations dues à la pose, aux expressions faciales, et à l'éclairage,

Le prochain chapitre sera dédié à la reconnaissance faciale, explorant les différentes techniques utilisées, les défis auxquels elles font face.

## **CHAPITRE 2: La Reconnaissance Faciale.**

### **2.1. Introduction**

La biométrie du visage est l'une des domaines de recherche les plus actifs en vision par ordinateur et en reconnaissance de formes, avec de nombreuses applications pratiques et commerciales telles que le contrôle d'accès, la médecine légale, et les interactions homme-machine. Au cours des dernières années, plusieurs techniques, algorithmes, approches, et bases de données ont été proposés pour étudier la reconnaissance faciale contrainte et non contrainte. Les approches proposées ont atteint un certain degré de maturité et ont rapporté des taux de reconnaissance très élevés. Cette performance est obtenue dans des environnements contrôlés où les paramètres d'acquisition sont maîtrisés, tels que l'éclairage, l'angle de vue, et la distance entre la caméra et le sujet. Cependant, si les conditions ambiantes (par exemple, l'éclairage) ou l'apparence faciale (par exemple, la pose ou l'expression faciale) changent, cette performance se dégrade considérablement.

Ce chapitre aborde le visage en tant que technologie biométrique, présente les principales phases d'un système de reconnaissance de visage, examine les problèmes ouverts, et met en évidence les raisons pour lesquelles nous avons considéré dans cette thèse l'étude de la reconnaissance faciale à partir d'un seul échantillon comme un défi ouvert. Nous concluons ce chapitre en présentant l'état de l'art de la reconnaissance faciale à partir d'un seul échantillon.

### **2.2. Le visage en tant que modalité biométrique**

La reconnaissance de visage a suscité une attention considérable au cours des trois dernières décennies, car elle est considérée comme une application simplifiée d'analyse d'images et de reconnaissance de formes. Il existe au moins deux raisons de comprendre cette tendance: (1) la grande variété de demandes commerciales et juridiques, en plus (2) de la disponibilité des technologies pertinentes (par exemple, les smartphones, les appareils photo numériques, les GPU...). Bien que les systèmes d'apprentissage automatique et de reconnaissance existants aient atteint un certain degré de maturité, leurs performances sont limitées aux conditions imposées dans les applications du monde réel [17]. Par exemple, l'identification d'images faciales obtenues dans un environnement non contraint (par exemple, les changements d'éclairage, de posture, ou d'expression faciale, ainsi que les occlusions partielles, les déguisements, ou les mouvements de la caméra) pose encore plusieurs défis. En

d'autres termes, les technologies existantes sont encore loin des capacités du système visuel humain.

Dans notre vie quotidienne, le visage est peut-être la caractéristique biométrique la plus courante et familière. Avec l'invention de la photographie, les départements gouvernementaux, et les entités privées ont conservé des photographies faciales (issues de documents d'identité personnels, de passeports, ou de cartes d'adhésion). Ces collections ont été utilisées dans des enquêtes médico-légales, en tant que bases de référence, pour faire correspondre et comparer des images faciales d'une personne interrogée (par exemple, un suspect, un témoin, ou une victime). De plus, l'utilisation généralisée des appareils photo numériques et des smartphones a facilité la production quotidienne d'images faciales; ces images peuvent être facilement distribuées et échangées par le biais de réseaux sociaux rapidement établis tels que Facebook et Twitter.

La reconnaissance faciale a une histoire longue et suscite l'intérêt de plusieurs domaines, tels que la neurologie, la psychologie, et l'informatique [18]. Comparé à d'autres traits biométriques, le visage humain n'est pas considéré comme une modalité idéale. Il est généralement moins précis que des modalités biométriques telles que l'iris ou l'empreinte digitale, et peut être influencé par des facteurs tels que les cosmétiques, les déguisements, et l'éclairage [19]. Cependant, malgré ces limitations, le visage présente des avantages qui en font l'une des caractéristiques biométriques les plus populaires pour la reconnaissance d'identité. Certains de ces avantages incluent :

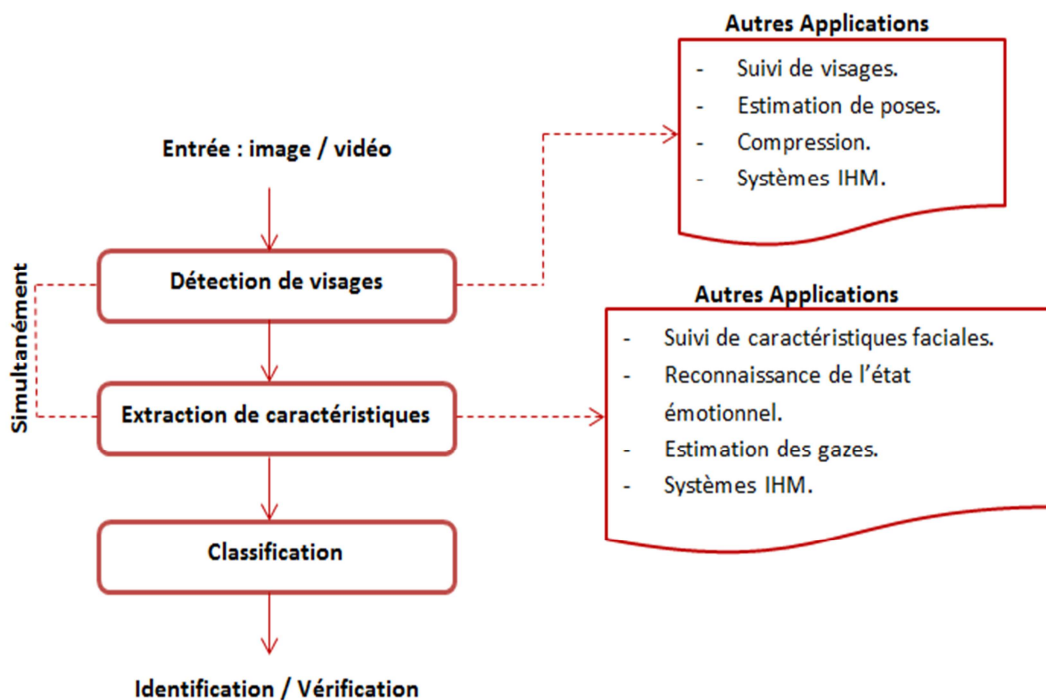
- **Caractère naturel:** La reconnaissance faciale est une caractéristique biométrique couramment utilisée par les humains pour l'authentification et l'identification, ce qui en fait l'une des plus pertinentes. Par exemple, dans le domaine du contrôle d'accès, les administrateurs peuvent facilement surveiller et évaluer les personnes autorisées en utilisant leurs caractéristiques faciales [20]. En revanche, l'identification des empreintes digitales ou de l'iris nécessite généralement des compétences spécialisées pour obtenir une confirmation précise. Cependant, il est important de noter que la reconnaissance faciale n'est pas sans limitations, et des préoccupations légitimes concernant la vie privée et l'utilisation abusive des données biométriques doivent être prises en compte.
- **Non intrusif:** Contrairement aux empreintes digitales ou aux images de l'iris, les images faciales peuvent être rapidement obtenues sans contact physique; les personnes se sentent plus détendues lorsqu'elles utilisent le visage comme identifiant

biométrique. De plus, un dispositif de reconnaissance faciale peut collecter des données de manière conviviale que les gens acceptent couramment [21].

- **Moins de coopération:** La reconnaissance faciale nécessite moins d'assistance de la part de l'utilisateur par rapport à l'iris ou aux empreintes digitales. Pour certaines applications limitées telles que la surveillance, un dispositif de reconnaissance faciale peut reconnaître un individu sans intervention active de la part du sujet [21].

### 2.3. Etapes essentielles pour les systèmes de reconnaissance faciale

En ingénierie, le problème de la reconnaissance automatique de visage comprend trois étapes essentielles [22] (comme présentées dans la Figure 2.1): (1) détection et normalisation approximative du visage, (2) extraction des caractéristiques et normalisation précise du visage, et (3) classification (vérification ou identification).



**Figure 2.1.** La conception standard d'un système de reconnaissance automatique de visage.

La détection des visages est la première étape d'un système de reconnaissance automatique des visages. Elle permet de déterminer si une image contient un ou plusieurs visages. Si c'est le cas, sa fonction est de repérer la ou les positions des visages dans l'image [23].

L'étape d'extraction des caractéristiques consiste à extraire du visage détecté un vecteur de caractéristiques appelé "signature", qui doit être suffisamment représentatif d'un visage. Il est nécessaire de vérifier l'individualité du visage et la capacité à distinguer deux personnes

distinctes. Il convient de noter que l'étape de détection des visages peut accomplir ce processus.

La classification implique la vérification et l'identification. La vérification nécessite de comparer un visage à un autre pour autoriser l'accès à une identité demandée. Cependant, l'identification compare un visage à plusieurs autres visages qui sont fournis avec plusieurs possibilités pour trouver l'identité du visage en question.

Parfois, certaines étapes ne sont pas distinctes. Par exemple, les caractéristiques faciales telles que les yeux, la bouche et le nez, qui sont utilisées pour extraire des informations pertinentes, sont souvent utilisées dans le processus de détection de visage. Dans certains cas, la détection et l'extraction des caractéristiques peuvent être effectuées simultanément, comme illustré dans la Figure 2.1.

Selon la complexité de l'environnement d'application, certains facteurs externes peuvent entraîner des distributions d'identité intra-visages élevées (ou des distributions d'identité inter-visages faibles) et dégrader la précision de la reconnaissance. Parmi ces facteurs, on peut citer la taille de la base de données, un éclairage faible ou élevé, la présence de bruit ou de flou, les déguisements, l'occultation partielle et certains facteurs secondaires qui sont souvent communs, inévitables, et très difficiles à gérer [24]. Dans un environnement bruyant, il peut être nécessaire d'effectuer un prétraitement de l'image [25-27].

Chaque étape du processus de reconnaissance faciale automatique, à savoir la détection des visages, l'extraction des caractéristiques faciales, et la vérification/identification, est considérée comme une question de recherche cruciale. Ces étapes nécessitent des améliorations continues et jouent un rôle essentiel dans de nombreuses applications, comme illustré dans la Figure 2.1. Par exemple, la détection des visages est nécessaire pour activer la surveillance faciale, tandis que l'extraction des caractéristiques faciales est cruciale pour identifier l'état émotionnel d'une personne, ce qui est essentiel dans les systèmes d'interaction homme-machine (IHM). En abordant chaque étape individuellement, il devient plus facile d'évaluer et de faire progresser l'état de l'art dans chaque domaine.

Dans cette thèse, nous nous concentrons principalement sur l'extraction des caractéristiques (et éventuellement la sélection des caractéristiques) ainsi que la classification.

#### **2.4. Défis ouverts de la reconnaissance faciale**

Comme dans la plupart des applications biométriques, les variations d'apparence causées par des environnements non contrôlés posent des défis ouverts en matière de reconnaissance

faciale. Dans les paragraphes suivants, nous citerons certains défis à relever dans un avenir proche.

#### 2.4.1. Reconnaissance faciale et occlusion

Le visage peut être capturé dans différentes poses et environnements sans assistance de l'utilisateur, ce qui peut entraîner des images ne montrant qu'une partie du visage. Les occlusions faciales peuvent prendre diverses formes, telles que des ombres, des foulards, des chapeaux, des voiles, des objets tenus par la personne (comme une main, de la nourriture ou un téléphone portable), une mauvaise qualité d'éclairage (comme une forte lumière créant une ombre), une auto-occlusion (lorsque le visage n'est pas orienté de face), ou une mauvaise qualité d'image (comme un flou). Cela est illustré dans la Figure 2.2.

Par exemple, lorsqu'il s'agit d'identifier un suspect dans une enquête à partir d'une image enregistrée, il peut être nécessaire de faire correspondre un visage partiellement occulté avec d'autres visages dans une foule. La reconnaissance des occlusions faciales présente des défis particuliers. Tout d'abord, les occlusions altèrent les caractéristiques distinctives du visage et augmentent la distance dans l'espace des caractéristiques entre deux images de la même personne. Les variations intra-classe deviennent plus importantes que les variations inter-classes, ce qui affecte négativement les performances de reconnaissance. Deuxièmement, lorsque les repères faciaux sont occultés, des erreurs d'alignement significatives peuvent se produire, ce qui réduit les taux de reconnaissance [28].



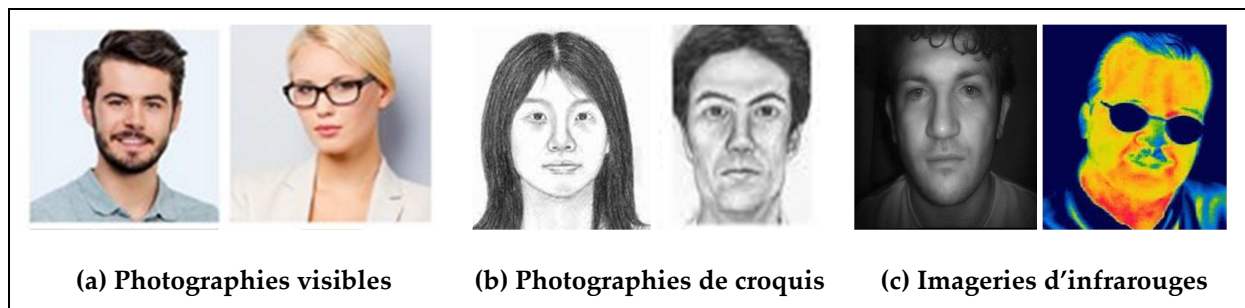
**Figure 2.2.** Quelques exemples d'occultation par chapeau, lunettes, masque, main, ombre et auto-occultation.

#### 2.4.2. Reconnaissance de visages hétérogènes

La reconnaissance de visages hétérogènes consiste à établir une corrélation entre deux représentations faciales provenant de méthodes d'imagerie différentes, ce qui s'avère particulièrement utile dans des situations légales. Par exemple, l'imagerie infrarouge (illustrée dans la Figure 2.3.c) peut être la seule méthode permettant d'acquérir une image utilisable d'un suspect dans des environnements nocturnes, mais les fichiers enregistrés par les forces de l'ordre sont généralement des images visibles [29]. Un autre exemple est la correspondance

entre des photographies de croquis (Figure 2.3.b) et des photographies faciales. Lorsqu'aucune image du suspect n'est disponible, un croquis artistique est créé à partir de la description fournie par un témoin oculaire. La corrélation entre les croquis et les photographies faciales est essentielle dans le cadre d'enquêtes légales.

Ces techniques de correspondance entre différentes modalités d'imagerie permettent d'exploiter pleinement les informations disponibles pour identifier des individus dans des situations où les images traditionnelles ne sont pas suffisantes ou indisponibles. Cela ouvre des possibilités précieuses pour les enquêtes et le maintien de l'ordre dans le domaine légal.



**Figure 2.3.** Quelques modalités d'images présentant des visages hétérogènes. (a) Photographies simples, (b) Images de croquis, et (c) Images infrarouges.

### 2.4.3. Reconnaissance faciale et vieillissement

Le processus de vieillissement du visage est en effet complexe et affecte à la fois la forme et la texture du visage, tels que le teint de la peau et l'apparition des rides. L'application courante des systèmes de reconnaissance faciale face à l'effet du vieillissement vise à détecter la présence d'une personne spécifique dans une base de données préalablement enregistrée. Cela peut être utilisé, par exemple, pour l'identification d'enfants disparus ou pour le contrôle de suspects figurant sur une liste de surveillance. Cependant, à mesure que l'écart d'âge entre une image de requête et une image de référence de la même personne augmente, la précision des systèmes de reconnaissance tend à diminuer, comme illustré dans la Figure 2.4 [30].



**Figure 2.4.** Exemple de vieillissement du visage.



#### 2.4.4. Reconnaissance faciale 3D

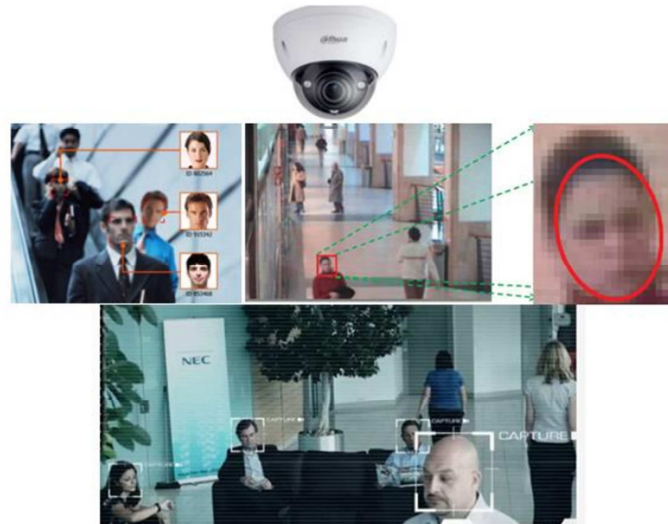
Le visage est un candidat fascinant pour les humains en raison de son acquisition sans contact et de sa nature non intrusive. Bien que les approches en 2D aient donné d'excellentes performances, elles restent très sensibles aux problèmes liés aux changements des conditions d'éclairage, de la posture, de l'expression faciale, des cosmétiques, et de l'occlusion. L'implémentation de la forme 3D du visage dans l'identification des individus est théoriquement considérée comme une manière très positive de résoudre ces problèmes. Cependant, certains problèmes ne sont pas encore bien résolus, surtout lorsque le sujet n'est pas coopératif pendant le processus d'acquisition, ce qui peut entraîner une différence de posture, d'expression faciale, et générer des occlusions par des corps étrangers à la surface du visage [31].

Concernant la constitution de bases de données, les données faciales en 2D peuvent être obtenues facilement en recherchant sur Internet, tandis que les données faciales en 3D nécessitent une collecte physique auprès de sujets réels, ce qui restreint leur utilisation et leur évolution. Nous pensons que la création de bases de données de reconnaissance faciale en 3D à grande échelle pourrait ouvrir la voie à des applications étendues dans le monde de la biométrie [12].

#### 2.4.5. Reconnaissance faciale dans la vidéosurveillance

Les systèmes de reconnaissance faciale sont de plus en plus répandus et sont utilisés dans divers domaines, notamment dans le domaine de la vidéosurveillance (comme illustré dans la Figure 2.5). Dans ce contexte, la performance des systèmes de reconnaissance faciale dépend principalement des conditions d'acquisition des images. Les changements de posture peuvent avoir un impact sur la qualité des images capturées, et les techniques d'acquisition elles-mêmes peuvent introduire des artefacts indésirables [32]. L'un des problèmes couramment rencontrés est la mise au point de la caméra, qui peut entraîner des images floues, de faible résolution, ou des erreurs liées à la compression et aux effets de blocs.

Le défi pour les systèmes de reconnaissance faciale dans ces cas spécifiques est de pouvoir distinguer les individus à partir de photographies capturées par des caméras de vidéosurveillance présentant des flous, une faible résolution, des artefacts de blocs, ou des visages avec des poses variables. Ce défi demeure un problème non résolu et nécessite des recherches supplémentaires afin de développer des techniques plus robustes et efficaces pour traiter ces images de qualité inférieure.



**Figure 2.5.** Exemples de reconnaissance faciale dans la vidéosurveillance.

#### **2.4.6. Reconnaissance faciale à partir d'un seul échantillon**

L'une des situations les plus exclusives et réalistes dans lesquelles s'applique la reconnaissance faciale est celle de l'échantillon unique par personne (*Single Sample Per Person* (SSPP)), ou simplement la reconnaissance faciale à partir d'un seul échantillon (*Single Sample Face Recognition* (SSFR)) [33]. Il s'agit là d'un des problèmes les plus difficiles de reconnaissance faciale, dans lequel il n'y a qu'une seule représentation faciale par individu dans l'ensemble d'apprentissage. Il est bien connu que :

- Dans les applications du monde réel (par exemple, les passeports, les systèmes d'immigration), un seul modèle de chaque individu est enregistré dans la base de données et accessible pour la tâche de reconnaissance.
- Les systèmes de reconnaissance de formes nécessitent une grande quantité de données d'entraînement pour assurer la généralisation des systèmes d'apprentissage.
- Les approches basées sur l'apprentissage profond sont considérées comme des techniques puissantes en reconnaissance faciale. Cependant, elles nécessitent une quantité significative de données d'entraînement pour bien performer [33].

En résumé, on peut conclure que la reconnaissance faciale à partir d'un seul échantillon reste une question non résolue et fait partie des sujets les plus courants dans les milieux universitaires et industriels.

## 2.5. Pourquoi la reconnaissance faciale à partir d'un seul échantillon?

La plupart des méthodes actuelles dans la littérature utilisent plusieurs images faciales (échantillons) par personne dans l'ensemble d'entraînement. Cependant, dans les systèmes du monde réel (par exemple, dans le suivi des fugitifs, les cartes d'identité, la gestion de l'immigration, ou les passeports), seuls les systèmes SSFR sont utilisés en raison des limitations de stockage et des politiques de confidentialité. Ces systèmes utilisent un seul échantillon par personne lors de l'étape d'entraînement (généralement des images neutres acquises dans des conditions contrôlées). En d'autres termes, un seul exemple de la personne à reconnaître est enregistré dans la base de données et accessible pour la tâche de reconnaissance [34]. Étant donné qu'il y a peu de données disponibles (c'est-à-dire que nous n'avons pas plusieurs échantillons par personne) pour effectuer un apprentissage supervisé, de nombreux algorithmes bien connus peuvent ne pas fonctionner correctement. Par exemple, les réseaux neuronaux convolutifs [35] sont considérés comme des techniques puissantes de reconnaissance faciale à plusieurs échantillons. Cependant, ils nécessitent une quantité considérable de données d'entraînement pour bien fonctionner, ce qui est en contradiction avec le principe du SSFR. Il est bien connu qu'un grand volume de données d'entraînement est nécessaire pour assurer la généralisation des systèmes d'apprentissage.

L'utilisation de l'imagerie tridimensionnelle (3D) au lieu de la représentation bidimensionnelle (2D) a permis de résoudre plusieurs problèmes liés aux conditions d'acquisition d'images, en particulier les variations de pose, d'éclairage, et de maquillage. Bien que les modèles 3D offrent une meilleure représentation de la forme du visage pour une distinction claire entre les personnes [33] (même au cas des SSFR), ils ne sont souvent pas adaptés aux applications en temps réel car ils nécessitent des calculs coûteux et sophistiqués ainsi que des capteurs spécifiques.

Nous concluons que la reconnaissance faciale à partir d'un seul échantillon par personne reste un problème non résolu et nécessite des efforts considérables pour que ce sujet doive être évolué pour le milieu académique et pour une exploitation commerciale, gouvernementale, ou légale.

## 2.6. Etat de l'art en reconnaissance faciale à partir d'un seul échantillon

Plusieurs méthodes et algorithmes ont été suggérés dans la littérature pour surmonter le problème de SSFR. Ils peuvent être subdivisés en deux approches fondamentales en fonction

de la méthode utilisée pour l'extraction des caractéristiques et la classification, à savoir : les méthodes traditionnelles et les méthodes basées sur l'apprentissage profond.

### 2.6.1. Méthodes traditionnelles

Les méthodes actuelles visant à résoudre le problème de la reconnaissance faciale à partir d'un seul échantillon (SSFR) peuvent être aussi subdivisées en trois classes fondamentales [33], à savoir: la génération d'échantillons virtuels, l'apprentissage générique, et la partition d'images.

#### 2.6.1.1. Méthodes de génération d'échantillons virtuels

Les méthodes de cette catégorie produisent des échantillons d'entraînement virtuels supplémentaires pour chaque individu afin d'augmenter la galerie (c'est-à-dire l'augmentation des données), de sorte que l'apprentissage discriminatif en sous-espace puisse être utilisé pour extraire des caractéristiques.

Par exemple, Vetter (1998) [36] a proposé un algorithme robuste de SSFR en générant des modèles faciaux 3D par la récupération d'une fidélité élevée de la réflectance et de la géométrie. Zhang et al. (2005) [37] et Gao et al. (2008) [38] ont développé deux techniques pour résoudre le problème de SSFR basé sur la décomposition en valeurs singulières (en anglais, singular value decomposition (SVD)). Hu et al. (2015) [39] ont suggéré un système SSFR différent basé sur l'algorithme lower-upper (LU). Dans leur approche, chaque sujet individuel a été décomposé et transposé en utilisant la procédure LU et chaque image brute a été réarrangée selon son énergie. Dong et al. (2018) [40] ont proposé une méthode efficace pour la réalisation de tâches de SSFR appelée "K-nearest neighbors virtual image set-based multimanifold discriminant learning" (KNNMMDL). Ils ont également suggéré un algorithme appelé "K-nearest neighbor-based virtual sample generating" (KNNVSG) pour augmenter les informations sur la variation intra-classe dans les échantillons d'entraînement. Ils ont également proposé l'algorithme "image set-based multimanifold discriminant learning" (ISMMDL) pour exploiter les informations sur la variation intra-classe.

Bien que ces méthodes puissent atténuer quelque peu le problème de SSFR, leur principal inconvénient réside dans la forte corrélation entre les images virtuelles, qui ne peuvent pas être considérées comme des exemples indépendants pour la sélection des caractéristiques.

### 2.6.1.2. Méthodes d'apprentissage générique

Les méthodes de cette catégorie extraient d'abord des caractéristiques discriminantes à partir d'un ensemble d'entraînement générique supplémentaire qui comprend plusieurs exemples par individu, puis utilisent ces caractéristiques pour les tâches de SSFR.

Par exemple, Deng et al. (2012) [41] ont développé la technique "extended sparse representation classifier" (ESRC), dans laquelle le dictionnaire variant intra-classe est créé à partir de personnes génériques non incorporées dans l'ensemble de la galerie afin d'augmenter l'efficacité du processus d'identification. Dans une méthode appelée "sparse variation dictionary learning" (SVDL), Yang et al. (2013) [42] ont entraîné un dictionnaire de variation parcimonieux en tenant compte de la relation entre l'ensemble d'entraînement et l'ensemble générique externe, en ignorant les caractéristiques distinctives des différents organes du visage humain. Zhu et al. (2014) [43] ont proposé un système pour le SSFR basé sur la représentation générique locale (en anglais, local generic representation (LGR)), qui tire parti des avantages à la fois de la partition d'image et de l'apprentissage générique, et prend en compte le fait que la variation intra-classe du visage peut être répartie entre différents sujets.

### 2.6.1.3. Méthodes de partition d'images

Les méthodes de cette classe divisent les images de chaque individu en blocs locaux, extraient les caractéristiques discriminantes, et effectuent enfin des classifications basées sur les caractéristiques discriminantes sélectionnées.

Par exemple, Zhu et al. (2012) [44] ont développé un algorithme nommé PCRC (c'est-à-dire, patch collaborative representation-based classification), qui applique à chaque bloc la méthode originale proposée par Zhang et al. (2011) [45], appelée (CRC). Lu et al. (2012) [46] ont suggéré une technique appelée discriminant multi-manifold analysis (DMMA) qui divise chaque image en plusieurs blocs non chevauchants, puis apprend plusieurs espaces de caractéristiques pour optimiser les différentes marges des différents individus. Zhang et al. (2018) [47] ont développé des opérateurs d'images de visage basés sur des histogrammes locaux. Ils ont découpé chaque image en différents blocs non chevauchants. Ensuite, ils ont dérivé une matrice pour projeter les blocs dans un sous-espace optimal afin de maximiser les différentes marges des différents individus. Chaque colonne a ensuite été transformée en un filtre d'image pour traiter les images faciales et les réponses des filtres ont été binarisées à l'aide d'un seuil fixe. Gu et al. (2018) [48] ont proposé une méthode appelée local robust sparse representation (LRSR). L'idée principale de cette technique est de fusionner un modèle

de représentation parcimonieuse locale avec un modèle d'apprentissage de dictionnaire de variation générique basé sur les blocs pour déterminer les variations intra-classes possibles des images de tests. Zhang et al. (2021) [49] ont introduit une nouvelle mesure de distance pour le classificateur nearest neighbor classifier (NNC) pour résoudre les problèmes de SSFR. La technique proposée, nommée dissimilarity-based nearest neighbor classifier (DNNC), divise toutes les images en blocs égaux non chevauchants et produit un ensemble de blocs d'images organisé. Les dissimilarités entre l'ensemble de blocs d'images de requête donné et les ensembles de blocs d'images d'entraînement sont calculées et prises en compte par la métrique de distance du NNC.

## **2.6.2. Méthodes basées sur l'apprentissage profond**

### *2.6.2.1. C'est quoi l'apprentissage profond ?*

Les réseaux de neurones artificiels profonds, connus sous le nom d'apprentissage profond (ou deep learning en anglais), ont remporté de nombreuses compétitions en matière d'apprentissage automatique et de reconnaissance de motifs ces dernières années [50]. L'apprentissage profond, qui fait partie d'une classe d'apprentissage automatique, utilise des couches cachées successives de niveaux de traitement de l'information, organisées de manière hiérarchique pour la représentation ou la classification de motifs, ainsi que pour l'apprentissage de caractéristiques [51]. Selon Deng et Yu [52] (2014), il y avait trois raisons principales qui ont permis l'évolution de l'apprentissage profond: la croissance spectaculaire des capacités de traitement (par exemple, les unités GPU), la baisse considérable des coûts matériels informatiques, et enfin les progrès récents dans les études sur l'apprentissage automatique. De nombreux chercheurs ont obtenu des résultats satisfaisants en utilisant l'apprentissage profond dans diverses applications de vision par ordinateur, de reconnaissance de la parole conversationnelle, de reconnaissance phonétique, de recherche vocale, de codage de caractéristiques de la parole et de l'image, de reconnaissance d'écriture manuscrite, de classification sémantique des énoncés, de reconnaissance d'objets visuels et de traitement audio, et de recherche d'informations [52].

Les architectures profondes discriminatives ou l'apprentissage supervisé sont censées différencier plusieurs parties des données pour la classification. Le réseau neuronal convolutif (en anglais, convolutional neural network (CNN)) est le meilleur exemple d'apprentissage supervisé; il permet une efficacité architecturale exceptionnelle pour la reconnaissance d'images [53]. La reconnaissance faciale est couramment étudiée en vision par ordinateur, et le CNN a connu un grand succès, devenant une référence incontournable dans ce domaine.

### 2.6.2.2. Réseau neuronal convolutif ou Convolutional neural network (CNN)

Les réseaux de neurones convolutifs sont une forme de réseaux de neurones qui ont connu un succès dans des domaines tels que la reconnaissance et la classification d'images. Les CNNs se composent d'un ensemble de filtres/noyaux/neurones avec des paramètres ou des poids et des biais apprenables. Chaque filtre prend des entrées, effectue une convolution, puis applique une fonction de non-linéarité. La structure des CNNs comprend des couches de convolution, de pooling, d'unité linéaire rectifiée (ReLU), et de connexions entièrement connectées (fully connected).

- **Couche de convolution (Conv):** Il s'agit du bloc de construction essentiel du CNN qui vise à extraire des caractéristiques à partir des données d'entrée. Chaque couche utilise une opération de convolution pour obtenir une carte des caractéristiques. Ensuite, les cartes d'activation ou de caractéristiques sont transmises à la couche suivante en tant que données d'entrée [54].
- **Couche de pooling:** Il s'agit d'une forme de sous-échantillonnage non linéaire qui réduit la dimensionnalité de la carte des caractéristiques tout en conservant les informations essentielles. Il existe différentes fonctions de pooling non linéaires, parmi lesquelles le max-pooling est le plus efficace au sous-échantillonnage [55].
- **Couche ReLU:** Il s'agit d'une opération non linéaire qui utilise des unités rectifieuses.
- **Couche fully connected (FC):** Le raisonnement de haut niveau dans le réseau neuronal est réalisé à travers des couches FC après l'application de différentes couches de convolution et de couches de pooling [56].

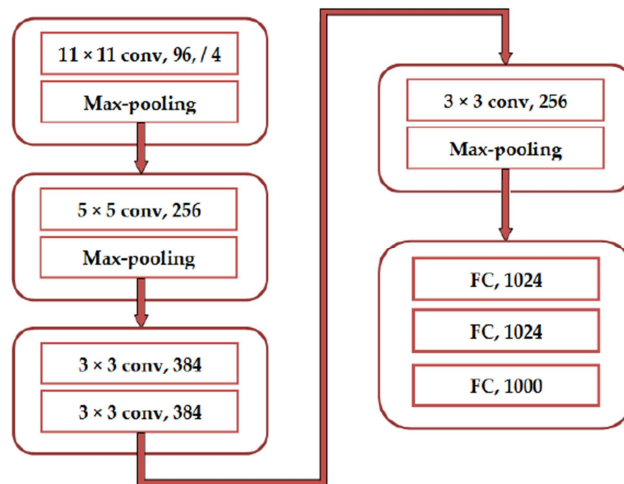
### 2.6.2.3. Architectures populaires des CNNs

Dans ce qui suit, nous présentons quelques architectures de CNNs largement utilisées.

#### - AlexNet

Le défi ILSVRC (*ImageNet Large Scale Visual Recognition Challenge*) est basé sur une base de données pour la reconnaissance d'objets à grande échelle organisée sous forme de compétitions annuelles qui ont débuté en 2010 jusqu'à aujourd'hui. Krizhevsky et al. [57] ont gagné l'ILSVRC-2012 en utilisant un grand modèle de CNN profond, appelé AlexNet, qui a obtenu des résultats révolutionnaires en vision par ordinateur par rapport à tous les algorithmes d'apprentissage automatique traditionnels. AlexNet comprend cinq couches de convolution, certaines étant suivies d'un max-pooling, ainsi que trois couches de type FC avec

une softmax à 1000 classes, comme illustré dans la Figure 2.6, et d'autres techniques telles que le dropout, ReLU, et l'augmentation de données.



**Figure 2.6.** Architecture de l'AlexNet.

#### - VGGNet

Lors de l'ILSVRC-2014, Simonyan et al. [58] ont exploré l'impact de la profondeur du réseau de convolution sur l'exactitude de la reconnaissance d'images à grande échelle. Leur contribution principale a été d'utiliser une architecture appelée VGGNet avec de petits filtres de convolution (3×3) et de doubler le nombre de cartes de caractéristiques après le max-pooling de type (2×2). La profondeur du réseau a été augmentée à 16-19 couches de poids, améliorant ainsi la flexibilité de l'architecture profonde pour apprendre des cartes non linéaires continues, comme illustré dans la Figure 2.7.



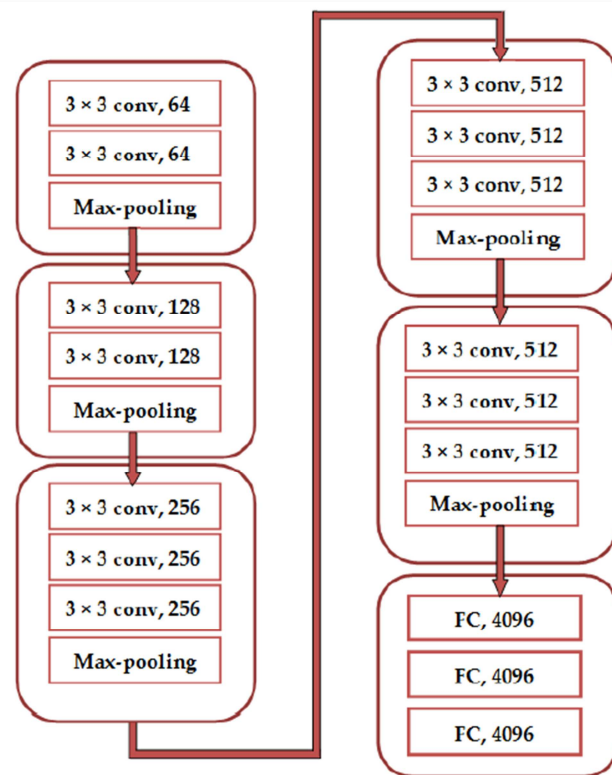


Figure 2.7. Architecture du VGGNet.

- ResNet

He et al. [59] (2015) ont introduit une nouvelle architecture novatrice appelée réseau neuronal résiduel (ResNet) pour faciliter l'entraînement de réseaux ultra-profonds par rapport aux réseaux déjà utilisés. ResNet a remporté l'ILSVRC-2015; il a été développé avec des "connexions de raccourci" et comprend une normalisation par lots. Il a été capable d'entraîner un réseau neuronal avec différents nombres de couches: 34, 50, 101, 152, voire même 1202. La Figure 2.8 illustre le diagramme de blocs de base de l'architecture ResNet.

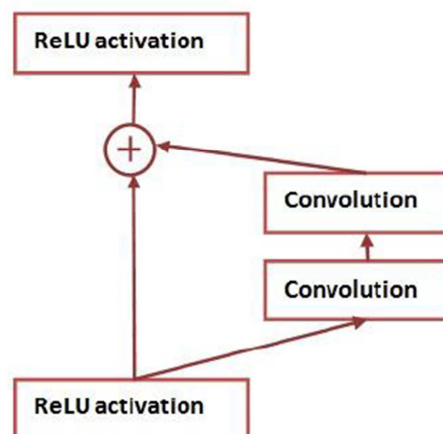


Figure 2.8. Architecture du ResNet.

#### 2.6.2.4. Méthodes de SSFR basées sur l'apprentissage profond

Zeng et al. (2017) [60] ont proposé une méthode utilisant un réseau CNN. Tout d'abord, ils proposent d'utiliser une technique d'échantillonnage expansif pour augmenter l'ensemble d'échantillons d'entraînement, puis un modèle CNN entraîné est mis en œuvre et affiné par ces échantillons expansifs pour être utilisé dans le processus de classification. Ding et al. (2017) [61] ont développé une technique d'apprentissage profond focalisé sur un réseau d'analyse en composantes principales à noyau (en anglais, kernel principal component analysis network (KPCANet)) et une nouvelle technique de vote pondéré. Tout d'abord, l'image faciale alignée est segmentée en plusieurs blocs non superposés pour créer l'ensemble d'entraînement. Ensuite, un KPCANet est utilisé pour obtenir des filtres et des banques de caractéristiques. Enfin, la reconnaissance de la classe non étiquetée est obtenue en appliquant la forme de vote pondéré. Zhang et Peng (2018) [62] ont introduit une méthode différente pour générer des variances intra-classes à l'aide d'un auto-encodeur profond. Ils ont ensuite utilisé ces variations intra-classes pour étendre les nouveaux exemples. Tout d'abord, un auto-encodeur profond généralisé est utilisé pour entraîner les images faciales de la galerie. Ensuite, un auto-encodeur profond spécifique à la classe (en anglais, class-specific deep auto-encoder (CDA)) est affiné avec un seul exemple. Enfin, le CDA correspondant est utilisé pour étendre les nouveaux échantillons. Du et Da (2020) [63] ont proposé une méthode intitulée apprentissage de dictionnaire par blocs (en anglais, block dictionary learning (BDL)) qui fusionne la représentation parcimonieuse (en anglais, sparse représentation (SR)) avec les CNNs. La SR est mise en œuvre pour améliorer l'efficacité du CNN en améliorant les variations de caractéristiques inter-classes et en créant un processus d'apprentissage de dictionnaire global à local pour augmenter la robustesse de la méthode.

Il est clair que l'approche basée sur l'apprentissage profond pour la reconnaissance faciale a suscité une attention particulière ces dernières années, mais elle souffre considérablement avec les systèmes SSFR car ils nécessitent encore une quantité importante d'informations dans l'ensemble d'entraînement.

## 2.7. Conclusion

Dans ce chapitre, nous avons présenté de manière exhaustive le nouvel état de l'art dans la recherche sur la reconnaissance faciale. Les défis ouverts dans ce domaine ont été clairement exposés et des perspectives d'amélioration ont été proposées. Naturellement, les nouveaux développements doivent répondre à quatre objectifs essentiels: une rapidité constante (offrant

une réponse immédiate perçue par l'utilisateur), une précision proche de 100 %, une sécurité optimale, ainsi que le développement d'équipements miniaturisés et portables.

Dans le cadre de cette thèse, nous avons donné une attention particulière au défi SSFR. Pour résoudre le problème SSFR, des discussions détaillées avec des comparaisons de nos contributions par rapport aux méthodes concurrentes sont fournies dans les chapitres suivants.

## **CHAPITRE 3: Caractéristiques Statistiques d'Images Binarisées en Multi-Bloc-Couleur pour la Reconnaissance Faciale à Echantillon Unique**

### **3.1. Introduction**

Le présent chapitre expose une méthode efficace pour résoudre le problème de la reconnaissance faciale à partir d'un seul échantillon (SSFR) en conditions non contrôlées, qui repose sur une variante du descripteur binarized statistical image features (BSIF), nommée multi-block color-binarized statistical image features (MB-C-BSIF). Cette méthode exploite plusieurs types d'information, à savoir: locale, régionale, globale, et de texture-couleur afin d'obtenir une représentation des visages précise et fiable.

Dans ce chapitre, nous présentons en détail le principe de notre méthode proposée. Ensuite, nous évaluons sa performance et comparons ses résultats avec ceux d'autres méthodes en littérature.

### **3.2. Méthode proposée**

Cette section détaille la méthode MB-C-BSIF qui est une approche basée sur le partitionnement d'images et qui comprend trois étapes clés: le prétraitement de l'image, l'extraction de caractéristiques basée sur MB-C-BSIF, et la classification. Une description détaillée de chaque étape est réalisée dans les sous-sections suivantes:

#### **3.2.1. Prétraitement**

Avant de commencer le processus d'extraction et de classification des caractéristiques de l'image, un prétraitement est nécessaire pour améliorer la qualité visuelle de l'image d'entrée.

Pour améliorer l'image faciale, deux étapes de prétraitement ont été effectuées. La première étape consiste à appliquer une normalisation d'histogramme pour égaliser la distribution des niveaux de gris dans l'image, ce qui peut aider à améliorer le contraste et la qualité de l'image. La deuxième étape consiste à appliquer un filtre non linéaire pour minimiser le bruit et améliorer la qualité de l'image tout en préservant l'apparence du visage. Le filtre médian [64] a été choisi pour cette tâche car il peut être efficace pour minimiser le bruit tout en préservant les contours et les détails importants de l'image.

Ces étapes de prétraitement peuvent aider à améliorer la qualité de l'image faciale et à faciliter le processus d'extraction et de classification des caractéristiques.

### 3.2.2. Extraction de caractéristiques basée sur MB-C-BSIF

Notre technique d'extraction de caractéristiques est basée sur la représentation couleur en multi-blocs du descripteur BSIF, appelée multi-block color BSIF (MB-C-BSIF). L'opérateur BSIF proposé par Kannala et Rahtu [65] est en effet un descripteur efficace et robuste pour l'analyse de texture, comme l'ont montré plusieurs études [66-67]. Le descripteur BSIF est un outil puissant pour extraire des informations de texture à partir d'images, qui peuvent ensuite être utilisées pour décrire les régions d'image sous forme d'histogrammes. Pour chaque pixel de l'image, la méthode calcule un code binaire en projetant un bloc d'image local centré sur ce pixel sur un sous-espace dont ses filtres ont été appris à partir d'images naturelles à l'aide de l'analyse en composantes indépendantes (ICA) [68]. Cette projection linéaire produit des coordonnées qui sont ensuite binarisées par seuillage afin d'obtenir le code binaire final pour ce pixel. La longueur de la chaîne du code binaire pour chaque pixel dépend du nombre de filtres appris par l'ICA, et les régions d'image peuvent être facilement représentées avec des histogrammes des codes binaires des pixels. Le BSIF s'inspire de descripteurs d'image qui génèrent des codes binaires, tels que le motif binaire local (en anglais, local binary pattern (LBP)) [69] et la quantification de phase locale (en anglais, local phase quantization (LPQ)) [70]. Cependant, contrairement à ces descripteurs, BSIF utilise des filtres appris à partir d'images naturelles, ce qui permet de mieux capturer les propriétés statistiques des images naturelles. Ce qui suit décrit brièvement les détails du descripteur BSIF.

Le calcul du code BSIF, comme décrit dans la référence [65], se fait de la manière suivante:

1. Etant donné un patch d'image  $X$  de taille  $l \times l$  pixels et un filtre linéaire  $W_i$  de la même taille, la réponse au filtre  $s_i$  est obtenu par :

$$s_i = \sum_{u,v} W_i(u, v) X(u, v) = w_i^T x \quad (3.1)$$

où

$W_i$  : Filtre linéaire,

$s_i$  : Résultat de la convolution du filtre  $W_i$  avec le patch  $X$ ,

$u, v$  : Les coordonnées spatiales,

$i$  : Indice du filtre  $W_i$ ,

$w$  et  $x$  : Deux vecteurs contenant les pixels de  $W_i$  et  $X$ , respectivement.

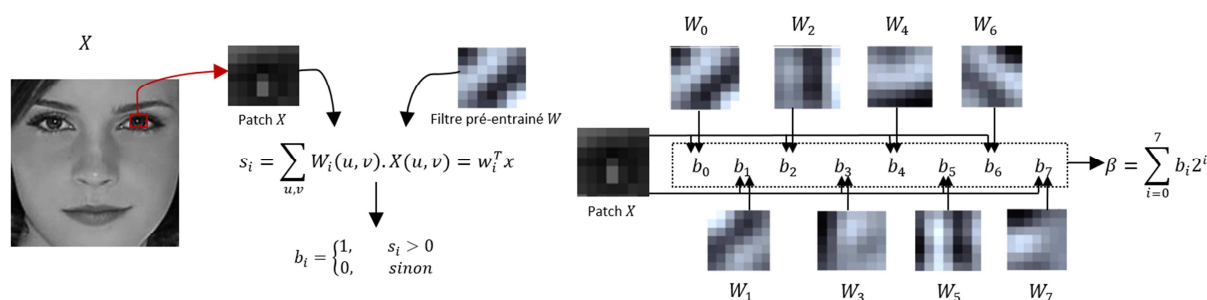
2. Le calcul des caractéristiques de binarisation se fait par la fonction  $b_i$ :

$$b_i = \begin{cases} 1, & \text{si } s_i > 0 \\ 0, & \text{sinon} \end{cases} \quad (3.2)$$

Après avoir calculé toutes les caractéristiques binarisées correspondantes à chaque filtre  $W_i$ , nous obtenons le « libellé du descripteur BSIF »  $\beta$  comme suit:

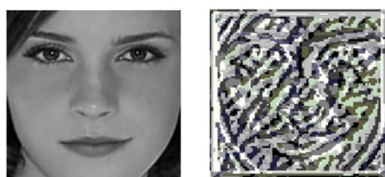
$$\beta = \sum_{i=0}^7 b_i 2^i \quad (3.3)$$

Le descripteur BSIF est caractérisé par deux paramètres clés: la taille du filtre  $l \times l$  et la longueur de la chaîne de bits  $n$ .



(a) Extraction d'une étiquette du descripteur

(b) Calcul d'un modèle binarisé unique en utilisant le descripteur



(c) Exemple d'une image originale et sa représentation après l'application de BSIF.

**Figure 3.1.** Illustration sur le fonctionnement du descripteur BSIF avec un filtre de taille  $7 \times 7$  et 8 bits.

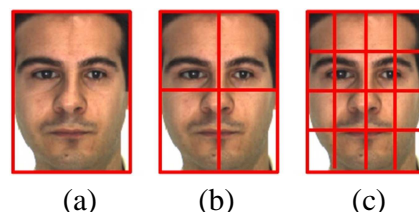
En utilisant ICA, les filtres  $W_i$  sont appris en minimisant la dépendance statistique de  $s_i$ , ce qui peut améliorer la performance du descripteur et augmenter la capacité de capturer les caractéristiques pertinentes de l'image. Pour entraîner ces filtres  $W_i$ , différents choix de valeurs de paramètres ont été considérés. En particulier, chaque ensemble de filtres a été entraîné à l'aide

de 50000 patch d'images. Figure 3.1 montre une illustration sur le fonctionnement du descripteur BSIF avec un filtre de taille  $7 \times 7$  et 8 bits.

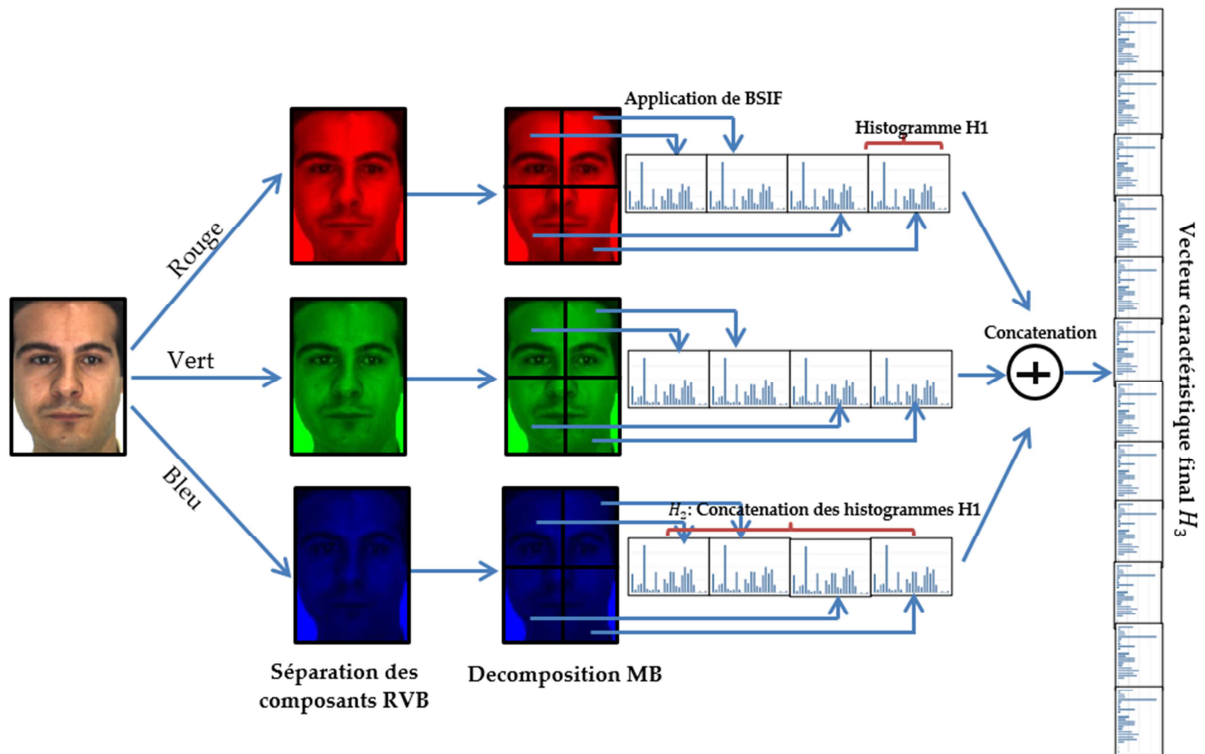
Comme pour les méthodologies LBP et LPQ, les cooccurrences des codes BSIF sont collectées dans un histogramme  $H1$ , qui est utilisé comme vecteur de caractéristiques pour représenter l'image. Le vecteur de caractéristiques  $H1$  est invariant à l'échelle, à la rotation, et à la translation de l'image, ce qui le rend approprié pour une utilisation dans des tâches de classification d'images.

L'opérateur BSIF simple basé sur un seul bloc peut ne pas être suffisamment robuste pour certaines tâches de traitement d'image, comme l'occlusion et la grande rotation de l'image. Afin de pallier certaines de ses limitations, une extension nommée BSIF en multi-blocs (MB-BSIF) est employée. Cette méthode consiste à partitionner l'image originale en blocs non superposés. Une image faciale peut être divisée de manière égale selon les directions horizontale et verticale, ce qui permet une représentation plus précise de la texture de l'image. À titre d'exemple, il est possible de segmenter l'image en grilles de  $1 \times 1$ ,  $2 \times 2$  ou  $4 \times 4$  afin de dériver respectivement 1, 4 ou 16 blocs, comme illustré dans la Figure 3.2. Chaque bloc renferme des informations relatives aux caractéristiques du visage, telles que le nez, les yeux, et les sourcils, ainsi que des données sur les relations spatiales entre ces éléments, comme la position du nez par rapport à la bouche ou l'oeil par rapport à l'autre oeil. Ces blocs et leurs données interconnectées sont essentiels pour la tâche de SSFR.

Notre proposition est de diviser l'image en blocs égaux et non superposés, puis de calculer les histogrammes de l'opérateur BSIF pour chaque bloc. L'histogramme  $H2$  est obtenu en combinant les histogrammes individuels des blocs réguliers, comme illustré à la Figure 3.2.



**Figure 3.2.** Exemple de décomposition d'une image en multi-blocs (MB): (a)  $1 \times 1$ , (b)  $2 \times 2$ , et (c)  $4 \times 4$ .



**Figure 3.3.** Cadre de l'approche proposée pour l'extraction des caractéristiques: MB-C-BSIF.

Dans le domaine de la reconnaissance faciale, certaines études se sont limitées à l'analyse des détails de luminance des images faciales en niveaux de gris. En revanche, ce chapitre présente une méthode intéressante qui utilise les informations de la texture couleur et démontre que l'analyse de la chrominance peut être bénéfique pour les systèmes de SSFR. Pour se faire, il est possible de diviser l'image faciale RVB en trois canaux distincts, à savoir le rouge, le vert et le bleu, puis de calculer le MB-BSIF pour chaque canal séparément. Le vecteur de caractéristiques final est obtenu en concaténant les histogrammes de chaque canal dans un histogramme global  $H_3$ . Cette approche est appelée multi-block color BSIF (MB-C-BSIF). La Figure 3.3 fournit une illustration schématique de l'architecture proposée MB-C-BSIF.

Bien que le modèle RVB soit couramment utilisé pour détecter, modéliser, et afficher des images en couleur, il présente certaines limitations dans l'interprétation d'images. L'une des principales limitations du modèle RVB est que les canaux de couleur rouge, vert, et bleu sont fortement corrélés les uns avec les autres, ce qui peut rendre difficile la distinction des détails en termes de luminance et de chrominance. Pour ces raisons, nous avons étudié et testé deux espaces colorimétriques supplémentaires, HSV et YCbCr, pour exploiter les détails de la texture



couleur. Le HSV (en anglais, hue (H), saturation (S), and value (V)) sépare la chrominance et la luminance en trois dimensions: teinte (T), saturation (S), et luminosité (L), ce qui permet de décrire la texture couleur avec plus de précision. De même, l'espace colorimétrique YCbCr divise les composantes RVB en trois canaux: luminance (Y), chrominance bleue (Cb) et chrominance rouge (Cr), offrant une autre façon de séparer les informations de couleur et de fournir des descriptions de texture plus détaillées. En utilisant ces espaces colorimétriques supplémentaires, il est possible d'exploiter les détails de la texture couleur de manière plus efficace et précise, ce qui est particulièrement utile dans les systèmes SSFR.

### 3.2.3. Classification

Lors du processus de la reconnaissance faciale, chaque image faciale est comparée aux images stockées dans la base de données. Pour identifier l'identité correspondante de l'image testée, nous avons utilisé le classificateur K-nearest neighbors (K-NN) [71] associé à une mesure de distance. Les K-NNs sont largement utilisés dans diverses applications, car ils sont flexibles et faciles à utiliser. En termes techniques, le classificateur K-NN est utilisé pour attribuer une étiquette à une personne de test  $x' \in R^D$  sur la base de données d'apprentissage présentée  $\{(x_i, y_i) \mid i = 1, 2, \dots, s\}$ . Chaque  $x_i \in R^D$  représente le vecteur de caractéristiques de la  $i^{\text{ème}}$  personne et  $y_i$  désigne son étiquette correspondante,  $D$  est la dimension du vecteur de caractéristiques, et  $s$  représente le nombre de personnes. L'algorithme K-NN calcule la distance entre  $x'$  et chaque  $x_i$  pour trouver le voisin le plus proche, puis attribue l'étiquette de la classe majoritaire parmi ces voisins à  $x'$ .

Il est possible de mettre en place différentes mesures de distance pour l'algorithme K-NN. Dans ce travail, nous avons comparé et évalué trois mesures de distance couramment utilisées: la distance de Hamming, la distance Euclidienne et la distance de Manhattan (aussi appelée distance City-Block).

La distance de Hamming entre  $x'$  et  $x_i$  est calculée selon la formule suivante:

$$d(x', x_i) = \sum_{j=1}^D (x'_j - x_{ij})^2 \quad (3.4)$$

La distance Euclidienne entre  $x'$  et  $x_i$  est calculée selon la formule suivante:

$$d(x', x_i) = \sqrt{\sum_{j=1}^D (x'_j - x_{ij})^2} \quad (3.5)$$

La distance de Manhattan entre  $x'$  et  $x_i$  est calculée selon la formule suivante:

$$d(x', x_i) = \sum_{j=1}^D |x'_j - x_{ij}| \quad (3.6)$$

où  $x_{ij}$  est la  $j^{\text{eme}}$  caractéristique de  $x_i$ .

L'étiquette correspondante à  $x'$  peut être déterminée par:

$$y' = y_{i^*} \quad (3.7)$$

où

$$i^* = \arg_{i=1, \dots, S} \min d(x', x_i) \quad (3.8)$$

La métrique de distance dans SSFR implique le calcul de la différence (ou l'écart) entre l'échantillon de test et les échantillons d'apprentissage.

L'algorithme suivant résume notre méthode proposée. De plus, une illustration graphique résume l'approche proposée est présentée dans la Figure 3.4.

**Algorithme 1:** SSFR basé sur MB-C-BSIF et KNN

---

**Entrée:** Une image faciale X

1. Normaliser l'histogramme de l'image X
2. Appliquer un filtre médian sur l'image X
3. Diviser l'image X en trois composantes (rouge, vert, bleu):  $C^n, n = 1, 2, 3$
4. **Pour**  $n = 1$  jusqu'à 3 **faire**
5.     Diviser  $C^n$  en  $K$  blocs équivalents:  $C_k^n, k = 1, \dots, K$
6.     **Pour**  $k = 1$  jusqu'à  $K$  **faire**
7.         Appliquer le descripteur BSIF sur le bloc-composant  $C_k^n$ :  $H1_{(k)}^{(n)}$
8.     **Fin pour**
9.     Concaténer les caractéristiques MB-BSIF calculées de la composante  $C_n$ :  $H2^{(n)} = H1_{(1)}^{(n)} + H1_{(2)}^{(n)} + \dots + H1_{(K)}^{(n)}$
10. **Fin pour**
11. Concaténer les caractéristiques MB-C-BSIF calculées:  $H3 = H2^{(1)} + H2^{(2)} + H2^{(3)}$
12. Appliquer le classificateur K-NN en utilisant une distance de mesure appropriée

**Sortie:** Renvoyer la décision d'identification.

---

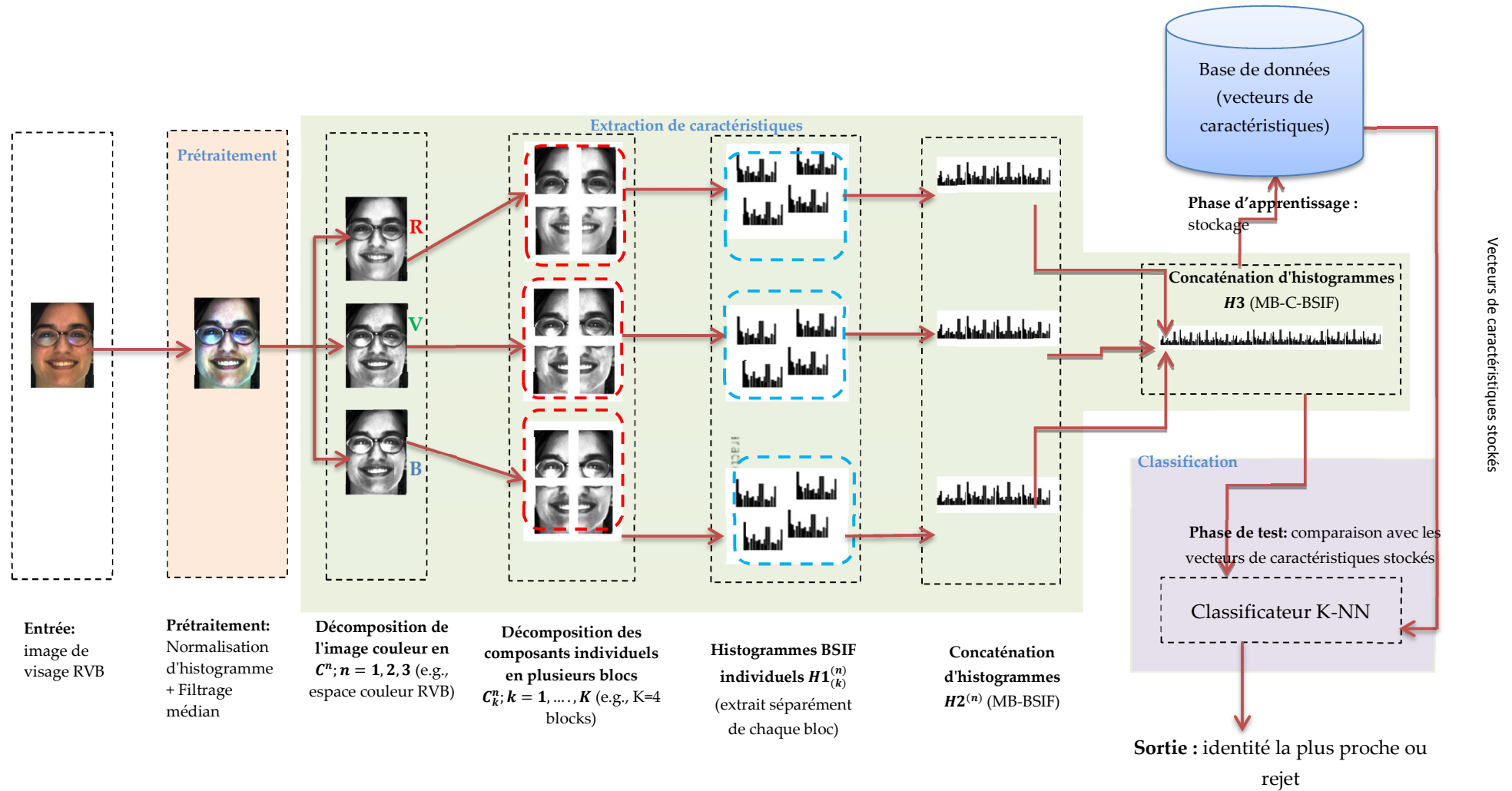


Figure 3.4. Schéma du système SSFR proposé, basé sur le descripteur MB-C-BSIF.

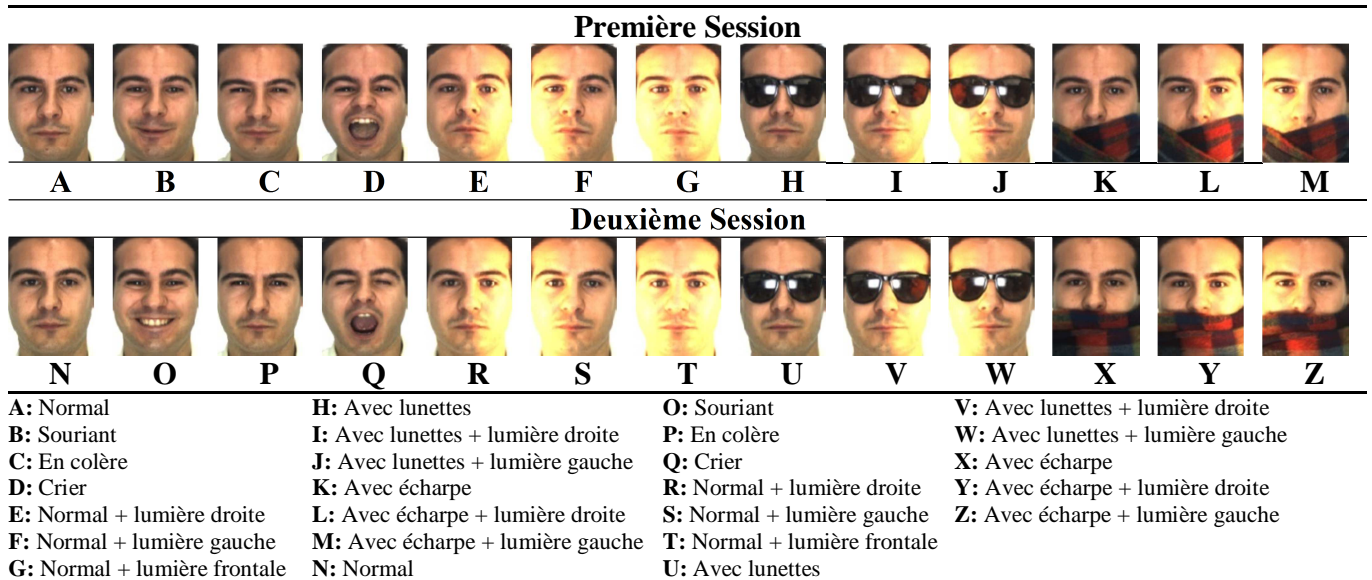
### 3.3. Analyse expérimentale

Notre système de reconnaissance faciale à partir d'un seul échantillon proposé a été évalué en utilisant deux bases de données, à savoir: Alex & Robert (AR) [72] et labeled faces in the wild (LFW) [73]. Dans cette section, nous décrivons les spécifications de chaque base de données et les configurations expérimentales utilisées pour l'évaluation du SSFR. Nous présentons également une analyse détaillée des résultats obtenus et effectuons une comparaison de nos performances avec celles de plusieurs méthodes de reconnaissance faciale de l'état de l'art.

#### 3.3.1. Expérimentations sur la base de données AR

##### 3.3.1.1. Description de la base de données

La base de données faciale d'Alex & Robert (AR) [72] est une base de données très utilisée dans les travaux de la reconnaissance faciale. Elle est composée de plus de 4000 images faciales en couleur de 126 individus, comprenant 56 femmes et 70 hommes. Chaque individu a 26 images différentes avec un visage frontal pris avec plusieurs expressions faciales, conditions d'éclairage, et occlusions. Les images ont été prises dans un intervalle de deux semaines et leur acquisition s'est faite en deux sessions (capture 1 et 2). Chaque session comprenait 13 images faciales par sujet.



**Figure 3.5.** Les 26 images faciales du premier individu de la base de données AR, accompagnées de leurs descriptions détaillées.

Dans notre étude, nous avons sélectionné un sous-ensemble des images faciales de 100 individus distincts, comprenant 50 hommes et 50 femmes, pour nos expérimentations ultérieures. La Figure 3.5 affiche les 26 images faciales du premier individu de la base de données AR, ainsi que des descriptions détaillées de celles-ci.

### 3.3.1.2. Configurations

L'objectif de cette étude vise à évaluer l'efficacité de la méthode MB-C-BSIF dans la présence des changements en expression faciale et occlusions. Pour cela, quatre ensembles de 100 classes ont été créés, chacun étant constitué de 50 hommes et 50 femmes. Le premier ensemble a été utilisé pour l'apprentissage est composé d'une seule image par individu (image A), représentant la première image de chaque personne. Les trois autres ensembles sont destinés pour les tests. Le premier ensemble de test contient 700 images qui sont divisées en sept sous-ensembles (sous-ensembles B, C, D, N, O, P, et Q) et comprenant chacun 100 images. Ces images présentent des visages avec diverses expressions faciales. Le deuxième ensemble de test est constitué de 600 images de personnes portant des lunettes de soleil, divisées en six sous-ensembles (sous-ensembles H, I, J, U, V, et W), comprenant chacun 100 images. Enfin, le troisième ensemble de test comprend 600 images de personnes portant des écharpes, réparties en six sous-ensembles (sous-ensembles K, L, M, X, Y, et Z), comprenant également chacun 100 images.

### 3.3.1.3. Expérimentations #1 (Effets des paramètres du descripteur BSIF)

Selon la section 3.2.2, l'opérateur BSIF dépend de deux paramètres: la taille du noyau de filtre  $l \times l$  et la longueur de la chaîne de bits  $n$ . Pour évaluer la méthode proposée, différentes configurations de l'opérateur BSIF ont été testées afin d'obtenir le meilleur taux de reconnaissance. Les images ont été converties en niveaux de gris, sans aucune segmentation en blocs non superposés (soit un bloc  $1 \times 1$ ), et la distance de Manhattan a été utilisée avec K-NN en classification. Les Tableaux 3.1 à 3.3 fournissent des détails complets et des comparaisons des résultats obtenus en utilisant les principales configurations de l'opérateur BSIF pour les sous-ensembles de variations en expression faciale, occlusion par des lunettes de soleil, et occlusion par des écharpes, respectivement. Les meilleurs résultats sont marqués en gras.

**Tableau 3.1.** Comparaison des résultats obtenus en utilisant six configurations de l'opérateur BSIF avec des changements dans l'expression faciale.

$l \times l$ (Pixels)	$n$ (Bits)	Taux de Reconnaissance (%)							Taux de Reconnaissance Moyen (%)
		B	C	D	N	O	P	Q	
3 × 3	5	70	72	38	36	20	24	14	39.14
5 × 5	9	94	97	59	75	60	66	30	68.71
9 × 9	12	100	100	91	95	90	92	53	88.71
11 × 11	8	97	99	74	85	70	75	43	77.57
15 × 15	12	100	100	96	97	96	96	73	94.00
17 × 17	12	100	100	98	97	96	97	71	<b>94.14</b>

**Tableau 3.2.** Comparaison des résultats obtenus avec six configurations de l'opérateur BSIF avec occlusion par lunettes de soleil.

$l \times l$ (Pixels)	$n$ (Bits)	Taux de Reconnaissance (%)						Taux de Reconnaissance Moyen (%)
		H	I	J	U	V	W	
3 × 3	5	29	8	4	12	4	3	10.00
5 × 5	9	70	24	14	28	14	8	26.50
9 × 9	12	98	80	61	80	38	30	61.50
11 × 11	8	78	34	23	48	26	15	37.33
15 × 15	12	100	84	85	87	50	46	75.33
17 × 17	12	100	91	87	89	58	46	<b>78.50</b>

**Tableau 3.3.** Comparaison des résultats obtenus avec six configurations de l'opérateur BSIF avec occlusion par écharpes.

$l \times l$ (Pixels)	$n$ (Bits)	Taux de Reconnaissance (%)						Taux de Reconnaissance Moyen (%)
		K	L	M	X	Y	Z	
3 × 3	5	7	4	2	3	2	2	3.33
5 × 5	9	22	9	6	12	6	2	9.50
9 × 9	12	88	54	34	52	31	15	45.67
11 × 11	8	52	12	90	22	9	7	32.00
15 × 15	12	97	69	64	79	48	37	65.67
17 × 17	12	98	80	63	90	48	31	<b>68.33</b>

Les résultats de l'expérimentation montrent que les meilleurs résultats en termes de taux de reconnaissance ont été obtenus avec les paramètres  $l \times l = 17 \times 17$  et  $n = 12$  de l'opérateur BSIF. Les performances d'identification augmentent également avec une

augmentation de la valeur de  $l$  ou de  $n$ . Bien que cette configuration ait donné de bons résultats pour les changements en expression faciale avec les sept sous-ensembles, elle a obtenu un taux de reconnaissance assez faible (71 %) pour le sous-ensemble Q, qui présente une variation importante de l'expression faciale. Par ailleurs, cette configuration n'a pas donné de résultats satisfaisants dans les conditions d'occlusion par des objets, notamment avec l'occlusion par écharpe, et nécessite donc d'autres améliorations.

### 3.3.1.4. Expérimentation #2 (Effets de la distance)

Dans cette expérimentation, nous avons évalué les performances de la dernière bonne configuration (qui utilise des images en niveaux de gris, des blocs de taille  $l \times l = 17 \times 17$  et  $n = 12$ ) en évaluant cette fois différentes mesures de distance associées au classificateur K-NN. En effet, les résultats obtenus avec la distance de Manhattan ont été comparés avec ceux obtenus par d'autres mesures de distance bien connues pour la reconnaissance faciale. Les trois sous-ensembles de test ont été pris en considération pour cette expérimentation, à savoir: les variations en expression faciale, l'occlusion par des lunettes de soleil, et l'occlusion par des écharpes. Les résultats les plus performants sont indiqués en gras dans les Tableaux 3.4 à 3.6.

**Tableau 3.4.** Comparaison des résultats obtenus par différentes mesures de distance avec des changements en expression faciale.

Distance	Taux de Reconnaissance (%)							Taux de Reconnaissance Moyen (%)
	B	C	D	N	O	P	Q	
Hamming	63	79	9	69	23	40	6	41.29
Euclidienne	99	100	80	90	83	82	43	82.43
Manhattan	100	100	98	97	96	97	71	<b>94.14</b>

**Tableau 3.5.** Comparaison des résultats obtenus par différentes mesures de distance avec occlusion par lunettes de soleil.

Distance	Taux de Reconnaissance (%)						Taux de Reconnaissance Moyen (%)
	H	I	J	U	V	W	
Hamming	37	5	6	11	4	2	10.83
Euclidienne	96	68	42	68	31	17	53.67
Manhattan	100	91	87	89	58	46	<b>78.50</b>



**Tableau 3.6.** Comparaison des résultats obtenus par différentes mesures de distance avec occlusion par écharpes.

Distance	Taux de Reconnaissance (%)						Taux de Reconnaissance Moyen (%)
	K	L	M	X	Y	Z	
Hamming	34	5	8	20	4	4	12.50
Euclidienne	79	32	16	41	22	5	32.50
Manhattan	98	80	63	90	48	31	<b>68.33</b>

Il a été observé que la distance de Manhattan a produit les performances de reconnaissance les plus fiables par rapport aux autres mesures de distance analysées dans cette étude, telles que les distances de Hamming et Euclidienne. Par conséquent, nous pouvons conclure que la distance de Manhattan est la plus appropriée pour notre méthode de reconnaissance faciale. En d'autres termes, l'utilisation de la distance de Manhattan a conduit à une performance de reconnaissance plus fiable que les autres distances évaluées dans cette étude.

### 3.3.1.5. Expérimentation #3 (Effets de la segmentation d'image)

L'objectif de cette expérience était d'améliorer le taux de la reconnaissance, en particulier dans des conditions d'occlusion en proposant de segmenter l'image en plusieurs blocs non superposés. Nous avons testé les performances d'identification en utilisant les caractéristiques MB-BSIF de manière segmentée plutôt que globale sur l'image entière, comme expliqué dans la section 3.2.2.

Dans cette étude, nous avons comparé trois méthodes de segmentation d'image en divisant chaque image originale en blocs de tailles différentes:  $1 \times 1$  pour représenter les informations globales,  $2 \times 2$  et  $4 \times 4$  pour représenter les informations locales. Cela a donné respectivement 1 bloc pour l'image originale, 4 blocs et 16 blocs. Pour les deux derniers cas, nous avons fusionné les vecteurs de caractéristiques (histogrammes de type  $H1$ ) dérivés de chaque bloc pour créer un vecteur de caractéristiques global (Histogramme de type  $H2$ ) représentant l'image entière.

Les Tableaux 3.7 à 3.9 présentent et comparent les taux de reconnaissance du MB-BSIF testés pour divers blocs avec des sous-ensembles de variations en expression faciale, des sous-ensembles d'occlusion pour les lunettes de soleil et des sous-ensembles d'occlusion pour les écharpes, respectivement (avec des images en niveaux de gris, la distance Manhattan,  $l \times l = 17 \times 17$ , et  $n = 12$ ). Les meilleurs résultats sont indiqués en gras.

**Tableau 3.7.** Comparaison des résultats obtenus à l'aide de différents blocs divisés avec des changements en expression faciale.

Segmentation	Taux de Reconnaissance (%)							Taux de Reconnaissance Moyen (%)
	B	C	D	N	O	P	Q	
(1 × 1)	100	100	98	97	96	97	71	94.14
(2 × 2)	100	100	95	98	92	91	60	90.86
(4 × 4)	100	100	99	98	92	97	76	<b><u>94.57</u></b>

**Tableau 3.8.** Comparaison des résultats obtenus à l'aide de différents blocs divisés avec occlusion par lunettes de soleil.

Segmentation	Taux de Reconnaissance (%)						Taux de Reconnaissance Moyen (%)
	H	I	J	U	V	W	
(1 × 1)	100	91	87	89	58	46	78.50
(2 × 2)	100	99	98	91	83	71	90.33
(4 × 4)	100	99	99	93	81	79	<b><u>91.83</u></b>

**Tableau 3.9.** Comparaison des résultats obtenus à l'aide de différents blocs divisés avec occlusion par écharpes.

Segmentation	Taux de Reconnaissance (%)						Taux de Reconnaissance Moyen (%)
	K	L	M	X	Y	Z	
(1 × 1)	98	80	63	90	48	31	68.33
(2 × 2)	98	95	92	92	79	72	88.00
(4 × 4)	99	98	95	93	84	77	<b><u>91.00</u></b>

Les résultats de cette expérimentation montrent que la segmentation d'image en blocs améliore légèrement le taux de reconnaissance pour les sous-ensembles de variation en expression faciale, bien que les performances étaient déjà raisonnables avec l'approche globale. Cependant, le sous-ensemble Q, qui se caractérise par des changements importants dans l'expression faciale, a bénéficié d'avantage de la segmentation d'image, avec un taux de reconnaissance amélioré de 71% à 76%.

Les résultats ont montré une augmentation significative dans les taux de reconnaissance pour tous les sous-ensembles occlus lorsque le nombre de blocs augmentait. Par exemple, en utilisant de 1 à 16 blocs, le taux est passé de 31% à 71% pour le sous-ensemble Z, de 46% à 79% pour le sous-ensemble W, et de 48% à 84% pour le sous-ensemble Y.

L'ajout d'informations locales est une stratégie importante pour améliorer la précision de la reconnaissance faciale en cas d'occlusion partielle. Cela permet d'obtenir des informations plus détaillées sur la structure du visage et la relation des positions entre les différentes composantes du visage, ce qui est essentiel pour une reconnaissance précise et fiable.

Les résultats montrent que les blocs de  $4 \times 4$  ont fourni la configuration la plus performante en termes de taux de reconnaissance pour les sous-ensembles d'expression faciale, d'occlusion par des lunettes de soleil et d'occlusion par foulards. En d'autres termes, utiliser des blocs de  $4 \times 4$  pour extraire les caractéristiques du visage a permis d'obtenir les meilleurs résultats pour la reconnaissance de ces sous-ensembles spécifiques, en comparaison avec d'autres configurations de blocs.

#### 3.3.1.6. *Expérimentation #4 (Effets des informations de la texture couleur)*

Nous avons procédé à une évaluation des performances de la dernière bonne configuration (c'est-à-dire la segmentation de l'image en blocs de  $4 \times 4$ , K-NN avec la distance de Manhattan,  $l \times l = 17 \times 17$ , et  $n = 12$ ). Dans le cadre de cette évaluation, nous avons testé trois espaces colorimétriques différents: RVB, HSV et YCbCr, plutôt que de convertir l'image en niveaux de gris. La méthode d'extraction de caractéristiques utilisée est appelée MB-C-BSIF, telle que décrite dans la section 3.2.2. Comme les images de la base de données AR sont déjà en format RVB, il n'est donc pas nécessaire de procéder à une transformation du premier espace couleur. Cependant, pour les autres espaces colorimétriques, il est nécessaire de convertir les images de RVB en HSV et en YCbCr. Les Tableaux 3.10 à 3.12 comparent les taux de reconnaissance du MB-C-BSIF en utilisant plusieurs espaces colorimétriques pour les différents sous-ensembles de variations en expression faciale, d'occlusion par des lunettes de soleil et d'occlusion par des écharpes, respectivement. Les meilleurs résultats obtenus sont mis en évidence en gras.

**Tableau 3.10.** Comparaison des résultats obtenus en utilisant différents espaces colorimétriques avec des changements dans l'expression faciale.

Espace colorimétrique	Taux de Reconnaissance (%)							Taux de Reconnaissance Moyen (%)
	B	C	D	N	O	P	Q	
Niveau de gris	100	100	99	98	92	97	76	94.57
RVB	100	100	95	97	92	93	67	92.00
HSV	100	100	99	97	96	95	77	<b><u>94.86</u></b>
YCbCr	100	100	96	98	93	93	73	93.29

**Tableau 3.11.** Comparaison des résultats obtenus en utilisant différents espaces colorimétriques avec occlusion par lunettes de soleil.

Espace colorimétrique	Taux de Reconnaissance (%)						Taux de Reconnaissance Moyen (%)
	H	I	J	U	V	W	
Niveau de gris	100	99	99	93	81	79	91.83
RVB	100	99	100	93	85	84	<b><u>93.50</u></b>
HSV	100	97	99	96	82	80	92.33
YCbCr	100	99	98	93	81	80	91.83

**Tableau 3.12.** Comparaison des résultats obtenus en utilisant différents espaces colorimétriques avec occlusion par écharpes.

Espace colorimétrique	Taux de Reconnaissance (%)						Taux de Reconnaissance Moyen (%)
	K	L	M	X	Y	Z	
Niveau de gris	99	98	95	93	84	77	91.00
RVB	99	97	97	94	88	81	<b><u>92.67</u></b>
HSV	99	96	90	95	75	74	88.17
YCbCr	98	98	96	93	87	78	91.67

En analysant les résultats obtenues, il est possible de constater que:

- Les résultats pour les sous-ensembles de variations en expression faciale sont presque identiques pour tous les espaces colorimétriques testés. En effet, une légère amélioration est observée avec l'espace colorimétrique HSV, tandis que de légères dégradations sont notées avec les espaces colorimétriques RVB et YCbCr.
- Les performances de la reconnaissance ont été améliorées pour tous les espaces colorimétriques par rapport à la norme en niveaux de gris pour les sous-ensembles

d'occlusion par des lunettes de soleil. L'espace colorimétrique RVB présente le meilleur rendement avec un taux de reconnaissance moyen augmentant de 91.83 % à 93.50 %.

- Pour les images avec une écharpe, l'espace colorimétrique HSV a montré une certaine régression tandis que les espaces RVB et YCbCr ont présenté des améliorations par rapport à la norme en niveaux de gris. L'espace RVB a donné la précision la plus élevée.
- La remarque la plus importante est que l'espace colorimétrique RVB a montré une amélioration significative des performances pour les sous-ensembles V, W, Y et Z (de 81 % à 85 % pour V, de 79 % à 84 % pour W, de 84 % à 88 % pour Y et de 77 % à 87 % pour Z).
- Enfin, nous notons que l'espace colorimétrique optimal, offrant un équilibre parfait entre la restauration de l'éclairage et l'amélioration de l'identification, était le RVB.

#### 3.3.1.7. Comparaison #1 (Protocole I)

Afin de confirmer que notre méthode proposée produit une performance de reconnaissance supérieure avec des variations d'expression faciale, nous avons comparé nos résultats avec plusieurs méthodes récemment utilisée pour résoudre le problème de SSFR. Le Tableau 3.13 présente les taux les plus élevées obtenues en utilisant les mêmes sous-ensembles et le même protocole d'évaluation : le sous-ensemble A comme ensemble d'apprentissage et les sous-ensembles de variations d'expression faciale B, C, D, N, O, et P constituant l'ensemble de test. Les résultats présentés dans le Tableau 3.13 sont extraits de plusieurs références [46, 49, 74, 75]. Les " - - " indiquent que la méthode considérée n'a pas de résultats expérimentaux. Les meilleurs résultats sont en gras.

**Tableau 3.13.** Comparaison avec 18 méthodes de l'état de l'art en utilisant des sous-ensembles de variations en expression faciale.

Auteur	Année	Méthode	Précision						Taux de Reconnaissance Moyen (%)
			B	C	D	N	O	P	
Turk, Pentland [76]	1991	PCA	97.00	87.00	60.00	77.00	76.00	67.00	77.33
Wu and Zhou [77]	2002	(PC) <sup>2</sup> A	97.00	87.00	62.00	77.00	74.00	67.00	77.33
Chen et al. [78]	2004	E(PC) <sup>2</sup> A	97.00	87.00	63.00	77.00	75.00	68.00	77.83
Yang et al. [79]	2004	2DPCA	97.00	87.00	60.00	76.00	76.00	67.00	77.17
Gottumukkal and Asari [80]	2004	Block-PCA	97.00	87.00	60.00	77.00	76.00	67.00	77.33
Chen et al. [81]	2004	Block-LDA	85.00	79.00	29.00	73.00	59.00	59.00	64.00
Zhang and Zhou [82]	2005	(2D) <sup>2</sup> PCA	98.00	89.00	60.00	71.00	76.00	66.00	76.70
Tan et al. [83]	2005	SOM	98.00	88.00	64.00	73.00	77.00	70.00	78.30
He et al. [84]	2005	LPP	94.00	87.00	36.00	86.00	74.00	78.00	75.83
Zhang et al. [85]	2005	SVD-LDA	73.00	75.00	29.00	75.00	56.00	58.00	61.00
Deng et al. [86]	2010	UP	98.00	88.00	59.00	77.00	74.00	66.00	77.00
Lu et al. [46]	2012	DMMA	99.00	93.00	69.00	88.00	85.00	85.50	79.00
Mehrasa et al. [74]	2017	SLPMM	99.00	94.00	65.00	--	--	--	--
Ji et al. [75]	2017	CPL	92.22	88.06	83.61	83.59	77.95	72.82	83.04
Zhang et al. [47]	2018	DMF	100.00	99.00	66.00	--	--	--	--
Chu et al. [87]	2019	MFSA+	100.00	100.00	74.00	93.00	85.00	86.00	89.66
Pang et al. [88]	2019	RHDA	97.08	97.00	96.25	--	--	--	--
Zhang et al. [49]	2021	DNNC	100.00	98.00	69.00	92.00	76.00	85.00	86.67
<b>Notre méthode</b>	<b>2021</b>	<b>MB-C-BSIF</b>	<b>100.00</b>	<b>100.00</b>	<b>95.00</b>	<b>97.00</b>	<b>92.00</b>	<b>93.00</b>	<b><u>96.17</u></b>

Les résultats obtenus dans l'étude confirment la fiabilité et la robustesse de la méthode proposée par rapport aux méthodes d'état de l'art pour l'identification en présence des changements en expression faciale. Les performances de la technique proposée MB-C-BSIF ont été évaluées avec six sous-ensembles différents, et les résultats montrent une précision élevée pour l'identification avec des variations en expression faciale pour chacun de ces sous-ensembles.

En particulier, la technique proposée a atteint un niveau de précision de 100% pour les sous-ensembles B et C, 95% pour le sous-ensemble D, 97% pour le sous-ensemble N, 92% pour le sous-ensemble O et 93% pour le sous-ensemble P. Ces résultats sont supérieurs à ceux obtenus avec les méthodes analysées dans cette étude.

### 3.3.1.8. Comparaison #2 (Protocole II)

Pour démontrer encore davantage l'efficacité de notre système SSFR proposé, nous avons effectué une comparaison avec des travaux récemment publiés dans des conditions non contraintes en utilisant la meilleure configuration du MB-C-BSIF. Cette configuration consiste en un espace colorimétrique RVB, une segmentation de l'image en blocs de  $4 \times 4$ , une distance de Manhattan, une taille de filtre  $l \times l = 17 \times 17$  et un nombre de bits  $n = 12$ . Nous avons suivi le même protocole expérimental que celui décrit dans les travaux [43, 49]. Les taux de reconnaissance des travaux comparés sur les sous-ensembles H + K (qui impliquent une occlusion par des lunettes de soleil et une écharpe) et les sous-ensembles J + M (qui impliquent une occlusion par des lunettes de soleil et une écharpe avec des variations d'éclairage) sont présentés dans le Tableau 3.14. Les meilleurs résultats sont indiqués en gras.

**Tableau 3.14.** Comparaison avec 12 méthodes en utilisant des sous-ensembles d'occlusion et éclairage + occlusion.

Auteur	Année	Méthode	Occlusion (H + K) (%)	Eclairage + Occlusion (J + M) (%)	Taux de Reconnaissance Moyen (%)
Zhang et al. [45]	2011	CRC	58.10	23.80	40.95
Deng et al. [41]	2012	ESRC	83.10	68.60	75.85
Zhu et al. [44]	2012	PCRC	95.60	81.30	88.45
Yang et al. [42]	2013	SVDL	86.30	79.40	82.85
Lu et al. [46]	2012	DMMA	46.90	30.90	38.90
Zhu et al. [43]	2014	LGR	98.80	96.30	97.55
Ref. [89]	2016	SeetaFace	63.13	55.63	59.39
Zeng et al. [60]	2017	DCNN	96.5	88.3	92.20
Chu et al. [87]	2019	MFSA+	91.3	79.00	85.20
Cuculo et al. [90]	2019	SSLD	90.18	82.02	86.10
Zhang et al. [49]	2021	DNNC	92.50	79.50	86.00
Du and Da [63]	2020	BDL	93.03	91.55	92.29
<b>Notre méthode</b>	<b>2021</b>	<b>MB-C-BSIF</b>	<b><u>99.5</u></b>	<b><u>98.5</u></b>	<b><u>99.00</u></b>

Selon les résultats du Tableau 3.14, la méthode LGR proposée par Zhu et al. [43] et la méthode MB-C-BSIF ont des performances élevées. Toutefois, le taux d'identification de la

méthode MB-C-BSIF a été significativement supérieur à toutes les autres méthodes envisagées pour les deux sessions de tests.

En comparaison avec les SSFR apparentés telles que les méthodes d'apprentissage génériques (par exemple, ESRC [41], SVDL [42] et LGR [43]), les méthodes de partitionnement d'image (par exemple, CRC [45], PCRC [44] et DNNC [49]) ou encore les méthodes d'apprentissage en profondeur (par exemple, DCNN [60] et BDL [63]), notre méthode se distingue par sa capacité à exploiter différentes formes d'information pour l'identification de personnes. En effet:

- Le descripteur BSIF analyse l'image pixel par pixel, en prenant en compte les informations locales.
- L'image est divisée en plusieurs blocs pour exploiter les informations régionales.
- Les occurrences du descripteur BSIF sont ensuite accumulées dans un histogramme global pour traiter l'information globale.
- Notre méthode utilise le MB-BSIF sur tous les composants d'image RVB pour exploiter les informations de texture couleur.

Ainsi, notre méthode est capable de combiner des informations locales, régionales, globales et de texture couleur pour améliorer la précision d'identification des personnes. Cette capacité de notre méthode est différente des SSFR apparentés, qui utilisent généralement des méthodes d'apprentissage génériques, des méthodes de partitionnement d'image ou des méthodes d'apprentissage en profondeur pour extraire des informations à partir des images.

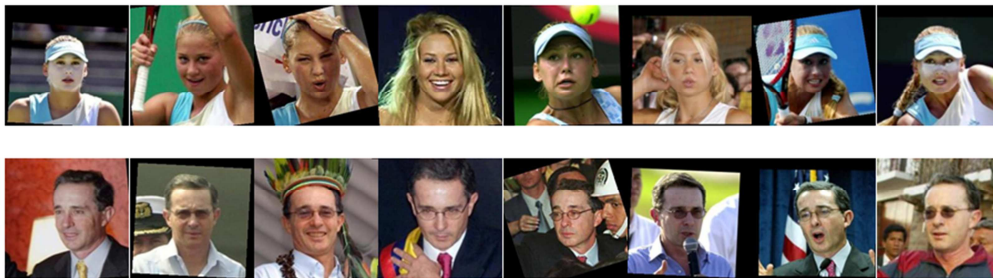
### **3.3.2. Expérimentations sur la base de données LFW**

#### *3.3.2.1. Description de la base de données*

La base de données labeled faces in the wild (LFW) est une collection de plus de 13000 images faciales de 5749 individus. Les images ont été collectées sur le World Wide Web dans des situations difficiles, avec 1680 individus ayant deux photos ou plus. Les tests effectués ont utilisé la variante LFW-a, où les images faciales sont alignées à l'aide d'un outil de normalisation commercial. Les différences intra-classes dans cette base de données sont très élevées par rapport aux bases de données contraintes bien connues et une normalisation des visages a été effectuée.



Chaque image a une taille de  $250 \times 250$  pixels et utilise l'extension jpeg. LFW est une base de données très difficile car elle vise à étudier les problèmes sans contraintes de la reconnaissance faciale, tels que les changements d'éclairage, l'âge, les vêtements, la mise au point, l'expression faciale, la saturation des couleurs, la posture, la race, la coiffure, l'arrière-plan, la qualité de l'appareil photo, le sexe, l'origine ethnique et d'autres facteurs. La Figure 3.6 présente ces différents facteurs.



**Figure 3.6.** Exemples de deux sujets différents de la base de données Labeled Faces in the Wild (LFW)-a.

### 3.3.2.2. Protocole expérimental

Le protocole expérimental suivi dans cette étude est basé sur celui présenté dans les articles [44, 42, 40]. Pour la base de données LFW-a, seuls les sujets ayant plus de 10 images ont été sélectionnés, ce qui a donné un sous-ensemble de 158 individus. Chaque image a été recadrée à une taille de  $120 \times 120$  pixels, puis redimensionnée à  $80 \times 80$  pixels. Les photographies faciales des 50 premiers sujets ont été utilisées pour créer l'ensemble d'entraînement et l'ensemble de test. Pour l'ensemble d'apprentissage, une image aléatoire a été sélectionnée pour chaque sujet, tandis que les images restantes ont été utilisées dans l'ensemble de test. Ce processus a été répété cinq fois avec différentes permutations, et la moyenne des résultats pour chaque permutation a été prise en compte.

### 3.3.2.3. Limite des systèmes SSFR

L'objectif de cette section est d'évaluer les limites de notre approche. Nous avons comparé les performances de notre méthode avec celles de plusieurs approches concurrentes en termes de taux d'identification. Notre meilleure configuration a utilisé le descripteur BSIF avec une taille de filtre  $l \times l = 17 \times 17$  et une longueur de chaîne de bits  $n = 12$ , un classificateur K-NN associé avec la distance de Manhattan, une segmentation de l'image en blocs de  $4 \times 4$  pixels, ainsi qu'un espace de couleur RVB.

**Tableau 3.15.** Taux d'identification utilisant la base de données LFW.

Auteur	Année	Méthode	Taux d'identification (%)
Chen et al. [81]	2004	Block LDA	16.40
Zhang et al. [85]	2005	SVD-FLDA	15.50
Wright et al. [91]	2009	SRC	20.40
Su et al. [92]	2010	AGL	19.20
Zhang et al. [45]	2011	CRC	19.80
Deng et al. [41]	2012	ESRC	27.30
Zhu et al. [44]	2012	PCRC	24.20
Yang et al. [42]	2013	SVDL	28.60
Lu et al. [46]	2012	DMMA	17.80
Zhu et al. [43]	2014	LGR	30.40
Ji et al. [75]	2017	CPL	25.20
Dong et al. [40]	2018	KNNMMDL	32.30
Chu et al. [87]	2019	MFSa+	26.23
Pang et al. [88]	2019	RHDA	32.89
Zhou et al. [93]	2019	DpLSA	37.55
<b>Notre méthode</b>	<b>2021</b>	<b>MB-C-BSIF</b>	<b>38.01</b>
Zeng et al. [94]	2018	TDL	<b><u>74.00</u></b>

Nous pouvons observer que les approches traditionnelles n'ont pas atteint des niveaux de précision d'identification élevés. Cela est principalement dû au fait que les images de la base de données LFW ont été prises dans des conditions non-contrôlées, ce qui génère des images faciales avec des différences intra-classes importantes, et augmente la complexité de la reconnaissance faciale. En conséquence, l'efficacité de SSFR est réduite. Cependant, notre solution proposée est meilleure par rapport aux autres approches traditionnelles concurrentes en raison de l'exploitation de différentes formes d'informations ce qui permettent d'améliorer les performances du SSFR. Elle intègre des informations locales, régionales, globales et de texture couleur ce qui représente une avancée significative par rapport aux solutions concurrentes, qui ne prennent généralement pas en compte autant de formes d'informations simultanément.

Dans une étude menée par Zeng et al. [94], les auteurs ont combiné des caractéristiques traditionnelles et d'apprentissage en profondeur (TDL) pour surmonter les limitations de chaque classe. Ils ont atteint un taux d'identification d'environ 74%, ce qui représente une avancée significative dans ce domaine difficile.

Notre étude comparative présentée dans [12] met en évidence que les systèmes actuels de reconnaissance faciale, qui utilisent plusieurs exemples dans l'ensemble d'apprentissage, obtiennent une très grande précision avec la base de données LFW, en particulier avec les méthodes basées sur l'apprentissage en profondeur. Cependant, les systèmes SSFR ont des difficultés considérables lors de l'utilisation de la base de données complexe LFW, et des recherches supplémentaires sont nécessaires pour améliorer leur fiabilité dans cette situation. En outre, lorsque l'étape d'apprentissage est basée sur des millions d'images, la technique SSFR proposée ne peut pas être utilisée. Dans ce cas, la référence [94], qui utilise respectivement des techniques d'apprentissage profond avec augmentation de données et des caractéristiques d'apprentissage profond combinées à des caractéristiques traditionnelles, permettent d'obtenir une meilleure précision.

### **3.4. Conclusion**

Dans ce chapitre, nous avons introduit une méthode de reconnaissance faciale à échantillon unique (SSFR) basée sur le descripteur MB-C-BSIF (Multi-Block Color-Binarized Statistical Image Features). Cette méthode permet l'extraction de caractéristiques en utilisant différents types d'informations, y compris des informations locales, régionales, globales et de texture couleur. Dans nos expérimentations, le MB-C-BSIF a été évalué sur plusieurs sous-ensembles d'images des bases de données sans contraintes, à savoir: AR et LFW. Des expérimentations menées sur la base de données AR ont montré que notre méthode améliore significativement les performances de la classification SSFR face à plusieurs variations de reconnaissance faciale. La stratégie d'extraction de caractéristiques proposée atteint une grande précision, avec un taux moyen variant entre 96.17 % et 99 % pour la base de données AR avec les protocoles I et II, respectivement. Ces résultats valident l'efficacité de la méthode proposée par rapport aux méthodes de l'état de l'art.

Dans le prochain chapitre de notre recherche, nous explorerons une approche hybride qui combine des méthodes traditionnelles et des méthodes profondes, en combinant notre descripteur

MB-C-BSIF avec des réseaux de neurones convolutifs (CNNs) pour la reconnaissance faciale à échantillon unique (SSFR).

## **CHAPITRE 4: Combinaison de Caractéristiques de Texture et d'Apprentissage Profond pour la Reconnaissance Faciale à partir d'un Seul Echantillon**

### **4.1. Introduction**

Ce chapitre est composé de deux parties. La première partie se concentre sur la comparaison de plusieurs modèles SSFR construits à l'aide de descripteurs de texture et de caractéristiques basées sur l'apprentissage profond. La deuxième partie présente le principe de notre méthode hybride qui combine le descripteur de texture identifié comme le plus performant avec les CNNs les plus performants, tels qu'identifiés dans l'étude comparative réalisée dans la première partie. Ensuite, nous évaluons sa performance et comparons ses résultats avec ceux d'autres méthodes en littérature.

### **4.2. Etude comparative entre les caractéristiques basées sur la texture et celles basées sur l'apprentissage profond**

Dans la première partie de ce chapitre, nous testons et comparons la performance de différentes méthodes SSFR. Nous explorons les caractéristiques de certaines méthodes de texture les plus performantes ainsi que celles de certaines architectures CNN pré-entraînées pour représenter l'image du visage sous forme d'un vecteur de caractéristiques qui quantifie le contenu de l'image faciale.

#### **4.2.1. Méthodes**

Dans cette section, nous détaillons les matériels, outils et méthodes utilisés pour mettre en œuvre les méthodes SSFR et menons une étude comparative.

##### *4.2.1.1. Augmentation des données et prétraitement*

La technique d'augmentation des données est un excellent moyen de compenser le problème de SSFR. L'augmentation des données consiste à générer de nouveaux échantillons à partir des échantillons existants en appliquant des transformations simples à ces échantillons. Le concept principal derrière l'augmentation des données est d'élargir l'ensemble de données d'apprentissage en créant des variations des échantillons existants, ce qui permet au modèle d'apprentissage de généraliser mieux et d'éviter le sur-ajustement. En générant de nouveaux

échantillons, on peut introduire une plus grande diversité dans l'ensemble de données, ce qui aide le modèle à apprendre des caractéristiques plus robustes et à mieux s'adapter à des scénarios variés [95]. Dans notre travail, nous avons utilisé des méthodes courantes d'augmentation des données, telles que la rotation d'image et le zoom. Avant de soumettre les images augmentées aux extracteurs de caractéristiques, il est essentiel d'effectuer une étape de prétraitement afin d'améliorer la qualité graphique des images. Dans notre approche, nous avons utilisé la technique de *contrast limited adaptive histogram equalization* (CLAHE) [96] pour le traitement des images faciales. Le CLAHE effectue une égalisation adaptative de l'histogramme en limitant la hauteur de l'histogramme local. Cette technique garantit que le contraste de l'image est amélioré tout en limitant l'amplification du bruit et la sur-amélioration du contraste local. En d'autres termes, le CLAHE permet de renforcer le contraste des régions locales de l'image tout en préservant la qualité globale de l'image.

#### 4.2.1.2. *Extraction de caractéristiques*

L'extraction de caractéristiques constitue une phase cruciale dans la construction de tout système de reconnaissance faciale. Son objectif est d'extraire les informations pertinentes qui caractérisent chaque image faciale. Dans cette sous-section, nous décrivons les méthodes d'extraction de caractéristiques mises en œuvre dans ce travail, qui se divisent en deux principales catégories en fonction des caractéristiques extraites : la texture et l'apprentissage profond.

##### **A. Méthodes basées sur la texture**

- **Local binary patterns (LBP) [69]:** LBP est un descripteur de texture largement utilisé pour l'extraction de caractéristiques à partir d'une image. Le descripteur LBP compare les intensités de pixel entre le pixel central et ses huit pixels environnants, en utilisant des comparaisons binaires. Cela permet de capturer les motifs locaux de la texture.
- **Local phase quantization (LPQ) [70]:** LPQ est un descripteur utilisé pour la quantification de la phase dans des voisinages locaux d'une image. Plus précisément, il opère en appliquant la transformée de Fourier aux zones locales de l'image pour extraire les informations de phase. Ces informations sont ensuite quantifiées, c'est-à-dire discrétisées en valeurs numériques. LPQ a la particularité d'être invariant au flou dans

certaines conditions, ce qui signifie que les informations de phase restent relativement stables même lorsque l'image est floue.

- **Binarized statistical image features (BSIF) [65]:** BSIF est une méthode proposée pour la reconnaissance faciale et la classification de textures qui s'appuie sur les concepts de LBP et LPQ. Contrairement à LBP et LPQ, qui utilisent des filtres conçus manuellement, BSIF introduit l'idée d'apprentissage automatique d'un ensemble fixe de filtres à partir d'un petit ensemble d'images naturelles. Ce processus d'apprentissage permet l'extraction de représentations statistiquement significatives des images, ce qui évite la nécessité d'un réglage manuel. Pour plus de détails sur le descripteur BSIF, vous pouvez vous référer à la section 2 du chapitre 3.
- **Histograms of oriented gradients (HOG) [97]:** Le descripteur HOG est l'une des meilleures méthodes utilisées pour analyser la texture locale des images, et il constitue l'une des meilleures caractéristiques pour l'information sur la forme et les contours. Le descripteur HOG peut décrire la forme du visage en analysant la répartition des directions de contour ou des gradients d'intensité lumineuse. Le processus de cette technique consiste à diviser l'image du visage en cellules (de petites régions ou zones) et à générer un histogramme des directions de contour ou des gradients de chaque cellule. Finalement, les histogrammes de toutes les cellules sont combinés pour extraire les caractéristiques de l'image du visage. Cette approche a connu un grand succès dans la reconnaissance faciale.

Pour chaque descripteur de texture, les concepts d'information multi-blocs (MB) et de couleur (C) ont été exploités, car ils fournissent d'excellents outils pour améliorer les taux de reconnaissance. Pour plus de détails sur le concept MB-C, les lecteurs peuvent se référer aux références [98-100]. Le Tableau 4.1. récapitule les configurations utilisées par chaque descripteur de texture dans nos différentes expériences.

**Tableau 4.1.** Récapitulatif des configurations utilisées par chaque opérateur de texture.

Descripteur	Configuration
<b>MB-C-LBP</b>	Rayon : 2 pixels, Voisinage : 8
<b>MB-C-LPQ</b>	Taille de la fenêtre : 5x5 pixels
<b>MB-C-BSIF</b>	Longueur de la chaîne de bits : 12, Taille du filtre : 17x17 pixels
<b>MB-C-HOG</b>	Orientations : 9, Taille du bloc : 16x16 pixels

## B. Méthodes basées sur l'apprentissage profond

Les CNNs typiques sont composés de trois types de couches: les couches de convolution, de pooling et fully connected. La couche de convolution a pour fonction d'extraire différentes représentations de caractéristiques à partir de l'image en entrée. Les couches de pooling réduisent considérablement le nombre de paramètres, diminuent le surajustement, économisent des ressources computationnelles et améliorent les performances du modèle. Les couches fully connected ont pour rôle principal d'intégrer les caractéristiques des couches de convolution précédentes [101]. Nous décrivons ici les différentes stratégies d'apprentissage utilisées pour extraire des caractéristiques faciales discriminantes à partir d'images en utilisant quatre architectures de CNN profonds (certaines avec plusieurs variantes) : les réseaux du groupe de géométrie visuelle (VGGNet-16 et VGGNet-19), les réseaux résiduels (ResNet-50, ResNet-101 et ResNet-152), les réseaux convolutionnels densément connectés (DenseNet-121) et Inception-V3. Nous décrivons par la suite chacun de ces CNN pré-entraînés.

- **VGGNet [58]:** C'est une architecture de réseaux neuronaux convolutifs (CNN) qui utilise des filtres de convolution de petite taille ( $3 \times 3$ ) et double le nombre de cartes de caractéristiques après chaque couche de pooling ( $2 \times 2$ ). Il a été préalablement entraîné sur la base de données ImageNet, qui comprend plus de 14 millions d'images classées en 1000 diverses catégories. VGGNet reste l'une des architectures les plus populaires pour la reconnaissance d'images. Nous avons adapté la profondeur de VGGNet en utilisant des configurations de poids de 16 et 19 couches, ce qui améliore la flexibilité de l'architecture profonde. Cela permet une extraction plus efficace des caractéristiques à différentes échelles. En exploitant les couches profondes de VGGNet, nous avons pu bénéficier de sa capacité à extraire des caractéristiques de manière plus précise et discriminante.
  - **VGG-16:** Ce modèle tire son nom du fait qu'il comporte 16 couches de poids; 13 couches de convolution et 3 couches de fully-connected.
  - **VGG-19:** L'architecture de VGG-19 se compose de 19 couches; 16 couches de convolution et 3 couches fully-connected.

La taille d'entrée de ces modèles est de  $224 * 224 * 3$  pour les images RVB. Le réseau comprend cinq couches de max-pooling, la couche Softmax servant de dernière couche de l'architecture.



- **ResNet [59]:** Cette architecture a introduit le concept de bloc résiduel; l'idée fondamentale derrière le bloc résiduel est d'introduire des connexions de saut (skip connections) qui permettent à l'information d'être directement transmise à travers les couches sans subir des transformations supplémentaires, ce qui facilite l'apprentissage des fonctions résiduelles. ResNet est souvent défini avec des variantes de profondeur différentes, telles que ResNet-18, ResNet-34, ResNet-50, etc. Le nombre après le nom "ResNet" indique le nombre de couches résiduelles dans le réseau. Dans le cadre de ce travail, nous avons utilisé les architectures pré-entraînées de ResNet-50, ResNet-101 et ResNet-152 en tant qu'extracteurs de caractéristiques:
  - **ResNet-50:** ResNet-50 est en effet l'une des variantes de l'architecture ResNet, comportant 50 couches résiduelles. Cette architecture a été conçue pour surmonter le problème de la dégradation des performances dans les réseaux neuronaux profonds, et elle a apporté une amélioration significative dans diverses applications de vision par ordinateur.
  - **ResNet-101:** ResNet-101 est une autre variante de l'architecture ResNet avec 101 couches résiduelles, permettant d'apprendre des représentations très riches et complexes des données.
  - **ResNet-152:** Il s'agit d'une variante de l'architecture ResNet qui atteint sa profondeur totale de 152 couches en utilisant une combinaison de blocs résiduels de différentes profondeurs.
- **DenseNet [102]:** DenseNet est une architecture CNN qui se distingue par ses connexions denses entre les couches. Contrairement à d'autres architectures, DenseNet établit des connexions directes entre chaque couche et toutes les couches précédentes. Cela permet un flux d'informations direct entre les couches et favorise la réutilisation des caractéristiques à différents niveaux du réseau. DenseNet est construit autour de blocs denses, où chaque couche reçoit les caractéristiques de toutes les couches précédentes du même bloc. Par exemple, l'architecture DenseNet-121 est constituée de 121 couches de convolution, et elle est connectée avec une couche entièrement connectée (fully connected) de 1000 unités en tant que dernière couche de sortie.

- **Inception-V3 [103]:** C'est une architecture CNN de la famille Inception qui présente plusieurs améliorations, notamment l'utilisation du lissage des étiquettes (label smoothing), des convolutions factorisées  $7 \times 7$  et l'utilisation d'un classifieur auxiliaire pour propager les informations d'étiquetage plus loin dans le réseau. Elle utilise également la normalisation par lots (batch normalization) pour les couches en tête.

Une comparaison de chacune de ces diverses variantes de CNN peut être trouvée dans le Tableau 4.2.

**Tableau 4.2.** Comparaison entre plusieurs modèles CNN utilisés dans notre étude.

Modèle CNN	Date d'apparition	Total de couches	Couches de convolution	Paramètres (poids) entraînaibles
<b>VGG-16</b>	2014	16	13	14,714,688
<b>VGG-19</b>	2014	19	16	20,024,384
<b>ResNet-50</b>	2015	50	48	23,534,592
<b>ResNet-101</b>	2015	101	99	42,552,832
<b>ResNet-152</b>	2015	152	150	58,219,520
<b>Inception-V3</b>	2016	48	22	21,768,352
<b>DenseNet-121</b>	2017	121	121	6,953,856

#### 4.2.1.3. Classification

Les différents modèles de texture et d'apprentissage profond sont classifiés à l'aide de l'algorithme K-NN [71]. Pour mesurer la similarité entre les vecteurs de caractéristiques, nous avons utilisé la distance de Manhattan (décrite en chapitre 3). Le calcul du taux d'identification utilise l'équation suivante:

$$\text{Taux d'identification} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total de prédictions}} \times 100 \quad (4.1)$$

#### 4.2.2. Expérimentations & Etude comparative

Afin d'explorer et comparer les performances des caractéristiques de texture et de l'apprentissage profond pour la tâche de SSFR, nous avons sélectionné quatre descripteurs de caractéristiques de texture performants et évalué leur performance par rapport à sept architectures de réseaux de neurones convolutifs (CNN) pré-entraînées. Nous avons rapporté les

résultats obtenus sur l'ensemble de données AR, qui contient plusieurs variations d'images reflétant des conditions d'imagerie contraintes et non contraintes.

#### 4.2.2.1. Description de la base de données et son protocole d'évaluation

L'ensemble de données AR [72] se compose de plus de 4000 images faciales en couleur de 126 personnes (70 hommes et 56 femmes). Chaque individu dispose de 26 images de son visage frontal prises dans différentes conditions d'éclairage, affichant différentes expressions faciales telles que neutre, sourire, cri et colère, ainsi que des occlusions par des écharpes et des lunettes (Voir Figure 3.5 pour plus de détails).

Le Tableau 4.3 résume le protocole AR utilisé dans cette étude, qui comprend quatre sous-ensembles : un pour l'apprentissage et trois pour les tests. Chaque sous-ensemble comprend les références des images par personne, le nombre total de personnes par sous-ensemble, le nombre d'images par personne, ainsi que le nombre d'images utilisées pour entraîner et tester les performances des différents modèles.

**Tableau 4.3.** Répartition des personnes et des images entre l'ensemble d'entraînement et les ensembles de test pour chaque sous-ensemble.

	Sous-ensemble d'apprentissage	Sous-ensemble de test		
		Expression Faciale	Occlusion par lunette	Occlusion par écharpe
<b>Images Référentielle</b>	A	B, C, D, N, O, P, et Q	H, I, J, U, V, et W	K, L, M, X, Y, et Z
<b>#Personnes</b>	100	100	100	100
<b>#Images / Personnes</b>	1	7	6	6
<b>#Images</b>	100	700	600	600

#### 4.2.2.2. Expérimentations & Comparaison

Dans cette expérimentation, nous avons testé et comparé les performances de plusieurs modèles de SSFR. Tout d'abord, nous avons évalué les performances de différents extracteurs de caractéristiques de texture en utilisant le protocole décrit précédemment. Plus précisément, nous avons utilisé les descripteurs MB-C-BSIF, MB-C-LPQ, MB-C-LBP et MB-C-HOG. Chacun de ces descripteurs a été classifié à l'aide du classificateur K-NN avec la distance de Manhattan. Les résultats de cette expérimentation sont présentés dans le Tableau 4.4.

**Tableau 4.4.** Résultats expérimentaux des descripteurs basés sur la texture utilisés dans notre étude.

N°	Extracteurs des caractéristiques	Classificateur	Taux d'identification (%)		
			Expression Faciale	Occlusion par lunette	Occlusion par écharpe
1	MB-C-LBP	K-NN	88.86	51.67	32.33
2	MB-C-LPQ		87.86	55.17	42.17
3	MB-C-BSIF		<b><u>92.00</u></b>	<b><u>93.50</u></b>	<b><u>92.67</u></b>
4	MB-C-HOG		88.43	84.33	75.33

Ensuite, nous avons évalué les performances des CNN classiques pré-entraînés en utilisant le même protocole. Plus précisément, nous avons utilisé les architectures VGG-16, VGG-19, ResNet-50, ResNet-101, ResNet-152, DenseNet-121 et Inception-V3 en tant qu'extracteurs de caractéristiques. Dans ce cas, le CNN joue deux rôles: celui d'extracteur de caractéristiques et de classificateur. Afin de comparer les résultats en fonction de l'extraction de caractéristiques avec les descripteurs de texture, nous avons utilisé le classificateur K-NN à la place du classificateur CNN standard (c'est-à-dire dense). Le Tableau 4.5 présente les résultats de la deuxième expérimentation.

**Tableau 4.5.** Résultats expérimentaux des modèles basés sur l'apprentissage profond.

N°	Extracteurs des caractéristiques	Classificateur	Taux d'identification (%)		
			Expression Faciale	Occlusion par lunette	Occlusion par écharpe
1	VGG-16	K-NN	<b><u>84.14</u></b>	50.50	45.66
2	VGG-19		81.57	<b><u>58.66</u></b>	<b><u>46.16</u></b>
3	ResNet-50		62.00	11.33	22.16
4	ResNet-101		54.00	12.16	23.33
5	ResNet-152		56.86	13.00	18.17
6	Inception-V3		57.14	06.16	13.83
7	DenseNet-121		79.57	13.66	33.00

Les résultats présentés dans le Tableau 4.4 montrent que les quatre descripteurs de texture testés fonctionnent bien avec le sous-ensemble d'expressions faciales. Cependant, les performances de reconnaissance diminuent significativement avec les sous-ensembles occlusifs en utilisant les descripteurs MB-C-LBP et MB-C-LPQ. En revanche, MB-C-HOG et MB-C-

BSIF présentent une excellente robustesse face au défi de l'occlusion. Enfin, le descripteur de texture MB-C-BSIF est le meilleur opérateur avec les meilleurs résultats pour tous les sous-ensembles par rapport aux autres descripteurs de texture.

En examinant les résultats du Tableau 4.5, nous pouvons constater que VGG-16, suivi de VGG-19, présentent les meilleures performances avec tous les sous-ensembles, atteignant respectivement des taux d'identification de 84.14% et 81.57% pour la classification d'images sur le sous-ensemble des expressions faciales. Par conséquent, nous concluons que VGG-16 et VGG-19 peuvent mieux représenter les caractéristiques d'image SSFR que d'autres modèles plus profonds. Nous pensons que lorsque l'architecture CNN est plus profonde et plus complexe (par exemple, DenseNet, ResNet et Inception), le taux de performance diminue dans le cas de SSFR en raison du manque de données dans l'ensemble d'apprentissage.

Enfin, nous avons comparé les performances de la reconnaissance SSFR entre les approches basées sur la texture et l'apprentissage profond. MB-C-BSIF donne les meilleurs résultats pour les descripteurs de texture, tandis que VGG-16 obtient les meilleurs résultats pour les approches basées sur l'apprentissage profond. Entre MB-C-BSIF et VGG-16, il existe une différence significative. Nous pouvons en déduire que les descripteurs basés sur la texture sont plus utiles que les caractéristiques d'apprentissage profond dans le cas de SSFR, caractérisé par des données d'apprentissage limitées.

### **4.3. Combinaison des caractéristiques de texture et d'apprentissage profond pour la reconnaissance faciale à échantillon unique**

Les méthodes traditionnelles de la reconnaissance faciale, telles que les descripteurs de texture, peuvent rencontrer des difficultés qui dégradent les performances lorsqu'elles sont appliquées à des images non contrôlées. Cette limitation peut être attribuée à la variabilité présente dans les conditions d'éclairage, les poses, les expressions faciales, et d'autres facteurs, ce qui conduit à une dégradation des performances.

D'un autre côté, les CNNs sont des modèles d'apprentissage profonds qui nécessitent généralement un grand nombre d'échantillons en apprentissage pour obtenir de bonnes performances. Dans le cas du SSFR, où seulement un échantillon par individu est disponible lors de l'apprentissage, les performances du CNN sont limitées.

Pour surmonter ces limitations, nous avons décidé de combiner les caractéristiques des méthodes traditionnelles avec les caractéristiques profondes extraites par les CNNs afin d'améliorer les performances du SSFR en tirant parti des avantages de chaque méthode.

### 4.3.1. Méthodes

Dans cette section, nous présentons tout d'abord les opérations de prétraitement utilisées. Ensuite, nous décrivons les deux types de caractéristiques utilisées, à savoir les caractéristiques traditionnelles et les caractéristiques profondes extraites par les CNNs. Enfin, nous détaillons le schéma de la classification.

#### 4.3.1.1. *Augmentation des données et prétraitement*

Avant de fournir les images faciales à l'extracteur de caractéristiques, une phase de prétraitement doit être réalisée. Nous avons utilisé un filtrage médian et une égalisation d'histogramme adaptative à contraste limité (CLAHE) [96] pour améliorer la qualité de l'image, ainsi qu'une augmentation des données pour produire davantage d'échantillons faciaux.

#### 4.3.1.2. *Extraction de caractéristiques*

##### **A. Caractéristiques traditionnelles**

Dans cette partie de l'étude, le descripteur MB-C-BSIF a été sélectionné pour l'analyse des caractéristiques traditionnelles. Cette décision découle des résultats de l'étude comparative menée précédemment, qui a révélé que le descripteur MB-C-BSIF était le meilleur opérateur parmi tous les descripteurs de texture et caractéristiques profondes évalués.

##### **B. Caractéristiques profondes**

Selon l'étude comparative, il a été constaté que les modèles VGG-16 et VGG-19 étaient plus performants pour représenter les caractéristiques des images SSFR que d'autres modèles plus profonds. Par conséquent, nous avons choisi d'utiliser des versions pré-entraînées des réseaux VGG-16, VGG-19 et leur fusion pour extraire les caractéristiques des images. Le modèle développé dans cette étude examine l'efficacité de l'extraction des caractéristiques à partir de la dernière couche convolutive du réseau, également appelée *Flatten Layer*. Ces caractéristiques extraites sont utilisées comme caractéristiques d'apprentissage profond dans la méthode proposée.

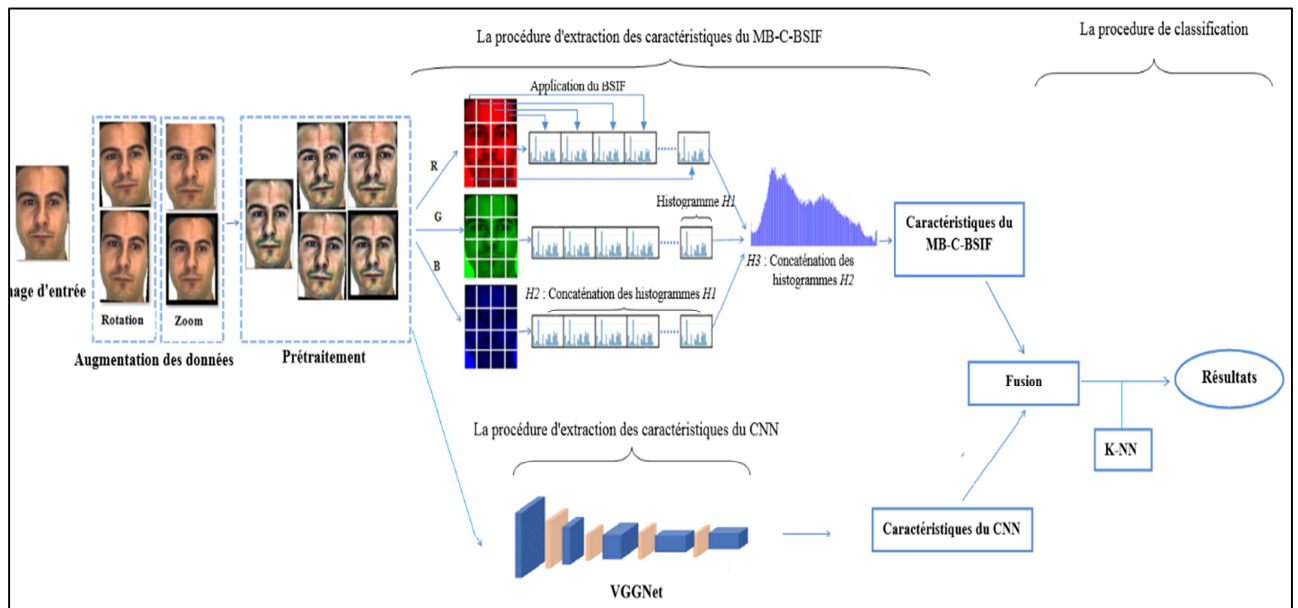
#### 4.3.1.3. *Classification des caractéristiques*

Le classificateur K-NN [71] est appliqué aux caractéristiques profondes et traditionnelles extraites. Pour mesurer la similarité entre les vecteurs de caractéristiques, nous avons utilisé la distance de Manhattan, également connue sous le nom de distance City Block (décrite en chapitre 3). Cette mesure de distance est considérée comme la plus appropriée pour notre méthode MB-C-BSIF. Nous avons également effectué des expériences en utilisant un classificateur de réseau de neurones convolutif appelé Dense, mais seulement avec des réseaux pré-entraînés. Cela signifie que nous avons utilisé des modèles de réseaux de neurones convolutifs pré-entraînés, tels que VGG-16 ou VGG-19, pour la classification des images. Ces modèles sont pré-entraînés sur de grandes bases de données d'images et ont déjà appris des caractéristiques discriminantes.

#### 4.3.2. **Système proposé**

Cette étude propose un système de reconnaissance faciale basé sur l'extraction des caractéristiques à partir de deux méthodes différentes: l'apprentissage profond (deep learning) et les méthodes traditionnelles.

Avant de procéder à l'extraction des caractéristiques, nous avons utilisé l'augmentation des données pour augmenter la taille de l'ensemble de données et éviter le surajustement. Pour ce faire, nous avons généré quatre patches supplémentaires à partir de chaque image de visage d'apprentissage en appliquant des zooms et des rotations (vers la droite et vers la gauche). Au total, nous avons obtenu 500 patches d'apprentissage augmentés à partir des 100 images d'apprentissage.



**Figure 4.1.** Le cadre général du modèle hybride développé.

Dans la première méthode, nous avons utilisé notre méthode traditionnelle précédente (notre première contribution), MB-C-BSIF, où nous avons extrait des caractéristiques de textures locales en appliquant le descripteur BSIF à partir de 16 blocs non superposés pour chaque composante (rouge, vert, bleu), comme illustré dans la Figure 4.1.

Dans la deuxième méthode, nous avons utilisé deux réseaux de neurones convolutifs pré-entraînés, VGG-16 et VGG-19, pour extraire les caractéristiques profondes.

Dans la dernière étape, nous avons combiné les caractéristiques extraites à partir du CNN pré-entraîné et celles du MB-C-BSIF en un seul vecteur de caractéristiques global. Cette fusion a pour objectif d'améliorer les performances du système de SSFR en exploitant les informations provenant des deux méthodes. La Figure 4.1 présente un schéma graphique détaillé du cadre développé, accompagné d'une description complète de ses différentes composantes.

### 4.3.3. Analyse expérimentale

Pour évaluer les performances de notre modèle, nous avons effectué une série de tests en utilisant un ensemble des images faciales de la base de données AR.



#### 4.3.3.1. Base de données

Nous avons mené plusieurs expériences pour évaluer les performances de SSFR en utilisant divers sous-ensembles de la base de données AR. La base de données AR est décrite en détail dans le chapitre précédent.

#### 4.3.3.2. Expérimentations

Dans notre première série d'expérimentations, nous avons utilisé le même protocole (Protocole I) que celui utilisé dans l'étude comparative (voir Tableau 4.3). Nous avons utilisé quatre sous-ensembles de 100 classes (50 hommes et 50 femmes) pour évaluer les performances de notre approche SSFR. Dans chaque sous-ensemble, une seule image (image A) par personne a été réservée à l'apprentissage, tandis que les trois autres sous-ensembles ont été réservés aux tests.

Le premier sous-ensemble de test comprenait des images avec diverses expressions (7 images par personne, liées aux images B, C, D, N, O, P et Q). Le deuxième sous-ensemble de test comprenait des images avec des personnes portant des lunettes de soleil (6 images par personne, liées aux images H, I, J, U, V et W). Le dernier sous-ensemble de test comprenait des images où les visages étaient masqués par des foulards (6 images par personne, liées aux images K, L, M, X, Y et Z).

#### **A. VGG-16, VGG-19 et fusion VGG-16+VGG-19 avec une classification dense**

Dans cette expérimentation, nous avons utilisé des architectures de réseaux de neurones convolutifs, tels que VGG-16, VGG-19 et une combinaison de VGG-16 et VGG-19. Dans ces trois cas, le CNN a été utilisé à la fois comme extracteur de caractéristiques et comme classificateur.

La phase d'extraction de caractéristiques est une étape où le CNN parcourt l'image en appliquant plusieurs opérations de convolution profonde. Ces opérations de convolution permettent au CNN de détecter des motifs et des caractéristiques dans l'image. Une fois que les caractéristiques sont extraites par le CNN, elles sont ensuite utilisées pour la classification. Pour cela, nous avons utilisé le classifieur dense traditionnel, qui est une couche de neurones entièrement connectés (en anglais, *Fully Connected*). Ce classificateur prend en entrée les caractéristiques extraites par le CNN et effectue la classification finale en assignant une étiquette

ou une classe à l'image. Les résultats de ces expériences sont présentés dans les trois premières lignes du Tableau 4.6.

### **B. MB-C-BSIF avec une classification par K-NN**

Dans la quatrième ligne du Tableau 4.6, nous avons présenté les résultats obtenus en utilisant l'extracteur de caractéristiques traditionnel MB-C-BSIF, tout en suivant le même protocole d'évaluation précédent. Pour cette configuration, la classification a été réalisée en utilisant le classificateur K-NN avec la distance de Manhattan.

### **C. VGG-16, VGG-19 et fusion VGG-16+VGG-19 avec une classification par K-NN**

Dans le troisième type d'expérimentation, nous avons opté pour une conception modifiée du CNN. Dans cette configuration, le rôle principal du CNN était de réaliser l'extraction des caractéristiques par apprentissage profond à l'aide des opérations de convolution profondes. Nous avons apporté une modification à la conception du CNN en remplaçant le classificateur dense traditionnel par un classificateur K-NN. Dans cette expérimentation nous avons évalué les performances du CNN modifié, qui utilise un classificateur K-NN, en utilisant le même protocole d'évaluation précédent. Les résultats de cette configuration ont été présentés dans le Tableau 4.6, aux lignes 5 à 7, en utilisant les architectures VGG-16, VGG-19 et VGG-16+VGG-19 combinées, respectivement.

### **D. Fusion VGG-16+MB-C-BSIF, Fusion VGG-19+MB-C-BSIF, et Fusion VGG-16+VGG-19+MB-C-BSIF avec une classification par K-NN**

Dans la dernière expérimentation, nous avons mesuré les performances en combinant les modèles VGG-16, VGG-19 et la combinaison VGG-16+VGG-19 avec l'extracteur MB-C-BSIF. La combinaison a été faite au niveau des caractéristiques, et le K-NN associé à une distance de Manhattan a été utilisé pour classifier les caractéristiques combinées et reconnaître les visages. Les résultats de cette expérience ont été présentés dans le Tableau 4.6, aux lignes 8 à 10.

**Tableau 4.6.** Comparaison des résultats obtenus à l'aide de différentes approches  
d'extraction/classification de caractéristiques.

N°	Extracteur de caractéristiques	Classificateur	Taux d'identification (%)		
			Expression Faciale	Occlusion par Lunette	Occlusion par Echarpe
1	VGG-16	Dense	73.14	27.25	53.67
2	VGG-19		69.86	14.83	49.83
3	VGG-16+VGG-19		72.57	22.67	47.50
4	MB-C-BSIF	K-NN	92.00	93.50	92.67
5	VGG-16		84.14	50.50	45.66
6	VGG-19		81.57	58.66	46.16
7	VGG-16+VGG-19		85.57	59.16	47.00
8	VGG-16+MB-C-BSIF		92.71	<b><u>94.50</u></b>	93.83
9	VGG-19 + MB-C-BSIF		<b><u>93.00</u></b>	94.33	93.66
<b>10</b>	<b>VGG-16+VGG-19+MB-C-BSIF</b>		<b><u>93.00</u></b>	<b><u>94.50</u></b>	<b><u>94.16</u></b>

Les observations suivantes peuvent être déduites du Tableau 4.6:

- Les résultats du CNN modifié sont supérieurs à ceux du CNN standard, quel que soit le sous-ensemble de test et quelle que soit l'architecture. Cela signifie que les modifications apportées à l'architecture du CNN, c.-à-d. l'utilisation d'un classificateur K-NN au lieu de dense, ont amélioré ses performances dans la tâche du SSFR.
- Le descripteur MB-C-BSIF surpasse les modèles d'apprentissage profond que ce soit avec des CNN standards ou modifiés, dans tous les sous-ensembles de test et avec toutes les architectures. Une explication possible est que les modèles d'apprentissage profond nécessitent un grand ensemble de données d'entraînement pour converger efficacement, ce qui n'a pas été atteint dans cette expérience malgré l'utilisation d'une stratégie d'augmentation des données.
- La combinaison des caractéristiques des CNNs avec des caractéristiques traditionnelles a amélioré les performances du SSFR; en effet, il est évident dans le Tableau 4.6 que la combinaison des modèles VGG-16, VGG-19 et VGG-16+VGG-19 avec les caractéristiques du MB-C-BSIF en utilisant le classificateur K-NN a amélioré les

performances avec tous les sous-ensembles de test. En particulier, la combinaison des caractéristiques VGG-16+VGG-19+MB-C-BSIF a donné les meilleurs résultats. Cela signifie que la fusion des informations extraites par différents types d'extracteurs peut fournir une représentation plus complète des visages et améliorer ainsi la précision du SSFR.

#### 4.3.3.3. Comparaison #1

Le Tableau 4.7 présente une comparaison entre les performances de notre système SSFR proposé, basé sur la combinaison VGG-16+VGG-19+MB-C-BSIF, et les résultats de certains articles récemment publiés dans le domaine. Pour cette comparaison, nous avons utilisé le protocole d'évaluation I qui consiste à utiliser la première image de chaque personne (image A) comme ensemble d'apprentissage, et six images d'expression faciale (images B, C, D, N, O et P) comme ensemble de test. Tous les travaux présentés dans le Tableau 4.7 ont suivi ce même protocole d'évaluation. Les résultats montrent que notre approche a obtenu d'excellentes performances en termes de taux d'identification, qui sont compétitives et les plus élevées, particulièrement en présence de variations en expression faciale.

**Tableau 4.7.** Comparaison des performances de l'approche proposée par rapport à d'autres méthodes utilisant le protocole I.

Références	Méthodes	Taux d'identification (%)
Lu et al. [46] (2012)	DMMA	79.00
Ji et al. [75] (2017)	CPL	83.04
Mehrasa et al. [74] (2017)	SLPMM	86.00
Pang et al. [88] (2019)	RHDA	96.78
Chu et al. [87] (2019)	MFSA	89.66
Zhang et al. [49] (2021)	DNNC	86.67
Adjabi et al. [98] (2021)	MB-C-BSIF	96.17
<b>Notre méthode [99] (2022)</b>	<b>VGG-16+VGG-19+MB-C-BSIF</b>	<b><u>97.33</u></b>

#### 4.3.3.4. Comparaison #2

Dans cette partie, nous avons utilisé un deuxième protocole d'évaluation (Protocole II) pour évaluer les performances de notre méthode développée face à des variations d'occlusion et d'éclairage. Le Protocole II utilise l'image A de chaque personne pour créer l'ensemble

d'apprentissage, tandis que deux sous-ensembles sont considérés comme ensembles de tests. Le premier ensemble de tests comprend deux images occluses avec des lunettes de soleil et un foulard, représenté par les images H+K de chaque personne. Le deuxième ensemble de tests comprend deux images de chaque personne affectée à la fois par des variations d'éclairage et d'occlusion, représenté par les images J+M. Au total, nous avons utilisé 100 images en apprentissage, 200 images pour le test I et 200 images pour le test II.

Le Tableau 4.8 présente les résultats de notre système SSFR, ainsi que ceux de certains articles récemment publiés, tous évalués selon le même protocole. Les résultats du Tableau 4.8 mettent en évidence que le modèle VGG-16+VGG-19+MB-C-BSIF obtient le score le plus élevé parmi les méthodes de l'état de l'art considérées.

**Tableau 4.8.** Comparaison des performances de l'approche proposée avec d'autres méthodes utilisant le protocole II.

Références	Méthodes	Occlusion (H+K) (%)	Eclairage + Occlusion (J+M) (%)	Taux d'identification moyen (%)
Lu et al. [46] (2012)	DMMA	46.90	30.90	38.90
Zeng et al. [60] (2017)	DCNN	96.5	88.3	92.20
Chu et al. [87] (2019)	MFSa	91.30	79.00	85.20
Cuculo et al. [90] (2019)	SSLD	90.18	82.02	86.10
Zhang et al. [49] (2021)	DNNC	92.50	79.50	86.00
Du and Da [63] (2021)	BDL	93.03	91.55	92.29
Adjabi et al. [98] (2021)	MB-C-BSIF	99.50	98.50	99.00
<b>Notre méthode (2022) [99]</b>	<b>VGG-16+VGG-19+MB-C-BSIF</b>	<b><u>99.66</u></b>	<b><u>98.83</u></b>	<b><u>99.25</u></b>

#### 4.4. Conclusion

Dans ce chapitre, nous avons introduit une méthode hybride de reconnaissance faciale à échantillon unique (SSFR) qui combine le descripteur MB-C-BSIF (*Multi-Block Color-Binarized Statistical Image Features*) et des réseaux de neurones convolutifs. Cette méthode permet

l'extraction de caractéristiques faciales en utilisant différentes sources d'informations, notamment des informations locales, régionales, globales et de texture couleur fournies par le MB-C-BSIF, ainsi que des informations plus profondes extraites par le CNN.

Lors de nos expérimentations, nous avons évalué notre méthode proposée sur différents sous-ensembles d'images provenant de la base de données AR. Les résultats obtenus à partir de cette base de données ont clairement démontré que notre méthode améliore de manière significative les performances de la classification SSFR face à diverses variations dans la reconnaissance faciale.

## **CHAPITRE 5: La Reconnaissance Faciale à partir d'un Seul Echantillon en Utilisant la Transformée en Ondelettes Discrète et les Caractéristiques de Texture**

### **5.1. Introduction**

Ce chapitre propose une solution pour améliorer la performance des systèmes de reconnaissance faciale à partir d'un seul échantillon (SSFR) en proposant une méthode d'analyse en multi-résolution. Cette méthode utilise la transformée en ondelettes discrète (en anglais, *discrete wavelet transform* (DWT)) pour extraire des caractéristiques d'image statistiques et binarisées (BSIF) en couleur à différentes résolutions. Cette approche vise à capturer davantage d'informations et à améliorer la capacité du système à reconnaître les visages dans des conditions non-contrôlées.

Dans ce chapitre, nous fournissons une description détaillée des méthodes utilisées. Ensuite, nous abordons l'utilisation de l'analyse en multi-résolution pour extraire des caractéristiques BSIF en couleur à différentes résolutions. Nous décrivons aussi l'étape de classification qui consiste à catégoriser les caractéristiques extraites dans les classes correspondantes en utilisant le classificateur K-NN. Nous présentons ensuite les résultats de notre analyse expérimentale en évaluant les performances de notre approche sur l'ensemble de données AR et en comparant sa performance par rapport à d'autres méthodes de l'état de l'art. Enfin, nous concluons le chapitre en résumant les principales synthèses de cette étude.

### **5.2. Préliminaires**

Dans cette section, nous présentons les cadres théoriques des méthodes et outils utilisés dans ce travail.

#### **5.2.1. Transformée en ondelettes discrète à 2 dimensions (2D-DWT)**

Les ondelettes ont été utilisées dans de nombreuses applications, notamment l'extraction des caractéristiques, la compression, le dé-bruitage et la détection des contours. La transformée en ondelettes discrète (DWT) divise un signal donné en plusieurs trames, chacune étant une série temporelle de coefficients caractérisant l'évolution temporelle du signal dans une bande de fréquence appropriée [104-105].

Le cadre théorique pour appliquer une DWT sur une image 2D est exprimé mathématiquement par les équations (5.1) à (5.7) et expliqué comme suit :

Supposons que  $\psi(t)$  est une fonction de l'ondelette mère. La famille de fonctions d'ondelettes  $\psi(s, p)(t)$  peut être obtenue comme suit [106-107]:

$$\psi(s, p)(t) = \frac{1}{\sqrt{s}} \Psi \left( \frac{t-p}{s} \right) \quad (5.1)$$

Où  $s$  est le paramètre d'échelle,  $t$  est une instance et  $p$  est le paramètre de position. Soit  $f(x, y)$  une image de taille  $M \times N$ . La 2D-DWT s'exprime comme suit :

$$W_\varphi(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \varphi_{j_0, m, n}(x, y) \quad (5.2)$$

$$W_\psi^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \psi_{j, m, n}(x, y) \quad (5.3)$$

Où  $i = \{horizontal, vertical, diagonal\}$ ,  $W_\varphi(j_0, m, n)$  sont les coefficients qui définissent une approximation de  $f(x, y)$  à une échelle  $j_0$ ,  $W_\psi^i(j, m, n)$  sont les coefficients qui ajoutent des détails horizontaux, verticaux et diagonaux pour une échelle  $j \geq j_0$ ,  $\varphi$  définit la fonction d'échelle, et  $\psi$  définit la fonction d'ondelette.

Dans une image, plusieurs résolutions sont représentées par des cycles répétitifs de mise à l'échelle (passe-bas) et de transformée en ondelettes (passe-haut), comme illustré dans la Figure 5.1. La mise à l'échelle capture les informations de basse fréquence de l'image, tandis que l'ondelette collecte les informations de haute fréquence de l'image. À chaque cycle de la transformée en ondelettes, une image à basse résolution et une image de détails fins, chacune ayant une taille deux fois plus petite que l'image originale, sont générées. Étant donné que les informations dans l'image de détails fins sont souvent limitées, la version à plus basse résolution capture la plupart des informations de l'image originale, ce qui entraîne une efficacité élevée de représentation de l'image [108].

La transformée en ondelettes 2D divise une image en quatre sous-images de sous-bandes: basse-basse (*Low-Low* LL), basse-haute (*Low-High* LH), haute-basse (*High-Low* HL) et haute-



haute (*High-High* HH), comme indiqué dans la Figure 5.1. Les basses fréquences représentent les coefficients d'approximation, tandis que les hautes fréquences représentent les coefficients de détails (horizontal, vertical et diagonal). Mathématiquement, cette procédure peut être présentée comme suit :

Soit  $\varphi$  qui définit la fonction d'échelle et  $\psi$  qui définit la fonction d'ondelette. Comme indiqué dans les équations (5.4) à (5.7), la transformée en ondelettes discrète (DWT) produit quatre images de quart de la taille à chaque niveau de décomposition: une image d'aperçu  $\varphi(x, y)$ , une image de détails horizontaux  $\psi^H(x, y)$ , une image de détails verticaux  $\psi^V(x, y)$  et une image de détails diagonaux  $\psi^D(x, y)$  [106]:

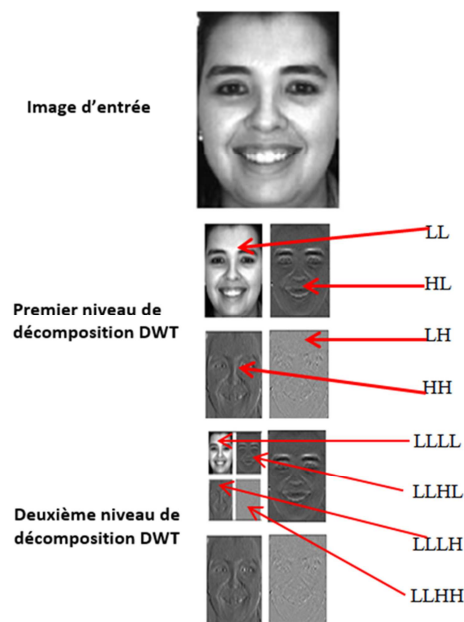
$$\varphi(x, y) = \varphi(x)\varphi(y) \rightarrow LL \quad (5.4)$$

$$\psi^H(x, y) = \varphi(x)\psi(y) \rightarrow HL \quad (5.5)$$

$$\psi^V(x, y) = \psi(x)\varphi(y) \rightarrow LH \quad (5.6)$$

$$\psi^D(x, y) = \psi(x)\psi(y) \rightarrow HH \quad (5.7)$$

Dans l'exemple suivant, des composantes de fréquence DWT pour deux valeurs d'échelle distinctes sont générées, et les sections de basse fréquence des composantes acquises (LL, LLLL) sont utilisées dans le processus de la reconnaissance.



**Figure 5.1.** Exemples d'application de la DWT aux premiers et deuxièmes niveaux de décomposition.

### 5.2.2. Binarized statistical image features (BSIF)

BSIF est un descripteur de texture développé par Kannala et Rahtu [65] pour décrire et analyser les caractéristiques de texture dans une image. Son fonctionnement est détaillé dans la section 2 du chapitre 3 de cette thèse.

## 5.3. Approche proposée

Dans ce chapitre, une approche en trois étapes est proposée pour aborder le défi SSFR. La première étape consiste en un prétraitement des images utilisant un filtrage médian et une égalisation d'histogramme adaptative limitée au contraste (CLAHE) afin d'éliminer les bruits potentiels et d'harmoniser l'éclairage des images. Ensuite, une analyse en multi-résolution est appliquée pour extraire les caractéristiques BSIF à différentes échelles en utilisant la transformée en ondelettes discrète (DWT). Enfin, une étape de classification est réalisée pour catégoriser les caractéristiques extraites dans la classe correspondante en utilisant le classificateur K-NN. Cette section détaille les concepts fondamentaux de cette approche.

### 5.3.1. Prétraitement

Nous avons appliqué des techniques de prétraitement telles que le filtrage médian [64] et le CLAHE [96] dans ce travail, car elles peuvent considérablement améliorer les performances et la robustesse de notre approche. Le filtrage médian réduit efficacement le bruit, tandis que le CLAHE est une technique d'amélioration du contraste qui peut améliorer la visibilité des caractéristiques dans l'image.

### 5.3.2. Extraction des caractéristiques

La phase d'extraction des caractéristiques constitue la principale contribution de notre travail. Le cadre général de notre approche proposée pour l'extraction des caractéristiques est présenté dans la Figure 5.2. Il peut être résumé comme suit :

- Nous divisons l'image d'entrée en composantes rouge, verte et bleue afin d'extraire des informations de couleur plutôt que des informations en niveaux de gris.
- Nous appliquons la DWT deux fois pour chaque composante afin d'obtenir les premier et deuxième niveaux de décomposition.

- Nous appliquons le descripteur BSIF pour chaque composante de l'image originale, la sous-bande LL et la sous-bande LLLL (LL et LLLL contiennent les basses fréquences ce qui signifie les informations les plus pertinentes).
- Nous concaténons les trois histogrammes extraits de chaque composante pour générer un histogramme régional.
- Nous concaténons les trois histogrammes régionaux pour obtenir le vecteur de caractéristiques final, qui sera utilisé comme modèle de caractéristiques.

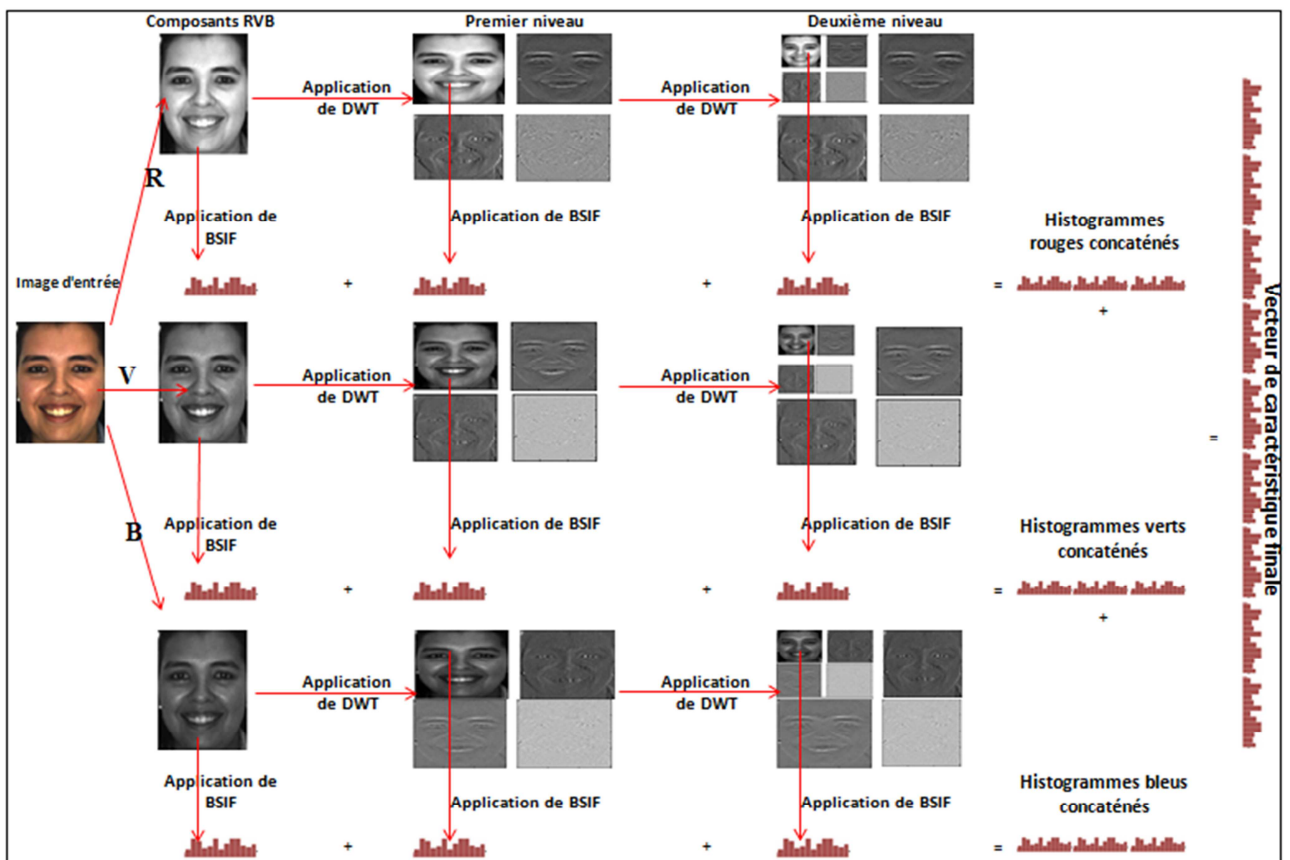


Figure 5.2. Organigramme graphique de l'approche d'extraction des caractéristiques proposée.

### 5.3.3. Classification

Enfin, une étape de classification est réalisée pour catégoriser les caractéristiques extraites dans leurs classes respectives en utilisant le classificateur K-NN [71] avec la distance de Manhattan (décrite en détails en chapitre 3).

## 5.4. Analyse expérimentale

### 5.4.1. Base de données

Nous avons mené plusieurs expérimentations pour évaluer les performances de l'approche proposée en utilisant divers sous-ensembles de la base de données AR. La base de données AR est décrite en détail dans le chapitre 3 (voir la Figure 3.5).

### 5.4.2. Expérimentation #1

Dans notre première série d'expérimentations, nous avons utilisé quatre sous-ensembles de 100 classes (50 hommes et 50 femmes) pour évaluer les performances de notre approche SSFR. Dans chaque sous-ensemble, une seule image (image A) par personne a été réservée à l'apprentissage, tandis que les trois autres sous-ensembles ont été réservés aux tests. Le Tableau 5.1 récapitule le nombre d'individus, le nombre d'images par individu, et le nombre d'images utilisées pour entraîner et tester les performances de notre approche.

**Tableau 5.1.** Répartition des personnes et des images entre l'ensemble d'apprentissage et les ensembles de test pour chaque personne.

	Sous-ensemble d'apprentissage	Sous-ensemble de test		
		Expression Faciale	Occlusion par lunette	Occlusion par écharpe
<b>Images Référentielle</b>	A	B, C, D, N, O, P, et Q	H, I, J, U, V, et W	K, L, M, X, Y, et Z
<b>#Personnes</b>	100	100	100	100
<b>#Images / Personnes</b>	1	7	6	6
<b>#Images</b>	100	700	600	600

De plus, le taux de reconnaissance (*Recognition rate* (RR)) a été utilisé comme une mesure statistique pour évaluer et comparer les performances de notre approche proposée par rapport à d'autres travaux de l'état de l'art:

$$\text{Taux de reconnaissance} = \frac{\text{Nombre d'images de test correctement identifiées}}{\text{Nombre total d'images de test}} \times 100 \quad (5.8)$$

Les résultats de cette expérimentation sont présentés dans le Tableau 5.2. Pour mener une analyse comparative, nous avons également inclus dans le Tableau 5.2 les résultats de nos travaux précédents, tels que VGG-16, VGG-19, VGG-16+VGG-19 [58] et MB-C-BSIF [98].

**Tableau 5.2.** Résultats et comparaison de l'expérimentation #1.

Extraction des caractéristiques	Classificateur	RR (%)		
		Expression Faciale	Occlusion par lunette	Occlusion par écharpe
VGG-16	Dense	73.14	27.25	53.67
VGG-19		69.86	14.83	49.83
VGG-16+VGG-19		72.57	22.67	47.50
MB-C-BSIF	K-NN	92.00	<b><u>93.50</u></b>	92.67
<b>Notre approche</b>		<b><u>95.57</u></b>	87.33	<b><u>93.50</u></b>

Nous pouvons observer dans le tableau que les performances de l'approche proposée sont parfaites, avec les variations d'expression faciale et d'occlusion par une écharpe. Cependant, elles se classent à la deuxième position en présence des lunettes. Il convient de noter que les méthodes comparées appartiennent aux méthodes d'apprentissage profond et de méthodes de texture.

### 5.4.3. Expérimentation #2

Dans la deuxième expérimentation, nous avons utilisé un protocole d'évaluation différent, consistant à utiliser l'image A de chaque individu pour former l'ensemble d'apprentissage, et les images B, C, D, N, O et P pour former l'ensemble de test. Au total, nous avons 100 images dans l'ensemble d'apprentissage et 600 images dans l'ensemble de test. Nous avons également utilisé le taux de reconnaissance moyen pour évaluer les performances de notre approche et les comparer à certaines approches concurrentes récemment publiées, utilisant le même protocole d'évaluation. Les résultats de cette expérimentation sont présentés dans le Tableau 5.3. Nous pouvons observer dans le Tableau 5.3 que notre résultat (98%) est très compétitif et surpasse nos travaux précédents ainsi que les approches de l'état de l'art.

**Tableau 5.3.** Résultats et comparaison de l'expérimentation #2.

Références	Méthodes	RR (%) Moyen
Mehrasa et al. [74] (2017)	SLPMM	86.00
Pang et al. [88] (2019)	RHDA	96.78
Chu et al. [87] (2019)	MFSA	89.66
Zhang et al. [49] (2021)	DNNC	86.67
Adjabi et al. [98] (2021)	MB-C-BSIF	96.17
Adjabi et al. [99] (2022)	MB-C-BSIF+VGG-16+VGG-19	97.33
<b>Notre approche [109]</b>	<b>DWT+ BSIF</b>	<b><u>98.00</u></b>

## 5.5. Conclusion

Dans cette contribution, nous avons proposé une stratégie d'extraction des caractéristiques pour relever le défi de SSFR. Pour résoudre ce problème, nous avons appliqué la transformée en ondelettes discrète (2D-DWT) sur l'image faciale afin de produire plusieurs représentations de l'image à différentes résolutions. Ensuite, nous avons appliqué le descripteur BSIF sur l'image couleur originale et aux sous-bandes d'images générées avec des résolutions de basses fréquences. Enfin, nous avons combiné les histogrammes extraits de chaque niveau pour créer le vecteur des caractéristiques final. Cette méthode capture les caractéristiques faciales et les informations de texture à différents niveaux, ce qui améliore potentiellement les performances de la reconnaissance faciale à partir d'un seul échantillon. Les résultats expérimentaux sur l'ensemble de données AR ont montré que la méthode proposée a atteint un taux de reconnaissance élevé et surpasse les travaux de l'état de l'art existants, en particulier dans les variations d'expression faciale.

## *Conclusion Générale*

La biométrie fait référence aux caractéristiques physiques ou comportementales des êtres humains qui peuvent être utilisées à diverses fins, allant de la reconnaissance des actions humaines à l'identification et à la vérification des individus. L'exactitude biométrique a été un sujet de recherche depuis des décennies. Les scientifiques cherchent à améliorer les performances des modèles biométriques en intégrant autant d'obstacles et de défis que possible, tels que la pose, les occlusions, les expressions, et les arrière-plans. La reconnaissance faciale a attiré une attention significative au cours des trente dernières années, en raison de sa simplicité d'application dans le domaine de l'analyse d'images et de la reconnaissance de formes. Cette tendance peut être expliquée par deux principales raisons: premièrement, la demande croissante dans les secteurs commerciaux et juridiques, et deuxièmement, la disponibilité de technologies pertinentes telles que les smartphones, les appareils photo numériques, et les GPU. La reconnaissance faciale à partir d'un seul échantillon (SSFR) est un défi en vision par ordinateur. Dans ce scénario, il n'y a qu'un seul exemple de chaque individu sur lequel entraîner le système, ce qui rend difficile l'identification des personnes dans des environnements non contraints, en particulier lorsqu'il s'agit de variations d'expression faciale, de posture, d'éclairage, et d'occlusion.

Dans cette thèse, nous étions particulièrement intéressés à proposer de nouvelles idées et des schémas pour améliorer les tâches de la reconnaissance faciale à partir d'un seul échantillon. Bien que les solutions proposées aient surpassé les méthodes connexes, plusieurs difficultés complexes nécessitent un développement supplémentaire.

Pour résoudre cette problématique, nous avons proposé trois contributions majeures dans cette thèse:

- Dans notre première contribution, nous avons proposé une méthode novatrice, appelée "multi-block color binarized statistical image features (MB-C-BSIF)". Cette méthode exploite les informations locales, régionales, globales, et de couleur pour obtenir une représentation précise et fiable des visages, en s'appuyant sur le descripteur de texture BSIF.
- Dans notre deuxième contribution, nous avons proposé une approche hybride combinant deux méthodes, l'une traditionnelle et l'autre basée sur l'apprentissage profond. Plus spécifiquement, nous avons fusionné le descripteur de texture couleur

MB-C-BSIF avec un réseau de neurones convolutifs (CNN), afin de bénéficier des avantages complémentaires de ces deux approches.

- Dans notre troisième contribution, nous avons proposé une méthode d'analyse en multi-résolution basée sur la transformée en ondelette discrète (DWT), visant à extraire des caractéristiques statistiques binarisées à différentes échelles. Cette approche vise à capturer un plus grand nombre d'informations et à améliorer la capacité du système à reconnaître les visages dans des conditions variables.

Ces contributions visent à améliorer les performances de la reconnaissance faciale en surmontant les défis liés à la disponibilité limitée d'échantillons par individu, à la diversité des expressions faciales, à la posture, à l'éclairage, et aux occlusions. Les résultats obtenus ont démontré une supériorité de nos approches par rapport aux méthodes connexes actuelles, ouvrant ainsi de nouvelles perspectives de recherche et de développement dans le domaine de la reconnaissance faciale.

Bien que ce travail puisse être étendu à de nombreux types de recherches, les futures recherches pourraient donc se concentrer sur:

- Appliquer une stratégie de sélection de caractéristiques basée sur la fusion des traits MB-C-BSIF et des réseaux de neurones convolutifs peut permettre d'améliorer la performance globale de la reconnaissance faciale en réduisant la taille des vecteurs de caractéristiques et en ne conservant que les informations les plus pertinentes. Un exemple de technique de sélection des vecteurs est l'Analyse en Composantes Principales (en anglais, Principal Component Analysis (PCA)). Une autre approche est la méthode du "Maximum Relevance - Minimum Redundancy" (en anglais, Maximum Relevance - Minimum Redundancy (MRMR)).
- Développer des méthodes d'apprentissage profond basées sur des informations sémantiques, telles que l'âge, le genre, et l'origine ethnique, afin de résoudre le problème de SSFR. C'est un domaine qui mérite une étude approfondie.
- Investiguer et analyser le problème de SSFR dans des environnements non contraints en utilisant des bases de données à grande échelle contenant des millions d'images faciales.



## *Références Bibliographiques*

- [1] Micheli-Tzanakou, E., Plataniotis, K. N., & Donida Labati, R. (2021). Biometrics: Terms and Definitions. In *Encyclopedia of Cryptography, Security and Privacy* (pp. 1-5). Springer.
- [2] Oyeleye, C. A., Fagbola, T. M., Babatunde, R. S., & Adigun, A. A. (2012). An exploratory study of odor biometrics modality for human recognition. *International Journal of Engineering Research & Technology*, 1(9), 1-9.
- [3] Hadid, A., Evans, N., Marcel, S., & Fierrez, J. (2015). Biometrics systems under spoofing attack: an evaluation methodology and lessons learned. *IEEE Signal Processing Magazine*, 32(5), 20-30.
- [4] Adeoye, O. S. (2010). A survey of emerging biometric technologies. *International Journal of Computer Applications*, 9(10), 1-5.
- [5] Kataria, A. N., Adhyaru, D. M., Sharma, A. K., & Zaveri, T. H. (2013, November). A survey of automated biometric authentication techniques. In *2013 Nirma university international conference on engineering (NUiCONE)* (pp. 1-6). IEEE.
- [6] Meng, W., Wong, D. S., Furnell, S., & Zhou, J. (2014). Surveying the development of biometric user authentication on mobile phones. *IEEE Communications Surveys & Tutorials*, 17(3), 1268-1293.
- [7] Alsaadi, I. M. (2015). Physiological biometric authentication systems, advantages, disadvantages and future development: A review. *International Journal of Scientific & Technology Research*, 4(12), 285-289.
- [8] BENZAOU, A. (2015). *Identification Biométrique par Descripteurs de Texture Locaux: Application au Visage & Oreille* (Thèse de Doctorat en Electronique, Université 8 Mai 1945-Guelma).
- [9] Benzaoui, A., Khaldi, Y., Bouaouina, R., Amrouni, N., Alshazly, H., & Ouahabi, A. (2023). A comprehensive survey on ear recognition: databases, approaches, comparative analysis, and open challenges. *Neurocomputing*.
- [10] Saini, R., & Rana, N. (2014). Comparison of various biometric methods. *International Journal of Advances in Science and Technology*, 2(1), 24-30.
- [11] Jain, A., Bolle, R., & Pankanti, S. (Eds.). (1999). *Biometrics: personal identification in networked society* (Vol. 479). Springer Science & Business Media.
- [12] Adjabi, I., Ouahabi, A., Benzaoui, A., & Taleb-Ahmed, A. (2020). Past, present, and future of face recognition: A review. *Electronics*, 9(8), 1188.
- [13] Sabhanayagam, T., Venkatesan, V. P., & SenthamaraiKannan, K. (2018). A comprehensive survey on various biometric systems. *International Journal of Applied Engineering Research*, 13(5), 2276-2297.
- [14] Kaur, N. (2021, March). A study of biometric identification and verification system. In *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 60-64). IEEE.
- [15] Khaldi, Y. (2022). *Biometric Identification using Deep Ear Features* (Doctoral dissertation in Computer Science, University AKLI MOHAND OULHADJ-Bouira).
- [16] Sharif, M., Raza, M., Shah, J. H., Yasmin, M., & Fernandes, S. L. (2019). An overview of biometrics methods. *Handbook of multimedia information security: techniques and applications*, 15-35.

- 
- [17] Kortli, Y., Jridi, M., Al Falou, A., & Atri, M. (2020). Face recognition systems: A survey. *Sensors*, 20(2), 342.
- [18] O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in cognitive sciences*, 6(6), 261-266.
- [19] Dantcheva, A., Chen, C., & Ross, A. (2012, September). Can facial cosmetics affect the matching accuracy of face recognition systems?. In *2012 IEEE Fifth international conference on biometrics: theory, applications and systems (BTAS)* (pp. 391-398). IEEE.
- [20] Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11), 1948-1962.
- [21] Ouamane, A., Benakcha, A., Belahcene, M., & Taleb-Ahmed, A. (2015). Multimodal depth and intensity face verification approach using LBP, SLF, BSIF, and LPQ local features fusion. *Pattern Recognition and Image Analysis*, 25, 603-620.
- [22] Chihaoui, M., Elkefi, A., Bellil, W., & Ben Amar, C. (2016). A survey of 2D face recognition techniques. *Computers*, 5(4), 21.
- [23] Benzaoui, A., Bourouba, H., & Boukrouche, A. (2012, October). System for automatic faces detection. In *2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA)* (pp. 354-358). IEEE.
- [24] Martinez, A. M. (2002). Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern analysis and machine intelligence*, 24(6), 748-763.
- [25] Sid Ahmed, S., Messali, Z., Ouahabi, A., Trepout, S., Messaoudi, C., & Marco, S. (2015). Nonparametric denoising methods based on contourlet transform with sharp frequency localization: Application to low exposure time electron microscopy images. *Entropy*, 17(5), 3461-3478.
- [26] Ouahabi, A. (2013). Image Denoising using Wavelets: Application in Medical Imaging. *Advances in Heuristic Signal Processing and Applications*, 287-313.
- [27] Ouahabi, A. (2013, May). A review of wavelet denoising in medical imaging. In *2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA)* (pp. 19-26). IEEE.
- [28] Zeng, D., Veldhuis, R., & Spreuwers, L. (2021). A survey of face recognition techniques under occlusion. *IET biometrics*, 10(6), 581-606.
- [29] Fu, C., Wu, X., Hu, Y., Huang, H., & He, R. (2021). Dvg-face: Dual variational generation for heterogeneous face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 44(6), 2938-2952.
- [30] Wang, H., Gong, D., Li, Z., & Liu, W. (2019). Decorrelated adversarial learning for age-invariant face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3527-3536).
- [31] Shi, L., Wang, X., & Shen, Y. (2020). Research on 3D face recognition method based on LBP and SVM. *Optik*, 220, 165157.
- [32] Zhang, D. X., An, P., & Zhang, H. X. (2018). Application of robust face recognition in video surveillance systems. *Optoelectronics Letters*, 14(2), 152-155.

- [33] Kumar, N., & Garg, V. (2019). Single sample face recognition in the last decade: A Survey. *International Journal of Pattern Recognition and Artificial Intelligence*, 33(13), 1956009.
- [34] Min, R., Xu, S., & Cui, Z. (2019). Single-sample face recognition based on feature expansion. *IEEE Access*, 7, 45219-45229.
- [35] Wang, M., & Deng, W. (2021). Deep face recognition: A survey. *Neurocomputing*, 429, 215-244.
- [36] Vetter, T. (1998). Synthesis of novel views from a single face image. *International journal of computer vision*, 28, 103-116.
- [37] Zhang, D., Chen, S., & Zhou, Z. H. (2005). A new face recognition method based on SVD perturbation for single example image per person. *Applied Mathematics and computation*, 163(2), 895-907.
- [38] Gao, Q. X., Zhang, L., & Zhang, D. (2008). Face recognition using FLDA with single training image per person. *Applied mathematics and computation*, 205(2), 726-734.
- [39] Hu, C., Ye, M., Ji, S., Zeng, W., & Lu, X. (2015). A new face recognition method based on image decomposition for single sample per person problem. *Neurocomputing*, 160, 287-299.
- [40] Dong, X., Wu, F., & Jing, X. Y. (2018). Generic Training Set based Multimanifold Discriminant Learning for Single Sample Face Recognition. *KSII Transactions on Internet & Information Systems*, 12(1).
- [41] Deng, W., Hu, J., & Guo, J. (2012). Extended SRC: Undersampled face recognition via intraclass variant dictionary. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(9), 1864-1870.
- [42] Yang, M., Van Gool, L., & Zhang, L. (2013). Sparse variation dictionary learning for face recognition with a single training sample per person. In *Proceedings of the IEEE international conference on computer vision* (pp. 689-696).
- [43] Zhu, P., Yang, M., Zhang, L., & Lee, I. Y. (2015). Local generic representation for face recognition with single sample per person. In *Computer Vision--ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part III 12* (pp. 34-50). Springer International Publishing.
- [44] Zhu, P., Zhang, L., Hu, Q., & Shiu, S. C. (2012, October). Multi-scale Patch Based Collaborative Representation for Face Recognition with Margin Distribution Optimization. In *ECCV* (1) (pp. 822-835).
- [45] Zhang, L., Yang, M., & Feng, X. (2011, November). Sparse representation or collaborative representation: Which helps face recognition?. In *2011 International conference on computer vision* (pp. 471-478). IEEE.
- [46] Lu, J., Tan, Y. P., & Wang, G. (2012). Discriminative multimanifold analysis for face recognition from a single training sample per person. *IEEE transactions on pattern analysis and machine intelligence*, 35(1), 39-51.
- [47] Zhang, W., Xu, Z., Wang, Y., Lu, Z., Li, W., & Liao, Q. (2018). Binarized features with discriminant manifold filters for robust single-sample face recognition. *Signal Processing: Image Communication*, 65, 1-10.
- [48] Gu, J., Hu, H., & Li, H. (2017). Local robust sparse representation for face recognition with single sample per person. *IEEE/CAA Journal of Automatica Sinica*, 5(2), 547-554.

- [49] Zhang, Z., Zhang, L., & Zhang, M. (2021). Dissimilarity-based nearest neighbor classifier for single-sample face recognition. *The Visual Computer*, 37, 673-684.
- [50] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117.
- [51] Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA transactions on Signal and Information Processing*, 3, e2.
- [52] Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and trends in signal processing*, 7(3-4), 197-387.
- [53] Kim, K., & Aminanto, M. E. (2017, September). Deep learning in intrusion detection perspective: Overview and further challenges. In *2017 International Workshop on Big Data and Information Security (IWBIS)* (pp. 5-10). IEEE.
- [54] Guo, G., & Zhang, N. (2019). A survey on deep learning based face recognition. *Computer vision and image understanding*, 189, 102805.
- [55] Scherer, D., Müller, A., & Behnke, S. (2010). Evaluation of pooling operations in convolutional architectures for object recognition. In *Artificial Neural Networks–ICANN 2010: 20th International Conference, Thessaloniki, Greece, September 15-18, 2010, Proceedings, Part III 20* (pp. 92-101). Springer Berlin Heidelberg.
- [56] Coşkun, M., Uçar, A., Yildirim, Ö., & Demir, Y. (2017, November). Face recognition based on convolutional neural network. In *2017 International Conference on Modern Electrical and Energy Systems (MEES)* (pp. 376-379). IEEE.
- [57] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.
- [58] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [59] 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).
- [60] Zeng, J., Zhao, X., Qin, C., & Lin, Z. (2017, December). Single sample per person face recognition based on deep convolutional neural network. In *2017 3rd IEEE International Conference on Computer and Communications (ICCC)* (pp. 1647-1651). IEEE.
- [61] Ding, C., Bao, T., Karmoshi, S., & Zhu, M. (2017). Single sample per person face recognition with KPCANet and a weighted voting scheme. *Signal, Image and Video Processing*, 11, 1213-1220.
- [62] Zhang, Y., & Peng, H. (2017). Sample reconstruction with deep autoencoder for one sample per person face recognition. *IET Computer Vision*, 11(6), 471-478.
- [63] Du, Q., & Da, F. (2021). Block dictionary learning-driven convolutional neural networks for fewshot face recognition. *The visual Computer*, 37(4), 663-672.
- [64] Ataman, E., Aatre, V., & Wong, K. (1980). A fast method for real-time median filtering. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), 415-421.
- [65] Kannala, J., & Rahtu, E. (2012, November). Bsif: Binarized statistical image features. In *Proceedings of the 21st international conference on pattern recognition (ICPR2012)* (pp. 1363-1366). IEEE.

- [66] Benzaoui, A., Hadid, A., & Boukrouche, A. (2014). Ear biometric recognition using local texture descriptors. *Journal of electronic imaging*, 23(5), 053008-053008.
- [67] Zehani, S., Ouahabi, A., Oussalah, M., Mimi, M., & Taleb-Ahmed, A. (2021). Bone microarchitecture characterization based on fractal analysis in spatial frequency domain imaging. *International Journal of Imaging Systems and Technology*, 31(1), 141-159.
- [68] Stone, J. V. (2002). Independent component analysis: an introduction. *Trends in cognitive sciences*, 6(2), 59-64.
- [69] Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7), 971-987.
- [70] Ojansivu, V., & Heikkilä, J. (2008). Blur insensitive texture classification using local phase quantization. In *Image and Signal Processing: 3rd International Conference, ICISP 2008*. Cherbourg-Octeville, France, July 1-3, 2008. *Proceedings 3* (pp. 236-243). Springer Berlin Heidelberg.
- [71] Jiang, L., Cai, Z., Wang, D., & Jiang, S. (2007, August). Survey of improving k-nearest-neighbor for classification. In *Fourth international conference on fuzzy systems and knowledge discovery (FSKD 2007)* (Vol. 1, pp. 679-683). IEEE.
- [72] Martinez, A. M., & Benavente, R. (1998). The AR Face Database. *CVC Technical Report*, 24, 1-10.
- [73] Huang, G. B., 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*.
- [74] Mehra, N., Ali, A., & Hodayun, M. (2017). A supervised multimanifold method with locality preserving for face recognition using single sample per person. *Journal of Central South University*, 24, 2853-2861.
- [75] Ji, H. K., Sun, Q. S., Ji, Z. X., Yuan, Y. H., & Zhang, G. Q. (2017). Collaborative probabilistic labels for face recognition from single sample per person. *Pattern Recognition*, 62, 125-134.
- [76] Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1), 71-86.
- [77] Wu, J., & Zhou, Z. H. (2002). Face recognition with one training image per person. *Pattern Recognition Letters*, 23(14), 1711-1719.
- [78] Chen, S., Zhang, D., & Zhou, Z. H. (2004). Enhanced (PC) 2A for face recognition with one training image per person. *Pattern Recognition Letters*, 25(10), 1173-1181.
- [79] Yang, J., Zhang, D., Frangi, A. F., & Yang, J. Y. (2004). Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE transactions on pattern analysis and machine intelligence*, 26(1), 131-137.
- [80] Gottumukkal, R., & Asari, V. K. (2004). An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letters*, 25(4), 429-436.
- [81] Chen, S., Liu, J., & Zhou, Z. H. (2004). Making FLDA applicable to face recognition with one sample per person. *Pattern recognition*, 37(7), 1553-1555.
- [82] Zhang, D., & Zhou, Z. H. (2005). (2D) 2PCA: Two-directional two-dimensional PCA for efficient face representation and recognition. *Neurocomputing*, 69(1-3), 224-231.

- [83] Tan, X., Chen, S., Zhou, Z. H., & Zhang, F. (2005). Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble. *IEEE Transactions on Neural Networks*, 16(4), 875-886.
- [84] He, X., Yan, S., Hu, Y., Niyogi, P., & Zhang, H. J. (2005). Face recognition using laplacianfaces. *IEEE transactions on pattern analysis and machine intelligence*, 27(3), 328-340.
- [85] Zhang, D., Chen, S., & Zhou, Z. H. (2005). C based on SVD perturbation for single example image per person. *Applied Mathematics and computation*, 163(2), 895-907.
- [86] Deng, W., Hu, J., Guo, J., Cai, W., & Feng, D. (2010). Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach. *Pattern Recognition*, 43(5), 1748-1762.
- [87] Chu, Y., Zhao, L., & Ahmad, T. (2019). Multiple feature subspaces analysis for single sample per person face recognition. *The Visual Computer*, 35, 239-256.
- [88] Pang, M., Cheung, Y. M., Wang, B., & Liu, R. (2019). Robust heterogeneous discriminative analysis for face recognition with single sample per person. *Pattern Recognition*, 89, 91-107.
- [89] Seetafaceengine. (2016). Available online: <https://github.com/seetaface/SeetaFaceEngine> (accessed on 1 June 2023).
- [90] Cuculo, V., D'Amelio, A., Grossi, G., Lanzarotti, R., & Lin, J. (2019). Robust single-sample face recognition by sparsity-driven sub-dictionary learning using deep features. *Sensors*, 19(1), 146.
- [91] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2008). Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2), 210-227
- [92] Su, Y., Shan, S., Chen, X., & Gao, W. (2010, June). Adaptive generic learning for face recognition from a single sample per person. In *2010 IEEE computer society conference on computer vision and pattern recognition* (pp. 2699-2706). IEEE.
- [93] Zhou, D., Yang, D., Zhang, X., Huang, S., & Feng, S. (2019). Discriminative probabilistic latent semantic analysis with application to single sample face recognition. *Neural Processing Letters*, 49, 1273-1298.
- [94] Zeng, J., Zhao, X., Gan, J., Mai, C., Zhai, Y., & Wang, F. (2018). Deep convolutional neural network used in single sample per person face recognition. *Computational intelligence and neuroscience*, 2018.
- [95] Lv, J. J., Shao, X. H., Huang, J. S., Zhou, X. D., & Zhou, X. (2017). Data augmentation for face recognition. *Neurocomputing*, 230, 184-196.
- [96] Musa, P., Al Rafi, F., & Lamsani, M. (2018, October). A Review: Contrast-Limited Adaptive Histogram Equalization (CLAHE) methods to help the application of face recognition. In *2018 third international conference on informatics and computing (ICIC)* (pp. 1-6). IEEE.
- [97] Huang, C., & Huang, J. (2017). A Fast HOG Descriptor Using Lookup Table and Integral Image. *arXiv preprint arXiv:1703.06256*.
- [98] Adjabi, I., Ouahabi, A., Benzaoui, A., & Jacques, S. (2021). Multi-block color-binarized statistical images for single-sample face recognition. *Sensors*, 21(3), 728.

- [99] Adjabi, I. (2022, May). Combining hand-crafted and deep-learning features for single sample face recognition. In 2022 7th International Conference on Image and Signal Processing and their Applications (ISPA) (pp. 1-6). IEEE.
- [100] Adjabi, I., & Benzaoui, A. (2022, December). Single Sample Face Recognition: A Comparative Study between Texture and Deep-Learning-based Features. In 2022 IEEE 21st international Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA) (pp. 380-385). IEEE.
- [101] Coşkun, M., Uçar, A., Yildirim, Ö., & Demir, Y. (2017, November). Face recognition based on convolutional neural network. In 2017 International Conference on Modern Electrical and Energy Systems (MEES) (pp. 376-379). IEEE.
- [102] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
- [103] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).
- [104] Parida, P., & Bhoi, N. (2017). Wavelet based transition region extraction for image segmentation. *Future Computing and Informatics Journal*, 2(2), 65-78.
- [105] Hardalac, F., Yaşar, H., Akyel, A., & Kutbay, U. (2020). A novel comparative study using multi-resolution transforms and convolutional neural network (CNN) for contactless palm print verification and identification. *Multimedia Tools and Applications*, 79, 22929-22963.
- [106] Arbaoui, A., Ouahabi, A., Jacques, S., & Hamiane, M. (2021). Concrete cracks detection and monitoring using deep learning-based multiresolution analysis. *Electronics*, 10(15), 1772.
- [107] Zhang, D., & Zhang, D. (2019). Wavelet transform. *Fundamentals of Image Data Mining: Analysis, Features, Classification and Retrieval*, 35-44.
- [108] Amrouni, N., Benzaoui, A., Bouaouina, R., Khaldi, Y., Adjabi, I., & Bouglimina, O. (2022). Contactless Palmprint Recognition Using Binarized Statistical Image Features-Based Multiresolution Analysis. *Sensors*, 22(24), 9814.
- [109] Adjabi, I. (2023, July). Single sample face recognition using multi-resolution analysis of color texture features. In 2023 International Conference on Nonlinear Science and Complexity (ICNSC2023). Istanbul, Turkey..