REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Farhat Abbas Sétif

Faculté des Science

Département d'Informatique

Traitement automatique d'images de visages algorithmes et architecture

Thèse présentée par :

Farid AYECHE

En vue de l'obtention du diplôme de

Doctorat en Sciences en Informatique

Soutenue devant le jury composé de :

Fouzi	SCHEMESSEDINE	Prof. Université Ferhat Abbas Sétif 1	Président
Adel	ALTI	MCA. Université Ferhat Abbas Sétif 1	Directeur
Makhlouf	DERDOUR	Prof. Université de Oum El-bouaqhi	Examinateur
Farid	NOUIOUA	MCA. Université de B.B.A	Examinateur
Salim	BOUAMAMA	MCA. Université Ferhat Abbas Sétif 1	Examinateur
Lamri	LAOUAMER	MCA. Université Al Qassim Arabie Saoudite	Examinateur

25/04/2021

REMERCIEMENTS

Je remercie et je loue tout d'abord mon grand DIEU de m'avoir aidé à défier tous les obstacles et qui m'a donné la patience, la volonté et le courage afin de compléter ce modeste travail dans le cadre de cette thèse de recherche scientifique.

C'est avec un immense plaisir que j'exprime mes remerciements, les plus chaleureux à mon directeur de thèse Monsieur **Adel ALTI** maitre de conférences à l'université Sétif 1, qui a toujours suivi ce travail avec intérêt et avec une grande patience qui l'a témoigné par des encouragements de toutes sortes. Avec son expérience dans la recherche scientifique, avec ses conseils, j'ai pu découvrir le monde de la recherche scientifique dans le domaine du traitement d'images et des techniques de la reconnaissance faciale. Il est effectivement un grand honneur et avec un grand plaisir pour moi de travailler sous la supervision de Monsieur **Adel ALTI**. Peu importe ce que je dis, je ne peux pas lui exprimer mes sincères remerciements et ma gratitude.

Un remerciement spécial est aussi adressé à Monsieur : **Abdou Allah BOUKKERAM** professeur à l'université Abderrahmane Mira à Bejaia que Dieu ait pitié de lui. Je le remercie très sincèrement pour son aide et surtout ses encouragements durant toute la période que j'ai travaillé avec lui dans le cadre de la préparation de cette thèse. Je demande à Dieu d'avoir pitié de lui avec sa grande miséricorde et de le récompenser pour tout ce qu'il a fait pour moi.

Je tiens également à remercier profondément les membres de jury qui me font l'honneur de bien vouloir évaluer mon travail :

Monsieur, **Fouzi SCHEMESSEDINE** Professeur à l'Université de Sétif 1, pour l'honneur qu'il me fait, en acceptant la présidence de ce jury. Monsieur **Makhlouf DERDOUR**, Professeur à l'Université de Tébessa, Monsieur **Farid NOUIOUA**, MCA à l'Université de Bordj Bouarij, Monsieur **Salim BOUAMAMA**, MCA à l'Université de Sétif 1, Monsieur **Lamri LAOUAMER** MCA à l'Université Al Qassim de Arabie Saoudite, pour avoir accepté de juger mon travail.

C'est avec un grand plaisir que je réserve ces lignes comme un signe de gratitude et de reconnaissance à tous ceux qui ont contribué, de près ou de loin, au bon déroulement de mon travail. Qu'ils trouvent ici l'expression de mes sincères remerciements.

Résumé

Le potentiel de reconnaissance faciale et d'expressions faciales a suscité un intérêt accru pour les interactions sociales et l'identification biométrique. La plupart des méthodes d'identification existantes souffrent d'un consensus entre le taux de précision et le temps de calcul. L'objectif de notre travail est de construire un nouveau système de reconnaissance des visages et de leurs expressions rapide et robuste aux diverses conditions d'éclairage. Pour cela, nous avons proposé un système automatique qui permet de fournir aux utilisateurs un classificateur approprié pour identifier de manière plus précise les visages et les expressions faciales. Le système est basé sur deux étapes fondamentales de l'apprentissage automatique, à savoir l'extraction des caractéristiques et la classification des caractéristiques. Le modèle des contours actifs (Active Shape Model, ASM) se compose des points de repère est utilisé pour sélectionner des traits de visage tandis que l'algorithme de classification des visages et des expressions faciales se base sur le classificateur le plus performant parmi sept classificateurs standard. Les résultats d'expérimentations montrent que le classificateur d'analyse quadratique discriminant (Quadratic Discriminant Analysis, QDA) offre d'excellentes performances et surpasse les autres classificateurs avec un taux de reconnaissance de 94,25% sur le même jeu de données. Ensuite, nous avons proposé de combiner les descripteurs de texture LBP et LGC sous forme un nouveau descripteur appelé LGN (Local Gradient Neighborhood) pour une reconnaissance efficace des visages et des expressions faciales. Tout d'abord, l'image du visage est divisée en plusieurs blocs. Pour chaque bloc, le vecteur des caractéristiques de l'image est extrait via LGN. Puis, les classificateurs SVM et K-NN sont utilisés pour classifier les visages de différentes personnes dans des benchmarks standards JAFFE et ORL. Un taux de reconnaissance égale à 98,50% et un temps d'exécution réduit à l'aide d'un vecteur de taille réduit sont atteints. Enfin, pour avoir une reconnaissance efficace des visages et des expressions faciales avec un vecteur de caractéristiques réduit, nous avons proposé deux nouveaux descripteurs appelés histogramme du gradient directionnel (Histogram of Directional Gradient,HDG) et histogramme du gradient directionnel généralisé (Histogram of Directional Gradient Generalized, HDGG) pour extraire les caractéristiques discriminantes des expressions faciales en termes des taux de reconnaissance d'une bonne classification par rapport aux classificateurs existants. Les descripteurs proposés sont basés sur les gradients locaux directionnels combinés avec le classificateur SVM. Les résultats d'expérimentations montrent que le descripteur HDG donne un taux de précision égale à 92,12% dans tous les ensembles de données testées par rapport aux travaux existants. Il démontre également un temps d'exécution rapide pour la reconnaissance faciale allant de 0.4 s à 0.7 s dans toutes les bases de données évaluées.

Mots clés : reconnaissance des expressions faciales, vecteur de caractéristiques, histogramme de gradient directionnel, reconnaissance des visages, classificateur SVM.

Abstract

The potential of facial expression recognitions have gained increased interest in social interactions and biometric identification. Most of the existing identification methods suffer in a consensus between the identification accuracy rates and execution time. The main objective of this work is to build a new fast and robust system for recognizing faces and facial expressions that shifts lighting conditions. Therefore, we have proposed an automatic system that provides users with a well-adopted classifier for recognizing facial expressions more accurately. The system is based on two fundamental machine-learning stages, namely, feature selection and feature classification. Active Shape Model (ASM) consists of landmarks is used to select features while the face and facial expression classification algorithm depends on the top-performing classifier from seven standard classifiers. The experimental results show that the Quadratic discriminant classifier provides excellent performance and it out performs other classifiers with the highest recognition rate of 94.25% for the same dataset. Then we have combined LBP and LGC descriptors to form a novel Local Gradient Neighborhood (LGN) descriptor for both effective face and facial expression recognition. Firstly, the face image is divided into blocks. For each block, the vector of image features is extracted via LGN. Then, Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) are used to discriminate faces of different persons in JAFFE and ORL benchmarks. A recognition rate of 98.50% and reduced execution time are achieved. Finally, to build an efficient face and facial expression recognition using little training, we have proposed two new descriptors called Histogram of Directional Gradient (HDG) and Histogram of Directional Gradient Generalized (HDGG) to extract discriminant facial expressions features for better enhancing the classification accuracy and efficiency compared with existing classifiers. The proposed descriptors are based on the directional local gradients combined with SVM linear classification. The experiment results reveal significant accuracy ratios that equal to 92.12% in all experimented datasets compared with existing works. It demonstrates a fast execution time for face recognition ranging from 0.4s to 0.7s in all evaluated databases.

Keywords: Facial Expression Recognition, Feature Vector, Histogram of Directional Gradient, Generalized, LGN, Face recognition, SVM Classifier.

Préface

Les travaux présentés dans cette thèse ont été publiés dans les journaux suivants :

— Ayeche, F., Alti, A., & Boukerram, A. (2020). ***Improved Face and Facial Expression Recognition Based on a Novel Local Gradient Neighborhood***, *Journal of Digital Information Management*, 18(1), 33-45.

— Ayeche, F., & Alti, A. (2020). ***Performance Evaluation of Machine Learning for Recognizing Human Facial Emotions***, *Revue d'Intelligence Articficelle*, 34(3), 267-278. (SCOPUS IF 0.256).

— Ayeche, F., & Alti, A. (2020). ***Novel Descriptors for Effective Recognition of Face and Facial Expressions,*** *Revue d'Intelligence Articficelle*, 34(5), 521-530. (SCOPUS IF 0.256).).

— Ayeche, F., & Alti, A. ***HDG and HDGG: An Extensible Feature Extraction Descriptor for Effective Face and Facial Expressions Recognition,*** *Pattern Anal Applic (2021).*

https://doi.org/10.1007/s10044-021-00972-2

Table des matières

Liste des figures

Liste des tableaux

Liste des abréviations

LBP	Local Binary Pattern
SIFT	Scale Invariant Feature Transform
LGN	Local Gradient Neighborhood
HOG	Histogram Oriented Gradient
HDG	Histogram of Directional Gradient
HDGG	Histogram of Directional Gradient Generalized
PCA	Principal Component Analysis.
LDA	Linear Discriminant Analysis
LDP	Local Directional Pattern
LGP	Local Gradient Pattern
GDP	Gradient directional Pattern
LPQ	Local Phase Quantization
LTP	Local Ternary Pattern
GLTP	Gradient Local Ternary Pattern
WLD	Weber Local Descriptor
LGC	Local Gradient Code
LDN	Local Directional Number
KNN	k-nearest neighbors
NB	Naïve Bayesien
SVM	Support Vector Machine
QDA	Quadratic Descriminant Analysis
MLP	Multi-Layered Perceptron
RF	Random Forest

Introduction générale

1. Contexte de la thèse

Le monde connait une évolution rapide liée aux révolutions technologiques et aux développements culturels, imposé ainsi des besoins indispensables de se protéger et des mesures fiables et efficaces pour assurer et forcer la sécurité dans notre vie quotidienne. Face à ces besoins de sécurité et ainsi qu'à l'augmentation des méthodes d'escroquerie et de fraude qui ne cessent de se croitre et se développe d'une part, et d'autre part pour simplifier les différentes activités quotidiennes telles que le contrôle d'accès aux outils informatiques, l'e-commerce, les transactions bancaires, …, etc., on doit offrir des méthodes avancées et efficaces qui répondent aux attentes des individus.

Les méthodes classiques d'identification basées sur les informations à priori des personnes à savoir le mot de passe, le code PIN (Personal Identification Number) ou même celles basées sur la possession d'un objet tels que les pièces d'identités, la carte bancaire, les clefs, badge, …, etc. Tous ces moyens s'avèrent inefficaces et peuvent être volés, devinés ou même perdus. Toutes ces méthodes traditionnelles ne sont pas suffisamment fiables et ne garantissent pas la sécurité.

Pour faire face à tous ces problèmes d'inefficacité et d'insécurité, la révolution technologique fait référence à l'utilisation d'une nouvelle approche d'identification, il s'agit bien de la biométrie. Cette nouvelle tendance technologique ne se cesse de se croitre et se développe depuis son apparition au début du dernier siècle. La biométrie permet l'identification d'une personne à partir des mesures comportementales et physiques. L'avantage primordial de la technologie biométrique est d'être universelle et unique pour tous les individus, en plus elle est permanente et stable avec le temps. Les caractéristiques biométriques facilitent le mode de vie et il est impossible de les volés.

Bien que la biométrie soit une technologie émergente, en réalité, les humains ont exploité leurs caractéristiques physiques bien avant. En effet, la civilisation babylonienne a permis de sceller des accords commerciaux en utilisant l'empreinte de pousse laissée sur une poterie d'argile. Un autre exemple dans la civilisation égyptienne, où ils utilisent les caractéristiques physiques pour déterminer la différence entre ceux qui sont connu et ceux qui sont nouveaux sur le marché. Initialement la technique biométrique la plus utilisée était l'empreinte digitale pour identifier les criminels au profit des services de la police. Par la suite, d'autres techniques biométrique ont prouvé leurs grandes importances et leurs efficacités aux domaines économiques et sociaux tels que la voix, le visage, l'iris et la forme de la main. Alors que cette technologie autrefois, mieux utilisée dans les services de sécurité notamment la police, aujourd'hui, le besoin d'identification d'un individu dans multitude lieux essentiellement ceux dont la sécurité est nécessaire, par exemple pour accéder à un aéroport, dans un endroit militaire, pour entrer à son lieu de travail, pour retirer de l'argent à un distributeur automatique, …etc.

2. Problématique

La reconnaissance faciale (Facial Recognition, FR) et la reconnaissance des expressions faciales (Facial Expression Recognition, FER) sont devenues de nos jours de plus en plus très importants dans de nombreux domaines socio-économiques. En effet, ils sont impliqués dans de multiples applications de sécurité à savoir, les applications de contrôle d'accès (aux zones autorisées, ordinateurs personnels, aéroports, etc.) , les dispositifs de surveillance dans les espaces publics (tels que les stades de football, les stations et les grands centres commerciaux), les applications médico-légales (telles que la vérification et/ou gestion d'identité pour le système de justice pénale, l'identification des victimes de catastrophes), les applications d'interrogation d'identité de la personne dans les bases de données image / vidéo, les applications d'interaction homme-machine et les solutions de cartes à puce (sécurité avancée du distributeur automatique, passeport biométrique). Une autre raison de faire FR et FER plus intéressante est que les images de visage peuvent nous fournir beaucoup d'informations utiles à nos domaines d'application, y compris le sexe humain, les expressions faciales, l'âge humain, l'ethnie humaine et la direction du regard pour n'en nommer que quelques-uns. De plus étant donné que l'évolution rapide des appareils numériques photo/caméscopes, ainsi que l'apparence et les développements florissants de services de partage d'images/vidéos et réseaux sociaux sur Internet, ils ont promu et fait progresser les recherches dans le domaine de reconnaissance faciale. Parce que les objets les plus envahissants qui peuvent être trouvés dans les données image / vidéo sont les visages humains.

Les domaines de FR et FER ont été étudiés par de nombreuses méthodes de recherche qui ont été développées et appliquées par plusieurs sociétés industrielles depuis plus de 25 ans. Bien que ces méthodes aient tentés de contribuer par des solutions exploitables dans la réalité pour les systèmes FR et FER, ce domaine représente plusieurs défis et problèmes à résoudre, en particulier dans les environnements non contrôlés. Les principaux défis qui peuvent diminuer considérablement la précision et l'efficacité d'un système de reconnaissance faciale et expressions faciales sont :

- **Changement d'illuminations :** les images de visages sont généralement acquises dans des environnements d'illumination différentes, en particulier dans des situations non contrôlées ce qui affectent et dégradent énormément la performance globale des systèmes FR et FER.
- **Variation de pose :** les images faciales varient considérablement selon les ongles de pose des individus. Quand les angles de pose sont supérieurs à 45°, la précision du système est considérablement dégradée.
- **Expressions faciales :** provoquent des déformations des traits cruciaux du visage tels que les yeux, les sourcils, la bouche, le nez et affectent donc les résultats du système de reconnaissance.

- **Les conditions de vieillissement :** les visages humains évoluent énormément au fil du temps, le processus d'identification des images faciales et des expressions faciales devient à long terme un véritable défi, même pour les humains.
- **Images à faible résolution :** les caméras de surveillance se caractérisent par une qualité médiocre ainsi que de mémoire limitée et par conséquent les images seront floues avec petites taille et bien sûre cela provoque des problèmes réels pour un système de FR et FER.
- **Système multi-échelles :** Dans le cas d'une large base de données des visages, la réalisation d'un système de reconnaissance devient très complexe en termes de fiabilité et efficacité avec en temps réel est un défi reste à soulever.
- **Eclairage proche infrarouge :** les images réalisées dans le proche infrarouge sont fortement affectées par la température ambiante, la variation de pose, les expressions faciales et l'état de santé de l'individu. Ce qui rend complexe la tâche de la correspondance entre une image capturée en éclairage naturel et celle réalisées dans le proche infrarouge du système de reconnaissance.

Dans les environnements non contrôlés les problèmes mentionnés ci-dessus surviennent rarement séparément mais plutôt groupé ensemble, même pour une personne on peut trouver des images de visage extrêmement variées. Dans un système FR ou FER, on distingue généralement trois phases primordiales : la phase de détection du visage, la phase d'extraction de caractéristiques et la phase de classification des résultats. La sortie de chaque étape représente l'entrée pour l'étape suivante. Une étape de prétraitement est ajouté avant la phase de la détection, la phase d'extraction des caractéristiques ou même avant la phase de classification qu'on a besoin d'opérer une réduction des données volumineuse afin d'aboutir à un algorithme de reconnaissance fiable, efficace, robuste et précis. Généralement, les techniques les plus utilisées pour la phase :

- de détection et localisation du visage, est l'algorithme de Viola-Jones [1].
- d'extraction des traits du visage sont les ondelettes de Gabor [2], Local Binary Pattern (LBP) [3] et Histogram Oriented Gradient (HOG) [4].
- de réduction de la dimensionnalité des données on utilise la méthode Analyse en Composantes Principales (PCA) [5, 24] et le Discriminant Linéaire de Fisher (FLD) [6].
- l'identité (étiquette) de l'image du visage ou de l'expression faciale est identifiée dans l'étape de classification sur la base des vecteurs projetés résultaient de l'étape précédente. Pour réaliser la classification, les machines à vecteurs de support (*Support Vector Machine, SVM*) [7] et les k-plus proches voisins (*k Nearest Neighbors,k-NN*) [8] sont les deux choix les plus populaires.

Parmi toutes les étapes de reconnaissance faciale et expressions faciales, la phase d'extraction des caractéristiques est l'étape primordiale pour construire un système de FR et FER efficace, fiable et robuste, la phase dont laquelle le système extrait les caractéristiques discriminantes des images de visage pour construire des vecteurs de traits propres pour chaque visage. La performance d'un système FR ou FER dépend de toutes les étapes indiquées précédemment, mais l'étape la plus importante est l'extraction de caractéristiques qui détermine l'efficacité et l'effectivité des systèmes.

3. Objectives et contributions

L'objectif principal dans cette thèse est de concevoir et réaliser des nouvelles méthodes d'extraction de caractéristiques robustes pour faire face à de nombreux défis susmentionnés et suffisamment rapide pour être appliquées dans des applications du monde réel. En effet, nous avons conçus de nouvelles méthodes basées sur l'analyse des inconvénients et des avantages des approches contemporaines ainsi que sur l'exploitation des résultats des recherches obtenues sur la perception visuelle et divers algorithmes de traitement d'image, applicables à l'amélioration de l'efficacité et l'effectivité des méthodes d'extraction des traits du visage. Le travail réalisé dans le cadre de cette thèse représente un travail de recherche de la méthode de classification la plus approprié pour un système FR et FER, fiable et efficace. Les principales contributions de cette thèse sont :

- **La première contribution**

  Dans la première contribution, nous proposons un nouveau système de reconnaissance des expressions faciales qui utilise les points critiques (Landmarks) pour décrire l'expression faciale sous forme d'un vecteur de caractéristiques de landmarks. Le but principal est d'évaluer les différents classificateurs existent pour sélectionner le classificateur efficace et effectif qui convient mieux pour notre système.

- **La deuxième contribution**

  Dans la deuxième contribution, nous proposons un nouveau descripteur de la texture Local Gradient Neigaboord (LGN), fondé sur la combinaison des qualités des deux descripteurs LBP [3] et LGC [9]. Nous avons appliqué LGN dans deux axes de recherches différents FR et FER. L'objectif principal de descripteur LGN est d'augmenter le taux de reconnaissance que ce soit pour le cas FR ou FER tout en réduisant le temps d'exécution pour cibler les applications en temps réel.

- **La troisième contribution**

  Dans la troisième contribution, nous présentons deux nouveaux descripteurs HDG (Histogram of Directional Gradient) et HDGG (Histogram of Directional Gradient Generalized) qui exploitent les informations de la direction des contours, appliqués sur deux systèmes FR et FER.

Nous avons comparés nos descripteurs avec les descripteurs les plus populaires et les plus utilisés tels que LBP [3], HOG [4], LPQ [10], LDP [11], …. etc.

4. Plan de la thèse

Cette thèse a été organisée en six chapitres. Les trois premiers chapitres sont plus théoriques et consacrés à une revue de la littérature liée à nos recherches, tandis que les chapitres 4, 5 et 6 présentent nos contributions. Chaque chapitre est décrit comme suit :

- Le **chapitre 1** examine l'état de l'art des systèmes FR et FER, les différentes étapes utilisées pour les systèmes de FR et FER. Il détaille la détection de visage et les approches les plus populaires de localisation d'un visage. Ensuite il traite les méthodes d'extraction des connaissances les plus répondues dans le domaine de reconnaissance. Enfin il fait une synthèse et comparaison des différentes approches et méthodes de reconnaissances du visage.
- Le **chapitre 2** détaille les méthodes d'extraction des caractéristiques dans une image d'une façon générale et plus précisément dans une image de visage. Le chapitre concentre sur la description de la texture, la qualité d'image la plus utilisée pour exploiter et extraire l'information pertinente de l'image. Ensuite il examine les descripteurs des textures les plus utilisés pour la reconnaissance générale d'un objet et en particulier le visage.
- Le **chapitre 3** présente et étudie les approches les plus populaires et les plus utilisées pour la classification et la reconnaissance dans les domaines FR et FER.
- Le **chapitre4** décrit une nouvelle méthode pour la reconnaissance des expressions faciales en utilisant les landmarks, des points spécifiques simulent la géométrie du visage en utilisant les ASM (Active Shape Model). Ensuite on a évalué et comparé les résultats des sept classificateurs pour avoir le classificateur le plus approprié en terme de taux de reconnaissance et de temps d'exécution.
- Le **chapitre 5** décrit un nouveau descripteur LGN (Local Gradient Neighborhood) pour extraire l'information la plus pertinente en combinant entre les qualités des descripteurs LBP [3] et LGC [9]. Nous avons appliqués ce nouveau descripteur dans les deux axes de recherches FR et FER.
- Le **chapitre 6** décrit deux nouveaux descripteurs HDG (Histogram of Directional Gradient) et HDGG (Histogram of Directional Gradient Generalized) en exploitant l'information du contour d'un objet visage, en utilisant la direction du gradient. Nous avons appliqués ces descripteurs dans les deux axes de recherches FR et FER.

Partie I : Etat de L'art

Chapitre 1

La reconnaissance de visage et expressions faciales

Sommaire

1.1 Introduction

Actuellement, les systèmes d'analyse faciale, y compris la détection, la reconnaissance des visages, la reconnaissance des expressions faciales sont parmi les applications les plus pertinentes de l'analyse d'images. C'est un vrai défi de construire un système automatisé qui équivaut à la capacité cognitive humaine tout en améliorant la détection et la reconnaissance des visages. Bien que les humains soient assez intelligents pour identifier les visages connus, nous ne sommes pas assez compétents pour traiter un grand nombre de visages inconnus. Les ordinateurs, avec une mémoire et une vitesse de calcul presque illimitée, devraient surmonter les limitations humaines. La reconnaissance efficace et précise du visage et leurs expressions faciales reste un grand défi et une technologie indispensable pour un grand nombre de secteurs industriels. Cette technologie pourrait offrir de nouveaux services. Certains incluent l'interaction homme-machine, les caméras photo, la vidéo surveillance, la réalité virtuelle ou l'application des lois. Cet intérêt multi disciplinaire pousse la recherche et suscite l'intérêt de diverses

disciplines. L'analyse faciale est un sujet pertinent à la reconnaissance de formes, aux réseaux de neurones, à l'infographie, au traitement d'images et à la psychologie [12]. Ainsi, de nombreuses méthodes ont été développées [13].

Un des problèmes centraux de la reconnaissance des visages et des expressions faciales est de choisir une représentation pertinente et discriminante des images donnant naissance à des traits significatifs et fiables qui traduisent le contenu sémantique des visages et expressions faciales. Il semblerait qu'il n'existe pas des traits qui puissent modéliser parfaitement des visages tout en réalisant la classification avec une précision élevée et un temps d'exécution rapide. Les systèmes actuels tentent de combiner des traits variés pour améliorer la discrimination et la classification. Le visage et les expressions faciales comportent en effet une information riche et universelle, primordiale pour l'exploration du contenu visage. La classification et la reconnaissance des expressions faciales, encore nouvelle, utilise une méthodologie empruntée principalement à l'analyse d'images nécessitant l'extraction de descripteurs forme et texture.

1.2 Historique

Les premiers travaux sur la reconnaissance faciale ont débuté depuis les années 1950 [14]. Ils se focalisent sur la reconnaissance des expressions faciales, l'interprétation des émotions et la perception des gestes. Dès 1960, Bledsoe *et al.* [15] lancent Panoramic Research à Palo Alto, en Californie au profit du ministère de la défense de l'USA. En 1964 et 1965, Bledsoe *et al.* [16] s'intéressent à la reconnaissance faciale automatisée par ordinateurs pour reconnaître les visages humains. Depuis que ces recherches ont été financées par une agence de renseignement anonyme, peu de travaux ont été publiés. Ils ont ensuite poursuivi plus tard leurs recherches au Stanford Research Institute. Bledsoe a conçu et mis en œuvre un système semi-automatique. Certaines coordonnées de visage ont été sélectionnées manuellement par un opérateur humain, puis les ordinateurs ont utilisé ces informations pour la reconnaissance. Il a décrit les problèmes centraux de la reconnaissance faciale, liées aux changements de luminosité, à la rotation de la tête, aux expressions faciales et vieillissement. L'extraction des caractéristiques discriminantes du visage comme la taille de l'oreille ou la distance entre les yeux ont été aussi proposées [17]. Ces travaux ont décrit un vecteur de caractéristiques de taille 21. Ces caractéristiques définissent la protrusion de l'oreille, le poids des sourcils ou la longueur du nez, comme base pour reconnaître les visages à l'aide de techniques de classification des motifs.

En 1973, Fischler *et al.* [18] ont essayé d'extraire et de mesurer automatiquement des caractéristiques similaires des images faciales. Ils ont utilisé pour la reconnaissance faciale un algorithme de correspondance locale calculé depuis un score globale d'ajustement pour trouver et mesurer les caractéristiques faciales. En outre, d'autres approches qui caractérisent un visage et leurs attributs géométriques sont utilisées dans la reconnaissance des formes. Mais le premier a développé un système

de reconnaissance faciale automatisé était Kenade en 1973 [19]. Il a conçu et mise en œuvre un système de reconnaissance faciale autonome. Le système extrait automatiquement 16 paramètres faciaux. La comparaison de cette extraction automatisée à une extraction humaine ou manuelle, ne montrant qu'une petite différence. Il a obtenu de 45 à 75% d'identification correcte. Des meilleurs résultats ont été obtenus lorsque des caractéristiques non pertinentes n'étaient pas utilisées.

Dans les années 80, il y avait une diversité d'approches activement poursuivies, dont la plupart continuaient avec les tendances antérieures. Certains travaux ont tenté de mesurer les caractéristiques pertinentes pour améliorer la discrimination et la classification. Nous pouvons citer le travail du Mark Nixon [20] qui a proposé une métrique géométrique de l'espacement des yeux. L'approche de modèle correspondant a été améliorée avec des stratégies telles que les «modèles déformables». Cette décennie a également apporté de nouvelles approches. Certains chercheurs développent des algorithmes de reconnaissance faciale utilisant des réseaux de neurones artificiels [21]. La première utilisation des visages propres (Eigen Faces) dans le traitement d'images a été réalisée par Sirovich *et al.* en 1986 [22], cette technique deviendrait l'approche largement dominante dans les années ultérieures, leurs méthodes étaient basées sur l'analyse des composants principaux (Principal Component Analysis, PCA). Leur objectif était de représenter une image dans une dimension réduite sans perdre beaucoup d'informations, puis de la reconstruire [23]. Leur travail sera plus tard le fondement de la proposition de nombreux nouveaux algorithmes de reconnaissance faciale.

Les années 1990 voient le véritable lancement de la reconnaissance faciale basée sur la philosophie approche des visages propres (PCA) citée comme base des premières applications industrielles et l'état de l'art. Dans1992, Mathew Turk et Alex Pentland du MIT [24] présentent les visages propres pour la localisation, le suivie et la classification de la tête d'un sujet. Dès les années 1990, la technologie de reconnaissance du visage comme axe de recherche a reçu beaucoup d'attention, avec une augmentation notable du nombre de publications. De nombreuses approches ont été adoptées qui ont conduit à des algorithmes variées : PCA (Principal Component Analysis) [25], ICA (Independent Component Analysis) et LDA (Linear Discriminant Analysis) [26] et leurs dérivés était des plus pertinents.

1.3 Reconnaissance de visage et expressions faciales

1.3.1 Définitions préliminaires

1.3.1.1 Définition de la reconnaissance de visage

La reconnaissance faciale (Face Recognition FR) est un système qui fournit les techniques d'intelligence artificielle et les descripteurs biométriques pour comparer et analyser le visage d'une personne afin de l'identifier [27]. La reconnaissance faciale a favorisé la création et le développement avancé du Big Data, des réseaux de neurones et des GPUs.

1.3.1.2 Définition de la reconnaissance des expressions faciales

La reconnaissance des expressions faciales (Face Expression Recognition FER) est un ensemble de mécanismes, de techniques d'intelligence artificielle et des descripteurs nécessaires pour classifier des expressions faciales émotionnelles dans une seule classe d'émotion. Un ensemble standard de sept catégories d'émotions sont la *colère (AN), le dégoût (DI), la peur (FE), le bonheur (HA), la tristesse (SA) et la surprise (SU)* en tant que expressions atomiques.

1.3.2 Système de reconnaissance des visages et de leurs expressions faciales

Un système de reconnaissance faciale ou reconnaissance des expressions faciales est une application logicielle permettant l'extraction des caractéristiques des visages et de leurs expressions d'émotions et l'identification des visages qui sont classifiés automatiquement en plusieurs classes selon leurs caractéristiques de textures. Le résultat du système soit une identification des personnes ou une authentification (*i.e. vérifier l'identité de sujets qui apparaissent dans l'image ou la vidéo*) [27]. Un système de reconnaissance de visage et expressions faciales ont plusieurs cas d'usages :

1. **Suivi des fonctionnalités individuelles :** on s'intéresse au suivi de la tête. Le système suivi un visage entier ou sous partie de visage (bouche, œil, nez).
2. **Connaissance 2D :** les systèmes bidimensionnels suivent un visage et produisent l'espace où se trouve le visage.
3. **Connaissance 3D :** Les systèmes tridimensionnels effectuent une modélisation 3D du visage. Cette approche permet d'estimer les variations de pose ou d'orientation.

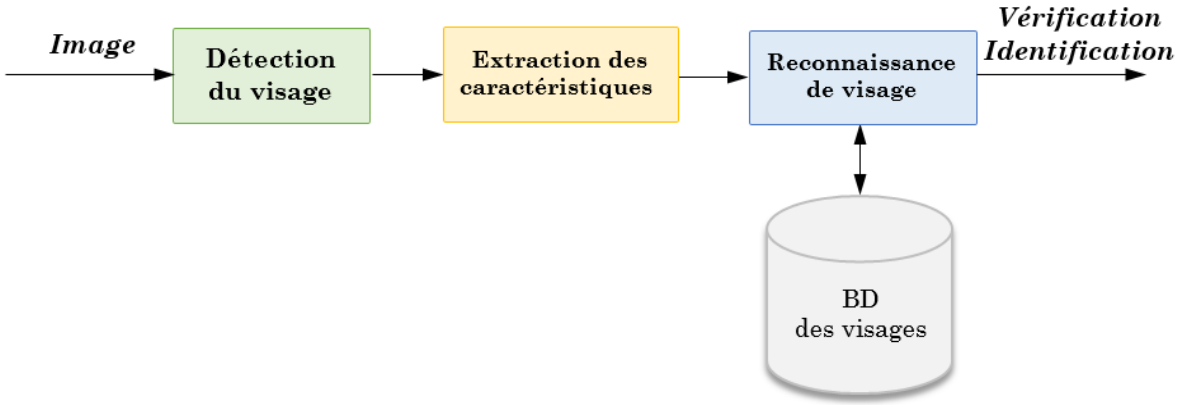

Figure 1.1 Système de reconnaisse de visage et expressions faciales [41].

1.3.3 Les étapes de la reconnaissance de visage et leurs expressions faciales

Le mécanisme de reconnaissance des visages doit suivre les étapes suivantes de l'acquisition à l'identification afin d'obtenir une identité correcte de visage.

1. **La détection de visage :** est le processus de détection, de localisation et d'extraction des visages dans une scène. Le système identifie un visage ou ses éléments facials. Ce processus adapté à de nombreuses applications comme le suivi du visage, l'estimation de pose ou la compression.
2. **L'extraction des caractéristiques :** est la technique d'extraction qui doit être capable d'obtenir une signature du visage pertinente à travers l'exploitation des caractéristiques des images. Elle permet l'extraction des caractéristiques qui peuvent être certaines régions de visage, des variations de l'illuminosité, des angles de directions ou des gradients, qui peuvent ou non être pertinents pour l'homme (par exemple, espacement des yeux). Cela permet de faire le suivi des traits du visage ou la reconnaissance des émotions.
3. **La reconnaissance de visage :** le système retrouve une identité dans une base de données vidéo ou images. Cette phase implique une méthode de comparaison, un algorithme de classification et une mesure de précision. Elle utilise des méthodes communes à de nombreux autres domaines qui font des processus de classification, ingénierie sonore, exploration de données, reconnaissance de formes, etc. Ces phases peuvent être fusionnées ou de nouvelles peuvent être ajoutées. Par conséquent, nous pourrions trouver de nombreuses approches et techniques différentes pour les problèmes de reconnaissance faciale et expressions faciales. La détection et la reconnaissance des visages peuvent être effectuées en tandem, ou procéder à une analyse d'expression avant la normalisation du visage [28].

1.3.4 Détection du visage

De nos jours, certaines applications de reconnaissance faciale ne nécessitent pas de détection de visage. Dans certains cas, les images des visages stockées dans les bases de données sont déjà normalisées. Il existe un format d'entrée d'image standard, il n'y a donc pas besoin d'une phase de détection. Un exemple de ceci pourrait être une base de données criminelle. Généralement l'agence sécuritaire stocke les visages des personnes avec un rapport criminel. S'il y une nouvelle personne, la police possède sa photo de passeport, alors la détection de visage n'est pas nécessaire. La détection de visage est un concept qui comprend de nombreux sous-problèmes. Certains systèmes détectent et localisent les visages en même temps, d'autres effectuent d'abord une détection et ensuite, s'ils sont positifs, ils essaient de localiser le visage. Ensuite, des algorithmes de suivi peuvent être nécessaires (voir figure 1.2).

Les algorithmes de détection de visage partagent généralement des étapes communes. Premièrement, une réduction de la dimension des données est effectuée, afin d'aboutir à un temps de réponse admissible. Un prétraitement peut également être effectué pour adapter l'image d'entrée au pré requis de l'algorithme. Ensuite, certains algorithmes analysent l'image telle qu'elle est, et d'autres tentent d'extraire des régions faciales pertinentes. La prochaine étape consiste généralement à l'extraction des

traits ou des mesures du visage. Ceux-ci seront alors être pondérés, évalué ou comparé pour décider s'il y a un visage. Enfin, certains algorithmes ont une routine d'apprentissage et incluent de nouvelles données à leurs modèles.

Figure 1.2 Les étapes essentielles de détection du visage.

La détection de visage est, par conséquent, un problème à deux classes où nous devons décider s'il y a un visage ou pas sur une image. Cette approche peut être vue comme un problème de reconnaissance de visage. La reconnaissance faciale doit classifier un visage donné, et il y a autant de classes que de candidats. Par conséquent, de nombreuses méthodes de détection de visage sont très similaires aux algorithmes de reconnaissance de visage. En d'autres termes, les techniques utilisées dans la détection de visage sont souvent utilisées dans la reconnaissance de visage.

1.3.5 Méthodes de détection de visage

Les chercheurs se focalisent sur deux critères de classification pour le problème de la détection des visages : soit en fonction de scenario utilisé, auquel cas les approches peuvent être regroupées en différents scénarios, ou bien selon l'algorithme de détection du visage. Ce dernier est divisé en quatre catégories : *méthodes basées sur la connaissance*, *méthodes de correspondance des modèles (Template Matching)*, méthodes *basées sur l'apparence et les méthodes basées sur la théorie des informations* (voir figure 1.3). Dans ce qui suit nous décrivons chacune de ces méthodes.

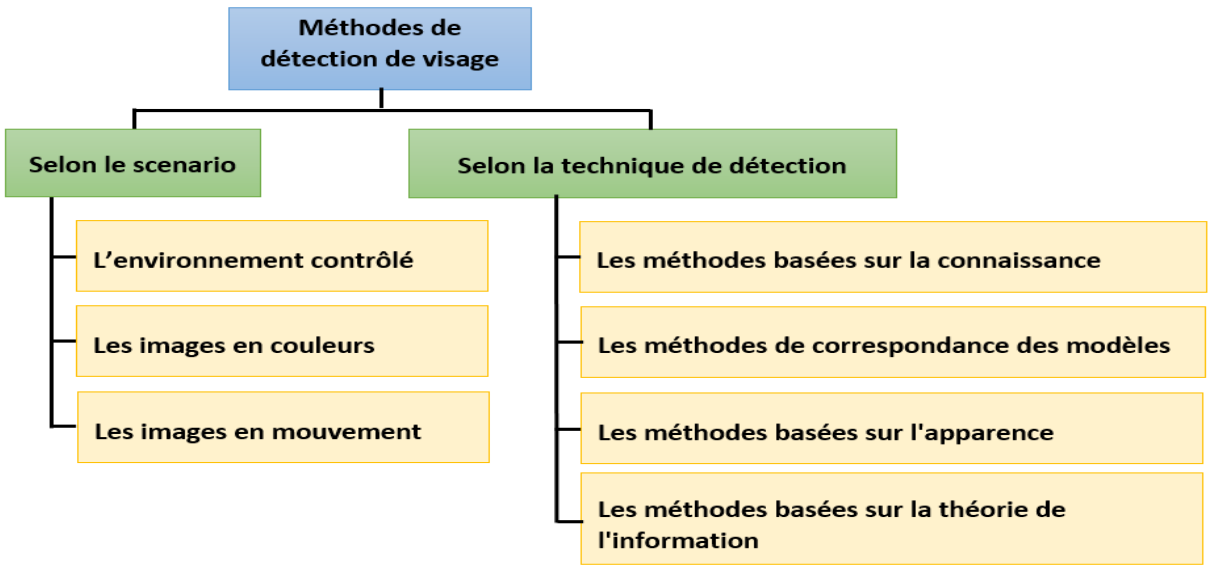

Figure 1.3 Méthodes de détection de visage.

1.3.5.1 Classification selon le scenario

Selon le scénario de la détection de visage, cette dernière peut être faite de trois types différents :

- **L'environnement contrôlé.** C'est le cas le plus simple ou les photos sont prises sous la lumière contrôlée depuis un arrière-plan. Les techniques de détection des contours peuvent être utilisées pour détecter des visages [29].
- **Les images en couleurs.** Les couleurs de la peau typique peuvent être utilisées pour détecter des visages. Ils peuvent également être faibles si les conditions d'éclairage changent. De plus, la peau change de couleur du presque blanche au presque noir. Mais, plusieurs études montrent que la différence majeure réside dans leurs intensités lumineuses, donc la chrominance est une caractéristique fondamentale. Ce n'est pas facile d'établir une représentation humaine solide de la couleur de la peau. Cependant, ils existent d'autres travaux pour améliorer la détection de visage basés sur la couleur de la peau [30].
- **Les images en mouvement.** La vidéo en temps réel donne aux développeurs la possibilité de détecter des mouvements pour localiser les visages. La plupart des systèmes commerciaux doivent tenir compte des vidéos pour localiser des visages. En effet, il y a un défi permanent pour atteindre des résultats efficaces et performants. On note un intérêt dans une approche basée sur le mouvement est la détection des yeux fermés, qui a de nombreuses utilisations outre la détection de visage [31].

1.3.5.2 Classification selon la technique de détection des visages

Les méthodes de classification selon la technique de détection des visages sont généralement divisées en quatre types : *les méthodes basées sur la connaissance*, *les méthodes d'appariement des modèles*, *les méthodes basées sur l'apparence etles méthodes basées sur la connaissance*.

- **Les méthodes basées sur la connaissance :** elles se basent sur la connaissance des différents éléments du visage et les relations entre eux. Ces relations sont traduites en un ensemble de règles. Par exemple, un visage à généralement deux yeux symétriques et le contour des yeux est plus sombre que les joues. Les traits du visage peuvent être la distance entre les yeux ou la différence d'intensité de couleur entre le contour des yeux et la zone inférieure. Le grand problème avec ces méthodes est la difficulté d'élaborer un ensemble approprié de règles. Il pourrait y avoir de nombreux faux positifs si les règles étaient trop générales. D'un autre côté, il pourrait y avoir de nombreux faux négatifs si les règles étaient trop détaillées. Une solution consiste à créer des méthodes hiérarchiques fondées sur les connaissances pour surmonter ces problèmes. Cependant, cette approche est très limitée par le fait de trouver de nombreux visages dans une image complexe. D'autres techniques sont tentées de trouver des fonctionnalités cohérentes pour la détection des visages. L'idée peut atteindre les limites de la connaissance instinctive des visages. Un premier algorithme a été développé en 1997 par Han *et al.* [32]. La méthode est divisée en plusieurs étapes. Cette méthode commence par une étape de

recherche d'un analogue oculaire pixel, de sorte qu'il supprime les pixels indésirables de l'image. Après avoir effectué le processus de segmentation, l'algorithme considère chaque segment analogue de l'œil comme un candidat d'un des yeux. Ensuite, un ensemble de règles est exécuté pour déterminer la paire potentielle d'yeux. Une fois les yeux sélectionnés, cette méthode calcule la zone du visage sous la forme d'un rectangle de délimitation. Les quatre sommets du visage sont déterminés par un ensemble de fonctions. Par conséquent les visages potentiels sont normalisés à une taille et une orientation fixes. Ensuite, les régions du visage sont vérifiées à l'aide d'une propagation arrière réseau neuronal. Enfin, l'approche applique une fonction de coût pour rendre la sélection finale. Le taux de réussite est de 94%, même sur des images avec de nombreux visages. Cette méthode se révèle fiable et efficace avec des entrées simples mais cette elle devient insuffisante lorsque on a des images contentent un homme portant des lunettes. Il existe d'autres techniques qui peuvent résoudre ce problème.

- **Les méthodes de correspondance des modèles :** ces méthodes de correspondance de modèles « Template Matching » utilisent un visage comme **«** un modèle **»**. Ces méthodes consistent à trouver un gabarit standard de tous les visages. Différentes modèles caractéristiques d'un visage peuvent être définies indépendamment. Un visage peut être divisé en yeux y compris le contour du visage, le nez et la bouche. Un modèle de visage peut également être construit par des arêtes. La détection se fait ensuite par la comparaison de ces modèles avec les candidats. Ces méthodes prouvent ses performances dans les faces frontales et non occultées. D'autres modèles utilisent les relations entre les régions de visage en termes de luminosité et d'obscurité. Cette approche est simple à mettre en œuvre, mais devient insuffisante aux variations de pose, aux changements d'échelles et formes. Cependant, les modèles déformables ont a été proposé pour résoudre ces problèmes.
- **Les méthodes basées sur l'apparence :** ces méthodes se fondent sur des techniques d'analyse statistique et d'apprentissage automatique pour trouver les traits du visage pertinents. Certaines méthodes considèrent une probabilité qu'un visage ou un vecteur des caractéristiques appartenant à une classe du visage. D'autres méthodes considèrent une fonction discriminante entre les classes faciales et non faciales. Ces méthodes sont également utilisées pour extraire les caractéristiques pour la reconnaissance faciale et seront discutées plus en détails ultérieurement.
- **Les méthodes basées sur la théorie de l'information :** les champs aléatoires de Markov (MRF) peuvent être utilisés pour modéliser les contraintes contextuelles d'un motif du visage et ses caractéristiques. Le processus de Markov maximise la discrimination entre classes (une image contient un visage ou non) en utilisant la divergence Kullback–Leibler(KL) [33]. Par conséquent, cette méthode peut être appliquée à la détection de visage.

1.3.6 Extraction des caractéristiques

Les humains peuvent reconnaître les visages dès l'âge de 5 ans. Cela semble être un processus automatique dédié àchaque cerveau [34], bien que soit une question débattu. Il est tout aussi évident que nous pouvons reconnaître l'identité des personnes qui portent une lunette ou un chapeau. On peut aussi reconnaître l'identité des personnes qui portent une barbe. Il n'est pas difficile aussi de reconnaître l'identité de notre grand-mère, bien qu'elle ait 23 ans. Tous ces processus semblent triviaux, mais ils représentent en fait un défi pour les machines. Le principal problème de la reconnaissance faciale est d'extraire la principale information pertinente et discriminative. Ce processus d'extraction de caractéristiques peut être défini comme le processus d'extraction d'informations à partir d'une image de visage. Ces informations doivent être précises pour l'étape d'identification du sujet avec un taux d'erreur faible. Le processus d'extraction de caractéristiques doit être efficace en termes de temps de calcul et utilisation des ressources de mémoire. La sortie doit être optimisée au profit de l'étape de classification pour garantir la bonne précision et le cout faible dans de grandes bases de données images.

L'extraction de caractéristiques se fait en trois étapes : **réduction de la dimensionnalité**, **extraction et sélection de fonctionnalités** (voir figure 1.4). La réduction de la dimensionnalité est une tâche essentielle dans tout système de reconnaissance de formes. Les performances d'un classificateur dépendent de la quantité d'échantillon images, du nombre des caractéristiques et de la complexité du classificateur. On pourrait penser que le taux de faux positifs d'un classificateur n'augmente pas avec l'augmentation du nombre des caractéristiques. Cependant, des caractéristiques supplémentaires peuvent dégrader les performances d'un algorithme de classification. Cela peut se produire lorsque le nombre d'échantillons d'apprentissage est petit par rapport au nombre des caractéristiques. Ce problème est appelé «malédiction de la dimensionnalité» ou «phénomène de pic». Pour résoudre ce problème, les systèmes de reconnaissance des images faciales utilisent au moins dix fois plus les échantillons d'apprentissage par classe que le nombre d'entités. Cette exigence doit être également satisfaite pendant la phase de construction d'un classificateur. Plus le classificateur est complexe, plus le taux mentionné est élevé [35]. La construction d'un classificateur est complexe en raison de la dimension importante des caractéristiques et la vitesse de calcul qui empêchent le recours aux structures gourmandes. De plus, une dimension importante des caractéristiques peut entraîner un faux positif et une mauvaise précision du système de reconnaissance lorsque ces caractéristiques sont redondantes. Enfin, compte tenu du nombre des caractéristiques disponibles, les systèmes de reconnaissance devraient les choisies soigneusement en fonction des aspects distinctifs des caractéristiques du visage.

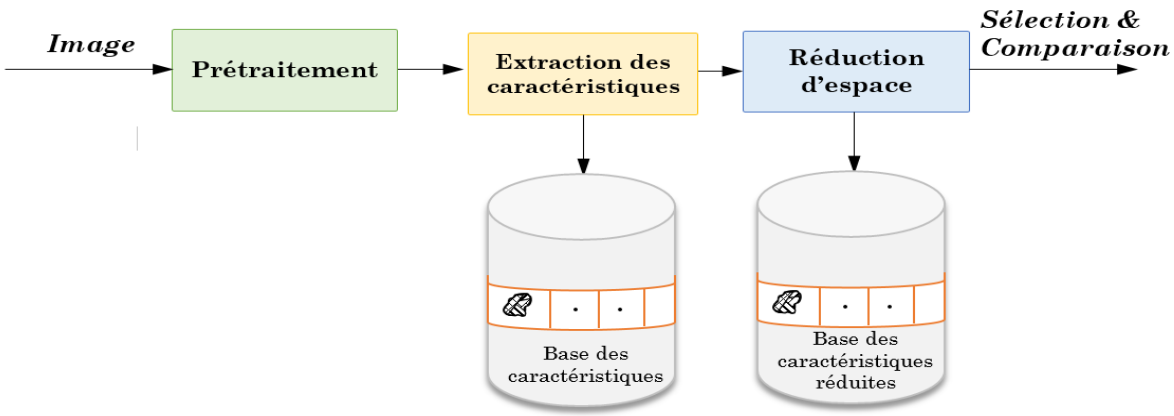

Figure 1.4 Les étapes d'extraction des caractéristiques

1.3.7 Méthodes d'extraction et de sélection des caractéristiques

1.3.7.1 Méthodes d'extraction des caractéristiques

Une méthode d'extraction des caractéristiques permet d'extraire des informations particulières de l'image faciale. Des nouvelles caractéristiques sont obtenues en appliquant des transformations ou des combinaisons des données originales. En d'autres termes, un algorithme d'extraction de caractéristiques transforme ou combine les données afin de sélectionner les meilleurs sous-espaces de l'espace originale des caractéristiques. Il existe de nombreux algorithmes permettant d'extraire les caractéristiques. Ils seront discutés plus en détails plus tard dans cette thèse (chapitre 2). On distingue trois catégories des méthodes d'extraction de caractéristiques (voir figure 1.5) :

- **Les méthodes basées sur la géométrie des visages :** ces méthodes se fondent sur une image de visage et la définition d'un modèle adéquat pour représenter ce visage. Un certain nombre de mesures ont utilisées entre les éléments de visage : les yeux, le nez, la bouche ; etc.
- **Les méthodes basées sur la couleur de la peau et la correspondance des modèles :** ces méthodes se basent sur la couleur de la peau en tant que facteur discriminative des visages. Il existe plusieurs espaces couleurs : espace RVB, espace HSV, espace XYZ et espace lab.
- **Les méthodes basées sur la connaissance :** ces méthodes s'appuient sur les heuristiques de modélisation des différentes connaissances à priori des visages. Ils extraient de manière hiérarchique les paramètres de description multi-attributs des images de visage.

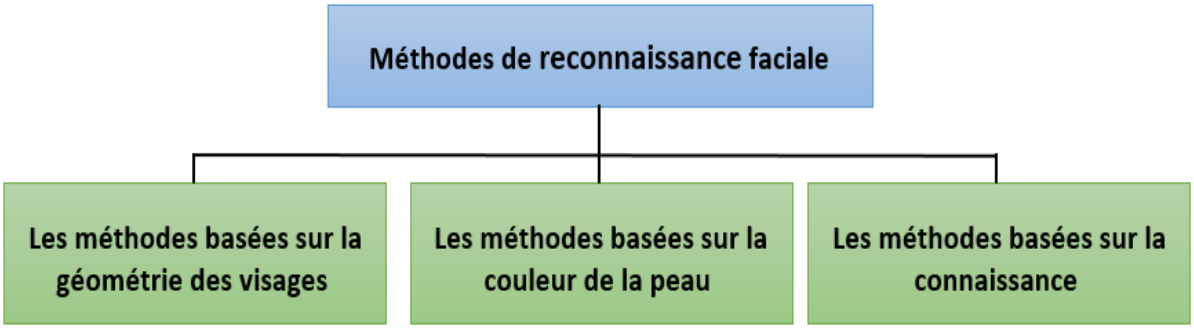

Figure 1.5 Méthodes d'extraction des caractéristiques.

1.3.7.2 Méthodes de sélection des caractéristiques

Une méthode de sélection de caractéristiques consiste à choisir le meilleur sous-ensemble de l'ensemble de caractéristiques d'entrée. Il élimine les caractéristiques non pertinentes. Souvent la sélection des caractéristiques est effectuée après l'extraction des caractéristiques. Puisque les caractéristiques sont extraites des images de visage, un sous-ensemble optimal de ces caractéristiques est choisi. Le processus de réduction de la dimensionnalité peut être intégré dans ces étapes, ou exécutées avant eux. Généralement l'approche du processus d'extraction des caractéristiques est largement adoptée dans les travaux de recherche (voir figure 1.6). La sélection des caractéristiques est un problème NP-difficile en raison de la complexité de calcul. Les chercheurs fourniront un algorithme satisfaisant, plutôt que de rechercher un algorithme optimal. Un algorithme qui sélectionne un sous-ensemble des caractéristiques le plus satisfaisant, tout en minimisant la dimensionnalité et la complexité de calcul. On va décrire quatre catégories des méthodes de sélection des caractéristiques : les méthodes basées sur la théorie de l'information [36], les méthodes basées sur les courbes [37], les méthodes basées sur les distances [38] et les méthodes basées sur les feuilles des calculs [39].

- **Les méthodes basées sur la théorie de l'information** qui utilisent le coefficient de ressemblance ou le taux de satisfaction [36] comme un critère primordial de sélection des caractéristiques dans l'Algorithme Génétique Quantique (QGA).
- **Les méthodes basées sur les courbes** qui utilisent des courbes faciales 3D. Ces courbes peuvent être explicitement analysées et comparées.
- **Les méthodes basées sur les distances** qui se basent sur un vecteur de distance entre les points d'intérêts (landmarks) de visage. Les points d'intérêts sont sélectionnés par « brute-forcing ». Divers combinaisons des points sélectionnées/non-sélectionnées sont possibles. Ils sont comparés au moment de la reconnaissance faciale par des taux de reconnaissance.
- **Les méthodes basées sur les feuille de données** qui utilisent les règles d'associations pour sélectionner les caractéristiques importantes et suffisamment discriminantes pour différentier les classes de visages.

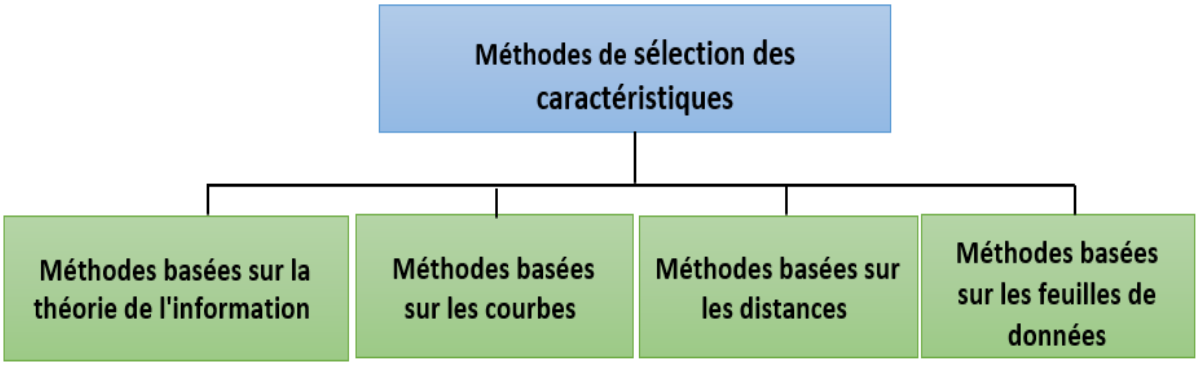

Figure 1.6 Les méthodes de sélection des caractéristiques.

1.3.8 Approches de la reconnaissance de visage et expressions faciales

Dans cette section, nous classerons ces systèmes en trois approches en fonction de leur méthodes de détection et de reconnaissance (voir figure 1.7) : (1) locale, (2) holistique (sous-espace) et (3) hybrides. La première approche utilise certaines caractéristiques du visage, sans considérer le visage entier. La seconde approche utilise le visage entier comme données d'entrée, puis se projeté dans un petit sous-espace ou dans un plan de corrélation. La troisième approche utilise des caractéristiques locales et globales dans le but d'améliorer la précision de la reconnaissance faciale.

- **Les approches locales** ne traitent que certains traits du visage. Elles sont robustes aux expressions faciales, aux occlusions et à la variation de pose [40]. L'objectif principal de ces approches est de découvrir des traits distinctifs. Généralement, ces approches peuvent être divisées en deux catégories : (1) les techniques basées sur l'apparence locale sont utilisées pour extraire les caractéristiques locales, tandis que l'image du visage est divisée en petites régions (patchs) [37]. (2) les techniques basées sur des points clés sont utilisées pour détecter les points d'intérêt dans l'image du visage, après que les caractéristiques localisées sur ces points sont extraites.
- **Les approches holistiques ou sous-spatiales** sont censées de traiter un visage entier sans nécessiter d'extraire des composants du visage (yeux, bouche, nez, etc.) ou des points caractéristiques. Le but principal de ces approches est de représenter l'image du visage par une matrice de pixels. Cette matrice est souvent convertie en un vecteur de caractéristiques pour faciliter leurs traitements. Ensuite ces vecteurs de caractéristiques sont implémentés dans un espace de petite dimension. Cependant, les techniques holistiques sont sensibles aux variations d'expressions faciales et poses et aux changements d'illuminations. Ces avantages rendent ces approches largement utilisées dans la pratique.
- **Les approches hybrides** sont basées sur des caractéristiques locales et spatiales pour bénéficier des avantages des deux techniques et d'offrir de meilleures performances pour les systèmes de reconnaissance faciale.

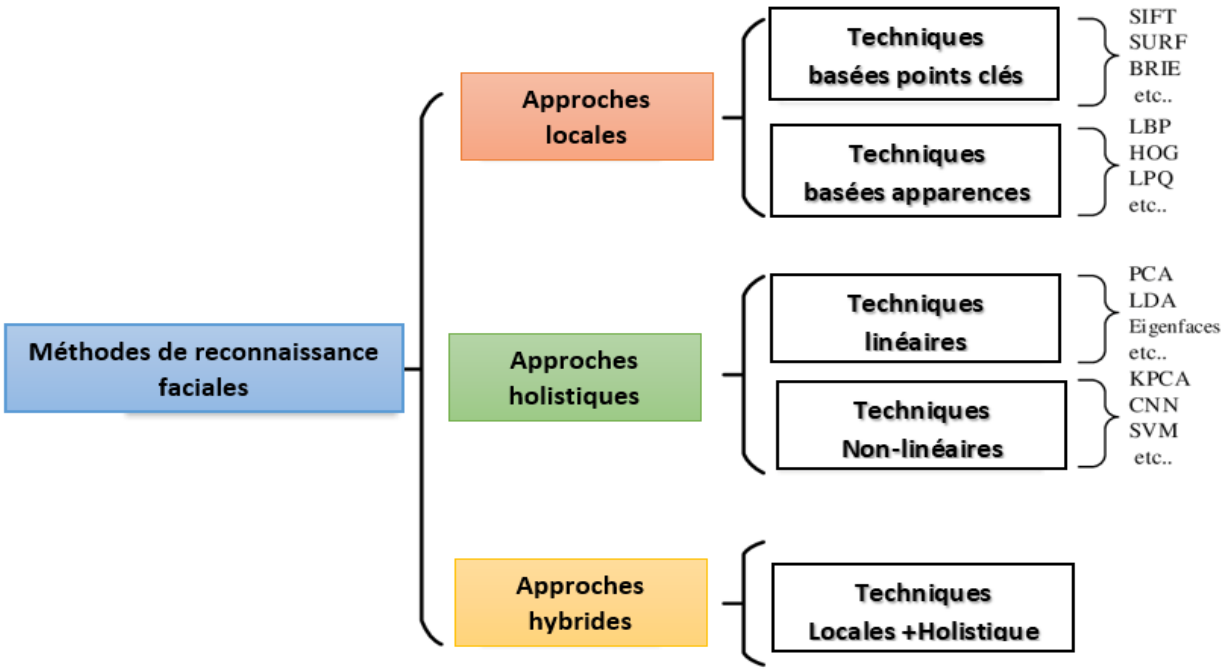

Figure 1.7 Les approches de reconnaissance de visages et expressions faciales

1.4 Défis de la reconnaissance de visage et expressions faciales

Bien que plusieurs défis à surmonter aient été détecté à chaque étape de reconnaissances de visage et des expressions faciales telles que les conditions environnementales et les contraintes matérielles, plusieurs combinaisons possibles des algorithmes de reconnaissance faciales ont été proposés pour faire face aux conditions environnementales. Cela conduit le processus de reconnaissance à donner une meilleure identification. Les défis de la détection de visage sont nombreux et dans cette section on va détailler les principaux défis qui sont communs à d'autres domaines d'apprentissages.

1.4.1 Changement d'éclairage

De nombreux algorithmes s'appuient sur les couleurs pour reconnaître les visages. Les caractéristiques sont extraites d'images en couleur, bien que certaines d'entre elles puissent être extraites d'images en niveaux de gris. En effet, la couleur dérive de la variation d'intensité lumineuse d'une couleur, noir–0% et blanc–100%. Il existe des changements d'éclairage importantes sur des images prises dans un environnement non contrôlé. Cependant, la chromacité est un facteur essentiel pendant le processus de reconnaissance de visages et le processus de reconnaissance des expressions faciales [41]. L'intensité lumineuse de la couleur des pixels peut varier considérablement en fonction des conditions d'éclairage. Un des problèmes centraux des méthodes d'extraction des caractéristiques est de choisir une représentation robustes des images faciales qui associait les mesures de changement de couleur / intensité entre les pixels. Ces mesures montrent une corrélation forte et efficace qui corrige les effets des changements d'éclairage. En plus, comme de nouvelles sources lumineuses sont apparues, des intensités lumineuses peuvent augmenter ou diminuer. D'autre part, les régions du visage entier sont

obscurcies ou ombrées, et l'extraction de leurs entités est une tâche complexe à cause de la solarisation. Cette complexité est clairement visible lorsque deux visages du même sujet sont définis avec des variations d'éclairage différentes. Une décision devrait être prise pour différencier entre eux que par rapport à un autre sujet. La figure 1.8 montre un exemple de variation d'éclairage.

Plusieurs solutions ont été proposées pour confronter le problème de changement d'éclairage. Nous pouvons citer par exemple les approches heuristiques. En effet, Il a été suggéré que dans le domaine de l'espace propre (Eigen Space), les trois premières composantes principales peuvent être jetés [42]. Ainsi, nous devons supposer que les trois premières composantes principales capturent uniquement les variations rencontrés dus aux effets d'éclairage. D'ailleurs, nous devons maintenir les performances du système avec des images de visages éclairées. D'autres travaux de recherches favorisent les approches statistiques telles que le modèle bayésien et le modèle linéaire de Fisher (LDA) afin de résoudre problème de changement d'éclairage. Certaines méthodes de reconnaissance utilisent un modèle d'éclairage afin de construire des algorithmes invariants aux variations d'éclairage.

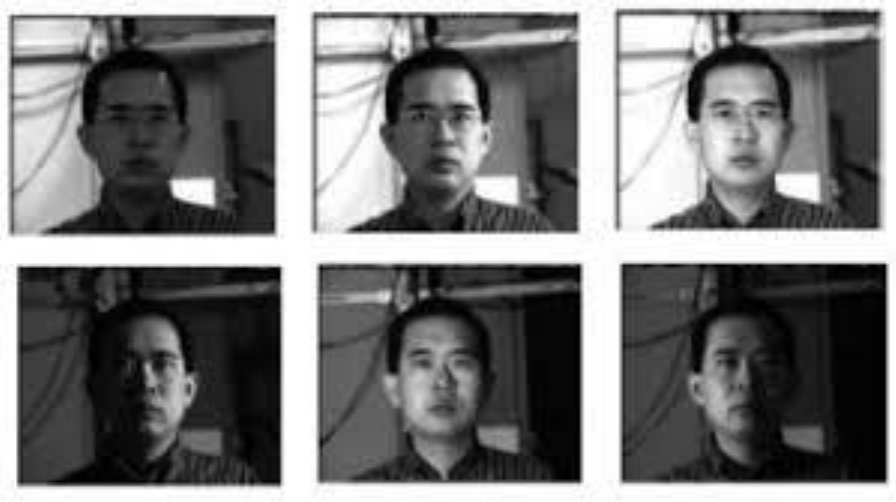

Figure 1.8 Exemple du changement d'éclairage.

1.4.2 Variations de pose

Dans le domaine de la reconnaissance du visage et la reconnaissance des expressions faciales, la variation de pose et les conditions d'éclairage sont les deux principaux problèmes rencontrés par les chercheurs. La grande majorité des méthodes de reconnaissance faciale sont basées sur les images de face frontale. Ces ensembles d'images peuvent fournir une base de recherche solide. On peut noter que 90% des travaux de recherche cités dans ce chapitre utilisent ce type de bases de données. Les méthodes de représentation d'image, les algorithmes de réduction dimensionnelle, les techniques de reconnaissance de base, les méthodes invariantes d'illumination et de nombreux autres sujets sont bien testées en utilisant des faces frontales. En revanche, les algorithmes de reconnaissance doivent respecter les contraintes des requêtes de reconnaissance faciale. Beaucoup d'entre eux, tels que la vidéo surveillance, les systèmes de sécurité vidéo ou les systèmes de divertissement à domicile en réalité augmentée,

prennent les données d'entrée d'un environnement non contrôlé. L'environnement non contrôlé comprend plusieurs obstacles à la reconnaissance faciale (ex. problème de changement d'éclairage et la variation de pose). Plusieurs variations de pose (voir figure 1.9) peuvent être résolues par des méthodes similaires. Les méthodes de traitement du visage sont classées en deux catégories :

- **augmentation d'un à plusieurs :** consiste à générer de nombreuses patchs ou images de variation de pose à partir d'une seule image afin de permettre aux systèmes FR et FER de décrire des représentations invariantes de la pose et au changement d'orientation.
- **normalisation de plusieurs à un :** consiste à récupérer la vue canonique des images faciales à partir d'une ou plusieurs images non frontales afin de permettre aux systèmes FR et FER d'exécuter comme dans des conditions contrôlées.

Figure 1.9 Exemple de variation de pose.

1.4.3 Les occlusions

Dans le domaine de la reconnaissance faciale et expressions faciales, certaines parties du visage ne peuvent pas être obtenues. Pour exemple, une photo du visage prise à partir d'une caméra de surveillance pourrait cacher derrière un poteau. Le processus de reconnaissance peut s'appuyer fortement sur la disponibilité complète d'un visage d'entrée. Par conséquent, l'absence de certaines parties du visage peut conduire à une mauvaise classification. La figure 1.10 montre un exemple d'occlusion du visage. Il existe également des objets qui peuvent obscurcir les traits du visage comme les lunettes, les chapeaux, les barbes, certaines coupes de cheveux, etc.

Figure 1.10 Exemple d'occlusion du visage.

1.4.4 Expressions faciales

Les expressions faciales sont d'une grande importance à la communication sociale car elles ont tendance à transmettre des émotions, des énergies et des expressions sans utiliser des mots. Le visage humain est capable de générer des milliers d'expressions faciales. Les approches d'apprentissage automatique du FER nécessitent un ensemble d'images d'entrainement. Chacun étiqueté avec une seule catégorie d'émotion. Un ensemble standard de sept catégories d'émotions est la colère (AN), le dégoût (DI), la peur (FE), le bonheur (HA), la tristesse (SA) et la surprise (SU) en tant que expressions atomiques. La figure 1.11 montre les sept catégories d'émotions. Bien que les expressions d'émotions offrent des avantages pour les systèmes d'interaction homme-machine, la classification des émotions humaines peut être une tâche difficile pour les machines. Cependant, un meilleur compromis entre précision et temps d'exécution pose toujours des problèmes pendant l'extraction des caractéristiques des émotions humaines.

Figure 1.11 Exemple de variation des expressions faciales.

1.4.5 Les conditions d'acquisition des images

Un système de reconnaissance faciale doit être conscient du format dans lequel des images d'entrée sont fournies. Il existe différents caméras, avec différentes fonctionnalités, faiblesses et problèmes. La plupart des systèmes de reconnaissance incluent une étape de prétraitement qui résout ce problème.

1.5 Etude des principaux travaux dans le domaine de la reconnaissance de visage et expressions faciales

De nombreuses travaux de recherches sur la reconnaissance faciale sous les environnements contrôlés ont été proposés, mais peu de travaux sur les environnements non contrôlés. Le problème reste ouvert en raison de grandes changements de luminosité, des expressions faciales, de l'âge, du fond dynamique, etc. Dans cette section, nous essayons d'aborder les différentes techniques de reconnaissance faciale les plus avancées proposées dans des environnements contrôlés / non contrôlés en utilisant différentes bases de données. Plusieurs systèmes sont mis en œuvre permettant d'identifier ou de vérifierd'une personnevia une base de données images 2D ou 3D.

1.5.1 Les techniques basées sur l'apparence

Les techniques basées sur l'apparence sont des techniques géométriques, également appelées techniques fonctionnelles ou techniques analytiques. Dans ces techniques, l'image du visage est représentée par un ensemble de vecteurs distinctifs de faibles dimensions ou de petites régions (*patchs*). Les techniques basées sur l'apparence locale se concentrent sur les éléments critiques du visage tels que le nez, la bouche et les yeux pour générer plus de détails. Aussi, il prend en compte la particularité du visage comme forme naturelle pour identifier et utiliser un nombre réduit de paramètres. De plus, ces techniques décrivent les caractéristiques locales à travers les orientations des pixels [4], les histogrammes [43], les propriétés géométriques et les plans de corrélation [44].

Dans un premier travail, Ahonen*et al.* [3] ont utilisé le modèle binaire local (Local Binary Pattern LBP) pour extraire des vecteurs de caractéristiques texture du visage. Dans la littérature, les descripteurs LBP sont considérés comme un descripteur excellent de texture couramment utilisés pour la reconnaissance faciale 2D. LBP combiné avec plusieurs autres descripteurs augmente la performance du système. Cependant, le problème de LBP est celui des classificateurs basés sur des caractéristiques, c'est-à-dire que le vecteur de caractéristiques de haute dimension conduit à un coût élevé. Le résultat obtenu est affecté par les changements de rotation, ce qui est confondu avec le processus de classification.

Hussain *et al.* [45] ont utilisé la technique de quantification de phase (Local Phase Quantization LPQ) pour extraire la propriété d'invariance du spectre de phase à partir de la transformée de Fourier à court terme (STFT). Cette technique produit un code LPQ invariant au flou pour chaque pixel de l'image.

Karaaba et *al.* [46] ont proposé un histogramme de gradient orienté (Histogram of Oriented Gradients HOG) pour décrire la forme des visages et ses contours externes. Le principe de ce travail consiste à diviser l'image entière du visage en régions. L'histogramme de direction de contour de pixel ou de gradients de direction est généré pour chaque région. Enfin, les histogrammes des régions sont combinés pour extraire les caractéristiques des contours présentes dans l'image du visage. La moyenne des distances minimales entre une image de test et une image de référence est utilisée pour la

reconnaissance de visage. Cependant, cette technique montre une faible précision de la reconnaissance faciale.

Arigbabu et *al.*[47] ont amélioré la précision de la reconnaissance à l'aide du descripteur PHOG (Pyramid ofHistogram Oriented Gradient) et le classificateur SVM. Chaque descripteur PHOG est transmis au classificateur SVM pour l'identification de visage. Les expériences menées ont reconnu l'efficacité supérieure de la méthode proposée par rapport à des méthodes similaires. Néanmoins, il requiert beaucoup de temps du calcul.

Leonard *et al.* [48] ont introduit la technique du corrélateur de Vander Lugt pour calculer le degré de similitude entre l'image test et les images de référence. La reconnaissance se fait par pic de corrélation. L'image du visage est traitée par la transformée de Fourier rapide. Cette technique montre une capacité élevée de discrimination, une robustesse au bruit, une invariance de décalage et un parallélisme inhérent

Napoléon *et al.* [49] ont introduit un nouveau système d'identification et de vérification des champs basé sur une modélisation 3D optimisée dans différentes conditions d'éclairage afin d'assurer la reconstruction des visages dans différentes poses. Le modèle de forme actif (Active Shape Model ASM) [50] est construit pour détecter un ensemble de points clés sur le visage comme elle montre la figure 1.12. Le corrélateur Vander Lugt est adopté [51] pour réaliser l'identification du visage et le descripteur LBP [3] est utilisé pour optimiser les performances d'une technique de corrélation dans différentes conditions d'illuminosité. Les expériences sont réalisées sur la base de données PHPID (Pointing Head Pose Image Database) avec une élévation allant de -30 ° à + 30 °.

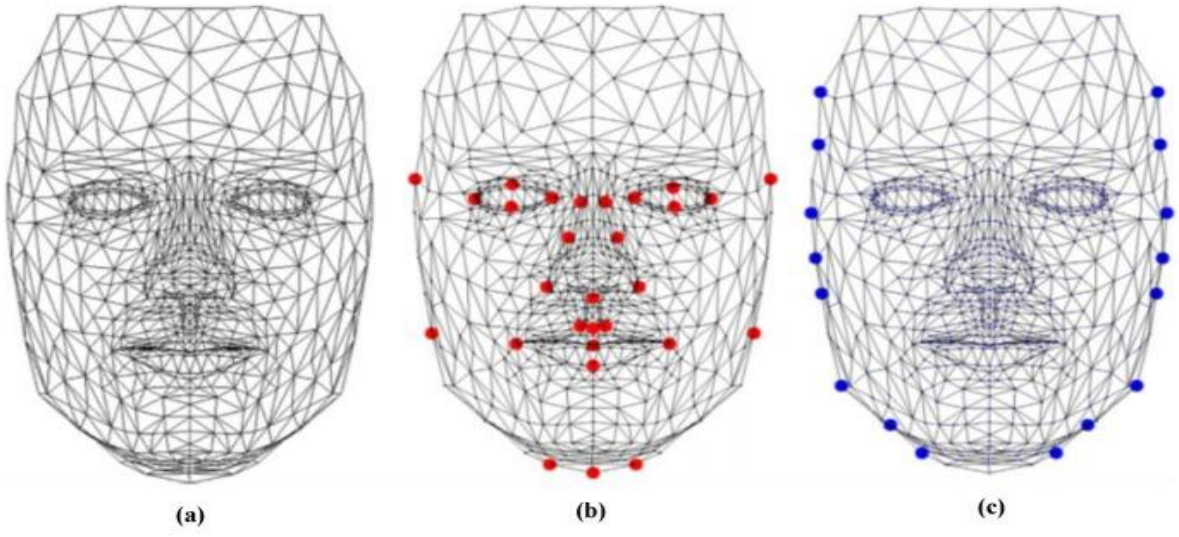

Figure 1.12 (a) Création du visage 3D d'une personne (b) résultats de la détection de 29 repères d'un visage à l'aide du modèle de forme active, (c) résultats de la détection de 26 repères d'un visage [49].

1.5.2 Les techniques basées sur les points clés

Ces techniques détectent d'abord les traits du visage (la distance entre les yeux, la largeur de la tête, etc.) puis extraient les caractéristiques [52]. La détection des traits du visage se base sur les performances des détecteurs afin de déterminer les éléments clés de l'image du visage. L'extraction des caractéristiques se base sur la représentation directe des informations véhiculées par les éléments clés de l'image du visage. Bien que ces techniques puissent résoudre les parties manquantes du visage et les occlusions. L'extraction des caractéristiques utilise, par ailleurs, les techniques de transformation SIFT (Scale Invariant Feature Transform), BRIEF (Binary Robust Independent Elementary Features) et SURF (Speeded-Up Robust Features).

Lenc *et al.* [53] ont développé un framework de détection des points clés basé sur le descripteur SIFT. L'image est convertit en un ensemble de points d'intérêt. Ces points contiennent les informations caractéristiques de l'image du visage. Les auteurs utilisent la technique SIFT en combinaison avec une approche Kepenekci pour la reconnaissance faciale. Ce travail présente de nombreux avantages notamment celui de l'invariance de l'échelle et de la rotation. D'ailleurs le framework soufre du temps de mise en correspondance des points critiques. La figure 1.13 montre un exemple de calcul des points SIFT.

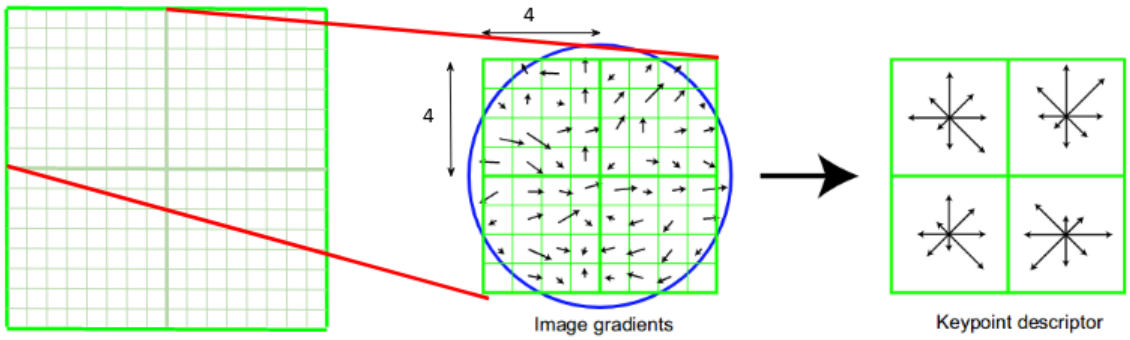

Figure 1.13. Calcul des points SIFT [53].

Işık *et al.* [54] ont combiné SURF avec une approximation du déterminant de Hesse. Ils utilisent SURF pour définir les points d'intérêt et trouver le maximum d'une approximation de la matrice de Hesse en utilisant des images intégrales. La technique proposée réduire considérablement le temps du calcul du traitement. Cependant, comparées aux descripteurs basés sur SIFT, elles sont moins précises dans la reconnaissance faciale.

Calonder *et al.* [54] ont proposé un descripteur BRIEF (Binary Robust Invariant Scalable Keypoints) basé sur les différences entre les intensités des pixels qui sont similaires à la famille des descripteurs binaires tels que BRISK et FREAK (Fast Reatina Keypoints) en termes d'évaluation. Ce travail adoucit d'abord les zones du visage puis utilise les différences d'intensité de pixel pour

représenter le descripteur BRIEF. Ce descripteur à des meilleures performances et une meilleure précision dans la reconnaissance de formes.

1.5.3 Les techniques linéaires

Une des techniques les plus utilisés pour la reconnaissance faciale est la technique d'analyse en composantes principales (ACP ou Eigenfaces) [11] qui joue un rôle important au calcul des vecteurs propres de la matrice de covariance et représente un visage par un vecteur de caractéristiques de petite dimension. Le but principal de PCA est la réduction de dimensionnalité de l'espace de données (*variables observées*) à la plus petite dimensionnalité intrinsèque de l'espace des caractéristiques (*variables indépendantes*).

Une autre technique dite « Fisherface» qui permet de réduire l'espace de représentation d'image à dimension élevée par la technique d'analyse discriminative linéaire (LDA) [57]. Cette technique de classification supervisée basée sur l'analyse discriminante linéaire, en général, à travers un système manipulant une base d'images étiquetées. Pour toutes les échantillons réparties en classes, la matrice de diffusion intra-classe S_W et la matrice de diffusion interclasses S_B sont définies comme suit :

$$S_B = \sum_{I=1}^{C} M_i \, (x_i - \mu) \, (x_i - \mu)^T \quad (1.1)$$

$$S_W = \sum_{I=1}^{C} \sum_{x_k \in X_i} M_i \, (x_k - \mu) \, (x_k - \mu)^T \quad (1.2)$$

Où

- μ : représente le vecteur moyen des échantillons appartenant à la classe i,
- X_i : représente l'ensemble des échantillons appartenant à la classe i avec x_k étant l'image numérique de cette classe,
- c : nombre de classes distinctes,
- M_i : nombre d'échantillons d'apprentissage dans classe i.
- S_B : dispersion des entités autour de la moyenne globale pour toutes les classes de visage
- S_W : dispersion des entités autour de la moyenne de chaque classe de visage.

Le but est de maximiser le rapport $det|S_B| \, / \, det|S_W|$, c'est-à-dire de minimiser S_W tout en maximisant S_B. La figure 1.14 montre les cinq premiers Fisher faces obtenus à partir de la base de données ORL [56].

Ainsi, on retrouve la technique d'analyse en composantes indépendantes (ICA) [58] qui permet de calculer les vecteurs de base d'un espace donné. Cette technique qui joue un rôle important dans la réalisation d'une transformation linéaire afin de réduire la dépendance statistique entre les différents vecteurs de base, permettant une analyse facile de composantes indépendantes. En outre, des images provenant de différentes sources sont recherchées dans des variables non corrélées, ce qui conduit à une

grande efficacité, car ICA acquiert des images au sein de variables statistiquement indépendantes. La figure 1.14 montre les cinq Eigen faces et Fisher faces construites à partir de la base de données ORL [56].

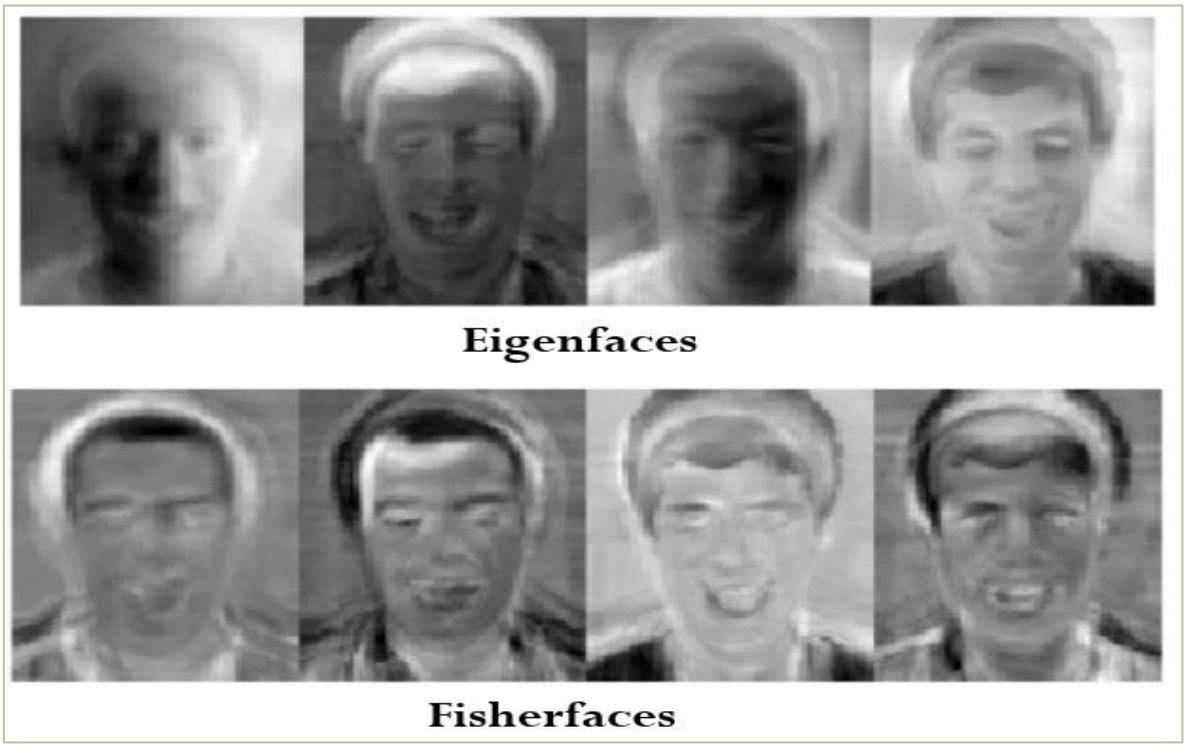

Figure 1.14 Eigen faces vs. Fisher faces [56].

Les systèmes de reconnaissances faciales actuels tentent de développer des nouveaux algorithmes pour améliorer les techniques linéaires. Cui *et al.* [59] proposent une nouvelle méthode de descripteur de région de visage spatial (SFRD) pour extraire la région de visage et traiter la variation de bruit. La méthode consiste à diviser les images du visage en régions de plusieurs niveaux spatiaux pour extraire les caractéristiques de fréquence (TF) existant dans chaque région**.** Ensuite, en regroupant les coefficients de reconstruction sur les patchs dans chaque région. Enfin, le SFRD des images de visage est extrait en appliquant une variante de la technique PCA appelée **«analyse en composantes principales blanchies (WPCA)»** pour réduire la dimension des traits et supprimer le bruit dans les principaux vecteurs propres. De plus, Prince *et al.*[60] ont proposé une extension de la méthode LDA nommée Analyse Discriminante Linéaire Probabiliste (PLDA) pour trouver des directions dans l'espace qui ont une discriminabilité maximale, et donc les plus appropriées à la fois pour la reconnaissance faciale et la reconnaissance frontale du visage sous différentes poses. Perlibakas *et al.* [61] ont élaboré un nouvel algorithme hybride basé sur le filtre de Gabor et la méthode d'analyse en composantes principales (PCA). L'algorithme Gabor-PCA est appliqué pour supprimer les redondances et obtenir la meilleure description des images de visages. Un métrique cosinus est utilisé pour évaluer la similitude entre l'image test et les images de référence.

D'autres techniques linéaires utilisées pour la reconnaissance faciale sont les techniques fréquentielles, nous pouvons citer la transformée de Fourier discrète (DFT) [62], la transformée en ondelettes discrète (DWT) [63] et la transformée en cosinus discrète (DCT) [64]. Ces techniques offrent une représentation du visage humain en termes d'énergie dans les régions à haute fréquence.

1.5.4 Les techniques basées sur l'apprentissage automatique

Ces techniques sont fondées sur les techniques d'analyse statistique et l'apprentissage automatique pour la reconnaissance faciale. Ces techniques considèrent une image ou un vecteur de caractéristiques comme une variable aléatoire avec une certaine probabilité d'appartenir à une classe du visage. D'autres techniques définir une fonction discriminante entre les classes faciales et non faciales. Ces méthodes sont également utilisé dans l'extraction de caractéristiques pour la reconnaissance faciale et sera discuté plus en détails plus tard. On présente ci-dessous les méthodes et les systèmes de reconnaissance les plus pertinents.

Sirovich *et al.* [65] ont développé une méthode pour représenter efficacement des images faciales à l'aide de la technique d'analyse en composantes principales (ACP). Leur objectif est de modéliser un visage comme un système de coordonnées. Les vecteurs qui composent ce système de coordonnées ont été référencés en tant qu'images propres. Plus tard, Turk *et al.*[24] ont utilisé cette approche pour développer un algorithme de reconnaissance basé sur les faces propres. Osuna *et al.*[66] ont appliqué pour la première fois le classificateur SVM avec succès à la détection du visage.

Yang *et al.*[67] ont utilisé un réseau clairsemé de Winnows (SNoW)pour la détection du visage. Ils ont défini un réseau clairsemé de deux unités linéaires ou de deux nœuds cibles, l'un pour les motifs faciaux et l'autre pour les motifs non-faciaux. Le SNoW avait un espace de caractéristiques appris de manière incrémentale. Les nouveaux cas étiquetés étaient des exemples positifs pour une cible et négatifs pour une autre. Ce système capable de détecter un visage en temps réel.

Moghaddam *et al.*[68] ont introduit un algorithme de reconnaissance d'objets qui a été modélisé et évalué par un classificateur bayésien. Les auteurs ont calculé la probabilité qu'un visage soit présent dans l'image en comptant la fréquence d'apparition d'une série de motifs sur les images d'entrainements. Ils ont mis l'accent sur des modèles comme l'intensité autour des yeux. Le classificateur a capturé les statistiques d'apparence et la position du visage dans le monde visuel. Dans l'ensemble, leur algorithme a montré de bons résultats dans la détection frontale du visage. Les classificateurs Bayes ont également été utilisé comme partie intégrante pour d'autres algorithmes de détection.

Bevilacqua *et al.* [69] ont utilisé un modèle statistique a été utilisé pour détecter le visage. Le défi consiste à établir un HMM (Hidden Markov Model) approprié afin que la probabilité puisse être fiable. Les états du modèle seraient les traits du visage, qui sont souvent définis comme des bandes de pixels. Généralement, les probabilités de transition entre états sont les frontières entre ces bandes de pixels. Les HMMs sont utilisés en conjonction avec d'autres méthodes pour construire des algorithmes de détection.

Arashloo *et al.* [70] ont proposé un classificateur d'analyse discriminante du noyau binaire non linéaire (CS-KDA) basé sur l'analyse discriminante du noyau de régression spectrale. D'autres techniques non linéaires ont également été utilisées dans le cadre de la reconnaissance faciale : transformée en ondelettes (WT) et réseaux de neurones cellulaires (CNN) [71], l'analyse discriminante de Kernel Fisher (KFD) et KPCA [54], l'enrobage localement linéaire (LLE) et l'analyse discriminative linéaire (LDA) [73].

1.5.5 Les techniques hybrides

Fathima *et al.* [74] ont proposé une approche hybride combinant l'ondelette de Gabor et l'analyse discriminante linéaire (HGWLDA) pour la reconnaissance faciale. L'image du visage en niveaux de gris est approximée et réduite en dimension. Les auteurs ont convolué l'image du visage en niveaux de gris avec un ensemble de filtres Gabor avec différentes orientations et échelles. Ensuite une technique 2D-LDA est appliquée pour maximiser la distance interclasses et minimiser la distance intra-classes. Pour trouver la classe à laquelle appartient toute nouvelle image faciale, le classificateur applique k-plus proche voisin (KNN). Une distance entre le vecteur de caractéristiques de la requête avec celles des images de la base d'images est calculée, les images le plus proches à l'image de test fourni par l'utilisateur sont extraites. Les résultats expérimentaux montrent la robustesse de cette approche dans différentes conditions d'éclairage.

Barkan *et al.* [75] ont proposé une nouvelle variante du descripteur LBP nommée OCLBP «Opposite Color Local Binary Patterns» qui améliore les performances dela reconnaissance faciale. OCLBP est une représentation à plusieurs échelles du descripteur LBP. De plus, OCLBP est utilisée pour réduire les représentations à haute dimensionnalité. La normalisation de covariance intra-classe (WCCN) est la technique d'apprentissage métrique qui a été utilisée pour la reconnaissance faciale.

Juefei *et al.* [76] ont développé une technique de reconnaissance faciale robuste à l'alignement périoculaire à échantillon unique basée sur le descripteur WLBP (Walsh-Hadamard Local Binary Pattern). Cette technique utilise un seul échantillon par classe de sujet et génère de nouvelles images faciales sous une large gamme de rotations 3D à l'aide de modèle élastique générique 3D. Formellement, cette technique est prouvé précis et moins complexe. La base de données LFW est utilisée à propos de l'évaluation des images faciales 3D et la méthode proposée domine les algorithmes de pointe sous quatre protocoles d'évaluation avec une précision élevée de 89,69%.

Yan *et al.* [77] ont proposé une technique d'extraction de caractéristiques efficace pour une reconnaissance de visage robuste, appelée banques de filtres de corrélation de visage multi-régions (MS-CFB). MS-CFB extrait les caractéristiques locales au niveau des régions prédéfinies du visage. Ensuite les diverses caractéristiques locales sont concaténées pour donner des sorties de corrélation globale optimales. Cette technique réduit la complexité, atteint des taux de reconnaissance plus élevés et offre une meilleure représentation des caractéristiques pour la reconnaissance faciale par rapport à de nombreuses techniques avancées sur diverses bases de données de visages publiques.

Simonyan *et al.* [78] ont développé une nouvelle méthode de reconnaissance faciale basée sur le descripteur SIFT et les vecteurs de Fisher. Les auteurs proposent une réduction de dimensionnalité discriminante en raison de la forte dimensionnalité des vecteurs de Fisher. Ensuite ces vecteurs sont projetés linéairement dans un sous-espace de faible dimension. Le but principal de cette méthode est de caractériser une image par des vecteurs de caractéristiques SIFT et Fisher pour améliorer les performances sur les ensembles de données LFW dans des conditions restreintes et non restreintes.

Ding *et al.* [79] ont proposé un nouveau framework par fusion multimodale profond du visage (MM-DFR) basé sur la technique des réseaux de neurones convolutifs (CNN). A partir de l'image du visage holistique originale, la face frontale est rendue par un modèle de visage 3D et des patchs d'image échantillonnés uniformément. Le modèle de visage 3D représente respectivement les traits du visage holistiques et les traits locaux du visage. Le framework MM-DFR ont proposé de combiner deux techniques : une technique CNN d'extraction des caractéristiques et une technique de codage automatique empilé à trois couches (SAE) est utilisée, qui permet de compresser les caractéristiques profondes de haute dimension en une signature faciale compacté. La base de données LFW est utilisée pour évaluer les performances d'identification du MM-DFR.

Moussa *et al.* [80] ont développé un système de reconnaissance faciale rapide basé sur les techniques DCT et PCA. La technique de l'algorithme génétique (GA) est utilisée pour extraire les traits du visage, ce qui permet de supprimer les traits non pertinents et de réduire le nombre de traits. De plus, la technique DCT–PCA est utilisée pour extraire les caractéristiques et réduire la dimensionnalité. La distance euclidienne minimale (ED) est utilisée pour la décision. Différentes bases de données de visages sont utilisées pour démontrer l'efficacité de ce système.

Mian *et al.* [81] ont présenté un système de reconnaissance faciale multimodal (2D et 3D) utilisant l'appariement hybride pour l'efficacité et la robustesse des expressions faciales. Cette technique est basée sur la transformation Hotelling qui corrige automatiquement la pose d'une face 3D en utilisant sa texture. Ensuite, afin de former un classificateur de rejet, une nouvelle représentation de visage sphérique (SFR) 3D avec le descripteur SIFT est utilisée, qui fournit une reconnaissance efficace dans de grandes galeries en éliminant un grand nombre de visages des candidats. Un algorithme du plus proche point itéré (ICP) est utilisé pour la décision de la classe pour une image faciale. Ce système preuve ses résultats d'expérimentations et sa robustesse sur une base d'images complète FRGC v2. Il atteint un taux de vérification de 98,6% et un taux d'identification de 96,1%.

Cho *et al.* [82] ont proposé un système de reconnaissance de visage hybride efficace basé sur les caractéristiques holistiques et locales de l'image. La technique PCA est utilisée pour réduire la dimensionnalité et la technique de séquence locale d'histogramme de motif binaire de Gabor (LGBPHS) pour réduire la complexité causée par les filtres de Gabor afin d'améliorer la qualité de reconnaissance. Les résultats expérimentaux montrent un meilleur taux de reconnaissance par rapport aux techniques

d'ondelettes PCA et Gabor sous différentes conditions d'éclairage. La base de données Extended Yale Face Database B été utilisée pour démontrer l'efficacité de ce système.

Sing *et al.* [83] proposent une nouvelle technique hybride pour la représentation et la reconnaissance des visages, qui exploite à la fois les caractéristiques locales et spatiales. L'image entière est divisée en plusieurs régions et les caractéristiques locales sont associées à chaque région, tandis que les caractéristiques globales sont extraites directement de l'image entière. Ensuite les techniques de discrimination linéaire PCA et de Fisher (FLD) sont appliquées sur le vecteur de caractéristiques pour réduire la dimensionnalité. Les résultats obtenus à partir des bases de données de visage CMU-PIE, FERET et AR sont satisfaisants.

Kamencay *et al.* [84] développent une nouvelle méthode de reconnaissance faciale basée sur les techniques SIFT, PCA et KNN. Le détecteur Hessian – Laplace ainsi que le des descripteurs SPCA sont appliqués pour extraire les caractéristiques locales du visage. Le classificateur KNN est utilisé pour identifier les visages humains les plus proches à partir des fonctionnalités entraînées. Cette technique atteint un taux de reconnaissance de 92% pour la base de données ESSEX non segmentée et de 96% pour la base de données segmentée avec 700 images d'entrainement.

1.5.6 Les techniques basées sur le Deep Learning

Depuis 2014 et avec l'apparition de l'apprentissage en profondeur du visage (Deep Face Recognition), le domaine de reconnaissance faciale a connu une révolution réelle. En effet, le DeepFace a remodelé les axes de recherches *FR* et *FER*. Le DeepFace (DF) applique plusieurs couches de traitement pour apprendre des représentations de données avec plusieurs niveaux d'extraction de traits du visage. DF s'appuie sur une architecture hiérarchique pour assembler les pixels en une représentation invariante des visages. Il a considérablement amélioré les performances et favorisé le succès des applications du monde réel. Inspirées par le succès retentissant du le défi Image Net, les architectures typiques CNN, telles que AlexNet, VGGNet, GoogleNet, ResNet et SENet, sont introduites et largement utilisées comme un modèle de base en RF et FER. Les architectures actuels CNN ou encore les architectures nouvelles conçues pour FR et FER sont adoptéespour améliorer les performances. Combinées avec des réseaux de base, les méthodes de RF et FER forment souvent des réseaux assemblés avec des entrées multiples ou des tâches multiples. Un réseau est destiné à un type particulier d'entrée ou à un type particulier de tâche.

Sun *et al.* [85] proposent une approche hybride d'apprentissage profonde appelée CNN – LSTM – ELMafin d'obtenir une reconnaissance séquentielle de l'activité humaine (HAR). Leur structure CNN – LSTM – ELM proposée est évaluée à l'aide de l'ensemble de données OPPORTUNITY, qui contient 46 495 échantillons d'apprentissage et 9894 échantillons de test. Chaque échantillon est associé à une séquence. Les deux phases d'entraînement et des tests du modèle s'exécutent sur un GPU avec 1536 cœurs, une vitesse d'horloge de 1050 MHz et 8 Go de RAM.

Hu *et al.*[86] ont montré qu'il est possible d'améliorer les performances après accumulation des résultats des réseaux assemblés. Une fois que les réseaux profonds sont exécutés avec de grandes quantités de données et une fonction de perte appropriée, chaque image de test est passée à travers les réseaux pour obtenir une représentation profonde des caractéristiques. Après avoir extrait les caractéristiques profondes, la plupart des méthodes calculent directement la similarité entre deux caractéristiques à l'aide d'une distance spécifique, puis le plus proche voisin (K-PP) et la comparaison de seuil sont utilisés pour identifier et vérifier de l'objet test.

1.5.7 Comparaison et synthèse

Ces dernières années, diverses techniques de reconnaissance [45 - 90] ont été développées pour améliorer la précision et le temps de traitement d'images en utilisant différentes approcheset techniques de reconnaissance. Le tableau 1.1 illustre une comparaison entre plusieurs approches en termes de précision de reconnaissance, bases de données des tests, méthode utilisée, en présentant leurs inconvénients et avantages. Nous avons conclu que les techniques de reconnaissance basées sur l'apparence [45 - 49] prennent en charge les caractéristiques visuelles et modélise ses derniers avec de nouveaux espaces de représentation. Cependant, ils souffrent d'une faible précision par rapport aux techniques de reconnaissance basées sur les points clés. En effet, les techniques basées sur les points clés [53 - 55] fournissent une reconnaissance plus précise mais souffrent du temps de traitement long. Les approches hybrides existantes [74 - 79] visent à combiner les avantages des différentes approches de reconnaissance pour produire de meilleurs taux de reconnaissance à tous les niveaux. Cependant, ces approches traitent seulement un nombre prédéfinies de catégories d'expressions faciales. Ainsi de nombreuses techniques d'apprentissage automatique ont été développées combinant les avantages de la réduction d'espace de représentation des techniques linéaires pour. Dans le but d'optimiser les traitements qui doivent s'effectuer en temps réel pendant le processus de reconnaissance et améliorer le taux de reconnaissance, plusieurs approches d'apprentissage en profondeur adoptent l'extraction profonde [86 - 90] des caractéristiques pour obtenir une précision de reconnaissance élevée. Cependant, ces approches souffrent du temps de calcul et espace mémoire importants dans la construction et la représentation des caractéristiques de visages. Notre travail consiste à réduire efficacement le vecteur de caractéristiques de visage et leurs expressions faciales, cela revient à améliorer les descripteurs standards d'images dans l'espace de caractéristiques dans le but de réduire la taille de vecteur des caractéristiques sans perte d'information pertinente (compromis temps de traitement / précision de classification).

Table 1.1 : Comparaison des méthodes de reconnaissance des visages et expressions faciales.

Méthodes de reconnaissance de visage		BD de test	Méthode	Avantages	Inconvénients	Résultat
basées sur l'apparence						
▸ LBP	Ojala*et al.* [3]	*TDF – CF1– LFW*	*MAP*	*Robustesse frontale*	*Asymétrie de l'image*	*5% –13.03% – 90.95%*
▸ LPQ	Hussain *et al.* [45]	*FRET –LFW*	*Cosinus*	*Robustesse à l'éclairage*	*Beaucoup d'Infos discriminantes*	*99.20% –75.30%*
▸ HOG et MMD	Karaaba *et al.* [46]	*FRET –LFW*	*MMD / MLPD*	*Aligner les difficultés*	*Faible précision de reconnaissance*	*68.59% – 23.49%*
▸ PHOG et SVM	Arigbabu *et al.* [47]	*LFW*	*SVM*	*Invariantà la pose*	*Complexité et temps de calcul*	*88.50%*
▸ *Corrélateur VLC*	Leonard *et al.* [48]	*PHPID*	*ASPOF*	*Robuste aux bruits*	*Faible nombre d'image de réf.*	*92%*
▸ *PCA-FCF*	Napoléon *et al.* [49]	*CMU_PIE –FRGC2*	*Fil. corrélation*	*Insensible à l'occlusion*	*Basé seulement-méthode linéaire*	*96.60% –91.92%*
basées sur les points clés						
▸ SIFT	Lenc*et al.* [53]	*FRET – AR – LFW*	*Probabilité Post.*	*Robustesse acceptable*	*Encore loin d'être parfait*	*97.3%–95.8%–98.4%*
▸ SURF	Du *et al.* [91]	*LFW*	*Distance FLANN*	*Robuste et distinctif*	*Temps de calculs important*	*95.60%*
▸ SURF et SIFT	Vinay*et al.* [27]	*LFW – Face94*	*Distance FLANN*	*Robustesse acceptable*	*Temps de calculs importants*	*78.86% –96.67%*
▸ BRIEF	Calonder *et al.* [55]	*Exemples d'images*	*k-NN*	*Temps de traitement réduit*	*Taux de reconnaissance faible*	*48%*
Méthodes hybrides						
▸ GW-LDA	Fathima *et al.* [74]	*AT – FACE – MITIND*	*k-NN*	*Invariantà l'illuminosité*	*Temps de calculs importants*	*88%– 94.02%–88.12%*
▸ OCLBP	Barkan *et al.* [75]	*LFW*	*WCCN*	*Réduire la dimension*	-	*87.85%*
▸ WLBP –ACF	Juefei *et al.* [76]	*LFW*		*Invariantà la pose*	*Complexité élevée*	*89.69%*
▸ Fisher – SIFT	Simonyan *et al.* [78]	*LFW*	*Mat.Mahalanobis*	*Robustesse élevée*	*Un seul vecteur de caractéristique*	*87.74%*
▸ DCT–PCA	Ojala *et al.* [3]	*ORL – UMIST – YALE*	*Distance Eucl.*	*Réduire la dimension*	*Complexité élevée*	*92.6%– 99.4%–95.6%*
▸ SIFT –ICP	Mian*et al.* [81]	*FRGC*	*ICP*	*Robustesse aux expressions*	*Complexité*	*99.74 %*

▸ PCA–FLD	Ojala *et al.*[3]	*CMU– FRET – AR*	*SVM*	*Pose et illuminosité et Exp.*	*Robustesse faible*	*71.9%– 94.7%–68.6%*
Apprentissage automatique						
▸ ACP	Sirovich *et al.* [65]	*Yale – ORL – CMU*	*PCA*	*Principe simple*	*Précision faible*	*78 %– 88.5%–75.7%*
▸ SVM	Osuna *et al.*[66]	*BD Propre*	*SVM*	*Large classes d'images*	*Instabilité des calculs*	*74.2%*
▸ SNoW	Yang *et al.*[67]	*ORL – CMU*	Winnows	*Meilleurs performances*		*87.24%–87.52%*
▸ Bayésien	Moghaddam *et al.*[68]	*FRET*	*Probabilité*	*Réduire la dimension*	*instables chang. d'illuminations*	*85.42%*
▸ HMM	Bevilacqua *et al.*[69]	*ORL*	*HMM – CNN*	*Robustesse aux expressions*	*Temps de calculs importants*	*99.50%*
▸ WT - CNN	Vankayalapati*et al.* [71]	*FRET – ORL*	*WaveletTrans.*	*Amélioration de précision*	*N'est pas invariantà la pose*	*89.27%–90.02%*
Deep Learning						
▸ DeepFace	Hu *et al.* [86]	*LFW– YTF*	*DNN*	*Précision élevée*	*Temps de calculs importants*	*97.35%–91.4%*
▸ Light-CNN	Wu*et al.*[87]	*LFW– VAL1– VAL2*	*MFM256-D*	*Meilleurs performances*	*Temps de calculs très importants*	*98.8%–95.3%–92.9%*
▸ L-softmax	Liu*et al.*[88]	*MNIST – LFW*	*CNN-Softmax*	*Interprétation géom.*	*Complexité élevée*	*96.53%–92.9%*
▸ Face-Net	Schroff*et al.*[89]	*LFW–Youtube Faces*	*CNN-SVM*	*Meilleurs performances*	*Complexité élevée*	*99.63%–95.12%*
▸ CosFace	Wang *et al.*[90]	*LFW–YTF*	*Deep CNN*	*Meilleurs performances*	*Temps de calculs très importants*	*99.73%–97.6%*

1.6 Conclusion

Au cours de ce chapitre, nous avons présenté les différentes méthodes de reconnaissance faciales et expressions faciales, on a cité ces inconvénients et ces avantages. Nous avons étudié et comparés certaines approches existantes dans la littérature, de l'extraction des caractéristiques et la reconnaissance faciale. Chacune des approches étudiées traite un point particulier et utilise des techniques appropriées. Nous avons détaillé plusieurs défis liées aux méthodes de la reconnaissance des visages et expressions faciales. Les performances de ces méthodes sont comparées à travers de nombreux facteurs : base de données de test, taux de reconnaissance et méthodes utilisées. De l'étude comparative de ces méthodes, nous concluons que le problème majeur des techniques de reconnaissance est le développement d'une nouvelle technique qui offre un temps de traitement rapide tout en réduisant la taille des vecteur des caractéristiques discriminantes pour distinguer les visages des personnes entre eux.

Nous sommes particulièrement intéressés aux descripteurs textures des images faciales présentées et détaillées dans le chapitre suivant.

Chapitre 2

Texture et descripteurs de la texture

Sommaire

2.1 Introduction

Après avoir introduit dans le chapitre précédent les différentes approches d'analyse et de traitement des visages, essentielles à nos sociétés. Nous présentons dans ce chapitre les approches et les techniques utilisées dans nos contributions. Ces dernières s'appuient sur l'exploitation et l'extraction de la texture à travers des descripteurs standards pour avoir des informations issues du visage ou des expressions faciales. Nous commençons dans ce chapitre par une description générale de la texture et les modèles utilisés pour extraire et classifier la texture. Ensuite nous décrivons les descripteurs les plus utilisés dans le domaine d'identification du visage et les expressions faciales.

L'analyse de la texture joue un rôle primordial dans le domaine de l'analyse et du traitement d'images, car les images traitées sont souvent formées des objets présentent à la fois un aspect textural et structurel. L'analyse de la texture d'image peut être trouvée dans de nombreuses applications du monde réel telles que la reconnaissance d'objets, l'analyse des images médicales pour aide au diagnostic et dans le domaine industriel pour le contrôle et la classification de qualité. La texture est considérée comme une variation spatiale d'intensités de pixels, ou d'une façon plus générale c'est la reproduction d'un motif de base dans plusieurs directions spatiales, ou la texture est une structure spatiale constituée par l'organisation des motifs (primitives) de base ayant chacune un aspect aléatoire. Elle est considérée comme un phénomène bidimensionnel : la description primitive de base à partir desquels sont formée la texture et la description des relations spatiales entre ces primitives. On distingue généralement deux types de texture tactile et optique. La texture tactile fait référence à la sensation tangible d'une surface et la texture visuelle fait référence à la forme ou au contenu de l'image. L'analyse de texture dans un système de vision humaine est facilement réalisable mais dans le domaine de la vision industrielle et le traitement d'image ont leurs propres complexités. La figure**2.1** montre un exemple d'image texturale. La figure **2.1** montre une image d'un mur en bois avec un motif entièrement répétitif. Ce motif est répété sur toute l'image. Le motif en bambou est également illustré à la figure **2.1.b**.

(a) **(b)**

Figure 2.1 : Un exemple d'image texturale : a) Image de texture ; b) Motif répétitif de texture.

L'analyse de la texture est essentielle dans le processus de détection, de classification et de reconnaissance des objets ou dans le traitement d'image de façon générale, car elle représente une caractéristique importante de la structure et la surface interne d'un objet dans une scène. Son principe général est basé sur la recherche et la détermination des paramètres affectés à la texture afin qu'ils soient

le plus discriminants et pertinents que possible et puissent caractériser la texture entre plusieurs autres et la distinguerde façon simple et optimale.

Dans le traitement d'image, la texture représente la variation spatiale d'intensité lumineuse des pixels. La texture représente également les variations d'intensités, qui mesurent des caractéristiques telles que le lissé, la grossièreté et la régularité de chaque surface dans différentes directions de l'espace. Les images texturales dans le traitement d'image et la vision industrielle se référant aux images dans lesquelles un modèle spécifique de distribution et de dispersion de l'intensité lumineuse des pixels est répété spatialement sur toute l'image. La classification, la segmentation et la synthèse des textures et la forme des textures font partie des principaux problèmes traitées par l'analyse de texture [92, 171] comme l'illustre la figure **2.2**. L'objectif du domaine « Texture Shape Extraction », est d'extraire des images 3D qui recouvrent différentes régions de l'image avec une texture spécifique. Dans ce chapitre on va focaliser sur les approches qui consistent à extraire la structure et la forme des éléments de l'image en analysant leurs propriétés textuelles et l'espace reliant les uns avec les autres. La synthèse de texture permet de produire des images qui ont la même texture que la texture d'origine. Les applications visées concernent la création d'images graphiques et de jeux informatiques, l'élimination d'une partie de l'image et le placement de l'arrière-plan texture, création d'une scène avec éclairage et un angle de vue différent ou même la création d'effets artistiques sur des images contenant les textures en relief.

La segmentation de texture consiste à diviser une image en régions, chacune ayant une texture spécifique. Les contours des différentes régions sont déterminés lors de la segmentation de la texture. En d'autres termes, dans la segmentation de texture, les caractéristiques des limites et des zones sont comparéeset si leurs caractéristiques de texture sont suffisamment différentes, la limite est trouvée.

La classification des textures est l'un des domaines importants dans le contexte de l'analyse des textures. La classification permette de fournir des descripteurs de textures pour catégoriser les images texturales. La classification de texture consiste à affecter une image échantillon inconnue à l'une des classes de textures prédéfinies. Les questions clés de l'analyse de la texture posent le problème de la classification *efficace* et *efficiente* de la texture dans de grandes bases de données à travers des descripteurs de texture pertinents et discriminants.

Ces dernières années, l'analyse de la texture est effectuée dans plusieurs travaux de recherche par plusieurs techniques et différentes approches. Certaine utilisent les éléments géométriques locales, d'autres utilisent les notions des modèles tels que les fractales, les champs aléatoires et les techniques statistiques de texture. Il existe également d'autre techniques basées sur l'analyse des fréquences temporelles ou sur le filtrage telles que la transformée de Huang-Hilbert, la transformée de Hermite et la transformé en ondelettes.

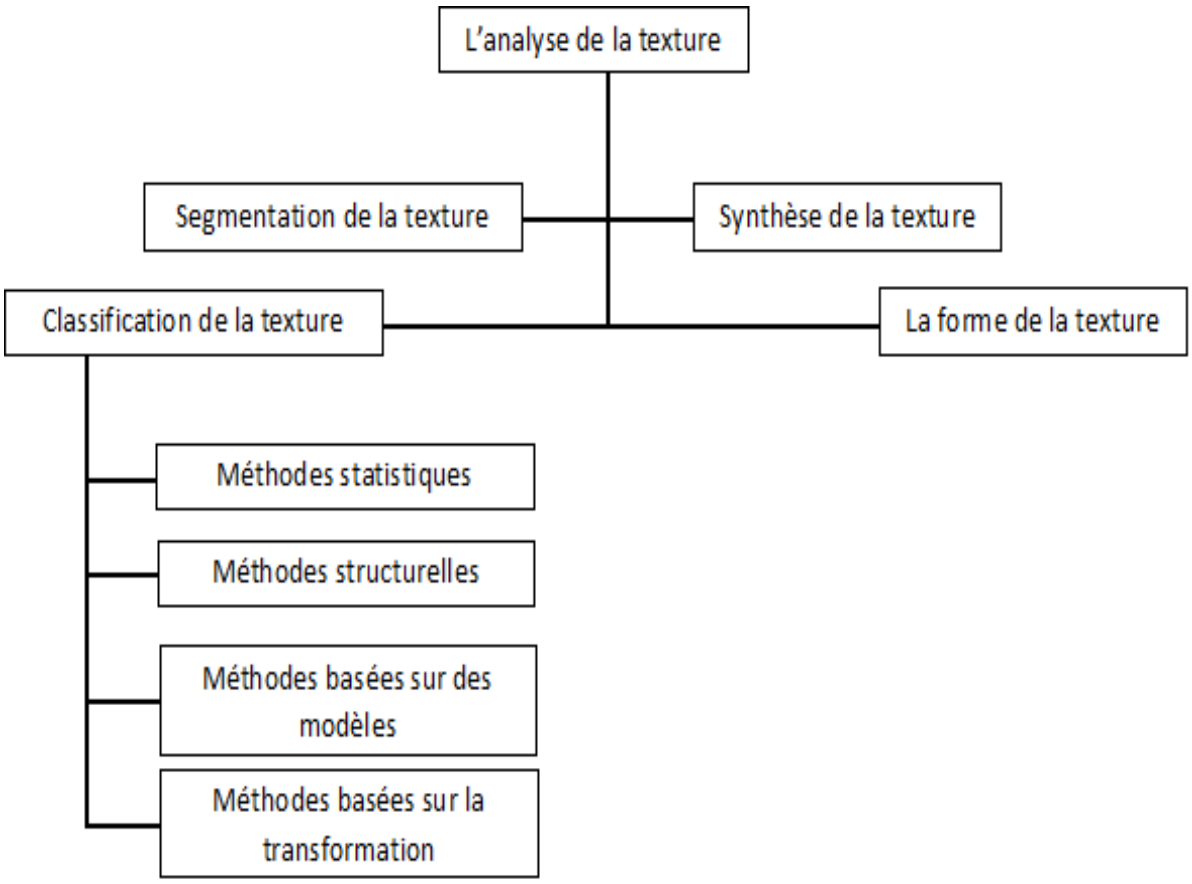

Figure2.2 : Vue générale de l'analyse de texture.

2.2 Défis de la classification d'images texturées

2.2.1 Robustesse

Dans les situations réelles, les deux défis majeurs dans l'analyse et la classification des images sont de nombreux effets destructeurs. Ces deux phénomènes importants sont liés à la rotation des images et les effets de bruit. Dans la pratique, les méthodes de classification qui prend en considérations ces phénomènes ne sont pas durables et l'exactitude des résultats peut être gravement réduite. Par conséquent, dans des circonstances réelles, les méthodes d'analyse et de classification des images doivent être aussi robustes et stables que possible face à ces deux phénomènes et neutraliser leurs effets dévastateurs. En plus des problèmes ci-dessus, les images peuvent différer les unes des autres en termes d'échelle, point de vue ou intensité de la lumière. C'est l'un des principaux défis du système de classification des textures. Pour cette raison, divers des méthodes ont été proposées, chacune essayant de couvrir ces aspects.

2.2.2 Extraction des descripteurs discriminants de texture

La classification de texture signifie l'affectation d'une image à un groupe de texture qu'on l'a défini précédemment. La texture pouvant être un critère très discriminant dont le challenge est d'extraire

d'une manière efficace et pertinente par le descripteur. Un processus de classification se fait généralement en deux phases :

- **Extraction des fonctionnalités** : qui extraire les propriétés texturales. Le but est de créer un modèle pour chacune des textures qui existent dans la plate-forme de formation.
- **Classification** : qui consiste à déterminer la classe de la texture de l'image à partir d'échantillon de test par un algorithme de classification. L'organigramme général des méthodes de classification des images de texture est indiqué sur la figure **2.3**, sur la base des deux phases précédentes.

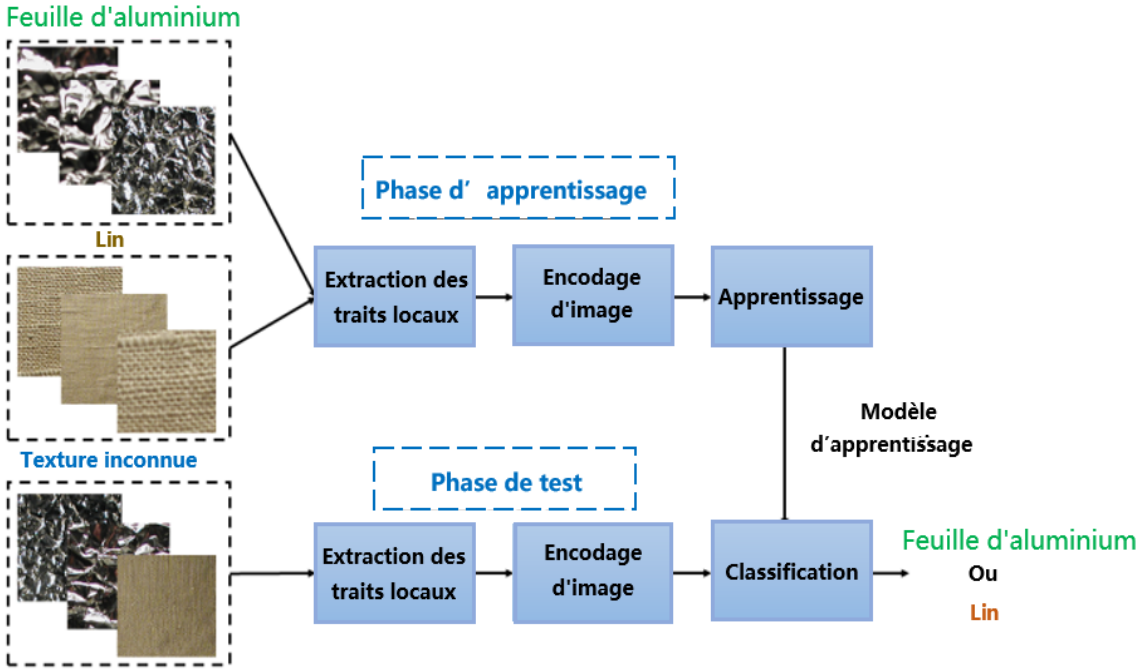

Figure2.3 : L'organigramme général de la classification de la texture.

a. Extraction des fonctionnalités

La première étape de l'extraction des caractéristiques de texture consiste à établir un modèle pour chacune des textures obtenues dans la phase d'apprentissage. Elles peuvent être décrient par des histogrammes numériques et discrets, des distributions empiriques, des caractéristiques de texture telles que le contraste, structure spatiale, direction, etc. Elles sont utilisées pour préparer l'étape de la classification. Jusqu'à présent, de nombreuses méthodes de classification de la texture ont déjà été proposées, mais l'efficacité de la plupart ces approches dépend, bien sûr du type de caractéristiques extraites et ses mesures. Elles sont généralement classifier en quatre approches principales (1) – les méthodes statistiques, (1) – les méthodes structurelles, (3) – les méthodes basées sur des modèles et (4) les méthodes des transformations. Chacune de ces méthodes extrait différentes caractéristiques basés texture [92 – 93, 171]. Il convient de noter qu'aujourd'hui, il est difficile de mettre certaines méthodes dans un groupe particulier en raison de la complexité des méthodes et l'utilisation des propriétés combinées car la plupart d'entre elles se répartissent en plusieurs groupes. Les types de méthodes

d'extraction de caractéristiques sont populaires et largement utilisées dans les systèmes actuels. Ils seront décrits en détail dans la section suivante.

b. Classification des textures

Nous nous intéressons dans cette section à la classification des textures basée sur l'utilisation des algorithmes d'apprentissage automatique supervisés ou algorithmes de classification non-supervisés. On détermine la classe appropriée pour chaque image par la comparaison du vecteur de caractéristiques extraites dans la phase d'apprentissage avec le vecteur des caractéristiques de la phase de test. Ce processus est répété de manière itérative pour chaque image en phase de test. À la fin, les classes estimées pour les tests avec leur classe réelle sont déterminées et le taux de reconnaissance est calculé ce qui reflète l'efficacité de la méthode. On détermine le taux de reconnaissance de chaque algorithme ; résultant de la mise en œuvre en fonction du stimulus d'entrée. Enfin, nous pouvons évaluer l'efficacité de son algorithme avec d'autres méthodes disponibles en comparant leurs taux de reconnaissance.

$$Precision = \frac{nombre\ de\ correspondances\ correctes}{nombre\ total\ d'imagesde\ test} \text{ x } 100 \qquad (2.1)$$

2.3 Domaines d'applications

La texture de l'image nous donne beaucoup d'informations utiles sur le contenu de l'image, les objets à l'intérieur, l'arrière-plan contexte, arrière-plan, etc. L'analyse de texture touche la plupart des domaines du traitement d'images, en particulier dans le processus d'apprentissage et l'extraction de la fonctionnalité est en cours de discussion lorsque nous voulons comparer les images telles que :

- Détection et reconnaissance de visage [94]
- Suivi des objets dans les vidéos [95]
- Diagnostics de la qualité des produits [96]
- Analyse d'images médicales [97]
- Télédétection [98]
- Végétation [99]

2.4 Principaux méthodes de classification des textures

Nous nous intéressons dans cette section à l'exposition de différentes méthodes basées sur l'information de la texture. La majorité des méthodes sur l'extraction et la classification des textures sont regroupées principalement dans quatre approches principales : les méthodes statistiques, les méthodes structurelles, les méthodes à base des modèles et les méthodes à base des transformations, il existe plusieurs façons pour le gérer. Le tableau **2.1** présente une liste de certaines méthodes présentées dans ce domaine. Nous allons essayer de tirer de ces inconvénients et de ces avantages, nos propres méthodes pour extraire et classifier des images faciales et des expressions faciales.

2.4.1 Méthodes statistiques

Les méthodes de traitement statistique de la texture contiennent un nombre important des méthodes développés dans le domaine de la vision industrielle. Ces méthodes d'analyse effectuent une série de calculs statistiques à base des fonctions de distribution des niveaux de gris des pixels. Elles sont bien adaptées aux approches stochastiques pour l'analyse et le traitement de la texture. Le principe des méthodes statistiques est basé sur l'extraction des relations entre un pixel et ses voisins. Les paramètres sont calculés selon le nombre de pixel impliqué pour un ordre statistique donné. On peut noter qu'l y a des paramètres mathématiques qui reflètent des propriétés visuelles de la texture, tant que d'autres ne correspondent pas forcément à des aspects visuels réels. En général, les méthodes de dérivation du vecteur d'objet issues de calculs statistiques s'entrent dans ce groupe. Les caractéristiques statistiques de premier, deuxième et ordre supérieur qui caractérisent le contenus texture de l'image font partie de ces méthodes. La différence entre ces trois types est que la spécification de pixel au premier ordre est unique sans tenir compte des dépendances spatiales entre les pixels de l'image. Alors celui de la caractéristique statistique de deuxième ordre et ordre supérieur, la spécification est calculée en tenant compte de la dépendance entre deux pixels ou plus. La matrice de cooccurrence connue sous le nom de l'histogramme de deuxième ordre est l'une des méthodes à inclure dans ce groupe.

2.4.1.1 Caractéristiques et spécifications de l'histogramme

L'une des techniques les plus simples pour décrire une texture est l'utilisation de couples statistiques liés à l'histogramme de l'intensité d'une image ou région. L'histogramme est une représentation graphique de la densité optique de l'image. L'histogramme de l'image est une représentation bidimensionnelle de distribution des niveaux de gris dans l'image. Les indicateurs statistiques de premier ordre (i.e. fréquence)sont calculés directement à partir des niveaux de gris des pixels de l'image d'origine, quel que soit la relation spatiale entre eux.

Catégorie	Sous-catégorie	Méthodes
Statistiques	Propriétés de l'histogramme	• Modèle binaire de Gabor [100] • GLCM et filtres de Gabor [96] • Gabor et LBP [101] • Transformée en ondelettes et GLCM [102] • Variation d'énergie [93] • Motifs binaires locaux et sélection de points significatifs [103] • Combinaison de modèle primitif et caractéristique statistiques [104] • Hybrid color local binary patterns [111]
	Matrice de cooccurrence	
	Descripteurs binaires locaux	
	Basé sur l'inscription	
	Energie de la texture	
Structural	Mesure primitive	• Variation d'énergie [93] • Détection de granularité de texture basée contours [105] • Filtre morphologique • Squelette primitif et ondelettes
	Caractéristiques de contours	
	Représentation du squelette	
	Opérations morphologiques	
	SIFT	
Basée modèle	Autorégressive (AR) model	• Analyse multi-fractale dans la pyramide d'ondelettes multi-orientation [106] • Modèles de texture de champ aléatoire de Markov [107] • Modèles autorégressifs simultanés [108]
	Modèles fractals	
	Modèle de champ aléatoire	
	Modèle Texem	
Basée transformation	Spectral	• Modèle de Gabor binaire [100] • Les canaux d'ondelettes et banque de filtres des canaux LL [109] • GLCM et filtres de Gabor [96] • Gabor et LBP [101] • Transformée en ondelettes et GLCM [102] • SVD et DWT [110] • Squelette primitif et ondelettes
	Gabor	
	Wavelet	
	Transformation de Curvelet	

Tableau 2.1 : Catégorisation des méthodes de classification des textures [171]

En règle générale, les indices statistiques de premier ordre sont calculés des moments statistiques de l'histogramme de l'image pour caractériser la quantité de lumière et d'obscurité de l'image. De nombreuses caractéristiques peuvent en être extraites qui sont utilisées pour calculer le n^{ième} couple autour de la moyenne comme suit :

$$u_n(z) = \sum_{i=0}^{l}(z_i - m)^n p(z_i) \quad (2.2)$$

$$m = \sum_{i=0}^{l-1} z_i p(z_i) \quad (2.3)$$

Où : z_i sont les variables aléatoires liées à la gamme d'intensité de l'image,

$p(z_i)$ représente le nombre de pixels dans l'image qui ont un niveau de gris égal à z_i .

La variance est une mesure de la diversité du contraste, elle peut être aussi utilisée pour créer des descripteurs de lissage relatif.

$$\sigma = \sum_{i=0}^{l}(z_i - m)^2 p(z_i) \quad (2.4)$$

L'asymétrie est un critère pour décrire le degré de symétrie de l'histogramme

$$\text{Asymétrie} = \sum_{i=0}^{l}(z_i - m)^3 p(z_i) \quad (2.5)$$

L'entropie est un critère de variabilité, nul pour une image fixe.

$$\text{Entropie} = -\sum_{i=0}^{l} p(z_i)\, log_2 p(z_i) \quad (2.6)$$

2.4.1.2 Matrice de cooccurrence

L'un des opérateurs les plus connus pour extraire des caractéristiques de texture est la matrice de cooccurrence introduite par Haralick [112]. Ces opérateurs sont similaires à des histogrammes bidimensionnels, ils représentent le nombre d'occurrences de pixels spéciaux dans l'image. Les matrices de cooccurrence sont basées sur le principe que le niveau de gris d'un pixel d'une image est fortement dépendant des niveaux de gris des pixels voisins. Elles mesurent la probabilité pour qu'un couple de niveaux de gris apparaisse dans l'image selon une loi spatiale donnée. La matrice de cooccurrence d'une image est calculée sur la base des corrélations entre les pixels de l'image. On s'intéresse aux matrices de cooccurrence en couples de pixels qui sont, par définition, séparés par une distance (1, 2, 3,4…).

Pour une image $n\ bits$ avec une luminosité $L = 2 \times n$ niveaux, une matrice $L \times L$ est calculée dont les éléments sont le nombre des paires de pixels ayant une luminosité a, b séparée par d. Après avoir calculé la matrice de cooccurrence, les caractéristiques textuelles statistiques (Haralick) sont calculées.

La figure 2.4 montre un exemple du calcul de la matrice cooccurrence d'une image à 3 bits avec 8 niveaux d'intensité dont sa matrice de cooccurrence à 8 lignes et 8 colonnes. Les éléments de cette matrice sont le nombre d'occurrences de pixels à niveaux de gris i, j qui sont représentées par un déplacement de un pixel dans la direction de zéro degré. En générale, la matrice de cooccurrence est

définie pour les quatre directions principales (0, 45, 90 et 135).

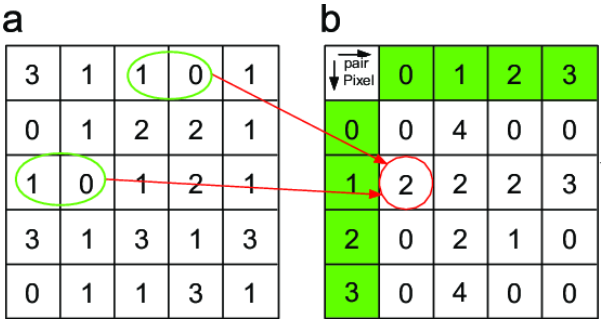

Figure 2.4 : Exemple de calcul d'une matrice cooccurrence [172].

La figure **2.5** montre les quatre angles possibles entre deux pixels (0°, 45°, 90° et 135°). Ces angles sont représentés avec un déplacement de 3 entre deux pixels.

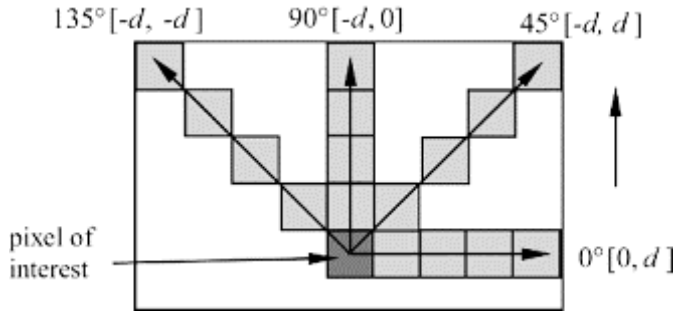

Figure 2.5 : Quatre directions différentes avec déplacement 3 entre deux pixels [185].

Après avoir construire la matrice de cooccurrence, les propriétés statistiques de Haralick peuvent être obtenues à partir de cette dernière, caractérisant la dépendance spatiale entre les niveaux de gris. Par exemple, six propriétés statistiques : le *contraste*, la *corrélation*, l'*énergie*, l'*homogénéité*, l'*entropie* et la *probabilité maximale* peuvent être calculées de la matrice de cooccurrence.

$$\text{Contraste} = -\sum_{i,j} |i-j|^2 \, p_{ij} \tag{2.7}$$

$$\text{Correlation} = -\sum_{i,j} \frac{(i-u_i)(j-u_j)\, p_{ij}}{\sigma_i \, \sigma_j} \tag{2.8}$$

$$\text{Energie} = \sum_{i,j} p_{i,j}^2 \tag{2.9}$$

$$\text{Homogénité} = \sum_{i,j} \frac{p_{ij}}{1+|i-j|} \tag{2.10}$$

$$\text{Entropie} = -\sum_{i,j} p_{i,j} \, log_2 p_{i,j} \tag{2.11}$$

$$\text{Probabilité maximale} = \max(p_{i,j}) \tag{2.12}$$

Où :

$p_{i,j}$: la matrice de cooccurrence normalisée.

u_i : la moyenne par ligne.

u_j : la moyenne par colonne.

σ_i, σ_j : écarts types par ligne et par colonne.

2.4.1.3 Motifs binaires locaux

Le motif binaire local (Local Binary Pattern LBP) [3] est un descripteur d'images texturales qui peuvent définir la structure spatiale et le contraste local d'une image ou d'une partie de l'image. LBP est devenu un descripteur de texture standard. Ce descripteur est largement utilisé pour sa mise en œuvre simple de l'extraction du vecteur de caractéristiques avec une dégrée de précision élevée de classification. Pour plus de détails sur LBP voir la section 2.5.3.

2.4.2 Méthodes structurelles

Dans cette catégorie de méthodes, la texture est introduite en fonction des primitives initiales et de leur disposition spatiale. Les primitives initiales peuvent être simplement un pixel, une région ou une forme. L'analyse structurelle consiste à déterminer la texture primitive et les règles de placement spatialde ces primitives. Les règles de placement permettent d'arranger un ensemble des primitives en calculant des relations géométriques et examiner leurs propriétés statistiques. En d'autres termes, les méthodes structurelles considèrent la texture comme une combinaison de motifs initiaux, une fois la texture primaire est détectée, puis les propriétés statistiques de la texture primaire sont calculées, alors elles sont utilisées comme des traits de caractéristique de l'image. Ces méthodes sont bien adaptées aux textures avec une structure régulière, mais comme les images traitées sont généralement dotées d'une texture irrégulière, alors on peut les classifier comme des méthodes non optimales.

2.4.2.1 Caractéristiques du contour

Le contour d'un objet est la frontière entre cet objet et son arrière-plan. En d'autres termes, les contours sont les bords des deux niveaux de gris, qui sont les variations d'illuminations de deux pixels adjacentes qui se produisent à un endroit spécifique de l'image et se diffèrent considérablement. La détection des contours a pour but d'identifier les points d'une image dans lesquels l'intensité de la lumière se transforme brusquement. Les bords peuvent être des changements géométrique de la scène ou de l'objet correspond à un taux de variations d'illuminations qui détermine les caractéristiques des objets, tels que les marques et la forme de la surface. La détection des contours peuvent être effectué par divers opérateurs comme Sobel [113], Perwit [114], Robert, etc.

2.4.2.2 Transformation de caractéristiques visuelles invariante à l'échelle

Actuellement, la méthode SIFT (Scale Invariant Feature Transform) est l'un des meilleurs moyens puissants pour extraire et décrire des traits. Cette méthode a été introduite par David Loo en 1999 [116]. Cette méthode a l'avantage d'avoir des spécificités locales, donc il a une bonne précision d'occlusion différentielle. Pour plus de détails sur SIFT voir la section **2.5.1**.

2.4.3 Méthodes basées modèles

Les méthodes basées sur des modèles permettent de concevoir et d'élaborer des modèles pour représenter des images de texture. Les méthodes les plus populaires sont la méthode Autorégressive (AR) [117], les chaines de Markov (Hidden Markov Modèle (HMM)) [117], Gibbs RMF [118] et le modèle fractal [119]. De cette façon, un modèle de l'image est créé puis ce modèle est utilisé pour décrire l'image de synthèse. Un modèle est un ensemble de paramètres représentant les propriétés de qualités essentielles de la texture. Le modèle basé sur la géométrie fractal est un modèle exemple, utilisé pour analyser de nombreux phénomènes naturels et physiques.

2.4.3.1 Modèle Fractal

En 1970, Mendel Bert, introduit un nouveau domaine des mathématiques appelé fractale [119]. Il introduit une nouvelle classe de collections mathématique appelé fractale. Cette collection contient de nombreux objets compliqués qui ont été produits en répétant des règles simples. En général, on peut dire qu'un objet fractal est un objet géométrique constitué d'éléments semblables à toute échelle, chacune de leur partie reproduit leur totalité. Il a une complexité intrinsèque, une propriété d'autosimilarité basée sur une homothétie interne et qui se manifeste d'une irrégularité fondamentale à toutes les échelles d'observation. En pratique, des algorithmes géométriques itératifs ont utilisés pour définir la structure d'un objet fractal. En revanche, on dit qu'un objet est non fractal si l'on ne peut pas reproduire une nouvelle forme similaire à l'origine chaque fois qu'on agrandit une de ses parties (Figure 2.6). Les fractales peuvent être utilisé pour modéliser la grossièreté, la dureté et l'auto-similitude dans une image de texture.

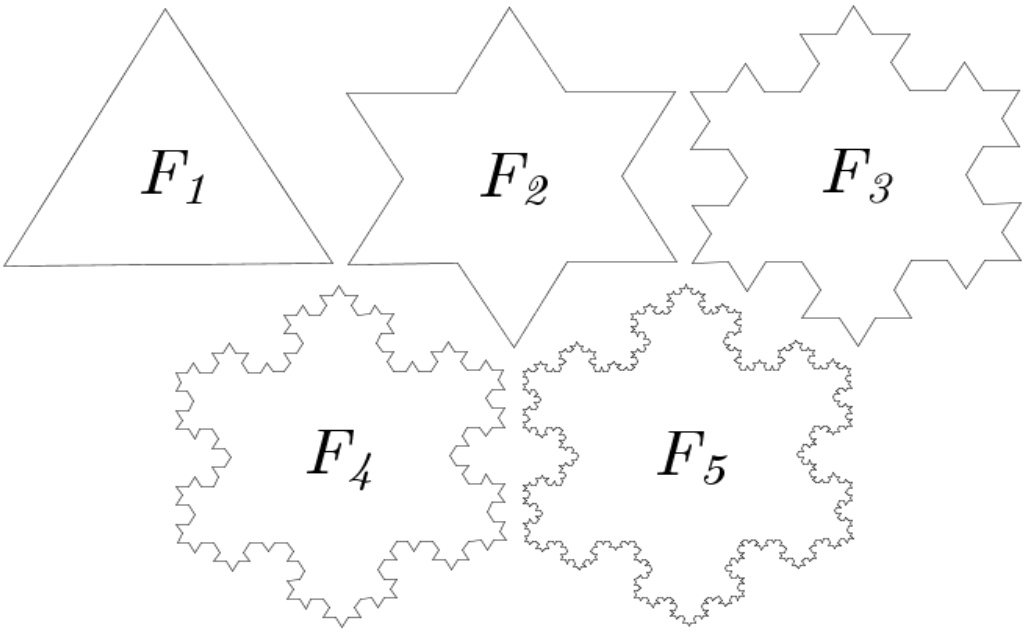

Figure 2.6 : Exemple de génération d'image fractal à partir d'un objet.

2.4.5 Méthodes basées transformés

Dans cette classe des méthodes, l'image est représentée dans un autre espace à l'aide d'une fonction de transformation.En effet, la texture est facilement distinguée dans le nouvel espace. Les approches les plus utilisées, les transformées en ondelettes [120], Curvelet [121], Ridgelet [122] ,Gabor [123] et le spectre de Fourier [124].

2.4.5.1 Le spectre texture de Fourier

L'analyse de la texture dans une image peut être réalisée de plusieurs manières selon le type d'information représentée afin de mieux exploiter certaines propriétés des images. Les informations contenues dans une image peuvent être représentées dans le domaine spatial, ou dans le domaine fréquentiel selon l'objectif et selon la nature de traitement pour aboutir à un traitement fiable, optimal de l'analyse de texture. On peut même combiner les deux domaines spatial et fréquentiel. La représentation spatiale d'une image numérique correspond à une matrice rectangulaire dont les valeurs de chaque pixel représentent la distribution normale de la luminosité. Pour celle de fréquentielle, on utilise la transformée de Fourier, qui le sujet de cette section. La Transformée de Fourier (TF) pour un signal discret 2D telle qu'une image, permet de passer d'une représentation spatial de l'image vers une représentation fréquentielle.

L'intérêt principal de spectre texture de Fourier est de décrire les motifs alternés ou les motifs bidimensionnels. Ces motifs de texture peuvent être bien identifiés comme étant les niveaux des concentrations supérieures à l'énergie dans le spectre. Trois caractéristiques du spectre texture de Fourier qui sont utiles pour décrire la texture sont les suivantes [124] :

- Le pic dominant dans le spectre de la direction principale de la texture.
- Le pic sur la plaque de fréquence alternative principale de la texture.
- Élimination de tout composant proportionnel en filtrant les composants d'image non alternatifs.

Le spectre est symétrique autour de la source, donc seule la moitié de la plaque de fréquence est prise en compte. Par conséquent, pour l'analyse de chaque modèle alternatif, un seul pic est lié au spectre. Elle est généralement réalisée en affichant le spectre des coordonnées polaires en fonction de$s(r,\theta)$ où s est la fonction du spectre et r sont les variables de ce système de coordonnées. Pour chacun, θ et $s(r,\theta)$ peut être considérés comme une dimension de $S_\theta(r)$. De même, pour chaque fréquence r, $s_r(\theta)$ est une fonction unidimensionnelle. L'analyse de $s_r(\theta)$ pour une valeur constante de θ montre le comportement du spectre (la présence de pics dans le spectre) au long de la radiale direction de l'origine. Il est plus facile de distinguer et interpréter les propriétés spectrales en exprimant le spectre en coordonnées polaires.Sest la fonction spectrale ; r et θ sont des variables coordonnées polaires. R_0est le rayon du cercle correspondant à l'origine ; $S(r)$ est constant vers la rotation [152].

$$R_0 = round\left(min\frac{m,n}{2}\right) - 1 \quad (2.13)$$

$$S(\theta) = \sum_{i_0}^{R_0} S_r(\theta)\,,\ 1 < r < 80 \quad (2.14)$$

$$S(r) = \sum_{i_0}^{R_0} S_\theta(r)\,,\ 1 < r < R_0 \quad (2.15)$$

2.4.5.2 Filtre de Gabor

Le filtre de Gabor[2], nommé d'après Dennis Gabor physicien anglais d'origine, est un filtre linéaire largement utilisé dans le domaine de traitement d'image pour l'analyse de texture, dont la réponse impulsionnelle est une sinusoïde modulée par une fonction gaussienne (également appelée ondelette de Gabor). Le filtre de Gabor est utilisé dans de nombreuses applications d'analyse d'images, y compris la détection des contours, la segmentation de texture, l'extraction de caractéristiques, la classification, la reconnaissance d'objets dans une image, la reconnaissance de l'alphabet, l'enregistrement d'image, l'orientation et le mouvement. Ils sont des classes spéciales de filtres passe-bande, qui autorisent une certaine «bande» de fréquences et en rejettent les autres. Un filtre de Gabor peut être considéré comme un signal sinusoïdal de fréquence et d'orientations particulières, modulé par une onde gaussienne. Les filtres de Gabor sont des filtres bidimensionnels, définis dans les domaines spatial et fréquentiel. Il analyse s'il y a un contenu de fréquence spécifique dans l'image dans des directions spécifiques dans une région localisée autour du point ou de la région. Le filtre de Gabor est similaire à la transformée en ondelette, suit une distribution gaussienne. Par conséquent, cette transformation est optimale dans le domaine fréquentiel [2]. L'ondelette de Gabor est une transformation optimale pour minimiser l'incertitude bidimensionnelle associée à l'emplacement et les domaines de fréquence. Cette ondelette peut être utilisée comme détecteurs d'échelle directionnels et comparables pour délimiter les frontières et les bords d'images [123]. De plus, les propriétés statistiques de cette transformation peuvent être utilisées pour déterminer la structure et le contenu visuel des images. De nombreux scientifiques contemporains de la vision confirment que les représentations de fréquence et d'orientation des filtres de Gabor sont similaires à celles du système visuel humain. Ils se sont avérés être particulièrement appropriés pour la représentation et la discrimination de texture.

Un filtre de Gabor est défini comme suit [171]:

$$G = \frac{e^{\frac{-\left(\frac{x-x_0}{\sigma_x}\right)^2\left(\frac{y-y_0}{\sigma_y}\right)^2}{2}} e^{i[\,u_0(x-x_0)+\,v_0(y-y_0)]}}{2\pi\sigma_x\sigma_y} \quad (2.16)$$

Où :

- σ_x et σ_y:coordonnées spatiales de la fonction gaussienne x, y.
- (x_0, y_0) : centre de fonction.
- (u_0, v_0) : centre de fréquence de la fonction.

Dans la pratique, pour analyser la texture et/ou obtenir des caractéristiques à partir de l'image, une banque de filtres de Gabor avec un nombre d'orientations différentes est utilisée. Dans la figure**2.7**, nous observons un exemple d'une banque de 16 filtres Gabor orientés à un angle de 11.250 (c'est-à-dire si le premier filtre est à 00, alors le second sera à 11.250, le troisième sera à 22.50, et ainsi de suite). Divers filtres sont utilisés dans l'analyse de texture d'une image comme le montre la figure 2.7.

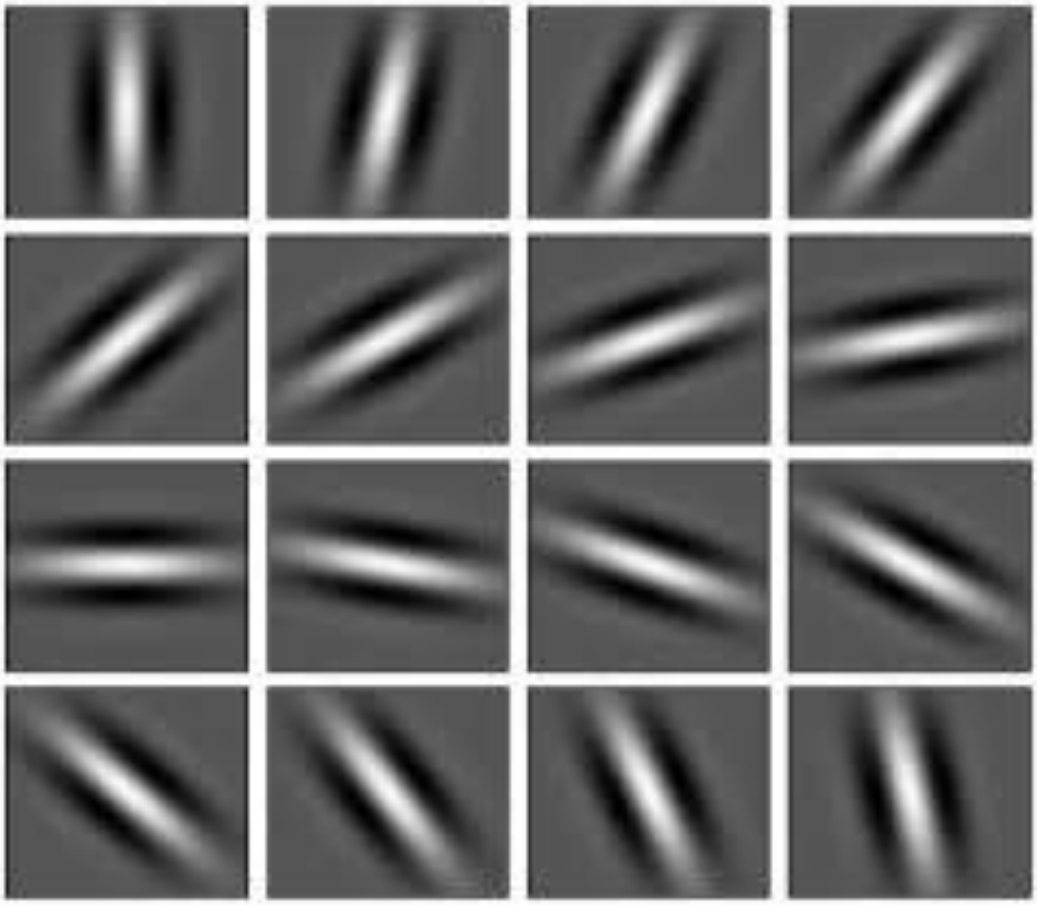

Figure 2.7 : Un exemple d'une banque de 16 filtres de Gabor.

Le filtre Gabor a été largement utilisé dans différentes contextes d'applications de la vision par ordinateur, telles que l'analyse de la texture et la détection des contours. Le filtre de Gabor est un filtre linéaire et local. Le noyau convolutionnel du filtre de Gabor est le produit d'une fonction exponentielle et gaussienne complexe. La fonction de Gabor est un moyen de distinguer les caractéristiques de la texture et ses bords. Il est invariant à la rotation et la mise en échelle. Un filtre Gabor est l'une des méthodes basées sur les filtres standards pour extraire correctement le contenu texture de l'image. Ce filtre fonctionne dans les domaines spatial et fréquentiel. Dans le domaine spatial, le noyau du filtre est obtenu à partir du produit d'une fonction gaussienne avec une fonction sinusoïdale dirigée. En conséquence, le filtre produit des réponses exceptionnelles aux points de l'image qui ont localement une certaine orientation et une certaine fréquence. Il peut être aussi utilisé pour obtenir des directions qui font la correspondance entre les propriétés extraite de l'image et les changements de rotation.

2.5 Descripteurs de la texture

Dans cette section on va présenter les descripteurs les plus populaires et les plus utilisés pour extraire et analyser la texture dans le domaine de reconnaissance de formes et plus spécialement dans les domaines de reconnaissance de visages (FR) et d'expressions faciales et (FER). Ces descripteurs sont basés sur l'extraction de l'information de la texture à partir d'une image ou d'une partie d'une image pour une utilisation ultérieure notamment la classification et la reconnaissance de formes.

2.5.1 SIFT : Transformation de caractéristiques visuelles invariante à l'échelle

Actuellement, le descripteur SIFT (Scale Invariant Feature Transform) est l'un des meilleurs moyens puissants pour extraire et décrire des caractéristiques de texture dans une image de visage. Cette méthode a été introduite par David Loween 1999 [115]. Ce descripteur à l'avantage d'avoir des particularités locales, donc il a une bonne précision d'occlusion et différentiable. Le processus d'extraction des fonctionnalités sur l'image basé sur l'opérateur SIFT est illustré dans la figure **2.8**. La première étape consiste à identifier les points clés de l'image. Elles sont les points de l'image dont la différence de gaussienne (DoG) est maximisée ou minimisée [116]. Le processus de recherche de ces points commence par la construction d'une pyramide d'images et la convolution de l'image d'origine avec les filtres gaussiens $G(x, y, \sigma)$. Donc l'espace de l'échelle est organisé comme suit [116].

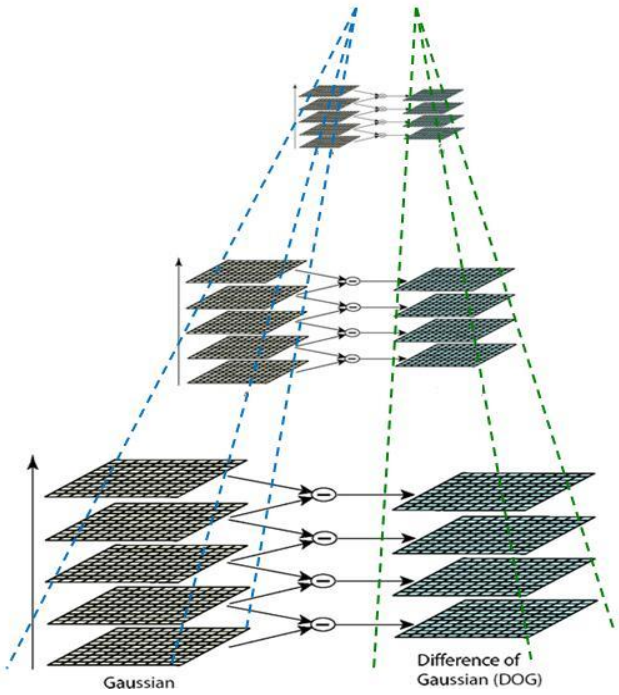

Figure 2.8 : Construction de l'image DoG [116]

- Détection des extrema d'espace échelle.
- Localisation des points clés.
- Définition des orientations.
- Descripteur de points clés.

Le niveau de maturité est contrôlé par le paramètre d'écart type σ de la fonction gaussienne. L'espace d'échelle DoG est également obtenu en soustrayant les niveaux d'échelle adjacents. La figure **2.9** décrit les étapes de construction du DoG. La deuxième étape d'extraction des caractéristiques consiste à trouver les points maximum ou minimum dans chaque octave. Cela se fait en comparant chaque pixel avec le $26^{ème}$ voisin dans la région 3×3 de tous les niveaux DoG adjacents dans la même octave. Si le point souhaité est plus grand ou plus petit que tous ses voisins, il est choisi comme un point cible [116].

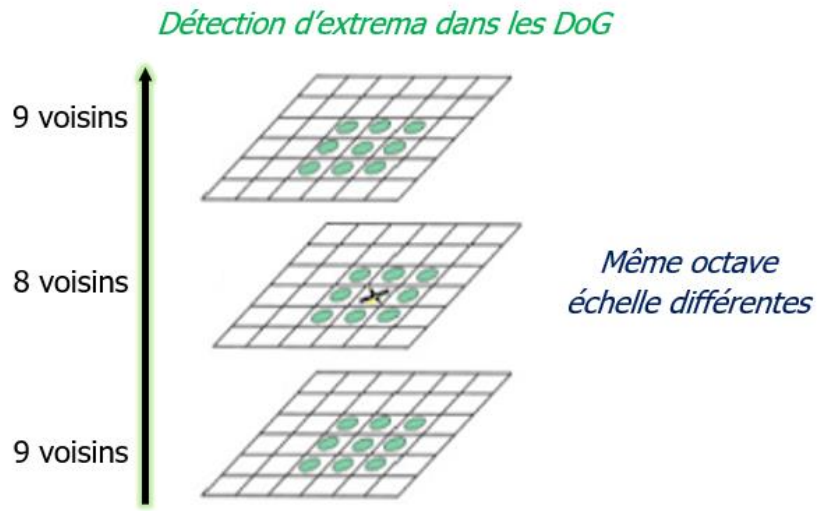

Figure 2.9 : Les maxima et les minima de l'image DoG [116].

La dernière étape consiste à supprimer tous les points avec un contraste faible et instable et les points sur le bord. Dans cette étape, le vecteur de propriété principal sera extrait ;il joue le rôle de la clé d'une image. Initialement, le champ de gradient et le chemin autour du point clé sont échantillonnés en utilisant la matrice 4×4 avec 8 directions pour l'histogramme. Par conséquent, le vecteur à une longueur de 128 pour chaque point. Ces vecteurs d'extraction des propriétés sont appelés points SIFT [141], La figure **2.10** montre un exemple de création des points clés de SIFT.

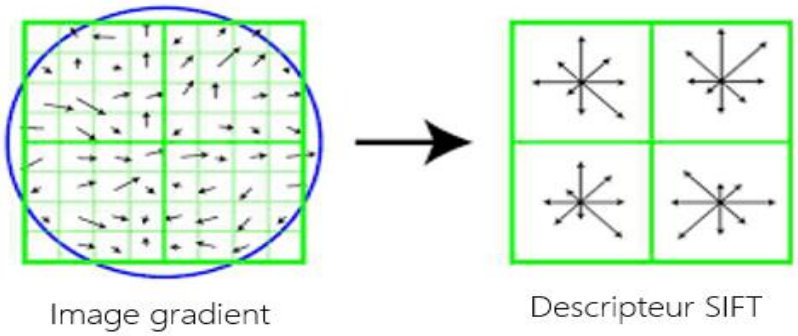

Figure 2.10 : Descripteur SIFT des points clés [116].

2.5.2 HOG : Histogramme de gradient orienté

Dala*et al.* [4] ont introduit un nouveau descripteur nommé Histogram Oriented Gradient (HOG)

qui utilise seulement l'information du contour. Le principe de ce descripteur est basé sur la distribution d'intensité des gradients ou de direction des contours pour décrire la forme d'un objet et l'apparence locale dans une image. HOG subdivise l'image en cellules, chaque cellule peut être d'écrite par un histogramme simule la distribution locale des orientations et des amplitudes du gradient pour chaque intervalle (bin). Il exploite l'information liée à l'amplitude et l'orientation du gradient de tous les pixels da la cellule. Le descripteur HOG présente l'avantage de pouvoir être calculé directement sur des cellules localisées dans l'image via deux composants du gradient, *amplitude* et *orientation*. Ce descripteur assure une invariance aux transformations géométriques et photométriques des larges régions d'espaces de la texture.

Pour mieux comprendre le principe du descripteur HOG, on commence par introduire la notion du gradient qui est capitale pour l'histogramme orienté gradient [4]. Soit $f(x_1, x_2, \dots, x_n)$ une fonction de n paramètre**,** un gradient $\boldsymbol{\nabla} f$ de cette fonction est défini comme suit :

$$\boldsymbol{\nabla} f = \left(\frac{\delta f}{\delta x_1}, \frac{\delta f}{\delta x_2}, \dots, \frac{\delta f}{\delta x_n}\right) \tag{2.17}$$

Le vecteur de gradient d'une fonction à deux dimensions (image) dans un plan est défini par deux composantes importantes : une amplitude (longueur) et une direction (orientation du contour). Dans la pratique, une image peut être considérée comme une fonction discrète a deux dimensions, ou les deux paramètres d'entrée x et y de cette fonction décrivent les coordonnées de chaque pixel de l'image, et la valeur d'intensité lumineuse par pixel est le résultat de la fonction. Le gradient est défini par les dérivées partielles de cette fonction. Dans les algorithmes de traitement d'image, la dérivée d'une image peut être approchée par le biais d'un filtrage différentiel. Deux filtres gradient utilisés pour calculer le gradient d'image, un pour chaque dimension, horizontale et verticale. On fait une combinaison des différentes entrées des deux cartes en un seul vecteur gradient représentant la position(x, y). L'approche standard dans ce cas est celle de Sobel [138] définie par deux filtres S_x et S_y comme ci-dessous :

$$S_x = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix} \tag{2.18}$$

et

$$S_y = S_x^T \tag{2.19}$$

Ces filtres sont appliqués à une image via une convolution discrète, donnée par :

$$I * F(x, y) = \sum_{i=1}^{3} \sum_{j=1}^{3} I(x, y) * F(x - i, y - j) \tag{2.20}$$

Où I représente l'image et F le filtre de Sobel. Pour chaque point (x, y) de l'image, on calcule l'amplitude $G(x, y)$ définiecomme suit :

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{2.21}$$

Où G_x et G_y représentent, respectivement, le gradient dans les directions x et y. L'orientation du contour en ce point est définie par la formule :

$$\theta(x,y) = \ arctg(\frac{G_y(x,y)}{G_x(x,y)}) \tag{2.22}$$

Pour construire le vecteur de descripteur de l'histogramme du gradient orienté, chaque image est subdivisée en cellules homogènes, voir la figure **2.11**. Les histogrammes de contour sont agrégés partialement pour chaque cellule. L'agrégation spatiale offre une robustesse vis-à-vis des petites variations de forme et réduit ainsi la taille du vecteur du descripteur, ceci rend le descripteur HOG très simple et surtout efficace pour les systèmes temps réels. L'idée de base alors est de projeter chaque pixel dans chaque cellule. Ensuite pour chaque cellule, on construit un histogramme et on concatène tous les histogrammes partiels dans un histogramme global HOG.

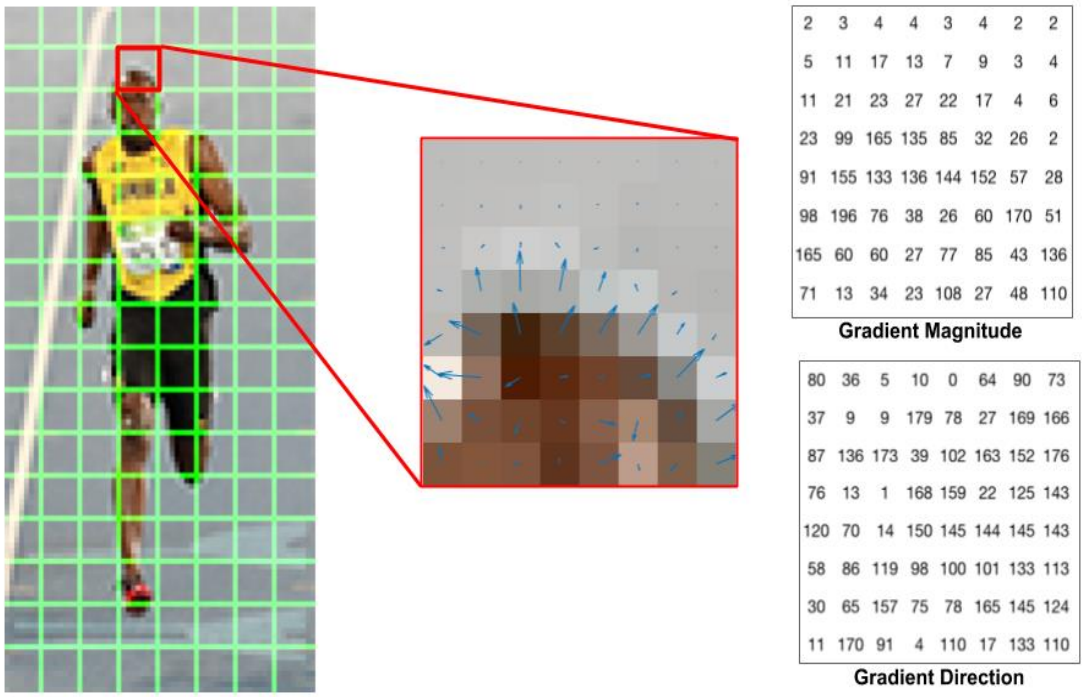

Figure 2.11 : La subdivision d'une image HOG en cellules [186].

En pratique, le nombre de cellules est huit et l'histogramme de chaque cellule est représenté par un vecteur de 9 classes. Chaque pixel de la cellule vote alors pour une classe de l'histogramme, en fonction de l'orientation du gradient à ce pixel. Le vote du pixel est pondéré par l'intensité du gradient en cepoint. Les histogrammes sont uniformes de 0 à 180 (cas non signé) ou de 0 à 360 (cas signé). Chaque classe est un intervalle d'orientations, ceci rend le descripteur HOG parmi les opérateurs les plus réduits en terme de taille plus sa robustesse élevée vis-à-vis de plusieurs paramètres : le changement de luminosité, la rotation et le mise en échelle. Une étape finale, essentielle dans la philosophie de l'opérateur de HOG est de normaliser indépendamment chaque histogramme de bloc cellule, pour avoir une norme constante par la division de chaque classe dans l'histogramme, voir la figure **2.12**.

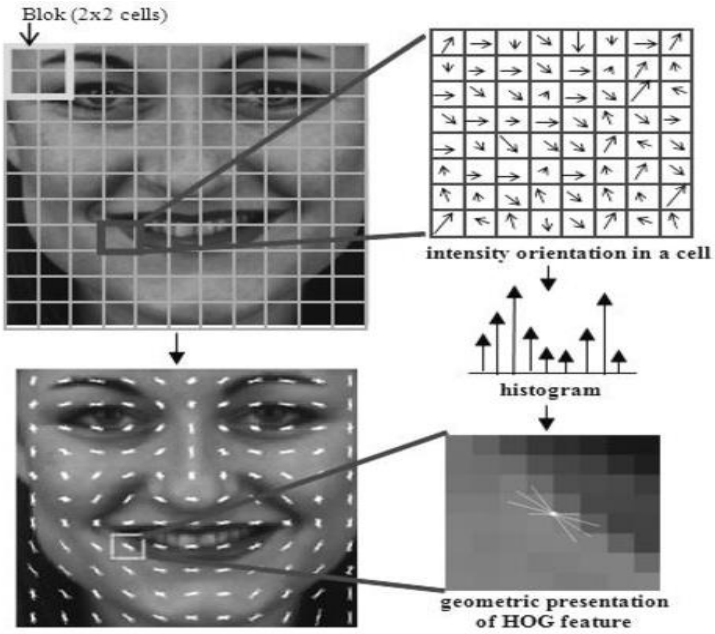

Figure 2.12 L'extraction du vecteur caractéristique HOG à partir d'un visage [187].

2.5.3 LBP : Méthode du motif binaire local

Le motif binaire local (Local Binary Pattern, LBP) est un descripteur d'images texturales qui peuvent définir la structure spatiale et le contraste local d'une image ou d'une partie de l'image. LBP est devenu un descripteur de texture standard. Ce descripteur est largement utilisé pour sa mise en œuvre simple de l'extraction du vecteur de caractéristiques avec un degré de précision élevé de classification. L'opérateur LBP est robuste vis à vis des changements uniformes de l'échelle de gris. Il est considéré aussi parmi les descripteurs les plus standard et efficace dans la description de la texture d'images. En effet LBP est considérée l'une des méthodes d'analyse de texture d'image les plus appropriées. Ojala *et al.* [3] ont introduit le descripteur LBP pour la première fois, il est indépendant de la rotation. Dans cette méthode, un voisinage est considéré en premier lieu pour chaque point de l'image. L'intensité du pixel central est ensuite comparée à l'intensité des pixels voisins. Si l'intensité lumineuse du pixel voisin est plus grande que le pixel central, alors la valeur de ce voisin dans le motif binaire extrait est considérée comme égale à un, sinon elle est nulle. Enfin, avec une somme binaire pondérée des valeurs dans le modèle d'extraction binaire, nous obtenons des valeurs à la base de dix. Cette somme est appelée la valeur LBP. Le calcul du LBP final est défini dans l'équation suivante :

$$LBP(I_c) = \sum_{i=1}^{p-1} S(I_p - I_c) * 2^p \qquad (2.23)$$

Où :

- P : le nombre de voisins pour le pixel central
- I_c: intensité lumineuse du pixel central.
- I_p: intensité lumineuse du pixel voisin.

- $S(x)$: fonction de signe pour générer des valeurs binaires. Elle est définie selon l'équation (27)

$$S(x) = \begin{cases} 1, & si \quad x \geq 0 \\ 0, & si \quad x < 0 \end{cases} \tag{2.24}$$

Le descripteur LBP est appliqué sur tous les pixels de l'image comme le montre la figure **2.13**. L'histogramme LBP peut être utilisé pour extraire les caractéristiques de textures de l'image. Afin d'éviter la sensibilité de l'opérateur LBP à la rotation, le voisinage est considéré comme un cercle et les intensités lumineuses des points dont les coordonnées ne correspondent pas exactement au centre du pixel mais elles sont déterminées par interpolation.

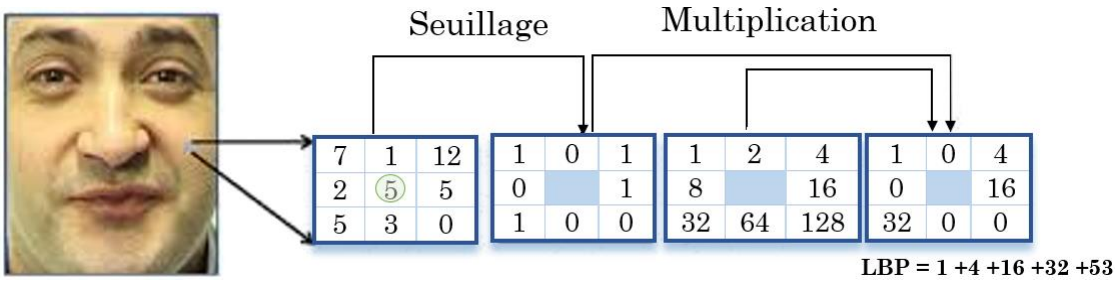

Figure 2.13 Exemple de calcul d'un code LBP [3].

La figure 2.14 montre un voisinage circulaire avec un rayon R différent et un nombre de voisins du point P. La propriété la plus importante de LBP réside dans son invariance vis à vis des changements monotones de l'illumination causés par des variances d'éclairage pour les applications du monde réel. Une autre propriété aussi importante réside dans sa simplicité de calcul en temps réel.

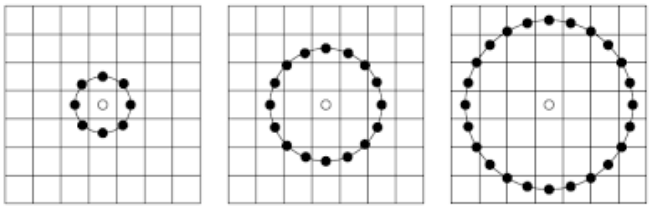

Figure2.14 Voisins symétriques circulaires pour différentes valeurs de Pet R [3].

Le descripteur LBP est un nombre binaire de $P - bits$ de sorte que soit 2^pvaleurs différentes [5]. Lorsque on aaucune image pivot, tous les voisins tourne autour du pixel central $I_c$dans une seule direction, ce qui entraîne la production de valeurs différentes pour $LBP_{p,r}$. Cela signifie que la valeur de $LBP_{p,r}$ dépend de l'indexation des pixels voisins. Un opérateur de rotation du bit vers la droite est introduit de sorte qu'une valeur unique à chacun des motifs est affectée pour éliminer les destructeurs dus aux effets de rotation [3]. Cet opérateur de rotation indépendant est calculé comme suit :

$$LBP^{ri}_{P,R}(x,\ y) = \min\left(ROR\left(LBP^{ri}_{P,R}(x,\ y), i\right), i\ \in [0, P-1]\right) \quad (2.25)$$

Où : RORest la fonction de rotation d'un P_bits vers la droite et r_i est l'insensibilité de l'opérateur à la rotation.

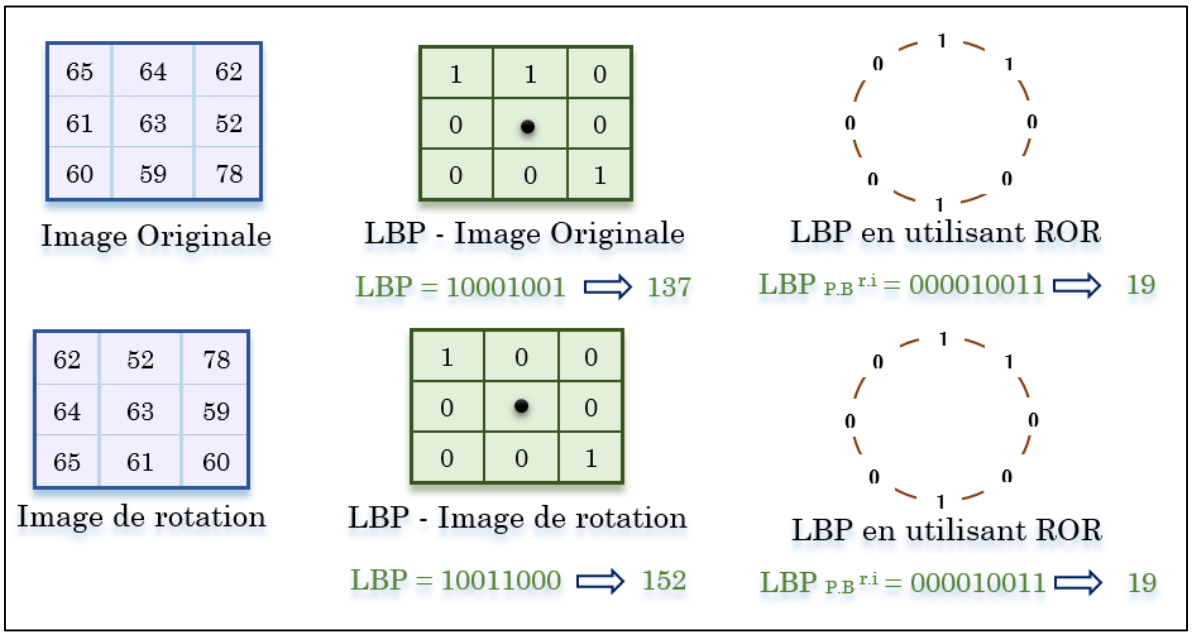

Figure 2.15 Un processus LBP utilisant la rotation de bits vers la droite ($LBPP^{ri}_{P,R}$) [3].

La rotation binaire est effectuée i fois et le nombre minimum obtenu pour chaque i entre 0 et $P\ -\ 1$ est la valeur finale du LBP. De nombreuses extensions ont été proposées pour améliorer LBP. La première extension de LBP est appelée LBP uniforme (MLBP), a été introduite par Ojala [3].

Description du visage à l'aide de LBP

Le descripteur locale LBP et ses variantes ont suscité un intérêt ces derniers temps, ce qui est compréhensible étant donné les limites des représentations holistiques. Ces méthodes basées sur les caractéristiques locales sont plus résistantes aux variations de pose ou d'éclairage que les méthodes holistiques. Le principe de l'approche LBP pour la classification des textures est basé sur le regroupement des occurrences des codes LBP à partir d'une image de visage et les collectées dans un histogramme. Ensuite, La classification est réalisée en calculant de simples similitudes d'histogramme. Cependant, considérer une approche similaire pour la représentation de l'image faciale entraîne une perte d'informations spatiales et par conséquent, il convient de codifier les informations de texture tout en conservant également leurs emplacements. Une façon d'atteindre cet objectif consiste à utiliser les descripteurs de texture LBP pour créer plusieurs descriptions locales du visage et les combiner en une description globale. Le principe de base pour la description du visage basée sur la LBP proposée Ahonen et al. [3] est le suivant : l'image faciale est divisée en blocs locales et les descripteurs de texture LBP sont extraits de chaque cellule indépendamment sous forme d'histogrammes locaux. Les histogrammes ainsi

construits sont ensuite concaténés pour former un Histogramme globale du visage, comme indiqué dans la figure 2.16.

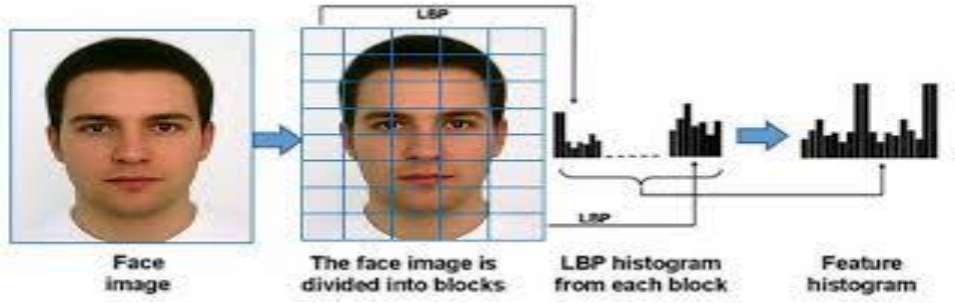

Figure 2.16 Description du visage avec des motifs binaires locaux LBP[202].

Dans ce qui suit on va présenter les opérateurs variantes du LBP qui sont : MLBP, LTP, LDP, LDN, LGC, LGP, LPQ et WLD. Tous ces descripteurs utilisent le même principe que celui du LBP en ce qui concerne sa philosophie général,ils se diffèrent seulement dans la façon du codage binaire des pixels d'image.

MLBP : Méthode du motif binaire local modifié

LBP modifié (Modified Local Binary Pattern, MLBP) correspondant au caractère uniforme de nombre de mutations qui se produise entre 0 et 1 (et vice versa) dans le motif binaire local extrait à partir des voisins, il est représenté par le symbole U. Par exemple, les nombres binaires 00000000 et 11111111 ne montrent aucune mutation par contre le nombre binaire 11010110 montre six mutations. De cette façon, l'opérateur LBP modifié est invariant à la rotation. Un nombre de mutations entre les bits est définie comme suit :

$$U(LBP) = \left|S(g_{p-1} - g_c) - S(g_0 - g_c)\right| + \sum_{p=0}^{p-1}\left|S(g_{p-1} - g_c) - S(g_0 - g_c)\right| \qquad (2.26)$$

Dans MLBP, les motifs avec une homogénéité inférieure ou égale à UT sont définis comme des motifs uniformes, et les motifs avec une homogénéité supérieure à UT sont définis comme des motifs hétérogènes. L'homogénéité des motifs est définie par :

$$LBP_{p,r}^{riu2} = \begin{cases} \sum_{p=0}^{p-1} S(g_p - g_c), & si \quad U(LBP) < U_T \\ P + 1, & sinon \end{cases} \qquad (2.27)$$

Où les étiquettes 0 à P sont attribuées à des voisins homogènes et l'étiquette $P + 1$ est attribué à un voisin hétérogène. Après avoir appliqué l'opérateur MLBP sur des images texturales où différentes étiquettes sont attribuées, la probabilité d'une étiquette particulière peut être approximée par le ratio du nombre de points étiquetés au nombre total de points dans l'image entière. Finalement, un vecteur propre de probabilité de $P + 2$ pour chaque image d'entrée est obtenue. Selon Ojala [3], UT est

généralement considéré comme $P/4$, et la petite quantité de motifs dans la texture est étiquetée comme $P+1$.

2.5.4 LTP : Modèle ternaire local

Le descripteur LBP est très sensible au bruit aléatoire pour surmonter ce problème, Tan et al.[126] ont développé une généralisation de LBP appelée Local Ternary Pattern (LTP). Au contraire du LBP standard qui un code à 2 valeurs, le LTP est codé sur 3 valeurs, dans lesquelles les niveaux de gris dans une zone de largeur $\pm\, t$ autour de I_c sont quantifiés à zéro, au-dessus sont quantifié à $+1$ et en dessous sont quantifiés à -1 (voir la figure 2.16). Pour le descripteur LTP, $S(x)$ est calculée par la formule (2.28). Le problème avec cette méthode est qu'il faut toujours fixer la valeur seuil qui n'est pas assez simple.

$$S(x) = \begin{cases} 1, & si \quad (I_c \geq I_p + t) \\ 0, & si \quad (I_c > I_p - t) \ et \ (I_c < I_p + t) \\ -1, & si \quad (I_c \leq I_p - t) \end{cases} \qquad (2.28)$$

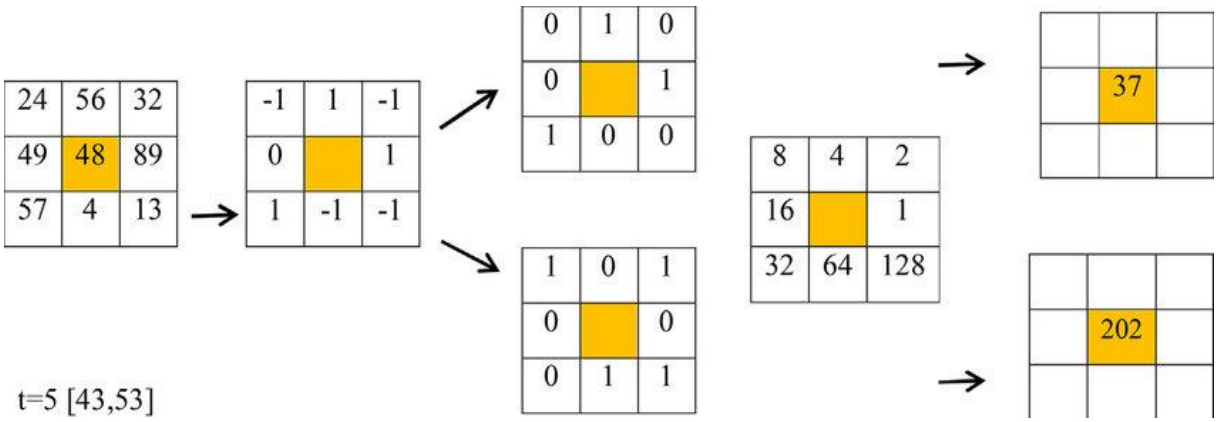

Figure 2.17 Un exemple de calcul du code LTP pour un pixel (t = 5) [126].

2.5.5 LDP : Modèle directionnel local

Le Modèle directionnel local (Local Directionnel Pattern, LDP) [11] appartient à la même famille des opérateurs LBP, c'est un motif de texture en niveaux de gris qui caractérise la structure spatiale d'une texture locale d'une image. L'opérateur LDP calcule les valeurs de réponse de contour dans les huit directions pour chaque pixel et produit un code de la magnitude de force relative. Les réponses du contour sont invariantes aux changements d'illumination et au bruit. La fonctionnalité LDP résultante décrit les primitives locales, y compris différents types des courbes, des coins et des jonctions, de manière plus stable et conserve plus d'informations.

Soit un pixel central d'une image, les huit valeurs de réponse du contour directionnel $M_i \ \ i = 0, 1, \ldots, 7$ sont calculées à partir des masques de Kirsch $M_i$dans huit orientations différentes centrées

sur sa position. Les masques sont illustrés à la figure **2.17**. Les valeurs de réponse ne sont pas tous de la même importance dans les directions. La présence d'un coin ou d'un bord provoque des valeurs de réponse élevées dans certaines directions. Il est intéressant que les k-directions les plus importantes (top kbits)seront générés par LDP [11], Le top k bits directionnel sont les réponses b_i qui sont définies par 1. Les $(8 - k)$ bits restants du modèle LDP sont définis par 0.Enfin, le code LDP est calculé selon l'équation 1. La figure **2.18** montre les réponses du masque et les positions des bits LDP, et la figure 3.14 montre un exemple de code LDP avec $k = 3$.

Figure 2.17 Les huit réponses du masque de Krish.

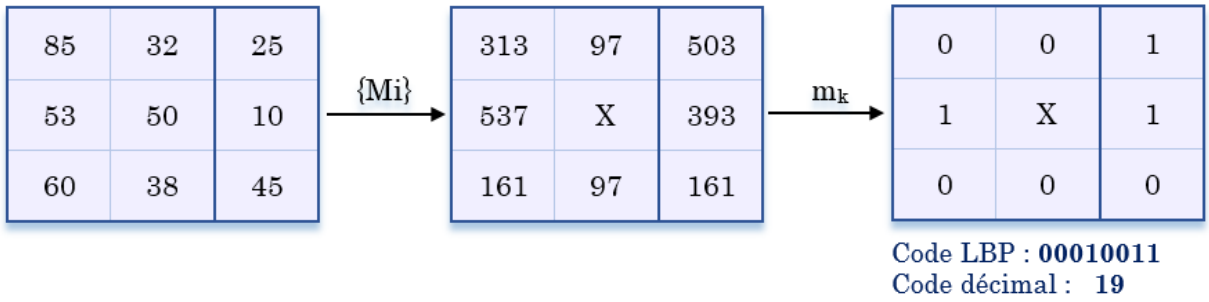

Figure 2.18 Exemple de calcul du code LDP pou k = 3.

2.5.6 LDN : Modèle de nombre directionnel local

Beaucoup de descripteurs soufferts du problème de luminosité qui influe largement les magnitudes des contours dans beaucoup de descripteurs tel que LBP et LDP. A cet effet, Rivera *et al.* [127] proposent un nouvel opérateur appelé Local Directional Number (LDN). LDN est un code binaire à six bits qui est attribué à chaque pixel de l'image d'entrée représentant ses transitions d'intensité ainsi que la structure de la texture. Le micro-motif est obtenu en calculant la réponse de contour du voisinage à l'aide d'un masque de Kirsch et en prenant les nombres directionnels supérieurs, c'est-à-dire les nombres

de direction les plus positifs et négatifs de ces réponses de bord qui indiquent les directions du brillant et les zones sombres du voisinage, voir la figure **2.19**.

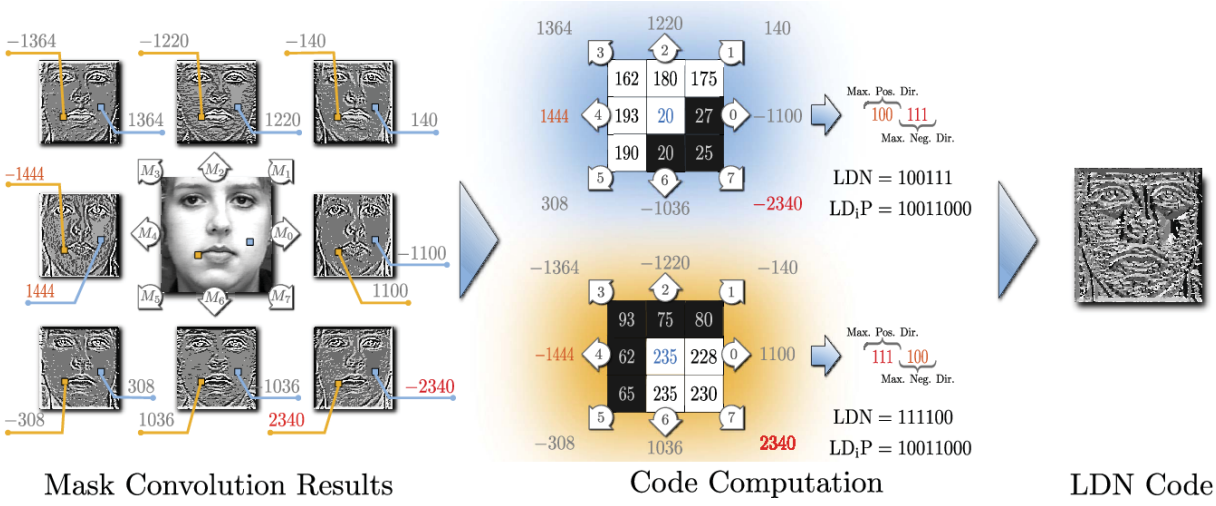

Figure 2.19 Exemple de calcul du code LDN [175].

A partir de ces directions positives et négatives supérieures, un micro-motif LDN à 6 bits est généré et on obtient une image LDN. Cette image LDN est ensuite divisée en blocs et à partir de chaque bloc un histogramme LDN (LH) est calculé. Ces LH sont concaténés pour former le vecteur de caractéristiques appelé descripteur de visage. Rivera *et al.* [127] a étendu le LDN au motif de texture directionnelle locale (LDTP) pour distinguer les expressions des personnes et les différentes scènes de paysages.

2.5.7 LGC : Codage Gradient Locale

LGC (Local Gradient Coding) proposé par [9]. Au contraire du LBP qui compare la valeur d'intensité du pixel du centre avec les intensités de ses voisins, LGC compare les valeurs de l'intensité des pixels au voisinage d'un pixel de centre loin de ce pixel, Il calcul les gradients des pixels du voisinage d'un pixel donné. LGC regroupe les gradients horizontaux et verticaux et diagonaux respectivement en utilisant une modèle 3*3 comme montré dans la figure. La valeur de chaque pixel I_c est définie comme suit :

g_1	g_2	g_3
g_4	g_c	g_5
g_6	g_7	g_8

Figure 2.20. Une région locale de 3x3 pixels[9].

$$LGC(x_c, y_c) = s(g_1 - g_3) * 2^7 + s(g_4 - g_5) * 2^6 + s(g_6 - g_8) * 2^5 + s(g_1 - g_6) * 2^4 + s(g_2 - g_7) * 2^3 + s(g_3 - g_8) * 2^2 + s(g_1 - g_8) * 2^1 + s(g_3 - g_6) * 2^0 \quad (2.29)$$

Tong *et al.*[9] proposent une variante de l'opérateur LGC, appelée LGC-HD. Les auteurs prouvent que l'opérateur LGC-HD (LGC-based Horizontal and Diagonal Gradient Prior Principal) est plus puissant et plus rapide. Cet opérateur selon Tong *et al.* [9] est capable de réduire le temps d'exécution et augmente le taux de reconnaissance par rapport au LGC original.

L'opérateur LGC-HD est défini comme suit :

$$LGC = s(g_1 - g_3) * 2^4 + s(g_4 - g_5) * 2^3 + s(g_6 - g_8) * 2^2 + s(g_1 - g_8) * 2^1 + s(g_3 - g_6) * 2^0 \quad (2.30)$$

2.5.8 LGP : Local Gradient Pattern

Le descripteur Motif de Gradient Local (LGP), appelée Local Gradient Pattern, proposé par Islam [169], est conçu de maniéré simple et efficace pour l'extraction de l'information de texture dans une image. Il appartient à la famille des LBP. La méthode de représentation de texture est basée sur les différences de gradient de couleur du visage d'une région locale de 3x3 pixels (voir figure **2.20**).

a_1	b_1	a_2
b_4	C	b_2
a_4	b_3	a_3

Figure 2.21.Une région locale de 3x3 pixels[169].

Les représentations fonctionnent sur des images faciales en niveau de gris ou sur des objets en général. Le pixel « C » sur la figure 2.20 est représenté par deux motifs binaires à 2 bits. Les valeurs de a_1 à a_1 et les valeurs de b_1 à b_4 représentent les intensités lumineuses des pixels correspondants. Le « bit-1» de « Pattern-1» est 0 si $a_1 \leq a_3$ sinon, il vaut 1. De manière similaire, d'autres bits sont également calculés à l'aide de la formule illustrée à la figure **2.22**. Le « Pattern-1» peut avoir $2^2 = 4$ combinaisons possibles. Ainsi, quatre bits sont nécessaires pour le « motif-1». De même, quatre autres bits sont nécessaires pour le motif-2 (voir Figure 2.23). Cette nouvelle représentation s'appelle LGP. LGP est calculé pour chaque pixel de l'image. Un exemple d'obtention de LGP est illustré à la figure **2.24**. Donc pour le pixel de valeur d'intensité 25, les deux cases 3 et 6 seront augmentées de un. LGP permet de réduire en taille en plus il est très efficace comme opérateur d'extraction de l'information.

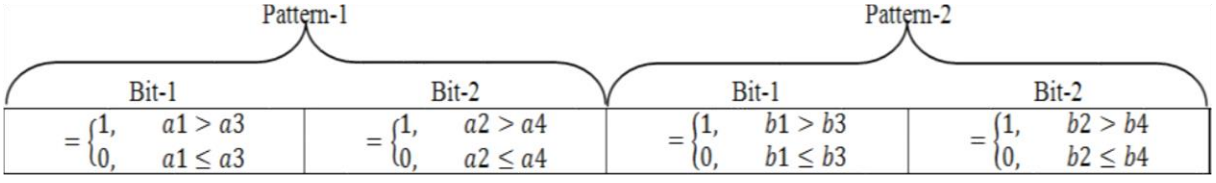

Figure 2.22 Représentation de pixel unique utilisant deux modèles distincts [169].

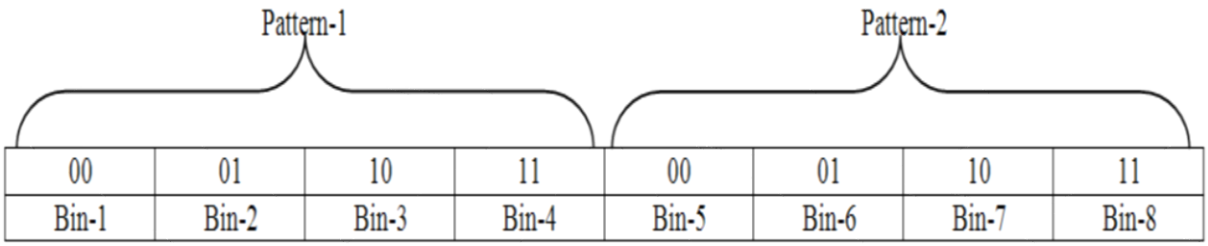

Figure 2.23 Représentations bin pour les motifs [169].

18	**25**	**14**
85	**25**	**86**
45	**65**	**14**

Pattern 1 = **10**

Pattern 2 = **01**

Figure 2.24. Exemple de codage de LGP [169].

2.5.9 LPQ : Quantification par phase locale

Le descripteur LPQ ou quantification par phase locale a été proposé pour la première fois en 2008 par Ojansivu *et al.* [10]. Ce descripteur permet de classifier des images comportant des textures floues, ce qui ne peut pas être réalisé par le descripteur standard LBP. LPQ est invariant au texture fou causé par le mouvement des capteurs ou par le mouvement des expressions faciales telles que le rire, le sourire, la colère, la surprise, etc. Ce descripteur a connu un large intérêt dans ces dernières années dans le domaine de reconnaissance de visage et il a été démontré qu'il est très efficace dans la reconnaissance de visages avec des images floues ou nettes. La figure **2.25** montre un exemple de codage d'un pixel image avec LPQ.

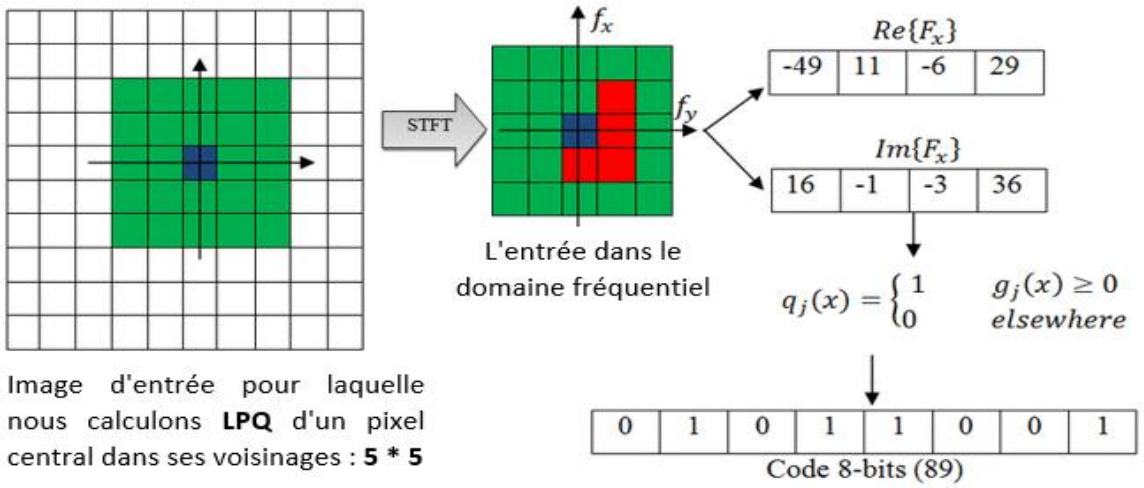

Figure2.25 : Exemple de codage d'un pixel image avec LPQ [176].

Le descripteur de quantification par phase local est basé sur les propriétés de flou-invariant du spectre de phase de Fourier. Il utilise l'information locale de la phase extrait par le 2D DFT, ou plus

précisément la transformé de Fourier court terme (STFT Short-Term Fourier Transform) calculé sur un rectangle N_x de $N * M$ voisins pour chaque pixel x de l'mage $f(x)$ définie par :

$$F(x,y) = \sum_{y \in N_x} f(x-y)\, e^{2j\pi u^T y} = w_u^T f_x \quad (2.31)$$

Où

- w_u : vecteur de base de la décomposition à la fréquence u.
- f_x : les valeurs de l'image appartenant au voisinage deN_x.

LPQ ne considère que quatre coefficients complexe pour le calcul de la transformé de Fourier en 2D : $u1 = [a;\ 0]^T, u2 = [0;\ a]^T, u3 = [a;\ a]^T, u4 = [a; a]^T$, a est une fréquence scalaire, soit :

$$F_x^c = [F(u_1,x), F(u_2,x), F(u_3,x), F(u_4,x)] \; et \; F_x = [Re(F_x^c),\ Im(F_x^c)] \quad (2.32)$$

Avant l'étape de la quantification**,** les coefficients doivent être décolérer, parce qu'il peut être montré que l'information est préservée au maximum dans la quantification scalaire si les échantillons quantifiés sont statistiquement indépendants. En supposant que la distribution est gaussienne, l'indépendance peut être obtenue à l'aide d'une transformation de blanchiment.

$$G_x = V^T F_x \quad (2.33)$$

OùV est une matrice orthonormale dérivée de la décomposition en valeurs singulières (SVD) de la matrice D qui est définie par :

$$D = U \sum V^T \quad (2.34)$$

Par la suite, G_x est calculé pour chaque position de l'image et les vecteurs résultants sont quantifiés à l'aide d'un simple quantificateur scalaire :

$$q_j = \begin{cases} 1 & si \quad g_i < 0 \\ 0 & sinon \end{cases} \quad (2.35)$$

Où q_j est la j[ème] composante de G_x. Les coefficients quantifiés sont représentés sous forme de valeurs entières comprises entre 0 et 255 utilisant le codage binaire.

$$LPQ(x) = \sum_{j=1}^{8} q_j\, 2^{j-1} \quad (2.36)$$

2.5.10 WLD : Descripteur Weber Locale

WLD [125] est un descripteur efficace et très robuste. Ce descripteur regroupe deux composantes : excitation et orientation différentielle. La composante d'excitation différentielle est une fonction du rapport entre deux termes : l'une représente les différences d'intensités relatives d'un pixel courant par rapport à ses voisins ; l'autre est l'intensité du pixel courant. La composante d'orientation est quant à elle, l'orientation du gradient du pixel courant est inspirée de la loi de Weber. Ernst Weber, un psychologue du 19[eme] siècle, a observé que le rapport du seuil d'incrément à l'intensité de fond est une

constante [125]. Cette relation, connue depuis sous le nom de la loi de Weber et peut être s'exprimer comme suit :

$$\frac{dI}{I} = k \tag{2.37}$$

Où dI : Seuil d'incrémentation ;

I : Intensité initiale du stimulus.

k : signifie que la proportion sur le côté gauche de l'équation reste constante malgré les variations du terme I.

La fraction dI/I est connue sous le nom de Fraction de Weber.La loi de Weber, plus simplement dit que la taille d'une différence juste perceptible (dI) est une proportion constante de la valeur de stimulus d'origine.

WLD calcul pour chaque pixel d'une image deux composants, l'excitation différentielle et l'orientation du gradient. Une image ou une région d'une image sera représentée par un histogramme qui est une combinaison des traits des pixels. Le descripteur WLD utilise les avantages du descripteur SIFT en ce qui concerne le calcul de l'histogramme en utilisant les gradients et ses orientations et celle de l'opérateur LBP pour l'efficacité de calcul et les petites régions utilisés dans la représentation. Le descripteur SIFT est un histogramme 3D des gradients et des orientations des gradients dans lesquelles l'image est représentée par des coordonnées spatiales plus une dimension pour l'orientation des gradients. La figure **2.24** illustre le processus de reconnaissance des visages par WLD. Mais au contraire du descripteur SIFT qui est appliqué sur des régions séparées, WLD utilise et exploite chaque pixel de l'image et dépend à la fois de la variation d'intensité locale et de la magnitude de l'intensité du pixel central, autrement dit il combine aussi le principe des descripteurs HOG et LBP. En ce qui concerne le descripteur LBP, il représente une image d'entrée en construisant des statistiques sur les variations locales de micro-motifs. Ces motifs locaux peuvent correspondre à des tâches claires/ sombres, des bords et des zones planes, etc. En revanche, WLD calcule d'abord les micro-motifs saillants, puis construit des statistiques sur ces motifs saillants ainsi que l'orientation du gradient du point actuel.

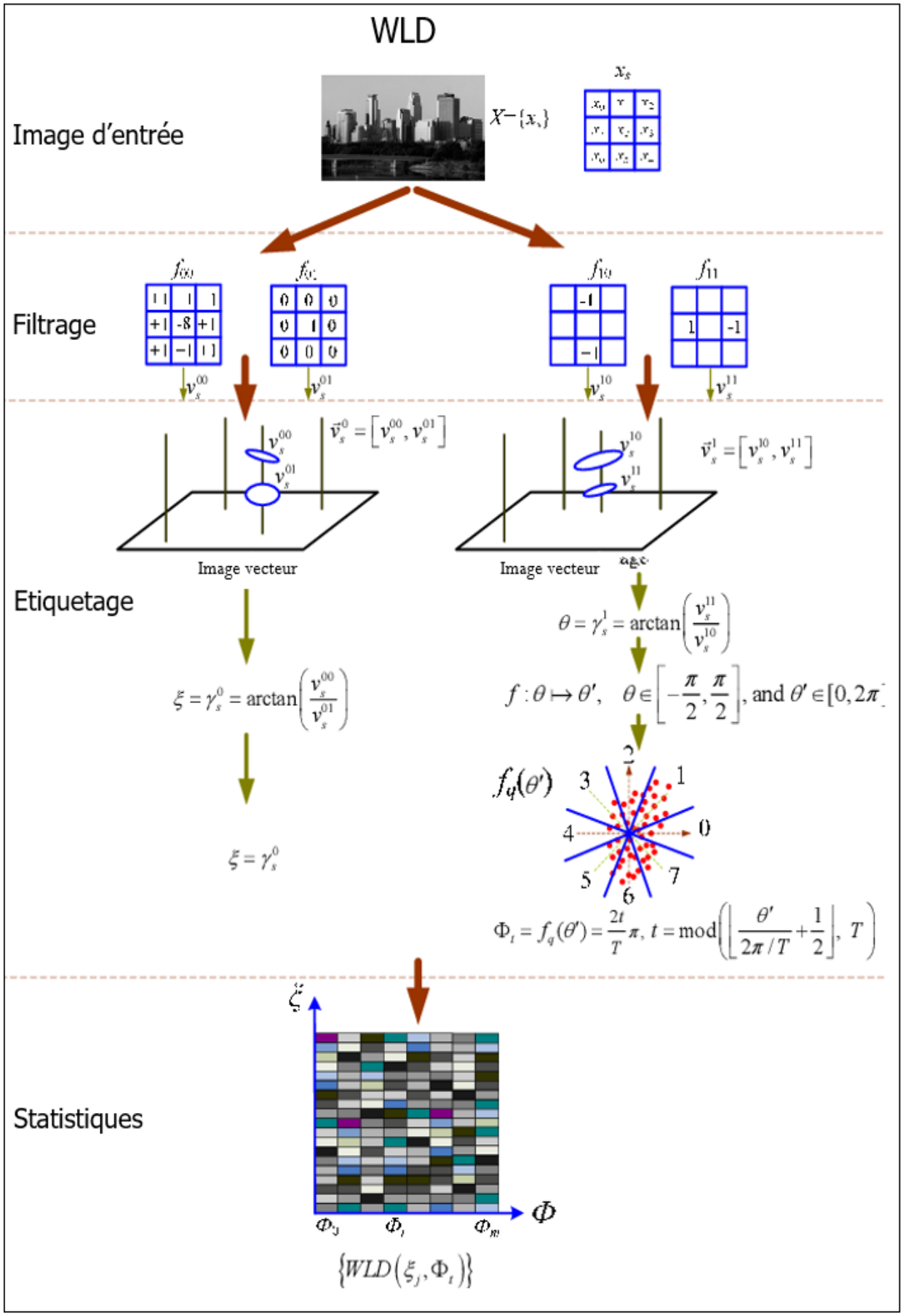

Figure 2.26 : Reconnaissance des visages par le Descripteur Weber Locale [167].

2.6 Conclusion

A travers ce deuxième chapitre nous avons présentés le principe de l'analyse de la texture en vision ordinateur et son application dans le domaine de la classification de texture. On a défini les quatre catégories principales des méthodes de classification des textures. Dans la matrice de cooccurrence de catégorie statistique et les méthodes de Patten binaires locales sont les plus populaires, et pour la catégorie basée sur le modèle, les modèles fractals sont les plus connus, Gabor et Wavelet sont plus applicables via la catégorie basée sur la transformation. La majorité des méthodes sont classées dans les méthodes statistiques et basées sur la transformation ou une combinaison de ces méthodes. Une raison principale de la diversité des méthodes est que le changement concerne l'analyse de l'image de texture (par exemple, bruit, rotation, échelle, éclairage, point de vue). En conséquence, chaque nouvelle méthode recherchée surmonte certains défis. Presque toutes les méthodes sont invariantes par rotation ; cependant, la majorité des méthodes sont sensibles au bruit. Ensuite on a présenté les descripteurs orienté traitement et extraction des caractéristiques de la texture. On a concentré sur les descripteurs les plus populaires et les plus utilisés dans le domaine de reconnaissance du visage et la reconnaissance des expressions faciales. On peut dire que Les deux opérateurs HOG et LBP sont devenus maintenant des descripteurs standard et utilisés dans plusieurs plateformes comme primitive essentiels dans l'analyse du visage (détection, suivi, reconnaissance et expressions faciales).

Chapitre 3

Apprentissage automatique

Sommaire

3.1 Introduction

En 1950, Alan Turing [129] a été le premier à introduit le terme d'Intelligence Artificielle (IA), pour rendre les machines intelligentes. Mais son application effective n'a jamais été aussi efficace que ces dernières années. En1956, après la conférence scientifique qui s'était tenu au Collège Darmouth, IA vient de commencer à entrer dans le domaine scientifique. Dans cette décennie les applications d'IA prouvent ses performances d'algorithmes d'intelligence. En 2012, l'IA a réalisé une véritable révolution scientifique avec l'apparition de l'apprentissage profond (*en anglais, Deep Learning DL*) et le développement rapide des composants informatiques, qui sont devenus de plus en plus puissants et moins chers. Tout cela marque la naissance d'une quatrième révolution industrielle. D'un autre côté et

avec le coût faible de stockage des données, cette révolution scientifique d'IA a lancé des plans pour optimiser la collecte de données multimédias : *images*, *vidéos*, *dossiers médicaux*, *données cartographiques*. Aujourd'hui, l'IA attire de plus en plus de nouveaux domaines dans le traitement d'images, la programmation informatique, la médecine et même l'agriculture. L'intelligence artificielle est l'ensemble des théories et des techniques mises en œuvre en vue de fabriquer des machines capables de simuler l'intelligence humaine [130]. Autrement dit, il s'agit d'un ensemble de techniques permettant aux machines d'effectuer des tâches et de résoudre des problèmes normalement réservés aux humains. L'apprentissage automatique (*Machine Learning ML*) est un domaine de l'IA qui permet à la machine d'analyser, de résoudre des problèmes et de mettre en œuvre des solutions grâce à l'utilisation massive de données et d'algorithmes d'apprentissage [131]. L'apprentissage profond (*Deep Learning DL*) est un sous-domaine de Machine Learning. Le Deep Learning simule les réseaux de neurones biologiques du cerveau de l'être humain pour permet à la machine d'apprendre et de progresser à travers son expérience [132]. La figure **3.1** montre l'Intelligence Artificielle, Machine Learning et Deep Learning. Les intersections entre ces domaines sont illustrées dans la figure **3.2**.

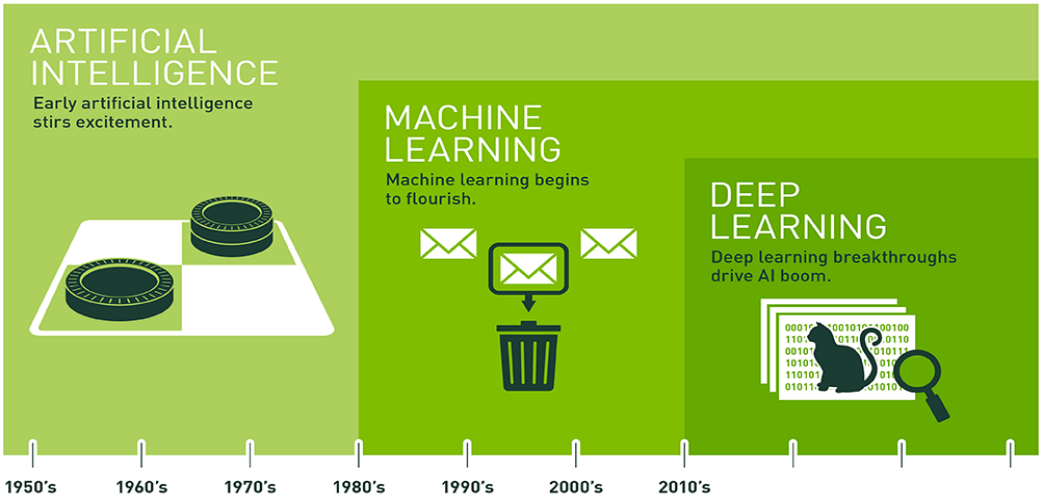

Figure 3.1 La relation entre l'Intelligence Artificielle, le ML et l'apprentissage en profondeur [189].

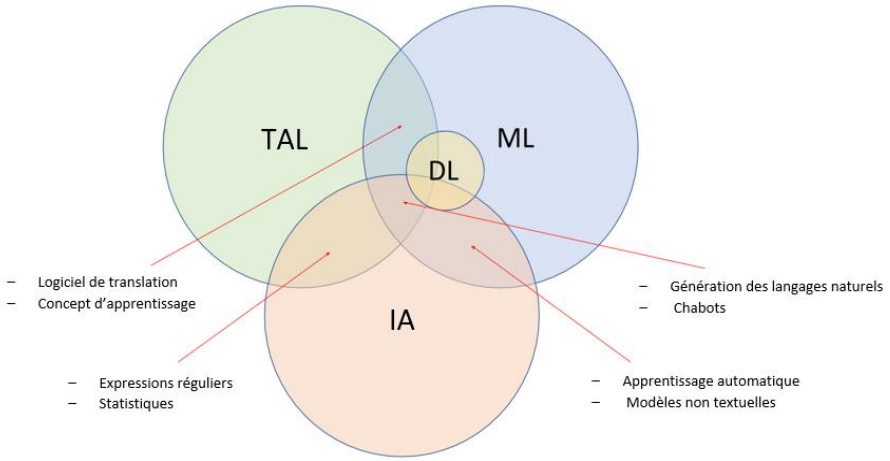

Figure 3.2 L'intersection de l'Intelligence Artificielle, le ML et l'apprentissage en profondeur [190].

3.2 Apprentissage automatique

Dès qu'un concept d'apprentissage entre en jeu, de nombreux chercheurs scientifiques rêvent de passer le test de Turing, mentionné dans un article publié en 1950 sur les ordinateurs et l'intelligence [133]. Ce test consiste à initier une conversation masquée entre un humain, une machine et un autre humain. Une fois que le premier humain est incapable d'identifier l'interlocuteur alors qu'elle est une machine, cet dernier réussit au test de Turing.

En 1959, l'informaticien Arthur Samuel utilise pour la première fois le terme « Machine Learning», lors de la présentation de son programme au profit de la société IBM. Il a créé un système qui a permis au programme d'apprendre de chaque position qu'il avait déjà vue via les opportunités qu'il offrait. Il est également joué contre lui-même comme un autre moyen d'apprendre. Grâce à ces données, Arthur Samuel a évolué son programme et a pu battre le 4ème joueur aux États-Unis. C'est également le premier programme informatique à jouer à un jeu de société à un niveau avancé [134].

En 1997, avec l'apparition du système Deep Blue, le monde a connu le véritable couronnement de l'apprentissage automatique. En effet, le supercalculateur a pu vaincre le champion du monde dans le jeu d'échecs Garry Kasparov. Cet incident à donner naissance à plusieurs projets d'apprentissage d'IA. En 2014, Google Deep Mind une société britannique, développe AlphaGo, un autre exemple de succès d'IA qui défi l'être humain et le vaincu.

De manière générale, les systèmes d'apprentissage automatique sont des systèmes de classification ou de reconnaissance de formes. Un système d'apprentissage automatique est une boite noire avec une entrée, par exemple une image, un son, ou un texte, et une sortie qui peut représenter la classe de l'objet dans l'image, la parole ou le sujet dans le texte [135]. Les internautes contribuent depuis de nombreuses années à former des systèmes d'apprentissage automatique. Les premiers programmes d'apprentissage automatique servent à identifier un chat, un panneau STOP ou un feu de circulation. Bien sûr, au début l'AI faisait à peine la différence entre l'image d'un chat et celle d'un chien, pour que le programme comprenne ce qui est un chat, un superviseur apprend au programme que cette image contient un chat. Mais pour progresser, il faut développer le bon algorithme d'apprentissage, et cela prend du temps très important. Aujourd'hui l'apprentissage automatique se développe de plus en plus particulièrement dans les systèmes experts ou les super calculateurs avec le grand saut qu'il a fait grâce aux réseaux de neurones artificiels et au Deep Learning [136].

Dans sa forme la plus simple, l'apprentissage automatique est en mode supervisé [135]. Nous présenterons à l'entrée de la machine une photo d'un objet, par exemple une voiture, et on lui donne la sortie souhaitée pour une voiture. Ensuite, nous lui montrons la photo d'un chien avec la sortie souhaitée pour un chien. Après chaque exemple, la machine ajuste ses paramètres internes afin de rapprocher sa sortie de la sortie souhaitée. Après avoir montré des milliers ou des millions d'instances étiquetés à la machine par sa catégorie, la machine devient impérative de classer correctement la plupart

d'entre eux. Mais ce qui est plus intéressant, c'est qu'il peut également classer correctement les images de voiture ou dechien qu'elle n'a jamais vues pendant la phase d'apprentissage. C'est ce qu'on appelle la *capacité de généralisation* [135].

Les systèmes classiques de reconnaissance de forme sont devisés en deux modules principaux : un module d'extraction de caractéristiques et un module de classification [135 – 137]. Le module d'extraction de caractéristiques est un programme qui transforme la matrice de nombres représentant l'image, par exemple, en un vecteur de caractéristiques, dont chacune indique la présence ou l'absence d'un motif simple dans l'image. Ce vecteur est envoyé au classificateur qui calcule une somme pondérée des caractéristiques. Si la somme dépasse un seuil, la classe est reconnue. L'estimation des poids est relative de la classe à laquelle le vecteur de caractéristiques est comparé. Les poids varient pour les classificateurs de chaque catégorie, ce sont ceux qui sont modifiés pendant l'apprentissage. Les premières méthodes de classification linéaire entraînable datent de la fin des années 1950 et sont encore largement utilisées aujourd'hui. Ils prennent les deux noms de perception ou de régression logistique.

Dans le domaine de l'apprentissage automatique, il existe deux principaux types de tâches (voir figure **3.3**): supervisées et non supervisées [135]. L'apprentissage supervisé est la technique permettant d'accomplir une tâche en fournissant aux systèmes des modèles d'apprentissage, d'entrée et de sortie [135]. Tandis que l'apprentissage non supervisé est une technique d'auto-apprentissage dans laquelle le système doit découvrir par lui-même les caractéristiques de la population d'entrée [135]. La principale différence entre les deux types est que l'apprentissage supervisé est basée sur une connaissance préalable de ce que devraient être les valeurs de sortie de nos échantillons. Par conséquent, le but de l'apprentissage supervisé est d'apprendre une fonction à partir d'un échantillon de données et des résultats désirés, la meilleure approximation de la relation entre l'entrée et la sortie observables dans les données. En revanche, l'apprentissage non supervisé n'a pas de résultats étiquetés. Son objectif est donc de déduire la structure naturelle présente dans un ensemble de points de données. Il est important de souligner que ces deux types d'apprentissage ne sont pas, par nature, adaptés aux mêmes types de situations.

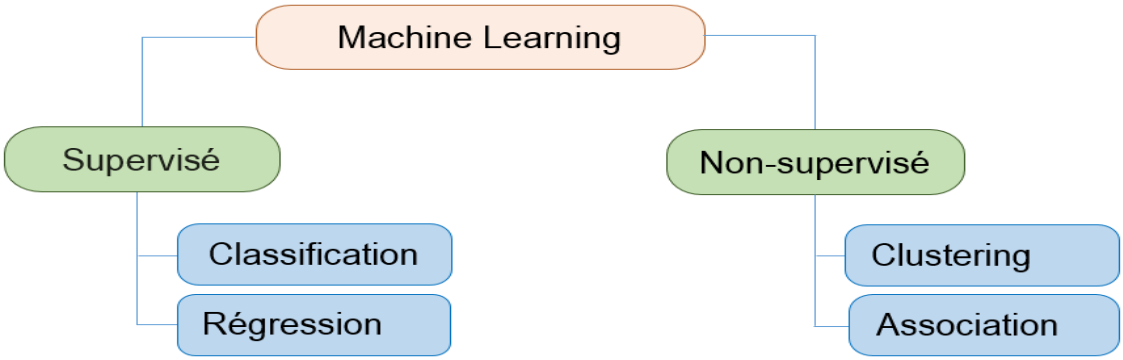

Figure 3.3 Classification des méthodes d'apprentissage automatique.

3.2.1 Apprentissage automatique supervisée

L'apprentissage supervisé, dans le contexte d'IA, est un système qui fournit les entrées et les sorties attendues. Les données en entrée et en sortie sont étiquetées pour la classification, afin d'établir une base d'apprentissage pour un traitement ultérieur des données [135]. Les systèmes d'apprentissage automatique supervisés alimentent des algorithmes d'apprentissage avec des quantités connues qui soutiendront les futures décisions. C'est ce qu'on appelle l'apprentissage supervisé, car le processus d'un algorithme est tiré de l'ensemble de données, et peut être considéré comme un contrôleur supervisant le processus d'apprentissage. Nous connaissons les bonnes réponses, l'algorithme fait des prédictions itératives sur les données d'apprentissage et elles sont ajustées parle superviseur. L'apprentissage s'arrête lorsque l'algorithme atteint un niveau de performance acceptable. Prenons, par exemple, des photos de chats puis des photos de chiens. Pendant l'apprentissage, nous superviserons le modèle d'apprentissage automatique en lui montrant des photos de chats lui disant qu'il s'agit bien des chats, puis des photos de chiens et lui disant qu'ils sont des chiens. Le but principal est de former un modèle qui facilite la reconnaissance et l'interprétation de l'image originale et à pouvoir distinguer une image d'un chat d'une image d'un chien [138]. Les étapes d'un système d'apprentissage supervisé sont décrites dans la figure **3.4**. Les problèmes d'apprentissage supervisés peuvent être classés en problèmes de classification et de régression [139].

- **Classification :** le problème de classification est un problème où nous utilisons des données pour prédire dans quelle catégorie l'objet en question appartient. Un exemple de problème de classification pourrait être l'analyse d'une image pour déterminer si elle contient une voiture ou une personne, ou l'analyse de données médicales pour déterminer si une certaine personne fait partie d'un groupe à haut risque pour une certaine maladie ou non. En d'autres termes, nous essayons d'utiliser des données pour faire une prédiction sur un ensemble discret de valeurs ou de catégories.

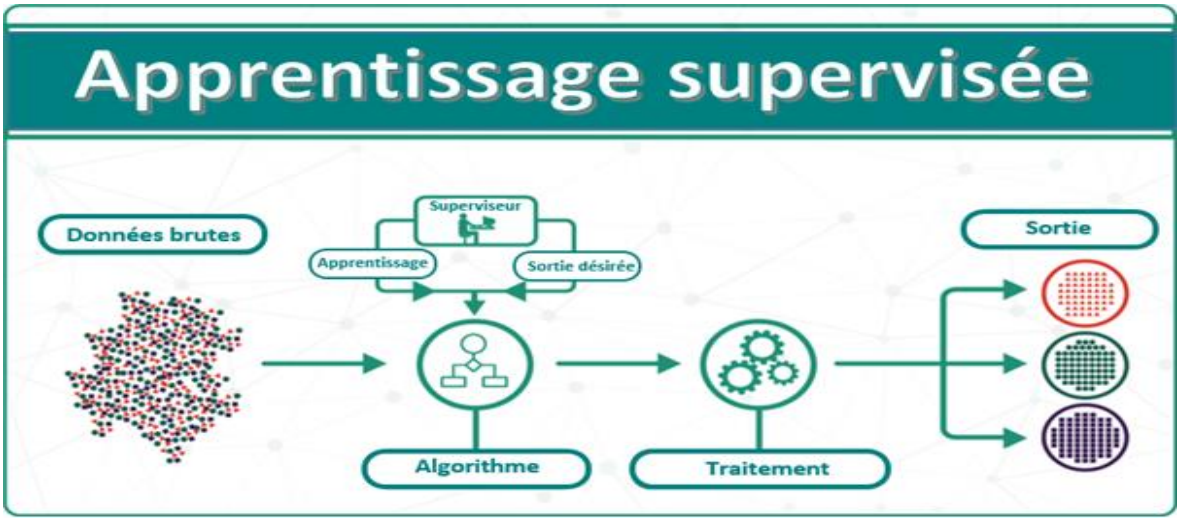

Figure 3.4 Schéma général d'un système d'apprentissage supervisé [191].

- **Régression : l**es problèmes de régression, par contre, sont des problèmes où nous essayons de faire une prédiction sur une échelle continue. Des exemples pourraient être de prédire le cours des actions d'une entreprise ou de prédire la température de demain sur la base de données historique.

La principale différence entre la régression et la classification est que la variable de sortie dans la régression est numérique (ou continue) tandis que celle pour la classification est catégorielle (ou discrète).

3.2.2 Apprentissage automatique non supervisée

Dans le cas d'un apprentissage non supervisé, l'apprentissage automatique se déroule de manière autonome. Des données sont alors communiquées à la machine sans lui fournir les exemples de résultats attendus en sortie. Il appartient à l'algorithme de déterminer les critères les plus pertinents pour classer les données. Le but de l'apprentissage non supervisé est de modéliser la structure de base ou la distribution sous-jacente dans les données afin d'en savoir plus sur les données. Celles-ci sont appelées apprentissage non supervisé car, contrairement à l'apprentissage supervisé ci-dessus, il n'y a ni bonne réponse ni superviseur. Les algorithmes sont laissés à leurs propres mécanismes pour découvrir et présenter la structure intéressante des données. La machine procède aux rapprochements en fonction de ces caractéristiques qu'elle est capable d'identifier sans avoir besoin d'une intervention extérieure. Cela ouvre également la possibilité de développer un système de recommandation (le système peut par exemple recommander un livre ou un film à un utilisateur selon les souhaits des utilisateurs partageant des caractéristiques communes) ainsi que la possibilité de développer un système de détection d'anomalies. Les étapes d'un système d'apprentissage non supervisée sont décrites dans la figure **3.5**.

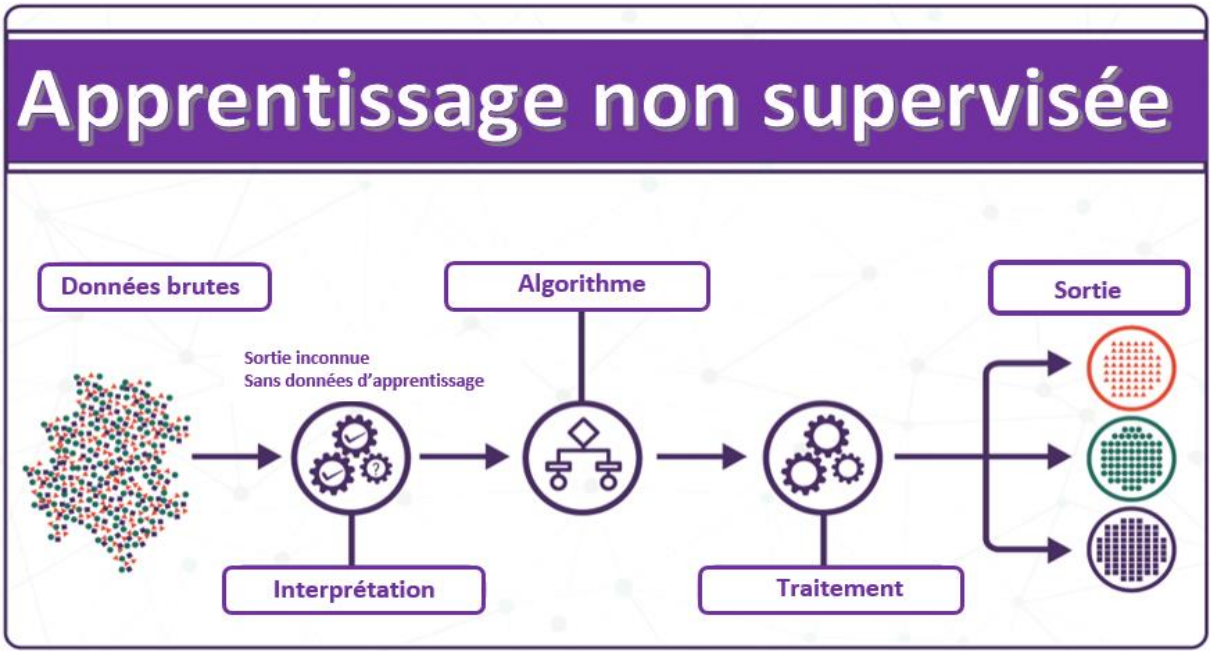

Figure 3.5 Schéma général d'un système d'apprentissage non supervisé [191].

L'apprentissage non supervisé désigne le regroupement ou Clustering. Il est définit par un processus de regroupement d'un ensemble d'individus hétérogènes sous forme de sous-groupes homogènes ou liés par des caractéristiques communes. Les problèmes d'apprentissage non supervisés peuvent être classés en problèmes de Clustering (regroupement) et d'association [141].

- Clustering : permet de découvrir les regroupements inhérents aux données, tels que le regroupement des clients en fonction du comportement d'achat.
- Association : permet de découvrir des règles décrivant une grande partie de données, telles que les acheteurs de X ont également tendance à acheter Y.

Prenons notre exemple d'images de chats et de chiens. Contrairement à l'apprentissage supervisé, le modèle ne s'appuiera pas sur les humains pour classer ces images mais sur ce qu'on appelle la « variance ». Le concept de variance consiste à mesurer le degré de différence entre les données.

3.3 Principaux techniques d'apprentissage automatique

Dans cette section on va présenter les approches d'apprentissage de type supervisé les plus populaires et les plus utilisées qui s'adaptent bien à notre sujet traitées dans cette thèse, la classification et la reconnaissance des visages et des expressions faciales

3.3.1 K-plus proches voisins

La méthode des K-plus proches voisins [8] est une méthode d'apprentissage supervisé, elle est utilisée pour la régression et la classification. Pour faire la prédiction, contrairement à la régression linéaire, les K instances de l'ensemble de données les plus proches de sens des caractéristiques des observations sont obtenues grâce au module de prédiction. Le principe intuitif du K-NN est illustré dans la figure 3.6
K-NN se basera sur leurs variables de sortie pour calculer la valeur de la variable d'observation que nous souhaitons prédire [8]. Autrement dit :

- Si $K-NN$ est utilisé pour la classification, c'est le *mode* des K variables plus proches observations qui serviront pour la prédiction
- Si $K-NN$ est utilisé pour la régression, c'est la *moyenne* des K variables plus proches observations qui serviront pour la prédiction

3.3.1.1 Algorithme de K-NN

Le principe de fonctionnement de la méthode KNN est décrit par l'algorithme 1.

Algorithme K-NN

Entrées :

- X : jeu de données
- d : une fonction de distance.
- K : un nombre entier

Sorties :

- Y : prédiction de la valeur de sortie

Début

Pour un nouveau test X pour lequel on va prédire sa variable de sortie **Faire**

1. Calculer toutes les distances de cette observation X avec les autres éléments du jeu de données
2. Conserver les K éléments de l'ensemble de données les proches de X par l'application de fonction de distance d.
3. Garder juste les valeurs Y des K observations sélectionnées.

Si nous effectuons une régression **alors**

calculer la moyenne des K observations.

Fsi

Si nous effectuons une classification **alors**

calculez le mode des K observations sélectionnées.

Fsi

4. Retourner la valeur calculée à l'étape 3 en tant que valeur prédite par KNN pour le test X.

Fin Pour

Fin

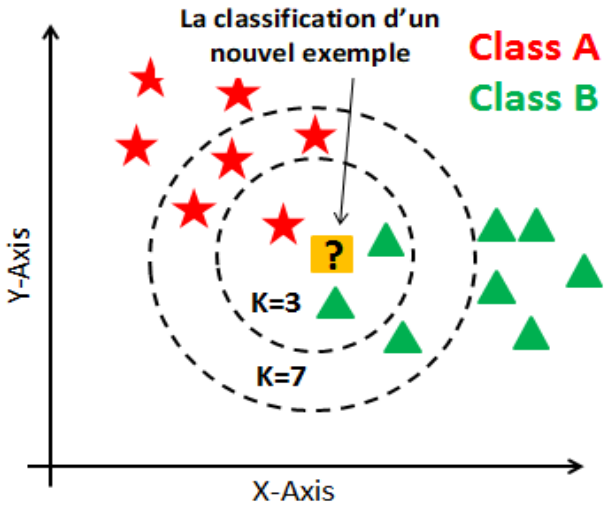

Figure 3.6 Schéma représentant le principe intuitif d'un K-NN [192].

3.3.1.2 Le choix de la fonction de similarité

Dans $K-NN$, la distinction entre deux observations, similaires ou non est faite grâce au calcul de la distance. Plus les deux points sont proches l'un de l'autre, plus ils sont semblables et vice versa. Il existe plusieurs fonctions de calcul de distance, notamment la distance euclidienne, la distance de Manhattan, la distance de Minkowski, celle de Jaccard, la distance de Hamming, etc. En choisissant la fonction de distance en fonction des types de données qu'on manipule. Pour les données quantitatives (*exemple : poids, salaires, taille, montant panier électronique, etc.*) et du même type, la distance euclidienne est un bonne candidat. Quant à la distance de Manhattan, c'est une bonne mesure à utiliser lorsque les données ne sont pas du même type (*exemple : âge, sexe, longueur, poids, etc.*).

Les distances les plus utilisées sont :

- **La distance Euclidienne** : calcule la racine carrée de la somme du carré différence entre les coordonnées de deux points.

$$D_e(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (3.1)$$

- **La distance de Manhattan :** calcule la somme des valeurs absolues des différences entre les coordonnées de deux points.

$$D_m(x,y) = \sum_{i=1}^{n}|x_i - y_i| \quad (3.2)$$

- **La distance de Hamming** : est une distance entre deux points donnés simulée par une différence maximale entre leurs coordonnées sur une cote.

$$D_h(x,y) = max|x_i - y_i| \quad (3.3)$$

3.3.1.3 Le choix de la valeur K

Dans les algorithmes $K-NN$ on choisit le K selon la nature, le type et le jeu de données utilisées. On peut dired'une façon générale, si le K est petit c'est-à-dire moins qu'on utilise de voisins, plus on aura un sous-apprentissage. Par contre si le nombre K choisit est grand c'est-à-dire plus on utilise de voisins, on risque d'avoir un sur-apprentissage (over fitting) et par conséquent on aura un modèle qui se généralise d'une façon incorrecte sur des éléments non encore vu. Les algorithmes $K-NN$ prendraient en générale la valeur K égale à 3. $K-NN$ est un algorithme assez simple qui se base sur la distance de voisinage et le nombre de voisins K. Nous devons faire plusieurs combinaisons et adaptation de l'algorithme $K-NN$ pour obtenir un résultat satisfaisant. De plus, nous devons donc faire attention à la taille du jeu d'entraînement. La figure **3.7** montre l'influence de K sur la classification.

3.3.1.4 Avantages et inconvénients

- **Avantages**
  - Simple à mettre en œuvre et facile.
  - Efficace pour des classes réparties de manière irrégulière.
  - Reste toujours efficace même si les données sont étaient incomplètes.
  - La classification est rapide par rapport aux autres approches.
- **Inconvénients**
  - Nécessite une capacité de stockage importante.
  - De nombreuses données de références sont nécessaires pour classifier les nouvelles entrées.

3.3.2 Classification bayésienne

La classification bayésienne naïve est un type de classification bayésienne probabiliste simple basée sur le théorème de Bayes avec une forte autonomie (dite naïve) d'hypothèses, issue des travaux du Thomas Bayes (1702-1761) [143]. Il implémente un classificateur bayésien naïf appartenant à la famille des classificateurs linéaires dont le rôle est de classer en groupes (classes) des échantillons aux propriétés similaires, mesurés sur des observations[144]. Le classificateur naïf de Bayes est basé sur le théorème de Bayes, Il s'agit d'un résultat de base en théorie des probabilités, Cette dernière est un classique de la théorie des probabilités. Ce théorème est basé sur des probabilités conditionnelles. Une probabilité conditionnelle peut être définie intuitivement par : "Quelle est la probabilité qu'un événement se produise sachant qu'un autre événement s'est déjà produit".

3.3.2.1 Principe mathématique

Les classificateurs de Naïve Bayes sont une famille d'algorithmes reposant sur le principe commun selon lequel la valeur d'une fonctionnalité spécifique est indépendante de la valeur de toute autre fonctionnalité. Ils nous permettent de prédire la probabilité qu'un événement se produise en fonction de conditions que nous connaissons pour les événements en question. Le principe mathématique peut être formulé [143] :

$$P(A \setminus B) = \frac{P(B \setminus A)\, P(A)}{P(B)} \tag{3.4}$$

Où

A et B : événements et $P(B) > 0$.

$P(A|B)$: Probabilité que l'événement A se produise sachant quel'événement$B$ s'est déjà produit.

$P(B|A)$: probabilité que l'événement B se produise sachant que l'événement A s'est déjà produit.

$P(A)$ et $P(B)$: sont les probabilités des événements A et B indépendamment les uns des autres.

3.3.2.2 Avantages et inconvénients

- **Avantages**
  - Robuste aux caractéristiques non pertinentes.
  - Relativement simple à appréhender et à implémenter.
  - Facile et pratique même avec un petit jeu de données
  - La classification est rapide par rapport aux autres MLs.
- **Inconvenients**
  - Théoriquement chaque fonctionnalité est indépendante, ce qui n'est pas toujours le cas dans les exemples réels.
  - En général, ce type de classificateur permet de faire le même travail de classification que les autres algorithmes qui existent déjà, mais ces performances sont limitées quand il s'agit d'une grande quantité de données à traiter. En effet, si le nombre de données augmente, alors le nombre des dépendances entre l'ensemble des mots augmentent, et donc, la vérification de l'hypothèse de Naïve Bayes diminue.

3.3.3 Machine à vecteurs de support

Machine à vecteurs de support (en anglais Support Vector Machine, SVM) est l'un des algorithmes d'apprentissage supervisé les plus populaires. Il est utilisé pour les problèmes de classification et de régression. Cependant, il est utilisé principalement pour les problèmes de classification automatique. Le classificateur SVM est formellement défini par un hyperplan de séparation. Si les données d'apprentissage sont étiquetées (apprentissage supervisé), l'algorithme génère un hyperplan optimal qui classe les nouveaux exemples [7, 145].Dans un espace bidimensionnel, cet hyperplan est une ligne qui divise le plan en deux parties où, dans chaque classe, se trouvait de chaque côté. Le but de l'algorithme SVM est de créer la meilleure ligne ou limite de décision qui peut séparer l'espace à n dimensions en classes afin que nous puissions facilement mettre le nouveau point de données dans la bonne classe à l'avenir. Cette meilleure frontière de décision est appelée ***hyperplan***. SVM choisit les points/vecteurs extrêmes qui aident à créer l'hyperplan. Ces cas extrêmes sont appelés vecteurs de support, et donc l'algorithme est appelé machine de vecteur de support [7, 145]. La figure **3.7** montre le schéma général d'un SVM.

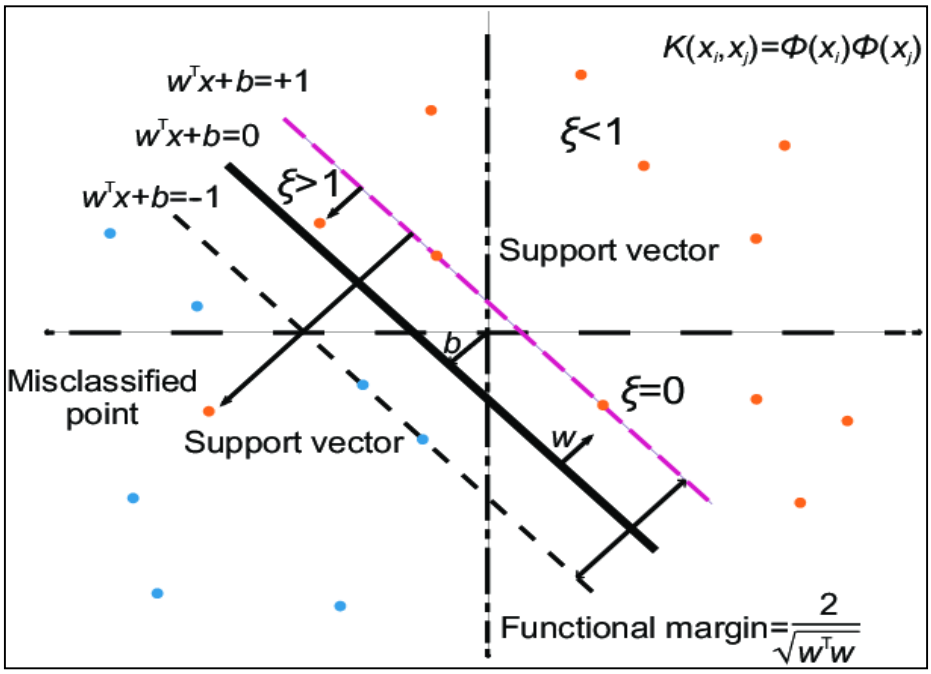

Figure 3.7 Schéma général d'un classificateur SVM [177].

Il existe deux classes différentes classées à l'aide d'une frontière de décision ou d'un hyperplan. SVM peut être classifié en deux types : linéaire et non linéaire (figure **3.9**).

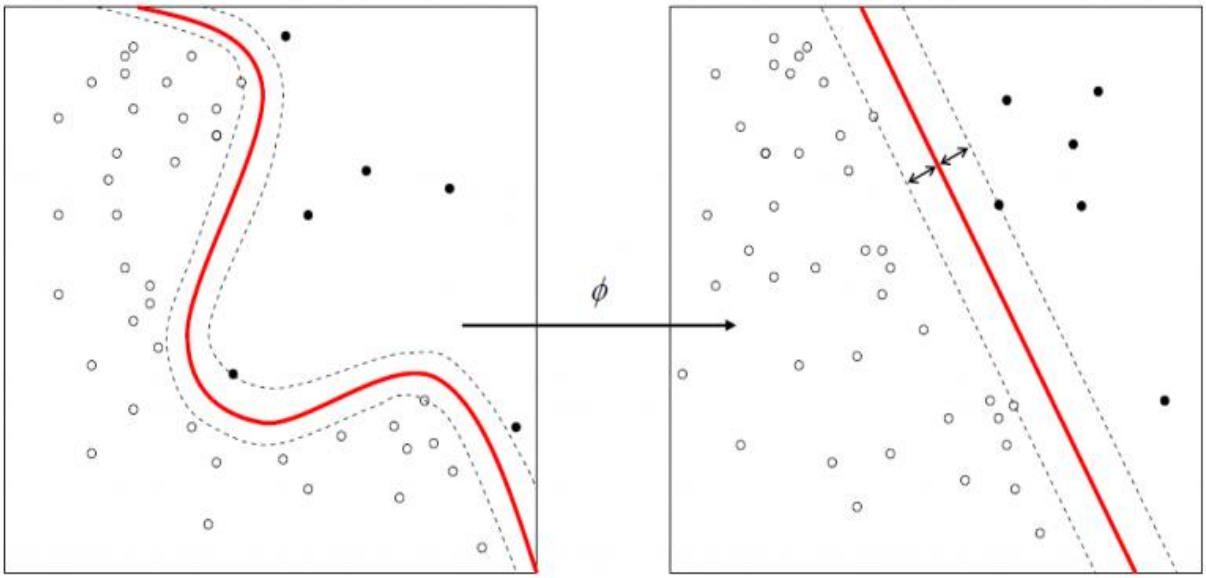

Figure 3.8 SVM linéaire et non linéaire [193].

- SVM linéaire est utilisé pour les données linéairement séparables, ce qui signifie que si un ensemble de données peut être classé en deux classes en utilisant une seule ligne droite, alors ces données sont appelées données linéairement séparables, et le classificateur est utilisé comme classificateur SVM linéaire [7].
- SVM non linéaire est utilisé pour les données non linéairement séparables, ce qui signifie que si un ensemble de données ne peut pas être classé en utilisant une ligne droite, ces données

sont qualifiées de données non linéaires et le classificateur utilisé est appelé classificateur SVM non linéaire [7].

3.3.3.1 Principe mathématique

Le principe du SVM est basé sur la minimisation du risque structurel d'écrit par la théorie d'apprentissage statistique de Vapnik [7]. Les SVMs établissent un mappage des données d'entrée vers un espace de plus grande dimensionnalité, dans lequel elles deviennent linéairement séparables. Soit l'ensemble de données $x_i, i = 1, \ldots, l$ et les étiquettes correspondantes $y_i \in [-1, +1]$, la tâche principale pour entrainer un SVM consiste à résoudre le problème d'optimisation quadratique suivant [7, 145].

$$maximise \ \sum_{i=1}^{n} \alpha_i \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \ K(x_i, \ y_j) \tag{3.5}$$

$$\alpha_i y_i = 0 \tag{3.6}$$

$$0 \leq \alpha_i \leq C, \quad i = 1 \ldots n \tag{3.7}$$

Où :

x_i : Les exemples d'apprentissage.

$y_i = \pm 1$: Les classes perspectives de xi.

α_i : Les multiplicateurs de Lagrange à déterminer.

K : La fonction noyau (Kernal) utilisée.

C : La constante de régularisation représente la limite maximale de toutes les variables.

La solution mathématique sert à chercher une fonction d'optimisation qui permet de trouver un plan optimal séparant les instances positives des instances négatives. L'ensemble de données est dit optimalement séparable, si elles sont séparées sans erreurs et que la distance séparant le vecteur le plus proche et l'hyperplan est maximal. La fonction noyau K surmonte le problème de non-linéarité, ce qui permet de traiter des relations non linéaires avec des machines linéaires. En pratique, plusieurs fonctions standard sont considérées parmi lesquelles on peut citer ici :

- La fonction de noyau linéaire définie par :

$$K(x_i, \ y_j) = x_i^T x_j \tag{3.8}$$

- La fonction du noyau polynomial définie par :

$$K(x_i, \ y_j) = (x_i^T x_j + C)^d \tag{3.9}$$

- La fonction noyau RBF (Radial Basis functions) est définie comme :

$$K(x_i, \ y_j) = exp(-\frac{|x_i - x_j|^2}{2\sigma^2}) \tag{3.10}$$

Pour un problème de discriminations multiples, dans de nombreux cas comme la classification des visages et des expressions faciales, « un contre tous » et « un contre un » sont les deux approches les plus utilisées pour la construction de SVM multi-classes à partir d'un ensemble de SVM binaires. La première se base sur le principe du vote majoritaire (*winner-take-all*), dans lequel la classe est assignée par le SVM ayant le plus grand score en sortie.

- Approche « **un contre un** » consiste à utiliser un SVM par couple de catégories. Chaque fois que le SVM assigne une instance à l'une des deux classes, la classe de vote majoritaire sera bénéficié de un en plus et ainsi de suite. La classe avec un maximum de votes détermine l'étiquette de l'instance [145].
- Approche « **un contre tous** » peut souffrir de la répartition déséquilibrée des exemples positifs par rapport aux exemples négatifs. C'est la raison pour laquelle, dans notre travail dans le cadre de cette thèse, nous nous servirons de la stratégie « un contre un » pour résoudre le problème de classification multiple de visages expression faciale [145].

3.3.3.2 Algorithme SVM

Le principe de fonctionnement de SVM est décrit par l'algorithme suivant :

Algorithme SVM

Entrées :

- x_i : l'ensemble des données d'apprentissage.

Sorties :

- Y : résultats en sortie ± 1

Début

1. Préparation la matrice des données.
2. Sélection de la fonction noyau K à utiliser.
3. Sélection des paramètres de la fonction noyau sélectionnée et les valeurs de C.
4. Lancer le processus d'apprentissage pour obtenir les valeurs de α_i .
5. Procéder à l'étape de test en utilisant les α_i et déterminer les valeurs de y_i .

Fin

3.3.3.3 Avantages et inconvénients

❑ **Avantages**

- Grande degré de précision de prédiction
- Facile et pratique avec un petit jeu de données.
- Manipule de grandes quantités de données.
- Classification d'images fiable et efficace.
- Elle est justifiée et fondée théoriquement et mathématiquement.

❑ **Inconvénients :**

- Temps d'entraînement très long avec des jeux de données volumineux.

- Moins efficace sur les jeux de données contenant des bruits.

3.3.4 Arbre de décision

L'arbre de décision (Decision Tree, DT), est une classe des techniques qui se basent sur l'apprentissage automatique supervisé. L'arbre de décision est très similaire au raisonnement humain. Cela nous éclaire également avec beaucoup d'informations sur les données et, surtout, l'arbre de décision est facile à interpréter. Le DT est une technique de classification simple et largement utilisée. Le DT applique une idée rigoureuse pour résoudre le problème de classification en posant une série de questions soigneusement conçues sur les attributs du test. Chaque fois qu'il reçoit une réponse, une autre question est posée jusqu'à ce qu'une conclusion soit atteinte concernant l'étiquette de la classe du dossier [146]. Un arbre de décision permet de créer un modèle d'apprentissage qui peut être utilisé pour prédire la classe ou la valeur de la variable cible en apprenant des règles de décision simples déduites de données antérieures (*données d'entraînement*). Dans les arbres de décisions, pour prédire une classe (figure **3.9**), la méthode commence de parcourir l'arbre de la racine. Ensuite elle compare les valeurs de l'attribut racine avec l'attribut de l'enregistrement. Sur la base de la comparaison, nous suivons la branche correspondant à cette valeur et passons au nœud suivant.

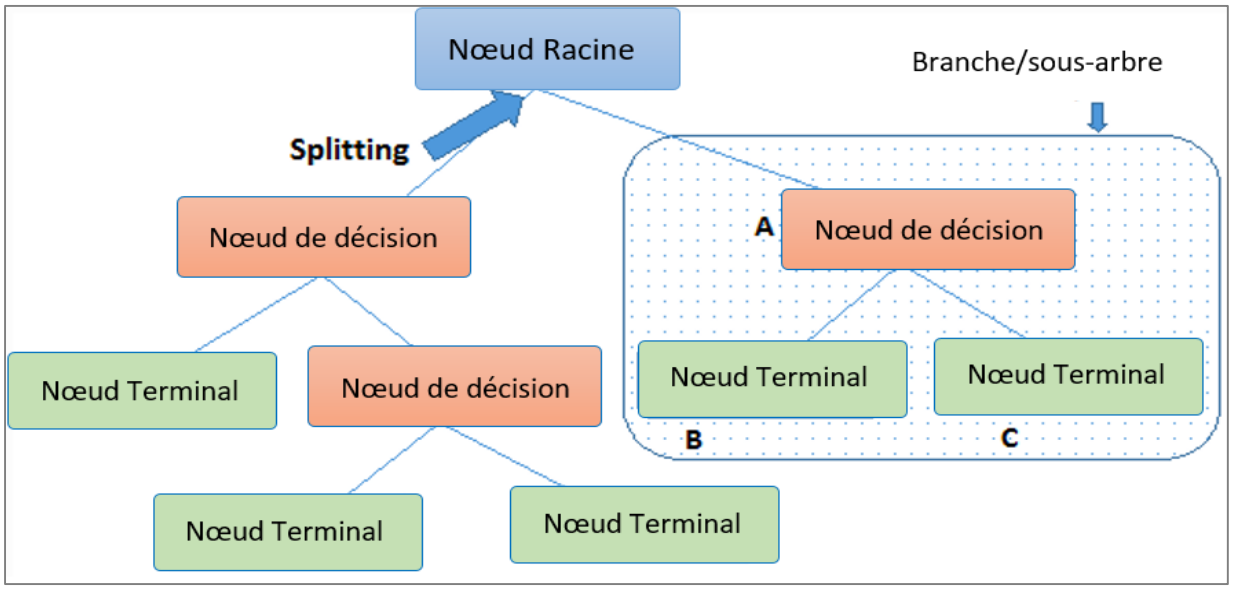

Figure 3.9 La structure générale d'un arbre de décision DT [194].

Pour aboutir à une bonne classification des échantillons, les arbres de décision classifient les échantillons en les triant dans l'arbre de la racine au nœud feuille/nœud feuille, nœud feuille/nœud terminal. Chaque nœud de l'arbre agit comme un cas de test pour un attribut, et chaque nœud correspond aux réponses possibles au cas de test. Ce processus est récursif par nature, répété pour chaque sous-arbre enraciné dans le nouveau nœud.

3.3.4.1 Algorithmes arbre de décision

Il existe plusieurs algorithmes utilisés dans les arbres de décision, l'algorithmeID3 [178, 180], C4.5 le successeur d'ID3 (Itérative Dichotomiser 3) [147], CART (Classification and Regression Tree) [147], CHAID (Détection automatique des interactions du chi-square Effectue des divisions à plusieurs niveaux lors du calcul des arbres de classification) [147] et MARS (Multivariate Adaptive Regression Splines) [147], nous détaillons dans cette section l'algorithmeID3.

L'algorithme ID3 [148] est le plus populaire, il construit des arbres de décision en utilisant une approche de recherche gloutonne descendante à travers d'espace de branches possibles sans retour en arrière. Cet algorithme permet de construire un arbre de décision de maniée récursive. Il calcule, parmi les attributs restants, celles qui génèrent plus d'informations et qui permettront la classification des exemples à n'importe quel niveau de l'arbre de décision. Les étapes de l'algorithme ID3 sont décrites comme suit :

1- Initialement on a un ensemble d'origine S comme nœud racine.
2- À chaque itération de l'algorithme, on parcourt l'attribut inutilisé de l'ensemble S et oncalcule l'entropie (H) et le gain d'information (IG) de cet attribut.
3- On sélectionne l'attribut qui a le plus petit gain d'entropie ou le plus grand gain d'informations.
4- L'ensemble S est divisé par l'attribut sélectionné pour produire un sous-ensemble des données.
5- L'algorithme continue de se reproduire sur chaque sous-ensemble, en en tenant compte des attributs qui n'étaient pas prises en compte auparavant.

3.3.4.2 Avantages et inconvénients

- **Avantages**
  - Aucun modèle et aucun présupposé à satisfaire.
  - Règles simples et facilement interprétables.
  - Traitement des valeurs manquantes.
  - Principe simple et peu sensible aux variables non discriminantes.
  - Facile de les associer à d'autres d'outils de décisions.
- **Inconvénients**
  - Soufre du problème sur apprentissage.
  - N'est pas robuste au nombre de classes.
  - Nécessite d'un grand nombre d'individus.

3.3.5 Forêt aléatoire

L'algorithme forêt aléatoire(en anglais Random Forest RF) [149] est un algorithme de classification supervisé. Il a en quelque sorte crée une forêt et l'a rendue aléatoire. Le RF crée un ensemble d'arbres de décision à partir d'un sous-ensemble sélectionné au hasard d'un ensemble d'apprentissage. Il agrège ensuite les votes des différents arbres de décision pour décider de la classe finale de l'objet de test. Il existe une relation directe entre le nombre d'arbres dans la forêt et les résultats qu'elle peut obtenir. Plus le nombre d'arbres est grand, plus le résultat est précis.

Le principe de fonctionnement du FR est très similaire à celle des arbres de décision, mais les FRs sont plus compliqués, car les arbres de décision unique sont très faciles à visualiser et à comprendre car ils suivent une méthode de prise de décision qui est très similaire à la façon dont nous, humains, prenons des décisions : avec une chaîne de règles simples. Cependant, ils ne sont pas très robustes, puisqu'ils ne se généralisent pas bien aux échantillons invisibles. C'est là que les forêts aléatoires entrent en jeu. La différence entre l'algorithme de forêt aléatoire et l'algorithme d'arbre de décision, réside dans le fait que dans la forêt aléatoire, les processus de recherche du nœud racine et de division des nœuds d'entités s'exécuteront de manière aléatoire.

3.3.5.1 Algorithme forêt aléatoire

L'algorithme Random Forest comporte deux étapes. La première consiste à créer une forêt aléatoire et la seconde consiste à prédire la classe à partir de forêt aléatoire créé lors de la première étape. L'ensemble du processus est illustré ci-dessous (figure **3.10**)

Algorithme Forêt aléatoire

Entrées :

- x_i : l'ensemble des données d'apprentissage.

Sorties :

- Y : résultats en sortie± 1

Début

1 - Sélectionner au hasard les K caractéristiques parmi les caractéristiques totales $\boldsymbol{m}$ où

$$k \ll m.$$

2 - Parmi les K caractéristiques, calculer le nœud $d$en utilisant le meilleur point de partage.

3 - Fractionner le nœud en sous-nœuds en utilisant le meilleur mécanisme de fractionnement.

4 - Répéter les étapes 1 à 2 jusqu'à ce que le nombre L de nœuds soit atteint.

5- Construire la forêt en répétant les étapes de 1 à 4 n fois afin de créer $\boldsymbol{n}$ arbres.

Fin

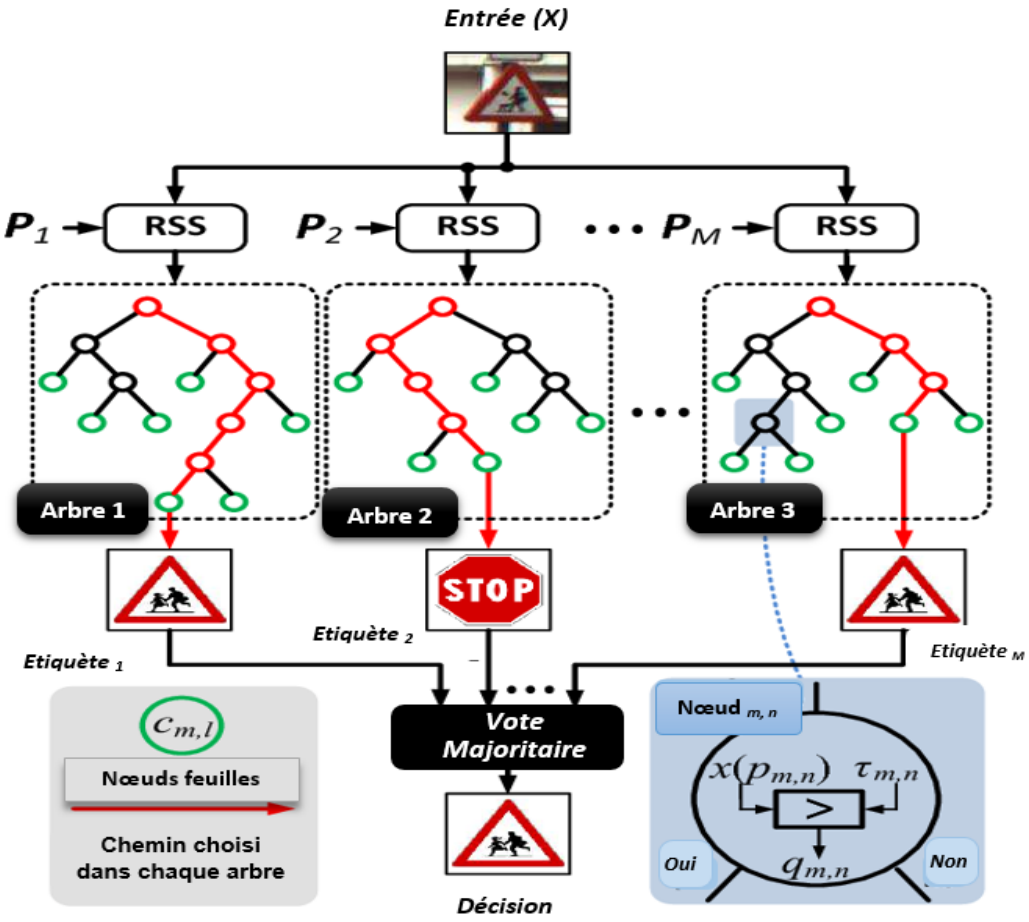

Figure 3.10 Exemple d'une forêt aléatoire [178].

3.3.5.2 Avantages et inconvénients

- **Avantages**
  - Random Forest évitera le problème de sur-ajustement.
  - Identification des caractéristiques dominantes de l'ensemble de données d'apprentissage
  - Utilisé pour la classification et la régression.
- **Inconvénients**
  - Nécessite un grand nombre d'individus.
  - FR est compliqué par apport les DT et couteuse en temps.

3.3.6 Classificateur d'analyse discriminante quadratique

Un classificateur d'Analyse discriminante quadratique (Quadratic Discriminant Analysis, QDA) [150] est une version plus générale du classificateur linéaire. Il est basé sur le principe d'un classificateur statistique qui utilise une surface de décision quadratique pour séparer les mesures de deux ou plusieurs classes d'objets ou d'événements.

La classification statistique considère un ensemble de vecteurs d'observations X d'un objet ou d'un événement, chacun ayant un type connu Y. Cet ensemble est appelé ensemble d'entraînement. Le problème est alors de déterminer pour un nouveau vecteur d'observation donné, quelle devrait être la meilleure classe. Pour un classificateur quadratique, la solution correcte est supposée quadratique en utilisant une fonction discriminante représentée par une surface quadratique de n-dimensions :

$$x^T Ax + b^T + c \tag{3.11}$$

Où la matrice $A = (a_{ij})$, le vecteur $b = (b_1, b_2, \ldots, b_n)^T$ et la constante c. Ces variables caractérisent le classificateur comme suit :

- Si la matrice A est définie semi-positive alors la fonction discriminante est unhyper-ellipsoïde ayant les axes principaux dans les directions des vecteurs propres de A.
- Si A égale à I (matrice identité) alors la fonction discriminante est une hyper-sphère (dans un espace euclidien)
- Si A est définie négative, alors la fonction discriminante est une hyperboloïde.

Les propriétés de la matrice A déterminent la forme et les caractéristiques de la fonction de décision. En particulier, chaque observation se compose de deux mesures, cela signifie que les surfaces séparant les classes seront des sections coniques peut être soit une ligne, un cercle, une ellipse, une parabole ou une hyperbole. L'utilisation du modèle quadratique est justifiée par sa capacité de représenter des surfaces de séparation plus complexes.

3.3.7 Réseaux de neurones et apprentissage profond

Les réseaux de neurones sont des modèles mathématiques basés sur les «neurones formels » inspirés du cerveau humain vu de leurs intérêts lors de l'identification des objets (ex*emple pour notre cas visages et des expressions faciales*). Les réseaux de neurones sont des séries d'algorithmes passent par le processus d'apprentissage et la reconnaissance des visages sur plusieurs couches neuronales [151]. Ils peuvent ainsi apprendre et s'adapter à l'évolution des entrées. Cette fonctionnalité permet au réseau une génération des meilleurs résultats sans tenir en compte les critères de sortie. Le concept de réseaux de neurones, qui a ses racines derrière l'intelligence artificielle, gagné rapidement en popularité dans le développement de multiple systèmes commerciaux de la vie sociale, à savoir la transcription de la parole en texte, la reconnaissance de l'écriture manuscrite pour le traitement des chèques, la prévision météorologique et la reconnaissance faciale, l'analyse des données et l'exploration pétrolière. Il existe plusieurs types d'architectures pour les réseaux de neurones :

- Les perceptrons multicouches (MLP) [152], les plus anciens et les plus simples.
- Les Réseaux de Neurones Convolutionnel (CNN) [153], adaptés au traitement d'images

- Les réseaux de neurones récurrents, adaptés au traitement de texte ou séries chronologiques [187]

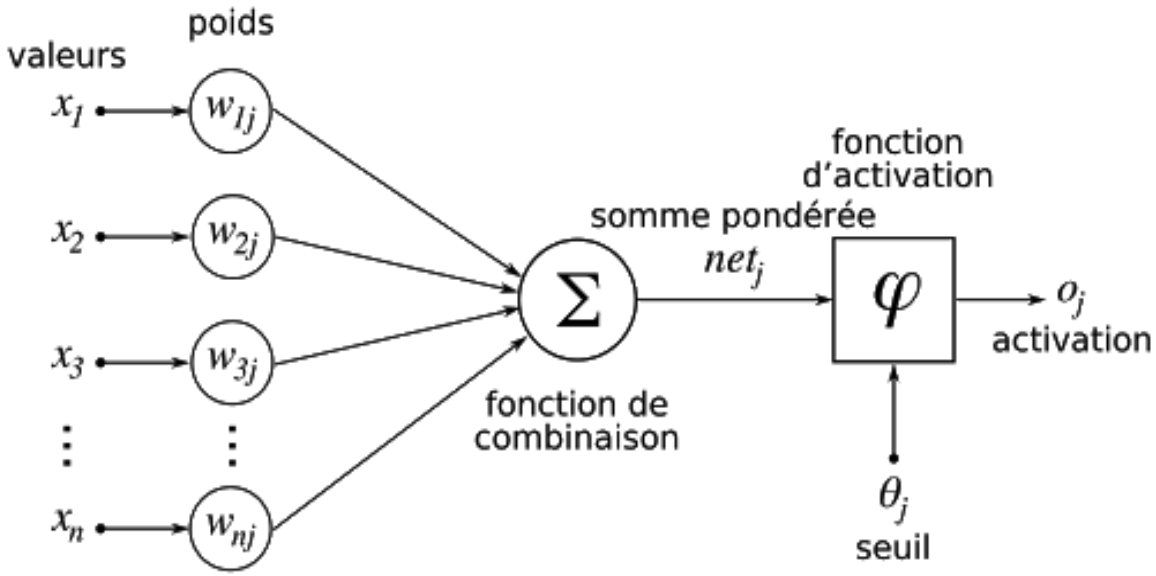

Figure 3.11 La structure d'un réseau de neurone [195].

3.3.8.1 Réseaux de neurones

Un réseau de neurones artificiels [153] est une application, non linéaire par rapport à ses paramètres θ qui associe à une entrée x une sortie $y = f(x, \theta)$. Par mesure de simplicité, nous supposons que y est unidimensionnel, mais il pourrait également être multidimensionnel. Cette application f a une forme particulière que nous préciserons. Les réseaux de neurones peuvent être utilisés pour la régression ou la classification. Comme d'habitude dans l'apprentissage statistique, les paramètres θ sont estimés à partir d'un échantillon d'apprentissage. La fonction de minimisation n'est pas convexe, conduisant à des minimiseurs locaux. Le succès de la méthode provient d'un théorème d'approximation universel dû à Cybenko (1989) et Hornik (1991). De plus, Cun (1986) a proposé un moyen efficace de calculer le gradient d'un réseau neuronal, appelé rétro-propagation du gradient, qui permet d'obtenir facilement un minimiseur local du critère quadratique.

Un neurone artificiel est une fonction f_j de l'entrée $x = (x_1, \ldots, x_d)$ pondérée par un vecteur de poids de connexion $w_j = (w_{1,j}, \ldots, w_{n,j})$, complétée par un biais de neurone b_j, et associée à une fonction d'activation φ, à savoir

$$y_j = f_j(x) = \phi(< hw_j, x_i > + b_j) \tag{3.12}$$

Plusieurs fonctions d'activation peuvent être envisagées :

- La fonction identitaire :

$$\phi(x) = x \tag{3.13}$$

- La fonction sigmoïde (ou logistique) :

$$\phi(x) = \frac{1}{1+\exp(-x)} \tag{3.14}$$

- La fonction tangente hyperbolique :

$$\phi(x) = \frac{\exp(x) - \exp(-x)}{\exp(x)+\exp(-x)} = \frac{\exp(2x) - 1}{\exp(2x) + 1} \tag{3.15}$$

- La fonction seuil :

$$\phi_\beta(x) = 1_x \geq \beta \tag{3.16}$$

- La fonction d'activation de l'unité linéaire rectifiée (ReLU) :

$$\phi(x) = \max(0, \quad x) \tag{3.17}$$

3.3.8.2 Les perceptrons multicouches

Un perceptron multicouche (Multi Layer Perceptron MLP) [152] est une structure composée de plusieurs couches de neurones cachées où la sortie d'un neurone d'une couche devient l'entrée d'un neurone de la couche suivante. De plus, la sortie d'un neurone peut également être l'entrée d'un neurone de la même couche ou d'un neurone de couches précédentes (*les réseaux de neurones récurrents*). Dansla dernière couche, qui s'appelle la couche de sortie, nous pouvons appliquer des fonctions d'activation différentes aux couches cachées selon le type de problèmes que nous confrontant : *régression* ou *classification*. La figure **3.12** montre un réseau neuronal avec trois variables d'entrée, une variable de sortie et deux couches cachées. Les perceptrons multicouches ont une architecture de base car chaque neurone d'une couche est liée à tous les neurones de la couche suivante mais n'a aucun lien avec les neurones de la même couche.

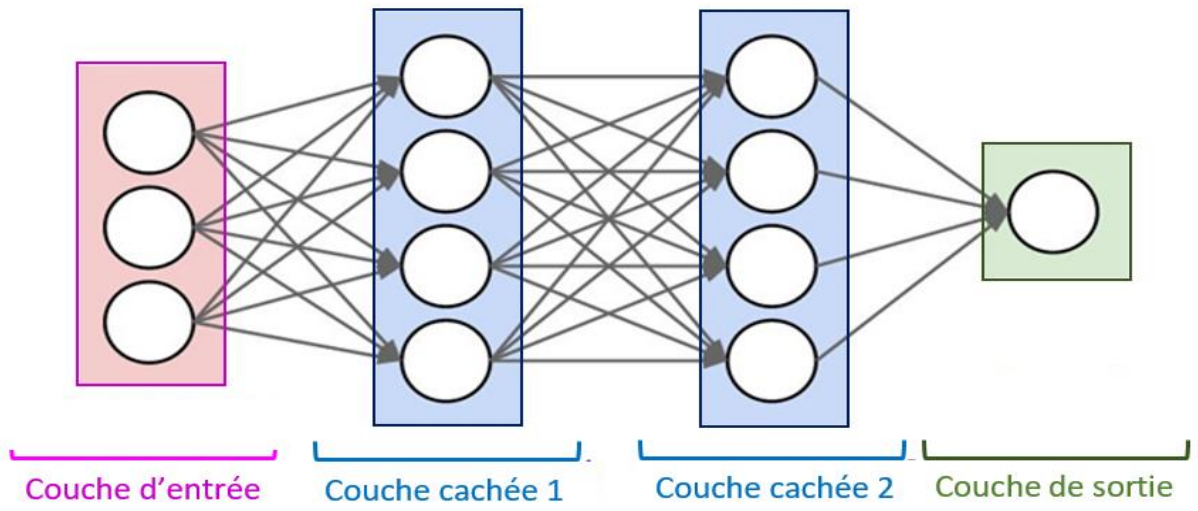

Figure 3.12 La structure d'un réseau de neurone MLP [196].

Les paramètres de l'architecture sont le nombre de couches cachées et de neurones dans chaque couche. Les fonctions d'activation sont également sélectionnées par l'utilisateur. Pour la couche de sortie, comme mentionné précédemment, la fonction d'activation est généralement différente de celle utilisée

sur les couches cachées. En cas de régression, nous n'appliquons aucune fonction d'activation sur la couche de sortie. Pour la classification binaire, la sortie donne une prédiction de $P(Y = 1/X)$ puisque cette valeur est dans $[0, 1]$, la fonction d'activation sigmoïde est généralement considérée. Pour la classification multi-classes, la couche de sortie contient un neurone par classe i, donnant une prédiction de $P(Y = i/X)$. La somme de toutes ces valeurs doit être égale à 1. La fonction multidimensionnelle soft max est généralement utilisée.

3.3.8.3 Réseaux de neurone convolutif

Les perceptrons multicouches MLPs sont définis pour les informations de type vecteur (unidimensionnel) comme des données d'entrée et ne sont pas définis pour d'autres types complexes. Nous nous intéressons en particulier aux images. Donc pour l'appliquer aux images, nous devons transformer les images en vecteurs, et par conséquent nous perdons les informations spatiales contenues dans les images, telles que les formes. Avant le développement de l'apprentissage en profondeur pour la vision par ordinateur, l'apprentissage était basé sur l'extraction de variables d'intérêt, appelées caractéristiques, mais ces méthodes nécessitent beaucoup d'expérience pour le traitement d'image. Les réseaux de neurones convolutifs (Convolutional Neural Networks CNN) introduit par LeCun [153] ont révolutionné le traitement d'image, éliminant l'extraction manuelle des fonctionnalités. CNN manipule directement des matrices, voire sur les tenseurs d'images RVB à trois couleurs. Les CNNs sont désormais largement utilisés pour la classification d'images, la segmentation d'images, la reconnaissance d'objets, la reconnaissance faciale. Un réseau neuronal convolutif est composé de plusieurs types de couches, couches convolutives (Convolutional Layers), couches de mise en commun (Pooling Layers) et couches entièrement connectées (Fully Connected Layers) voir la figure **3.13**.

3.3.8.4 Réseaux de neurones récurrents (RNN)

Les réseaux de neurones récurrents (Recurrent Neural Networks, RNN) [154] sont un type spécifique du réseau neurone interconnecté dans lequel la sortie de l'étape précédente est envoyée en entrée à l'étape actuelle. Dans les réseaux neuronaux traditionnels, toutes les entrées et sorties sont indépendantes les unes des autres, mais dans des cas comme lorsqu'il est nécessaire de prédire le mot suivant d'une phrase, les mots précédents sont requis et doivent donc être mémorisée (voir la figure **3.14**). C'est ainsi que RNN a vu le jour, qui a résolu ce problème avec l'aide d'une couche cachée. La caractéristique primordiale de RNN est l'état caché, qui mémorise de certaines informations spécifiques sur une séquence[154].

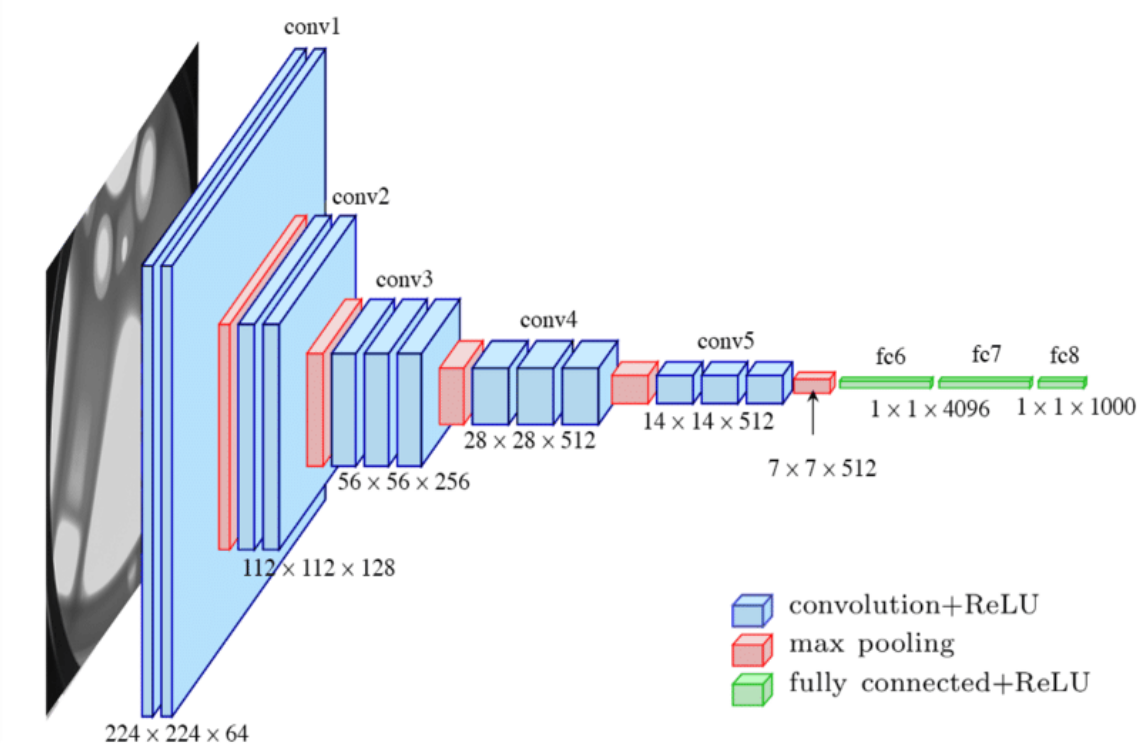

Figure 3.13 La structure d'un réseau de neurone CNN [179].

Les RNNs ont une mémoire qui se souvient de toutes les informations sur ce qui a été calculé. Il utilise les mêmes paramètres pour chaque entrée car il effectue la même tâche sur toutes les entrées ou couches cachées pour produire la sortie. Cela réduit la complexité des paramètres, contrairement aux autres réseaux de neurones.

3.3.8.5 Avantages et inconvénients

- **Avantages**
  - Meilleure prédiction des séries temporelles.
  - Possède une mémoire à court terme pour mémoriser également les entrées précédentes.
  - Voisinage effectif des pixels.
- **Inconvénients**
  - La souffrance des problèmes de fuite et d'explosion.
  - L'apprentissage est très complexe.
  - Temps de calcul très long dans la construction et la représentation des grandes séquences.

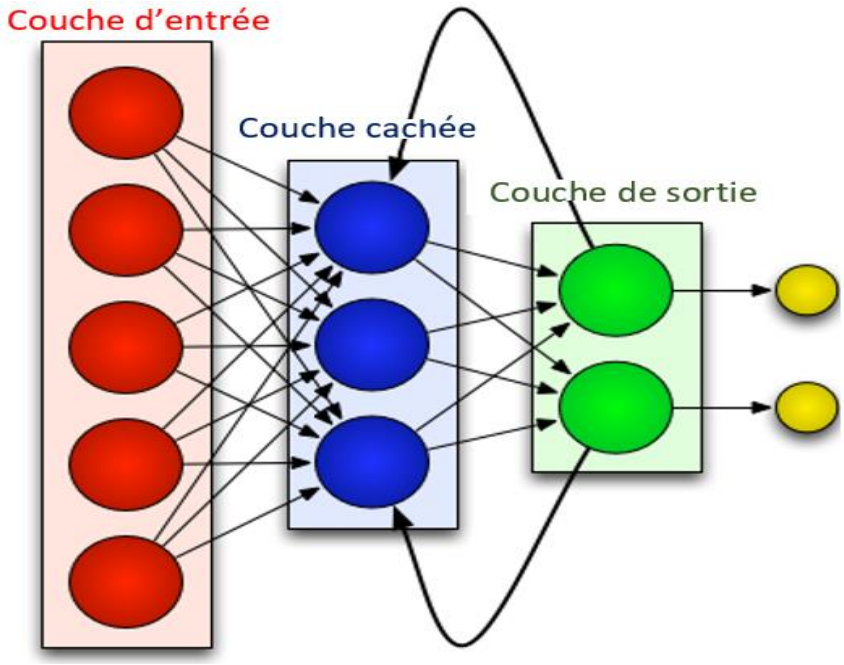

Figure 3.14 La structure d'un neurone récurrent [197].

3.3.8.6 Apprentissage en profondeur

Le Deep Learning (DL), ou apprentissage profond, est la principale technologie du ML et d'IA. Le DL repose sur la technologie des réseaux de neurones artificiels. Ils ont constitué de plusieurs couches artificielles interconnectés [155-157, 161]. Plus le nombre des neurones est élevé plus le réseau est profond, d'où vient le nom "Deep".

Le terme "Deep Learning" a été introduit pour la première fois au ML par Dechter (1986) [155], et aux réseaux neuronaux artificiels par Aizenberg et al (2000) [156]. Le Deep Learning est un nouveau domaine de recherche du ML, qui a été introduit dans le but de concevoir une machine procède de la même façon de celle d'un être humain. Ils peuvent apprendre plusieurs niveaux de représentation dans le but de modéliser des relations complexes entre les données. Il s'agit d'algorithmes inspirés par la structure et le fonctionnement du cerveau.

L'apprentissage profond repose sur l'utilisation d'une nouvelle technologie des réseaux de neurones artificiels pour exploiter et gérer des quantités massives de données tout en ajoutant des couches au réseau. Ces multiples couches permettent au réseau de neurones DL d'extraire des caractéristiques complexes à partir des données brutes à l'aide de multiples transformations linéaires et non linéaires à travers les couches multiples [157].

L'avantage primordial du Deep Learning réside dans ses performances énormes lorsqu'il s'agit de grande quantité de donnée, au contraire des algorithmes ML classiques qui sont très limité avec les grandes quantités des données. En effet, DL s'adapte bien avec une très large quantité de données, et il

peut aller jusqu' à dépasser la performance d'un être humain dans plusieurs domaines tels que le traitement d'image.

En réalité nous n'avons pas comprendre le potentiel réel du DL qu'après la disponibilité de grandes quantités de données grâce au Big Data et de nombreux objets connectés et avec le développement des machines performants qui sont devenues des supercalculateurs. A ce stage, le DL a pu dépasser les algorithmes d'apprentissage automatique traditionnels. Il est mieux adapté dans les domaines d'IA tels que la reconnaissance vocale et la reconnaissance d'objets. Une autre différence entre les algorithmes de ML traditionnels et les algorithmes de Deep Learning réside au niveau de l'étape d'extraction de caractéristiques complexes. Dans les algorithmes de ML traditionnels, l'extraction de caractéristiques est faite avec une conception indépendante avec des outils spatiaux tels que les descripteurs de la texture, c'est une étape difficile et encore beaucoup plus coûteuse en temps puisqu'ils requièrent un spécialiste en la matière alors qu'en Deep Learning cette étape est exécutée automatiquement par l'algorithme (voir figure **3.15**).

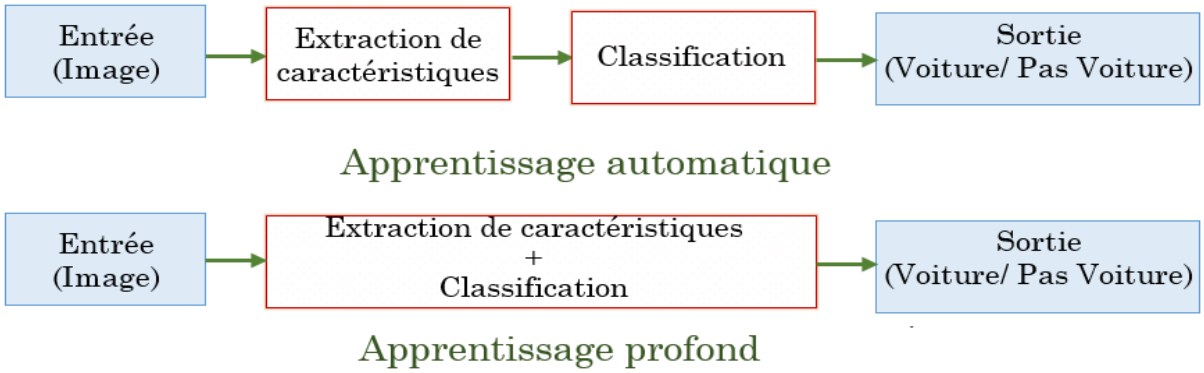

Figure 3.15 : Schéma général du ML comparé à celui du DL.

3.3.7.3 Avantages et inconvénients

- **Avantages**
  - Caractéristiques automatiquement déduites et optimisées pour le résultat souhaité.
  - robustesse aux variations apprises automatiquement.
  - Même niveau de réutilisabilité que les réseaux de neurones.
  - Calculs massivement parallèles grâce à l'utilisation de GPU évolutifs pour traiter de gros volumes de données
- **Inconvénients**
  - Technique difficile et ambigüe avec un calcul intensif.
  - Certains problèmes se prêtent mieux à l'apprentissage en profondeur que d'autres applications.
  - Modèles plus simples peuvent être suffisants pour certains domaines de problèmes.
  - Les modèles sont très difficiles à expliquer par rapport aux autres algorithmes ML.

3.4 Conclusion

Dans ce chapitre, nous avons présenté et détaillé le principe général de l'apprentissage automatique, les deux approches essentielles de l'apprentissage automatique, supervisé et non supervisé. Il existe deux types de problèmes de classification qui s'adaptent mieux à l'apprentissage automatique supervisé, tels que la reconnaissance faciale et la reconnaissance des expressions faciales. Cependant, il existe un autre type de problème qui nécessite un apprentissage non supervisé. Ensuite nous avons présenté les approches de classifications les plus connues et les plus utilisées dans le domaine de reconnaissance d'objets et plus particulièrement dans le domaine de la reconnaissance faciale et la reconnaissance des expressions faciales. Compte tenu des types des problèmes abordés dans cette thèse, nous nous sommes concentrés sur l'apprentissage automatique supervisé. Dans le domaine de la reconnaissance faciale et la reconnaissance des expressions faciales, le SVM et les réseaux de neurones (simples ou profonds) sont les plus favorisés et utilisés. D'une manière générale, les classes des problèmes traités et la nature des données utilisées et exploitées s'influent énormément sur le choix du classificateur et donc sur le temps d'exécution et la précision de la classification.

Une évaluation expérimentale des classificateurs d'apprentissage automatique sur la reconnaissance faciale et la reconnaissance des expressions faciales en termes du temps d'exécution et taux de précision, présentée et détaillée dans le chapitre suivant.

Partie II : Contributions

Chapitre 4

Evaluation des performances des modèles d'apprentissage automatique pour la reconnaissance émotionnelle faciale

Sommaire

4.1 Introduction

Les systèmes de reconnaissance des expressions faciales (Facial Expression Recognition FER) ont attiré beaucoup d'attention de la recherche scientifique dans les dernières décennies. FER a prouvé de nombreux avantages et a connu un grand succès dans le domaine de vision par ordinateur. Il est devenu de nos jours de plus en plus important, en effet ils sont appliquée dans divers domaines d'applications dans la vie quotidienne à savoir l'interface homme-machine (HCI), l'analyse psychologique automatique, la sécurité et la surveillance aéroportuaire et l'éducation robotique en particulier, il offre une meilleure expérience d'apprentissage avec une meilleure compréhension des émotions des étudiants. Des systèmes d'apprentissage en ligne existent permet d'estimer la tendance criminelle et la sécurité des conducteurs.

Les expressions faciales ont une importance capitale pour la qualité d'interaction humaine et la communication sociale. Les expressions faciales ont la capacité de transmettre des émotions, des énergies et des expressions sans utiliser de mots. D'ailleurs le visage humain est capable de générer des milliers d'expressions faciales. Toutes les approches d'apprentissage automatique FER nécessitent un ensemble d'images de test et d'entrainement, chacun étiqueté avec une seule catégorie d'émotion. Un ensemble standard de six catégories d'émotions est la colère (AN), le dégoût (DI), la peur (FE), le bonheur (HA), la tristesse (SA) et la surprise (SU) en tant que «expressions atomiques». Ces six

expressions sont uniques parmi les différentes races, religions, cultures et groupes [158]. Certains chercheurs considèrent le visage neutre comme une septième expression [159]. Malgré les avantages puissants qu'à procurent les systèmes d'interaction homme-machine, la classification des émotions humaines peut être une tâche difficile pour les machines. Cependant, une précision de classification plus élevée et un temps d'exécution plus court restent les principaux challenges lors de l'extraction des caractéristiques des émotions humaines. Donc, il a fallu trouver des techniques d'apprentissage précises et efficaces nous permettant de favoriser la création et le développement des systèmes de reconnaissance émotionnelle faciale pour les grandes bases de données des images faciales.

Généralement, les systèmes FER peuvent être classés en deux approches fondamentales, à savoir les approches basées sur la géométrie et les approches basées sur l'apparence [162 - 168]. Chacun a ses avantages et ses inconvénients. Dans ce chapitre, nous présentons notre première contribution, une approche d'évaluation expérimentale des classificateurs existants de reconnaissance des expressions faciales, pour le problème de recherche d'une technique approprié de classification. Elle est basée sur (i) l'extraction des caractéristiques discriminantes et pertinentes qui sont implicite des images des visages et (ii) la réalisation d'un apprentissage automatique comparatif aux émotions faciales humaines, y compris deux nouvelles techniques de classification, la première est **un contre autres** et la seconde est le **test contre entrainement**. Notre but est de trouver des classificateurs appropriés qui répondent au mieux à l'attente de l'utilisateur, réduisent le temps d'exécution et augmentent la précision du système.

Ce chapitre présente un module d'extraction automatique de points de repère amélioré avec Active Shape Model SVM [199]. L'évaluation du temps d'exécution et de la précision de sept classificateurs parmi les algorithmes les plus populaires dans FER : K-Nearest Neighbor (KNN) [8], Naive Bayes (NB) [143], Support Vector Machine (SVM) [7], Decision Tree (DT) [146], Quadratic Discriminant Analysis (QDA) [150], Random Forest (RF) [149] et MULTI-Layered Perceptron (MLP) [152]. En effet nous avons effectué plusieurs tests sur la base de données CK+[180] pour prédire la classe d'expression faciale. L'approche proposée conduisant à de nouveaux classificateurs appropriés pour explorer la reconnaissance automatique des émotions via l'apprentissage automatique.

4.2 Approche basée sur le modèle des contours actifs pour la reconnaissance des expressions faciales

Les systèmes de reconnaissance des expressions faciales recherchent des solutions effectives et efficaces. En effet, ils souhaitent avoir une technique d'apprentissage rapide et adaptée aux attentes des utilisateurs et aux changements d'expressions faciales. Le temps d'exécution rapide est imposé en raison de nombre important des images faciales. En plus des besoins de reconnaissance de qualité, plusieurs utilisateurs exigent des précisons élevées de classification. Pour répondre aux exigences de ces utilisateurs, les

techniques d'apprentissage automatique (*ou machine Learning*) existent sont des solutions pour la reconnaissance des expressions faciales. Chaque solution présente des avantages et des inconvénients. En outre, nous menons des évaluations de performance des techniques d'apprentissages automatiques existants en termes de taux de précisions et temps de calcul. Notre but est de comparer et d'analyser d'une manière expérimentale la technique d'apprentissage selon la précision, le temps d'exécution. Cependant, la plupart d'entre eux souffrent d'une mauvaise précision et/ou d'un temps d'exécution élevé dans la plupart des images faciales.

Pour remédier cela, nous proposons un système qui vise à offrir aux utilisateurs un meilleur classificateur avec une précision élevée et un temps d'exécution réduit adapté au domaine de la reconnaissance des expressions faciales. Ce système consiste en un module personnalisé appelé Active Shape Model ou modèle des contours actifs (ASM) [162] combiné avec sept classificateurs les plus populaires. Le système extrait les points de repères discriminantes des images de visages et les sélectionnés.

4.2.1 Architecture générale

L'architecture du système FER proposé est décrite dans la figure **4.1** Le système proposé se compose de deux phases fondamentales et successives : *apprentissage* et *tests*. Pendant la phase d'apprentissage, le système extrait les caractéristiques et les points de repère de toutes les images faciales en tant que vecteurs caractéristiques de forme déterminés par l'ASM. La normalisation et l'alignement sont introduites au niveau des vecteurs pour éliminer les changements des transformations géométriques (translations, rotation et de la mise à l'échelle). De même, le processus de test extrait les caractéristiques et les points de repère de l'image de test et sera comparé à ceux constituées à partir des vecteurs extraits dans la base de données pour identifier ses classes d'émotions.

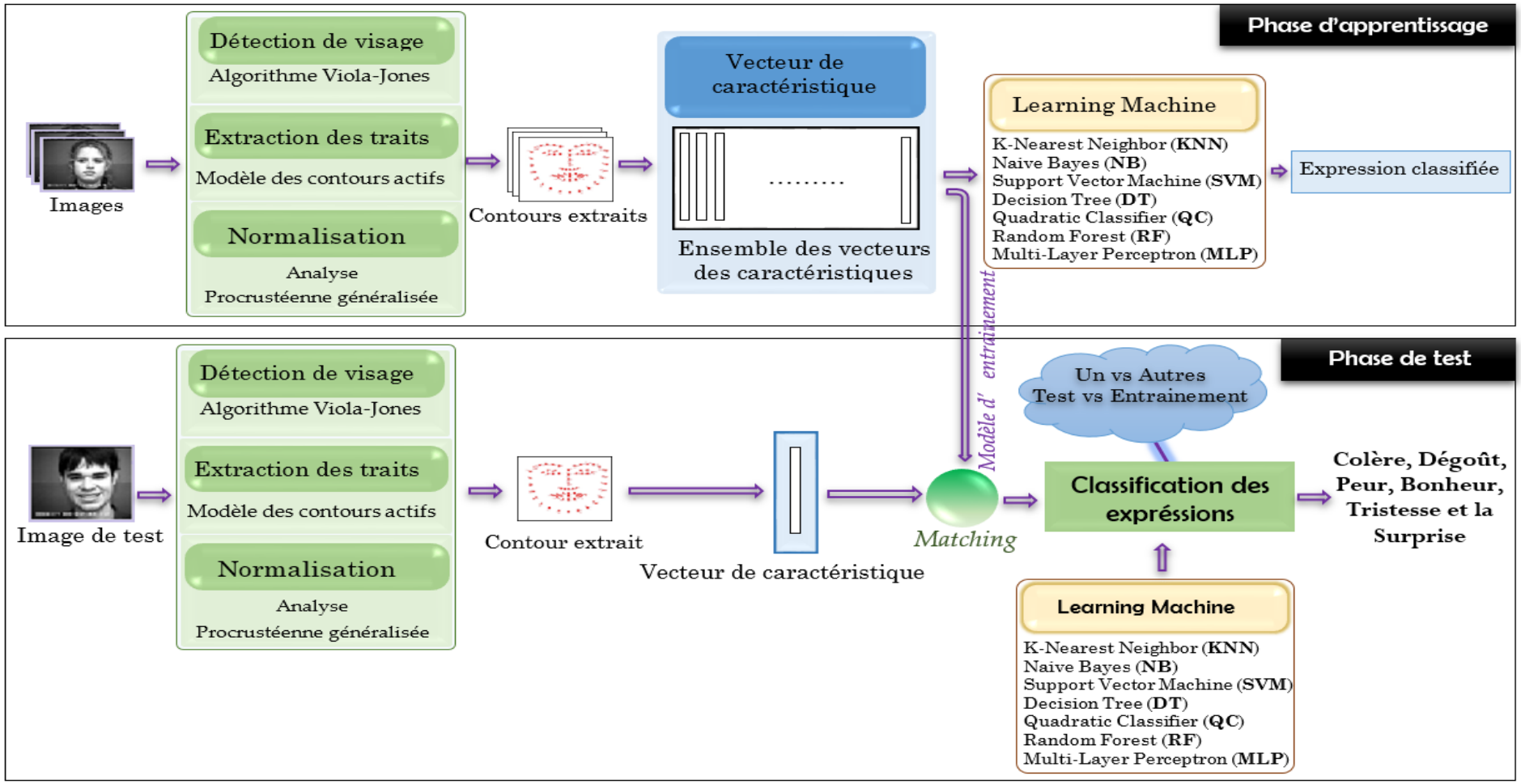

Figure 4.1 : Architecture générale du système proposé.

4.2.2 Phase de détection du visage

La phase de détection du visage localise le visage dans l'image d'entrée en utilisant l'algorithme de Viola-Jones [1]. Cet algorithme est souvent efficace pour résoudre des problèmes complexes de bord et des changements de luminosité. Un tel algorithme est invariant au bruit et la mise en échelle. L'algorithme de Viola-Jones est défini en deux étapes principales : l'extraction des caractéristiques HAAR et la classification à l'aide d'Adaboost [1].

4.2.3 Phase d'extraction de caractéristiques discriminantes

La phase d'extraction des caractéristiques correspondrait à l'extraction des repères faciaux (en anglais, Landmarks) les plus discriminants de l'image du visage. Cela signifie que les caractéristiques des expressions faciales locales et globales sont extraites à partir du modèle des contours actifs (ASM), pour améliorer l'extraction des contours les plus indicatifs d'une classe spécifique. Pour un objet visage, le modèle des contours actifs (ASM) prédit les arêtes optimales grâce à des transformations géométriques pour caractériser les formes des visages et leurs expressions faciales. L'utilisation de cet algorithme nous a permis de générer 68 points de repère afin d'obtenir des résultats plus précis. Il a été inclus dans notre système après l'étape de détection de visage. Le modèle des contours actifs (ASM) est une analyse statistique qui permet de déterminer la forme moyenne et ses variations admissibles dans un ensemble d'images $q_i, (q_i i = 1..m)$ représentant un ensemble d'objets constitués de coordonnées cartésiennes $(x_j, y_j, j = 1..n)$, dont chacune est un point de repère particulier. Les points de repère sont décrits dans un vecteur de forme et calculés par l'équation 5.1 :

$$q_i = \begin{pmatrix} x_1 & x_2 \dots & x_n \\ y_1 & y_2 \dots & y_n \end{pmatrix}, \quad for\ i = 1..n \tag{4.1}$$

n est le nombre de points de repère.

Le processus d'extraction de fonctionnalités est décrit comme suit (voir figure **4.2**) :

- **Etape 1 -** La forme moyenne $\bar{q}$ du vecteur de formes est calculée comme suit :

$$\bar{q} = \frac{1}{m}\sum_{i=1}^{m} q_i \tag{4.2}$$

- **Etape 2-** La matrice de covariance S pour capturer la variabilité de forme est calculée comme suit :

$$S = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{q} - q_i)(\bar{q} - q_i)^T \tag{4.3}$$

- **Etape 3-** Le vecteur propre et la valeur propre λ_i de la matrice de covariance S. Le vecteur propre définit le mode de variations de tous les points de la forme calculée comme suit :

$$Su_i = \boldsymbol{\lambda}_i u_i \tag{4.4}$$

λ_i est la ième valeur propre de la matrice de covariance S et $U = (u_1|u_2|\dots|u_t)$ contient t vecteurs propres de la matrice des covariances S

- **Etape 4-** Nous approchons n'importe quelle instance de la forme en projetant sur les t premiers vecteurs propres comme suit :

$$q = \bar{q} + U \times b \tag{4.5}$$

Où, $b = [b_1, b_2, \dots b_n]^T$ désigne le vecteur de poids qui est identifié comme une caractéristique de cette instance de la forme. Varier les poids b_i nous permet d'explorer les variations admissibles de la forme. Ainsi, nous trouvons une représentation de forme optimale dans les étapes suivantes.

- **Etape 5 -** Initialiser le vecteur de poids b à zéro.
- **Etape 6 -** Générer l'instance de modèle initiale à l'aide de **l'Eq. 5.**
- **Etape 7 -** Calculer respectivement les paramètres de translation, de rotation et de mise à l'échelle (X_t, Y_t, s, θ) qui alignent le mieux l'instance de modèle.
- **Etape 8 -** Mettre à jour le vecteur de poids *b* à l'aide de l'équation suivante :

$$b = U^t(q - \bar{q}) \tag{4.6}$$

- **Etape 9 - si** les poids ou les paramètres de position sont modifiés **alors** revenez à ***l'étape 4*** **sinon** arrêter.

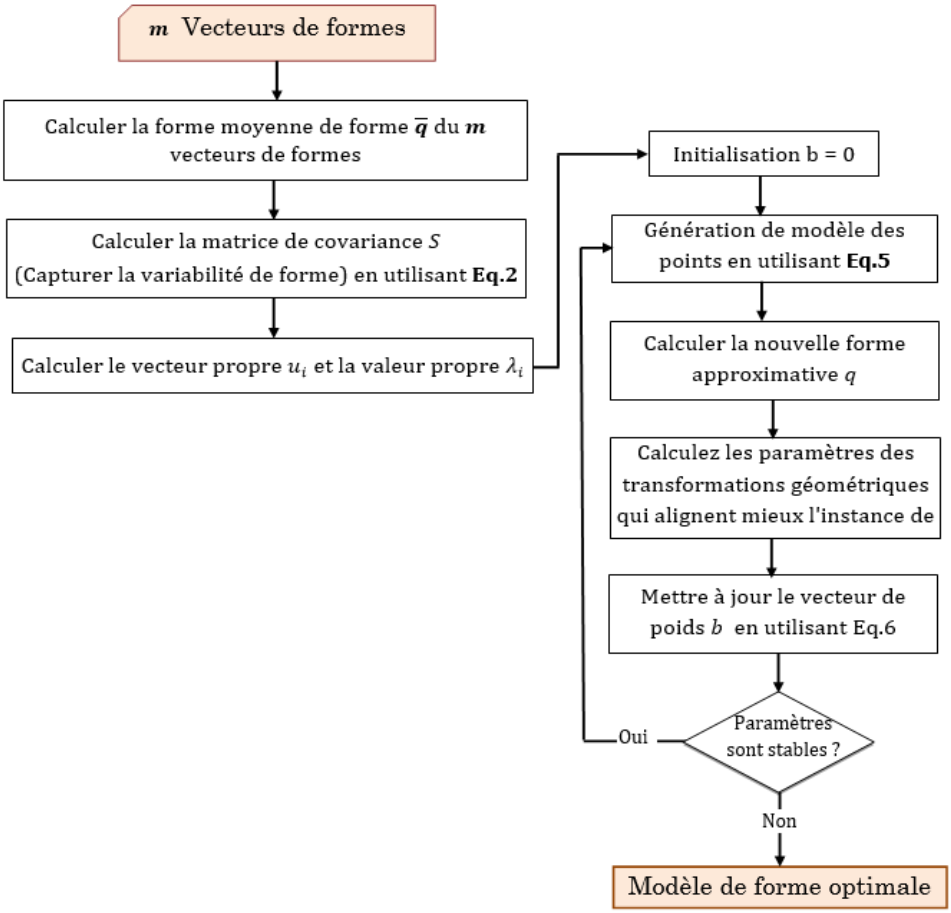

Figure 4.2. Organigramme de la phase d'extraction de caractéristiques.

4.2.5 Phase de normalisation des caractéristiques

Après l'étape d'extraction des fonctionnalités, nous passons à l'étape de normalisation des fonctionnalités. elle est utilisée pour éliminer les effets d'échelle, de rotation et de translation entre toutes les formes à l'aide du modèle ASM pour déterminer correctement les informations faciales pertinentes, ce qui augmentera et améliorera considérablement le taux de reconnaissance du système proposé. Nous avons utilisé l'analyse Procruste généralisée (GPA) qui a été proposée par [163]. La figure **4.3** montre un exemple de contours avant et après l'application du GPA.

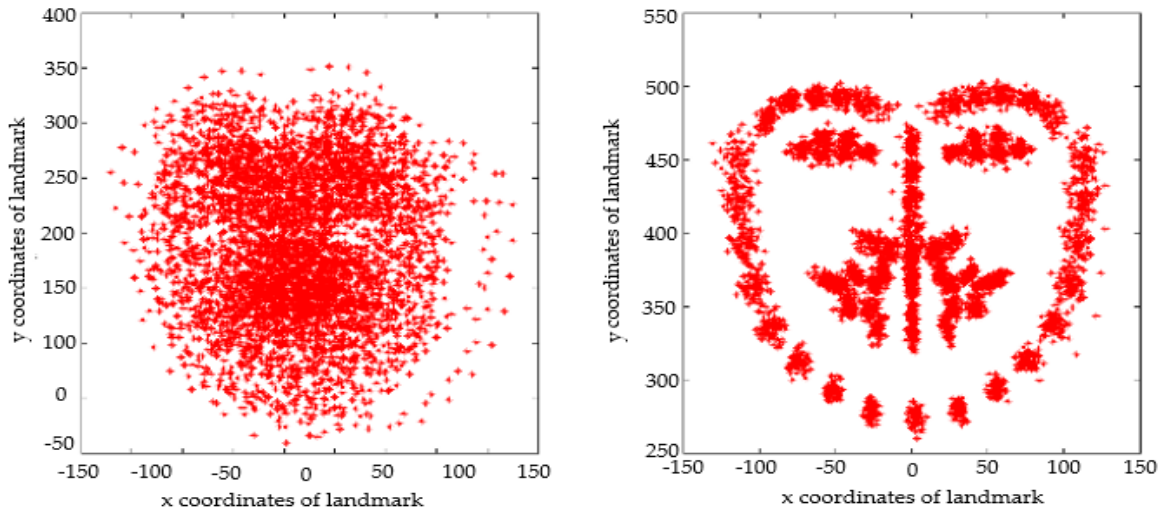

Figure 4.3 Normalisation de formes, exemple des formes de l'ensemble heureux **a -** avant, **b -**après.

4.2.6 Phase de construction et classification des vecteurs de caractéristiques

Après l'étape de normalisation des traits, nous sauvegardons les vecteurs de contours dans une base de données. Elle contienne tous les vecteurs de caractéristiques qui représentent les émotions humaines. Soit F la base de données contenant les vecteurs de caractéristiques définis par :

$$F = (q_1 ; q_2; \dots ; q_m) \tag{4.7}$$

Où, q ¡est un vecteur de forme défini et conçu à l'aide des étapes précédentes et m est le nombre d'images de forme.

La figure **4.4** montre un exemple des points de repère extraits par notre système sur un ensemble standard de six expressions faciales. Nous avons effectuée plusieurs expériences de performance des techniques d'apprentissage avec l'algorithme proposé et en comparons ses résultats en appuyant sur la précision et le temps d'exécution. Notre objectif est de sélectionner le meilleur classificateur parmi les classificateurs les plus standards pouvant être utilisé dans la reconnaissance des expressions faciales :

- K-plus proches voisins (**KNN**)
- Classification bayésienne (**NB**)
- Machine du vecteur de support (**SVM**)
- Arbre de décision (**DT**)
- Classificateur quadratique (Q**DA**)
- Forêt aléatoire (**RF**)
- Les perceptrons multicouches (**MLP**).

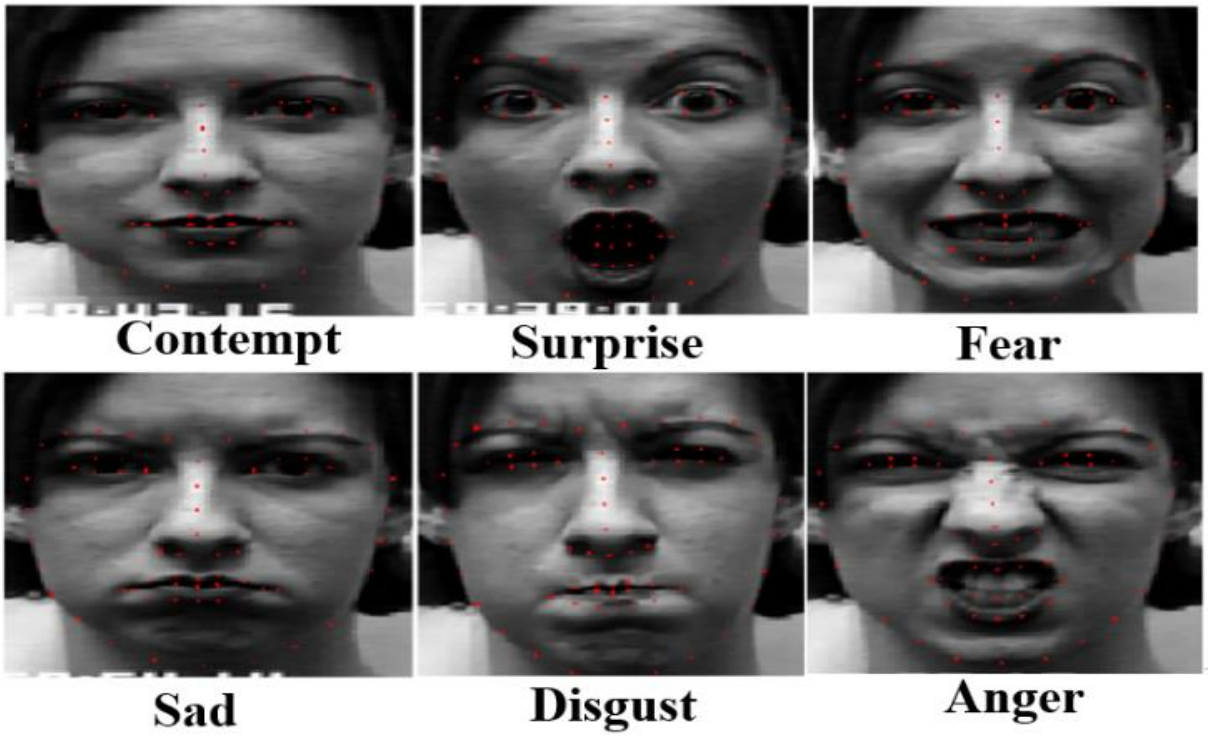

Figure 4.4. Images de repères d'expression faciale depuis de notre système dans la base CK+.

4.3 Base de données et Apprentissage

4.3.1 Base de données

Des expérimentations ont été menés sur une base de visages étendu Cohn-Kanade (CK +) [180] afin de mesurer la qualité de notre approche au regard d'autres classificateurs existants. La base de données CK+ contient 326 images d'expressions faciales pour sept catégories d'émotions : la colère (AN), le dégoût (DI), la peur (FE), le bonheur (HA), la tristesse (SA) et la surprise (SU) variant entre les expressions faciales posées et non posées de 210 adultes. L'ensemble d'images allant de 18 à 50 ans se compose de 69 femmes, 81euro-américaines, 13 afro-américaines et 6 autres groupes. La résolution de toutes les images d'apprentissage et tests de taille 256 × 256. Chaque catégorie d'émotions contient le nombre suivants des expressions :

- colère (45),
- mépris (18),
- dégoût (59),
- peur (25),
- bonheur (69),
- tristesse (28) et surprise (82).

Aucun objet du visage n'a été collecté plus d'une fois avec la même émotion. Nous considérons la meilleure qualité pour prendre en compte les caractéristiques d'une base de données.

4.3.2 Apprentissage

Nous considérons deux approches de classification ***un contre autres*** et ***test contre entrainement*** qui sont le meilleur moyen de réaliser une étude comparative approfondie de sept classificateurs parmi les algorithmes FER les plus populaires.

- **Un-Contre-Autres** : tout d'abord, nous prenons une image de visage d'une base de données comme image de test et les images restantes comme ensemble d'images d'entrainements. En appliquant le processus d'apprentissage de sept classificateurs aux images de visage originales, de nouvelles caractéristiques des expressions faciales sont extraites. Une fois l'extraction des caractéristiques terminée on applique la phase de test sur l'image retenue de la base de données pour prédire la classe résultat. Le même processus est appliqué sur toutes les images de visage de la base de données. Dans cette approche, le processus de test a classé toutes les images de visage et comparé une par une au reste des images de la base de données. Nous évaluons les performances et la précision du FER par chaque catégorie et comparons l'efficacité de chaque catégorie avec celles des catégories traditionnelles.
- **Test-Contre-Entrainement** : nous effectuons des expériences approfondies sur notre ensemble de données en utilisant sept classificateurs dans les images de visage de test et d'entraînement. L'ensemble de données contenant 326 images a été divisé en 2 parties : apprentissage et test. Nous construisons sept classification avec différentes paires de {*test*, *entrainement*} pour les sept classificateurs. Tout d'abord, nous prenons une image de visage de chaque classe dans l'ensemble de test, puis deux images de test de chaque classe dans la deuxième étape jusqu'à dix visages de chaque classe dans l'ensemble de test et le reste à chaque fois sera pour l'ensemble d'apprentissage. Enfin, nous évaluons les performances et la précision du FER en utilisant sept classificateurs dans toutes les images de visage des groupes de test.

4.3.3 Les mesures d'évaluation

Nous avons utilisé six mesures d'évaluation afin d'évaluer les performances de notre système :

- taux d'exactitude.
- précision par classe.
- rappel par classe.
- f-mesure par classe.
- temps d'exécution.
- matrice de confiance.

4.3.3.1 Exactitude

La notion d'exactitude, Acc, dépend du rapport entre le nombre de bonnes émotions, et le nombre total d'émotions.

$$Acc = \frac{VP}{TP+VP+FN+TN} \tag{4.8}$$

4.3.3.2 Précision

La précision, Pre, est un rapport entre le nombre d'émotions correctement classées et l'ensemble de toutes les émotions classées. La valeur de précision est entre 0 et 1 avec une valeur proche ou égale à 1 indiquant une meilleure performance de classification. On calcul la précision de classification comme suit :

$$Pre = \frac{VP}{VP+FP} \tag{4.9}$$

4.3.3.3 Rappel

Le rappel, Rec, est un rapport entre le nombre d'émotions correctement classées et l'ensemble de toutes les émotions classées attendues. Un meilleur rappel est exprimé lorsque Rec est égal à 1. On calcule le rappel de classification comme suit :

$$Rec = \frac{VP}{TP+FN} \tag{4.10}$$

4.3.3.4 F-Mesure

La mesure F1, est utilisée pour évaluer une moyenne pondérée entre Prec et Rec. Cela met indirectement en évidence les émotions classifiées attendues manquées, ce qui est un facteur important basé sur le rappel pondéré. F-Mesure atteint la meilleure valeur à 1. Elle est calculée comme suit :

$$F_1 = 2 * \frac{Pre*Rec}{Pre+Rec} \tag{4.11}$$

Où,

- VP: le nombre d'émotions correctement classées (vrais positifs)
- FP: le nombre d'émotions mal classées (vrais négatifs)

- VN: le nombre d'émotions correctement classés qui ne sont pas classés par l'approche donnée (vrais négatifs)
- FN: le nombre d'émotions mal classés qui ne sont pas classés par l'approche donnée (faux négatifs)

4.4 Evaluation des performances et comparaisons

Pour réaliser une étude comparative approfondie de sept classificateurs dans FER, il est nécessaire de les comparer en termes de temps d'exécution, d'exactitude, de précision et de score F1 à travers deux nouvelles approches de classification ***Un-Contre-Autres*** et ***Test-Contre-Entrainement***.

4.4.1 Comparaison du temps d'exécution

Nous avons évalué le temps d'exécution avec les algorithmes de classifications les plus courants de FER. Le temps d'exécution est le temps de réponse moyen nécessaire pour accomplir l'étape de test de tous les classificateurs à l'exception des SVM qui utilisent des étapes de test et de formation liées. La figure 4.5 montre la comparaison du temps d'exécution pour sept classificateurs formés sur l'ensemble de données CK +. Pour TREE, KNN, NB et QDA (DA) montrent une amélioration supérieure en termes de temps d'exécution. Le classificateur MLP et FR donne un temps d'exécution inférieur à celui des autres classificateurs. La taille de notre vecteur de fonctionnalités était plus petite par rapport aux autres descripteurs tels que LBP, HOG … etc.

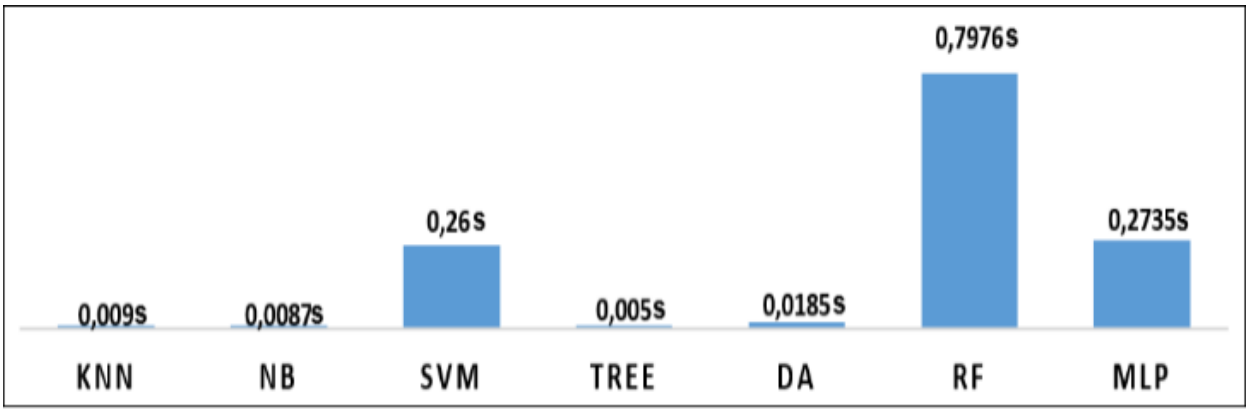

Figure 4.5 : Temps d'exécution pour chaque classificateur dans la base CK+.

4.4.2 Comparaisons des classificateurs par l'approche un-contre-autres

Le tableau **4.1** et la figure **4.6** montre les mesure d'exactitude, de précision, de rappel et de F-score pour les sept classificateurs appliqués sur l'ensemble de données CK+. Comme prévu, SVM, DA et FR présentent un taux de reconnaissance idéal de 100%, ce qui prouve leurs efficacités. On observe également que les taux de reconnaissance de KNN, NB, TREE et NN sont très satisfaisants et très proches les uns des autres. Nous pouvons constater que SVM, QDA et FR présentent des résultats de précision impressionnants.

Classificateur	Exactitude	Précision	Rappel	F-Mesure
KNN	96.93	95.48	95.09	95.29
NB	96.32	95.79	95.13	95.46
SVM	**100**	**100**	**100**	**100**
TREE	97.55	96.70	96.94	96.81
DA	**100**	**100.**	**100**	**100**
RF	**100**	**100**	**100**	**100**
MLP	98.16	98.10	98.19	98.14

Table 4.1. Comparaison des performances entre sept classificateurs par l'approche *un-contre-autres*

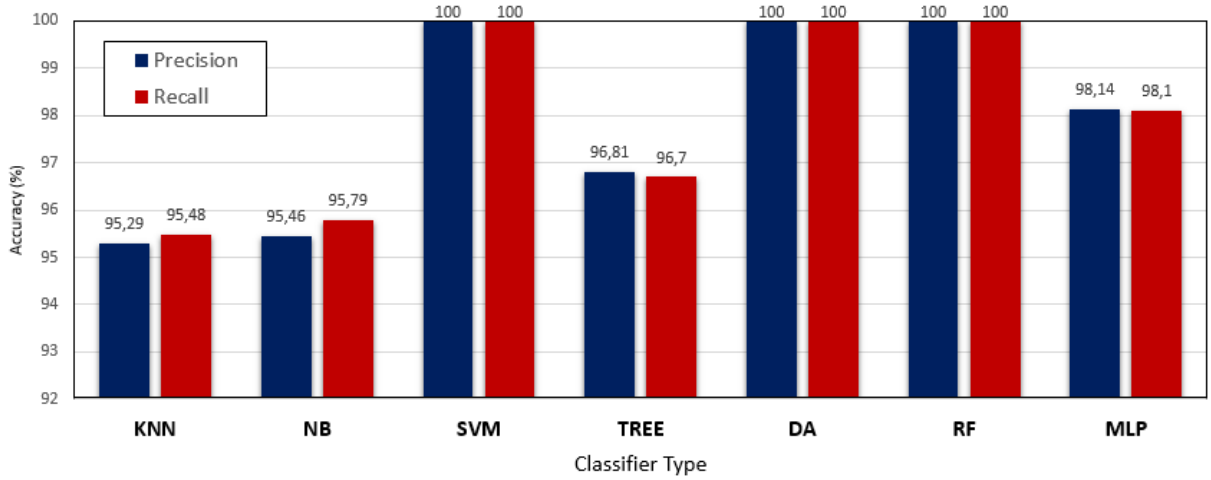

Figure 4.6 : Taux de précision et rappel pour chaque classificateur dans la base CK+.

L'évaluation des sept classificateurs se base sur la matrice de confusion sur un jeu de données CK+. Les tableaux **4.2, 4.3, 4.4** et **4.5** montrent la matrice de confusion sur un jeu de test en utilisant sept classificateurs. En particulier, les valeurs de correspondance des classificateurs SVM, QDA et FR sont idéales et aucun chevauchement entre les expressions faciales été constaté. A partir des expériences ci-dessus, on constate que le classificateur quadratique(DA) est le meilleur classificateur en termes de temps d'exécution et de taux de précision, tandis que les classificateurs KNN et NB viennent en dernier.

	AN	CO	DI	FE	HA	SA	SU
AN	**100.00%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CO	0.00%	**100.00%**	0.00%	0.00%	0.00%	0.00%	0.00%
DI	0.00%	0.00%	**100.00%**	0.00%	0.00%	0.00%	0.00%
FE	0.00%	0.00%	0.00%	**100.00%**	0.00%	0.00%	0.00%
HA	0.00%	0.00%	0.00%	0.00%	**100.00%**	0.00%	0.00%
SA	0.00%	0.00%	0.00%	0.00%	0.00%	**100.00%**	0.00%
SU	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**100.00%**

Table 4.2. Matrice de confusion des SVM, DA, and RF dans la base CK+.

	AN	CO	DI	FE	HA	SA	SU
AN	**95.56%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CO	0.00%	**94.44%**	0.00%	0.00%	0.00%	3.57%	0.00%
DI	2.22%	0.00%	**98.30%**	0.00%	0.00%	0.00%	0.00%
FE	0.00%	0.00%	0.00%	**92.00%**	0.00%	3.57%	0.00%
HA	0.00%	0.00%	0.00%	0.00%	**100.00%**	0.00%	0.00%
SA	2.22%	5.56%	0.00%	4.00%	0.00%	**89.29%**	0.00%
SU	0.00%	5.56%	0.00%	0.00%	0.00%	0.00%	**98.78%**

Table 4.3. Matrice de confusion de KNN dans la base CK+.

	AN	CO	DI	FE	HA	SA	SU
AN	**93.33%**	0.00%	5.08%	0.00%	0.00%	0.00%	0.00%
CO	0.00%	**94.44%**	0.00%	4.00%	0.00%	0.00%	0.00%
DI	8.89%	0.00%	**93.22%**	0.00%	0.00%	0.00%	0.00%
FE	0.00%	0.00%	0.00%	**92.00%**	0.00%	7.14%	0.00%
HA	0.00%	0.00%	0.00%	0.00%	**100.00%**	0.00%	0.00%
SA	0.00%	0.00%	0.00%	4.00%	0.00%	**100.00%**	0.00%
SU	0.00%	5.56%	0.00%	0.00%	0.00%	0.00%	**97.56%**

Table 4.4. Matrice de confusion de NB dans la base CK+.

	AN	CO	DI	FE	HA	SA	SU
AN	**100.00%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CO	0.00%	**100.00%**	0.00%	0.00%	0.00%	0.00%	0.00%
DI	2.22%	0.00%	**98.30%**	0.00%	0.00%	0.00%	0.00%
FE	2.22%	5.56%	0.00%	**92.00%**	0.00%	0.00%	0.00%
HA	2.22%	0.00%	0.00%	0.00%	**98.55%**	0.00%	0.00%
SA	4.45%	0.00%	0.00%	0.00%	1.45%	**89.29%**	0.00%
SU	0.00%	5.56%	0.00%	0.00%	0.00%	0.00%	**98.78%**

Table 4.5. Matrice de confusion de TREE dans la base CK+.

	AN	CO	DI	FE	HA	SA	SU
AN	**100.00%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CO	2.22%	**94.44%**	0.00%	0.00%	0.00%	0.00%	0.00%
DI	4.44%	0.00%	**96.61%**	0.00%	0.00%	0.00%	0.00%
FE	0.00%	0.00%	0.00%	**100.00%**	0.00%	0.00%	0.00%
HA	0.00%	0.00%	0.00%	1.69%	**95.65%**	3.57%	1.22%
SA	0.00%	0.00%	0.00%	0.00%	0.00%	**100.00%**	0.00%
SU	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**100.00%**

Table 4.6. Matrice de confusion de MLP dans la base CK+.

4.4.3 Comparaisons des classificateurs par l'approche test-contre-entrainement

Après avoir divisé la base CK+ selon le modèle de classification test-contre-entrainement, nous obtenons les paires testées : (7, 326-7), (14, 326-14)… (70, 326-70). Le tableau **4.7** montre les résultats de précision de l'approche proposée pour dix paires. De plus, après l'analyse des résultats, nous constatons que le classificateur quadratique(QDA) présente des résultats de précision très intéressants. Les classificateurs SVM, NB, RF et KNN avec de bons résultats peuvent favoriser les performances de reconnaissance de l'expression faciale. Puis TREE et MLP sont dans la dernière place.

4.4.4 Synthèse et discussion

L'objectif principal de ce travail est de réaliser un système permettant d'évaluer les performances des techniques d'apprentissage automatique basées sur le modèle des contours actifs(AS) pour extraire les caractéristiques discriminantes de visage et leurs expressions faciales. En effet, le système doit évaluer les techniques d'apprentissage automatique FER et de sélectionner le meilleur modèle d'apprentissage automatique qui donne des résultats très précis avec un temps d'exécution optimale. Le résumé de ce travail est décrit ci-dessous :

- La classification des expressions faciales est une tâche complexe pour les machines, et les solutions proposées ne manquent pas d'inconvénients. Donc, la sélection de meilleures techniques d'apprentissage automatique peut améliorer le taux de précision afin de reconnaître les expressions faciales. Dans ce travail, un système FER analytique est développé pour sélectionner le classificateur ayant un meilleur taux de précision. La classification basée sur le modèle des contours actifs (ASM) permet de garantir une bonne mise en œuvre et une bonne performance de la sélection des techniques d'apprentissage sur un même ensemble de données CK+.
- Le travail réalisé permettant d'identifier le meilleur classificateur d'expressions faciales avec des tests de comparaison entre les classificateurs. Depuis les figures **4.5, 4.6, 4.7** et **4.8** et les tableaux

4.2, 4.3, 4.4, 4.5 et **4.6**, la comparaison des performances a été effectuée selon différents paramètres en termes d'exactitude, précision, rappel et F-Score. Les expérimentations montrent que le classificateur d'analyse quadratique(QDA) fournit les meilleurs résultats par rapport à d'autres techniques d'apprentissages.

- Le système développé a pris en compte deux nouveaux processus de classification ***un contre autres*** et ***test contre entrainement***. Il est présenté dans les tableaux 4.1 et 4.7 considérant deux processus de classification. Le classificateur d'analyse quadratique(QDA) est nettement meilleur en termes des résultats obtenus par rapport aux classificateurs existants. Cette tâche de reconnaissance consiste à classifier des expressions émotionnelles en extrayant les traits du visage obtenus en 0.00185 secondes à l'aide du classificateur QDA avec un vecteur de caractéristiques de taille 136 éléments.

Taille		7	14	21	28	35	42	49	56	63	70
KNN	Acc	100	92.85	95.23	89.28	88.57	78.57	77.55	69.64	65.07	62.85
	Pre	100	92.85	95.23	89.28	88.57	78.57	77.55	69.64	65.07	62.85
	Rec	100	95.23	96.42	92.38	91.83	74.14	74.28	64.76	55.71	53.84
	F1s	100	94,03	95.82	90.80	90.17	76.29	75.88	67.12	60.03	58.00
NB	Acc	100	100	90.47	89.28	82.85	76.19	75.51	66.07	61.90	60.00
	Pre	100	100	90.47	89.28	82.85	76.19	75.51	66.07	61.90	60.00
	Rec	100	100	92.85	90.71	86.59	74.48	74.28	64.76	57.56	55.13
	F1s	100	100	91.65	89.99	84.68	75.33	74.89	65.41	59.65	57.46
SVM	Acc	100	100	95.23	82.14	85.71	83.33	75.51	64.28	61.90	58.57
	Pre	100	100	95.23	82.14	85.71	83.33	75.51	64.28	61.90	58.57
	Rec	100	100	96.42	86.42	87.44	78.57	71.59	70.30	55.84	53.23
	F1s	100	100	95.82	84.23	86.57	80.88	73.49	67.15	58.71	55.77
TREE	Acc	71.42	92.85	71.42	67.85	71.42	66.66	59.18	50.00	53.96	54.28
	Pre	71.42	92.85	71.42	67.85	71.42	66.66	59.18	50.00	53.96	54.28
	Rec	57.14	95.23	78.57	74.04	75.79	65.60	69.82	56.25	51.12	48.14
	F1s	63.49	94.03	74.82	70.81	73.54	66.13	64.06	52.94	52.50	51.03
DA	**Acc**	**100**	**100**	**100**	**96.42**	**94.28**	**85.71**	**77.55**	**73.21**	**71.42**	**68.57**
	Pre	**100**	**100**	**100**	**96.42**	**94.28**	**85.71**	**77.55**	**73.21**	**71.42**	**68.57**
	Rec	**100**	**100**	**100**	**97.14**	**95.91**	**78.57**	**74.28**	**71.90**	**57.14**	**54.76**
	F1s	**100**	**100**	**100**	**96.78**	**95.09**	**81.98**	**75.88**	**72.55**	**63.49**	**60.89**
RF	Acc	100	100	90,47	85,71	88.57	71.42	69.38	67.85	58.73	60.00
	Pre	100	100	90,47	85,71	88.57	71.42	69.38	67.85	58.73	60.00
	Rec	100	100	92,85	87.14	91.15	73.94	72.02	64.76	56.72	54,76
	F1s	100	100	91.65	86.42	89.84	72.66	70.68	66.27	57.70	57,26
MLP	Acc	71.42	50.00	57.14	53.57	60.00	69.04	53.06	58.92	52,38	54,28
	Pre	71.42	50.00	57.14	53.57	60.00	69.04	53.06	58.92	52,38	54,28
	Rec	57.14	40.47	47.85	47.78	52.55	56.29	46.25	53.90	51,36	42,92
	F1s	63.49	44.73	52.08	50.50	56.02	62.02	49.42	56.30	51,86	47,94

Tableau 4.7. Comparaison des performances de sept classificateurs par l'approche ***test-contre-entrainement***

4.5 Conclusion

Dans ce chapitre, nous avons montré un système de reconnaissance des expressions faciales simple, rapide et efficace. Le système est basé sur l'extraction des points de repère (Landmarks) complété par l'utilisation des techniques d'apprentissage automatiques pour extraire les formes géométriques des expressions faciales. Elles sont décrites par des modèles des contours actifs (ASM).

Le système proposé se compose de plusieurs phases qui montrent d'une part, sa robustesse aux transformations géométriques et, d'autre part, sa capacité de convergence vers des modèles de formes optimales. Le système développé est testé et validé avec deux nouveaux processus de classification **« un contre autres »** et **« test contre entrainement »** sur le même ensemble de données CK+ en utilisant sept modèles de classifications parmi les algorithmes FER les plus populaires. Les expérimentations montrent de bons résultats de mesure de performance et d'efficacité de classification des expressions faciales basées sur l'usage de l'analyse quadratique (QDA) en comparaison aux résultats obtenues avec l'usage des autres classificateurs standards. On remarque que la qualité de la méthode d'extraction des caractéristiques dépend de la qualité du classificateur et la taille de l'ensemble de données d'apprentissage. Pour ces raisons et d'autres, la reconnaissance des expressions faciales par les descripteurs classiques présente des inconvénients causé par le choix de tel ou tel descripteur ainsi la transformation géométriques sur l'image, aux taux de précision et aux exigences du temps de réponse des utilisateurs. Pour cela, nous intéressons à un système automatique pour extraire de caractéristiques de visage et leurs expressions faciales grâce à des techniques d'apprentissage en profondeur et montrer que l'extension des descripteurs standards existant peut être appliquée sur des données en temps réel dans les systèmes de la reconnaissance faciale.

Dans le prochain chapitre, nous présenterons une nouvelle technique de reconnaissance des visages et des expressions faciales basée sur une nouvelle extension du descripteur du gradient pour améliorer les performances du système en termes de temps d'exécution et de taux de précision. Ce système inclut également des résultats d'expérimentations sur l'ensemble des bases de données JAFFE et ORL.

Chapitre 5

LGN : Nouvelle descripteur gradient pour la reconnaissance efficace des visages et des expressions faciales

Sommaire

5.1 Introduction

Avec l'utilisation croissante et massive des technologies de l'information et de communication dans la vie quotidienne, le traitement efficace des informations multimédias devient une préoccupation primordiale. Dans l'urgence de traiter efficacement les informations multimédias, la reconnaissance faciale a été combinée à la reconnaissance avancée de formes, offrant de nouvelles solutions pour la reconnaissance des personnes. Ces bénéfices potentiels sont liés à des difficultés tout en garantissant des contrastes suffisants entre les images faciales tels que l'analyse des émotions, des expressions faciales, une interface homme-machine intelligente [165], …, etc.

Ce travail se concentre plus sur la reconnaissance des visages et la reconnaissance émotionnelle faciale. Les principales étapes du système de classification des visages et des émotions faciales sont l'extraction des traits caractéristiques du visage et la classification des traits grâce à une technique d'apprentissage. Dans la littérature, plusieurs techniques d'apprentissage sont utilisées pour les FR et les FER ont été proposés [11]. La plupart des techniques existantes visent à réduire le temps d'exécution des tâches de calcul dans différentes images de visages. Ils fournissent des solutions performantes pour la reconnaissance des expressions faciales et l'analyse des émotions des modèles de caractéristiques de base

dans les systèmes biométriques et la reconnaissance faciale en général [8, 165].Souvent les systèmes utilisent les techniques d'apprentissage standards comme les machines à vecteurs de support (Support Vector Machine, SVM) [7], les réseaux de neurones artificiels [152 – 154, 161] des modèles de Markov cachés (Hidden Markov Model, HMM) [69] et l'analyse en composants principaux (Principal Component Analysis, PCA) [5]. La plupart d'entre eux sont basées sur une technique de cascade améliorée avec des fonctionnalités plus avancées, permettant à construire des modèles économiques plus rentables.

Plusieurs approches existent, certains essayant d'améliorer les performances de leurs techniques de reconnaissance faciale. Cela reste un défi fondamental pour améliorer l'efficacité en termes de taux de précision. D'ailleurs l'utilisation des descripteurs d'extraction des traits caractéristiques du visage comme le modèle binaire local (Local Binary Patterns LBP) [3] et le modèle de gradient local (Local Gabor Gradient LGC) [9] preuve ses limites pour construire un systèmes de détection et de reconnaissance. Le modèle LBP ne considère que les relations entre le pixel central st ses voisins, et le descripteur LGC utilise seulement la relation entre les niveaux de gris des pixels de voisinage, qui sont éloignées du pixel central. Ces descripteurs ne réduisent pas la dimensionnalité des vecteurs des caractéristiques du visage et présentent encore des problèmes de classification moins précis. Cela nécessite de combiner des descripteurs (LBP et LGC) assignées à une image du visage pour construire le modèle de reconnaissance faciale et expressions faciales.

Ce chapitre a pour but de proposer un nouveau descripteur baptisé LGN (Local Gradient Neighborhood) [200] qui combine à la fois les qualités des descripteurs LBP et LGC appliqué à un système de référence connu. Ce descripteur sera alors appliqué sur l'image du visage pour obtenir l'image LGN divisée en sous-blocs. Chaque bloc est mappé sur un vecteur de taille 9. Ensuite, tous les vecteurs de blocs seront concaténés pour obtenir le vecteur de caractéristiques d'image. Enfin, les machines à vecteurs de support (SVM) [7] et les K-plus proches visions (KNN) [8] seront appliquées pour identifier les images d'entrée selon leurs vecteurs de caractéristiques. Cette approche est un peu plus rapide par la prise en considération des informations réduites. Il améliore la précision de la reconnaissance faciale et le temps d'exécution en utilisant une fonction de taille vectorielle réduite.

5.2 Approche de reconnaissance grâce au nouveau descripteur LGN

L'approche proposée est un nouveau descripteur appelé Local Gradient Neighborhood (LGN) pour l'extraction des vecteurs de caractéristiques du visage image et la réduction de la dimensionnalité de ces vecteurs tout en gardant un maximum d'information. L'approche consiste principalement en deux étapes principales : l'étape d'extraction des vecteurs de caractéristiques et l'étape de reconnaissance de visage et leurs expressions faciales. La figure 5.1 montre les différentes étapes de l'approche proposée. L'étape de reconnaissance applique les méthodes SVM et KNN pour identifier les images faciales

d'entrée en fonction des vecteurs de caractéristiques extraits. Notre objectif est d'améliorer un taux de reconnaissance plus précis et un temps d'exécution plus rapide sans affecter le coût.

5.2.1 Extraction des caractéristiques de gradient

L'extraction des traits est l'une des principales étapes du processus de reconnaissance du visage et des expressions faciales (Figure **5.1**). Dans cette étape, nous extrayons des caractéristiques uniques d'une image, ces caractéristiques appelées descripteurs qui servaient à décrire l'image.

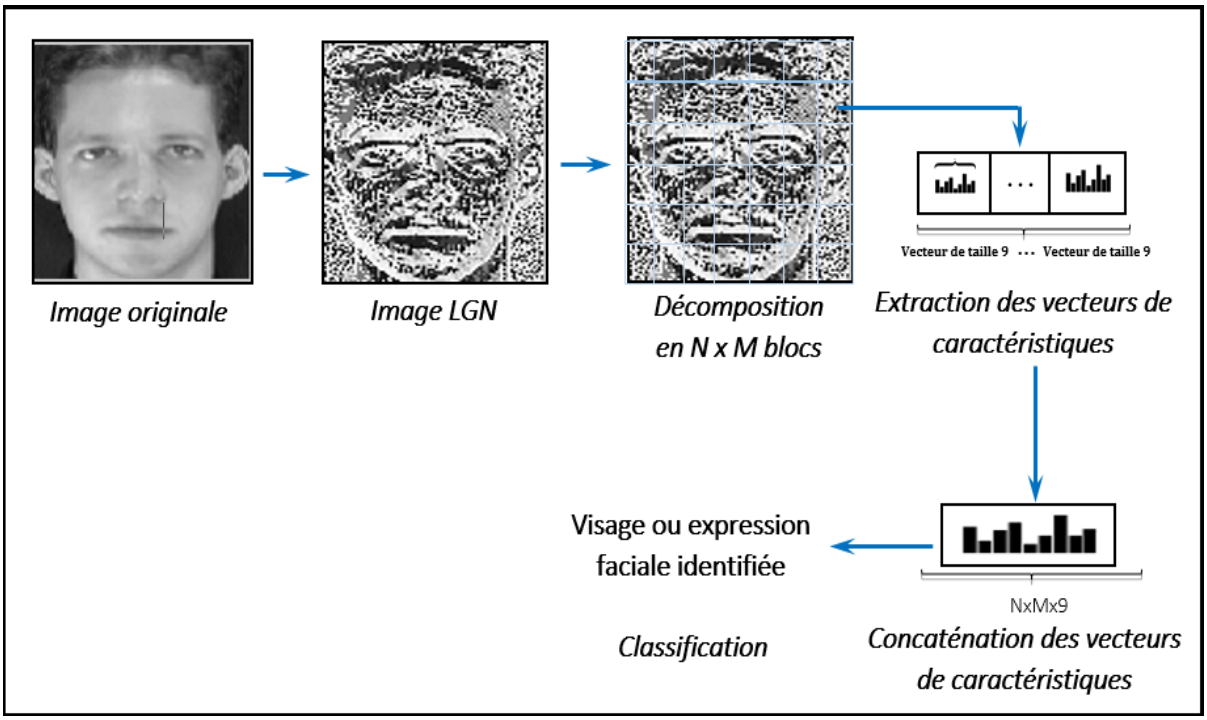

Figure 5.1 : Différentes étapes du système proposé.

Le descripteur LBP (Local Binary Patterns) [3] est l'un des descripteurs les plus connu pour la reconnaissance de visage. Il ne considère que la relation entre les pixels centraux et les pixels voisins, alors que le Local Gradient Code (LGC) [9] utilise la relation des niveaux de gris entre les pixels voisins. Dans ce travail, nous proposons un nouveau descripteur appelé Local Gradient Neighborhood (LGN) qui combine à la fois les qualités LBP et LGC. Il considère à la fois la relation du niveau de gris du pixel central avec ceux du voisinage ainsi que la relation des pixels du voisinage entre eux. L'approche de reconnaissance consiste à diviser l'image en plusieurs blocs. Le descripteur LGN est appliqué sur chaque bloc et l'image est représentée par la concaténation des différents histogrammes locaux. Après l'application de ce processus sur l'image entière, on utilise un algorithme de classification comme SVM (Support Vector Machine), KNN (K Nearest Neighbor) ou autre pour identifier l'image du visage demandée. D'abord, les niveaux de gris des pixels de voisinage, $g_i \; i = 1 \ldots 8$ sont centralisés par rapport à l'intensité du pixel central selon l'équation 5.1

$$g_i' = |g_i - g_c| \qquad i = 1..8 \tag{5.1}$$

La deuxième étape consiste à calculer la différence de niveaux de gris entre tous les deux pixels adjacents. Le nouveau niveau de gris des pixels centraux est calculé selon l'équation 5.2

$$LGN(x,y) = S(g'_1 - g'_2) * 2^7 + S(g'_2 - g'_3) * 2^6 + S(g'_3 - g'_4) * 2^5 + S(g'_4 - g'_5) * 2^4 + S(g'_5 - g'_6) * 2^3 + S(g'_6 - g'_7) * 2^2 + S(g'_7 - g'_8) * 2^1 + S(g'_8 - g'_9) * 2^0 \quad (5.2)$$

Où l'opérateur LGN est défini à l'aide d'une fonction $S(x)$ de variable x elle représente la différence de niveaux de gris entre deux pixels adjacents.

$$S(x) = \begin{cases} 1 & , \quad Si \ x \geq 0 \\ 0 & \quad Sinon \end{cases} \quad (5.3)$$

La figure **5.2** montre un modèle 3x3 pour le descripteur LGN et la figure **5.3** montre l'image originale et les résultats LGN.

g'_1	g'_2	g'_3
g'_8	g_c	g'_4
g'_7	g'_6	g'_5

Figure 5.2 : Modèle 3 * 3 pour descripteur LGN.

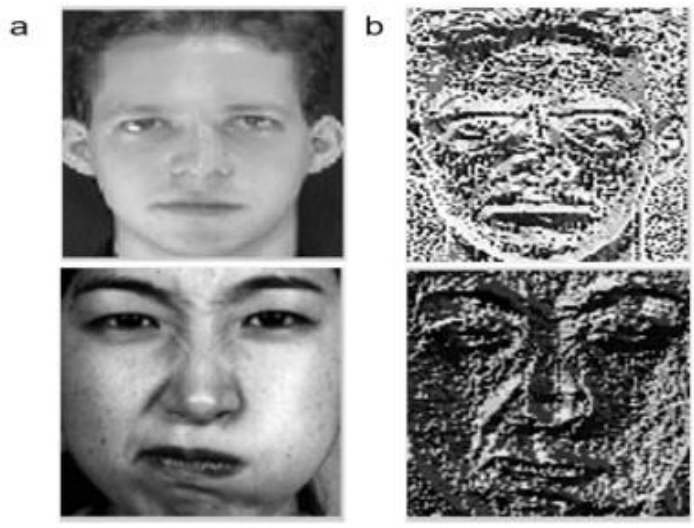

Figure 5.3 :(a) image originale (b) Résultat LGN

5.2.2 Réduction des caractéristiques de gradient

L'opérateur LGN est appliqué sur toute l'image du visage, puis en la subdivisant en blocs$n \times m$. Pour chaque bloc, un histogramme statistique est extrait, réduit et normalisé de la même manière que les

Histogrammes de Gradients Orientés (HOG) [4]. L'utilisation d'un histogramme statistique vise à améliorer le temps d'exécution et à atteindre une précision maximale en réduisant les informations traitées. A partir de chaque bloc, un vecteur de fonction de taille B est extrait de telle sorte que$B << 256$. À cet égard, mappez chaque bloc en intervalles B (bins) (Dahmane et Meunier, 2011). La valeur du $b^{ième}$ intervalle est donnée par :

$$\varphi_b(x,y) = \begin{cases} LGN(x,y) & , \quad Si\ LGN(x,y) \in bin_b \\ 0 & \quad Sinon \end{cases} \tag{5.4}$$

Où $\varphi_b(x,y)$ est calculé sur un bloc i,j et peut être effectué efficacement en utilisant l'image intégrale.Chaque vecteur est normalisé par la norme $L2$ selon l'équation 5.6

$$U = \frac{U}{\sqrt{\|U\|^2}} \tag{5.5}$$

Le vecteur de caractéristiques final est obtenu en concaténant tous les vecteurs d'histogrammes. Il est codé comme un vecteur unique de taille$n \times m \times 9$. Les vecteurs extraits sont classés pour déterminer la classe du visage ou la classe d'expression faciale grâce un algorithme d'apprentissage automatique supervisé. Un vecteur de caractéristiques est comparé aux vecteurs de caractéristiques construits. Nous utilisons deux algorithmes d'apprentissage automatique supervisé KNN et SVM.

5.2.3 Classification des visages et des expressions faciales

5.2.3.1 K plus proches visions (K-NN)

L'algorithme KNN [8] (K plus proches visions)est un algorithme de classification supervisé très simple qui consiste à sélectionner les K voisines les plus proches d'une nouvelle image selon une distance choisie. Cet algorithme est utilisé pour l'estimation statistique et la reconnaissance de formes. Le jeu de données complet est séparé soit dans les échantillons d'entraînement, soit dans les échantillons de test.

Pour plus de détail voir l'algorithme K-NN dans la section **3.3.2.1.**

5.2.3.2 Machine à vecteurs de support (SVM)

L'algorithme SVM [7] (Machine à vecteurs de support) cherche un hyperplan de séparation optimale, qui permet de maximiser la marge entre l'hyperplan et les échantillons d'images les plus proches. Une nouvelle image sera alors projetée et comparée à l'hyperplan des images existantes dans la base de données. Si cet hyperplan est optimal d'une projection existante dans la base, on peut dire que la personne est identifiée, sinon on pourra dire qu'il s'agit d'une nouvelle personne.

Les paramètres sont trouvés en résolvant un problème de programmation quadratique avec des contraintes linéaires d'égalité et d'inégalité. Toutes les données sont ensuite séparées à l'aide de

l'hyperplan. S'il y a trop de valeurs aberrantes à l'aide de l'hyperplan calculé, un nouvel hyperplan est calculé jusqu'à ce que l'hyperplan le plus simple soit formulé.

Algorithme 6 : LGN - SVM

Début

1- Préparer la matrice de motifs.
2- Sélectionnez la fonction du noyau à utiliser.
3- Sélectionner les paramètres de la fonction noyau à utiliser et la valeur de C.
4- Exécuter l'algorithme d'apprentissage pour avoir les α_i .
5- Classifier les données de test en utilisant les α_i et les vecteurs du support

Fin

5.3 Expérimentations

5.3.1 Matériel et mesures d'évaluation

Nous avons évalué la performance de notre approche de classification des visages et des expressions faciales, plusieurs expérimentations ont été réalisées sur un PC utilisant un Intel Core i5 2.67 GHz, 4 Go de RAM, Windows 7 (32 bits) et MATLAB afin d'évaluer deux critères :

1) - la performance de notre approche de reconnaissance des visages et des expressions faciales en termes de temps d'exécution.

2) - son efficacité (rappel, précision, F1-Mesure) par rapport aux standards existants en matière de taille des vecteurs de caractéristique et ses représentations, plus particulièrement LBP [3], LTP [154], GLTP [166], LGC [9].

5.3.2 Les bases de données de test

Deux bases de données de test sont utilisées, la première étant la base de données ORL [181] utilisée pour la reconnaissance faciale tandis que la seconde étant la base de données JAFFEE [182] utilisée pour la reconnaissance des expressions faciales.

La base de données ORL a été créée par les laboratoires AT&T Cambridge (Hoover et al 2000). Le jeu de données ORL contient 400 images de visage de taille 112 × 92.La base de données contient 40 personnes avec 10 vues par personne. Toutes les images sont de même taille et collectées sur un fond sombre. Certaines images sont collectées à des dates différentes avec des variations d'expressions faciales (expression neutre, sourire et yeux fermés) et une occultation partielle par les lunettes. Les poses de tête ont quelques variations de profondeur par rapport à la pose frontale. Cependant, ces variations ne concernent que certaines personnes et ne sont pas aussi systématiques. Chaque image de visage est

divisée en 6 × 8 blocs. La figure **5.4** montre un échantillon d'images de la base de données de visage ORL utilisée dans l'expérimentation.

Figure 5.4 : Exemple d'images faciales JAFFEE utilisées dans l'expérimentation

Le jeu de données de données JAFFEE contient 210 images d'expressions faciales de taille 100 ×100.La base de données JAFFE est constituée de 210 images collectées auprès de 10 femmes exprimées avec 2 à 4 images chaque expression (bonheur, tristesse, surprise, colère, dégoût, peur et neutre). Les résolutions des images en niveaux de gris sont de pixels.

Dans notre travail, on a utilisé la bibliothèque fdlibmex disponible sur Matlab pour la localisation du visage. L'image du visage utilisé a une taille uniforme de100 × 100 pixels. Afin d'améliorer la précision, chaque image est divisée en blocs de taille 11 × 11. La figure **5.5** montre un échantillon d'images utilisées en expérimentation sur la base de données JAFFE.

5.3.3 Description des expérimentations

Nous avons effectué les cinq expérimentations suivantes :

Test#1 : Extraction des vecteurs de caractéristiques par le descripteur LGN

Nous avons utilisé le descripteur LGN afin de représenter chaque image de taille N × M comme un vecteur de caractéristiques de taille N × M × 9.Ce test permis d'obtenir une image LGN segmentée en blocs de taille N × M et chaque bloc est représenté par un vecteur de caractéristiques de taille 9.

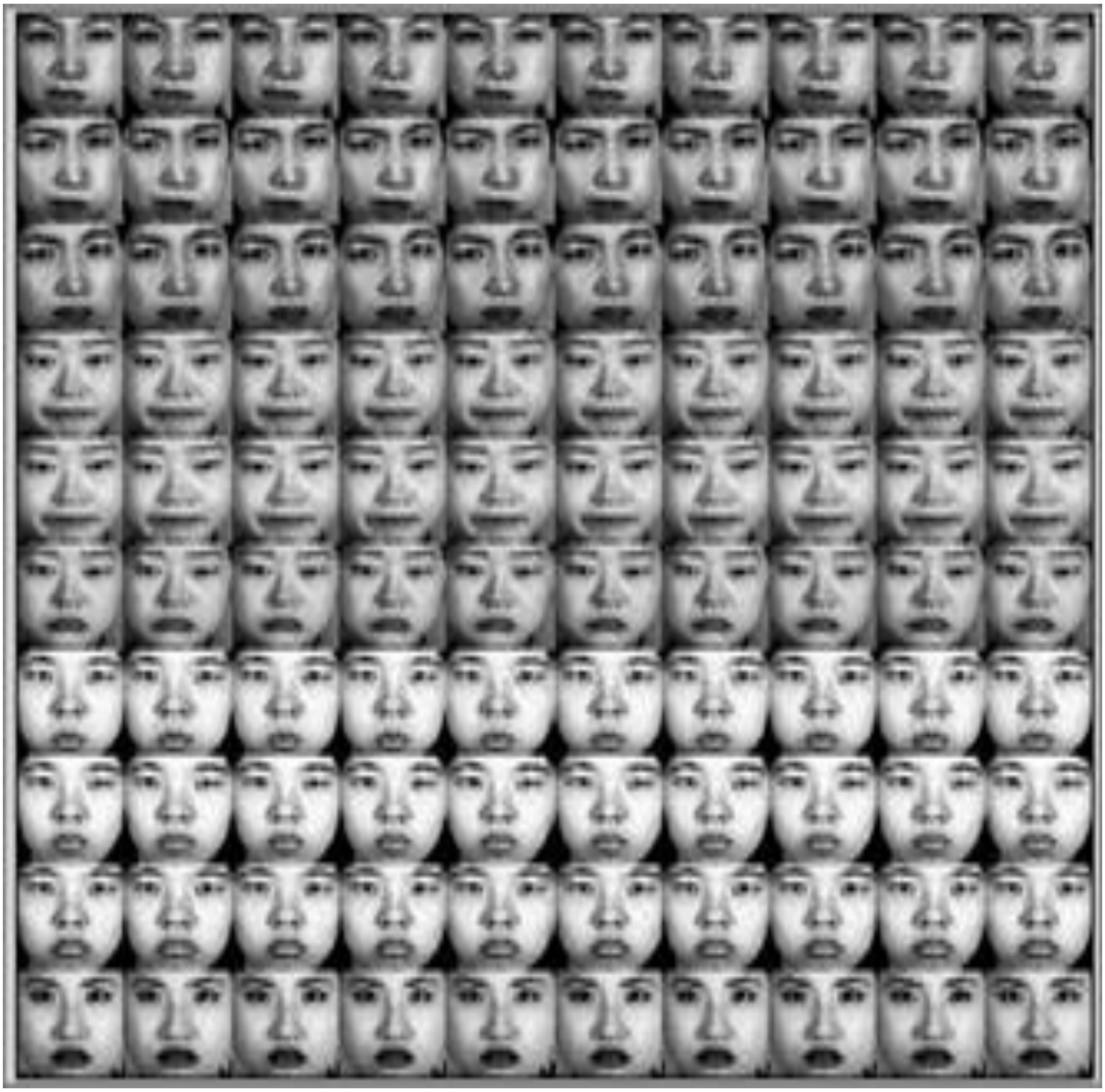

Figure 5.5 : Exemple d'images faciales JAFFEE utilisées dans l'expérimentation

Test#2 : Extraction des vecteurs de caractéristiques par descripteur LBP

Nous avons utilisé le descripteur LGN afin de représenter chaque image de taille N × M comme un vecteur de caractéristiques de taille N × M × 256.Ce test permis de d'obtenir une image LBP segmentée en blocs de taille N × M et chaque bloc est représenté par un vecteur de taille 256.

Test#3 : Extraction des vecteurs de caractéristiques par descripteur LGC

Nous avons utilisé le descripteur LGC afin de représenter chaque image de taille N × M comme un vecteur de caractéristiques de taille N × M × 256.Ce test permis d'obtenir une image LGC segmentée en blocs de taille N × M et chaque bloc est représenté par un vecteur de taille 256.

Test#4 : Extraction des vecteurs de caractéristiques par descripteur LTP

Nous avons utilisé le descripteur LTP afin de représenter chaque image de taille N × M comme un vecteur de caractéristiques de taille N × M × 512.Ce test permis d'obtenir une image LTP segmentée en blocs de taille N × M et chaque bloc est représenté par un vecteur de taille 512.

Test#5 : Application de descripteur GLTP

Nous avons utilisé le descripteur GLTP afin de générer, pour chaque image de taille N × M, un vecteur de caractéristiques taille N × M × 512. Ce test permis d'obtenir une image GLTP segmentée en blocs N × M et chaque bloc est représenté par un vecteur de caractéristiques de taille 512.

Le tableau 5.1 montre l'évaluation des descripteurs existants en termes de taille du vecteur de caractéristiques et de réduction temps d'exécution. La taille d'un vecteur de caractéristiques utilisant le descripteur LBP ou LGC est 11 × 11 × 256 = 30976 et 11 × 11 × 512 = 61952 pour LTP et GLTP, tandis que notre taille de vecteur de caractéristiques 11 × 11 × 9 = 1089. Chaque image faciale LGN est représentée par un vecteur de taille N × M × 9 au lieu d'utiliser une taille de N × M × 256(ou bien N × M × 512) afin de réduire le temps d'exécution jusqu'à 96.50%.

Table 5.1.Évaluation de la concision des annotations dans les différents jeux de tests.

Groupe de tests	NB. Blocs	Taille de vecteur
LBP	N × M	N × M × 256
LGC	N × M	N × M × 256
LTP	N × M	N × M ×512
GLTP	N × M	N × M ×512
LGN	**N × M**	**N × M × 9**

Nous avons utilisé les méthodes de classifications SVM et KNN pour identifier des visages et des expressions faciales. L'approche SVM est appliquée sur des images faciales par paires*« Un-Contre-Un »*avec une fonction de noyau radiale et l'approche standard KNN est appliquée avec k égal à 3. Afin de justifier l'utilisation du descripteur LGN, des comparaisons sont effectuées entre le descripteur LGN et les autres descripteurs standards comme le modèle ternaire local (LTP), le modèle directionnel local (LDP), le modèle binaire local (LBP) et le modèle de gradient local (LGC).

5.4 Résultats d'expérimentations

5.4.1 Evaluation et comparaison de l'efficacité

5.4.1.1 Base de données ORL

Nous avons évalué l'efficacité du descripteur LGN en termes de taux de précision, en considérant deux classificateurs SVM et KNN. Le tableau 5.2 montre les précisions de reconnaissance de chaque classificateur avec une taille réduite du vecteur de caractéristiques pour la base de données ORL. Les résultats montrent que la méthode SVM-LGN donne un taux de reconnaissance égale à 98.50 % bien que le taux de reconnaissance avec KNN-LGN soit toujours bon, 94.50%. Comme on peut observer que le taux le plus élevée pour la reconnaissance faciale est enregistrée pour SVM-LGN. Ceci en raison de la puissance de classification de SVM-LGN et l'aspect discriminatif de LGN. Nous avons également évalué et comparé le taux de reconnaissance du descripteur LGN avec d'autres descripteurs existants LBP, LGC, LTP et GLTP en utilisant les classificateurs SVM et KNN. Le tableau 5.2 montre les taux de précisions de reconnaissance des différents descripteurs dans la base de données de visages ORL.

Tableau 5.2 Evaluation de l'efficacité du descripteur LGN dans la base de données ORL.

Classificateur	Taux de reconnaissance(%)
SVM-LGN	98.50
KNN-LGN	94.50

Le tableau 5.3 montre clairement que le descripteur LGN donne des bonnes précisions de reconnaissance avec les deux méthodes de classification. D'ailleurs, le taux de reconnaissance en utilisant SVM-LGN sur la base de données de visage ORL est meilleur (98.50%), par rapport aux résultats obtenues par la méthode SVM-LTP(90.25%) et la méthode SVM-GLTP (83%).

Tableau 5.3 Comparaison de l'efficacité des différents descripteurs dans la base de données ORL.

Classificateur	Descripteur	Taux de reconnaissance(%)
SVM	LBP	81.50
	LTP	90.25
	GLTP	83.00
	LGC	94.50
	LGN	**98.50**
KNN	LBP	93.75
	LTP	92.50
	GLTP	93.00
	LGC	94.25
	LGN	**94.50**

Mêmes lorsque le classificateur KNN est combiné avec LGN, le taux de précision est légèrement supérieure de 92.50% (KNN-LBP) et 93%(KNN-GLTP). Ces résultats montrent que la méthode KNN-LGN réussit à obtenue des solutions précises et de haute qualité en comparaison avec d'autres descripteurs standards.

5.4.1.2 Base de données JAFFEE

Nous avons également évalué le taux de reconnaissance de la technique proposée en calculant la matrice de confusion d'expression faciale. Le tableau 5.3 montre les taux de reconnaissance de différentes méthodes en utilisant le descripteur LGN sur la base de données des expressions faciales JAFFEE. La matrice de confusion pour toutes les classes d'expressions est un facteur de performance important pour la reconnaissance des expressions faciales, reflétant la capacité du système à classifier correctement les

expressions faciales. En d'autres termes, la valeur entre les différentes expressions signifie une bonne classification de l'approche évaluée.

Tableau 5.4 Evaluation de l'efficacité du descripteur LGN dans la base de données JAFFEE.

Classificateur	Précision (%)
SVM-LGN	84.28
KNN-LGN	72.38

Le tableau 5.4 montre la classification avec sept classes d'expressions faciales en utilisant la correspondance de modèle appliquée avec SVM-LGN sur la base d'expressions faciales JAFFEE. Il montre que la combinaison du descripteur LGN avec SVM donne des taux de reconnaissances meilleures sur toutes les images faciales testées.

Tableau 5.5 Matrice de Confusion de SVM-LGN dans la base de données JAFFEE.

	AN	CO	DI	FE	HA	SA	SU
AN	**96.70**	0.00%	0.00%	0.00%	0.00%	0.00%	3.70 %
CO	3.30%	**75.80%**	9.30%	0.00%	0.00%	6.50%	3.70 %
DI	0.00%	3.40%	**65.60%**	3.20%	6.70%	19.3%	3.70 %
FE	0.00%	0.00%	0.00%	**90.00%**	0.00%	6.40%	3.70 %
HA	0.00%	0.00%	0.00%	3.20%	**90.00%**	0.00%	7.40 %
SA	0.00%	3.40%	6.20%	3.20%	3.30%	**80.60%**	3.70 %
SU	0.00%	0.00%	0.00%	0.00%	0.00%	6.4.0%	**92.50%**

Nous avons également évalué la matrice de confusion de classificateur KNN en utilisant le descripteur LGN sur les images de la base de données JAFFEE. Les résultats obtenus dans le tableau 5.5 montrent que la méthode proposée donne des taux de classification élevés. Ceci en raison des changements légers dans le descripteur LGN sur les images des expressions faciales.

Grâce à ces expérimentations, on constate que le taux d'expression faciale utilisant le classificateur SVM maintient des bons résultats de classification par rapport au classificateur KNN.

Table 5.5. Matrice de Confusion de KNN-LGN dans la base de données JAFFEE.

	AN	CO	DI	FE	HA	SA	SU
AN	**80.00%**	0.00%	0.00%	0.00%	0.00%	12.90	7.40%
CO	10.00%	**75.80%**	9.30%	0.00%	0.00%	3.20%	0.00%
DI	0.00%	3.40%	**71.80%**	0.00%	0.00%	19.30	7.40%

FE	0.00%	0.00%	0.00%	**80.00%**	3.30%	3.20%	14.80%
HA	0.00%	0.00%	3.10%	3.20%	**83.30%**	0.00%	11.10%
SA	0.00%	0.00%	3.10%	3.20%	0.00%	**74.10%**	22.20%
SU	0.00%	0.00%	3.10%	0.00%	13.30%	3.20%	**77.70%**

Le tableau 5.6 montre les résultats de classification des expressions faciales obtenues en combinant du classificateur KNN avec le descripteur LGN dans la base de données d'expression faciale JAFFEE. La méthode KNN combiné avec LGN donne un taux de reconnaissance élevé de 72,38% par rapport à d'autres combinaisons de KNN avec les descripteurs LBP, LGC, LTP et GLTP, et montrant une meilleure performance du descripteur proposé.

Tableau 5.6 Comparaison de l'efficacité des différents descripteurs sur la base de données JAFFEE.

Classificateur	Descripteur	Précision (%)
SVM	LBP	52.85
	LTP	66.19
	GLTP	52.85
	LGC	82.38
	LGN	**84.28**
KNN	LBP	71.42
	LTP	69.04
	GLTP	68.57
	LGC	72.85
	LGN	**72.38**

De plus, le classificateur SVM avec le descripteur proposé atteint un taux de précision intéressant de 84,28% dans la base de données des expressions faciales JAFFEE par rapport aux autres descripteurs LBP, LGC, LTP et GLTP. La précision obtenue avec les descripteurs existants LTP et GLTP est intéressante, puisque ils atteignant en moyenne de 67% de précision. La classification des expressions faciales grâce au descripteur LGN permet d'atteindre un taux de précision plus élevée que les descripteurs LTP et GLTP en temps d'exécution rapide appliquées sur toutes les images de la base de données JAFFEE.

5.4.2 Evaluation et comparaison de l'effectivité

5.4.1.3 Base de données ORL

La figure **5.6**montreles résultats de comparaison entre le temps d'exécution du classificateur KNN combiné avec LGN par rapport aux autres combinaisons existantes LBP, LGC, LTP et GLTP sur la base de données ORL. Les résultats obtenus montrent que le classificateur SVM prend un temps d'exécution

important par rapport au classificateur KNN par tous les classificateurs sur toutes les images de la base de données ORL. Cependant, l'algorithme de classification KNN s'exécute208fois plus rapide que le classificateur SVM pour la base de données ORL. Ces résultats prouvent la supériorité du descripteur LGN combiné avec KNN à classifier une nouvelle image beaucoup plus rapide que le classificateur SVM. Pour le cas des descripteurs LBP ou LGC, la taille du vecteur de caractéristiques est $6 \times 8 \times 256 = 12288$, $6 \times 8 \times 512 = 24576$ pour les descripteurs LTP et GLTP et $6 \times 8 \times 9 = 432$ pour le descripteur LGN.

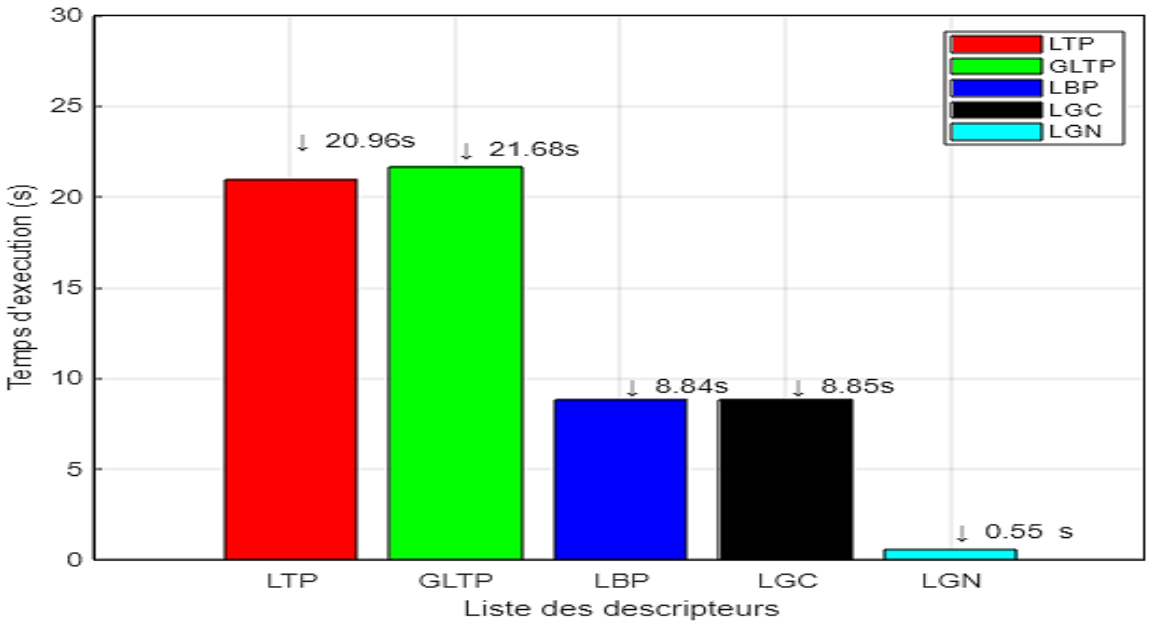

Figure 5.6 : Temps d'exécutions pour chaque descripteur sur la base de données ORL pour le classificateur KNN.

Les résultats montrent que le classificateur SVM combiné avec LGN et le classificateur KNN combiné avec LGN sont plus rapide que les autres descripteurs existants dans toutes les images testées. On note que le temps d'exécution du SVM sur la base de données ORL est de 7.75 secondes par le descripteur proposé par rapport aux autres descripteurs LBP, LGC, LTP et GLTP. En outre, le temps d'exécution obtenu de KNN sur la base de données ORL via le descripteur LGN est de 0.55 secondes, ce qui est plus de 16 fois moins au temps d'exécution de KNN avec LBP et LGC. Cela est dû à la réduction de la taille du vecteur des traits du visage.

5.4.1.4 Base de données JAFFEE

La figure **5.7**montreles résultats de comparaison entre le temps d'exécution de classificateur SVM combiné avec LGN par rapport aux autres descripteurs existants LBP, LGC, LTP et GLTP sur la base de données JAFFEE. Le temps de classification étendu par rapport au classificateur KNN pour toutes les images de la base de données JAFFE. Cependant, le classificateur KNN s'exécute 208 fois plus rapide que

le classificateur SVM sur la base de données JAFFE. Cela signifie qu'après l'entraînement et le test du classificateur KNN, il peut répondre à une nouvelle image beaucoup plus rapide que le classificateur SVM en utilisant la base de données JAFFE.

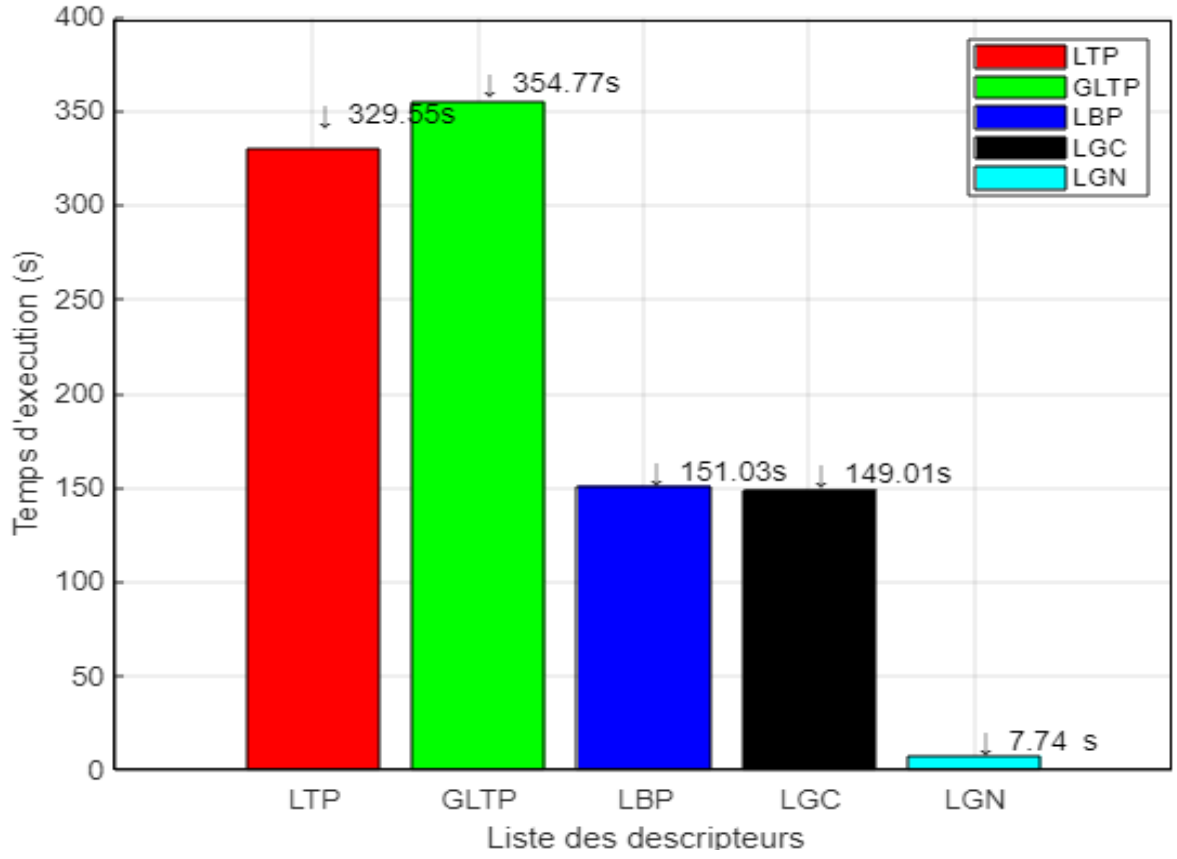

Figure 5.7 : Temps d'exécutions pour chaque descripteur sur la base de données JAFFE pour le classificateur SVM.

Depuis les expérimentations ci-dessus nous pouvons conclure que la méthode proposée est plus efficace par rapport aux autres méthodes existantes et diminue considérablement le temps d'exécution, ce qui rend également applicable dans des systèmes de reconnaissance en temps réel. Néanmoins, elle est moins efficace que les méthodes basées sur du Deep Learning dans le cas où les images sont prises dans des conditions de poses et de luminosité assez différentes.

5.5 Conclusion

Dans ce chapitre, nous avons montré une nouvelle méthode de reconnaissance des expressions faciales grâce à un nouveau descripteur de gradient local (LGN). Ce descripteur bénéficié des avantages des deux descripteurs d'extraction de caractéristiques LBP et LGC. Une approche de reconnaissance faciales et expressions faciales à deux étapes est également proposé pour l'amélioration du processus de reconnaissance en termes de temps d'exécution. La première étape consiste à extraire les vecteurs de

caractéristiques du visage qui respectent les contraintes standards de normalisations. La deuxième étape sert à classifier les vecteurs des visages obtenus à partir de la première phase afin de le stocker dans la base de données. Dans cette étape, on a utilisé les classificateurs SVM et KNN afin de classifier le nouveau visage et la nouvelle expression faciale, ayant le meilleur taux de précision. Pour cela nous avons effectuées plusieurs expérimentations sur deux jeux de données ORL et JAFFEE. Les résultats obtenus sont jugés encourageants et démontrent l'efficacité et l'effectivité des méthodes de classification SVM combiné avec LGN et KNN combiné avec LGN. Les résultats d'expérimentations montrent que KNN peut classer plus rapidement que SVM. Le taux de reconnaissance de SVM est nettement meilleur que KNN sur deux types des jeux de données, puisque la méthode SVM fait généralement plus de vecteurs de support, alors que KNN nécessite peu de vecteurs pour les mêmes ensembles de données.

Dans le chapitre suivant, nous proposons d'étendre le descripteur standard HOG par deux nouveaux descripteurs par l'intégration des attributs d'orientations et d'amplitudes, ce qui rend les résultats plus précis.

Chapitre 6

Reconnaissance efficace des visages et des expressions faciales grâce aux nouveaux descripteurs HDG et HDGG

Sommaire

6.1 Introduction

La reconnaissance faciale devient un domaine de recherche attractive qui rendre la communication entre l'homme et la machine plus naturelle et plus confortable. Ses enjeux technologiques des dimensions scientifique, sociale et économique évoluent rapidement. En particulier, les machines ont permis d'extraire et d'identifier les traits du visage grâce à la reconnaissance faciale (Facial Recognition FR) et à la reconnaissance des expressions faciales (Facial Expression Recognition FER) avec des degrés de précision et compréhension approfondie des expressions faciales. La reconnaissance d'expressions faciales est largement utilisée dans les systèmes actuels d'identification des expressions faciales émotionnelles et des activités physiologiques, comme la fatigue, la douleur, etc. En effet, la reconnaissance faciale représente un excellent moyen de suivre la personnalité en psychopathologie dans

un objectif de vérifier l'identité des personnes afin de mettre en évidence leur message verbal en «face à face» ainsi que les expressions faciales [168]. En plus, fournir davantage de communication «face à face» aux personnes ayant des besoins spécifiques concernant l'utilisation de l'analyse et la modélisation des expressions faciales, par exemple les handicapés dans le secteur privé et le secteur public. Il permet également d'améliorer la sécurité des systèmes, de réduire les coûts d'apprentissage et de renforcer la sécurité des utilisateurs [168].Les systèmes FR et FER se devisent en deux classes, les approches basées sur les caractéristiques géométriques ; le cas de localiser un visage et extraire des caractéristiques de distance (*yeux, nez et bouche*) et les approches basées sur l'apparence (Textures) [160]. Les approches basées sur les caractéristiques géométriques posent le problème de l'extraction incorrect et imprécise de différentes entités avec des résultats de correspondance incorrects [160]. Les approches basées sur l'apparence utilisent diverses méthodes d'extraction de caractéristiques comme les méthodes basées les différences des couleurs, les méthodes d'extraction de caractéristiques de texture, les méthodes d'extraction des directions du gradient, les méthodes d'extraction des traits de Gabor, chacun ayant des avantages de robustesse aux changements environnemental avec moins de complexité de calcul. Cependant, cette approche ne nous offre pas la possibilité d'extraire les caractéristiques les plus idéales des images de visage.

Dans le domaine de reconnaissances faciales, de nombreux systèmes de reconnaissance ont été le résultat de combinaison des technologies existantes. Comment l'utiliser et comment le combiner peut conduire à un nouveau système sophistiqué. Dans ce travail, le compromis entre la précision et le problème de complexité de calcula été abordé en proposant de nouveaux descripteurs HDG (Histogram of Directional Gradient) et HODG (Histogram of Directional Gradient Generalized)[201]. En effet, les principales contributions de notre travail sont :

- définition des deux nouveaux descripteurs appelée HDG (Histogram of Directional Gradient) et HDGG (Histogram of Directional Gradient Generalized), pour extraire des caractéristiques plus discriminantes en utilisant des cartes des amplitudes et des orientations. Ces descripteurs ayant une meilleure capacité de classification et huit valeurs de réponses de bord directionnelles entre les mêmes régions d'images faciales.
- proposition d'une nouvelle réduction de vecteur de caractéristiques de texture de taille 8 pour chaque région de l'image pour la reconnaissance rapide des visages et des expressions faciales. Les descripteurs proposés sont basés sur les cartes d'magnitudes et d'orientations du gradient.

- implémentation du prototype afin de valider les nouveaux descripteurs en termes d'efficacité et de précision des résultats de reconnaissance des expressions faciales sur les mêmes ensembles de données sur la même plateforme.
- amélioration des performances du modèle de généralisation, 11 descripteurs standards parmi les algorithmes les plus connus ont été comparés afin de valider les descripteurs proposés.
- réalisation d'une analyse qualitative et quantitative des performances des descripteurs proposés sur la base de données standard JAFFE [182] et la base de données YALE [183] qui contiennent une variété des images classifiées par le classificateur SVM [7].

6.2 HDG et HDGG : Nouveaux descripteurs pour la reconnaissance efficace et effective des visages et expressions faciales

De nombreux descripteurs pour la reconnaissance d'images de visage ont été déjà proposés dans la littérature et approuvées la réduction des informations contenues dans une image. L'histogramme de gradients orienté (HOG) [4] est un descripteur important et primordial pour la bonne extraction de la texture d'une image faciale et l'augmentation des performances. Il permet d'améliorer la précision de la classification par ses différents codes directionnels. Les cartes d'amplitudes et d'orientations sur les coordonnées horizontales et verticales permettent aussi d'extraire plus efficacement un ensemble de caractéristiques des images de visage. L'histogramme de gradients orienté (HOG) [4] est souvent considéré comme un standard des descripteurs les plus efficaces parmi ceux basés sur l'extraction des paramètres de texture de l'image utilise une transformé de gradient.

Dans cette section, nous allons proposer deux nouveaux descripteurs, HDG (histogramme de gradient directionnel) et HDGG (histogramme de gradient directionnel généralisé) qui présentent de nouvelles techniques d'extraction des contours et des cartes des magnitudes et d'orientations. Ils sont basés sur une philosophie similaire à celle de l'histogramme de gradients orienté (HOG). HDG et HDGG codent les informations directionnelles de la texture du visage de manière réduite pour produire un vecteur des caractéristiques plus discriminant que les méthodes existantes, alors que la classification des visages et des expressions faciales est effectuée grâce au classificateur SVM multi-classes. Le premier descripteur HDG introduit une nouvelle idée originale de représentation et d'extraction de texture en utilisant le concept de direction du gradient, à partir d'une image ou d'une partie d'image (de tout type d'image, pas seulement une image du visage). Le deuxième descripteur HDGG est une combinaison des deux descripteurs : HOG [4] et HDG. L'idée principale de cette hybridation est d'intégrer au descripteur HOG et HDG une relation de longue distance dans les vecteurs de caractéristiques. Les deux descripteurs HDG et HDGG utilisent un vecteur de caractéristiques unique et en le réduisent sa taille afin d'améliorer l'efficacité et l'effectivité des systèmes de reconnaissance faciale et expressions

faciales. Ils fournissent aux utilisateurs des temps de réponse plus rapide, conduisent à des performances de reconnaissance très stables et expriment plus clairement les limites des objets dans une relation de longue distance. Dans la session suivante, nous détaillons les descripteurs proposés puis nous détaillons toutes les étapes de la classification et de la reconnaissance des visages et expressions faciales.

6.2.1 Descripteur HDG

Un histogramme de gradients orienté (HOG) est un descripteur basé sur le calcul de toutes les directions des gradients des pixels de l'image. Cependant, il souffre du coût de calcul élevé [4] en raison de son calcul fréquent et complexe lors de la phase d'extraction de caractéristiques ou de reconnaissance, et par conséquent nous proposons un nouveau descripteur du gradient directionnel basé sur la concaténation de huit valeurs de gradient des pixels de l'image. L'histogramme du gradient directionnel (HDG) est un nouvel descripteur basé sur la direction du gradient. Le descripteur HDG est un opérateur simple que ce soit en termes de conception ou d'implémentation. Les résultats obtenus avec HDG sont très encourageants comme nous le verrons dans la partie expérimentale. Les caractéristiques extraites sont les distributions des gradients de directions de l'image. Les gradients sont généralement des structures micro-motifs de directions différentes extraites de l'image du visage pour décrire nettement les bords de l'objet en tenant compte des informations directionnelles. Ces informations sont codées sur huit indices de direction. Cela permet de distinguer des modèles structurels similaires ayants des transitions de gradient différentes. Nous intégrons également des algorithmes de réduction d'histogramme pour améliorer la performance du système de reconnaissance de visage des expressions faciales.

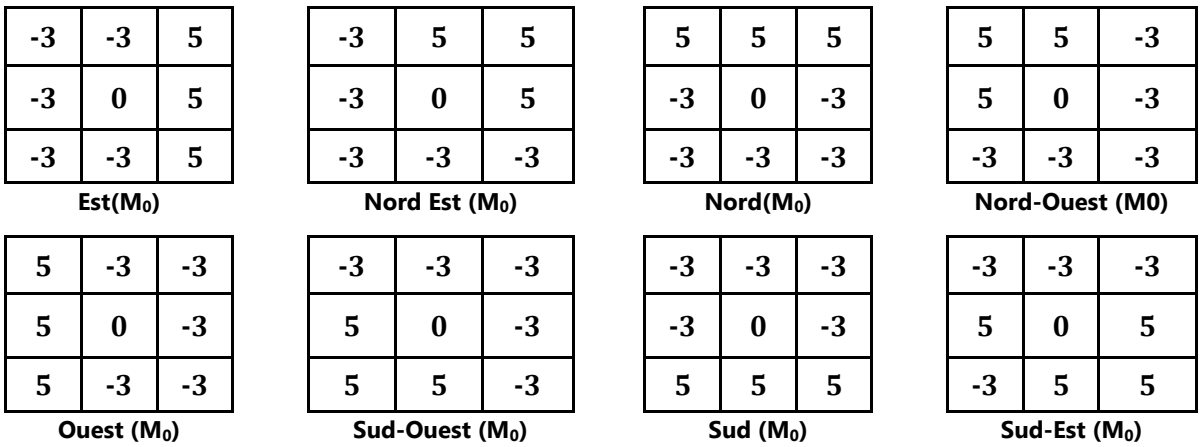

Figure 6.1 Les huit filtres de Kirsch.

Le descripteur HDG consiste à calculer les valeurs de réponse du contour en utilisant les masques de Kirsch, comme le montre la figure **6.1**, dans différentes directions par huit gradients différents pour chaque pixel d'une image. Ensuite, l'image est devisée en grille de blocs $n \times m$. Le descripteur HDG

est appliqué sur chaque bloc, et l'image sera représentée par une concaténation des différents histogrammes. HDG calcule la somme des différentes valeurs de gradient pour chaque direction. La figure **6.2** montre l'image originale et les résultats HDG après l'application de Kirsh. Les valeurs de réponse de contour se réfèrent aux primitives de base appelées texture macro extrait d'un bloc particulier.

Pour chaque pixel (x_c, y_c), l'intensité $I_c$est représentée à l'aide des huit directions de masques de Kirsch $M_i, i = 0, 1, 2, \dots, 7$. Les huit valeurs de réponse de contour sont utilisées pour représenter chaque pixel dans une direction spécifique notée par $m_i, i = 0, 1, 2, \dots, 7$. La figure 6.3 montre qu'un bloc$X$ est représenté par une seule valeur pour chaque direction, cette valeur est définie comme la somme de tous les réponses de ce bloc X pour une direction donnée. Ensuite on concatène les huit valeurs des huit directions pour former un vecteur de 8-valeurs représentant le bloc X.

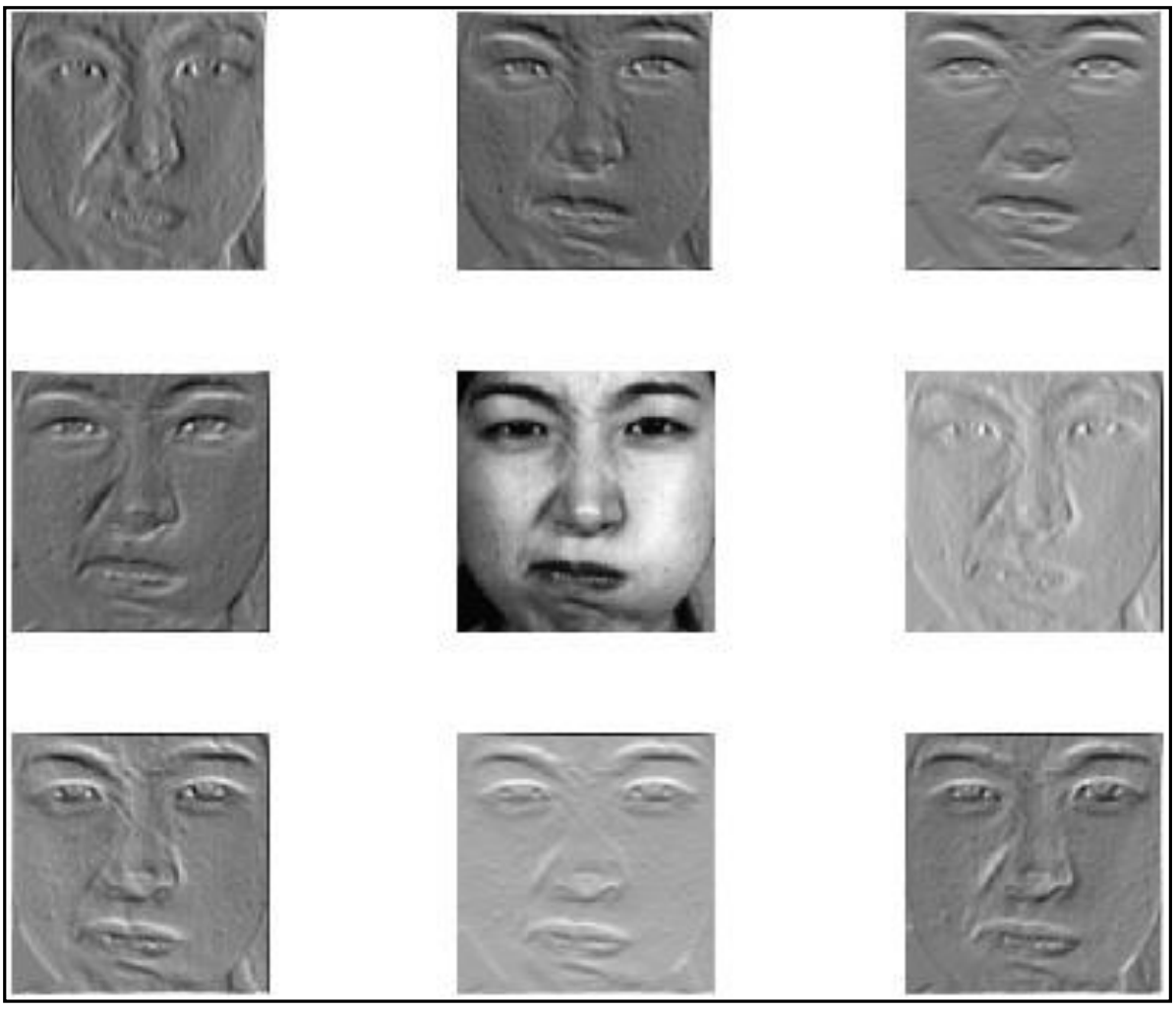

Figure6.2. Exemple d'une image originale et ses images issues des masques du Kirsh

On calcule la valeur de réponse$b_i$de chaque bloc *X* comme suit :

$$b_i = \sum_{m_i \in X} |m_i| \; i = 0,1,..,7 \tag{6.1}$$

Où,

i: une direction,

X : un bloc de l'image (partie d'une image ou l'image entière).

m_i: la valeur de réponse du pixel pour la direction i.

Chaque bloc de l'image est représenté par un histogramme B :

$$B = \{b_i\} i = 0,1,..,7 \tag{6.2}$$

Tous les histogrammes seront concaténés pour former un vecteur de caractéristiques de taille $n \times m \times 8$. Ce vecteur représente la signature du descripteur de visage, comme le montre la figure **6.3**.

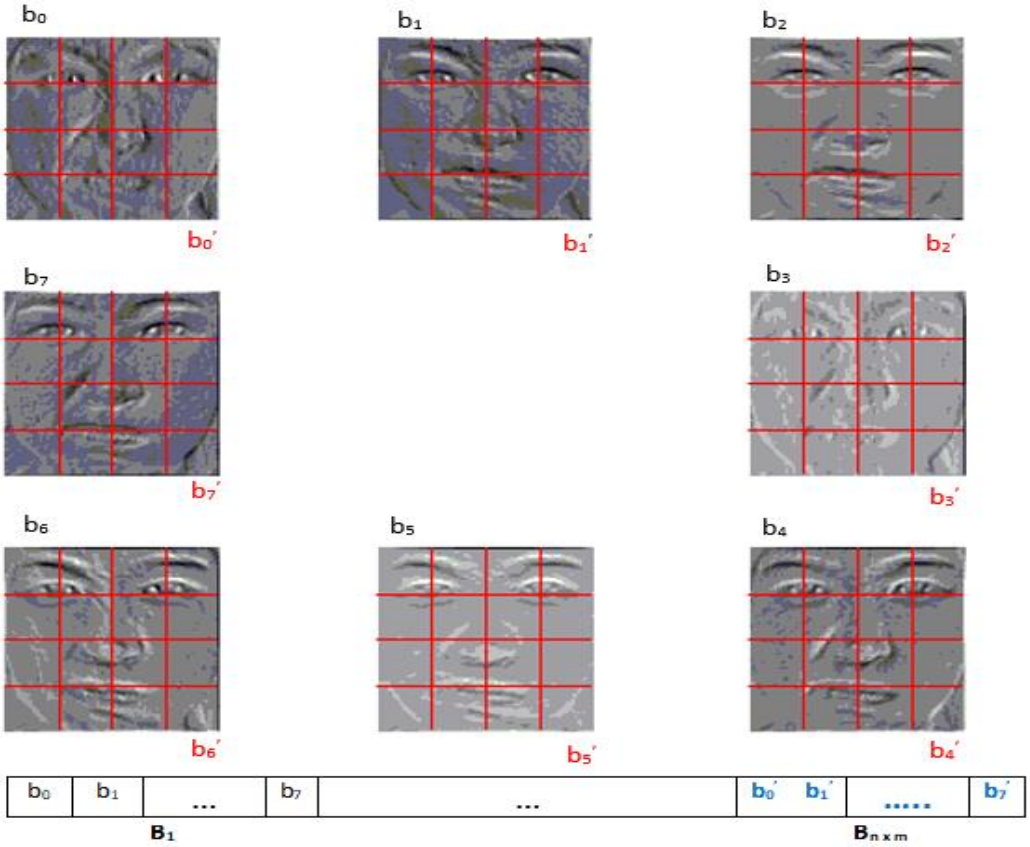

Figure 6.3 Processus de la création du vecteur des caractéristiques par l'opérateur HDG.

6.2.2 Descripteur HDGG

Le descripteur HDGG est une nouvelle extension du descripteur HDG qui utilise les cartes d'amplitudes et d'orientations pour améliorer les performances du système de reconnaissance faciale et de

reconnaissance des expressions faciales. HDGG est définit par la somme de toutes les valeurs de gradient des pixels d'image se référant à 8 directions à l'aide du filtre de Kirsch, qui sera projeté sur des cartes d'amplitudes et d'orientations. Les étapes de ce descripteur sont détaillées ci-dessous.

Etape 1 – Calcul des huit vecteurs directionnels

On suivi les mêmes étapes de filtre de Kirsch de HDG pour obtenir des huit directions des pixels de chaque bloc (voir figure **6.4.a**).

Etape 2 – Calcul de la somme de huit vecteurs directionnels

On calcul la somme des huit vecteurs de gradient pour avoir un seul vecteur représentant chaque pixel, comme le montre la figure 6.4.b, en utilisant l'équation suivante :

$$x_i = \sum_{i=0}^{7}(m_i \cos(i * \pi/4)) \quad (6.3)$$

$$y_i = \sum_{i=0}^{7}(m_i \sin(i * \pi/4)) \quad (6.4)$$

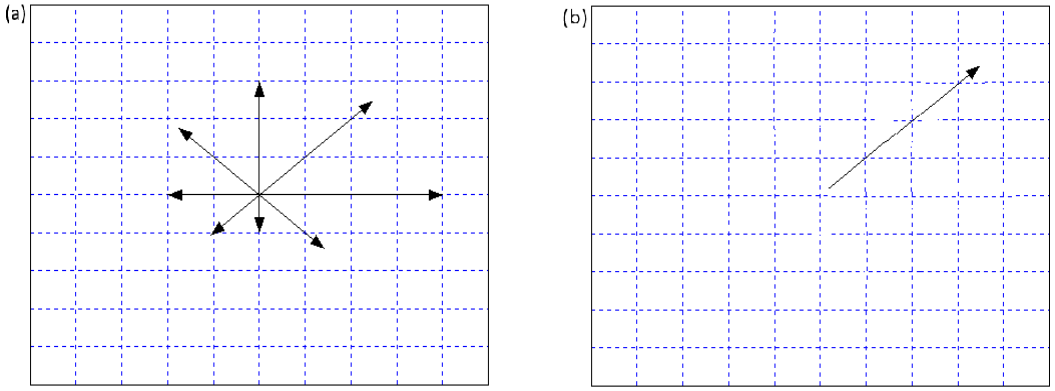

Figure 6.4(a) Un pixel est représenté par huit vecteurs directionnels et **(b)**un pixel est représenté par un seul vecteur qui définit par la somme des huit vecteurs directionnels

Etape 3-Elaboration des cartes d'amplitudes et d'orientations

On calcul les valeurs des cartes de magnitude et d'orientation sur les coordonnées horizontales et verticales de chaque vecteur de pixel en fonction des valeurs G_i et θ_i:

$$G_i = \sqrt{\mathrm{x_i}^2 + \mathrm{y_i}^2} \quad (6.5)$$

$$\theta_i = tan^{-1}(y_i/x_i) \quad (6.6)$$

Etape 4-Décomposition de l'image en blocs

Cette étape consiste à diviser l'image entière en une grille de blocs $n \times m$ sans chevauchementpuis les valeurs d'orientation sont quantifié sur chaque bloc par un histogramme de9 directions appelées bits d'orientations et les valeurs de magnitude sont les votes.

Etape 6 -Extraction et stockage de vecteur de caractéristiques

L'extraction des vecteurs des traits du visage de l'image est obtenue, avec la division de l'image en grille de blocs $n \times m$.Pour chaque bloc, nous utilisons des intervalles également espacés dans l'intervalle[0, π [. Un histogramme local est automatiquement généré et normalisé pour chaque bloc.

Etape 5-Concaténation des blocs d'histogramme

On concatène tous les histogrammes normalisés pour obtenir le vecteur de caractéristique de l'image final. Le vecteur ainsi conçu peut être utilisé comme un vecteur de caractéristique pour évaluer les performances de la méthode de reconnaissance faciale et leurs expressions faciales. La figure **6.5** montre l'image originale, l'image correspondante en magnitudes HDGG et l'image correspondante en orientations HDGG. Ce vecteur de caractéristique obtenu stocké avec l'image de visage dans la base de données des visages.

Figure 6.5. (a) Image originale ; (b) Image en magnitudes HDGG et (c) Image en orientations HDGG.

6.3 Système de reconnaissance grâce au nouveaux descripteur HDG et HDGG

Le système proposé extraire les vecteurs de caractéristiques de l'image du visage et fournit une technique de reconnaissance faciale à gradients directionnels pour identifier les visages dans les images qui garantit une grande précision et une bonne efficacité. Le processus de reconnaissance faciale proposé comprend deux parties principales : l'apprentissage des visages et la classification. La figure **6.6** illustre le système de reconnaissance des visages et des expressions faciales. Ce système trouve et manipule les traits du visage.

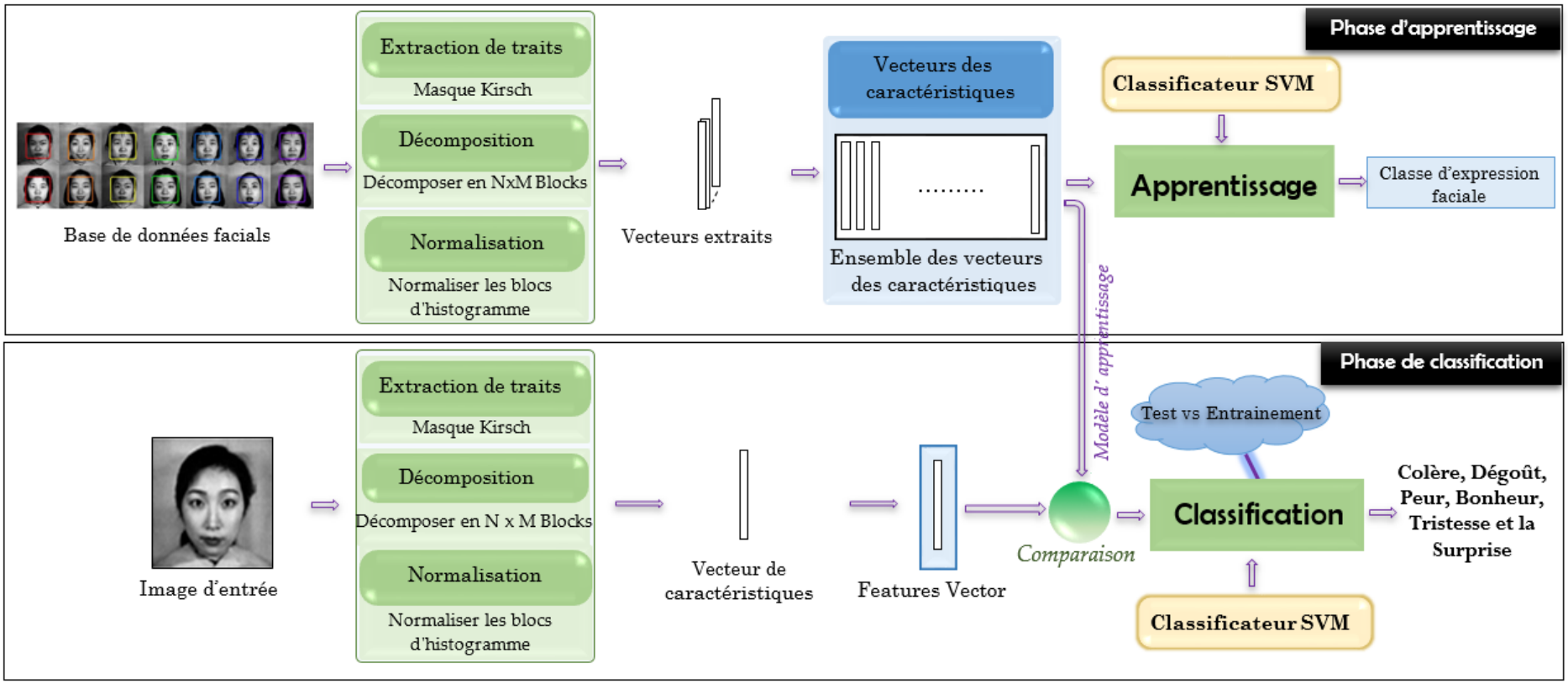

Figure 6.6 Système de reconnaissance faciale proposée.

6.3.1 Phase d'apprentissage

La phase d'apprentissage, c'est la phase au cours de laquelle les traits du visage sont extraits des images de visage à l'aide du descripteur HDG (HDGG). L'algorithme SVM (Support Vector Machine) est l'un des algorithmes de reconnaissance des visages le plus connus et la plus utilisée. On applique l'algorithme SVM avec un noyau multi-classes à toutes les images de la base de données des visages pour l'apprentissage et la classification. SVM prend en entrée un ensemble de données et produit en sortie un ensemble d'entrainements. L'algorithme d'entrainement de visage parcourt les images de visage et calcul le modèle directionnel de gradient en appliquant les descripteurs HDG et HDGG sur l'image entière. Le système divise l'image en grille de blocs $n \times m$ pour réduire la dimensionnalité des vecteurs de caractéristiques et rendre les descripteurs HDG et HDGG plus robustes face aux changements de luminosité et aux bruits. Ensuite, les descripteurs HDG et un HDGG sont appliqués sur chaque bloc. Le bloc est représenté par un histogramme de huit valeurs (9 pour le descripteur HDGG) des directions de gradient et leur magnitude. Enfin, toutes ces valeurs sont normalisées et concaténées dans un vecteur de caractéristiques final, qui est stocké avec l'image du visage dans une base de données.

6.3.2 Phase de classification

Dans cette phase, les vecteurs de caractéristiques des visages et des expressions faciales sont déjà stockés dans la base de données des images du visage. Le système calcul et normalise le vecteur de caractéristiques sur l'image d'entrée comme dans la phase d'apprentissage. Les vecteurs de caractéristiques stockées dans la base de données semblables au vecteur de caractéristiques sont calculés à partir de la distance de similarité. L'image du visage est classée dans la classe avec une distance minimale.

6.4 Expérimentations

6.4.1 Configuration matérielle et outils utilisées

Nos expérimentation sont été réalisées sur un PC Intel Core i5-5005U CPU 2 GHz, 4 Go de RAM et Windows 7 64 bits. Au cours de la phase d'implémentation, les images faciales sont collectées à partir de deux jeux de données constituées de différentes images faciales qui aident à évaluer l'efficacité et l'effectivité des descripteurs HDG et HDGG. Les performances du système sont évaluées et comparées en matière des taux de reconnaissance et du temps d'exécution avec les descripteur d'extraction de caractéristiques existants dans la littérature comme HOG [4], LBP [3], LDP [11], LDN [127], LPQ [10], WLD [125], LGP [169], GDP [170], LGC [9], LTP [126] et GLTP [166]. Tous les descripteurs sont implémentés et évalués sur la même plate-forme avec les mêmes benchmarks JAFFE [182] et YALE [183].

6.4.2 Les bases de données de test

Nous avons utilisé les benchmarks de JAFFE et YALE pour évaluer la performance des descripteurs proposés. Après avoir collecté les images, diverses caractéristiques sont extraites qui sont entraînées et classées à l'aide d'une machine à vecteurs de support multi classe (LIB-SVM) [184], une approche **« un contre un »**appliquée avec la fonction de noyau linéaire. LIB-SVM est une boîte à outils moderne qui contient plusieurs algorithmes d'apprentissage automatique qui aident à écrire des applications sophistiquées de reconnaissance faciale basées sur Matlab. Cette bibliothèque crée des vecteurs de caractéristiques à partir des faces et identifie une face spécifique sur les faces entraînées.

6.4.2.1 La base de données JAFFE

La base de données JAFFE [182] contient 213 expressions faciales (10sujets) et sept classes d'émotions sont considérées. Ce sont le bonheur, la tristesse, la surprise, la colère, le dégoût et la nature. Les images de niveau de gris sont de taille de 256×256. Nous avons utilisé la bibliothèque fdlibmex de Matlab pour la détection de visage. Nous avons également normalisé toutes les images de visage testées avant nos expérimentations à une taille de $128 X 128$ pixels. La figure **6.7** montre des exemples d'images extraits de la base JAFFE utilisées dans l'expérimentation

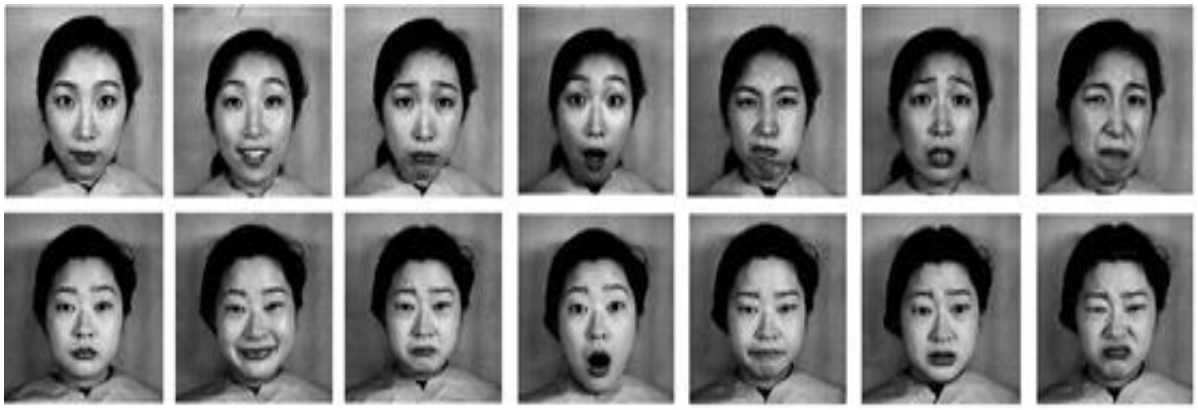

Figure 6.7. Exemples d'images extraites de la base JAFFE utilisées dans l'expérimentation.

6.4.2.2 La base de données YALE

La base des visages YALE [183] contient 165 images en niveaux de gris au format GIF de 15individus (ou 11 sujets). Il y a 11 images par sujet, un par expression ou configuration faciale différente : lumière centrale, pas de lunettes, heureuse, lumière gauche et normale, lumière droite, triste, somnolent, surpris et clin d'œil. Toutes les images sont de taille 64×64 divisées en blocs 8×8. Dans notre travail, nous n'avons considéré que les images frontales. La figure **6.8** montre des exemples d'images extraits de la baseYALE utilisées dans l'expérimentation.

Figure 6.8. Exemple d'images extraites de la baseYALE utilisées en expérimentation.

6.4.3 Paramètres et métriques d'évaluation

Afin d'évaluer l'efficacité et l'effectivité des descripteurs proposés, plusieurs paramètres et mesures ont été adoptés. Nous avons utilisé deux paramètres : la taille de vecteur de caractéristiques. Nous avons évalué l'efficacité et l'efficacité de notre approche de classification automatique du visage et des expressions faciales en termes de :

- **Temps d'exécution (T) :** est le temps nécessaire pour accomplir le processus de classification d'un jeu de données des visages.
- **Précision (P)** : est le rapport entre le nombre visages pertinents (expressions faciales pertinentes) et le nombre des visages (expressions faciales) correctement classées.
- **Rappel (R) : est** le rapport entre le nombre de visages pertinents (expressions faciales pertinentes) et le nombre total des visages pertinents (expressions faciales) attendus.
- **F-mesure :** est le moyenne pondérée de Pet R. C'est un facteur important basé sur le rappel pondéré.
- **Exactitude (A) :** est le rapport le nombre des prévisions correctes et le total des prévisions.

$$Vrai\ positif(VP) = \frac{nb.de\ visages\ identifiés}{nb.de\ visages\ dans\ le\ jeu\ de\ données} \times 100 \quad (6.7)$$

$$Faux\ positif(FP) = \frac{nb.de\ visages\ identifiés}{nb.de\ visages\ identifiés + nb.fause\ identifications} \times 100 \quad (6.8)$$

$$\text{Faux Négatif } (FN) = \frac{nb.de\ visages\ ratés}{nb.de\ visages\ dans\ le\ jeu\ de\ données} \times 100 \quad (6.9)$$

$$Précision(Pre) = \frac{VP}{VP+FP} \quad (6.10)$$

$$R = \frac{TP}{TP+FN} \tag{6.11}$$

$$F - mesure = 2 \times \frac{Précison \times Rappel}{Précison+Rappel} \tag{6.12}$$

$$\text{Exatitude} = \frac{\text{VP}}{\text{VP+VN+FP+FN}} \tag{6.13}$$

Où

- VP : est le nombre de visages pertinents (vrais positifs).
- FP : est le nombre de visages non pertinents (vrais négatifs).
- VN : est le nombre de visages pertinents qui ne sont pas classés par l'approche donnée (vrais négatifs).
- FN : est le nombre de visages non pertinents qui ne sont pas classés par l'approche donnée (faux négatifs).

6.5 Résultats d'expérimentations

Nous avons évaluée le système proposé en utilisant deux descripteurs différents, HDG et HDGG. Les descripteurs HDG et HDGG ont des stratégies différentes concernant l'extraction des vecteurs de caractéristiques des visages et l'identification des visages et la classification des expressions faciales des humains. HDG utilise les directions des gradients. En revanche, HDGG utilise les cartes d'amplitudes et d'orientations. Nous avons évalué et comparé les performances des descripteurs développés en utilisant deux types de critères : *l'efficacité* et *l'effectivité*. L'efficacité est évaluée en termes de temps d'exécution et l'effectivité en termes de l'exactitude, de la précision, de rappel et de F-mesure et. Le temps d'exécution, la précision et le rappel sont des facteurs cruciaux. Le taux de précision et le rappel élevés montrent une bonne performance du système. Notre objectif est de maximiser à la fois la précision et l'exactitude et de minimiser le temps d'exécution. Nous considérons qu'une approche fonctionne bien si elle peut vérifier correctement autant de visages dans un temps d'exécution rapide avec un taux de précision très élevé. Autrement dit, l'approche sera évaluée à l'aide de quatre métriques, l'exactitude, la précision, le rappel et le temps d'exécution. L'efficacité du processus de classification basé sur HDG et HDGG est comparée à tous les descripteurs d'extraction de caractéristiques traditionnels: HOG [4], LBP [3], LDP [11], LDN [127], LPQ [10], WLD [125], LGP [169], GDP [170], LGC [9], LTP [126] et GLTP [166]. Tous ces descripteurs ont été implémentés et intégrés dans notre système. Nous allons présenter quelques évaluations et comparaisons des descripteurs HDG et HDGG avec d'autres descripteurs existants sur les mêmes jeux de données et des mesures de performance considérées pour améliorer l'objectivité et l'analyse de qualité. Ensuite, nous allons également appliquer l'algorithme SVM avec les descripteurs HDG et HDGG pour améliorer le taux de précision et le temps d'exécution. Enfin, nous allons analyser les résultats de classification avec HDG et HDGG et montrer l'impact de la taille de bloc sur les résultats de classification.

6.5.1 Impact de la taille du bloc

6.5.1.1 Impact de la taille du bloc sur le taux de réduction mémoire

Les descripteurs proposés HDG et HDGG sont appliqués à l'ensemble de données JAFFE afin d'évaluer son efficacité. Dans cette expérimentation, nous avons utilisé une image comme image de test et le reste des images sont utilisés comme échantillons d'apprentissage pour le classificateur SVM. Le visage d'entrée est divisé en blocs de taille $n\ x\ m$ puisque un bloc peut donner plus d'informations de localisation. Dans cette expérimentation, le visage de taille 128×128 est également divisé en $8 \times 8 = 64$ blocs. Un vecteur de caractéristiques est extrait de chaque bloc. La concaténation des histogrammes de tous les blocs produit un vecteur de caractéristiques unique pour l'image entière. Le tableau 6.1 montre la taille de vecteur de caractéristiques et la taille de vecteur de caractéristiques pour 64 blocs en utilisant l'opérateur HOD par rapport aux autres descripteurs existantes. Dans les cas de descripteur LTP ou GLTP, la taille du vecteur de caractéristiques est$8 \times 8 \times 512 = 32768$ pour LBP et LGC, $8 \times 8 \times 256 = 16384$ pour LDP, $8 \times 8 \times 56 = 3584$ pour HDG et $8 \times 8 \times 8 = 512$et $8 \times 8 \times 9 = 576$pour HDGG.

Tableau 6.1 : Impact de la taille du vecteur de caractéristiques sur la réduction de la mémoire.

Descripteur	Taille des vecteurs	Taille des vecteurs sur 64 blocs
LTP	512	8*8*512 = 32768
GLTP	512	8*8*512 = 32768
LBP	256	8*8*256 = 16384
LDP	56	8*8*56 = 3584
LDN	56	8*8*56 = 3584
LGC	256	8*8*256 = 16384
LPQ	256	8*8*256 = 16384
WLD	32	8*8*32 = 2048
GDP	8	8*8*8 = 512
LGP	7	8*8*8 = 448
HOG	9	8*8*9 = 576
HDG	**8**	**8*8*8 = 512**
HDGG	**9**	**8*8*9 = 576**

6.5.1.2 Impact de la taille du bloc sur la précision de reconnaissance

Nous avons évaluée l'efficacité des descripteurs HDG et HDGG dans la classification des visages en termes de taux de précision, de taux de rappel, de F-mesure et de temps d'exécution. Le premier test

évalue l'impact de la taille des blocs sur le taux de précision. Les descripteurs HDG et HDGG sont appliqués à différentes tailles: $1 \times 1, 2 \times 2, 4 \times 4, 8 \times 8$ et 16×16. Les résultats sont présentés dans le tableau 6.2. La combinaison des blocs 8×8 est la meilleure pour une image de taille 128×128 pour les descripteurs HDG et HDGG. On peut observer que la petite dimension de bloc ignorait les limites des objets. Par conséquent, une grande dimension de bloc fournira une meilleure précision de reconnaissance.

Tableau 6.2 : Résultats de précision des descripteurs HDG et HDGG pour chaque taille du bloc.

Taille du bloc	Précision en utilisant HDG(%)	Précision en utilisant HDGG(%)
1 x 1	38.57	39.52
2 x 2	53.33	66.67
4 x 4	73.80	82.38
8 x 8	**90.00**	**91.43**
16 x 16	89.04	88.57

6.5.2 Évaluations de l'efficacité

6.5.2.1 Évaluation et comparaison de l'efficacité sur la base de données JAFFE

Dans cette section on démontre que les descripteurs proposés sont efficaces. Pour cela nous comparons les résultats d'efficacité des descripteurs proposés HDG et HDGG à certains descripteurs standards en termes de précision, rappel et F-mesure. Le tableau 5.3 montre la comparaison des taux de précision des descripteurs proposés avec d'autres descripteurs existants en utilisant la base de données JAFFE où le descripteur HDG dépasse 90% des images identifiées et le descripteur HDGG dépasse 91% des images identifiées. Il est clair que les descripteurs proposés offrent des meilleurs taux de précision sur la base de données JAFFE.

Le tableau 6.3 montre les taux de reconnaissance des différents descripteurs sur la base de données JAFFE. Comme le montre le tableau 6.3, les descripteurs proposés HDG et le HDGG reconnaissent avec succès les visages extraits à 91.21% grâce à l'efficacité de la méthode de classification par bloc. Ces résultats valident l'efficacité des descripteurs facials proposés. Les taux de reconnaissance plus élevés pour les descripteurs HDG et le HDGG ont été enregistrés par rapport aux autres méthodes sur la base de données YALE voir le tableau 6.4.

Tableau 6.3 : Résultats de précision pour chaque descripteur dans la base de données JAFFE.

Descripteur	Taux (%)
LTP	40.95
GLTP	38.09
LBP	42.00
LDP	52.38
LDN	68.57
LGC	78.57
LPQ	69.52
WLD	81.90
GDP	87.14
LGP	83.80
HOG	88.57
HDG	**90.00**
HDGG	**91.43**

Tableau 6.4 : Résultats de précision pour chaque descripteur dans la base de données YALE.

Descripteur	Taux (%)
LBP	87.88
LDP	56.36
LDN	79.39
LGC	70.00
LPQ	73.93
WLD	92.12
GDP	90.91
LGP	90.91
HOG	90.30
LTP	48.48
GLTP	36.36
HDG	**92.12**
HDGG	**92.12**

La matrice de confusion est un résumé des résultats de prédiction du problème de classification. La classe avec une valeur de confusion la plus élevée domine les autres classes. Les résultats des tableaux 6.5 et 6.6 montrent respectivement la matrice de confusion des expressions de sept classes, et la précision utilisant la correspondance de modèle appliquée à SVM sur une base de données JAFFE en utilisant les opérateurs HDG et HDGG. Il est clair que les opérateurs proposés HDG ou HDGG offrent la meilleure précision

sur toutes les images testées. De plus, les résultats des tableaux 6.5 et 6.6 démontrent l'efficacité des opérateurs HDG et du HDGG pour l'identification efficace des expressions faciales.

Tableau 6.5. Matrice de confusion de HDG-SVM dans la base de données JAFFE.

	AN	CO	DI	FE	HA	SA	SU
AN	**100%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00 %
CO	0.00%	**89.65%**	0.00%	3.22%	0.00%	6.45%	0.00 %
DI	0.00%	0.00%	**84.38%**	0.00%	3.33%	9.68%	3.70 %
FE	0.00%	0.00%	0.00%	**80.65%**	0.00%	16.12%	3.70 %
HA	0.00%	0.00%	3.13%	6.45%	**90.00%**	0.00%	0.00 %
SA	0.00%	0.00%	6.25%	3.20%	0.00%	**87.09%**	0.00 %
SU	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**100%**

Les taux de précision des descripteurs HDG et HDGG sont comparés avec ceux des autres descripteurs existants. Comme on peut le voir sur la figure **6.9**, les résultats obtenues montrent que les descripteurs proposés HDG et HDGG atteignent une précision plus élevée par rapport aux autres descripteurs traditionnels qui sont respectivement de 90.00% et 91.43% dans la base de données JAFEE. La figure **6.9**présente les tests de comparaison du F-mesure entre les descripteurs HDG et HDGG avec les autres descripteurs sur la base de données JAFFE. Le descripteur HDGG est évidemment donne un taux de F-mesure égale à91.84% et le descripteur HDG donne un taux de F-mesure égale à90.54%.

Tableau 6.6 : Matrice de confusion de HDGG -SVM dans la base de données JAFFE.

	AN	CO	DI	FE	HA	SA	SU
AN	**96.66%**	3.45%	0.00%	0.00%	0.00%	0.00%	0.00 %
CO	0.00%	**93.10%**	3.25%	0.00%	0.00%	3.23%	0.00 %
DI	0.00%	3.45%	**81.25%**	0.00%	3.33%	9.68%	3.70 %
FE	0.00%	0.00%	0.00%	**87.09%**	0.00%	9.68%	3.70 %
HA	0.00%	0.00%	6.25%	3.22%	**86.67%**	3.23%	0.00 %
SA	0.00%	0.00%	0.00%	3.22%	0.00%	**96.77%**	0.00 %
SU	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**100%**

A partir des expérimentations ci-dessous, les deux descripteurs de reconnaissance faciale proposés atteignent un bon taux de F-mesure (91.84%), une bonne précision (92.12%) et rappel (92.67%), montrant que le descripteur HDGG combiné avec SVM peut identifier efficacement un visage donné.

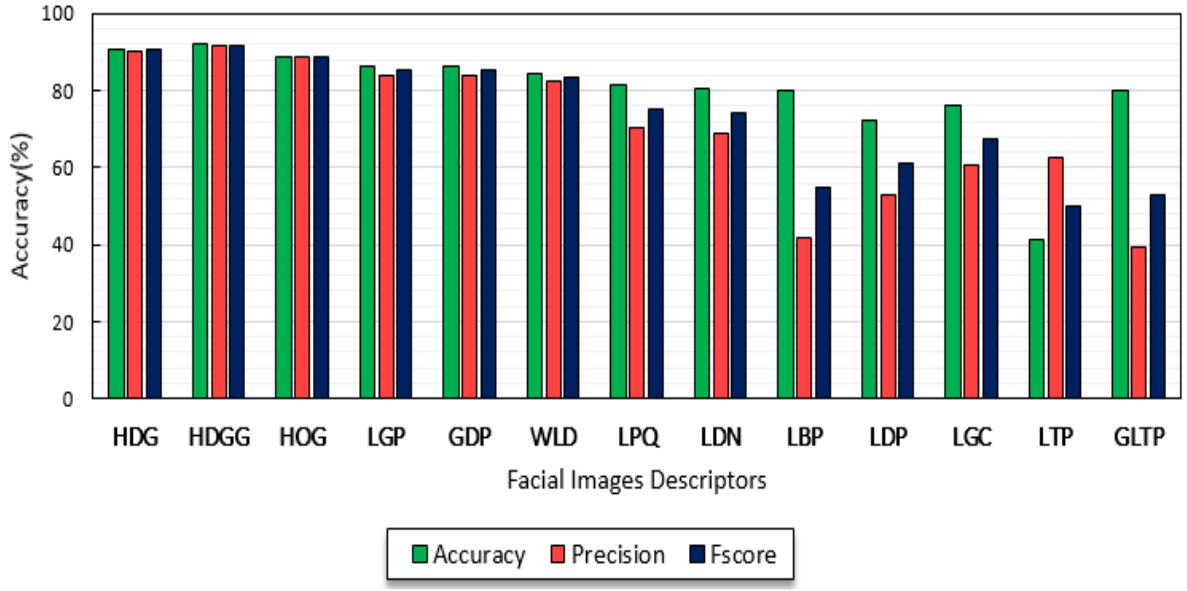

Figure 6.9. Exactitude, précision et F-score pour chaque descripteur dans la base de données JAFFE.

6.5.2.2 Évaluation et comparaison de l'efficacité sur la base de données YALE

Nous avons évalué les mesures de taux de précision, de taux de rappel et de F-mesure avec une variété d'images en utilisant une palette de descripteurs d'extraction de caractéristiques faciales (HDG, HDGG, HOG, LGP, GDP, WLD, LPQ, LDN, LBP, LGC, LTP et GLTP). La figure **6.10** montre une comparaison de taux de précision, de taux de rappel et de F-mesure. Comme on peut le voir que les taux de précision des descripteurs HDG et HDGG sont très élevés, ce qui donne de meilleures performances des descripteurs de reconnaissance faciale. Le taux de précision du descripteur HDG sur la base de données des visages YALE est de 92.12% et le taux de précision du descripteur HDGG sur la base de données YALE est de 92.12%. On peut également noter que les taux de rappel des descripteurs proposés sont nettement meilleurs que les autres descripteurs existants. En outre, le rappel du descripteur HDGG sur les images des expressions faciales YALE par rapport à d'autres descripteurs dépasse 92.16%. Le taux de rappel est faible dans le descripteur GLTP, ce qui explique le grand impact des effets de changement de luminosité sur les images YALE, où le contraste entre l'image YALE et l'image JAFEE devient plus élevée. Cela preuve la bonne efficacité des descripteurs d'extraction de caractéristiques proposés. Comme on peut le voir sur la figure **6.10**, les résultats obtenus de F-mesure par les descripteurs HDG et HDGG sont bien meilleurs que d'autres descripteurs existants. Dans la base de données YALE, SVM-HDG donne un meilleur taux de F-mesure 91.25% tandis que SVM-HDGG peut atteindre un taux de F-mesure92.67%. A partir des expérimentations ci-dessous, on peut conclure que les deux descripteurs de reconnaissance faciale proposés atteignent un bon taux de F-mesure (92.67%), une bonne précision (92.12%) et un rappel (92.67%), montrant que les descripteurs proposés identifient efficacement un grand nombre des images et récupèrent avec succès les caractéristiques de visage.

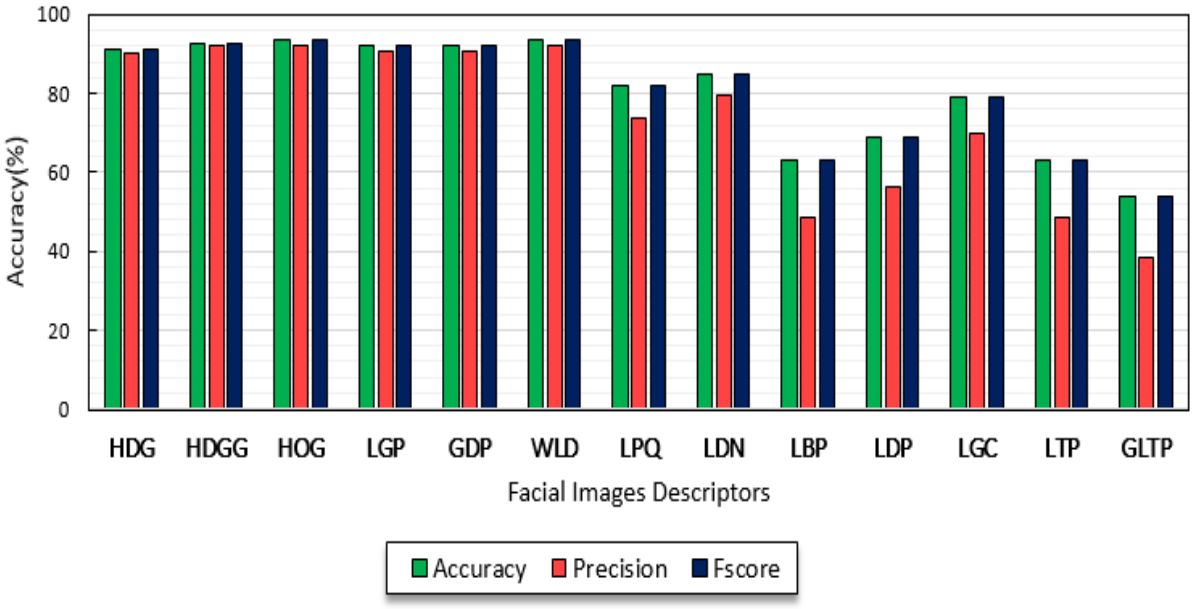

Figure 6.10 Exactitude, précision et F-score pour chaque descripteur dans la base de données YALE

6.5.3 Comparaison des temps d'exécutions

6.5.3.1 Comparaison des temps d'exécutions sur la base de données JAFFE

En se basant sur les descripteurs proposés, les vecteurs de caractéristiques des visages extraites sont classifiés en utilisant la méthode SVM. Il classe avec succès les caractéristiques des visages dans un temps optimal. La figure **6.11** montre la comparaison des temps d'exécution des différents descripteurs. On peut constater que les descripteurs HDG et HDGG classifient efficacement les caractéristiques faciales extraites, avec un temps d'exécution optimal par rapport aux autres descripteurs existants. On remarque aussi que le temps d'exécution des descripteurs HDG et HDGG est respectivement de 0.526s et 0.4067s. Les descripteurs HDG et HDGG ont une taille de vecteur plus petit que les autres descripteurs existants dans la littérature, ce qui peut réduire le temps de classification et le temps d'apprentissage.

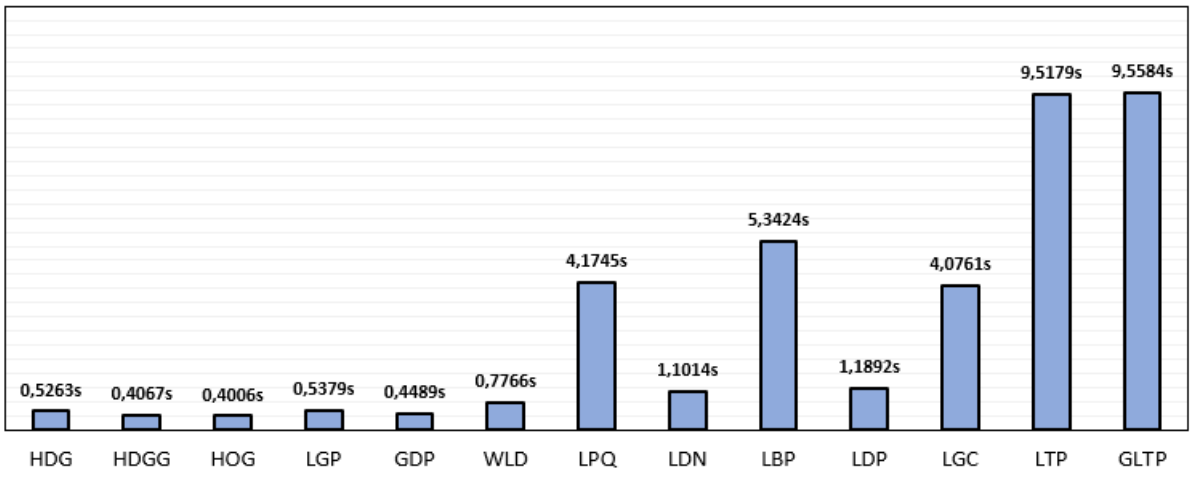

Figure 6.11. Temps d'exécution pour chaque descripteur appliqué sur la base de données JAFFE.

6.5.3.2 Comparaison du temps d'exécution sur la base de données YALE

Nous avons évalué le temps d'exécution de notre système sur la base de données YALE. La figure **6.12** montre la variation du temps d'exécution en fonction du descripteur d'extraction de caractéristiques appliqué. Il faut observer que le temps de réponse diminue significativement en fonction des différentes images faciales testées. En tant que résultats du temps d'exécution dans les cas d'images de visage YALE, la figure **6.12**garantit que nos propositions permettent une exécution rapide dans le cas d'images de visage YALE. En comparant les différents descripteurs de reconnaissance faciale, HDG et HDGG auront un temps d'exécution plus rapide par rapport aux descripteurs LBP et HOD.

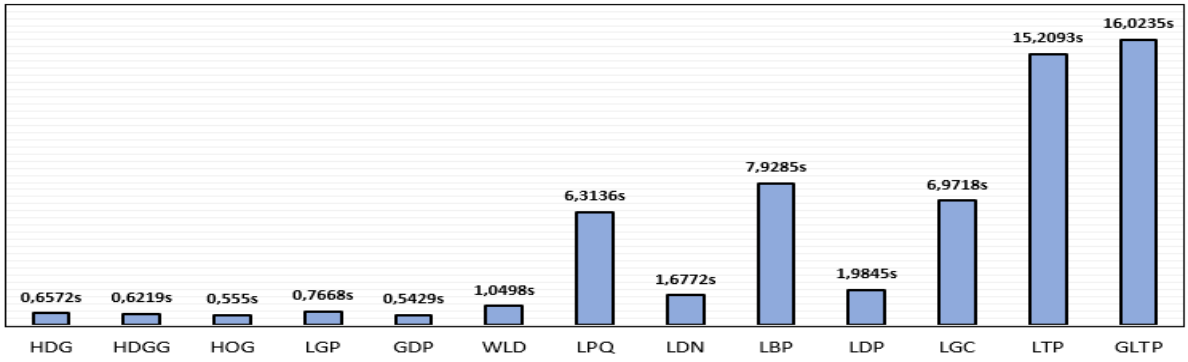

Figure 6.12. Temps d'exécution pour chaque descripteur appliqué sur la base de données YALE.

6.6 Conclusion

Ce chapitre est consacré aux deux nouveaux descripteurs d'extraction de caractéristiques de texture basés sur les directions de gradient HDG et HDGG. Nous avons commencé par décrire les descripteurs HDG et HDGG. Ensuite nous avons présenté un système de reconnaissance des visages et des expressions faciales basées sur les valeurs de réponses des orientations des gradients. En outre, nous avons positionné nos descripteurs par rapport aux descripteurs existants en comparant nos résultats avec ceux des autres descripteurs en appuyant sur quelques paramètres et métriques d'évaluations. En effet, nous avons comparé le taux de précision et le temps d'exécution des descripteurs sur deux benchmarks en termes de *taille de vecteur de caractéristiques* et de t*aux d'erreur*. Lorsque la taille du vecteur de caractéristiques ne dépassait pas 512, le taux de reconnaissance atteignant 92.12% et le taux d'erreur variait de 0.08 à 0.1. Nous avons mené de nombreuses évaluations empiriques sur plusieurs ensembles de données en utilisant les descripteurs HDG et HDGG. Nous avons commencé par évaluer l'impact de la taille du bloc sur la précision du système. Plus la taille du bloc est élevée, meilleure est la précision. Ensuite, nous avons appliqué également les descripteurs HDG et HDGG pour trouver la meilleure expression faciale dans l'ensemble de données JAFFE et la reconnaissance faciale dans l'ensemble de données YALE. Enfin, nous avons obtenir un classificateur de visage avec une précision de 92.12%.Nos travaux futurs se concentreront sur l'apprentissage automatique avancé utilisant Open CV et Face DL.

Conclusion générale

Avec l'explosion de la quantité d'information et l'évolution rapide des technologies, le besoin de sécurité des informations augmente continuellement dans les sociétés modernes. Le système biométrique est l'un des outils les plus efficaces et fiables pour satisfaire ce besoin. En effet, il permet une reconnaissance automatique de l'identité basée sur une analyse biologique de l'individu. La biométrie regroupe deux aspects principaux, l'identification et la vérification. Dans le cas d'identification, le dispositif biométrique requit une information biométrique et la compare avec toute les informations stockée dans la base de données, il s'agit d'une comparaison « un à plusieurs ». Alors que, pour le cas de vérification, l'utilisateur saisit son identité à travers une information biométrique, ensuite le système fait la comparaison entre les caractéristiques extraites à partir de l'information entrée, avec celle qui correspond à l'identité prétendue, il s'agit d'une comparaison « un à un ». Plusieurs modalités importantes des caractéristiques sont citées et étudiées dans la littérature. Nous avons abordé dans cette thèse l'une des modalités comportementale, en particulier le *visage*, puisqu'il permet une identification efficace, avec une grande richesse d'information sur le sexe, l'âge, l'état émotionnel, la race,…etc.

Le système de reconnaissance de visage et des expressions faciale est une tâche d'étude populaire dans le domaine du traitement d'image et de la vision par ordinateur. Cette popularité est augmentée en raison de son application potentiellement énorme ainsi que de sa valeur théorique. Ces systèmes sont déployés dans les différents domaines d'applications telles que la sécurité, la surveillance, la sécurité intérieure, le contrôle d'accès, la recherche d'images, l'interaction homme-machine et le divertissement. Cependant, ces applications sont confrontées à divers défis tels que les conditions d'éclairage et les expressions faciales. Pour relever ces défis, les industries adoptent un système de reconnaissance faciale comme un outil essentiel d'identification des individus en fournissant des informations faciales discriminatoires.

Au cours de la dernière décennie, les systèmes de reconnaissance faciale et expressions faciales sont devenues de plus en plus des méthodes d'authentification biométriques importantes. En effet, de nombreuses techniques ont été utilisées pour la reconnaissance faciale et expressions faciales basées sur les caractéristiques faciales. On distingue généralement trois types d'approches, selon le type de caractéristiques souhaitées :

1. **Approche locale :** approche basée sur les caractéristiques faciales locales telles que les yeux, la bouche et le nez. Les vecteurs de caractéristique du visage peuvent être représentés par des points de repère pour l'étape de reconnaissance.
2. **Approche holistique :** l'approche stochastique utilise des caractéristiques de visage global comme entrée à l'algorithme de reconnaissance.
3. **Approche hybride :** qui combine les approches locales et holistiques.

Le travail présenté dans cette thèse consiste à proposer un nouveau système de reconnaissance de visage et leurs expressions faciales qui donnent de meilleurs taux de précision tout en réduisant le temps d'exécution. Pour cela, nous avons proposé plusieurs descripteurs basés texture autour divers aspects de l'utilisation des visages et des expressions faciales comme identifiant biométrique. Les descripteurs de texture locaux décrient d'une façon optimale la combinaison de caractéristiques globales et locales des images de données biométriques bidimensionnelles, et ils permettent de simuler parfaitement le principe du système neurologique humain dans la perception et la reconnaissance des objets et des visages.

1. Intérêts du travail

Les intérêts du travail dans cette thèse sont résumés comme suit :

- Développement de nouvelles techniques de reconnaissance de visage et expressions faciales qui offrent un temps d'exécution rapide pour l'identification des personnes tout en assurant une précision élevée.
- Soutenir les applications temps réel afin de mettre en œuvre un système efficace de combinaison et d'utilisation des nouveaux descripteurs de manière expérimentale pour trouver les classificateurs les mieux adaptés dans le cas des FR et FER.

2. Bilan de la thèse

Nous avons divisé la thèse en deux parties, la première était consacrée à l'état de l'art lié à notre recherche et la deuxième exposé nos contributions. Dans la première partie, le chapitre 1donne une vue détaillé sur les systèmes de reconnaissance des visages et expressions faciales. Dans le chapitre 2, on a présenté la notion de la texture et les descripteurs de la texture qui représentent souvent le noyau d'un système de reconnaissance FR et FER. Dans le chapitre 3, on a présenté les outils d'apprentissage automatiques les plus utilisés dans le domaine FR et FER. Dans la deuxième partie on a focalisé sur nos contributions.

Dans le chapitre 4, on a proposé une étude comparative des sept classificateurs les plus utilisés et les plus populaires dans le domaine de reconnaissance de la forme d'une façon générale, et plus particulièrement dans les systèmes FR et FER. Cette étude comparative repose sur l'extraction de

connaissance à partir des visages et que les informations extraites sont simples et abstraites. Le vecteur de traits qui représente l'information entracte du visage n'est que les positions, spécifiées par deux coordonnées qui simulent le squelette du visage. Les résultats d'expérimentations montrent que le classificateur d'analyse discriminant quadratique (QDA) dépasse considérablement tous les autres classificateurs tels que le MLP, SVM, KNN, RF avec tous les types de tests différents qu'on a effectués. Sachant que dans la littérature les classificateurs SVM, MLP et K-NN sont les plus répondus et approprié. Après cette étude comparative, pour sélectionner un meilleur classificateur, nous avons constaté que le bon choix du classificateur dépend fonctionnellement des types de données extraites lors de l'étape d'extraction des caractéristiques.

Dans le chapitre 5, nous avons proposé un nouveau descripteur de texture appelé LGN(Local Gradient Neighborhood). Ce descripteur est le résultat d'une combinaison des qualités des deux descripteurs standards LBP et LGC. Il est comparé avec LBP, LGC, LTP et GLTP selon différents axes de recherches FR et FER en utilisant les classificateurs standard SVM et KNN, appliqués sur deux bases de données populaires ORL et JAFFE. Les résultats d'expérimentations ont montré que LGN dépasse légèrement les autres opérateurs. Généralement, il est difficile d'avoir un descripteur qui donne toujours le meilleur résultat par rapport aux autres. En effet, il y a toujours un domaine particulier dans lequel un descripteur est meilleur que d'autres descripteurs. En d'autres termes, les descripteurs comme d'ailleurs les classificateurs ont des spécificités d'applications les uns par rapport aux autres.

Dans le chapitre 6, nous avons focalisé sur une notion bien spécifique de la texture, c'est la notion de l'orientation ou la direction du gradient. Cette notion simple dans sa conception peut donner une richesse d'information et faire la différence lors de la classification et la reconnaissance des objets. Sur cette base, deux nouveaux descripteurs HDG et HDGG ont été proposés pour exploiter la direction du gradient d'une façon optimale et simple pour viser les différentes applications du FR et FER. Ils visent à minimiser le temps d'exécution et à maximiser de taux de reconnaissance. Dans ce travail on a comparé nos descripteurs avec dix autres descripteurs grâce à SVM dans deux axes de recherches différents FR et FER. Les résultats sont très encourageants, mais il reste beaucoup de travail à faire.

3. Perspectives

Le travail présenté dans cette thèse a abordé diverses techniques de reconnaissance faciales et expressions faciales concernant la réduction du temps du calcul et l'augmentation de taux de précision. A cet effet, quelques perspectives qui semblent pertinentes dans le futur à court et à long terme afin d'améliorer ses performances. Ces perspectives sont :

- **Vers une reconnaissance des expressions faciales dans des conditions réelles**

  Il serait intéressant de tester les approches proposées dans des conditions d'éclairage dégradées, dans un contexte de vidéo surveillance à faible résolution et sur des bases

d'images à grande échelle impliquant des millions d'identités, afin de consolider les résultats obtenus.

- **Intégration de la fusion multimodale au sein du système**

  Il serait également intéressant de proposer un système pour le FR et le FER sur la base d'une fusion multimodale d'une part basée sur les descripteurs au niveau de l'extraction des fonctionnalités et d'un autre part sur la base des classificateurs au niveau de la reconnaissance, pour améliorer encore le taux de reconnaissance et exploiter d'une façon optimale et pratique les qualités des descripteurs et classificateurs.

- **Amélioration des performances du système de reconnaissance faciale et expressions faciales**

  La tendance actuelle sera bien sûre vers un système en temps réel appliqué dans un environnement non contrôlé en utilisant le Deep Learning.

Bibliographie

[1] Viola, P., Jones, M. (2001). Robust real-time object detection. International journal of computer vision, 4 (4), 34-47.

[2] Daugman, J. G. (1988). Complete discrete 2-D Gabor transforms by neural networks for image analysis andcompression. IEEE Transactions on acoustics, speech, and signal processing, 36(7), 1169-1179.

[3] Ojala, T., Pietikäinen, M., Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis & Machine Intelligence, (7), 971-987.

[4] Dalal, N., Triggs, B. (2005). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE

[5] Jolliffe, I. T., Cadima, J. (2016). Principal component analysis: a review and recent developments. PhilosophicalTransactions of the Royal Society A: Mathematical, Physical and Engineering sciences, 374 (2065), 20150202.

[6] Simonyan K., Parkhi O.M., Vedaldi A. (2013). Zisserman A. Fisher vector faces in the wild; Proceedings of the BMVC British Machine Vision Conference; Bristol, UK. 9–13 September 2013.

[7] Fan, R. E., Chen, P. H., Lin, C. J. (2005). Working set selection using second-order information for training support vector machines. Journal of Machine Learning Research, 6 (December), 1889-1918.

[8] Nigsch, Florian; Bender, Andreas; van Buuren, Bernd; Tissen, Jos; Nigsch, Eduard; Mitchell, John B. O. (2006). "Melting point prediction employing k-nearest neighbor algorithms and genetic parameter optimization". Journal of Chemical Information and Modeling, 46 (6): 2412–2422

[9] Tong, Y., Chen, R., Cheng, Y. (2014). Facial expression recognition algorithm using LGC based on horizontal and diagonal prior principle. Optik-International Journal for Light and Electron Optics, 125 (16) 4186-4189.

[10] V. Ojansivu and J. Heikkila. Blur insensitive texture classification using local phase quantization. In Proc. Int. Conf. on Image and Signal Processing (ICISP08), pages 236-243, 2008.

[11] T. Jabid, M.H. Kabir, and O. Chae, (2009). Local directional pattern (LDP) A robust image descriptor for object recognition, Advanced Video and Signal Based Surveillance (AVSS), 2010 7thIEEE International Conference on, 2010, pp. 482-487.

[12] Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. ACM ComputingSurvey (CSUR), 35(4), 399-458.

[13] Kortli, Y., Jridi, M., Al Falou, A., Atri, M. (2020). Face recognition systems: A Survey. Sensors, 20(2), 342.

[14]J. S. Bruner and R. Tagiuri. (1954). Thepercepton of people. Handbook of Social Psycology, 2(17), 75 - 90.

[15] M Ballantyne, M., Boyer, R. S., & Hines, L. (1996). Woody bledsoe: His life and legacy. AI magazine, 17(1), 7-7.

[16] Bledsoe, W. W. (1968). Semi automatic facial recognition. *Technical report Sri Project* 6693.

[17] Goldstein, A. J., Harmon, L. D., &Lesk, A. B. (1971). Identification of human faces. Proceedings of the IEEE, 59(5), 748-760.

[18]Fischler, M. A., & Elschlager, R. A. (1973). The representation and matching of pictorial structures. IEEE Transactions on computers, 100(1), 67-92.

[19] Kanade, T. (1974). Picture processing system by computer complex and recognition of human faces.

[20] Nixon, M. (1985, December). Eye spacing measurement for facial recognition. *In Applications of Digital Image Processing VIII (Vol. 575, pp. 279-285). International Society for Optics and Photonics*.

[21] Stonham, T. J. (1986). Practical face recognition and verification with WISARD. In *Aspects of face processing* (pp. 426-441). Springer, Dordrecht.

[22] Sirovich, L., Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. Josa a, 4(3), 519-524.

[23] Kirby, M., & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern analysis and Machine intelligence*, 12(1), 103-108.

[24] Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, *3*(1), 71-86.

[25] Shah, J. H., Sharif, M., Raza, M., Azeem, A. (2013). A Survey: Linear and Nonlinear PCA Based Face Recognition Techniques. *Int. Arab J. Inf. Technol.*, *10*(6), 536-545.

[26] Seo, H. J., Milanfar, P. (2011). Face verification using the lark representation. *IEEE Transactions on Information Forensics and Security*, *6*(4), 1275-1286.

[27] Vinay, A., Hebbar, D., Shekhar, V. S., Murthy, K. B., & Natarajan, S. (2015). Two novel detector-descriptor based approaches for face recognition using sift and surf. Procedia Computer Science, 70, 185-197.

[28] Turk, M. (2001). A random walk through EigenSpace. IEICE transactions on information and systems, 84(12), 1586-1595.

[29] Louban, R. (2009). Image processing of edge and surface defects. Berlin Heidelberg.

[30] Liu, Q., Peng, G. Z. (2010, March). A robust skin color based face detection algorithm. In 2010 2ndInternational Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010) (Vol. 2, pp. 525-528). IEEE.

[31] Divjak, M., Bischof, H. (2008). Real-time video-based eye blink analysis for detection of low blink-rate during computer use. First international workshop on tracking humans for the evaluation of their motion in image sequences (THEMIS 2008) (pp. 99-107).

[32] C.-C. Han, H.-Y. M. Liao, K. Chung Yu, and L.-H. Chen (1997). Fast face detection via morphology-based pre-processing Lecture Notes in Computer Science, volume 1311, pages 469–476. Springer Berlin / Heidelberg.

[33] Goldberger, J., Gordon, S., & Greenspan, H. (2003, October). An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures. In null (p. 487). IEEE.

[34] Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. Journal of cognitive neuroscience, 8(6), 551-565.

[35] Jain, A. K., Duin, R. P. W., Mao, J. (2000). Statistical pattern recognition: A review. IEEE Transactions on pattern analysis and machine intelligence, 22(1), 4-37.

[36] Zhang, G., Jin, W., Hu, L. (2004). A novel feature selection approach and its application. In International Conference on Computational and Information Science (pp. 665-671). Springer, Berlin, Heidelberg.

[37] Bronstein, A. M., Bronstein, M. M., & Kimmel, R. (2005). Three-dimensional face recognition. International Journal of Computer Vision, 64(1), 5-30.

[38] Del Bimbo, A. (1999). Visual information retrieval. Morgan and Kaufmann.

[39]Pogačnik, D., Ravnik, R., Bovcon, N., &Solina, F. (2012). Evaluating photo aesthetics using machine learning.

[40] Liao, S., Jain, A. K., & Li, S. Z. (2012). Partial face recognition: Alignment-free approach. IEEE Transactions on pattern analysis and machine intelligence, 35(5), 1193-1205.

[41] Achecheb M., Khene Y. Segmentation couleur dans l'espace couleur, Thèse d'ingéniorat. USTHB, 1999.

[42] Zhao, W., Chellappa, R. (2002). Image-based face recognition: Issues and methods. Optical engineering-New York-marcel Dekker incorporated-, 78, 375-402.

[43]Awwal, A. A., Iftekharuddin, K. M., Javidi, B. (2007, September). Optics and Photonics for Information Processing. In SPIE (Vol. 6695).

[44] Karaaba, M., Surinta, O., Schomaker, L., Wiering, M. A. (2015, December). Robust face recognition by computing distances from multiple histograms of oriented gradients. 2015 IEEE Symposium Series on Computational Intelligence (pp. 203-209). IEEE.

[45] Hussain, S. U., Napoléon, T., Jurie, F. (2012, September). Face recognition using local quantized patterns.

[46] Karaaba, M., Surinta, O., Schomaker, L., Wiering, M. A. (2015, December). Robust face recognition by computing distances from multiple histograms of oriented gradients. In 2015 IEEE Symposium Series on Computational Intelligence (pp. 203-209).

[47] Arigbabu, O. A., Ahmad, S. M. S., Adnan, W. A. W., Yussof, S., Mahmood, S. (2017). Soft biometrics: Gender recognition from unconstrained face images using local feature descriptor. ArXiv preprint arXiv:1702.02537.

[48] Leonard I., Alfalou A., Brosseau C. (2012) Face Recognition: Methods, Applications and Technology. Nova Science Pub Inc. London, UK: Face recognition based on composite correlation filters: Analysis of their performances.

[49] Napoléon, T., Alfalou, A. (2017). Pose invariant face recognition: 3D model from single photo. Optics and Lasers in Engineering, 89, 150-161.

[50] Milborrow, S., & Nicolls, F. (2008, October). Locating facial features with an extended active shape model. In European conference on computer vision (pp. 504-513). Springer, Berlin, Heidelberg.

[51] Purwosumarto, P., Francis, T. S. (1997). Robustness of joint transform correlator versus Vander Lugt correlator. Optical Engineering, 36(10), 2775-2780.

[52]Heflin, B., Scheirer, W., & Boult, T. E. (2012, January). For your eyes only. In 2012 IEEE Workshop on the Applications of Computer Vision (WACV) (pp. 193-200). IEEE.

[53] Lenc, L., Král, P. (2015). Automatic face recognition system based on the SIFT features. Computers & Electrical Engineering, 46, 256-272.

[54] Işık, Ş., Özkan, K. (2014). A comparative evaluation of well-known feature detectors and descriptors. International Journal of Applied Mathematics, Electronics and Computers, 3(1), 1-6.

[55] Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., Fua, P. (2012). BRIEF: Computing a local binary descriptor very fast. IEEE transactions on pattern analysis and machine intelligence, 34(7), 1281-1298.

[56] Devi, B. J., Veeranjaneyulu, N., Kishore, K. V. K. (2010). A novel face recognition system based on combining eigenfaces with fisher faces using wavelets. Procedia Computer Science, 2, 44-51.

[57] Simonyan, K., Parkhi, O. M., Vedaldi, A., Zisserman, A. (2013). Fisher vector faces in the wild. In BMVC (Vol. 2, No. 3, p. 4).

[58] Annalakshmi, M., Roomi, S. M. M., Naveedh, A. S. (2019). A hybrid technique for gender classification with SLBP and HOG features. Cluster Computing, 22(1), 11-20.

[59] Cui, Z., Li, W., Xu, D., Shan, S., Chen, X. (2013). Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3554-3561).

[60] Prince, S., Li, P., Fu, Y., Mohammed, U., Elder, J. (2011). Probabilistic models for inference about identity. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(1), 144-157.

[61] Perlibakas, V. (2006). Face recognition using principal component analysis and log-gabor filters. ArXiv preprint cs/0605025.

[62] Samra, A. S., Allah, S. E. T. G., Ibrahim, R. M. (2003, December). Face recognition using wavelet transform, fast Fourier transform and discrete cosine transform. In 2003 46th Midwest Symposium on Circuits and Systems (Vol. 1, pp. 272-275).

[63] Huang, Z. H., Li, W. J., Shang, J., Wang, J., Zhang, T. (2015). Non-uniform patch based face recognition via 2D-DWT. Image and Vision Computing, 37, 12-19.

[64] Sufyanu, Z., Mohamad, F. S., Yusuf, A. A., Mamat, M. B. (2016). Enhanced Face Recognition Using Discrete Cosine Transform. Engineering Letters, 24(1).

[65] L. Sirovich; M. Kirby (1987). "Low-dimensional procedure for the characterization of human faces". Journal of the Optical Society of America A. 4 (3): 519–524.

[66] Osuna, E., Freund, R., Girosit, F. (1997, June). Training support vector machines: an application to face detection. In Proceedings of IEEE computer society conference on computer vision and pattern recognition (pp. 130-136).

[67] Yang, M. H., Roth, D., Ahuja, N. (2000). A SNoW-based face detector. In Advances in neural information processing systems (pp. 862-868).

[68] Moghaddam, B., Jebara, T., Pentland, A. (2000). Bayesian face recognition. Pattern recognition, 33(11), 1771-1782.

[69] Bevilacqua, V., Cariello, L., Carro, G., Daleno, D., Mastronardi, G. (2008). A face recognition system based on Pseudo 2D HMM applied to neural network coefficients. Soft Computing, 12(7), 615-621.

[70] Arashloo, S. R., Kittler, J. (2014). Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features. IEEE Transactions on Information Forensics and Security, 9(12), 2100-2109.

[71] Vankayalapati, H. D., Kyamakya, K. (2009). Nonlinear feature extraction approaches with application to face recognition over large databases. In 2009 second International Workshop on Nonlinear Dynamics and Synchronization (pp. 44-48).

[72] Yang, J., Frangi, A. F., Yang, J. Y. (2004). A new kernel Fisher discriminant algorithm with application to face recognition. Neurocomputing, 56, 415-421.

[73] Pang, Y., Liu, Z., & Yu, N. (2006). A new nonlinear feature extraction method for face recognition. Neurocomputing, 69(7-9), 949-953.

[74] Fathima, A. A., Ajitha, S., Vaidehi, V., Hemalatha, M., Karthigaiveni, R., Kumar, R. (2015, November). Hybrid approach for face recognition combining Gabor Wavelet and Linear Discriminant Analysis. In 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS) (pp. 220-225).

[75] Barkan, O., Weill, J., Wolf, L., Aronowitz, H. (2013). Fast high dimensional vector multiplication face recognition. In Proceedings of the IEEE international conference on computer vision (pp. 1960-1967).

[76] Juefei-Xu, F., Luu, K., Savvides, M. (2015). Spartans: Single-sample periocular-based alignment-robust recognition technique applied to non-frontal scenarios. IEEE Transactions on Image Processing, 24(12), 4780-4795.

[77] Yan, Y., Wang, H., Suter, D. (2014). Multi-subregion based correlation filter bank for robust face recognition. Pattern Recognition, 47(11), 3487-3501.

[78] Simonyan, K., Parkhi, O. M., Vedaldi, A., Zisserman, A. (2013, September). Fisher vector faces in the wild. In BMVC (Vol. 2, No. 3, p. 4).

[79] Ding, C., Tao, D. (2015). Robust face recognition via multimodal deep face representation. IEEE Transactions on Multimedia, 17(11), 2049-2058.

[80] Moussa, M., Hmila, M., Douik, A. (2018). A novel face recognition approach based on genetic algorithm optimization. Stud. Inform. Control, 27(1), 127-134.

[81] Mian, A., Bennamoun, M., Owens, R. (2007). An efficient multimodal 2D-3D hybrid approach to automatic face recognition. IEEE transactions on pattern analysis and machine intelligence, 29(11), 1927-1943.

[82] Cho, H., Roberts, R., Jung, B., Choi, O., Moon, S. (2014). An efficient hybrid face recognition algorithm using PCA and GABOR wavelets. International Journal of Advanced Robotic Systems, 11(4), 59.

[83] Sing, J. K., Chowdhury, S., Basu, D. K., Nasipuri, M. (2012). An improved hybrid approach to face recognition by fusing local and global discriminant features. International Journal of Biometrics, 4(2), 144-164.

[84] Kamencay, P., Zachariasova, M., Hudec, R., Jarina, R., Benco, M., Hlubik, J. (2013). A novel approach to face recognition using image segmentation based on spca-knn method. Radioengineering, 22(1), 92-99.

[85] Sun, J., Fu, Y., Li, S., He, J., Xu, C., Tan, L. (2018). Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors. Journal of Sensors, 2018.

[86] Hu, J., Lu, J., & Tan, Y. P. (2014). Discriminative deep metric learning for face verification in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1875-1882).

[87] Wu, X., He, R., Sun, Z., Tan, T. (2018). A light cnn for deep face representation with noisy labels. IEEE Transactions on Information Forensics and Security, 13(11), 2884-2896.

[88] Liu, W., Wen, Y., Yu, Z., Yang, M. (2016). Large-margin softmax loss for convolutional neural networks. In ICML (Vol. 2, No. 3, p. 7).

[89] Schroff, F., Kalenichenko, D., Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).

[90] Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J. Liu, W. (2018). Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5265-5274).

[91] Du G., Su F., Cai A. MIPPR 2009: Pattern Recognition and Computer Vision. Volume 7496. SPIE; Bellingham, WA, USA: 2009. Face recognition using SURF features; p. 749628. International Society for Optics and Photonics

[92] Eichkitz, C. G., Davies, J., Amtmann, J., Schreilechner, M.G., and De Groot, P. (2015). Grey level co-occurrence matrix and its application to seismic data, First Break, 33(3), 71-77.

[93] Fekri-Ershad, Sh. (2012). Texture classification approach based on energy variation. International Journal of Multimedia Technology, 2(2), 52-55.

[94] Rahim, M.A., Azam, M.S., Hossain, N., and Islam, M.R. (2013). Face recognition using local binary patterns (LBP). Global Journal of Computer Science and Technology, 1-12.

[95] Nagy, A.M., Ahmed, A., and Zayed, H. (2014). Particle filter based on joint color texture histogram for object tracking, Paper presented at the Proceeding of the First International Conference on Image Processing, Applications and Systems Conference (IPAS), pp. 1-6.

[96] Raheja, J. L., Kumar, S., and Chaudhary, A. (2013). Fabric defect detection based on GLCM and Gabor filter: A comparison, Optik-International Journal for Light and Electron Optics, 124(23), 6469-6474.

[97] Patterns, M. L. (2017). Breast density classification using multiresolution local quinary patterns in Mammograms, Paper presented at the Proceeding of the 21st Annual Conference in Medical Image Understanding and Analysis, pp. 365-376, Edinburgh, UK.

[98] Duque, J. C., Patino, J. E., Ruiz, L. A., and Pardo-Pascual, J. E. (2015). Measuring intra-urban poverty using land Cover and texture metrics derived from remote sensing data, Landscape and Urban Planning, 135, 11-21.

[99] Wood, E. M., Pidgeon, A. M., Radeloff, V. C., &Keuler, N. S. (2012). Image texture as a remotely sensed measure of vegetation structure. Remote Sensing of Environment, 121, 516-526.

[100] Zhang, L., Zhou, Z., Li, H. (2012, September). Binary Gabor pattern: An efficient and robust descriptor for texture classification. In 2012 19Th IEEE international conference on image processing (pp. 81-84). IEEE.

[101] Tan, X., Triggs, B. (2007, October). Fusing Gabor and LBP feature sets for kernel-based face recognition. In International Workshop on Analysis and Modeling of Faces and Gestures (pp. 235-249). Springer, Berlin, Heidelberg.

[102] Yang, P., Yang, G. (2016). Feature extraction using dual-tree complex wavelet transform and gray level co-occurrence matrix. Neuro computing, 197, 212-220.

[103] Fekri-Ershad, Sh., and Tajeripour, F. (2017). Color texture classification based on proposed impulse-noise resistant color local binary patterns and significant points selection algorithm. Sensor Review, 37(1), 33-42.

[104] Ershad, S. F. (2011). Color texture classification approach based on combination of primitive pattern units and statistical features. International Journal of Multimedia & Its Applications, 3(3), 1-13.

[105] Liang, H., Weller, D. S. (2016, September). Edge-based texture granularity detection. In 2016 IEEE International Conference on Image Processing (ICIP) (pp. 3563-3567). IEEE.

[106] Xu, Y., Yang, X., Ling, H., Ji, H. (2010, June). A new texture descriptor using multifractal analysis in multi-orientation wavelet pyramid. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 161-168).

[107] Cross, G. R., Jain, A. K. (1983). Markov random field texture models. IEEE Transactions on Pattern Analysis & Machine Intelligence, (1), 25-39.

[108] Mao, J., Jain, A. K. (1992). Texture classification and segmentation using multiresolution simultaneous autoregressive models. Pattern recognition, 25(2), 173-188.

[109] Ahmadvand, A., Daliri, M. R. (2016). Rotation invariant texture classification using extended wavelet channel combining and LL channel filter bank. Knowledge-Based Systems, 97, 75-88.

[110] Chang, B. M., Tsai, H. H., & Yen, C. Y. (2016). SVM-PSO based rotation-invariant image texture classification in SVD and DWT domains. Engineering Applications of Artificial Intelligence, 52, 96-107.

[111] Fekri-Ershad, Sh., and Tajeripour, F., (2017). Impulse-Noise resistant color-texture classification approach using hybrid color local binary patterns and kullback–leibler divergence. Computer Journal, 60(11), 1633-1648.

[112] Haralick, R. M., &Shanmugam, K. (1973). Textural features for image classification. IEEE Transactions on systems, man, and cybernetics, (6), 610-621.

[113] Sobel, I. (1990). An isotropic 3 image gradient operator, Machine Vision for Three-Dimensional Scenes (H. Freemaned.).

[114] Prewitt, J. M. (1970). Object enhancement and extraction. Picture processing and Psychopictorics, 10(1), 15-19.

[115] Lowe, D. G. (1999, September). Object recognition from local scale-invariant features. In iccv (Vol. 99, No. 2, pp. 150-1157).

[116] David G. Lowe, « Distinctive image features from scale-invariant keypoints », International Journal of Computer Vision, vol. 60, no 2,110–91 p. ,2004 .

[117] Povlow B, Dunn S. Texture classification using non causal hidden Markov models. IEEE Trans. Pattern Analysis and Machine Intelligence 1995; 17:10:1010-1014.

[118] Panjwani D, Healey G. Markov random field models for unsupervised segmentation of textured color images", IEEE Trans. Pattern Analysis and Machine Intelligence 1995; 17:11:939-954

[119] Mandelbrot, B. B. (1977). Ebookstore Release Benoit B Mandelbrot Fractals: Form, Chance And Dimension prc, Freeman, p. 365.

[120] Scheunders, P., Livens, S., Van-de-Wouwer, G., Vautrot, P., and Van-Dyck, D. (1998). Wavelet-based texture analysis. International Journal on Computer Science and Information Management, 1(2), 22-34.

[121] Shen, L., & Yin, Q. (2009). Texture classification using curvelet transform. In Proceedings. The 2009 International Symposium on Information Processing (ISIP 2009) (p. 319). Academy Publisher.

[122] Arivazhagan, S., Ganesan, L., & Kumar, T. S. (2006). Texture classification using ridgelet transform. Pattern Recognition Letters, 27(16), 1875-1883.

[123] Idrissa, M., &Acheroy, M. (2002). Texture classification using Gabor filters. Pattern Recognition Letters, 23(9), 1095-1102.

[124] Gonzalez, R. C., Woods, R. E., &Eddins, S. L. (1992). Representation and description. Digital Image Processing, 2, 643-692.

[125] J. S. Li, D. Gong, and Y. Yuan, Face recognition using Weber local descriptors. Neuro computing 122 (2013) 272- 283.

[126] Tan, and Bill Triggs,"Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions," IEEE Transaction on Image Processing, vol. 19, No. 6, June 2010, pp. 1635-1650.

[127] Adin Ramirez Rivera, Jorge Rojas Castillo, and OksamChae,"Local Directional Number Pattern for Face Analysis:Face and Expression Recognition," IEEE Transactions on image processing, vol.22, No.5, May 2013, pp.1740-1752.

[128] Adin Ramirez Rivera, Jorge Rojas Castillo, and OksamChae, "Local Directional Texture Pattern image descriptor", Elsevier, Pattern Recognition Letters, September 2014, pp. 94-100.

[129] Report on a general problem-solving program. Newell, A., Shaw, J.C. and Simon, H.A. 1959. Proceedings of the International Conference on Information Processing. pp. 256-264.

[130] Andreas Kaplan et Michael Haenlein, Siri, Siri in my Hand, who's the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence [archive], Business Horizons, vol. 62, no 1, 2018

[131] Dutton D, Conroy G (1996) A review of machine learning. KnowlEng Rev 12:341–367

[132] I. Arel, C. Rose, and T. Karnowski. Deep machine learning a new frontier in artificial intelligence. IEEE Computational Intelligence Magazine, 5:13 – 18, November 2010.

[133] Alan Turing, « Computing machinery and intelligence », Mind (en), Oxford University Press, vol. 59, no 236, 1950, p. 433-460

[134] A. L. Samuel, IBM Journal of Research and Development , Vol 3, July 1959, p. 210-229.

[135] Kevin P. Murphy. Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series). The MIT Press, 2012.

[136] Pat Langley and Herbert A. Simon. Applications of machine learning and rule induction. Communications ACM, 38(11):54-64, 1995.

[137] Peter Harrington. Machine Learning in Action. Manning Publications Co, Green-wich, CT, USA, 2012. ISBN 1617290181, 9781617290183.

[138] Tom M. Mitchell. Machine learning and data mining. Communications ACM, 42(11):30-36, 1999.

[139] Alpaydın, E. (2014). Introduction to machine learning. Cambridge, MA: MIT Press.

[140] Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. Informatica, 31, 249–268.

[141] Marshland, S. (2015). Machine learning: An algorithm perspective. Boca Raton, FL: CRC Press.

[142] Morgan, G. A., Vaske, J. J., Gliner, J. A., & Harmon, R. J. (2003). Logistic Regression and Discriminant Analysis: Use and Interpretation. Journal of the American Academy of Child & Adolescent Psychiatry, 42(8), 994–997.

[143] D. Berry, Statistics—A Bayesian Perspective, Duxbury Press, 1996.

[144] D. D. Lewis, Naive (Bayes) at forty: the independence assumption in information retrieval, in: C. N´edellec, C. Rouveirol (Eds.), Machine Learning: ECML-98: 10th European Conference on Machine Learning, Chemnitz, Germany, April 21–23, Springer, Berlin/Heidelberg, 1998, pp. 4–15.

[145] Dahmane, M., Meunier, J. (2011, March). Emotion recognition using dynamic grid-based HoGfeatures. In: Face and Gesture 2011 (p. 884-888).

[146] Gershman A, Meisels A, Lüke KH, Rokach L, Schclar A, Sturm A. A Decision Tree Based Recommender System. InIICS 2010 Jun 3 (pp. 170-179).

[147] Priyama A, Abhijeeta RG, Ratheeb A, Srivastavab S. Comparative analysis of decision tree classification algorithms. International Journal of Current Engineering and Technology. 2013 Jun; 3(2):334-7.

[148] Jayakameswaraiah M, Ramakrishna S. (2014) Implementation of an Improved ID3 Decision Tree Algorithm in Data Mining System. International Journal of Computer Science and EngineeringVolume-2, Issue-3 E-ISSN.

[149] L. Breiman. Random forests. (2001) Machine Learning, 45(1): 5–32.

[150] Friedman, Jerome, Hastie, Trevor, and Tibshirani, Robert. The elements of statistical learning, volume 2. Springer series in statistics, New York, NY, USA, 2009.

[151] Hinton G.E., Srivastava N., Krizhevsky A., Sutskever I., and Salakhutdinov R. Improving neural networks by preventing co-adaptation of feature detectors. CoRR, abs/1207.0580, 2012.

[152] Bishop (1995) : Neural networks for pattern recognition, Oxford University Press.

[153] LeCun Y., Jackel L., Boser B., Denker J., Graf H., Guyon I., Henderson D., Howard R., and Hubbard W. (1998) Handwritten digit recognition : Applications of neural networks chipsand automatic learning. Proceedings of the IEEE, 86(11):2278–2324.

[154] Hochreiter S. and SchmidhuberJ. (1997) Long short-term memory. Neural Computation, 9(8):1735–1780.

[155] R. Dechter and J. Pearl, The cycle-cutset method for improving search performance in AI applications. University of California, Computer Science Department, 1986.

[156] I. Aizenberg, N. N. Aizenberg, and J. P. Vandewalle, Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications. Springer Science & Business Media, 2013.

[157] J. Schmidhuber, "Deep learning.," Scholarpedia, vol. 10, no. 11, p. 32832, 2015.

[158] Ekman, P. (2004). Emotions revealed. Bmj, 328 (Suppl S5).

[159] Chuang, C. F., Shih, F. Y. (2006). Recognizing facial action units using independent component analysis and supportvector machine. Pattern recognition, 39(9), 1795-1798.

[160] I. Kotsia, S. Zafeiriou, I. Pitas, Texture and Shape Information Fusion for Facial Expression and Facial Action Unit Recognition, Pattern Recognition 41 (3) (2008) 833851.

[161] Shaveta Dargan, Munish Kumar, Maruthi Rohit Ayyagari, Gulshan Kumar. (2020) A Survey of Deep Learning and ItsApplications: A New Paradigm to Machine Learning. Archives of Computational Methods in Engineering. Vol 27, pages1071–1092.

[162] T. F. Cootes and C. J. Taylor, "Statistical models of appearance for computer vision," Wolfson Image Anal. Unit, Univ. Manchester, Manchester, U.K., Tech. Rep., 1999.

[163] Ten Berge, J. M. (1977). Orthogonal Procrustes rotation for two or more matrices. Psychometrika, 42(2), 267-276.

[164] Neto, J. B. C., Marana, A. N., Ferrari, C., Berretti, S., Del Bimbo, A. (2019). Deep Learning from 3DLBP Descriptors for Depth Image Based Face Recognition. In: the IEEE International Conference on Biometrics (ICB), p. 1–7, (June).

[165] Zhang, H., & Li, L. (2017, October). Facial Expression Recognition Using Histogram Sequence of Local Gabor Gradient Code-Horizontal Diagonal and Oriented Gradient Descriptor. In Chinese Intelligent Systems Conference (pp. 243-251). Springer, Singapore

[166] Ahmed, F., Hossain, E., Automated facial expression recognition using gradient-based ternary texture patterns. Chin. J. Eng. 1–8, 2013.

[167] Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X., Gao, W.: WLD: A robust local image descriptor. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(9), 1705–1720 (2010).

[168] S. Zhang, B. Hu, T. Li, and X. Zheng, A Study on Emotion Recognition Based on Hierarchical Adaboost Multi-class Algorithm. International Conference on Algorithms and Architectures for Parallel Processing, pp. 105-113. Springer, Cham, November 2018.

[169] M.S. Islam, Local gradient pattern-A novel feature representation for facial expression recognition. Journal of AI and Data Mining 2 (2014) 33-38.

[170] M.S. Islam, Gender Classification using Gradient Direction Pattern. Science International 25 (2013).

[171] LalehArmi, ShervanFekri-Ershad, Texture image analysis and texture classification methods - a review. International Online Journal of Image Processing and Pattern Recognition Vol. 2, No.1, pp. 1-29, 2019

[172]Manisha Verma, Balasubramanian Raman, SubrahmanyamMurala . Texture Local Extrema Co-occurrence Pattern for Color and Texture Image Retrieval. NeurocomputingVolume 165, 1 October 2015, Pages 255-269

[173] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: application to face recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 28, no. 12, pp. 2037–2041, 2006.

[174] VahidNaghashi, Co-occurrence of adjacent sparse local ternary patterns: A feature descriptor for texture and face image retrieval, Optik journal, Volume 157, March 2018, Pages 877-889

[175] Adin Ramirez Rivera, Jorge Rojas Castillo, and OksamChae,"Local Directional Number Pattern for Face Analysis: Face and Expression Recognition," IEEE Transactions on image processing, vol.22, No.5, May 2013, pp.1740-1752.

[176] M.Belahcene, M.Laid, A.Chouchane, A.Ouamane and S.Bourennane, Local descriptions and tensor local preserving projetion in face recognition, Mathematics Computer Science, (2016) 6th European Workshop on Visual Information Processing (EUVIP)

[177] Yuxin Ma1 , Wei Chen1, Xiaohong Ma1 , Jiayi Xu1 , Xinxin Huang1 , Ross Maciejewski2 , and Anthony K. H. Tung, Easy SVM: A visual analysis approach for open-box support vector machines, Springer, Computational Visual Media volume 3, pages161–175(2017)

[178] Mingu Kang, Sujan K. Gonugondla, Sungmin Lim, Naresh R. Shanbhag A 19.4-nJ/Decision, 364-K Decisions/s, In-Memory Random Forest Multi-Class Inference Accelerator, IEEE Journal of Solid-State Circuits. Volume: 53 , Issue: 7 , July 2018

[179] Karen Simonyan, Andrew Zisserman (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition, Computer Vision and Pattern Recognition.

[180]The CK+ Database:https://www.kaggle.com/c/visum-facial-expression-analysis.

[181] The ORL Database:http://cam-orl.co.uk/facedatabase.html.

[182] The JAFFE Database:https://zenodo.org/record/3451524#.X3irdcIzbIV.

[183] The Yale Database:http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html

[184] LIB-SVM:https://www.csie.ntu.edu.tw/cjlin/libsvm

[185] Image de co-occurrence : http://matlab.izmiran.ru/help/toolbox/images/enhanc15.html

[186]HOG en cellules : https://www.learnopencv.com/histogram-of-oriented-gradients

[187] HOG2:https://heartbeat.fritz.ai/introduction-to-basic-object-detection-algorithms-b77295a95a63

[188] Image LBP : http://what-when-how.com/face-recognition/local-representation-of-facial-features-face-image-modeling-and-representation-face-recognition-part-2/

[189] Image d'intelligence Artificielle : https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

[190] https://www.altoros.com/blog/natural-language-processing-saves-businesses-millions-of-dollars/

[191] https://datafloq.com/read/machine-learning-explained-understanding-learning/4478

[192] https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn

[193] https://en.wikipedia.org/wiki/Supportvector machine#/media/File:KernelMachine. Accessed: 2020-05-01.

[194] https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html

[195] https://commons.wikimedia.org/wiki/File:ArtificialNeuronModel_francais.png?uselang=fr.

[196] http://blog.christianperone.com.

[197] https://acadgild.com/blog/recurrent-neural-network-tutorial.

[198] https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063.

[199] Ayeche, F., & Alti, A. (2020). Performance Evaluation of Machine Learning for Recognizing Human Facial Emotions. Revue d'Intelligence Articficelle, 34(3), 267-278.

[200] Ayeche, F., Alti, A., & Boukerram, A. (2020). Improved Face and Facial Expression Recognition Based on a Novel Local Gradient Neighborhood. Journal of Digital Information Management, 18(1), 33-45.

[201] Ayeche, F., & Alti, A. Novel Descriptors for Effective Recognition of Face and Facial Expressions.Revue d'Intelligence Articficelle, 34(5), 521-530.

[202] Abuzneid M., Ausif M. (2018), Enhanced Human Face Recognition Using LBPH Descriptor, Multi-KNN, and Back-Propagation Neural; IEEE Access (V 6); pp : 20641 – 20651

Thèse : Traitement automatique d'images de visages algorithmes et architecture

Doctorant : Farid AYECHE **Directeur :** Dr. Adel ALTI

ملخص. حاليا، تسعى تقنيات التعرف الآلي على الوجه وتعبيراتهاللوصول إلى نتائج التعرف الدقيقةوتحسين وقتالتنفيذ. لضمان التعرفالهادف والفعال على الوجوه وتعبيراتهاالعاطفية، نقترحفي هذا العمل مجموعةمنالمساهماتفيمجالالتعرف على الوجه وتعبيراتهالعاطفية. المساهمة الأولى هي تقييملأداء عمل تقنيات التعلم الآلي السبعة الشائعة على قواعد البياناتللأوجه+CK. النتائج التجريبية أظهرت أن المصنف التربيعي يقدم أداءً مميزًا. المساهمة الثانية هي اقتراح اثنين من الواصفات الجديدة التي تقلل من التعقيد ووقت التنفيذوتضمندقة التعرف العالية. يُطلق على الواصف الأول اسم LGN (جوار التدرج المحلي) استنادًا إلى الجمع بين الموصفين LBPوLGC، والثاني يسمى HDG (مخطط التدرج الاتجاهي) الذي يحسن واصف HOG (مخطط التدرج الموجه). لقد تم إجراء العديد من التجارب باستخدام قواعد البيانات للوجوه. أظهرت النتائج التجريبية أن المصنفات SVM-LGNو SVM-HDG تعطي على التوالي معدلات دقة عالية تبلغ 94.50٪ و92.12٪ مع وقتالتعرفسريع على الوجه وتعبيراته العاطفية.

كلمات البحث :تعبيرات الوجه العاطفية، التعلمالآلي،LGN،HDG ، التعرف على مشاعرللوجه،SVM .

Résumé. De nos jours, lestechniques de reconnaissance automatique des visages et d'expressions facialesnécessitentdes résultats de reconnaissance précisetun temps d'exécution réduit. Pour assurer une reconnaissanceefficace et effective des visages et d'expressions faciales, nous proposons plusieurs contributions dans le domaine de la reconnaissance faciale et ses expressions émotionnelles. La première contributionest l'évaluation des performances de sept techniques d'apprentissage automatiquestandard sur la base de donnéesCK+.Les résultats expérimentaux montrent que le classificateur quadratique offred'excellentes performancesavec 94.25%. La deuxième contribution est la proposition de deux nouveauxdescripteursqui réduisant la complexité en termes de temps de calcul et assurant une précision de reconnaissance élevée.Le premier descripteur est appelé LGN (Local Gradient Neighborhood) basée sur la combinaison des deux descripteurs LBP et LGC, le deuxième est appelé HDG (Histogram Gradient Directional) qui étendrele descripteur HOG. Plusieurs expérimentations ont été menées en utilisant des bases de données standard des visages. Les résultats d'expérimentations montrent que lesclassificateursSVM-LGNet SVM-HDG donnent respectivementdes taux deprécisionsélevées de94.50% et92.12%et un temps de calcul rapidepour lareconnaissance de visage et leurs expressions émotionnelles.

Mots clés. Expressions faciales, apprentissage automatique, LGN, HDG, Reconnaissance Emotionnelle, SVM.

Abstract. Nowadays, automatic facial recognition techniques and their emotional expressions require more accurate recognition results and reduced execution time. To ensure efficient and effective recognition of faces andemotional facial expressions, we propose several contributions in the field of facial recognitionand their emotionalexpressions. The first contribution is the performance evaluation of seven standard machine-learning techniques on the CK + database. The experimental results show that the quadratic classifier offers excellent performance with 94.25%. The second contribution is the proposition of two new descriptors, which reduce the complexity in terms of computation time and ensure high recognition accuracy rates. The first descriptor is called LGN (Local Gradient Neighborhood) based on the combination of the two well-known descriptors LBP and LGC, the second is called HDG (Histogram Gradient Directional) which extend the HOG descriptor. Several experiments were carried outwith severalfaces databases. The experimental results show that the SVM-LGN and SVM-HDG classifiers respectively give high precision rates of 94.50% and 92.12% and fast computation time.

Keywords. Facial Expressions, Machine Learning, LGN, HDG, Recognition of Facial Emotions, SVM.