



Faculté de Technologies
Département d'Automatique, Télécommunication et d'Électronique (ATE)

Mémoire de Fin d'Étude

EN VUE DE L'OBTENTION DU DIPLOME DE MASTER EN TÉLÉCOMMUNICATIONS

Option : Réseaux et Télécommunications

Thème

Reconnaissance d'Expressions Faciales à l'aide des techniques d'apprentissage automatique

Réalisé par :

Mlle. HAMED Wissam

Mlle. MERABTENE Meriem Maria

Soutenu le 19/ 06/ 2025, Devant le jury composé de :

Président : Mr MEKHMOUKH Abdenour.

Examineur : Mr. KASMI Reda

Encadrant : Mr. TOUNSI Mohamed

Co-Encadrant : Mr. BELABBACI El Ouanas

Promotion : 2024/2025

Dédicace

Je tiens tout d'abord à exprimer ma profonde gratitude à mon Dieu pour m'avoir accordé la force et la patience nécessaires à l'accomplissement de ce travail.

Je dédie ce modeste travail :

À mon oncle Lounis HAMED, Pour avoir toujours cru en moi, pour m'avoir guidée, encouragée et élevée avec sagesse et amour. Vous avez su combler l'absence de mon père avec une force et une tendresse inestimables. Vous avez été un pilier, un conseiller, un modèle et surtout un père dans toutes les dimensions de ma vie. Ce travail est à votre honneur. À mes trois tantes paternelles chéries, Vous avez comblé l'absence de ma mère par votre tendresse et votre présence constante. Merci d'avoir toujours été là, dans mes joies comme dans les moments les plus sombres. Vous êtes mes sources d'amour dans cette vie.

À mon cousin Soufian HAMED, Un frère de cœur, une source d'inspiration, un exemple de persévérance et de réussite. Merci pour ton soutien sans faille et tes encouragements qui m'ont portée depuis l'enfance.

À mon encadrant Mr TOUNSI Mohamed, Le professeur qui a marqué mon parcours universitaire par sa générosité, sa patience et sa passion pour l'enseignement. Vous êtes, l'un de mes professeurs préférés que j'ai eu la chance de rencontrer tout au long de mes années d'études. Merci pour votre accompagnement précieux, votre bienveillance et vos encouragements constants.

À mon coencadrant Mr EL OUANAS BELABBACI, Pour sa disponibilité, ses conseils éclairés, et sa gentillesse. Merci pour votre accompagnement exceptionnel, qui a été l'une des raisons principales pour lesquelles j'ai gardé la motivation nécessaire pour mener ce mémoire à bien. Vous êtes parmi les meilleurs enseignants que j'ai eu la chance de rencontrer au cours de cette expérience, et je vous en suis profondément reconnaissante.

À ma binôme, mon alliée dans cette aventure, pour la collaboration, le respect et la solidarité tout au long de ce travail.

À mes amis(es), Les compagnons fidèles de ces années d'études, Vous avez rendu ce chemin plus lumineux. Et enfin, à tous mes professeurs, qui ont contribué, chacun à sa manière, à construire mes connaissances.

HAMED

Dédicace

En premier lieu je remercie profondément Allah, dont la guidance et les bénédictions m'ont accompagnée et soutenue tout au long de ce projet.

À mes parents. Vous m'avez soutenue sans relâche, encouragée à donner le meilleur de moi-même, non seulement dans mon parcours académique, mais dans tous les aspects de ma vie. Fiers de mes accomplissements, vous avez été une véritable source de force, de motivation et d'inspiration. Merci d'avoir toujours cru en moi, même lorsque moi-même je doutais. En témoignage de mon amour profond et de ma sincère reconnaissance, je vous dédie cet ouvrage. Qu'Allah exauce vos vœux les plus chers.

À mes grands-parents. Mes sources d'affection et de tendresse.

À mes deux frères, Abdelaziz et Zakaria. Merci d'avoir été à mes côtés tout au long de mon parcours. Votre participation à la réussite de ce projet a été précieuse. Je vous suis profondément reconnaissante. Votre présence est un cadeau que la vie m'a offert. Cette dédicace est pour vous, avec ma profonde gratitude.

À Chahra, Baya, Marcunda, Lina, Diha, Maissa, B. Rayane, Mohamed et CH. Rayane, mes compagnons d'études. Chaque éclat de rire partagé, chaque larme essuyée, chaque regard complice restera gravé en moi. Grâce à vous, ce chemin parfois difficile s'est transformé en une aventure remplie de vie et de lumière. Vous avez rempli mon cœur de souvenirs doux et sincères. Merci du fond du cœur pour tout.

À mon encadrant Mr TOUNSI Mohamed. Merci pour votre engagement, votre dévouement et votre accompagnement, qui ont grandement enrichi cette expérience. Vous faites partie des meilleurs professeurs que j'ai eu la chance d'avoir durant mon parcours universitaire, et je vous en suis profondément reconnaissante.

À mon co-encadrant Mr BELABBACI El Ouanas. Merci pour votre présence constante, pour votre soutien, vos efforts et vos précieuses explications tout au long de ce projet. Je vous dédie ce travail avec toute ma reconnaissance.

À ma chère binôme. Merci pour ton soutien, ta complicité et ta bonne humeur tout au long de cette aventure. Je suis reconnaissante d'avoir vécu cette réussite à tes côtés.

MERABTENE

Remerciements

Nous arrivons aujourd'hui au terme de notre parcours universitaire. À cette occasion, nous souhaitons exprimer notre profonde gratitude à nos chères familles. Depuis le début de nos études, elles nous ont constamment soutenues, avec des encouragements précieux et une présence réconfortante. Grâce à leur appui constant, nous avons pu évoluer dans un environnement favorable à l'apprentissage, à la concentration et à l'épanouissement personnel. Leur soutien a été essentiel tout au long de cette étape importante de notre vie.

Nous tenons à exprimer notre profonde gratitude envers notre encadrant de mémoire **Tounsi Mohamed**. Grâce à son accompagnement bienveillant, ses conseils judicieux et sa grande disponibilité, nous avons pu mener à bien ce projet dans les meilleures conditions. Son expertise et sa pédagogie nous ont permis de progresser et d'acquérir des compétences précieuses pour notre avenir professionnel.

Nous remercions également notre co-encadrant **Belabbaci El Ouanas**, qui a été un véritable pilier tout au long de la réalisation de ce travail. Sa patience, ses encouragements constants et son implication ont été d'un grand soutien. Grâce à lui, nous avons gardé la motivation nécessaire pour aller jusqu'au bout. Son encadrement de qualité et ses conseils avisés ont grandement contribué à la réussite de ce mémoire. Nous n'aurions pas pu espérer de meilleurs accompagnateurs pour cette étape importante de notre parcours.

Nous remercions également les membres du jury pour avoir accepté de consacrer de leur temps à l'évaluation de notre travail.

Nous souhaitons également remercier nos amis proches ainsi que nos camarades de promotion, qui ont été des compagnons précieux tout au long de ce parcours. Leurs encouragements, leur esprit d'entraide et leur soutien constant ont joué un rôle important dans la richesse de cette expérience académique. Partager ce chemin d'apprentissage à leurs côtés a apporté une dimension humaine et chaleureuse à nos années d'études, rendant cette période encore plus agréable et inoubliable.

Résumé

Ce mémoire porte sur la reconnaissance automatique des expressions faciales à partir d'images, un domaine crucial pour l'amélioration de l'interaction homme-machine et le développement d'applications dans des secteurs comme la sécurité, la santé mentale ou les interfaces intelligentes. Il propose une analyse approfondie des principaux défis, tels que la variabilité inter-sujets et les conditions d'acquisition d'images hétérogènes, et évalue de manière systématique plusieurs méthodologies d'apprentissage automatique. La recherche vise à concevoir une chaîne de traitement complète, permettant d'identifier les combinaisons les plus efficaces entre techniques d'extraction de caractéristiques et algorithmes de classification, afin d'optimiser la précision. Le système débute par une étape de prétraitement utilisant l'algorithme de Viola-Jones pour une détection fiable des visages. Une contribution majeure de ce travail réside dans l'évaluation comparative de plusieurs descripteurs de caractéristiques texturales discriminantes : les Patrons d'Orientation Optimaux Locaux (LOOP), dans leurs variantes basées sur les histogrammes (LOOP-MBH) et les statistiques (LOOP-MBSF), la Quantification de Phase Locale (LPQ) et les Caractéristiques d'Image Statistiques Binarisées (BSIF). Pour la classification, plusieurs algorithmes supervisés sont mis en œuvre et comparés : Machines à Vecteurs de Support (SVM), K-Plus Proches Voisins (KNN), Forêts Aléatoires (RF) et Arbres de Décision (DT). Chaque combinaison descripteur-classifieur est rigoureusement évaluée sur deux bases de données publiques d'expressions faciales, afin de garantir la généralisabilité des résultats. Les évaluations montrent que le choix judicieux du descripteur et du classifieur est déterminant pour atteindre une haute performance. Les meilleurs appariements identifiés permettent d'obtenir des taux de reconnaissance élevés, offrant un cadre méthodologique clair pour la conception de systèmes robustes et précis en conditions réelles.

Mots clés : Reconnaissance automatique des expressions, Viola-Jones, descripteur de texture, classifieur d'apprentissage automatique.

Table des matières

Liste des Figures	v
Liste des Tableaux	vi
Introduction Générale	1
I Fondements de la Reconnaissance d'Expressions Faciales	4
I.1 Introduction	5
I.2 Expressions faciales	5
I.2.1 L'expression faciale neutre	5
I.2.2 La tristesse :	6
I.2.3 La joie	6
I.2.4 La colère	7
I.2.5 Le dégoût	7
I.2.6 La peur	8
I.2.7 La surprise	8
I.3 Relations entre expressions faciales et émotions	9
I.4 Application	10
I.5 Difficultés	11
I.6 Détection automatique des expressions faciales	14
I.6.1 Les phases de détection automatique des expressions faciales	14
I.6.2 Prétraitement	15
I.6.3 Extraction de caractéristiques	15
I.6.4 Classification et prise de décision	15
I.7 Conclusion	16
II État de l'art et méthodologie	17
II.1 Introduction	17
II.2 Travaux connexes	17
II.3 Méthodologie proposée	20
II.3.1 Prétraitement	21

II.3.2	Extraction des caractéristiques	23
II.3.2.1	Local Optimal-Oriented Pattern (LOOP)	24
II.3.2.2	La méthode MBH-LOOP	25
II.3.2.3	La méthode MBSF-LOOP	27
II.3.3	Étape de normalisation	28
II.3.4	Classification et prise de décision	29
II.3.4.1	Les machines à vecteurs de support (SVM)	29
II.3.4.2	La forêt aléatoire (RF)	30
II.3.4.3	L'arbre de décision (DT)	31
II.3.4.4	K-Nearest Neighbors (KNN)	31
II.4	Conclusion	33
III	Expérimentations et Résultats	34
III.1	Introduction	35
III.2	Environnement de travail	35
III.3	Protocole	35
III.4	Métriques d'évaluation utilisées	36
III.4.1	La matrice de confusion	36
III.4.2	Métriques dérivées de la matrice de confusion	37
III.5	Bases de données utilisées	38
III.5.1	La base de données CK+ (<i>Extended Cohn-Kanade</i>	38
III.5.2	La base de données JAFFE (<i>Japanese Female Facial Ex-</i> <i>pression</i>)	39
III.6	Expérimentations et résultats	39
III.6.1	Choix de la technique d'extraction des caractéristiques	39
III.6.1.1	Performances avec le descripteur LOOP Histogramme MBH-LOOP	40
III.6.1.2	Performances avec le descripteur LOOP Statis- tiques MBSF-LOOP	40
III.6.1.3	Comparaison des descripteurs MBSF-LOOP, LPQ et BSIF	41
III.6.2	Choix du modèle de classification	41
III.6.3	Comparaison des performances du système avec les bases de données CK+ et JAFFE	42
III.7	Analyse des performances de notre système FER	43
III.7.1	L'avantage du descripteur LOOP statistique	44
III.7.2	L'avantage du classificateur SVM	44
III.7.3	L'avantage de la base CK+	44
III.8	Conclusion	45
	Conclusion Générale	46

Perspectives	47
A Interface graphique	53
A.1 Fenêtre d'accueil	53
A.1.1 Le système de reconnaissance des expressions faciales en temps réel.	55
A.1.2 Environnement de travail.	58
A.1.3 Bibliothèques.	58

Table des figures

I.1	Un exemple de l'expression faciale neutre	6
I.2	Un exemple de l'expression faciale de la tristesse	6
I.3	Un exemple de l'expression faciale exprimant la joie	7
I.4	Un exemple de l'expression faciale exprimant la colère	7
I.5	Un exemple de l'expression faciale exprimant le dégoût	8
I.6	Un exemple de l'expression faciale exprimant la peur	8
I.7	Un exemple de l'expression faciale exprimant la surprise	9
I.8	Exemple de la variation d'éclairage	12
I.9	Exemple de la variation de pose	12
I.10	Exemple de vieillissement	13
I.11	Exemple de la variation d'apparences	13
I.12	Exemple d'effets de flou progressif sur un portrait	14
I.13	Les étapes du processus de détection automatique des expressions faciales	16
II.1	Architecture globale du système de reconnaissance automatique des expressions faciales	21
II.2	Un exemple de la détection de visage avec Viola et Jones	22
II.3	Les motifs des descripteurs de Haar	23
II.4	Étapes de transformation d'une image en un vecteur de caractéris- tiques à l'aide du descripteur LOOP	27
II.5	Exemples de séparation avec le classifieur SVM	30
II.6	Principe de la classification Random Forest	31
II.7	Un exemple de mécanisme de classification avec KNN	32
III.1	Détection des visages de la base de données JAFFE	39
III.2	Matrice de confusion avec la base JAFFE.	43
III.3	Matrice de confusion avec la base CK+.	43
A.1	Fenêtre d'accueil.	54
A.2	Exemple de détection de l'expression colère.	55

A.3	Exemple de détection de l'expression surprise.	55
A.4	Exemple de détection de l'expression tristesse.	56
A.5	Exemple de détection de l'expression joie.	56
A.6	Exemple de détection de l'expression peur.	56
A.7	Exemple de détection de l'expression neutre.	57
A.8	Exemple de détection de l'expression dégoût.	57
A.9	Exemple du cas non détection de visage.	57
A.10	Logo de Jupyter.	58

Liste des tableaux

I.1	Comparaison entre expressions faciales et émotions	10
III.1	Performances de notre système FER à base du descripteur MBH- LOOP.	40
III.2	Performances de notre système FER à base du descripteur MBSF- LOOP.	41
III.3	Comparaison des performances avec différents descripteurs.	41
III.4	Comparaison des performances avec divers classifieurs.	42
III.5	Performances du système avec les bases de données CK+ et JAFFE.	42

Liste des acronymes

ABR	<i>Augmentation Based Resampling</i>
AdaBoost	<i>Adaptive Boosting</i>
ANN	<i>Artificial Neural Network</i>
AUC	<i>Area Under the Curve</i>
Bi-LSTM	<i>Bidirectional Long Short-Term Memory</i>
BSIF	<i>Binarized Statistical Image Features</i>
Ck+	<i>Cohn-Kanade+</i>
CNN	<i>Convolutional Neural Network</i>
CPU	<i>Central Processing Unit</i>
DLib	<i>Deep Learning Library</i>
DMD	<i>Driver Monitoring Dataset</i>
DNN	<i>Deep Neural Network</i>
DT	<i>Decision Tree</i>
Eception	<i>Extreme Inception</i>
EFHQ	<i>ExtremePose-Face-HQ</i>
ER	<i>Emotion Recognition</i>
FAN	<i>Face Alignement Facial</i>
FD	<i>Face Detection</i>
FER	<i>Facial Expression Recognition</i>
FERDERnet	<i>Facial Expression Recognition in Driving Environment using Deep ER network</i>
FN	<i>Faux Négatif</i>
FP	<i>Faux Positif</i>
GoogLeNet	<i>Inception-v1 CNN architecture from Google</i>
HOG	<i>Histogram of Oriented Gradients</i>

JAFFE	<i>Japanese Female Facial Expression</i>
KMU-FED	<i>Kyungpook National University Facial Expression Database</i>
KNN	<i>K-nearest neighbor</i>
LBP	<i>Local Binary Patterns</i>
LDP	<i>Local Directional Pattern</i>
LMRF	<i>Lightweight Multilayer Random Forest</i>
LOOP	<i>Local Optimal Oriented Pattern</i>
LPQ	<i>Local Phase Quantization</i>
LTP	<i>Local Ternary Pattern</i>
MBH-LOOP	<i>Multi-Block Histogram – Local Optimal Oriented Pattern</i>
MBSF-LOOP	<i>Multi-Block Statistical Features – Local Optimal Oriented Pattern</i>
MMI	<i>Man-Machine Interaction</i>
OMUP	<i>Optimiseur Multi-Univers Parallèle</i>
Open Cv	<i>Open Source Computer Vision</i>
REFC	<i>Reconnaissance des émotions faciales du conducteur</i>
ResNet	<i>Residual Network</i>
ResNet50	<i>Residual Network with 50 layers</i>
RF	<i>Random Forest</i>
SVM	<i>Support Vector Machines</i>
VGGNet	<i>Visual Geometry Group Network</i>
VN	<i>Vrai Négatif</i>
VP	<i>Vrai Positif</i>

Introduction Générale

Contexte et motivation

Les expressions faciales constituent un canal de communication non verbale fondamental dans les interactions humaines. Reflets visibles des états internes, notamment émotionnels, elles permettent de transmettre instantanément une multitude d'informations telles que la joie, la colère, la surprise ou encore la peur. Ces manifestations, traduites par les contractions des muscles du visage, sont l'un des moyens les plus naturels et universels pour communiquer [1].

L'avènement des technologies numériques et les progrès en intelligence artificielle ont ouvert la voie à l'automatisation de l'analyse de ces signaux. La reconnaissance automatique des expressions faciales (RAEF) est ainsi devenue un domaine de recherche majeur en vision par ordinateur. Elle vise à doter les machines de la capacité de détecter et d'interpréter les expressions humaines à partir d'images ou des vidéos. Cette technologie offre des perspectives d'applications considérables dans des secteurs variés tels que les interfaces homme-machine (IHM), la sécurité, le marketing, l'automobile ou l'analyse comportementale. Cependant, la conception d'un système de RAEF robuste présente des défis importants. Les expressions faciales varient considérablement d'un individu à l'autre et peuvent être influencées par des facteurs culturels, contextuels ou physiologiques. De plus, la qualité des images est souvent dégradée par des conditions d'acquisition non contrôlées : variations de luminosité, changements de pose de la tête, occultations partielles ou faible résolution.

Pour surmonter ces obstacles, les chercheurs se sont massivement tournés vers les techniques d'apprentissage automatique (machine learning). Ces approches permettent d'apprendre à partir de vastes ensembles de données pour construire des modèles capables d'identifier des motifs pertinents. Le processus typique de RAEF se décompose en plusieurs étapes : détection et prétraitement du visage, extraction de caractéristiques pertinentes, et enfin classification de l'expression. L'efficacité globale du système dépend de la performance de chacune de ces étapes. Dans le cadre de ce mémoire, nous nous intéressons à la conception et à l'évaluation d'un système de reconnaissance automatique des expressions faciales à partir

d'images statiques. Notre objectif est d'étudier et de mettre en œuvre une chaîne de traitement complète, en comparant différentes approches pour l'extraction de caractéristiques et la classification, afin de proposer une solution capable d'identifier avec précision les expressions du visage humain dans des conditions visuelles variables.

Problématique

À l'heure où les interactions homme-machine deviennent de plus en plus complexes, il est essentiel de doter les systèmes artificiels de la capacité à interpréter avec justesse les signaux humains. Parmi ces signaux, les expressions faciales constituent une source d'information particulièrement riche, ouvrant la voie à des applications innovantes qui rendent les technologies plus intuitives, réactives et adaptées à leurs utilisateurs. Cependant, la mise en œuvre d'un système fiable de reconnaissance faciale automatique demeure un défi de taille. La diversité interindividuelle des expressions, la subtilité des mimiques, ainsi que les conditions souvent imparfaites d'acquisition des images (variations de luminosité, de pose ou de résolution) compliquent considérablement la tâche d'identification précise.

Ce mémoire s'inscrit dans cette problématique en visant la conception d'un système de reconnaissance d'expressions faciales à la fois robuste et précis, fondé sur l'exploitation de techniques d'apprentissage automatique et sur l'extraction de caractéristiques discriminantes, afin de surmonter les contraintes liées à la variabilité des visages et des conditions de prise de vue.

Contribution

La contribution principale de ce mémoire réside dans le développement et l'évaluation d'une chaîne de traitement complète pour la reconnaissance d'expressions faciales. Notre approche s'articule autour des points suivants :

- **Détection de visage** : utilisation de l'algorithme de Viola-Jones, une référence robuste pour la localisation des visages dans les images.
- **Extraction de caractéristiques** : mise en œuvre du descripteur **LOOP (Local Optimal Oriented Pattern)** pour capturer les micro-motifs texturaux du visage. Nous explorons deux variantes pour agréger ces informations : l'une basée sur des histogrammes de motifs et l'autre sur des mesures statistiques, afin de générer un vecteur de caractéristiques discriminant.
- **Classification** : emploi d'une **machine à vecteurs de support (SVM - Support Vector Machine)**, reconnue pour son efficacité en classification supervisée, pour catégoriser les expressions faciales à partir des caractéristiques extraites.

Cette approche vise à évaluer l'efficacité de la combinaison du descripteur LOOP et du classifieur SVM pour améliorer la précision et la robustesse de la reconnaissance. Les résultats de cette étude pourraient contribuer à l'amélioration des systèmes existants dans des domaines d'application tels que les interfaces intelligentes, la santé ou l'analyse comportementale.

Organisation du mémoire

Ce mémoire est structuré en trois chapitres principaux, précédés de cette introduction générale et suivis d'une conclusion.

- **Chapitre 1 : Fondements de la reconnaissance d'expressions faciales.** Ce chapitre introductif pose les bases théoriques du domaine. Nous y présentons les définitions fondamentales, les défis techniques associés à la RAEF et nous détaillons les étapes clés d'un système de reconnaissance type, de l'acquisition de l'image à la décision finale.
- **Chapitre 2 : Méthodologie et approche proposée.** Le deuxième chapitre présente un état de l'art des travaux existants avant de détailler la méthodologie que nous avons adoptée. Nous y décrivons en profondeur chaque module de notre système : la détection de visages avec Viola-Jones, les méthodes d'extraction de caractéristiques basées sur LOOP (MBH-LOOP, MBSF-LOOP) et les algorithmes de classification étudiés (SVM, KNN, DT, RF).
- **Chapitre 3 : Expérimentations et Résultats.** Ce dernier chapitre est consacré à la validation expérimentale de notre approche. Nous y décrivons l'environnement de travail, les bases de données publiques utilisées et les métriques d'évaluation retenues. Ensuite, nous présentons, analysons et discutons les résultats obtenus à travers une série d'expériences visant à évaluer la performance de notre système.
- **Conclusion générale et Perspectives.** Le mémoire se termine par une conclusion qui synthétise le travail accompli et les résultats obtenus. Nous y discutons les apports de notre contribution et proposons des pistes pour des travaux futurs visant à perfectionner et enrichir le système présenté. Une annexe est également fournie pour présenter l'interface graphique développée pour ce projet.

Fondements de la Reconnaissance d'Expressions Faciales

I.1 Introduction

La reconnaissance des expressions faciales constitue un domaine en pleine expansion de l'intelligence artificielle, visant à permettre aux machines de détecter et d'analyser les états émotionnels subtils des êtres humains [2].

Ce premier chapitre a pour objectif de présenter les concepts fondamentaux relatifs à la reconnaissance des expressions faciales. Il convient de distinguer cette dernière de la reconnaissance faciale, dont le but est d'identifier une personne à partir de son visage, tandis que la reconnaissance des expressions faciales se concentre sur l'interprétation des émotions véhiculées par les traits du visage. Nous proposons également un aperçu des principaux défis associés à la détection et à l'analyse des expressions faciales, notamment à travers l'utilisation de techniques d'apprentissage automatique.

I.2 Expressions faciales

Les expressions faciales sont des configurations des différents mouvements des petits muscles du visage (micromoteurs) qui sont utilisés pour déterminer l'état émotionnel et psychologique discret d'une personne [3]. Elles sont caractérisées par trois éléments clés : la bouche, les yeux et les sourcils, qui influencent la nature de l'expression faciale. Selon ces caractéristiques, on définit sept types d'expressions faciales.

I.2.1 L'expression faciale neutre

C'est une expression vide, également connue sous le nom de 'Pocker Face', qui est une expression qui ne montre aucune émotion, caractérisée par une neutralité des traits du visage. Pour la forme des éléments du visage, un regard droit, les paupières légèrement ouvertes pour les yeux, les sourcils positionnés de manière neutre, la bouche fermée ou légèrement ouverte.

L'expression faciale neutre est illustrée par l'image ci-dessous.



FIGURE I.1 – Un exemple de l'expression faciale neutre

I.2.2 La tristesse :

La tristesse est un état émotionnel négatif résultant d'une souffrance morale[4], généralement provoquée par une perte, un regret, une douleur ou un problème [5]. Pour la forme des éléments du visage, les yeux sont tombants, les sourcils sont relevés en leur centre, et la bouche forme une courbe descendante. L'expression faciale de la tristesse est illustrée par l'image ci-dessous.



FIGURE I.2 – Un exemple de l'expression faciale de la tristesse

I.2.3 La joie

La joie est un état émotionnel positif se manifestant chez une personne dans une situation de satisfaction [4]. Elle peut résulter de la réalisation d'une réussite, d'un désir, d'un rêve, ainsi que du bonheur [5]. Pour la forme des éléments du visage, les yeux sont légèrement plissés, les sourcils détendus, et la bouche forme un large sourire. L'expression faciale de la joie est illustrée par l'image ci-dessous.



FIGURE I.3 – Un exemple de l'expression faciale exprimant la joie

I.2.4 La colère

La colère est une émotion intense qui survient en réponse à des frustrations ou des injustices [4]. Elle se manifeste par des sensations physiques comme une augmentation du rythme cardiaque et une tension musculaire [5]. Les signes du visage qui peuvent indiquer la colère : on retrouve les sourcils froncés, les yeux plissés, les lèvres serrées, la rougeur du visage. Ces signes peuvent varier d'une personne à l'autre. L'expression faciale de la colère est illustrée par l'image ci-dessous.



FIGURE I.4 – Un exemple de l'expression faciale exprimant la colère

I.2.5 Le dégoût

L'expression du dégoût est une réaction faciale et corporelle qui se manifeste en réponse à des stimuli perçus comme répugnants ou désagréables [4]. Les signes du visage qui peuvent manifester l'expression du dégoût sont les sourcils froncés, le nez plissé, la bouche tordue, les yeux détournés. L'expression faciale du dégoût est illustrée par l'image ci-dessous.



FIGURE I.5 – Un exemple de l'expression faciale exprimant le dégoût

I.2.6 La peur

La peur est une émotion fondamentale qui se déclenche en réponse à une menace ou un danger perçu [4]. Elle est souvent accompagnée de sensations physiques telles qu'une augmentation du rythme cardiaque, une respiration rapide, des sueurs et une tension musculaire [5]. Les signes du visage qui indiquent la peur incluent des sourcils levés, des yeux écarquillés, une bouche ouverte et un froncement de sourcils. L'expression faciale de la peur est illustrée par l'image ci-dessous.

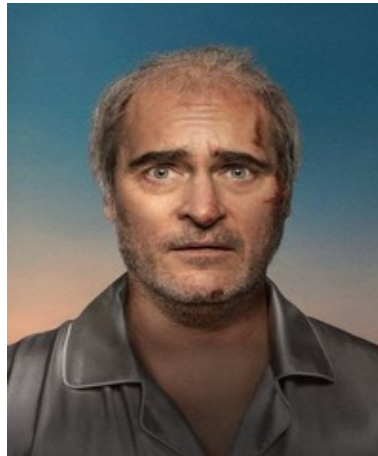


FIGURE I.6 – Un exemple de l'expression faciale exprimant la peur

I.2.7 La surprise

L'expression faciale de surprise est une réaction émotionnelle qui se produit en réponse à un événement inattendu ou surprenant [4]. Elle est souvent brève et peut être déclenchée par des stimuli positifs ou négatifs [5]. Les signes du visage qui indiquent la surprise sont les sourcils levés, des yeux écarquillés et une bouche ouverte, signalant une réponse à un événement inattendu. L'expression faciale de la surprise est illustrée par l'image ci-dessous.



FIGURE I.7 – Un exemple de l'expression faciale exprimant la surprise

I.3 Relations entre expressions faciales et émotions

L'expression faciale constitue un canal fondamental de communication non verbale, jouant un rôle clé dans la manifestation et la perception des émotions humaines. Elle permet de transmettre des indices sur l'état affectif d'une personne, facilitant ainsi les interactions sociales.

Les émotions, quant à elles, représentent des états internes complexes, engendrés en réponse à des stimuli spécifiques. Elles influencent le comportement, la cognition et les réponses physiologiques d'un individu [6]. Selon les travaux de Paul Ekman, certaines émotions dites de base (telles que la joie, la tristesse, la colère, la peur, la surprise et le dégoût) sont universellement exprimées par des expressions faciales similaires à travers différentes cultures.

Il est cependant important de souligner que, bien que les émotions et les expressions faciales soient intimement liées, elles ne sont pas toujours parfaitement synchrones. Il est tout à fait possible qu'une personne exprime volontairement une émotion qu'elle ne ressent pas, ou qu'elle masque une émotion réelle derrière une expression contrôlée. Ce phénomène peut être influencé par des facteurs sociaux, culturels ou contextuels. Une comparaison détaillée des caractéristiques distinctes entre les émotions et les expressions faciales est présentée dans le Tableau I.1.

TABLE I.1 – Comparaison entre expressions faciales et émotions

Expressions faciales	Émotions
Manifestations visibles sur le visage	États internes psychophysiologiques
Peuvent être observées directement	Ne peuvent être inférées que par des indices (comportement, parole, expressions, etc.)
Peuvent être contrôlées ou simulées volontairement	Généralement spontanées et automatiques
Utilisées pour communiquer avec autrui	Utilisées pour adapter le comportement à une situation
Universelles dans certaines expressions (ex. : sourire pour la joie)	Universelles dans certaines catégories fondamentales (Ekman)
Durée brève, souvent de quelques secondes (micro-expressions)	Peuvent durer plus longtemps et varier en intensité

I.4 Application

L'analyse automatique des expressions faciales a suscité de nombreuses études compte tenu de son utilité pratique dans plusieurs secteurs. Elle est mise en œuvre pour évaluer les émotions, optimiser l'interaction entre l'homme et la machine [7]. On la retrouve dans :

Surveillance et sécurité : La détection des émotions négatives aide à identifier les sentiments liés à des intentions malveillantes. Ainsi, l'identification d'émotions négatives comme la peur, la colère et l'agressivité dans des lieux publics permet de prévoir les comportements à risque et d'optimiser la sécurité.

Automobile : Identifier la fatigue, la distraction ou le stress du conducteur, ainsi que détecter la somnolence ou l'agitation du conducteur pour favoriser une interaction intuitive avec le véhicule à travers un assistant vocal ou un tableau de bord interactif dans le but d'augmenter la sécurité routière.

Évaluation de la santé mentale : Employée pour examiner les émotions et identifier des troubles psychologiques tels que l'autisme, la dépression ou le stress. Elle peut également être utilisée dans le cadre du suivi thérapeutique pour surveiller les réactions émotionnelles pendant les sessions.

Éducation : L'évaluation des émotions des étudiants permet aux enseignants de mieux comprendre leur engagement et leur compréhension, et aide à adapter les méthodes d'enseignement en fonction des réactions émotionnelles de ces derniers.

I.5 Difficultés

Les systèmes de détection des émotions rencontrent aussi des défis similaires à ceux de la reconnaissance faciale classique : des variations de lumière influençant la façon dont les traits du visage sont perçus, des masques partiels du visage (lunettes, cheveux, etc.) ou des positions et angles de vue non optimaux pour l'analyse, des expressions faciales intentionnellement cachées ou feintes [8]. Les expressions faciales des émotions diffèrent considérablement d'une personne à l'autre, ce qui fait que la détection des émotions est une tâche distincte de la reconnaissance faciale classique. Le défi dans la reconnaissance des émotions réside dans l'extraction des traits faciaux similaires entre diverses manifestations d'une même émotion chez différentes personnes. Trois enjeux spécifiques importants liés à la reconnaissance des émotions ont été repérés, facilitant ainsi la compréhension de la problématique et l'élaboration de systèmes informatiques spécialisés.

Variabilité individuelle : Chaque individu exprime ses émotions de manière unique. Cette variabilité rend parfois l'interprétation des émotions plus complexe et sujette à des erreurs.

Les micro-expressions : Certaines expressions peuvent être subtiles ou ambiguës, comme une micro-expression de peur dissimulée derrière un sourire. De plus, les expressions non naturelles peuvent tromper le système, qui risque alors de mal interpréter l'émotion réelle et croire, à tort, que la personne est réellement heureuse.

Complexité émotionnelle : Les émotions humaines sont souvent complexes et mélangées. Par exemple, une personne peut ressentir à la fois de la tristesse et du soulagement après avoir reçu une mauvaise nouvelle. Ce type de réaction rend difficile la classification en une seule émotion, ce qui peut entraîner une réponse inexacte du système.

On retrouve d'autres facteurs qui peuvent influencer différemment l'apparence des expressions. La reconnaissance des émotions est un sous-ensemble de l'analyse des expressions faciales, et l'apparence des expressions est sensible aux changements d'expressions faciales, aux occultations et aux poses.

Conditions d'éclairage variables : Un mauvais éclairage ou une caméra de mauvaise qualité, comme une image floue ou une faible luminosité dans une pièce sombre, peut réduire considérablement la précision du système, voire empêcher la détection correcte des traits du visage.



FIGURE I.8 – Exemple de la variation d'éclairage

Variations de pose : Les modifications de l'apparence et de la morphologie du visage peuvent être provoquées par des mouvements de tête comme l'inclinaison, le roulis et le lacet ainsi que par les changements d'angle de vue de la caméra. Ces fluctuations rendent extrêmement difficiles l'identification automatique des expressions faciales en fonction de l'angle de prise de vue (voir la figure I.9 pour des exemples d'angles variés) [9]. Il est donc crucial de corriger la pose, ce qui peut être réalisé grâce à des méthodes performantes qui permettent de faire tourner le visage et/ou de réajuster ses caractéristiques en fonction de l'axe de l'image.



FIGURE I.9 – Exemple de la variation de pose

Le vieillissement du visage : L'apparition de modifications sur le visage peut être attribuée au vieillissement, ce qui peut influencer de manière significative le mécanisme de reconnaissance des émotions. Voir la figure I.10 exemple de vieillissement du visage.



FIGURE I.10 – Exemple de vieillissement

Variations d'apparence : Dans des contextes non maîtrisés, la capacité à identifier les expressions émotionnelles doit demeurer efficace en dépit de l'usage de masques, d'objets ou d'obstructions. Des accessoires comme des casquettes, des lunettes ou une barbe peuvent en partie dissimuler les caractéristiques du visage et rendre l'interprétation des émotions plus complexe . Voir la figure I.11



FIGURE I.11 – Exemple de la variation d'apparences

Résolution et format d'image : Les performances de la détection des émotions sont directement influencées par la qualité et la résolution des images faciales, les spécifications techniques du matériel numérique, ainsi que l'emploi de divers appareils photographiques. Par exemple, la figure I.12 dépeint les difficultés engendrées par des images à faible résolution. Par ailleurs, les visages enregistrés dans des conditions réelles présentent des complexités additionnelles du fait de la variété des méthodes d'acquisition.

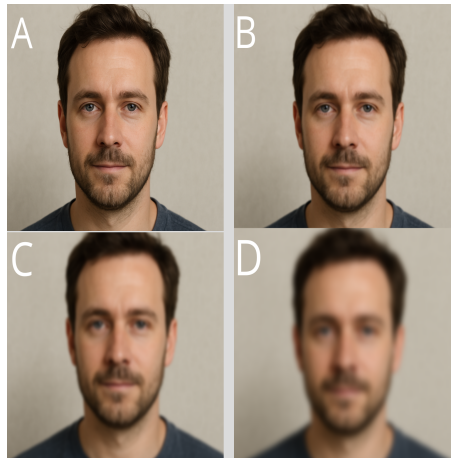


FIGURE I.12 – Exemple d'effets de flou progressif sur un portrait

I.6 Détection automatique des expressions faciales

L'objectif recherché ici est de présenter un système capable d'effectuer une analyse et une identification automatique des expressions faciales. Ce système opère généralement en trois étapes principales : La première consiste à détecter le visage dans une image, étape essentielle car elle permet d'isoler la région d'intérêt sur laquelle l'analyse sera effectuée. La seconde consiste à extraire les expressions faciales à partir des séquences d'images. Cette phase implique l'identification des éléments constitutifs du visage. Il est important de distinguer deux notions : les caractéristiques faciales et le modèle de caractéristique faciale. Les caractéristiques faciales désignent les parties spécifiques du visage comme les sourcils, les yeux, le nez, la bouche et le menton. Le modèle de caractéristique faciale représente la manière dont ces éléments sont combinés pour présenter un visage [10]. Enfin, ce système est capable de définir les catégories des expressions à détecter. Cela implique la mise en place d'un processus ou d'un mécanisme qui sur la base d'un prétraitement et de l'extraction des caractéristiques, permet une classification et une prise de décision, dans le but de la détection et de la reconnaissance automatique des expressions faciales [11].

I.6.1 Les phases de détection automatique des expressions faciales

Un système de détection automatique des expressions à partir d'images faciales en utilisant des techniques d'apprentissage automatique, fonctionne généralement selon deux grandes phases : la phase d'apprentissage et la phase de test [12]

Phase d'apprentissage : cette étape consiste à extraire à partir de données brutes (comme les images faciales), des descripteurs pertinents permettant de représenter les expressions. Ces descripteurs sont ensuite utilisés pour construire des modèles qui caractérisent les différentes expressions. L'objectif est de créer

une base de connaissances permettant de différencier les émotions en fonction de leurs signatures caractéristiques .

Phase de test : Lors de cette phase, les caractéristiques extraites à partir d'une nouvelle entrée (une expression faciale inconnue) sont comparées aux modèles des caractéristiques des expressions établis durant l'apprentissage. Un module de décision, s'appuyant sur une stratégie spécifique (un classificateur), détermine l'expression la plus probable. Il faut savoir que chacune de ces deux phases comprend les différentes étapes du processus de détection automatique des expressions faciales, à savoir le prétraitement, l'extraction des caractéristiques et la classification.

I.6.2 Prétraitement

Le prétraitement constitue une étape essentielle dans tout système de détection des expressions faciales, car il permet d'améliorer la qualité des données brutes en éliminant les informations non pertinentes et en réduisant le bruit pouvant nuire à la performance du modèle. L'objectif de cette phase est de concentrer l'analyse sur les régions comme le visage tout en minimisant l'influence des facteurs externes tels que l'arrière-plan, les variations d'éclairage ou les changements de pose [10]. Il est important de noter que les techniques de prétraitement doivent être adaptées en fonction du type de données utilisées. Dans ce mémoire, nous nous basons sur des données visuelles, qui sont les images faciales. Les données résultantes de cette étape de prétraitement seront utilisées dans les prochaines étapes de ce processus.

I.6.3 Extraction de caractéristiques

L'extraction des caractéristiques est une étape clé dans le processus de détection des expressions dans le but de les reconnaître automatiquement, car elle permet de transformer les données en représentations plus informatives, compatibles et adaptées à la classification. Son objectif principal est de faire ressortir les éléments les plus significatifs, en lien direct avec les expressions faciales, tout en réduisant la dimensionnalité des données afin d'optimiser l'efficacité des modèles d'apprentissage. L'extraction de caractéristiques facilite la tâche de classification automatique [10]. Dans le cadre de ce projet, nous avons exploré et mis en œuvre des techniques et des méthodes d'extraction adaptées aux types de données utilisés, notamment les images faciales. Les concepts théoriques associés à ces méthodes seront détaillés dans les sections suivantes.

I.6.4 Classification et prise de décision

Une fois les caractéristiques pertinentes extraites, celles-ci sont utilisées comme vecteurs d'entrée pour des algorithmes de classification. L'objectif de cette phase

est d'associer, pour chaque vecteur de caractéristiques, l'expression correspondante. Selon les approches choisies, cette classification peut s'appuyer sur des méthodes traditionnelles (comme les SVM, les K-plus proches voisins, etc.) ou sur des modèles plus avancés d'apprentissage automatique, notamment les réseaux de neurones profonds. L'étape de prise de décision repose sur l'entraînement de modèles prédictifs à partir d'un ensemble de données annotées, dans le but de distinguer les différentes expressions faciales [10]. Ce modèle apprend à reconnaître des schémas dans l'espace de représentation défini par l'extraction des caractéristiques. Une fois entraîné, il est capable de généraliser ses connaissances pour prédire l'expression exprimée dans une nouvelle donnée, jamais vue auparavant. Ainsi, l'efficacité du processus de détection repose sur la qualité de l'apprentissage supervisé réalisé à cette étape, mais aussi sur la pertinence des caractéristiques extraites dans le but d'obtenir des résultats fiables et corrects.

La figure suivante illustre le processus de détection automatique des expressions faciales.

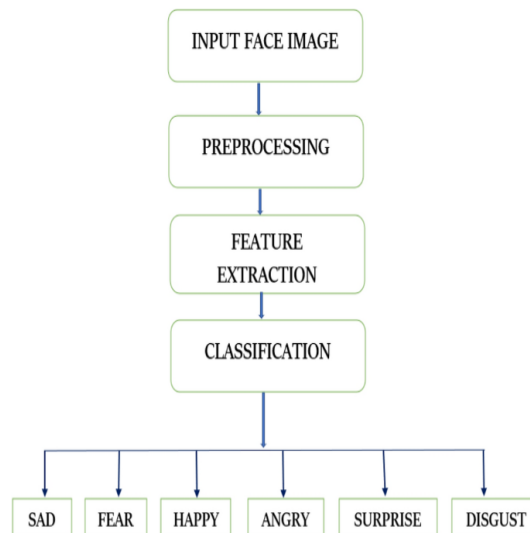


FIGURE I.13 – Les étapes du processus de détection automatique des expressions faciales

I.7 Conclusion

Dans ce chapitre, nous avons donné un aperçu sur la reconnaissance automatique des expressions faciales à partir des éléments du visage, à l'aide d'un système de détection automatique. Nous avons également abordé diverses applications potentielles de ce système, ainsi que ses principaux défis, tels que la variabilité des expressions, les conditions environnementales et les étapes du processus de détection et d'analyse automatique des expressions faciales.

Dans le prochain chapitre, nous présenterons quelques notions et plusieurs définitions relatives à la conception des systèmes de reconnaissance d'émotion.

État de l'art et méthodologie

II.1 Introduction

La détection automatique des expressions faciales constitue un champ de recherche essentiel en vision par ordinateur et en intelligence artificielle, en raison de ses nombreuses applications, notamment dans l'interaction homme-machine, la vidéosurveillance, la réalité augmentée et l'analyse du comportement humain. Pour garantir l'efficacité et la fiabilité de tels systèmes, il est crucial d'assurer un traitement rigoureux et optimal des données visuelles dès les premières étapes du processus [13].

Ce chapitre a pour objectif de passer en revue les principales approches existantes en matière de détection et de reconnaissance des expressions faciales. Nous mettrons particulièrement l'accent sur les étapes clés du processus : le prétraitement des données, l'extraction des caractéristiques pertinentes, et la classification des expressions. Ces phases permettent de transformer les images brutes en représentations numériques exploitables, conditionnant ainsi la performance globale du système. Enfin, nous présenterons la conception de notre propre approche, en détaillant les choix méthodologiques et technologiques retenus pour chacune des étapes de traitement.

II.2 Travaux connexes

La reconnaissance automatique des expressions faciales (*Facial Expression Recognition*, *FER*) est un domaine de recherche dynamique. Les approches développées peuvent être principalement classées en deux grandes catégories.

Approches basées sur des descripteurs conçus manuellement (classiques) :

Ces méthodes s'appuient sur des caractéristiques prédéfinies, extraites de manière experte pour être robustes aux variations d'éclairage, de pose ou d'orientation du visage. Elles incluent des techniques comme les textures

descriptrices ou les caractéristiques géométriques basées sur des points de repère faciaux.

Approches basées sur l'apprentissage profond (*Deep Learning*) : Fondées sur des architectures de réseaux de neurones profonds, notamment convolutifs (CNN), ces méthodes apprennent automatiquement les caractéristiques les plus discriminantes à partir des données brutes. Leur principal avantage réside dans leur capacité à optimiser conjointement l'extraction de caractéristiques et la classification, atteignant souvent des performances de pointe. Cependant, leur efficacité est conditionnée par la disponibilité de très grandes bases de données annotées.

Ces différentes méthodologies s'intègrent dans un processus global qui comprend généralement le prétraitement des images, l'extraction des caractéristiques, la classification et la prise de décision. Aujourd'hui, plusieurs entreprises comme Affectiva, NeurodataLab ou Microsoft Azure commercialisent des solutions basées sur ces technologies. Toutefois, leur efficacité demeure limitée dans des conditions non contrôlées, (*in the wild*), où les mouvements de tête, les occultations partielles du visage ou les variations d'éclairage sont fréquents.

Face aux défis posés par les réseaux de neurones profonds (DNN), tels que leur grand nombre de paramètres, la nécessité d'un réglage fastidieux des hyperparamètres et leur dépendance à de vastes ensembles de données, des recherches visent à concevoir des modèles plus efficaces. Dans ce contexte, Jeong *et al.* [14] ont proposé un modèle léger et efficace nommé *Lightweight Multilayer Random Forest* (LMRF). Ce modèle repose sur une structure à deux couches avec un nombre réduit d'arbres de décision, permettant une reconnaissance rapide. Leur système utilise d'abord la bibliothèque DLib pour détecter le visage et en extraire les points de repère. Des caractéristiques géométriques, basées sur des rapports de distance et des relations angulaires, sont ensuite calculées. Ces caractéristiques alimentent le modèle LMRF pour la classification de l'expression. Les expériences menées sur les bases de données de référence CK+, MMI et KMU-FED ont montré que le LMRF atteint des performances comparables à celles des DNN tout en étant beaucoup moins complexe. Il s'exécute également plus rapidement sur un processeur standard (CPU), le rendant adapté aux systèmes temps réel. Le modèle a atteint des taux de reconnaissance de 93,4 % sur CK+, 77,4 % sur MMI et 95,1 % sur KMU-FED.

Un champ d'application majeur de la FER est la sécurité routière, où l'analyse des émotions du conducteur peut permettre de prévenir des comportements à risque. Patil *et al.* [15] ont étudié l'impact de cinq émotions (colère, peur, joie, neutralité et tristesse) sur le comportement de conduite. Pour les détecter, ils ont proposé une approche combinant deux types de descripteurs : des caractéristiques géométriques issues des repères faciaux (extraits via DLib) et des caractéristiques

texturales extraites à l'aide de la méthode des motifs binaires locaux (LBP). Après détection du visage avec un classifieur de Haar, ces deux ensembles de caractéristiques sont fusionnés pour former un vecteur unique, qui est ensuite classifié par une machine à vecteurs de support (SVM). Évaluée sur la base de données CK+, leur méthode a atteint une précision de 86,7 %.

Dans une optique similaire de prévention, Azman *et al.* [16] ont développé un système temps réel visant spécifiquement à détecter la colère chez les conducteurs. Le système utilise la méthode de Viola-Jones pour la détection de visage et un classifieur SVM entraîné sur la base de données JAFEE pour la reconnaissance de la colère. Une webcam capture en continu les images du conducteur et, en cas de détection de colère, une alerte sonore est déclenchée. Le système a affiché une précision moyenne de 97 % dans leurs conditions expérimentales.

Un défi majeur dans les scénarios de conduite est l'occultation partielle du visage. Pour y remédier, Sudha *et al.* [17] ont proposé une méthode nommée Reconnaissance des Émotions Faciales du Conducteur (REFC). L'approche calcule d'abord le flux optique entre deux images consécutives à l'aide de l'algorithme de Farneback. Un optimiseur multi-univers parallèle (OMUP) est ensuite utilisé pour reconstruire les parties manquantes du flux optique. Le flux reconstruit est enfin analysé par un réseau de neurones convolutif profond (VGGNet) pour la classification. Sur les bases de données CK+ et KMU-FED, cette méthode a atteint des précisions de 95,92 % et 94,78 % respectivement.

Conscients que la plupart des approches sont validées en laboratoire, Xiao *et al.* [18] ont proposé FERDERnet, un système modulaire conçu pour fonctionner dans des conditions de conduite réelles. Il se compose de trois entités : un module de détection faciale (FD), un module de rééchantillonnage basé sur l'augmentation de données (ABR) pour équilibrer les classes d'émotions, et un module de reconnaissance des émotions (ER). En testant plusieurs architectures de CNN comme GoogLeNet, ResNet50 et InceptionV3, ils ont obtenu les meilleurs résultats avec Xception, atteignant une précision de 96,6 %.

Pour mieux capturer la nature dynamique des émotions, Varma *et al.* [19] ont développé une méthode hybride d'apprentissage profond qui traite à la fois les informations spatiales et temporelles à partir de vidéos. Contrairement aux approches analysant des séquences entières, leur modèle classifie les émotions image par image. L'architecture utilise un réseau d'alignement facial (FAN) pour extraire les visages, un CNN léger (SqueezeNet) pour les encoder, et un réseau de neurones récurrents bidirectionnel à mémoire longue et courte (Bi-LSTM) pour apprendre les dépendances temporelles. Validé sur les bases de données KMU-FED, DMD et un jeu de données expérimental, le modèle a montré d'excellentes performances, avec un score F1 de 0,958 sur KMU-FED.

Les recherches en reconnaissance automatique des expressions faciales ont connu

une forte accélération depuis les années 1990, succédant aux travaux précurseurs de Suwa *et al.* en 1978 [20]. Cette progression a été largement favorisée par les avancées dans la détection de visage [21] et la création de bases de données de référence. La publication de la base CK [22] en 2000, puis de sa version étendue CK+ [23] en 2010, a permis d'établir des protocoles d'évaluation standardisés.

Dans des environnements contrôlés (bon éclairage, vue frontale, sans occultation), les systèmes automatiques ont démontré des performances très élevées [24, 25, 26]. Cependant, comme le souligne la littérature récente, un fossé important subsiste entre les résultats obtenus en laboratoire et l'efficacité des systèmes dans des conditions réelles et non contraintes. Le contexte de la conduite automobile, avec ses défis uniques (variations d'éclairage, occultations, mouvements), représente un banc d'essai particulièrement pertinent pour développer et valider des approches plus robustes.

II.3 Méthodologie proposée

L'architecture de notre système de reconnaissance des expressions faciales est illustrée par la figure II.1. Le système opère en deux phases distinctes : l'apprentissage (entraînement) et le test. Au sein de chaque phase, trois étapes séquentielles sont exécutées.

1. Prétraitement du visage : la détection initiale des régions faciales dans les images d'entrée est effectuée à l'aide de l'algorithme standard de détection de visages de Viola-Jones.
2. Extraction de caractéristiques : après le prétraitement, des caractéristiques discriminantes sont extraites à l'aide de deux nouveaux descripteurs locaux : MBH-LOOP (histogramme multi-blocs de motifs orientés locaux) et MBSF-LOOP (caractéristiques statistiques multi-blocs – motifs orientés locaux).
3. Classification : les vecteurs de caractéristiques extraits sont ensuite fournis à un classifieur de machines à vecteurs de support (SVM), une technique d'apprentissage automatique bien établie et robuste, choisie pour son efficacité dans les espaces de grande dimension, afin de prédire l'étiquette d'expression faciale correspondante.

La performance globale de notre méthodologie a été rigoureusement évaluée à l'aide d'un ensemble de métriques de performance standards. Une description détaillée de chaque étape suit dans cette section.

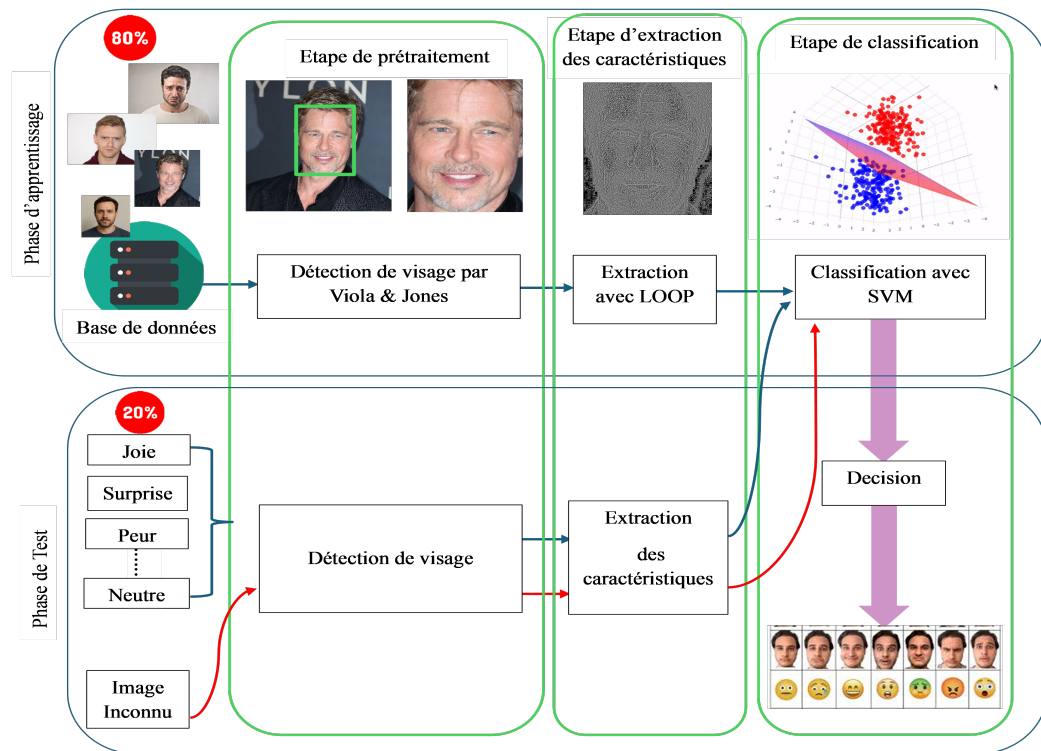


FIGURE II.1 – Architecture globale du système de reconnaissance automatique des expressions faciales

II.3.1 Prétraitement

Le prétraitement constitue une étape cruciale dans tout système de reconnaissance des expressions faciales. Il a pour objectif d'améliorer la qualité des données brutes en supprimant le bruit et en mettant en évidence les régions pertinentes, afin de les préparer efficacement pour les phases suivantes d'extraction de caractéristiques et de classification. Les techniques de prétraitement varient selon le type de données analysées [27].

Dans le cadre de ce travail, nous nous concentrerons particulièrement sur les méthodes appliquées aux images faciales. Une étape préalable de détection de visage est effectuée afin d'isoler automatiquement la région faciale pertinente dans chaque image, garantissant ainsi que seules les zones d'intérêt seront traitées lors des phases ultérieures.

La détection de visage est une étape fondamentale dans notre système, consistant à identifier et localiser automatiquement la présence d'un visage humain dans une image ou une séquence vidéo. Avant de pouvoir analyser ou reconnaître une expression faciale, il est indispensable de détecter d'abord la région du visage, c'est-à-dire de délimiter avec précision la zone d'intérêt sur laquelle les traitements ultérieurs seront appliqués. Cette étape permet d'encadrer le visage détecté au moyen d'un rectangle, facilitant ainsi l'extraction des caractéristiques faciales pertinentes.

Pour réaliser cette tâche, nous avons utilisé un détecteur de visage à la fois ra-

pide et robuste, basé sur l'algorithme développé par Paul Viola et Michael Jones en 2001 [28]. Ce détecteur repose sur l'utilisation des descripteurs de Haar, combinés à un classifieur en cascade, permettant une détection efficace en temps réel, même dans des conditions variées. Ce choix technologique assure un bon compromis entre précision et rapidité d'exécution. La figure suivante illustre la détection du visage sur une image (figure II.2).

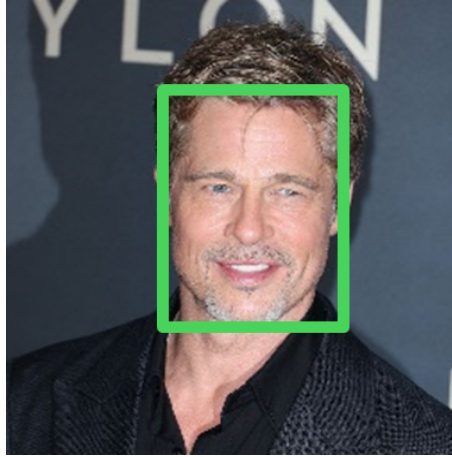


FIGURE II.2 – Un exemple de la détection de visage avec Viola et Jones

Les descripteurs de Haar : Les descripteurs de Haar sont utilisés pour détecter des objets ou des zones d'intérêt dans une image, notamment en temps réel. Ce sont des motifs rectangulaires simples composés de zones noires et blanches, utilisés pour en extraire des caractéristiques visuelles pertinentes. Plutôt que d'analyser les pixels individuellement, ces descripteurs s'appuient sur les écarts de contraste entre des groupes adjacents de pixels pour identifier les régions claires et sombres d'un visage, et pour repérer des structures caractéristiques, telles que les yeux, le nez ou la bouche (voir la figure II.3).

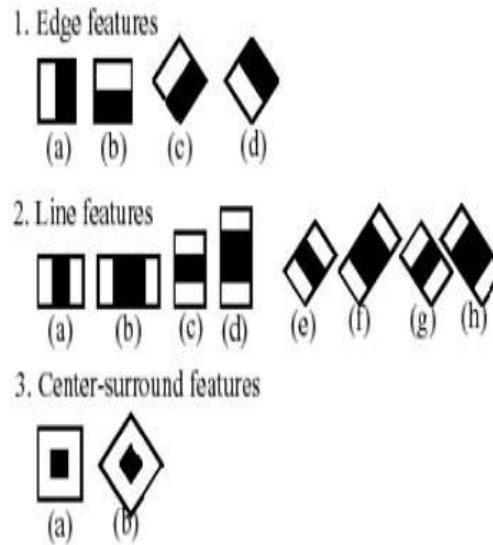


FIGURE II.3 – Les motifs des descripteurs de Haar

La cascade de classifieurs : Pour tirer pleinement parti des descripteurs, on les intègre dans une cascade de classifieurs, une méthode d'apprentissage supervisée particulièrement efficace pour la détection d'objets. Le principe repose sur l'entraînement du système à l'aide d'un grand nombre d'images positives (contenant un visage) et négatives (ne contenant pas de visage). Cette cascade est construite à l'aide de l'algorithme AdaBoost, qui sélectionne les descripteurs de Haar les plus pertinents et les combine pour former des classifieurs puissants. Grâce à sa structure hiérarchique, le système peut éliminer rapidement les zones non pertinentes, ne gardant que celles susceptibles de contenir un visage pour des analyses plus détaillées. Ce processus rend la détection à la fois rapide, précise et adaptée aux contraintes du traitement en temps réel.

II.3.2 Extraction des caractéristiques

L'extraction des caractéristiques constitue une étape essentielle dans tout système de reconnaissance automatique des expressions faciales. Elle vise à transformer les images prétraitées en vecteurs de caractéristiques, capables de représenter efficacement les variations d'expressions, et prêtes pour l'étape suivante de classification. Plusieurs méthodes ont été proposées à cet effet, parmi lesquelles on peut citer les descripteurs comme : Local Binary Patterns (LBP), Local Phase Quantization (LPQ) et Local Optimal Oriented Pattern (LOOP)).

Dans le cadre de notre travail, nous avons choisi d'utiliser le descripteur LOOP en raison de sa capacité à capturer de manière robuste les structures directionnelles présentes dans les images faciales, de sa robustesse et de son adaptation aux variations d'éclairage et de bruit.

Le descripteur LOOP repose sur une amélioration des méthodes LBP et LDP. Un bref rappel de ces deux approches est présenté ci-dessous.

Local Binary Pattern (LBP) : Le descripteur LBP [29] est largement utilisé pour capturer les variations locales d'intensité des pixels au sein d'une image. Il est reconnu pour sa simplicité et sa capacité à bien distinguer les différentes textures, ce qui en fait un outil efficace pour diverses applications de reconnaissance. Son expression est la suivante :

$$\text{LBP}(x_c, y_c) = \sum_{p=0}^{P-1} s(i_p - i_c) \cdot 2^p \quad (\text{II.1})$$

où :

$$s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (\text{II.2})$$

- i_c est l'intensité du pixel central à la position (x_c, y_c) .
- i_p est l'intensité des P pixels voisins.
- $s(x)$ est la fonction de seuillage binaire.

Local Directional Pattern (LDP) : Le descripteur LDP est une extension du précédent (LBP). Il introduit une composante directionnelle en utilisant les masques de Kirsch, qui sont des filtres orientés pour détecter les bords. Grâce à cette orientation directionnelle, le LDP s'avère plus résistant au bruit tout en fournissant une représentation plus riche des motifs locaux [30]. Son expression est la suivante :

$$\text{LDP}(x_c, y_c) = \sum_{k=0}^{K-1} s(m_k - T) \cdot 2^k \quad (\text{II.3})$$

où :

$$s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (\text{II.4})$$

- m_k est la réponse du filtre de Kirsch dans la k^{e} direction.
- T est un seuil fixé (souvent le k^{e} plus grand maximum).
- $K = 8$, car il y a 8 directions.

II.3.2.1 Local Optimal-Oriented Pattern (LOOP)

La méthode LOOP a été proposée pour surmonter les limitations des descripteurs LBP et LDP. En effet, dans le cas de LBP, l'utilisation de poids binaires attribués de manière arbitraire selon l'ordre des pixels voisins, rend la méthode sensible à l'orientation, ce qui limite sa robustesse aux rotations. De son côté, LDP repose sur l'introduction d'un paramètre de seuil fixé empiriquement, ce qui

réduit le nombre de motifs possibles et impose une contrainte artificielle sur le nombre de bits pouvant prendre la valeur 1.

LOOP se présente comme une fusion non linéaire de LBP et LDP, permettant de préserver leurs avantages tout en éliminant leurs inconvénients. L'idée est d'utiliser à la fois les variations d'intensité locales (comme dans LBP) et les informations directionnelles issues des masques de Kirsch (comme dans LDP), mais de manière plus flexible et optimisée.

Concrètement, pour chaque pixel central i_c situé en (x_c, y_c) , on considère ses huit voisins i_n pour $n = 0, 1, \dots, 7$ dans une fenêtre de 3×3 . On applique les huit masques de Kirsch, des filtres détectant les bords dans différentes directions, aux pixels voisins afin de mesurer la force de variation de l'intensité dans chaque direction. Chaque réponse de filtre de Kirsch m_n correspondant à chaque i_n , $n = 0, \dots, 7$ est alors classée selon son intensité, ce qui permet d'attribuer un poids exponentiel $w_n \in \{0, 1, \dots, 7\}$ à chaque pixel voisin, en fonction du rang de m_n parmi les huit directions. Cette pondération reflète l'importance relative de chaque direction locale autour du pixel central. La valeur LOOP est ensuite calculée à l'aide de la formule suivante :

$$\text{LOOP}(x_c, y_c) = \sum_{n=0}^7 (s(i_n - i_c) \cdot 2^n) \quad (\text{II.5})$$

où :

$$s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (\text{II.6})$$

Ainsi, pour chaque voisin, si son intensité est supérieure ou égale à celle du pixel central, on ajoute 2^n à la valeur finale : sinon on n'ajoute rien.

Grâce à cette stratégie, LOOP permet de :

- **Assurer une invariance à la rotation**, en adaptant dynamiquement les poids selon les directions dominantes.
- **Éviter les seuils arbitraires**, comme le paramètre k utilisé dans LDP.
- **Capturer des motifs discriminants plus riches**, en mettant l'accent sur les directions les plus significatives autour de chaque pixel.

Ce descripteur se révèle donc particulièrement adapté pour l'analyse fine des expressions faciales, notamment dans des conditions variables d'éclairage et d'orientation [30].

II.3.2.2 La méthode MBH-LOOP

Le MBH-LOOP (Multi Bloc Histogram Local Optimal Oriented Pattern) est l'une des méthodes utilisées pour l'extraction des caractéristiques avec le descripteur LOOP, souvent utilisé en reconnaissance d'images. Il sert à convertir les

valeurs locales des descripteurs calculés pixel par pixel en un vecteur de caractéristiques global. Ensuite, un histogramme est constitué, qui enregistre la fréquence d'apparition de chaque motif LOOP potentiel dans une image ou une région, utile pour des tâches de classification, de reconnaissance d'images ou d'analyse de textures.

Ce dernier comptabilise la fréquence d'apparition de chaque code LOOP dans l'image ou dans des zones spécifiques de celle-ci. Si l'image est segmentée en sous-blocs, un histogramme est calculé individuellement pour chaque sous-bloc. L'axe horizontal de cet histogramme montre les différentes valeurs possibles des motifs LOOP, tandis que l'axe vertical indique le nombre de fois que chaque motif se manifeste dans la zone concernée.

L'ensemble de ces histogrammes (un pour chaque sous-bloc ou un global pour l'image entière) est ensuite concaténé pour créer un vecteur de caractéristiques global. Ce vecteur représente de manière compacte et discriminatoire la structure locale de l'image via les motifs orientés détectés. Il peut alors servir d'entrée pour des algorithmes de classification ou de reconnaissance, facilitant ainsi l'identification d'objets, ou de classes d'images avec une grande précision [29].

On retrouve l'équation de calcul de l'histogramme : Soit $h(k)$ la fréquence du motif LOOP $k \in [0, 255]$, alors :

$$h(k) = \sum_{(x,y) \in R} \delta(\text{LOOP}(x, y) - k) \quad (\text{II.7})$$

où :

$$\delta(a) = \begin{cases} 1 & \text{si } a = 0 \\ 0 & \text{sinon} \end{cases} \quad (\text{II.8})$$

Tous les histogrammes des blocs sont concaténés en un vecteur final de caractéristiques :

$$F = [h1, h2, \dots, hm] \in R^{256 \times m} \quad (\text{II.9})$$

La figure suivante indique les étapes décrites précédemment :

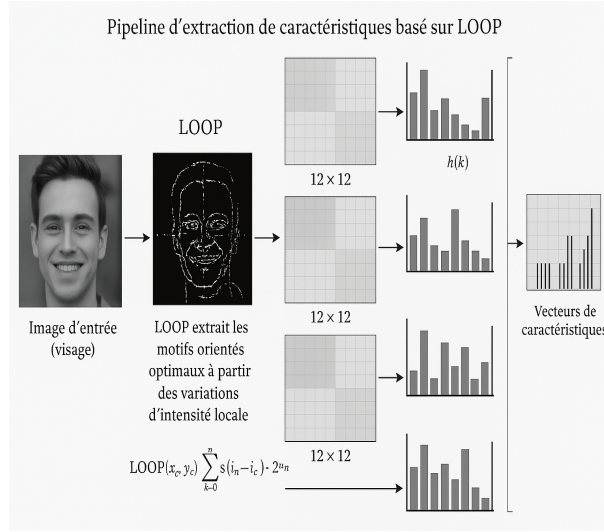


FIGURE II.4 – Étapes de transformation d'une image en un vecteur de caractéristiques à l'aide du descripteur LOOP

II.3.2.3 La méthode MBSF-LOOP

Le MBSF-LOOP (Multi Bloc Statistiques features Local Optimal Oriented Pattern) est une technique d'extraction de caractéristiques utilisée avec le descripteur LOOP. Après avoir effectué le calcul via le descripteur LOOP, une autre technique fréquemment adoptée pour extraire des caractéristiques globales est l'approche statistique. Cette méthode résume l'ensemble des valeurs des descripteurs locaux grâce à des mesures statistiques globales après avoir découpé l'image en sous-blocs. On en déduit des paramètres statistiques sur l'ensemble des codes ou des réponses dérivées de l'image entière ou des sous-blocs définis. Les mesures les plus couramment employées incluent :

- • **La moyenne**, qui indique la valeur centrale des descripteurs.

Son expression est :

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (\text{II.10})$$

- • **L'écart type (ou la variance)**, qui évalue la dispersion des valeurs autour de la moyenne.

Son expression est :

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (\text{II.11})$$

- • **L'asymétrie (skewness)**, qui montre si la distribution est déséquilibrée.

Son expression est :

$$\text{Skewness} = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^3 \quad (\text{II.12})$$

- • **La kurtosis**, qui détermine l'aplatissement ou la concentration de la distribution.

Son expression est :

$$\text{Kurtosis} = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^4 \quad (\text{II.13})$$

Et parfois d'autres mesures, comme **le minimum**, **le maximum**, ou **l'énergie**.

Dans ces expressions, x_i représente la valeur du code LOOP au pixel i ,

N est le nombre total de pixels (ou de codes LOOP),

μ est la moyenne,

σ est l'écart-type.

Ces mesures synthétisent la distribution des codes LOOP dans l'image sans se focaliser sur leur fréquence exacte, mais plutôt sur leur répartition générale [31].

En combinant ces statistiques, on crée un vecteur de caractéristiques global qui illustre le comportement global des motifs extraits par le descripteur sur l'image. Cette méthode est souvent plus concise qu'un histogramme et peut se révéler particulièrement efficace, surtout lorsque la distribution des motifs revêt une signification importante, comme dans la classification ou la détection des expressions faciales.

II.3.3 Étape de normalisation

La normalisation vient à chaque phase d'extraction des caractéristiques avec le descripteur LOOP, que ce soit avec la méthode histogramme ou statistiques. Elle est essentielle pour s'assurer que les vecteurs de caractéristiques sont comparables, résistants et discriminatoires. Cette dernière permet de :

- **Comparer les distributions de motifs**, peu importe le nombre total de pixels.
- **Limiter l'impact de l'éclairage ou du contraste** : puisque des images présentant des intensités variées (plus sombres ou plus lumineuses), peuvent générer des vecteurs de caractéristiques LOOP légèrement différents. Alors cette étape permet d'atténuer l'influence des variations générales de luminosité sur les motifs.
- **Harmoniser l'échelle des caractéristiques** : En effet, les classifieurs (SVM, k-NN, ...) sont sensibles aux échelles des données, donc si un histogramme n'est pas normalisé, les valeurs dominantes peuvent perturber l'apprentissage ou la mesure de distance.
- **Optimiser les performances de classification** : C'est-à-dire que des vecteurs normalisés facilitent l'utilisation des mesures de distance en garantissant

sant que chaque dimension a une contribution équitable. Cela mène souvent à une amélioration de la précision et de la stabilité des modèles.

La normalisation s'opère par l'équation suivante :

$$h_{\text{norm}}(k) = \frac{h(k)}{\sum_{j=0}^{255} h(j)}, \quad \text{pour } k = 0, 1, \dots, 255 \quad (\text{II.14})$$

II.3.4 Classification et prise de décision

La classification : une fois que les caractéristiques pertinentes ont été extraites du visage, l'étape suivante dans un système de reconnaissance automatique des expressions faciales consiste à classifier l'expression détectée. Cela signifie attribuer une étiquette émotionnelle (telle que joie, colère, tristesse, surprise, etc.) à partir des informations visuelles obtenues.

Il est essentiel de concevoir des mécanismes de classification efficaces pour différencier les expressions en fonction des caractéristiques extraites. Parmi les techniques couramment utilisées, on peut citer les machines à vecteurs de support (SVM), les k-plus proches voisins (KNN), les arbres de décision (DT), ainsi que les forêts aléatoires (Random Forest, RF). Ces méthodes se distinguent par leur complexité, leur capacité de généralisation et leur performance selon la nature des données utilisées. Tous ces mécanismes sont des techniques d'apprentissage automatique supervisé.

II.3.4.1 Les machines à vecteurs de support (SVM)

Les machines à vecteurs de support (SVM), sont largement utilisées comme algorithme de classification en raison de leur capacité à bien séparer différentes catégories d'expressions. Le principe de base des SVM consiste à trouver un hyperplan optimal qui sépare les données de différentes classes (par exemple, "joie" vs "colère") de manière à maximiser la marge, c'est-à-dire la distance entre cet hyperplan et les points les plus proches de chaque classe, appelés vecteurs de support.

Un hyperplan avec une large marge est généralement synonyme de meilleure capacité de généralisation du modèle, ce qui permet de mieux reconnaître des expressions faciales inédites ou légèrement différentes. À l'inverse, une marge étroite rend le modèle plus sensible aux variations ou au bruit présent dans les données, ce qui peut nuire à sa robustesse.

Ainsi, l'objectif des SVM est de trouver le séparateur optimal qui maximise cette marge entre les classes, afin d'obtenir une classification plus précise et fiable des expressions faciales, même dans des contextes complexes ou peu structurés [32]. La figure II.5 illustre des exemples de classification avec le SVM.

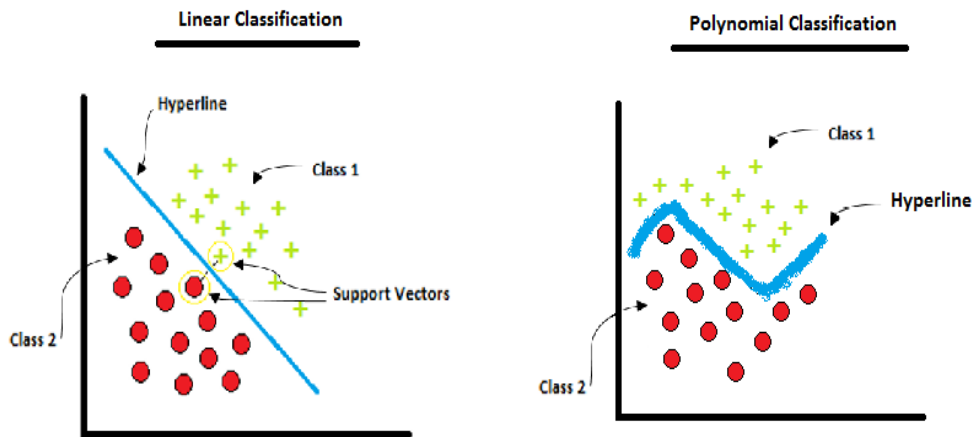


FIGURE II.5 – Exemples de séparation avec le classifieur SVM

II.3.4.2 La forêt aléatoire (RF)

La forêt aléatoire (RF) représente un algorithme d'apprentissage supervisé qui intègre plusieurs arbres de décision afin d'améliorer la précision des prévisions. Chaque arbre est formé de manière indépendante sur des sous-ensembles aléatoires des données d'apprentissage, suivant la méthode du bagging. Pour la création de chaque arbre, un sous-ensemble aléatoire de caractéristiques est utilisé, conformément au principe des "projections aléatoires". Les prédictions finales sont établies en faisant la moyenne des résultats de tous les arbres, ce qui diminue la variance des prévisions. En résumé, la forêt aléatoire regroupe divers avis pour mieux comprendre un problème donné. C'est un classifieur performant, sans grande exigence en prétraitement des données [33]. La figure suivante illustre la classification RF.

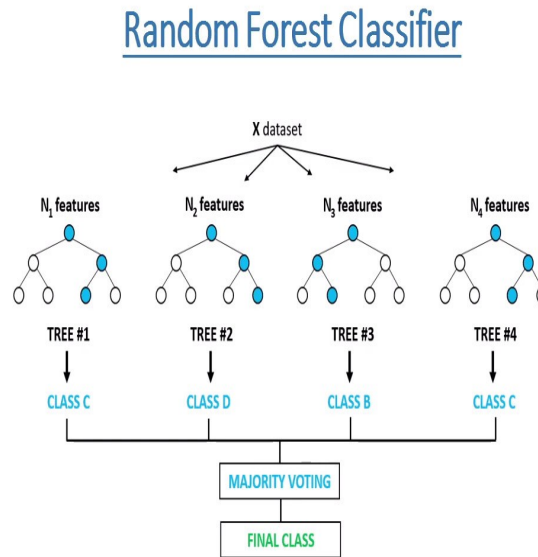


FIGURE II.6 – Principe de la classification Random Forest

II.3.4.3 L'arbre de décision (DT)

L'arbre de décision (DT) constitue un algorithme d'apprentissage supervisé qui permet de visualiser et de clarifier les décisions ainsi que le processus décisionnel relatif aux problèmes de classification et de régression. Il propose une représentation accessible et interprétable, tout en étant capable de générer des règles de classification logiques. Les arbres de décision sont fréquemment employés dans le domaine de la classification, en raison de leur aptitude à aborder des problèmes complexes. Un arbre de décision se compose de :

- Nœuds décisionnels, chacun contenant un test sur un attribut.
- Branches correspondant généralement à une valeur potentielle de l'attribut choisi.
- Feuilles regroupant les objets appartenant à la même catégorie.

L'application des arbres de décision dans les problèmes de classification se déroule en deux étapes clés :

- La création d'un arbre de décision basé sur un ensemble d'apprentissage.
- La classification ou l'inférence, qui consiste à attribuer une nouvelle instance en utilisant l'arbre de décision élaboré lors de la première étape [34].

II.3.4.4 K-Nearest Neighbors (KNN)

Le classifieur K-NN est un algorithme d'apprentissage supervisé couramment utilisé pour la classification. Il repose sur un principe simple mais efficace : attribuer une classe à une instance inconnue en se basant sur les k exemples les plus

proches d'elle dans l'espace des caractéristiques, selon une mesure de similarité. La classe assignée est généralement celle qui est la plus représentée parmi ces k voisins — on parle alors de vote majoritaire [35].

Dans le contexte de la reconnaissance automatique des expressions faciales, K-NN peut être utilisé pour associer un vecteur de caractéristiques extrait d'un visage à l'une des classes d'expression (joie, tristesse, surprise, etc.), en comparant ce vecteur aux exemples déjà étiquetés dans la base d'apprentissage.

K-NN est apprécié pour sa simplicité de mise en œuvre et sa capacité à gérer aussi bien des problèmes de classification que de régression. Il est fréquemment utilisé dans des applications industrielles, notamment lorsque la structure des données est complexe ou difficile à modéliser mathématiquement.

La performance de K-NN dépend fortement du choix de la mesure de distance utilisée pour définir la proximité entre les données. Parmi les mesures les plus courantes, on trouve : la distance euclidienne (la plus utilisée), la distance de Manhattan, la distance de Minkowski, et la distance de Hamming, adaptée aux variables catégorielles [35].

Le choix du nombre de voisins (k) et de la distance impacte directement la précision du classifieur. Un bon réglage de ces paramètres est essentiel pour obtenir des résultats fiables, notamment dans des contextes sensibles comme la reconnaissance faciale.

La figure suivante illustre le mécanisme de classification KNN.

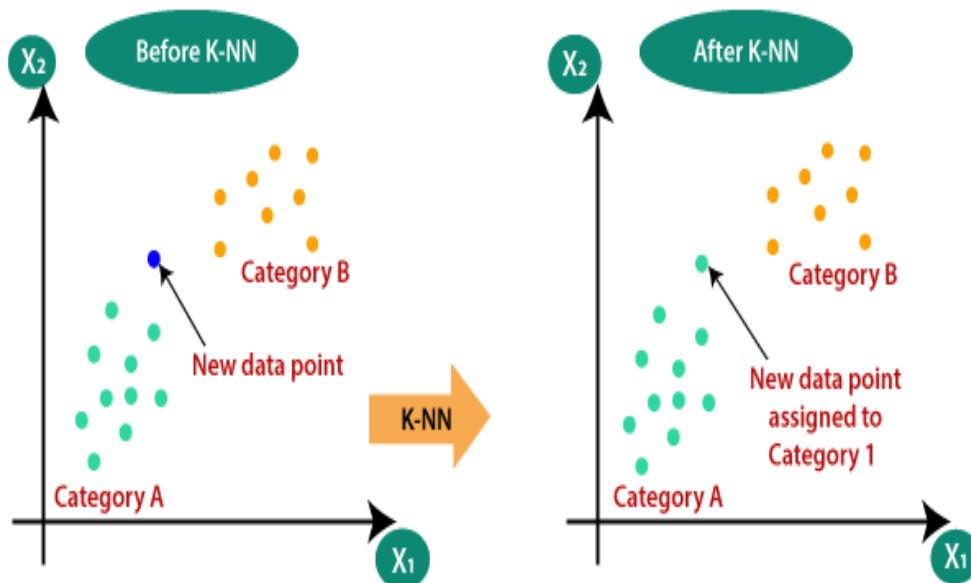


FIGURE II.7 – Un exemple de mécanisme de classification avec KNN

Prise de décision : Dans le cadre de la reconnaissance automatique des expressions faciales à partir d'images, la phase de décision constitue l'étape finale du

processus. Elle intervient après l'extraction des caractéristiques et la classification, et a pour objectif de valider l'expression faciale reconnue sur l'image analysée.

Concrètement, le système compare les caractéristiques extraites de l'image test, à celles contenues dans les modèles d'expressions faciales préalablement appris (comme 'joie', 'colère', 'tristesse', etc.). Cette comparaison produit une valeur scalaire, qui mesure à quel point l'image test est proche d'une classe particulière d'expression.

Deux cas peuvent se présenter :

- **Dans une tâche d'identification**, cette valeur permet de déterminer directement à quelle expression appartient l'image, en choisissant la classe ayant la plus forte similarité.
- **Dans une tâche de vérification**, plus rare dans ce contexte, la décision est binaire : il s'agit de valider ou non si l'expression faciale de l'image correspond à une classe spécifiée à l'avance.

La prise de décision repose généralement sur un seuil fixé à l'avance. Si la similarité entre l'image test et une expression du modèle dépasse ce seuil, l'expression est considérée comme reconnue. Si elle est inférieure, le système rejette cette correspondance.

Ainsi, cette étape assure que l'expression détectée est cohérente et fiable selon les critères définis par le système de reconnaissance [36].

II.4 Conclusion

Le présent chapitre visait à fournir un aperçu des solutions mises en œuvre jusqu'à ce jour pour la reconnaissance des expressions faciales, en insistant tout particulièrement sur les étapes cruciales du processus : le prétraitement, l'extraction de caractéristiques et la classification. Nous avons d'abord illustré l'importance du prétraitement pour améliorer la qualité des données visuelles, notamment grâce à la détection automatique des visages via l'algorithme de Viola-Jones. Par la suite, nous avons présenté plusieurs techniques d'extraction de caractéristiques locales, en mettant l'accent sur le descripteur LOOP et ses deux variantes (MBH-LOOP et MBSF-LOOP), qui sont capables de détecter avec précision les motifs directionnels des visages. Enfin, nous avons discuté des méthodes de classification telles que SVM, KNN, RF et DT, qui permettent de différencier efficacement les diverses expressions émotionnelles à partir des vecteurs extraits. Dans le chapitre suivant, nous aborderons la conception de notre système avec une validation numérique des approches que nous adopterons.

Expérimentations et Résultats

III.1 Introduction

Dans le domaine de la vision par ordinateur, la reconnaissance automatique des expressions faciales suscite un intérêt croissant, notamment grâce aux avancées récentes en apprentissage automatique. De nombreux chercheurs s'efforcent d'améliorer la performance des systèmes en explorant différentes méthodes d'extraction de caractéristiques et de classification.

Dans ce dernier chapitre, nous présentons la mise en œuvre complète de notre système de reconnaissance automatique des expressions faciales. Ce système repose exclusivement sur l'analyse d'images faciales. Afin de garantir la robustesse et la fiabilité du modèle proposé, nous avons mené une série d'expériences approfondies sur deux bases de données d'images faciales représentatives d'expressions faciales variées.

Ce chapitre détaille ainsi l'ensemble des étapes de conception avec les techniques utilisées dans la détection de visage, l'extraction des caractéristiques et la classification. Les réglages expérimentaux, ainsi que les résultats obtenus, visent à valider notre système de reconnaissance automatique des expressions faciales basé sur l'apprentissage automatique.

III.2 Environnement de travail

La langage de programmation utilisé dans cette étude est MATLAB, émulé par son environnement de programmation (dans notre cas, MATLAB 2024b) et conçu par The MathWorks. MATLAB permet d'implémenter des algorithmes de manière aisée et rapide, d'exécuter des tâches nécessitant une grande puissance de calcul, de manipuler et de visualiser des courbes, ainsi que de concevoir des interfaces graphiques. Les expérimentations ont été menées sur un ordinateur équipé d'un processeur Intel® Core™ i7-10750H, de fréquence 2,6 GHz avec 16 Go de RAM.

III.3 Protocole

La base de données utilisée dans notre système de reconnaissance automatique des expressions faciales a été divisée en deux ensembles distincts : 80 % des données ont été réservées à l'apprentissage du modèle, tandis que les 20 % restantes ont été utilisées pour les tests. Ce choix de répartition est justifié par de nombreuses recherches en intelligence artificielle et en apprentissage automatique, qui recommandent cette proportion comme étant la plus équilibrée et la plus efficace pour évaluer les performances d'un système de classification supervisé, notamment en reconnaissance faciale [27] et [37]. Dans notre cas, nous avons appliqué cette répartition à deux bases de données bien connues dans la littérature : CK+ et

JAFFE. Chacune de ces bases comprend sept classes d'expressions faciales (telles que la joie, la colère, la tristesse, la peur, la surprise, le dégoût et le neutre), représentées par un certain nombre d'images par classe. Cette organisation permet de simuler un système réaliste d'apprentissage supervisé, où le modèle est entraîné sur des données d'apprentissage et testé sur des données inconnues afin d'évaluer sa capacité.

III.4 Métriques d'évaluation utilisées

Pour évaluer quantitativement les performances de notre modèle de classification des expressions faciales, nous nous appuyons sur un ensemble de métriques standards issues de la matrice de confusion.

III.4.1 La matrice de confusion

La matrice de confusion est un tableau qui synthétise les résultats d'un classificateur. Elle croise les classes réelles (vérité terrain) avec les classes prédites par le modèle, permettant une analyse détaillée des erreurs de classification. Chaque cellule (i, j) de la matrice indique le nombre d'échantillons de la classe réelle i qui ont été prédits comme appartenant à la classe j .

La diagonale de la matrice contient les prédictions correctes, tandis que les éléments hors diagonale représentent les erreurs. À partir de cette matrice, on définit quatre termes fondamentaux pour une classe donnée :

Vrais Positifs (VP ou TP – True Positives) : Instances de la classe positive correctement prédites comme positives.

Faux Positifs (FP – False Positives) : Instances de la classe négative incorrectement prédites comme positives (erreur de type I).

Vrais Négatifs (VN ou TN – True Negatives) : Instances de la classe négative correctement prédites comme négatives.

Faux Négatifs (FN – False Negatives) : Instances de la classe positive incorrectement prédites comme négatives (erreur de type II).

Visuellement, pour une classe C , la matrice se structure comme suit :

		Classe Prédite	
		Classe C	Non-C
Classe	Classe C	VP	FN
Réelle	Non-C	FP	VN

L'analyse de la matrice de confusion est cruciale car elle permet non seulement de calculer un score de performance global, mais aussi d'identifier les confusions systématiques entre certaines expressions (par exemple, la peur et la surprise).

III.4.2 Métriques dérivées de la matrice de confusion

À partir des valeurs de VP, FP, VN et FN, plusieurs métriques de performance sont calculées.

Exactitude (Accuracy) : Mesure de la proportion de prédictions correctes sur l'ensemble des prédictions. C'est la métrique la plus intuitive, mais elle peut être trompeuse lorsque les classes sont déséquilibrées.

$$\text{Exactitude} = \frac{VP + VN}{VP + VN + FP + FN}$$

Précision (Precision) : Mesure de la fiabilité des prédictions positives. Elle répond à la question : "Parmi les prédictions positives, quelle proportion était correcte?". Une haute précision est cruciale lorsque le coût d'un faux positif est élevé.

$$\text{Précision} = \frac{VP}{VP + FP}$$

Rappel (Recall ou Sensibilité) : Mesure de la capacité du modèle à identifier toutes les instances positives pertinentes. Elle répond à la question : "Quelle proportion des instances réellement positives a été correctement identifiée?". Un rappel élevé est essentiel lorsque le coût d'un faux négatif est important (par exemple, ne pas détecter une expression de détresse).

$$\text{Rappel} = \frac{VP}{VP + FN}$$

Spécificité (Specificity) Évalue la capacité du modèle à identifier correctement les instances négatives. Elle est complémentaire au rappel et mesure la proportion de vrais négatifs.

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

Score F1 (F1-Score) Il s'agit de la moyenne harmonique de la précision et du rappel. Cette métrique offre un compromis équilibré entre les deux, ce qui la rend particulièrement robuste pour évaluer un classifieur, notamment en présence de classes déséquilibrées. Un score de 1 indique une précision et un

rappel parfaits.

$$F_1 = 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} = \frac{2 \times \text{VP}}{2 \times \text{VP} + \text{FP} + \text{FN}}$$

L'utilisation conjointe de ces métriques permet une évaluation complète et nuancée des performances du modèle, en surmontant les limites d'une seule mesure comme l'exactitude.

III.5 Bases de données utilisées

La performance d'un système d'apprentissage automatique dépend intrinsèquement de la qualité et de la pertinence des données utilisées pour son entraînement. Dans le domaine de la reconnaissance automatique des expressions faciales (FER), plusieurs bases de données ont été spécifiquement conçues pour la recherche et l'évaluation des modèles.

Pour les besoins de notre projet, nous avons sélectionné deux bases de données de référence, largement reconnues par la communauté scientifique : CK+ (Extended Cohn-Kanade) et JAFFE (Japanese Female Facial Expression). Le choix de CK+ s'explique par la qualité de ses annotations et par la nature séquentielle de ses données, qui permet une analyse dynamique des expressions [23]. La base JAFFE, quant à elle, offre un ensemble d'expressions prototypiques et posées, particulièrement adapté pour évaluer la capacité d'un modèle à discriminer des expressions clairement définies [38].

III.5.1 La base de données CK+ (*Extended Cohn-Kanade*)

CK+ est l'une des ressources les plus largement utilisées pour l'évaluation des systèmes FER en environnement contrôlé. Elle constitue une extension de la base Cohn-Kanade (CK) originale. CK+ comprend 593 séquences vidéo enregistrées auprès de 123 sujets. Chaque séquence illustre la transition d'une expression faciale neutre vers une expression émotionnelle pleinement exprimée. Les sept émotions de base y sont représentées : joie, colère, tristesse, surprise, peur, dégoût et mépris. Une des particularités majeures de cette base est l'annotation précise de chaque séquence, non seulement en termes d'émotion, mais également selon les unités d'action (Action Units - AUs) définies par le Facial Action Coding System (FACS). Les images, initialement en haute résolution, sont souvent transformées en niveaux de gris et redimensionnées pour les expériences. Dans de nombreux travaux, y compris le nôtre, une résolution de 48×48 pixels en échelle de gris sur 8 bits, est utilisée.

III.5.2 La base de données JAFFE (*Japanese Female Facial Expression*)

La base de données JAFFE est également très utilisée dans la recherche sur la reconnaissance FER. Elle contient 213 images statiques en niveaux de gris, capturant les expressions de 10 femmes japonaises. Chaque participante a exprimé les sept émotions universelles : joie, colère, tristesse, surprise, peur, dégoût, ainsi qu'une expression neutre. Les images ont une résolution de 256×256 pixels, offrant un niveau de détail suffisant pour permettre une analyse fine des caractéristiques faciales. Bien que plus restreinte en taille que CK+, la base JAFFE reste une référence précieuse pour les études menées en environnement contrôlé.

III.6 Expérimentations et résultats

Dans l'ensemble de nos expérimentations, l'évaluation du système de reconnaissance des expressions faciales a été réalisée en mesurant le taux de reconnaissance correcte, défini comme le rapport entre le nombre de prédictions correctes et le nombre total de tests effectués. Afin de mieux juger les performances globales de notre système, nous avons complété cette évaluation par les cinq métriques standards : exactitude (accuracy), précision, rappel (recall), spécificité et F1-score. Le processus expérimental s'est déroulé en plusieurs étapes clés. Pour la détection du visage, nous avons utilisé avec succès le détecteur Viola et Jones, comme le montre la figure suivante.

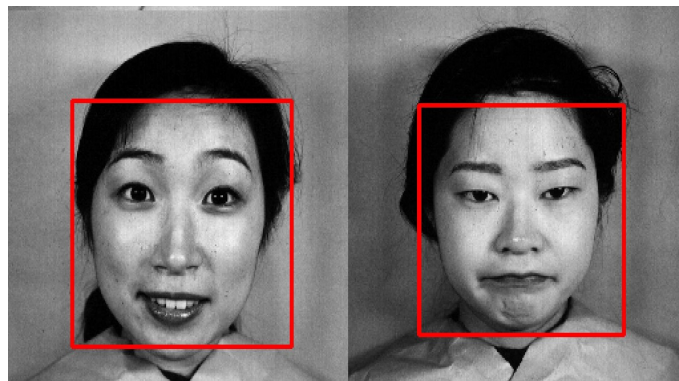


FIGURE III.1 – Détection des visages de la base de données JAFFE

III.6.1 Choix de la technique d'extraction des caractéristiques

Pour l'étape de l'extraction des caractéristiques, nous avons d'abord appliqué deux variantes des descripteurs LOOP (Local Optimal Oriented Pattern) : le LOOP histogramme et le LOOP statistique. Pour une meilleure caractérisation, ces techniques divisent l'image traitée (48×48 pixels) en plusieurs blocs afin de

capturer les motifs orientés locaux. Nous avons utilisé pendant nos tests six tailles de pas différentes.

III.6.1.1 Performances avec le descripteur LOOP Histogramme MBH-LOOP

Dans le but de déterminer le meilleur pas qui donne les meilleurs résultats de taux de reconnaissance (%) du système en utilisant la méthode LOOP Histogramme, nous avons réalisé six séries d'expériences correspondant à six tailles de sous-blocs différentes. En analysant ces résultats obtenus à travers les cinq métriques d'évaluation (accuracy, précision, rappel, spécificité et F1-score), nous avons constaté que le meilleur score a été atteint avec un pas de 10×12 pixels comme montré sur le tableau III.1. Cette configuration s'est avérée être la plus performante pour cette méthode.

Sous-blocs	Accuracy	Precision	Recall	Spécificité	F1-score
48×48	72.44	66.30	65.75	95.34	65.93
24×24	91.83	88.90	87.90	98.63	88.01
24×12	95.40	95.26	90.19	99.23	91.14
16×12	97.44	96.92	93.85	99.58	94.78
12×12	95.91	95.59	92.25	99.27	93.86
10×12	97.44	97.84	95.42	99.55	96.50

TABLE III.1 – Performances de notre système FER à base du descripteur MBH-LOOP.

III.6.1.2 Performances avec le descripteur LOOP Statistiques MBSF-LOOP

Dans le but de déterminer le pas optimal qui donne les meilleurs résultats de taux de reconnaissance (%) du système en utilisant la méthode LOOP Statistiques, nous avons réalisé six séries d'expériences correspondant à six tailles de sous-blocs différentes. En analysant ces résultats obtenus à travers les cinq métriques d'évaluation (accuracy, précision, rappel, spécificité et F1-score), nous avons constaté que le meilleur score a été atteint avec un pas de 12×12 pixels comme montré sur le tableau III.2. Cette configuration s'est avérée être la plus performante pour cette méthode.

D'après les résultats des cinq métriques de performances de notre système FER avec les deux méthodes d'extraction de caractéristiques, LOOP histogramme et LOOP statistique, il ressort clairement que la seconde méthode offre les meilleurs scores globaux notamment avec des sous-blocs de 12×12 pixels.

Sous-blocs	Accuracy	Précision	Rappel	Spécificité	F1-score
48×48	77.04	71.21	71.58	96.13	71.28
24×24	94.89	95.04	92.90	99.13	93.69
24×12	73.46	66.90	65.29	95.47	65.75
16×12	95.91	96.90	93.65	93.65	99.28
12×12	98.97	99.01	98.29	99.83	98.60
10×12	96.93	97.59	94.87	99.91	99.28

TABLE III.2 – Performances de notre système FER à base du descripteur MBSF-LOOP.

III.6.1.3 Comparaison des descripteurs MBSF-LOOP, LPQ et BSIF

Afin de déterminer la meilleure méthode d'extraction, les performances de notre système ont été évaluées avec trois descripteurs. D'abord avec le descripteur Loop Statistique retenu dans la section précédente, puis avec les descripteurs LPQ (Local Phase Quantization) et BSIF (Binarized Statistical Image Features). Lors des tests d'évaluation, les images ont été divisées chacune en 16 sous-blocs de 12×12 pixels, cette taille s'étant révélée optimale lors des tests précédents avec le Loop statistique. Les résultats sont montrés dans le tableau suivant :

Descripteurs	Accuracy	Précision	Rappel	Spécificité	F1-score
Loop Stat	98.97	99.01	98.29	99.83	98.60
LPQ	69.38	69.05	62.40	94.85	62.85
BSIF	91.83	90.89	88.08	98.54	88.83

TABLE III.3 – Comparaison des performances avec différents descripteurs.

En observant les résultats obtenus selon les cinq métriques d'évaluation (accuracy, précision, recall, spécificité et F1-score), nous constatons que la méthode LOOP statistique offre toujours les meilleures performances globales. Elle surpasse à la fois LPQ et BSIF en termes de taux de reconnaissance du système.

III.6.2 Choix du modèle de classification

Pour l'étape de classification dans notre processus FER, nous avons testé plusieurs classifieurs : les SVM (Support Vector Machine), les KNN (K-Nearest Neighbors), Random Forest (RF) et Decision Tree (DT). Nous utilisons dans nos simulations pour l'extraction des caractéristiques, le descripteur MBSF-LOOP avec le pas de 12×12 pixels (16 sous-blocs), précédemment identifié comme optimal. Les résultats de nos simulations sont consignés dans le tableau suivant.

Classificateur	Accuracy	Précision	Recall	Spécificité	F1-score
SVM	98.97	99.01	98.29	99.83	98.60
KNN	32.65	66.35	27.22	88.76	38.86
DT	56.63	49.25	47.57	92.69	47.61
RF	78.06	72.92	61.22	96.05	61.41

TABLE III.4 – Comparaison des performances avec divers classifieurs.

Les résultats obtenus montrent une différence significative entre les techniques utilisées pour la classification. En analysant les cinq métriques d'évaluation, nous constatons que le classifieur SVM se distingue clairement, surpassant nettement les autres modèles.

III.6.3 Comparaison des performances du système avec les bases de données CK+ et JAFFE

Nous avons testé notre système de reconnaissance automatique des expressions faciales sur deux bases de données distinctes, à savoir CK+ et JAFFE, dans le but d'évaluer la robustesse et la généralisation du système sur des ensembles de données différents. Cette étape vise également à déterminer si le système maintient des performances élevées indépendamment de la base utilisée, et à identifier quelle base permet d'obtenir les meilleurs résultats du système.

La base de donnée	Accuracy	Precision	Recall	Specificité	F1-score
CK+	98.97	99.01	98.29	99.83	98.60
JAFFE	76.19	82.43	76.19	96.03	76.04

TABLE III.5 – Performances du système avec les bases de données CK+ et JAFFE.

Les tests ont été menés en utilisant la méthode optimale identifiée précédemment, à savoir le descripteur LOOP statistique, un pas de 12×12 pixels et le classificateur SVM. Les résultats montrent que le système obtient de bonnes performances sur les deux bases, ce qui confirme sa fiabilité. Cependant, en comparant métriques d'évaluation obtenues, nous observons que la base CK+ a permis d'atteindre les meilleurs scores globaux. Pour une meilleure analyse comparative, nous avons dressé les matrices de confusion obtenues lors de nos tests de classification sur les deux bases de données.

	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise
Anger	4	0	0	0	0	2	0
Disgust	1	3	2	0	0	0	0
Fear	0	0	4	0	0	2	0
Happy	0	0	0	4	1	1	0
Neutral	2	0	0	0	3	1	0
Sadness	0	0	0	0	0	4	0
Surprise	0	0	1	0	0	0	4

FIGURE III.2 – Matrice de confusion avec la base JAFFE.

	Anger	Contempt	Disgust	Fear	Happy	Sadness	Surprise
Anger	27	0	0	0	0	0	0
Contempt	1	10	0	0	0	0	0
Disgust	1	0	34	0	0	0	0
Fear	0	0	0	15	0	0	0
Happy	0	0	0	0	41	0	0
Sadness	0	0	0	0	0	17	0
Surprise	0	0	0	0	0	0	50

FIGURE III.3 – Matrice de confusion avec la base CK+.

La matrice de confusion vise à comparer les vraies expressions présentes dans les images et les expressions prédites par le système. La diagonale montre les prédictions correctes tandis que les autres cases révèlent la confusion entre les différentes expressions faciales lors des prédictions du système. Nos résultats montrent que la classification automatique opérée par notre système avec la base CK+, assure moins d'erreurs de confusion entre les différentes classes d'expressions par rapport à celle avec la base JAFFE.

III.7 Analyse des performances de notre système FER

Les résultats de nos diverses investigations et simulations présentés dans les sections précédentes, révèlent que notre système de reconnaissance automatique des expressions faciales est performant en utilisant : - le descripteur MBSF-LOOP pour l'extraction des caractéristiques, la détection de visages étant faite avec le détecteur Viola et Jones ; - le classifieur SVM pour la classification des expressions faciales ; - la base de données CK+ permet moins d'erreurs de prédictions lors de la phase de test.

III.7.1 L'avantage du descripteur LOOP statistique

Parmi les différentes méthodes d'extraction de caractéristiques testées dans notre système, le descripteur LOOP statistique s'est révélé le plus performant. Contrairement à LOOP histogramme, qui se base uniquement sur la distribution des motifs binaires dans chaque sous-bloc, la version statistique exploite des mesures plus fines telles que la moyenne, l'écart-type ou d'autres indicateurs statistiques, permettant ainsi de mieux représenter les variations locales d'intensité au sein de l'image. Comparée aux méthodes LPQ (Local Phase Quantization) et BSIF (Binarized Statistical Image Features), le LOOP statistique présente l'avantage d'être moins sensible au bruit, tout en capturant des informations structurelles locales plus discriminantes. Cette richesse d'information permet une meilleure séparation entre les différentes expressions faciales. Cela est confirmé par les résultats expérimentaux obtenus dans les tableaux III.2 et III.3, montrant que le descripteur LOOP statistique fournit systématiquement les meilleurs scores.

III.7.2 L'avantage du classificateur SVM

SVM est particulièrement adapté aux problèmes de classification multi-classes avec un nombre limité d'exemples d'apprentissage. Il excelle dans la séparation des classes par des frontières optimales, même dans des espaces de données de grande dimension. Comparé à KNN, qui est sensible aux données bruitées et aux choix de distance, et aux arbres de décision (DT et RF), qui peuvent rapidement sur-apprendre (overfitting) en présence de données complexes, SVM offre une meilleure généralisation. Les résultats expérimentaux obtenus dans le tableau III.4 confirment cette supériorité. Le classifieur SVM a permis d'atteindre un taux de reconnaissance, qui est le choix optimal pour notre architecture de reconnaissance automatique des expressions faciales.

III.7.3 L'avantage de la base CK+

Les résultats obtenus dans le tableau III.5 et dans la matrice de confusion figureII.2, ont clairement montré que notre système est plus performant avec la base de données CK+. Cela peut s'expliquer par plusieurs facteurs. Tout d'abord, CK+ est une base plus riche et plus diversifiée, tant au niveau du nombre de sujets que de la variété des expressions faciales et des conditions d'acquisition. Elle contient également des images avec des transitions émotionnelles plus marquées, ce qui facilite l'apprentissage des modèles. En comparaison, JAFFE est plus restreinte, avec un nombre limité d'images et des expressions plus subtiles, issues uniquement de femmes japonaises, ce qui limite la capacité du système à généraliser. Ces limitations se reflètent dans les performances obtenues dans le tableauIII.5 et la matrice de confusion montrée par la figureII.1. Notre système

FER atteint le meilleur taux de reconnaissance sur CK+ et des scores moins élevés mais acceptables sur JAFFE .

III.8 Conclusion

Dans ce dernier chapitre, nous avons exposé les résultats obtenus suite à la mise en œuvre de nos architectures présentées dans le chapitre précédent (chapitre II). Tout d'abord, nous avons détaillé la mise en œuvre de notre travail, en présentant notre environnement de développement sur lequel le système a été réalisé. Ensuite, nous avons analysé les résultats de chaque expérience en calculant les métriques d'évaluation, facilitant ainsi la comparaison entre les différentes expérimentations. À partir des résultats obtenus, nous avons démontré que la combinaison du descripteur LOOP statistique avec un pas de 12×12 pixels (16 sous-blocs) et du classificateur SVM nous offre les meilleures performances. Le système s'est également révélé fiable sur différentes bases de données, en particulier CK+, confirmant son efficacité et sa capacité de généralisation. Cela constitue une solution particulièrement efficace pour la reconnaissance automatique des expressions faciales.

Conclusion Générale

Au terme de ce mémoire, nous avons proposé une démarche visant à améliorer la reconnaissance des expressions faciales à travers une chaîne de traitement composée de trois étapes principales : la détection du visage, l'extraction des caractéristiques et la classification.

Dans la phase de détection du visage, nous avons appliqué l'algorithme de Viola et Jones, reconnu pour sa rapidité et son efficacité dans des environnements contraints. Cette méthode nous a permis d'isoler avec précision le visage à partir d'images d'entrée, condition essentielle pour la suite du traitement.

Ensuite, dans le cadre de l'extraction des caractéristiques, nous avons employé le descripteur LOOP, qui exploite des informations statistiques locales orientées pour modéliser les variations de texture liées aux expressions faciales. Cette étape a permis de transformer l'image du visage détecté en un vecteur de caractéristiques représentatif.

Enfin, pour la classification des différentes expressions faciales, nous avons utilisé un classifieur SVM en raison de sa capacité à maximiser la marge entre les classes et à bien généraliser, même avec un nombre limité de vecteurs de caractéristiques. À travers différentes expérimentations, ce système a permis d'atteindre des taux de reconnaissance satisfaisants, confirmant la pertinence de notre approche.

Plusieurs investigations et simulations numériques nous ont conduites à la combinaison d'une détection robuste, d'un descripteur statistique adapté et d'un classifieur performant. Les résultats obtenus démontrent que la combinaison que nous avons retenue constitue une solution efficace pour la reconnaissance automatique des expressions faciales. Cette approche se veut prometteuse pour des applications en interaction homme-machine, en analyse comportementale, en surveillance intelligente et dans bien d'autres applications où une reconnaissance précoce d'expressions faciales revêt un intérêt capital.

Perspectives

Ce travail constitue une première étape vers la mise en place d'un système efficace de reconnaissance automatique des expressions faciales basé sur les images. Pour améliorer davantage ce système, plusieurs perspectives peuvent être envisagées. Il serait intéressant d'explorer d'autres bases de données plus variées, intégrant différentes conditions d'éclairage, différents angles de prise de vue ou différentes résolutions d'image. De plus, l'intégration de techniques d'apprentissage profond, comme les réseaux de neurones convolutifs (CNN), pourrait permettre une extraction automatique et plus performante des caractéristiques. Enfin, l'optimisation du système pour un traitement en temps réel représenterait un pas important vers des applications concrètes dans des environnements réels.

Bibliographie

- [1] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4) :169–200, 1992.
- [2] Shashidhar G Koolagudi and K Sreenivasa Rao. Emotion recognition from speech : a review. *International journal of speech technology*, 15 :99–117, 2012.
- [3] Paul Ekman. *Emotions Revealed : Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Times Books, New York, 2003.
- [4] Curtis Padgett and Garrison Cottrell. Representing face images for emotion classification. *Advances in neural information processing systems*, 9, 1996.
- [5] Andreas Lanitis, Christopher J. Taylor, and Timothy F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and machine intelligence*, 19(7) :743–756, 1997.
- [6] Iris B Mauss and Michael D Robinson. Measures of emotion : A review. *Cognition and emotion*, 23(2) :209–237, 2009.
- [7] Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1) :32–80, 2001.
- [8] Naman Kohli. *Automatic kinship verification in unconstrained faces using deep learning*. West Virginia University, 2019.
- [9] Simon Nash, Mark Rhodes, and Joanna Isabelle Olszewska. Ifr : Interactively pose corrected face recognition. In *International Conference on Bio-inspired Systems and Signal Processing*, volume 5, pages 106–112. SCITEPRESS, 2016.

- [10] Delphine Poux. *Reconnaissance automatique des expressions faciales en présence d'occultations partielles du visage*. PhD thesis, Université de Lille, 2021.
- [11] Naveed Ahmed, Zaher Al Aghbari, and Shini Girija. A systematic survey on multimodal emotion recognition using learning algorithms. *Intelligent Systems with Applications*, 17 :200171, 2023.
- [12] Payam Parsinejad. *Inferring mental workload changes of subjects unfamiliar with a touch screen game through physiological and behavioral measurements*. Northeastern University, 2016.
- [13] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. A survey of affect recognition methods : audio, visual and spontaneous expressions. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 126–133, 2007.
- [14] Mira Jeong, Jaeyeal Nam, and Byoung Chul Ko. Lightweight multilayer random forests for monitoring driver emotional status. *Ieee Access*, 8 :60344–60354, 2020.
- [15] Mrinalini Patil and S Veni. Driver emotion recognition for enhancement of human machine interface in vehicles. In *2019 international conference on communication and signal processing (ICCSP)*, pages 0420–0424. IEEE, 2019.
- [16] Afizan Azman, Kirbana Jai Raman, Imran Artwel Junior Mhlanga, Siti Zainab Ibrahim, Sumendra Yogarayan, Mohd Fikri Azli Abdullah, Siti Fatimah Abdul Razak, Anang Hudaya Muhamad Amin, and Kalaiarasi Sonai Muthu. Real time driver anger detection. In *Information Science and Applications 2018 : ICISA 2018*, pages 157–167. Springer, 2019.
- [17] SS Sudha and SS Suganya. On-road driver facial expression emotion recognition with parallel multi-verse optimizer (pmvo) and optical flow reconstruction for partial occlusion in internet of things (iot). *Measurement : Sensors*, 26 :100711, 2023.
- [18] Huafei Xiao, Wenbo Li, Guanzhong Zeng, Yingzhang Wu, Jiyong Xue, Jun-cheng Zhang, Chengmou Li, and Gang Guo. On-road driver emotion recognition using facial expression. *Applied Sciences*, 12(2) :807, 2022.
- [19] Harshit Varma, Nagarajan Ganapathy, and Thomas M Deserno. Video-based driver emotion recognition using hybrid deep spatio-temporal feature learning. In *Medical Imaging 2022 : Imaging Informatics for Healthcare, Research, and Applications*, volume 12037, pages 57–63. SPIE, 2022.

- [20] Motoi Suwa. A preliminary note on pattern recognition of human emotional expression. *Proc. IJPR*, pages 408–410, 1978.
- [21] Ciprian Adrian Corneanu, Marc Oliu Simón, Jeffrey F Cohn, and Sergio Escalera Guerrero. Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition : History, trends, and affect-related applications. *IEEE transactions on pattern analysis and machine intelligence*, 38(8) :1548–1568, 2016.
- [22] Takeo Kanade, Jeffrey F Cohn, and Yingli Tian. Comprehensive database for facial expression analysis. In *Proceedings fourth IEEE international conference on automatic face and gesture recognition (cat. No. PR00580)*, pages 46–53. IEEE, 2000.
- [23] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+) : A complete dataset for action unit and emotion-specified expression. In *2010 IEEE computer society conference on computer vision and pattern recognition-workshops*, pages 94–101. IEEE, 2010.
- [24] László A Jeni, Jeffrey M Girard, Jeffrey F Cohn, and Fernando De La Torre. Continuous au intensity estimation using localized, sparse facial feature space. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, pages 1–7. IEEE, 2013.
- [25] Xiangyun Zhao, Xiaodan Liang, Luoqi Liu, Teng Li, Yugang Han, Nuno Vasconcelos, and Shuicheng Yan. Peak-piloted deep network for facial expression recognition. In *Computer Vision–ECCV 2016 : 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 425–442. Springer, 2016.
- [26] Zhenbo Yu, Qinshan Liu, and Guangcan Liu. Deeper cascaded peak-piloted network for weak expression recognition. *The Visual Computer*, 34(12) :1691–1699, 2018.
- [27] Unqua Laraib, Arslan Shaukat, Rizwan Ahmed Khan, Zartasha Mustansar, Muhammad Usman Akram, and Umer Asgher. Recognition of children’s facial expressions using deep learned features. *Electronics*, 12(11) :2416, 2023.
- [28] Paul Viola and Michael Jones. Robust real-time object detection [j]. *International journal of computer vision*, 4 :34–47, 2001.
- [29] Timo Ojala, Matti Pietikainen, and David Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of

- distributions. In *Proceedings of 12th international conference on pattern recognition*, volume 1, pages 582–585. IEEE, 1994.
- [30] Tapabrata Chakraborti, Brendan McCane, Steven Mills, and Umapada Pal. Loop descriptor : local optimal-oriented pattern. *IEEE Signal Processing Letters*, 25(5) :635–639, 2018.
- [31] Jayaraman Murala, R. P. Maheshwari, and R. Balasubramanian. Local optimal oriented pattern (loop) for texture classification. *Neurocomputing*, 119 :382–396, 2013.
- [32] Vikramaditya Jakkula. Tutorial on support vector machine (svm). *School of EECS, Washington State University*, 37(2.5) :3, 2006.
- [33] Yanli Liu, Yourong Wang, and Jian Zhang. New machine learning algorithm : Random forest. In *International conference on information computing and applications*, pages 246–252. Springer, 2012.
- [34] N Ben Amor, Salem Benferhat, and Zied Elouedi. Reseaux bayesiens naifs et arbres de decision dans les systemes de detection d'intrusions. *TSI-Technique et Science Informatiques*, 25(2) :167–196, 2006.
- [35] Javaid Ahmad Wani, Sparsh Sharma, Malik Muzamil, Suhaib Ahmed, Surbhi Sharma, and Saurabh Singh. Machine learning and deep learning based computational techniques in automatic agricultural diseases detection : Methodologies, applications, and challenges. *Archives of Computational methods in Engineering*, 29(1) :641–677, 2022.
- [36] Luefeng Chen, Min Wu, Witold Pedrycz, and Kaoru Hirota. *Emotion recognition and understanding for emotional human-robot interaction systems*, volume 926. Springer Nature, 2020.
- [37] David Orozco, Christopher Lee, Yevgeny Arabadzhi, and Deval Gupta. Transfer learning for facial expression recognition. *semantics scholar*, 2018.
- [38] Michael Lyons, Shigeru Akamatsu, Miyuki Kamachi, and Jiro Gyoba. Coding facial expressions with gabor wavelets. In *Proceedings Third IEEE international conference on automatic face and gesture recognition*, pages 200–205. IEEE, 1998.
- [39] Paul Ekman. Facial expression and emotion. *American psychologist*, 48(4) :384, 1993.
- [40] Ian J Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, et al. Challenges in representation learning : A report on

- three machine learning contests. In *Neural information processing : 20th international conference, ICONIP 2013, daegu, korea, november 3-7, 2013. Proceedings, Part III 20*, pages 117–124. Springer, 2013.
- [41] Trung Tuan Dao, Duc Hong Vu, Cuong Pham, and Anh Tran. Efhq : Multi-purpose extremepose-face-hq dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22605–22615, June 2024.
- [42] Luoqi Liu, Chao Xiong, Hanwang Zhang, Zhiheng Niu, Meng Wang, and Shuicheng Yan. Deep aging face verification with large gaps. *IEEE Transactions on Multimedia*, 18(1) :64–75, 2015.
- [43] Abir Fathallah, Lotfi Abdi, and Ali Douik. Facial expression recognition via deep learning. In *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, pages 745–750. IEEE, 2017.
- [44] Hardie Cate, Fahim Dalvi, and Zeshan Hussain. Deepface : Face generation using deep learning. *arXiv preprint arXiv :1701.01876*, 2017.
- [45] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. Ieee, 2001.
- [46] AS Tolba, AH El-Baz, AA El-Harby, et al. Face recognition : A literature review. *International Journal of Signal Processing*, 2(2) :88–103, 2006.

Interface graphique

Dans le cadre de notre projet, nous avons conçu et développé une interface graphique permettant de faciliter l'interaction entre l'utilisateur et le système de reconnaissance automatique des expressions faciales. Cette interface vise principalement à détecter l'expression faciale automatiquement en temps réel.

Notre interface, conçue avec Python (Jupyter Notebook), est intuitive et conviviale.

A.1 Fenêtre d'accueil

La fenêtre principale de notre application s'ouvre automatiquement au démarrage, offrant à l'utilisateur un accès direct pour démarrer la caméra. Au centre de cette fenêtre d'accueil figure A.1.

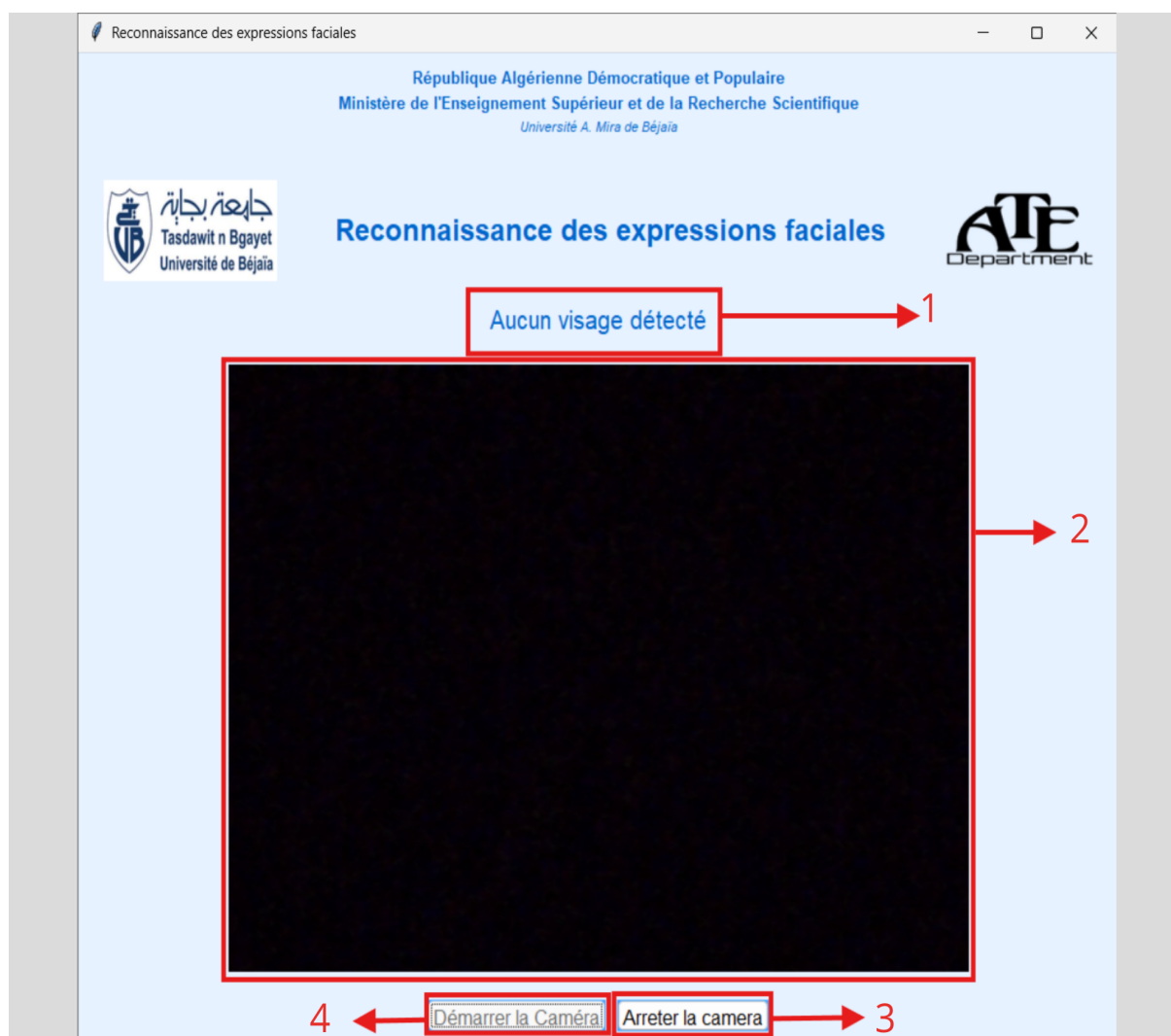


FIGURE A.1 – Fenêtre d'accueil.

1. Message indiquant que l'application soit : détecte l'expression ou aucun visage détecté.
2. Une zone d'affichage pour la capture vidéo en temps réel de la caméra.
3. Le bouton « Arrêter la caméra », permet à l'utilisateur d'arrêter la capture vidéo et de prendre une photo.
4. Le bouton « Démarrer la caméra », permet à l'utilisateur de lancer la capture vidéo en temps réel à partir de la webcam.

A.1.1 Le système de reconnaissance des expressions faciales en temps réel.

1. En cliquant sur le bouton « Démarrer la caméra », une zone s'affiche pour la capture vidéo en temps réel.
2. Détection de visage : Nous avons utilisé la méthode Viola-Jones.
Une fois le visage détecté, un cadre vert est affiché autour du visage de l'individu.
3. Extraction des caractéristiques et Classification : nous avons utilisé un modèle CNN préentraîné intégré dans la bibliothèque DeepFace.



FIGURE A.2 – Exemple de détection de l'expression colère.

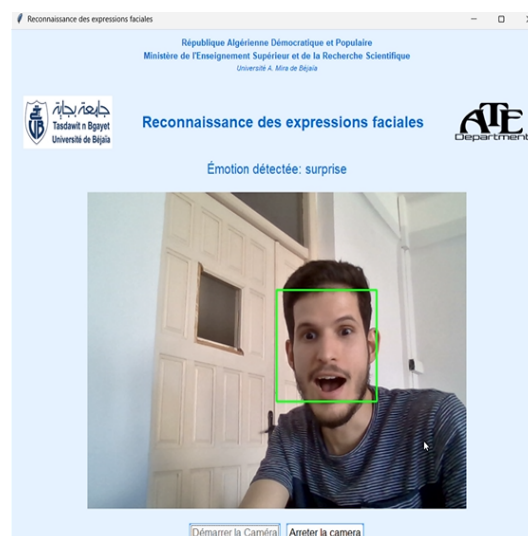


FIGURE A.3 – Exemple de détection de l'expression surprise.



FIGURE A.4 – Exemple de détection de l'expression tristesse.

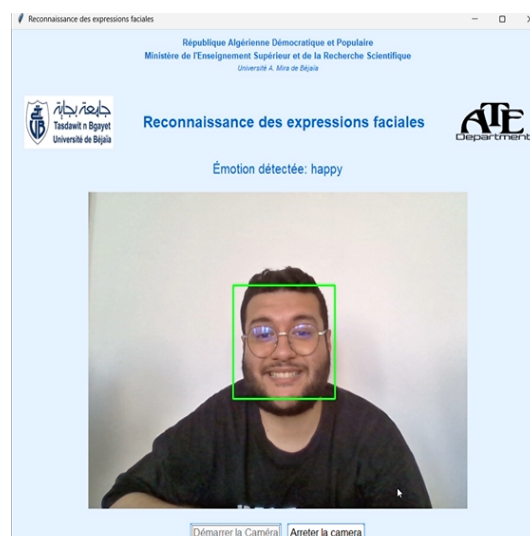


FIGURE A.5 – Exemple de détection de l'expression joie.

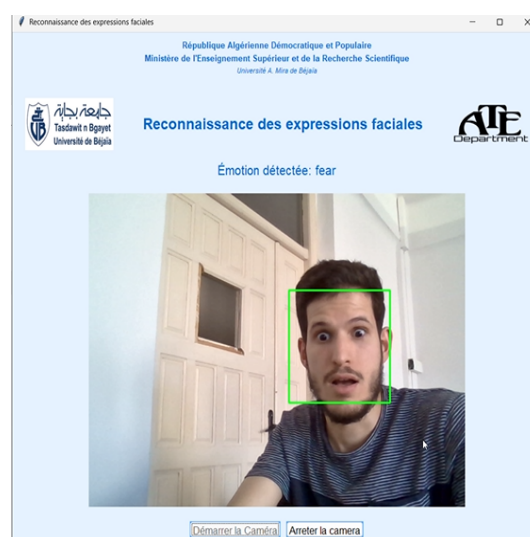


FIGURE A.6 – Exemple de détection de l'expression peur.



FIGURE A.7 – Exemple de détection de l'expression neutre.

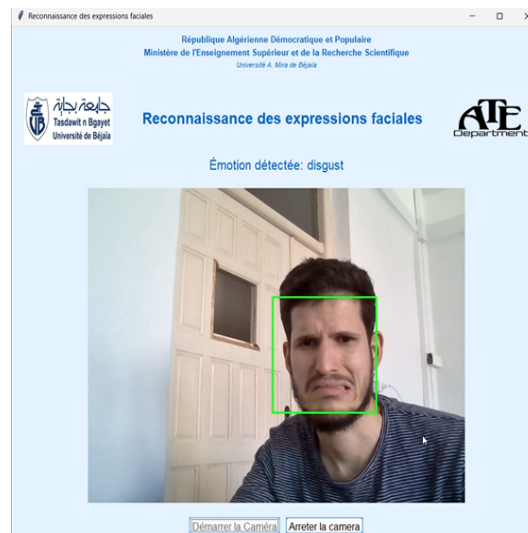


FIGURE A.8 – Exemple de détection de l'expression dégoût.



FIGURE A.9 – Exemple du cas non détection de visage.

A.1.2 Environnement de travail.

Jupyter Notebook est un environnement interactif open source qui permet d'écrire et d'exécuter du code, de la documentation, des visualisations et des équations mathématiques dans un seul document. Il est utilisé pour, le développement Python, la visualisation de données, l'apprentissage automatique, l'enseignement et la recherche scientifique.



FIGURE A.10 – Logo de Jupyter.

A.1.3 Bibliothèques.

1. **Tkinter** : Utiliser pour la création de l'interface graphique.
2. **Open Source Computer Vision Library (OpenCV)** : Utiliser pour l'accès à la caméra, détection de visage (Viola-Jones), traitement d'image.
3. **DeepFace** : Analyse d'expressions via CNN préentraîné [44].

Abstract

This thesis focuses on the automatic recognition of facial expressions from images, a field of critical importance for enhancing human-computer interaction and developing applications in areas such as security, mental health, and intelligent interfaces. It offers an in-depth analysis of the main challenges, such as inter-subject variability and heterogeneous image acquisition conditions, and systematically evaluates several machine learning methodologies. The research aims to design a complete processing pipeline to identify the most effective combinations of feature extraction techniques and classification algorithms to maximize accuracy. The system begins with a preprocessing step using the Viola-Jones algorithm for reliable face detection. A major contribution of this work lies in the comparative evaluation of several discriminative texture descriptors, including Local Optimal Oriented Patterns (LOOP), in their histogram-based (LOOP-MBH) and statistical (LOOP-MBSF) variants, Local Phase Quantization (LPQ), and Binarized Statistical Image Features (BSIF). For the classification phase, several standard supervised machine learning algorithms are implemented and compared: Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forests (RF), and Decision Trees (DT). Each descriptor-classifier combination is rigorously evaluated using two public facial expression datasets to ensure the generalizability of the results. The comparative assessments show that carefully selecting both the descriptor and the classifier is key to achieving high performance. The best-performing pairings yield high recognition rates, providing a clear methodological framework for designing robust and accurate facial expression recognition systems in real-world conditions.

Keywords: Automatic expression recognition, Viola-Jones, texture descriptor, machine learning classifier.

Résumé

Ce mémoire porte sur la reconnaissance automatique des expressions faciales à partir d'images, un domaine crucial pour l'amélioration de l'interaction homme-machine et le développement d'applications dans des secteurs comme la sécurité, la santé mentale ou les interfaces intelligentes. Il propose une analyse approfondie des principaux défis, tels que la variabilité inter-sujets et les conditions d'acquisition d'images hétérogènes, et évalue de manière systématique plusieurs méthodologies d'apprentissage automatique. La recherche vise à concevoir une chaîne de traitement complète, permettant d'identifier les combinaisons les plus efficaces entre techniques d'extraction de caractéristiques et algorithmes de classification, afin d'optimiser la précision. Le système débute par une étape de prétraitement utilisant l'algorithme de Viola-Jones pour une détection fiable des visages. Une contribution majeure de ce travail réside dans l'évaluation comparative de plusieurs descripteurs de caractéristiques texturales discriminantes : les patrons d'Orientation Optimaux Locaux (LOOP), dans leurs variantes basées sur les histogrammes (LOOP-MBH) et les statistiques (LOOP-MBSF), la Quantification de Phase Locale (LPQ) et les Caractéristiques d'Image Statistiques Binarisées (BSIF). Pour la classification, plusieurs algorithmes supervisés sont mis en œuvre et comparés : Machines à Vecteurs de Support (SVM), K-Plus Proches Voisins (KNN), Forêts Aléatoires (RF) et Arbres de Décision (DT). Chaque combinaison descripteur-classifieur est rigoureusement évaluée sur deux bases de données publiques d'expressions faciales, afin de garantir la généralisabilité des résultats. Les évaluations montrent que le choix judicieux du descripteur et du classifieur est déterminant pour atteindre une haute performance. Les meilleurs appariements identifiés permettent d'obtenir des taux de reconnaissance élevés, offrant un cadre méthodologique clair pour la conception de systèmes robustes et précis en conditions réelles.

Mots clés : Reconnaissance automatique des expressions, Viola-Jones, descripteur de texture, classifieur d'apprentissage automatique.

ملخص

تتناول هذه المذكرة موضوع التعرف التلقائي على تعابير الوجه انطلاقاً من الصور، وهو مجال بالغ الأهمية لتحسين التفاعل بين الإنسان والآلة، وتطوير التطبيقات في مجالات مثل الأمن، والصحة النفسية، والواجهات الذكية. يقترح هذا العمل تحليلاً معمقاً للتحديات الرئيسية، مثل اختلافات بين الأفراد وتفاوت ظروف التقاط الصور، كما يقدم تقييماً منهجياً لعدة منهجيات في التعلم الآلي. تهدف هذه الدراسة إلى تصميم سلسلة معالجة متكاملة تسمح بتحديد أنجع التوليفات بين تقنيات استخراج الخصائص وخوارزميات التصنيف، من أجل تحسين الدقة. يبدأ النظام بمرحلة ما قبل المعالجة باستخدام خوارزمية Viola-Jones للكشف الموثوق عن الوجوه. وتتمثل أبرز مساهمات هذا العمل في التقييم المقارن لعدة واصفات تعتمد على الخصائص التركيبية التمييزية، ومنها: الأنماط المثلى للاتجاهات المحلية (LOOP)، نسخاتها المعتمدتان على الهستوغرامات (LOOP-MBH) والإحصائيات، التكميم المحلي للطور (LPQ)، الخصائص الإحصائية الثنائية للصور (BSIF). فيما يخص التصنيف، تم تنفيذ ومقارنة عدة خوارزميات تعلم خاضع للإشراف، وهي: آلات المتجهات الداعمة (SVM)، خوارزمية أقرب الجيران (KNN)، الغابات العشوائية (RF)، أشجار القرار (DT). تم تقييم كل توليفة بين الواصف والمصنف بدقة عالية على قاعدتي بيانات عموميتين لتعابير الوجه، لضمان قابلية تعميم النتائج. أظهرت التقييمات أن الاختيار الدقيق للواصف والمصنف يلعب دوراً حاسماً في تحقيق أداء مرتفع، وقد تم التوصل إلى أفضل التوافقات التي تسمح بالحصول على معدلات تعرف عالية، مما يوفر إطاراً منهجياً واضحاً لتصميم أنظمة قوية ودقيقة في ظروف الاستخدام الحقيقية.

الكلمات المفتاحية: التعرف التلقائي على التعابير، خوارزمية فيولا-جونز، واصف القوام، مصنف التعلم الآلي.

