

THÈSE

En vue de l'obtention du Diplôme de Doctorat

Présenté par : TALEB Imène

Le système biométrique : détection et reconnaissance de visage

Faculté : Génie Electrique
Département : Electronique
Domaine : Sciences et Techniques
Filière : Electronique
Intitulé de la Formation : Techniques modernes de l'information et de la communication

Devant le Jury Composé de :

<i>Membres de Jury</i>	<i>Grade</i>	<i>Qualité</i>	<i>Domiciliation</i>
<i>ALIPACHA Adda-</i>	<i>Prof</i>	<i>Président</i>	<i>USTO -</i>
<i>OULD MAMMAR Madani -</i>	<i>MC-A</i>	<i>Encadrant</i>	<i>Univ_Mosta -</i>
<i>OUAMRI Abdelaziz -</i>	<i>Prof</i>	<i>Co-Encadrant</i>	<i>USTO -</i>
<i>-</i>			<i>-</i>
<i>LOUKIL Abdelhamid -</i>	<i>Prof</i>	<i>Examineurs</i>	<i>USTO-</i>
<i>MERAH Mustapha</i>	<i>MC-A</i>		<i>Univ_Mosta -</i>
<i>-</i>	<i>-</i>	<i>Invité(e)</i>	<i>-</i>

Année Universitaire : 2018/2019

A *la mémoire de mes défunts grands parents,*

A mes chers parents,

A mourad et amina,

REMERCIEMENTS

Louange à ALLAH, le tout Puissant, qui a permis que je parvienne à l'aboutissement de cette thèse qui a nécessité beaucoup d'efforts, de temps et de sacrifices.

Je débute ces remerciements par mon jury de thèse. Je remercie tout d'abord, le professeur **OUAMRI Abdelaziz**, pour m'avoir accueilli au sein du laboratoire signaux et images et d'avoir été à la fois mon co-directeur de thèse.

Le professeur **ALI PACHA Adda** pour l'honneur qu'il m'a accordé en acceptant de juger mon travail et avoir accepté la présidence du jury. Je remercie aussi tous les membres du jury pour leur participation et leur temps consacré, que messieurs **Abdelaziz OUAMRI**, **Mustapha MERAH** et **Abdelhamid LOUKIL** trouvent ici ma gratitude.

Je remercie particulièrement mon promoteur et directeur de thèse Mr. **OULD MAMMAR Madani** pour mon encadrement de thèse et avoir accepté de rapporter mes travaux. La pertinence de ses remarques m'ont permis d'avancer dans mes réflexions. Il a été celui que j'ai le plus sollicité pendant ces années. Je le remercie pour les conseils et les pistes de réflexion qu'il m'a soumis et pour ses qualités humaines. Je le remercie aussi pour m'avoir fait confiance dans les initiatives que j'ai prises.

Merci à tous mes collègues du laboratoire signaux et images, principalement à ceux qui m'ont éclairé, encouragé et motivé et à tous mes amis, source de bonté et de bonne énergie.

Je ne peux oublier d'exprimer ma reconnaissance à mes responsables au centre de développement des satellites d'Oran, pour leur support, conseils et tolérance quant à la finalisation de ma thèse.

Le meilleur pour la fin ! Un grand Merci à mes parents pour m'avoir soutenue tout au long de cette thèse, vous resterez à jamais mes premiers enseignants et formateurs des bonnes valeurs de la vie, du savoir être et du savoir vivre. Merci à vous parce que vous avez cru en moi.

De façon plus générale, merci à tous les membres de ma famille, en particulier mon frère Mourad, ma sœur Amina, ma chère grand-mère pour ses bonnes prières, ils sont d'ailleurs nombreux à avoir prié pour ma réussite, je vous en remercie de tout cœur.

Faut-il peindre ce qu'il y a sur un visage ? Ce qu'il y a dans un visage ? Ou ce qui se cache derrière un visage ?

-Pablo Picasso

RESUME

Le visage est l'un des moyens les plus importants de la communication humaine, aussi bien dans des contextes privés que professionnels. Nous y lisons souvent des informations pertinentes concernant l'humeur d'une personne ou ses intentions.

Une expression faciale est une manifestation visible sur un visage de l'état d'esprit (émotion, réflexion), de l'activité cognitive, de l'activité physiologique (fatigue, douleur), de la personnalité et de la psychopathologie d'une personne. Des travaux de recherche en psychologie ont démontré que les expressions faciales jouent un rôle prépondérant dans la coordination de la conversation humaine, et ont un impact plus important sur l'auditeur que le contenu textuel du message exprimé.

L'essentiel de l'information d'une expression faciale est contenue dans la déformation des traits permanents principaux du visage, caractérisée par un changement perceptible visuellement.

Les variations d'expression constituent des limitations importantes à la reconnaissance de visages en deux dimensions. Dans cette thèse, nous proposons d'augmenter la robustesse et performance des taux de reconnaissances d'expression faciales. Pour cela, nous proposons de reconnaître l'expression faciale pour un visage donné. Notre premier apport consiste à fusionner une méthode globale d'apparence, classique dans la reconnaissance faciale, à savoir, l'analyse en composantes principales (ACP) et un algorithme basé région moins classique, Angular Radial Transform (ART), qui est une décomposition polynomiale par moments.

Notre travail propose également une nouvelle forme d'ART pour la reconnaissance d'expression faciale utilisant un développement mathématique polaire basé sur la transformation radiale angulaire.

Notre deuxième apport combine la nouvelle forme de la transformation radiale angulaire polaire, que nous baptisons (P -ART) et l'analyse en composantes principales (ACP).

Le nouvel ART est un descripteur de forme présenté différemment que le descripteur cartésien conventionnel. L'entraînement de la machine à vecteurs de support (SVM) est utilisé pour reconnaître l'expression faciale pour une image de visage d'entrée.

Enfin, les résultats expérimentaux montrent la performance du P-ART et de l'ACP. La fusion de ces deux techniques peut être meilleure que d'autres méthodes existantes de reconnaissance de l'expression faciale. Au cours de l'expérience, la base de l'expression faciale féminine donnée au Japon (JAFFE) et les bases de données Cohn-Kanade ont été utilisées.

Mots clés : reconnaissance d'expression faciale ; Analyse en composantes principales (ACP) ; Angular Radial Transform (ART) ; transformation radiale angulaire polaire (P - ART) ; machine à vecteurs de support (SVM) ; (JAFFE) ; Cohn-Kanade.

ABSTRACT

Face is one of the common human communication tools used in private and professional world. We usually detect information regarding the humor of a person or its intentions.

A facial expression is a visible action of the face reflecting the state of mind (emotion, reflection), a cognitive activity, a psychological activity (tiredness, pain), personality and psychopathology. Research studies in psychology reveal that facial expression has a key role in coordination of human conversation, and has an important impact on auditor than a written message.

The useful information of a facial expression is held in the deformation of the main permanent facial features, characterized by a perceptible visual change. The expression variations present limitations for two-dimensional 2D facial recognition. In this thesis, we propose the improvement of the robustness and the enhancement of the performances of the facial recognition rates. For this purpose, we use a fusion between a global method of appearance, Principal Component Analysis (PCA) and a regional based algorithm called Angular Radial Transform (ART).

In this work, we present also a new method of facial expression recognition (FER) using a polar mathematical development based on Angular Radial Transformation. This method combines the Polar Angular Radial Transformation (P-ART) and Principal Component Analysis (PCA).

The new ART is a new and different form descriptor, than the convolutional Cartesian descriptor. Support Vector Machine (SVM) training is used to recognize the facial expression for an input face image.

The experimental results show the performances of P-ART and PCA. The fusion of these two techniques can be better than other existing methods of recognition of facial expression. During the experiment, the Japanese Female Facial Expression (JAFFE) and the Cohn-Kanade databases has been used.

Keywords: Fusion; Principal Component Analysis (PCA); Polar Angular Radial Transform (P-ART); support vector machine (SVM); facial expression recognition (FER); Japanese female expression (JAFFE); Cohn-Kanade.

الملخص

تعابير الوجه، هي واحدة من الوسائل المعتمدة في الإتصال البشري سواء في المعاملات الخاصة أو المهنية، حيث يتمكن عادة من استشعار حالة الشخص ومراده أو نواياه. تعابير الوجه هي تعبير مرئي على الحالة النفسية للشخص (مشاعر، تفكير)، على مستوى الإدراك، الحالة النفسية (الم، إرهاق)، عن الشخصية و الأمراض النفسية

تطُرقت العديد من الأبحاث العلمية في مجال علم النفس إلى مكانة تقنيات دراسة تعابير الوجه باعتبارها تلعب دورا مهما في علم الإتصال، إذ أنها تمثل تأثيرا قويا على المتلقي أكثر من التعبير الكتابي. المعلومات المهمة في تعابير الوجه تتركز في تحول تقاسيم الوجه الأساسية و التي يمك إدراكها مرئيا. التحولات في تعابير الوجه تمثل محدودية معتبرة في تقنية التعرف على تعابير الوجه ثنائي الأبعاد في هذه الأطروحة، نحاول تحسين دقة و فعالية معدل نتائج التعرف على تعابير الوجه، من خلال دمج النظام الكلي للمظهر الخارجي، تحليل المكونات الأساسية و خوارزمية تحديد المحيط تسمى تحويل الشعاع الزاوي من خلال هذا البحث، نقدم طريقة جديدة لتقنية التعرف على تعابير الوجه باستعمال تحليل رياضي قطبي يعتمد على محولة الشعاع الزاوي. هذه الطريقة تركز على التحويل الشعاع الزاوي القطبي و تحليل المكونات الأساسية. تحليل الشعاع الزاوي الجديد يعتبر واصفا فعلا للأشكال مقارنة بالواصف الديكارتي التقليدي

تستعمل آلة الدعم الموجه كمشغل للتعرف على التعابير المعروضة في المدخل. النتائج التجريبية تبين مدى دقة و فعالية التحويل الشعاعي الزاوي القطبي مع تحليل المكونات الأساسية بدمج هاتين التقنيتين يوفر نتائج جيدة مقارنة بالتقنيات الأخرى من خلال هذه التجارب، تم استعمال الوجه القاعدي للأنتى اليابانية جابفو قواعد البيانات كو هن كانادي

كلمات مفتاحية: تعابير الوجه؛ مشغل للتعرف على التعابير؛ دمج؛ التحويل الشعاع الزاوي القطبي؛ تحليل المكونات الأساسية؛ جابفو؛ كو هن كاناد

TABLE DES MATIERES

INTRODUCTION.....	17
CHAPITRE I : INTRODUCTION AUX SYSTEMES DE RECONNAISSANCE	
BIOMETRIQUE	20
1.1. Introduction	21
1.2. Préambule et cadre général	21
1.2.1. La Biométrie	21
1.2.2. Le système biométrique	22
1.2.3. L'architecture du système biométrique.....	22
1.2.4. Evaluation des performances des systèmes biométriques.....	24
1.2.5. La mesure de la performance du système biométrique.....	25
1.2.6. Les applications des systèmes biométriques	26
1.2.7. La biométrie du visage	26
1.2.8. Le système de reconnaissance faciale.....	28
1.2.9. Les principales difficultés de la reconnaissance faciale.....	30
1.2.10. La multimodalité	31
1.2.11. Les différentes formes de multimodalité	32
1.2.12. Les architectures.....	33
1.2.13. La fusion des données	33
1.3. La problématique	35
1.4. Les objectifs de la thèse	38
1.5. Nos contributions.....	37
1.7. Conclusion.....	38
CHAPITRE II : LES METHODES DE RECONNAISSANCE D'EXPRESSION	
FACIALE	39
2.1. Introduction	40

2.2. Les méthodes de reconnaissance faciale	40
2.2.1. Les méthodes globales	42
2.2.2. Les méthodes locales	41
2.2.3. Les méthodes hybrides	41
2.3. Les méthodes utilisées dans le FER.....	42
2.3.1. Le Système de Codification des Actions Faciales FACS	44
2.3.2. MPEG4.....	44
2.3.3. Les méthodes de reconnaissance d'expression faciale	45
2.3.3.1. L'Analyse en Composantes Principales (l'ACP).....	46
2.3.3.2. Les mesures de distance	49
2.3.3.3. Machine à Vecteur de Support.....	51
2.3.3.4. Description des formes	56
2.4. Le système de reconnaissance d'expression faciale	60
2.4.1. La détection du visage.....	60
2.4.2. Extraction des composantes du visage	63
2.5. Les bases de données	64
2.6. Approche proposée	65
2.7. Conclusion.....	65
CHAPITRE III : LES METHODES PROPOSEES.....	67
3.1. Introduction	68
3.2. L'algorithme proposé.....	68
3.3. Le système proposé pour la reconnaissance d'expression faciale	70
3.3.1. Le prétraitement.....	71
3.3.2. L'extraction des caractéristiques	72
3.3.2.1. Le Descripteur ART (Angular Radial Transform)	73
3.3.2.2. L'ART Polaire (P-ART) proposé	76

3.3.2.3.	L'Analyse en Composantes Principales.....	78
3.3.2.4.	Classification par SVM.....	79
3.3.2.5.	La fusion ART /ACP	80
3.4.	Conclusion	82
CHAPITRE IV : RESULTATS EXPERIMENTAUX.....		83
4.1.	Introduction.....	84
4.2.	Implémentation	84
4.2.1.	Base de données JAFFE	84
4.2.2.	Base de données Kohn-Canade.....	85
4.3.	Résultats et analyses	86
4.3.1.	Résultats des taux de reconnaissance d'expression faciale RR en utilisant L'ACP.....	87
4.3.2.	Résultats des taux de reconnaissance d'expression faciale RR en utilisant ART	88
4.3.3.	Résultats des taux de reconnaissance d'expression faciale RR de la fusion ACP+ ART en utilisant les bases de données JAFEE et CK	89
4.3.4.	Résultat obtenu en utilisant ART et P-ART (base de données JAFFE).....	91
4.3.5.	Résultat obtenu avec PCA / P-ART et fusion (PCA + P-ART) (bases de données JAFFE et CK).....	92
4.3.6.	La comparaison des taux de reconnaissance d'expression faciale de plusieurs méthodes	94
4.4.	Conclusion.....	95
CONCLUSION GENERALE		96
REFERENCES BIBLIOGRAPHIQUES.....		97
ANNEXE		105
PRODUCTION SCIENTIFIQUE DE L'AUTEUR.....		106

TABLE DES FIGURES

Figure 1.1: Exemples de modalités biométriques.....	21
Figure 1.2: L'architecture du système biométrique.....	23
Figure 1.3 : Illustration du FRR et du FAR en fonction de scores.	25
Figure 1.4 : Schéma synoptique d'un système de reconnaissance faciale	30
Figure 1.5 : Variabilité intra-classe, similarité inter-classe.....	33
Figure 1.6 : Les différents systèmes multimodaux.....	35
Figure 1.7: Les différents niveaux de fusion biométrique.....	37
Figure 1.8 : Variations dans l'apparence d'un visage (Base CMU MultiPIE)	38
Figure 2.1 : L'organigramme des méthodes de reconnaissance faciale	44
Figure 2.2 : Les expressions faciales universelles tirées de la base de données JAFFEE	45
Figure 2.3 : Modèle de visage mpeg-4 – Facial Animation.....	47
Figure 2.4 : Exemple d'EigenFaces	49
Figure 2.5 : Représentation d'une sphère avec la distance euclidienne et la distance City-Block	52
Figure 2.6 : Représentation de l'hyper plan optimal.....	55
Figure 2.7 : Représentation géométrique de la maximisation de marge (SVM)	56
Figure 2.8 : Exemple de similarités contour et région.....	59
Figure 2.9 : La forme binaire de base de l'ART.....	62
Figure 2.10 : Détection des régions d'intérêt	63
Figure 2.11 : Quelques détections par l'algorithme de Rowley.....	65
Figure 3.1 : Prétraitement : des visages de la base JAFFE avant (original image) et après simple redimensionnement.....	70
Figure 3.2 : Le processus de reconnaissance d'expression faciale.....	71
Figure 3.3 : Le bloc diagramme du système de reconnaissance d'expression faciale	72

Figure 3.4 : <i>Le prétraitement des images des bases de données utilisées (redimensionnement)</i>	73
Figure 3.5 : <i>La fonction de seuillage (la binarisation)</i>	73
Figure 3.6 : <i>La Grille cartésienne de pixels pour calculer les moments ART</i>	76
Figure 3.7 : <i>Schématisation des pixels polaires d'ART</i>	78
Figure 3.8 : <i>Le Block diagramme ACP</i>	80
Figure 3.9 : <i>Extraction de caractéristiques par fusion ART + ACP</i>	82
Figure 4.1 : <i>Echantillons de la base de données JAFFE</i>	85
Figure 4.2 : <i>Échantillons de la base de données Cohn-Canade</i>	85
Figure 4.3 : <i>Échantillons de la base synthétique 3D du logiciel FaceGen</i>	86
Figure 4.4 : <i>Echantillons de la base de données JAFFE (Problème de similarité)</i>	87
Figure 4.5 : <i>La comparaison des taux de reconnaissance d'expression faciale PCA, ART et fusion de PCA+ART</i>	90
Figure 4.6 : <i>Les taux de reconnaissance d'expression faciale pour (ART et P-ART)</i>	89

LISTE DES TABLEAUX

Tableau 2.1 : <i>Classification des techniques de descriptions de formes 2D</i>	60
Tableau 2.2 : <i>Les bases d'images d'expression faciales</i>	66
Tableau 4.1 : <i>Les taux de reconnaissance d'expression faciale donnés par L'ACP</i>	88
Tableau 4.2 : <i>Les taux de reconnaissance d'expression faciale donnés par ART</i>	89
Tableau 4.3 : <i>Les taux de reconnaissance d'expression faciale donnés par la fusion ACP+ART</i>	90
Tableau 4.4 : <i>Les taux de reconnaissance d'expression faciale donnés par P-ART en utilisant la base synthétique</i>	92
Tableau 4.5 : <i>Les taux de reconnaissance d'expression faciale donnés</i>	93
Tableau 4.6 : <i>Comparaison des taux de reconnaissance d'expression faciale entre plusieurs méthodes (JAFPE et CK /CK+)</i>	94

GLOSSAIRE

AAM : Active Appearance Model.

ACP : Analyse en Composantes Principales.

ART : Angular Radial Transform.

AUDN : AU-inspired Deep Network.

CSSD : Curvature Scale Space Descriptor

ADL/ FLD : Analyse discriminante linéaire de Fisher ("Linear discriminant analysis").

EER : Taux d'erreurs égales ("Equal Error Rate").

EFAR : Taux de fausses acceptations des inscrits ("Enrollee False Accept Rate").

FACS : Facial Action Coding System.

FAPU : Facial Animation Parameter Units.

FAR : Taux de fausses acceptations ("False Accept Rate").

FER : Facial Expression Recognition.

FFP : Facial Features Points.

FRR : Taux de faux rejets ("False Reject Rate").

HMM : Modèle de Markov caché ("Hidden Markov model").

ICA : Independent Component Analysis.

IHM : Interaction Homme-Machine.

LBP : Local Binary Pattern.

NN/ CNN : Neural Network / Convolutional Neural Network.

P-ART : Polar –Angular Radial Transform.

RBF : Fonction à base radiale ("Radial Basis Function").

ROC : Courbe représentant les taux d'erreur ("Receiver Operating Curve").

SVM : Séparateur à vaste marge ou machine à vecteurs de support (Support Vector Machine).

TR : Taux de reconnaissance.

ZM / PZMs : Zernick Moments / Pseudo Zernick Moments.

INTRODUCTION

Le visage est un stimulus dont l'importance et la particularité ont été fréquemment remarquées dans la littérature scientifique. Etant la partie la plus expressive et communicative d'un être humain, le visage représente un centre d'intérêt majeur dans les recherches actuelles concernant l'amélioration de l'IHM (Interaction Homme-Machine) pour l'établissement d'un dialogue entre les deux entités.

Les expressions du visage prennent une place importante dans le processus de communication humain et forment à elles seules un langage Co-verbal.

L'intérêt que porte la reconnaissance des expressions faciales réside dans ses applications à savoir : la santé, le médico-légal, le domaine de la linguistique, les sciences comportementales, la psychologie, la langue des signes et la sécurité.

C'est très utile d'en déduire si la personne ressent une des sept émotions universelles : la peur, la colère, la joie, la tristesse, le dégoût, la surprise ou le mépris. C'est un développement très intéressant pour le secteur des jeux vidéo mais aussi pour la médecine comme pour l'évaluation de la douleur chez les malades et les enfants, l'évaluation de nerfs faciaux, le marketing ou encore, moins attendu, la sécurité au volant.

Plusieurs autres domaines sont également concernés par l'analyse des expressions faciales notamment dans l'enseignement à distance (transmission au professeur de l'état des étudiants sous forme d'information de haut niveau), logiciels d'apprentissage, système de mesure objective, dispositif dans les voitures qui avertit le conducteur en cas de perte de vigilance, dispositif dans les avions (information supplémentaire pour la boîte noire), jeux et shopping on-line, la compréhension de l'image et dans la robotique, mais aussi la biométrie comme la reconnaissance de visages invariante à l'expression.

Les systèmes biométriques de reconnaissance de visage 2D permettent d'atteindre un taux de reconnaissance élevé sur des images acquises dans de bonnes conditions. Malheureusement, les performances obtenues avec les algorithmes actuels de reconnaissance de visage en deux dimensions diminuent fortement lorsque les

images sont issues d'acquisitions non contrôlées. Les taux de reconnaissance chutent considérablement quand les visages présentent une variation de pose ou une variation d'expression faciale. C'est dans cette optique que nous proposons d'augmenter la robustesse des taux de reconnaissance d'expression faciale, en fusionnant des méthodes classiques en premier lieu, puis nous présentons une forme nouvelle du descripteur utilisé.

Nous commencerons dans le chapitre I par présenter de manière générale les différentes méthodes d'authentification utilisées. Parmi ces méthodes, la biométrie permet une authentification fiable et sûre. Nous en présenterons les principaux avantages et inconvénients. Cadre de notre thèse, la reconnaissance de visage est l'une des modalités biométriques couramment utilisée. Nous en ferons une présentation plus détaillée dans le but de préciser les enjeux et les motivations de cette thèse. L'exposé des limitations actuelles de cette technologie nous permettra de justifier les motivations ayant conduit aux travaux présentés.

Nous procédons dans le chapitre II à une présentation des principaux travaux traitant de la reconnaissance faciale. Nous porterons une attention particulière aux méthodes de reconnaissance robustes aux variations d'expression faciale en dressant un état de l'art pour exposer par la suite notre orientation et approche effectuée au cours de cette thèse.

Dans une première partie, nous détaillerons l'outil que nous avons utilisé pour représenter un visage (Chapitre III). Celui-ci, nous mènera à exposer notre contribution pratique lors des différentes phases d'un système de reconnaissance d'expression faciale. De par sa construction, il est composé d'une phase de prétraitement, de détection et de reconnaissance faciale.

Nous présenterons les deux méthodes de reconnaissances d'expression utilisées dans notre travail séparément (l'ACP et l'ART), par la suite, nous allons fusionner les deux méthodes. La contribution de notre travail étant aussi de proposer une nouvelle forme mathématique pour la transformation angulaire radiale, baptisée (P-ART) et proposer une fusion des deux méthodes.

Dans la partie expérimentale (chapitre IV), nous aborderons l'aspect pratique après avoir étudié les différentes faces de notre projet. A cet égard, nous avons proposé deux applications pratiques, à savoir, la fusion des deux méthodes de reconnaissance (L'ACP et ART) et la proposition d'un nouvel ART en coordonnées polaires et proposer de fusionner ce dernier à l'ACP (L'ACP et P-ART).

Le dernier chapitre présente les résultats de notre implémentation (images de test et de validation) ainsi que les résultats des taux de reconnaissance d'expression faciale qui sont présentés sous forme de tableaux et de graphes.

CHAPITRE I :

INTRODUCTION AUX SYSTEMES DE RECONNAISSANCE BIOMETRIQUE

1.1. Introduction

La biométrie recense nos caractères physiques (et comportementaux) les plus uniques. Elle permet une authentification fiable et sûre. La reconnaissance de visages est l'une des modalités biométriques couramment utilisée. Nous allons, à travers cette introduction, exposer le cadre de notre thèse. Nous en ferons une présentation plus détaillée dans le but de préciser les enjeux et les motivations de notre travail. L'exposé des limitations actuelles de cette technologie nous permettra de justifier notre approche ainsi que les motivations ayant conduit aux travaux présentés.

1.2. Préambule et cadre général

1.2.1. La Biométrie

La biométrie consiste à identifier une personne à partir de ses caractéristiques physiques ou comportementales. Le visage, les empreintes digitales, l'iris, etc. sont des exemples de caractéristiques physiques. La voix, l'écriture, le rythme de frappe sur un clavier, etc. sont des caractéristiques comportementales (Figure 1.1). Ces caractéristiques, qu'elles soient innées comme les empreintes digitales ou bien acquises comme la signature, sont attachées à chaque individu et ne souffrent donc pas des faiblesses des méthodes basées sur une connaissance ou une possession.

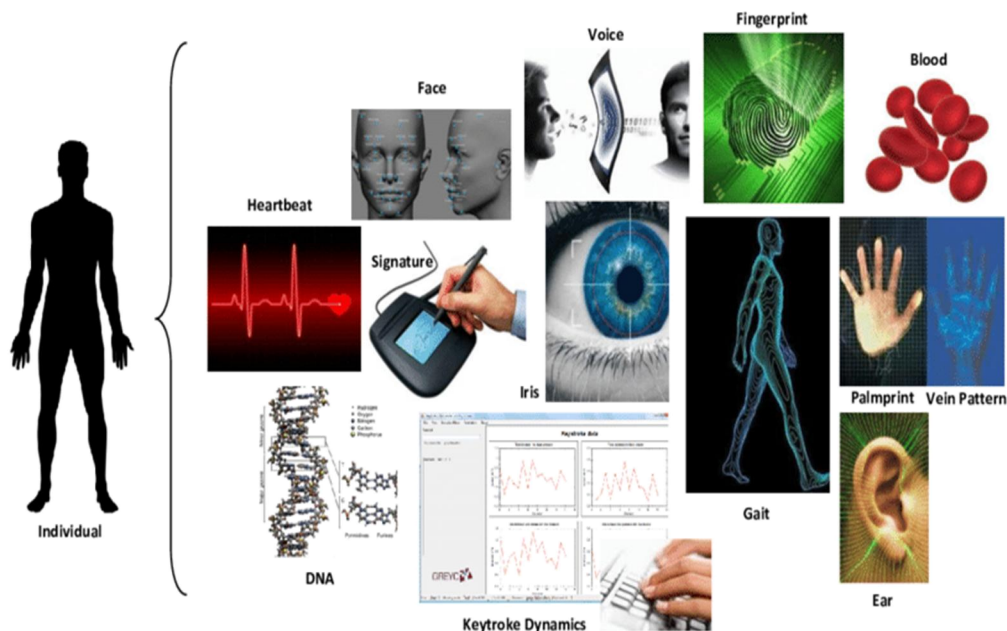


Figure 1.1: Exemples de modalités biométriques [1]

En effet, un attribut physique ou comportemental ne peut être oublié ou perdu. En général, ils sont très difficiles à « deviner » ou à « voler » ainsi qu'à « dupliquer » [1].

1.2.2. Le système biométrique

Un système biométrique est essentiellement un système de reconnaissance de formes qui fonctionne en acquérant des données biométriques à partir d'un individu, extrayant un ensemble de caractéristiques à partir des données acquises, et comparant ces caractéristiques contre la signature dans la base de données. Il sert à vérifier l'identité d'une personne à l'aide d'une ou plusieurs modalités qui lui sont propres (voix, iris, empreintes digitales, visage ...). On peut dire qu'un système de contrôle biométrique est un système automatique de mesure basé sur la reconnaissance de caractéristiques propres à l'individu. Aussi, on peut adapter la définition suivante : Un système de reconnaissance d'individus est un système biométrique qui permet l'identification d'une personne sur la base de caractères physiologiques ou de traits comportementaux automatiquement reconnaissables et vérifiables. On peut définir deux modes de fonctionnement d'un système biométrique [2] :

- Identification : Procédé permettant de déterminer l'identité d'une personne, il ne comprend qu'une étape : L'utilisateur fournit un échantillon biométrique qui va être comparé à tous les échantillons biométriques contenus dans la base de données biométrique du système. Si l'échantillon correspond à celui d'une personne de la base, on renvoie son numéro d'utilisateur.
- Authentification : Procédé permettant de vérifier l'identité d'une personne. Il comprend deux étapes :
L'utilisateur fournit un identifiant « Id » au système de reconnaissance (Numéro) ;
L'utilisateur fournit ensuite un échantillon biométrique qui va être comparé à l'échantillon biométrique correspondant à l'utilisateur « Id » contenu dans la base de données biométrique du système. Si la comparaison correspond, l'utilisateur est authentifié.

1.2.3. L'architecture du système biométrique

Bien que sous formes diverses et mesurant des caractéristiques différentes, les systèmes biométriques partagent tous la même architecture. Ce sont tous des systèmes de reconnaissance de formes. Ils sont composés d'un ou plusieurs systèmes d'acquisition qui vont mesurer le ou les traits physiques ou comportementaux de l'individu. Lorsque le système utilise plusieurs traits qu'il lie à un individu, on parle de systèmes multimodaux.

A l'inverse, si une seule caractéristique est utilisée, on utilise le terme de systèmes uni-modal. Un système d'information extrait, encode, stocke et compare ces données (figure 1.2).

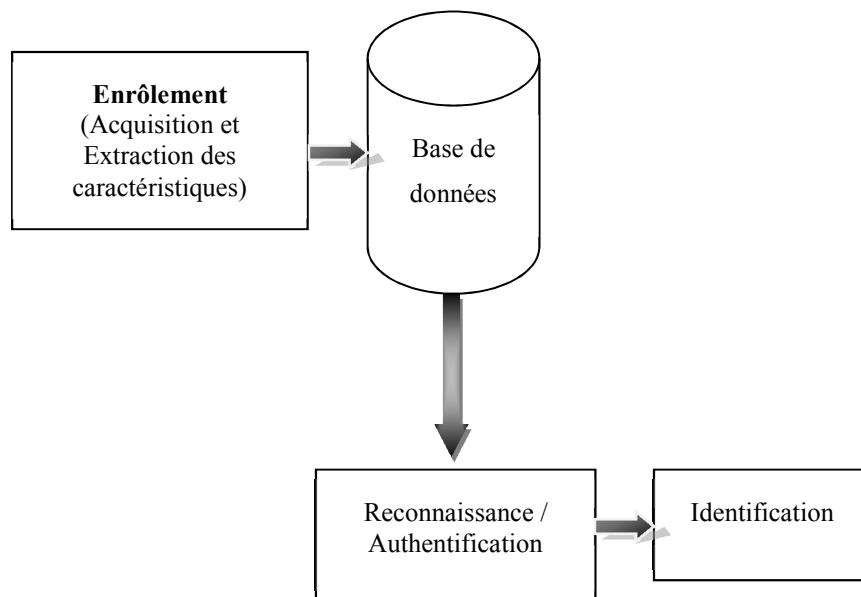


Figure 1.2: *L'architecture du système biométrique*

Les systèmes biométriques sont tous automatisés permettant un traitement rapide de l'information. Que le système soit utilisé en mode authentification ou identification, deux étapes sont nécessaires avant que celui-ci soit pleinement opérationnel [2,3]:

- **Apprentissage :** C'est la phase initiale qui s'effectue une seule fois. Au cours de l'apprentissage, on fait une capture (acquisition) de la caractéristique biométrique. En général cette capture n'est pas directement stockée, des transformations lui sont appliquées, le modèle étant une représentation compacte du signal permettant de faciliter la phase de reconnaissance et de diminuer la quantité des données à stocker. Les tâches essentielles de cette phase sont : l'extraction de paramètres et la construction d'un modèle de représentation (appelé gabarit).
- **Reconnaissance :** Opération se déroulant à chaque fois qu'une personne se présente devant le système, elle consiste en l'extraction d'un ensemble de paramètres comme pour l'étape d'apprentissage suivie d'une autre étape de comparaison et de prise de décision selon le mode opératoire du système : identification (partir de l'échantillon biométrique recherche du gabarit correspondant) ou vérification (échantillon biométrique correspond bien au gabarit).

1.2.4. Evaluation des performances des systèmes biométriques [2,3]

Bien que n'importe quelle mesure physiologique ou comportementale puisse être utilisée dans un système de reconnaissance biométrique, elles ne sont pas toutes aussi performantes les unes que les autres. Ainsi, toute bonne mesure biométrique doit satisfaire les critères suivants :

- **Universelle** : Par ceci, nous voulons dire que la caractéristique biométrique doit être présente pour tous les individus. Par exemple, il est impossible de recueillir l'empreinte digitale d'un individu qui a été amputé de la main ou bien de mesurer la démarche d'une personne quadriplégique.
- **Distinctive** : La mesure effectuée sur un individu doit être suffisamment différente de celles effectuées sur les autres individus pour permettre de les discriminer entre eux. Ici, on peut penser au code génétique qui varie significativement d'une personne à l'autre tandis que la taille est une mesure que plusieurs individus partagent.
- **Permanente** : Le fait que le corps humain vieillit implique qu'après un certain laps de temps, une mesure biométrique faite sur un individu peut être très différente de la mesure initiale utilisée pour l'inscription dans le système. Par conséquent, il est important de choisir une caractéristique biométrique qui reste stable durant la vie de l'individu. Un bon exemple de ceci est le motif présent dans l'iris ou même le code génétique de l'individu.
- **Facile à mesurer** : Ceci représente à quel point il est facile de recueillir et de quantifier la mesure biométrique. Il est facile d'enregistrer le son de la voix d'un individu à l'aide d'un microphone, mais il est beaucoup moins facile d'obtenir une image de bonne qualité de la rétine des individus.
- **Efficace** : L'efficacité fait référence à la quantité de ressources nécessaires afin d'obtenir le niveau de qualité désiré dans le temps requis. Un test sanguin chimique requiert du matériel sophistiqué et une longue période de temps avant d'obtenir un résultat comparativement à recueillir l'image du visage d'un individu à l'aide d'une caméra numérique.
- **Acceptable** : Ceci correspond aux aspects socioculturels de la reconnaissance biométrique. Même si avec la surabondance de caméras de sécurité, la population normale ne se soucie plus d'avoir son image recueillie par de multiples systèmes.
- **Robuste** : Par ceci, nous faisons référence à la difficulté de forcer le système à produire une prédiction erronée par l'utilisation d'une technique frauduleuse.

L'utilisation de technique de maquillage pour effets spéciaux a déjà permis de tromper autant les humains que les machines lors de la reconnaissance faciale.

1.2.5. La mesure de la performance du système biométrique

Tout d'abord, afin de comprendre comment déterminer la performance d'un système biométrique, il nous faut définir clairement trois critères principaux.

1. Le premier critère s'appelle le taux de faux rejet ("False Reject Rate" ou FRR). Ce taux représente le pourcentage de personnes censées être reconnues mais qui sont rejetées par le système.

$$FRR = \frac{\text{Nombre des clients rejetés (FR)}}{\text{Nombre total d'accès de clients}}$$

Tel que FR Le faux rejet correspond au cas où le système rejette un client légitime.

2. Le deuxième critère est le taux de fausse acceptation ("False Accept Rate" ou FAR). Ce taux représente le pourcentage de personnes censées ne pas être reconnues mais qui sont tout de même acceptées par le système.

$$FAR = \frac{\text{Nombre des imposteurs acceptés (FA)}}{\text{Nombre total d'accès imposteurs}}$$

Tel que FA correspond au cas où le système accepte un individu qui a proclamé une identité qui n'est pas la sienne.

3. Le troisième critère est connu sous le nom de taux d'égale erreur ("Equal Error Rate" ou EER). Ce taux est calculé à partir des deux premiers critères et constitue un point de mesure de performance courant. Ce point correspond à l'endroit où $FRR = FAR$, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations [2].

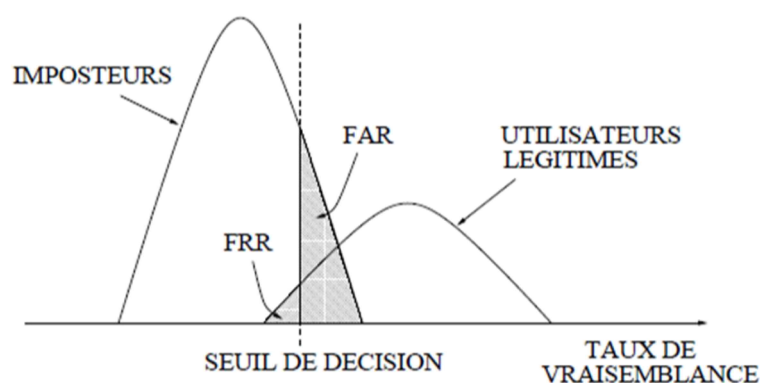


Figure 1.3 : Illustration du FRR et du FAR en fonction de scores

1.2.6. Les applications des systèmes biométriques

Les applications de la biométrie peuvent être divisées en trois groupes principaux [3] :

- Applications commerciales : telles que l'ouverture de réseau informatique, la sécurité de données électroniques, l'e-commerce, l'accès Internet, la carte de crédit, le contrôle d'accès physique, le téléphone cellulaire, la gestion des registres médicaux, l'étude à distance, etc.
- Applications gouvernementales: telles que la carte d'identité nationale, le permis de conduire, la sécurité sociale, le contrôle des frontières, le contrôle des passeports, etc.
- Applications légales : telles que l'identification de corps, la recherche criminelle, l'identification de terroriste, etc.

1.2.7. La biométrie du visage

Rien n'est plus naturel qu'utiliser le visage pour identifier une personne. Les images faciales sont probablement la caractéristique biométrique la plus communément employée par l'homme pour effectuer une identification personnelle. L'utilisation d'une caméra permet de capter la forme du visage d'un individu et d'en dégager certaines particularités.

Selon le système utilisé, l'individu doit être positionné devant l'appareil ou peut être en mouvement à une certaine distance. Les données biométriques qui sont obtenues sont par la suite comparées au fichier référence.

Durant les vingt dernières années, la reconnaissance automatique des visages est devenue un enjeu primordial, notamment dans les domaines de l'indexation de documents multimédias et surtout dans la sécurité, ceci est dû aux besoins du monde actuel mais aussi à ses caractéristiques avantageuses dont on peut citer :

- La disponibilité des équipements d'acquisition, leur simplicité et leurs coûts faibles.
- Passivité du système : un système de reconnaissance de visages ne nécessite aucune coopération de l'individu, du genre : mettre le doigt ou la main sur un dispositif spécifique ou parler dans un microphone. En effet, la personne n'a qu'à rester ou marcher devant une caméra pour qu'elle puisse être identifiée par le système.

Les modes de fonctionnement (Surveillance ou Recherche) [3]

Mode Surveillance : Le système peut détecter automatiquement la forme d'un visage, en extraire l'image, exécuter l'identification en s'appuyant sur une base de données d'individus préenregistrés.

Il calcule rapidement le degré de recoupement entre l'empreinte du visage réel qui vient d'être capté et ceux associés aux individus connus enregistrés dans une base de données biométrique d'images faciales.

Il peut retourner une liste d'individus possibles triée par score décroissant (images ressemblantes) ou il peut simplement retourner l'identité du sujet (résultat le plus haut) et un niveau de confiance associé. Ceci paramétrable au choix de l'opérateur.

En mode Surveillance il fonctionne en temps réel et dans le même temps, capte les images et recherche dans la base de données. L'architecture technique mise en place pour cela est fonction des besoins de l'application (nombre d'individus enregistrés, vitesse de défilement devant la caméra, flux...)

Une autre fonction du mode Surveillance permet de trouver des visages humains n'importe où dans le champ de vision du système et à toute distance. Il peut les suivre de façon continue et les extraire de l'image, en comparant le visage ainsi isolé avec une liste de visages stockés.

Mode **Recherche** : Le mode Recherche peut être utilisé en mode IDENTIFICATION (un à plusieurs) ou en mode VÉRIFICATION (un à un).

– Le système travaille à partir d'une image de visage stable. Cette image (photo) peut provenir de différentes sources : Appareil photo numérique, extrait de bande vidéo, photo scannée...etc.

– Il calcule alors rapidement le degré de recoupement entre l'empreinte du visage réel qui est présenté et ceux associés aux individus connus enregistrés dans une base de données biométrique d'images faciales. – Il peut retourner une liste d'individus possibles triée par score décroissant (images ressemblantes) ou il peut simplement retourner l'identité du sujet (résultat le plus haut) et un niveau de confiance associé.

En mode **vérification**, l'empreinte du visage peut être stockée sur une carte à puce ou sur un support informatisé. Le système associe simplement l'empreinte réelle à celle

enregistrée sur le support. Si le niveau de confiance dépasse un certain seuil, alors l'association est réussie et l'identité est vérifiée.

– Les systèmes peuvent compresser une empreinte du visage en moins de 100 octets pour usage dans les cartes à puce, codes-barres et autres appareils de stockage à capacité limitée.

1.2.8. Le système de reconnaissance faciale

Le système de reconnaissance exploite les caractéristiques du visage ainsi extraites pour créer une signature numérique qu'il stocke dans une base de données. Ainsi, à chaque visage de la base est associée une signature unique qui caractérise la personne correspondante. La reconnaissance d'un visage requête est obtenue par l'extraction de la signature requête correspondante et sa mise en correspondance avec la signature la plus proche dans la base de données. La reconnaissance dépend du mode de comparaison utilisé : vérification ou identification. On peut représenter le système de reconnaissance faciale par la figure suivante (Figure 1.4).

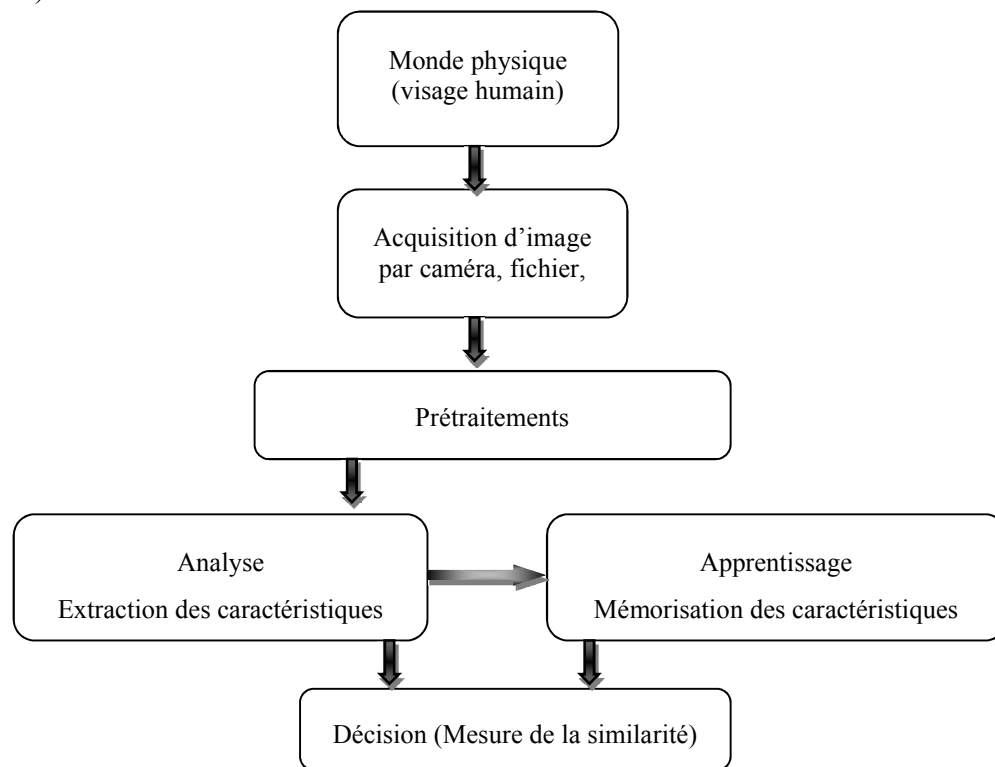


Figure 1.4 : Schéma synoptique d'un système de reconnaissance faciale

Donc pour être identifié, l'image d'une personne dans un système de reconnaissance faciale suit les étapes suivantes :

- Le monde physique (l'extérieur) : C'est le monde réel en dehors du système avant l'acquisition de l'image. Dans cette étape, on tient compte généralement de trois

paramètres essentiels : L'éclairage, la posture et l'échelle. La variation de l'un de ces trois paramètres peut conduire à une distance entre deux images du même individu, supérieure à celle séparant deux images de deux individus différents, et par conséquent une fausse identification.

- L'Acquisition de l'image : Cette étape consiste à capturer l'image de l'utilisateur du monde extérieur dans un état statique à l'aide d'un appareil photo ou dynamique à l'aide d'une caméra. Après, l'image extraite sera digitalisée ce qui donne lieu à une représentation bidimensionnelle au visage. L'image dans cette étape est dans un état brut ce qui engendre un risque de bruit qui peut dégrader les performances du système.
- Les prétraitements : Le rôle de cette étape est d'éliminer les parasites causés par la qualité des dispositifs optiques ou électroniques lors de l'acquisition de l'image en entrée, dans le but de ne conserver que les informations essentielles et donc préparer l'image à l'étape suivante. Elle est indispensable car on ne peut jamais avoir une image sans bruit à cause du background et de la lumière qui est généralement inconnue. Il existe plusieurs types de traitement et d'amélioration de la qualité de l'image, telle que : la normalisation, l'égénéralisation et le filtre médian. Cette étape peut également contenir la détection et la localisation du visage dans une image, surtout là où le décor est très complexe.
- L'extraction de paramètres : l'étape de l'extraction des paramètres représente le cœur du système de reconnaissance, elle consiste à effectuer le traitement de l'image dans un autre espace de travail plus simple et qui assure une meilleure exploitation de données, et donc permettre l'utilisation, seulement, des informations utiles, discriminantes et non redondantes.
- La classification (Modélisation) : Cette étape consiste à modéliser les paramètres extraits d'un visage ou d'un ensemble de visages d'un individu en se basant sur leurs caractéristiques communes. Un modèle est un ensemble d'informations utiles, discriminantes et non redondantes qui caractérise un ou plusieurs individus ayant des similarités.
- L'apprentissage : C'est l'étape où on fait apprendre les individus au système, elle consiste à mémoriser les paramètres, après extraction et classification, dans une base de données bien ordonnée pour faciliter la phase de reconnaissance et la prise d'une décision, elle est en quelque sorte la mémoire du système.
- La décision : C'est l'étape qui fait la différence entre un système d'identification d'individus et un autre de vérification. Dans cette étape, un système d'identification

consiste à trouver le modèle qui correspond le mieux au visage pris en entrée à partir de ceux stockés dans la base de données, il est caractérisé par son taux de reconnaissance. Par contre, dans un système de vérification il s'agit de décider si le visage en entrée est bien celui de l'individu (modèle) proclamé ou il s'agit d'un imposteur, il est caractérisé par son EER (Equal Error Rate).

1.2.9. Les principales difficultés de la reconnaissance faciale

Pour le cerveau humain, le processus de la reconnaissance de visages est une tâche visuelle de haut niveau. Bien que les êtres humains puissent détecter et identifier des visages dans une scène sans beaucoup de peine, construire un système automatique qui accomplit de telles tâches représente un sérieux défi. Ce défi est d'autant plus grand lorsque les conditions d'acquisition des images sont très variables. Il existe deux types de variations associées aux images de visages: inter et intra sujet. La variation inter-sujet est limitée à cause de la ressemblance physique entre les individus. Par contre la variation intra-sujet est plus vaste. Elle peut être attribuée à plusieurs facteurs que nous analysons ci-dessous.

- **Changement d'illumination :** Les variations d'éclairage rendent la tâche de reconnaissance de visage très difficile. En effet, le changement d'apparence d'un visage du à l'illumination, se révèle parfois plus critique que la différence physique entre les individus, et peut entraîner une mauvaise classification des images d'entrée.
- **Variation de pose :** Le taux de reconnaissance de visage baisse considérablement quand des variations de pose sont présentes dans les images. La variation de pose est considérée comme un problème majeur pour les systèmes de reconnaissance faciale. Quand le visage est de profil dans le plan image (orientation $< 30^\circ$), il peut être normalisé en détectant au moins deux traits faciaux (passant par les yeux). Cependant, lorsque la rotation est supérieure à 30° , la normalisation géométrique n'est plus possible.
- **Présence ou absence des composants structurels :** La présence des composants structurels telle que la barbe, la moustache, ou bien les lunettes peut modifier énormément les caractéristiques faciales telles que la forme, la couleur, ou la taille du visage. De plus, ces composants peuvent cacher les caractéristiques faciales de base causant ainsi une défaillance du système de reconnaissance.
- **Les vrais jumeaux :** Qui ont le même indicatif d'ADN, peuvent tromper les personnes qui ne les connaissent pas (les personnes familiers avec les jumeaux ont

reçu une grande quantité d'information sur ces derniers et sont donc beaucoup plus qualifiés à distinguer les jumeaux.).

- **Expressions faciales :** La déformation du visage qui est due aux expressions faciales est localisée principalement sur la partie inférieure du visage. L'information faciale se situant dans la partie supérieure du visage reste quasi invariable. Elle est généralement suffisante pour effectuer une identification. Toutefois, étant donné que l'expression faciale modifie l'aspect du visage, elle entraîne forcément une diminution du taux de reconnaissance. L'identification de visage avec expression faciale est un problème difficile qui est toujours d'actualité et qui reste non résolu.

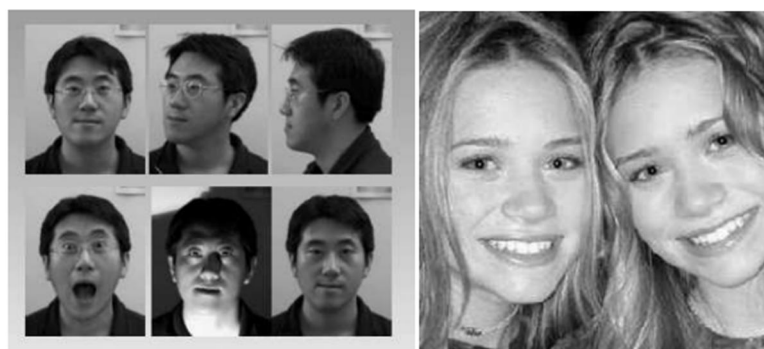


Figure 1.5 : *Variabilité intra-classe, similarité inter-classe*

1.2.10. La multimodalité

Face aux nombreuses limitations imposées par l'utilisation des systèmes biométriques unimodaux, la biométrie multimodale s'impose de manière indéniable comme une alternative d'avenir dans le domaine de la sécurité des personnes et leurs biens [4].

Bien que les techniques de reconnaissance biométrique promettent d'être très performantes, on ne peut garantir actuellement un excellent taux de reconnaissance avec des systèmes biométriques unimodaux, basés sur une unique signature biométrique. De plus, ces systèmes sont souvent affectés par les problèmes suivants :

- Bruit introduit par le capteur
- Non-universalité
- Manque d'individualité
- Manque de représentation invariante
- Sensibilité aux attaques.

Ainsi, à cause de tous ces problèmes pratiques, les taux d'erreur associés à des systèmes biométriques unimodaux sont relativement élevés, ce qui les rend inacceptables pour un

déploiement d'applications critiques de sécurité. Pour pallier à ces inconvénients, une solution est l'utilisation de plusieurs modalités biométriques ou de fusionner plusieurs algorithmes au sein d'un même système.

1.2.11. Les différentes formes de multimodalité

La multimodalité consiste à utiliser plusieurs modalités biométriques ou algorithmes dans le processus d'identification ou de vérification de personnes. Dans un système biométrique multimodal, la fusion peut se faire à des niveaux différents [3,4] :

- Systèmes multiples biométriques : Lorsque l'on considère plusieurs biométries différentes, par exemple visage et empreinte digitale. C'est le sens le plus classique du terme multimodal.
- Systèmes multiples d'acquisition : Par exemple utiliser deux scanners différents (l'un optique, l'autre thermique) pour la reconnaissance d'empreintes digitales.
- Mesures multiples d'une même unité biométrique : Lorsqu'ils associent plusieurs échantillons différents de la même modalité, par exemple deux empreintes digitales de doigts différents ou les deux iris.
- Instances multiples d'une même mesure : Faire une capture répétée du même attribut biométrique avec le même système d'acquisition, par exemple l'acquisition de plusieurs images de visage avec des changements de pose, d'expression ou d'illumination.
- Algorithmes multiples : Lorsque plusieurs algorithmes traitent la même image acquise, cette multiplicité des algorithmes peut intervenir dans le module d'extraction en considérant plusieurs ensembles de caractéristiques et/ou dans le module de comparaison en utilisant plusieurs algorithmes de comparaison.

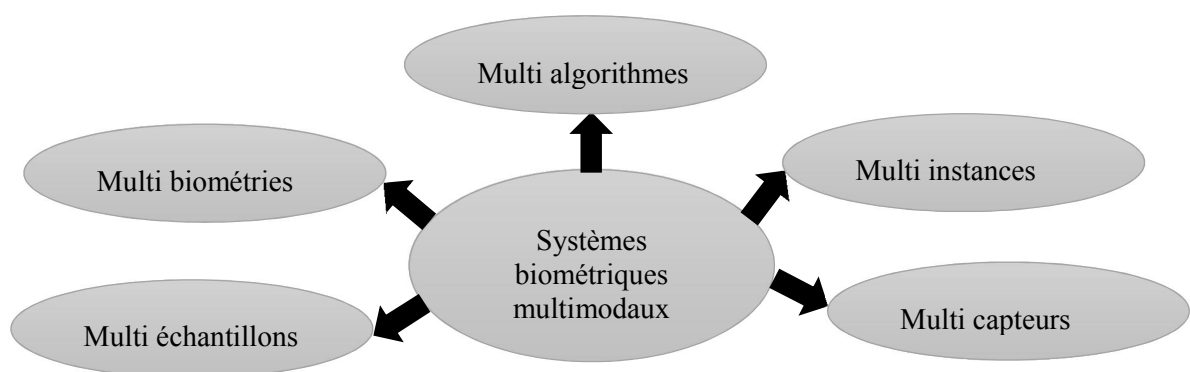


Figure 1.6 : Les différents systèmes multimodaux

1.2.12. Les architectures

Il existe deux types d'architectures :

- L'architecture en parallèle : correspond à l'acquisition et le traitement des données simultanément des systèmes biométriques associés pour la conception d'un système multimodal. C'est la plus utilisée car elle permet d'utiliser toutes les informations disponibles et donc d'améliorer les performances du système.
En revanche, l'acquisition et le traitement d'un grand nombre de données biométriques est coûteux en temps et en matériel, et réduit le confort d'utilisation.
- L'architecture en série : On parlera de l'architecture en série lorsque l'acquisition et le traitement des données de ces systèmes se font successivement. Elle peut être privilégiée dans certaines applications ; par exemple si la multimodalité est utilisée pour donner une alternative pour les personnes ne pouvant pas utiliser l'empreinte digitale. Pour la majorité des individus seule l'empreinte est acquise et traitée mais pour ceux qui ne peuvent pas être ainsi authentifiés on utilise un système à base d'iris alternativement.

1.2.13. La fusion des données

La combinaison de plusieurs systèmes biométriques peut se faire à quatre niveaux différents : au niveau des données, au niveau des caractéristiques extraites, au niveau des scores issus du module de comparaison ou au niveau des décisions du module de décision [4,5].

- Fusion au niveau de capteur : La fusion au niveau de capteur peut se faire uniquement si les diverses captures sont des instances du même trait biométrique obtenu à partir de plusieurs capteurs compatibles entre eux ou plusieurs instances du même trait biométrique obtenu à partir d'un seul capteur.

De plus, les captures doivent être compatibles entre elles et la correspondance entre les points dans les données brutes doit être connue par avance. Par exemple, les images de visage obtenues à partir de plusieurs caméras peuvent être combinées pour former un modèle 3D du visage. Un autre exemple de fusion au niveau de capteur consiste à mettre en mosaïque plusieurs images d'empreintes digitales afin de former une image d'empreinte digitale finale plus complexe. La fusion au niveau de capteur n'est généralement pas possible si les instances

des données sont incompatibles (par exemple, il est peut être difficile de fusionner des images de visages provenant de caméras ayant des résolutions différentes).

- Fusion au niveau de caractéristiques : cette fusion consiste à combiner différents vecteurs de caractéristiques qui sont obtenus à partir des sources suivantes ; plusieurs capteurs du même trait biométrique, plusieurs instances du même trait biométrique, plusieurs unités du même trait biométrique ou encore plusieurs traits biométriques. Quand les vecteurs de caractéristiques sont homogènes (par exemple, plusieurs images d’empreinte digitale du doigt d’un utilisateur), un unique vecteur de caractéristiques résultant peut être calculé comme une somme pondérée des vecteurs de caractéristiques individuels. Lorsque les vecteurs de caractéristiques sont hétérogènes (par exemple, des vecteurs de caractéristiques de différentes modalités biométriques comme l’iris et l’empreinte digitale), nous pouvons les concaténer pour former un seul vecteur de caractéristiques [4,5].
- Fusion au niveau de scores : Après les vecteurs de caractéristiques, les scores (de correspondance) donnés en sortie par les matchers contiennent l’information la plus riche à propos du modèle d’entrée. En fait, la fusion au niveau de scores donne le meilleur compromis entre la richesse d’information et la facilité d’implémentation. Aussi, il est relativement facile d’accéder et de combiner les scores générés par les différents matchers. En conséquence, l’intégration d’information au niveau scores est l’approche la plus courante dans les systèmes biométriques multimodaux.
- Fusion au niveau de décision : L’intégration d’information au niveau de décision peut être mis en place lorsque chaque matcher biométrique décide individuellement de la meilleure correspondance possible selon l’entrée qui lui est présentée. Les méthodes comme le majority voting, le behavior knowledge space, le weighted voting basé sur la théorie Dempster-Shafer, les règles ET et OU, etc. peuvent être utilisées afin d’arriver à la décision finale [4,5].

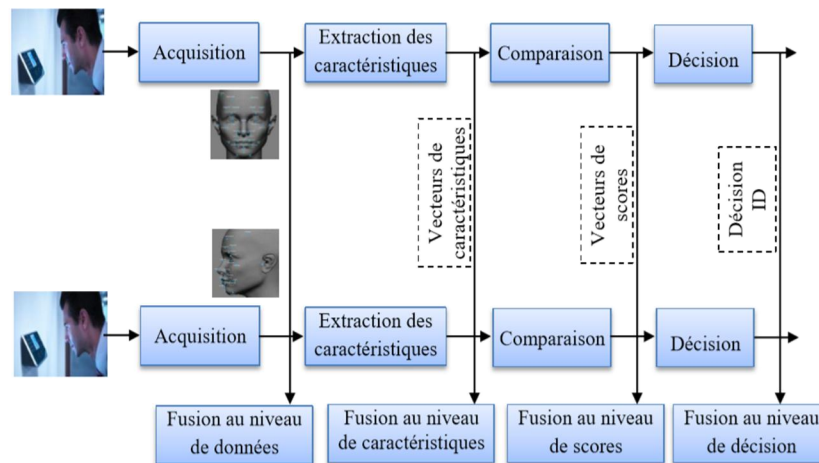


Figure 1.7: Les différents niveaux de fusion biométrique

1.3. La problématique

Tout ce qui a été présenté jusque-là nous a permis de constater que la biométrie, de manière générale, et que la reconnaissance de visage suscite de plus en plus l'intérêt de la communauté scientifique, car elle présente plusieurs challenges et verrous technologiques. Nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance automatique de visages, ce qui nous a permis de définir les problématiques traitées dans ce mémoire.

Divers changements au niveau des images présentent un grand défi face à un système de reconnaissance qui doit être robuste en ce qui concerne les grandes variabilités des images du visage comme les expressions faciales, la pose du visage et l'éclairage. Pour faire face à ce problème, il est important de choisir une représentation appropriée des images du visage. Cette représentation doit être compacte et significative.

Nous avons ensuite présenté une façon de réduire les limitations des systèmes biométriques monomodaux en combinant plusieurs systèmes ou algorithmes conduisant ainsi à fiabiliser le système de reconnaissance. Les systèmes multimodaux peuvent être de différentes natures, nous avons donc présenté dans ce chapitre, les différents types de combinaisons de modalités possibles, mais aussi les architectures et les niveaux de fusion qui peuvent être utilisés dans un tel système.

Dans ce travail, nous avons décidé de se focaliser sur l'une des majeurs problématiques de la reconnaissance faciale (le problème de variation d'expression faciale), cela consiste à reconnaître l'expression faciale sur une personne bien déterminée (dans un visage donné), dans le but d'augmenter les taux de reconnaissances et les performances du FER.

1.4. Les objectifs de la thèse

Rien n'est plus naturel qu'utiliser le visage pour identifier une personne. Les images faciales sont probablement la caractéristique biométrique la plus communément employée par l'homme pour effectuer une identification personnelle. De plus, la zone du corps la plus chargée de sens est sans conteste le visage, il est possible d'extraire une très grande quantité d'informations depuis cette zone, ce qui nous a nettement motivés pour la traiter comme modalité.

"Le problème de reconnaissance faciale est rendu difficile par la grande variabilité dans la pose du visage, l'intensité de l'éclairage et l'expression du visage."



Figure 1.8 : *Variations dans l'apparence d'un visage (Base CMU MultiPIE) [6]*

Dans cette thèse, nous axerons notre étude sur les problèmes liés aux variations d'expressions, l'objectif final étant de rendre les systèmes actuels de reconnaissance faciale robustes aux variations simultanées d'expression et de pose du visage. Pour traiter cette problématique, nous avons fusionné les caractéristiques de deux méthodes complémentaires, une classique, à savoir la méthode globale la plus incontournable l'ACP (Analyse des composantes Principales) et la deuxième moins classique ; qui est la (Transformation Angulaire Radiale) ART.

ART est un descripteur robuste à la rotation et à la variation d'échelle, d'où son utilisation dans notre système de reconnaissance d'expression faciale. Il est souvent utilisé dans sa forme cartésienne.

Le premier objectif de cette thèse est de reconnaître l'expression faciale sur une personne bien déterminée, en utilisant séparément les méthodes et algorithmes de reconnaissance ou en fusionnant les deux experts pour améliorer le taux de reconnaissance d'expression faciale.

Le deuxième objectif est de développer une nouvelle forme de la transformation angulaire radiale, nous présenterons une version polaire d'ART, baptisée (P-ART). En troisième lieu, nous fusionnons l'ACP et le P-ART. La fusion donne de meilleurs résultats que plusieurs autres méthodes utilisées dans la littérature scientifique.

1.5. Nos contributions

Ce travail présente en premier lieu, la fusion de plusieurs experts, à savoir (ACP+SVM), (ART+SVM), (ACP+ART). Ensuite, une nouvelle forme d'ART a été proposée pour la reconnaissance d'expression faciale (FER) utilisant un développement mathématique polaire basé sur la transformation radiale angulaire. La transformation radiale angulaire polaire, baptisée (**P-ART**) est combinée à l'analyse en composantes principales (ACP). Le nouvel ART est un descripteur puissant en termes de description, il représente une version nouvelle que celle du descripteur cartésien conventionnel.

L'entraînement de la machine à vecteurs de support (SVM) est utilisé pour reconnaître l'expression faciale pour une image de visage d'entrée. Enfin, les résultats expérimentaux montrent la performance de l'ACP et du P-ART. La fusion de ces deux techniques, peut être meilleure que d'autres méthodes existantes de reconnaissance de l'expression faciale. Au cours de l'expérience, la base japonaise de l'expression faciale féminine (JAFPE) et les bases de données Cohn-Kanade ont été utilisées.

Les principales contributions du document sont donc :

- L'utilisation du descripteur ART pour la reconnaissance de l'expression du visage.
- Étendre l'ART de sa forme cartésienne à une forme polaire.
- La fusion des caractéristiques d'ART (descripteur de forme) avec la méthode holistique basée sur l'apparence l'ACP.

1.6. Conclusion

Dans ce premier chapitre, nous avons présenté les technologies utilisées dans les systèmes biométriques pour l'identification de personnes. Nous avons aussi donné un aperçu sur les techniques de mesure de leurs performances. Cette étude nous a permis de constater que la reconnaissance de visage présente plusieurs challenges. Nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance automatique de visages, ce qui nous a permis de bien définir les problématiques traitées dans ce mémoire et les objectifs de ce travail. Les techniques utilisées aux différentes étapes de la reconnaissance de visage sont détaillées dans le chapitre suivant.

CHAPITRE II :

LES METHODES DE RECONNAISSANCE D'EXPRESSION FACIALE

2.1. Introduction

Dans ce chapitre, nous présenterons les méthodes de reconnaissances faciales en deux dimensions. Ces méthodes peuvent être classées en deux catégories. Les approches holistiques utilisent une représentation globale des visages tandis que les approches locales proposent de les caractériser par un ensemble de descripteurs locaux. Nous présenterons les principaux travaux relatifs à ces deux approches. Dans un second temps, nous nous concentrerons sur les méthodes traitant de la robustesse aux variations intra-classe du visage et plus précisément sur les variations d'expression.

Le but de ce chapitre est de donner un panorama des méthodes les plus significatives en reconnaissance de visages.

2.2. Les méthodes de reconnaissance faciale

L'étude de la reconnaissance automatique de visage a commencé à être étudiée depuis le début des années 1970. Les travaux effectués par Kanade [7] au cours de sa thèse sont considérés comme les premiers sur ce domaine. Depuis, de nombreux travaux de recherche ont eu lieu. Nous pouvons regrouper les méthodes de reconnaissance en trois groupes (figure 2.1) :

Les méthodes globales (holistiques) utilisent une représentation globale des visages, tandis que les approches locales proposent de les caractériser par un ensemble de descripteurs locaux. Les méthodes hybrides combinent les deux.

2.2.1. Les méthodes globales

Les méthodes globales sont basées sur des techniques d'analyse statistique bien connues. Il n'est pas nécessaire de repérer certains points caractéristiques du visage (comme Les centres des yeux, le centre de la bouche, etc.) à part pour normaliser les images. Dans ces méthodes, les images de visage (qui peuvent être vues comme des matrices de valeurs de pixels) sont traitées de manière globale et sont généralement transformées en vecteurs, plus faciles à manipuler. L'avantage principal des méthodes globales est qu'elles sont relativement rapides à mettre en œuvre et que les calculs de base sont d'une complexité moyenne. En revanche, elles sont très sensibles aux variations d'éclairement, de pose et d'expression faciale. Ceci se comprend aisément puisque la moindre variation des conditions de

l'environnement entraîne des changements inéluctables dans les valeurs des pixels qui sont traités directement. Ces méthodes utilisent principalement une analyse de sous-espaces de visages. Cette expression repose sur un fait relativement simple : une classe de "formes" qui nous intéresse (dans notre cas, les visages) réside dans un sous-espace de l'espace de l'image d'entrée. Ainsi, la représentation de l'image originale est très redondante et la dimensionnalité de cette représentation pourrait être grandement réduite si l'on se concentre uniquement sur les formes qui nous intéressent.

L'utilisation de techniques de modélisation de sous-espace a fait avancer la technologie de reconnaissance faciale de manière significative. Nous pouvons distinguer deux types de techniques parmi les méthodes globales, les techniques linéaires et les techniques non linéaires. Parmi les méthodes globale les plus connues : l'ACP, l'ADL, ...etc.

2.2.2. Les méthodes locales

Les méthodes locales, basées sur des modèles, utilisent des connaissances a priori que l'on possède sur la morphologie du visage et s'appuient en général sur des points caractéristiques de celui-ci. Kanade présenta un des premiers algorithmes de ce type [7] en détectant certains points ou traits caractéristiques d'un visage puis en les comparant avec des paramètres extraits d'autres visages. Ces méthodes constituent une autre approche pour prendre en compte la non-linéarité en construisant un espace de caractéristiques local et en utilisant des filtres d'images appropriés, de manière à ce que les distributions des visages soient moins affectées par divers changements. Toutes ces méthodes ont l'avantage de pouvoir modéliser plus facilement les variations de pose, d'éclairage et d'expression par rapport aux méthodes globales. Toutefois, elles sont plus lourdes à utiliser puisqu'il faut souvent placer manuellement un assez grand nombre de points sur le visage alors que les méthodes globales ne nécessitent de connaître que la position des yeux afin de normaliser les images, ce qui peut être fait automatiquement et de manière assez fiable par un algorithme de détection. Dans cette catégorie, on trouve plusieurs méthodes comme : les filtres de Gabor, Dynamic link architecture, HMM...etc.

2.2.3. Les méthodes hybrides

Les méthodes hybrides permettent d'associer les avantages des méthodes globales et locales en combinant la détection de caractéristiques géométriques (ou structurales) avec l'extraction de caractéristiques d'apparence locales. Elles permettent d'augmenter la stabilité

de la performance de reconnaissance lors de changements de pose, d'éclairement et d'expressions faciales.

Un état plus détaillé sur les différentes méthodes de reconnaissance faciale est présenté dans la section suivante.

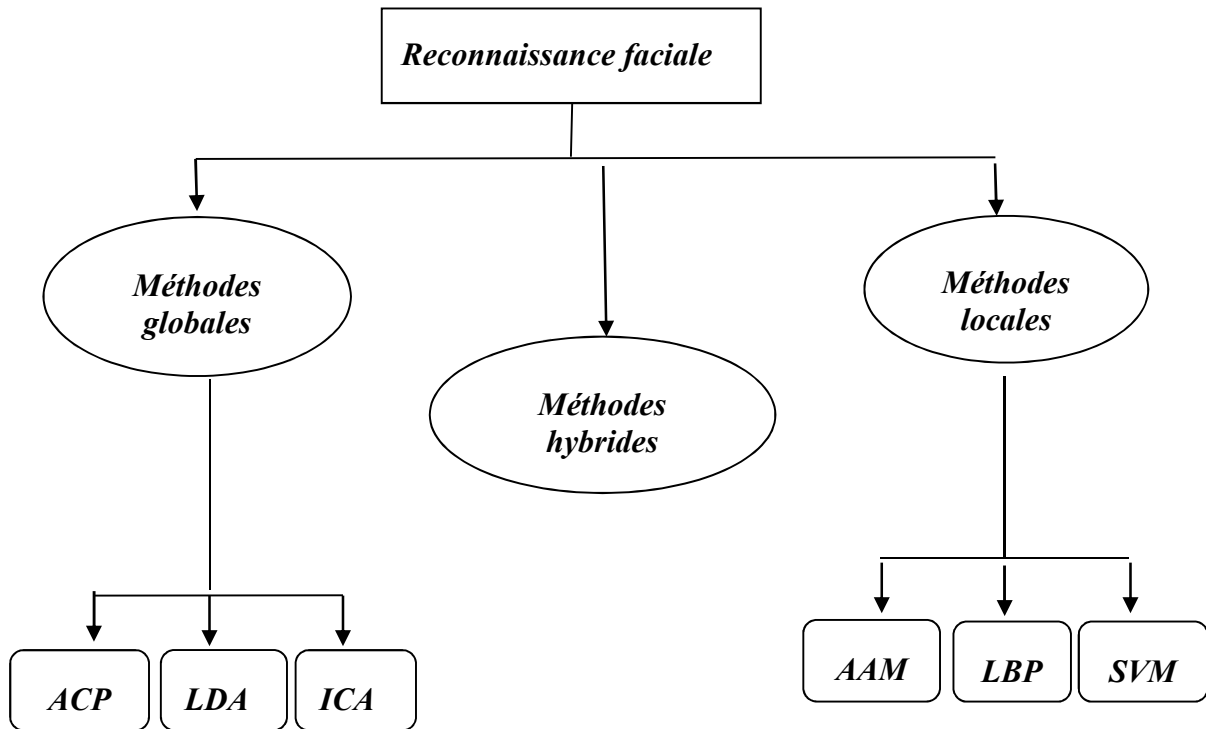


Figure 2.1 : *L'organigramme des méthodes de reconnaissance faciale*

2.3. Les méthodes utilisées dans le FER

Le visage transmet nos émotions, souvent de façon involontaire par ce que l'on appelle communément les expressions faciales. De nombreuses applications émergentes visent à remplacer l'expert humain dans l'analyse des comportements humains et en particulier de ses expressions faciales, en médecine, en psychologie et en sécurité.

Les expressions faciales jouent un rôle majeur, dans la mesure où elles sont un vecteur de la parole, de l'emphase des conversations et des émotions [8, 9]. Certaines études ont même montré que les expressions faciales aident la compréhension des conversations [10], [11], voire même sont la principale modalité des communications humaines : en effet, Mehrabian [12] indique que la partie verbale d'un message (c'est-à-dire les mots prononcés) constituent seulement 7% des effets du message, la partie vocale (par exemple les intonations),

contribuent à 38% et les expressions faciales et le langage corporel du locuteur ont un impact de 55% sur les effets du message verbal.

L'analyse automatique des expressions faciales reste actuellement un challenge. La principale difficulté est liée au fait que les expressions faciales sont le résultat de la déformation du visage par les muscles. L'enjeu majeur pour caractériser une expression faciale consiste alors à identifier de façon unique une expression à partir des informations visibles, qui mixent l'information utile (expression) à d'autres informations (texture, région, éclairage, etc.)

Un système de reconnaissance d'expression faciale assure un rôle significatif dans l'interaction homme-machine (IHM) [13] et l'interaction homme-robot (IHR). Pour une IHR efficace, il est recommandé pour un robot de reconnaître et interagir avec l'expression faciale de l'utilisateur.

Ce domaine a fait l'objet de plusieurs recherches scientifiques couvrant plusieurs disciplines et domaines d'expertise (ingénierie, informatique, psychologie, neurologies) [14]. Les sciences de la médecine utilisent des logiciels basés sur l'expression faciale comme un outil d'aide pour la mesure des douleurs sentis par les nouveaux nés [15].

C'est le psychologue Paul Ekman, qui s'est intéressé à partir du milieu des années 1960 aux expressions et émotions humaines. Il établit qu'il existe un nombre limité d'expressions reconnues par tous, indépendamment de la culture. Il met donc en évidence l'universalité de certaines émotions innées qui correspondent aux sept ou huit émotions : la neutralité, la joie, la tristesse, la surprise, la peur, la colère et le dégoût. Ce sont donc les travaux de Paul Ekman qui ont incité d'autres chercheurs à s'investir dans cette discipline qui ne cesse de prendre de l'importance dans différents domaines d'application [16, 17,18].

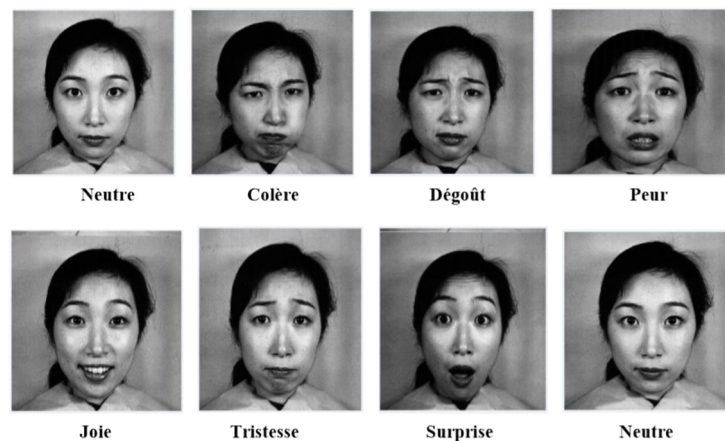


Figure 2.2 : *Les expressions faciales universelles tirées de la base de données JAFFEE*

2.3.1. Le Système de Codification des Actions Faciales (FACS)

Ekman développe un outil de codification des expressions du visage largement utilisé aujourd'hui. Il s'intéresse désormais à l'analyse des expressions de manière informatique.

En 1978, Ekman et Friesen présentent un système de codification manuelle des expressions du visage [19]. Leurs travaux d'observation leur permettent de décomposer tous les mouvements visibles du visage en termes de 46 unités d'Actions qui décrivent les mouvements élémentaires des muscles. N'importe quelle mimique observée peut donc être représentée sous la forme d'une combinaison d'unités d'Actions.

Ce système de codage est connu sous le nom de Facial Action Coding System (FACS). FACS s'est imposé depuis comme un outil puissant de description des mimiques du visage, utilisé par de nombreux psychologues. Bien que FACS soit un système de description bénéficiant d'une grande maturité (environ vingt années de développement), il souffre cependant de quelques inconvénients puisqu'il a été conçu au début pour des psychologues, certaines mesures restent floues, complexes et difficilement évaluable par une machine.

2.3.2. MPEG4

La norme de codage vidéo MPEG-4 dispose d'un modèle du visage humain développé par le groupe d'intérêt Face and Body AdHocGroup. C'est un modèle 3D articulé. Ce modèle est construit sur un ensemble d'attributs faciaux, appelés Facial Feature Points (FFP). Des mesures sur ces FFP sont effectuées pour former des unités de mesure (Facial Animation Parameter Units) qui servent à la description des mouvements musculaires (Facial Animation Parameters - équivalents des Actions Unitaires d'Ekman).

Les Facial Animation Parameter Units (FAPU) permettent de définir des mouvements élémentaires du visage ayant un aspect naturel. En effet, il est difficile de définir les mouvements élémentaires des muscles de manière absolue : le déplacement absolu des muscles d'une personne à l'autre change, mais leurs déplacements relatifs à certaines mesures pertinentes sont constants. C'est ce qui permet d'animer des visages de manière réaliste et peut permettre de donner des expressions humaines à des personnages non-humains. Comme exemples de FAPU, on peut citer la largeur de la bouche, la distance de séparation entre la bouche et le nez, la distance de séparation entre les yeux et le nez, etc. Par exemple, l'étirement du coin de la lèvre gauche (Facial Animation Parameter 6 stretch_l_cornerlip) est défini comme le déplacement vers la droite du coin de la lèvre gauche d'une distance égale à

la longueur de la bouche. Les FAPUs sont donc des mesures qui permettent de décrire des mouvements élémentaires et donc des animations.

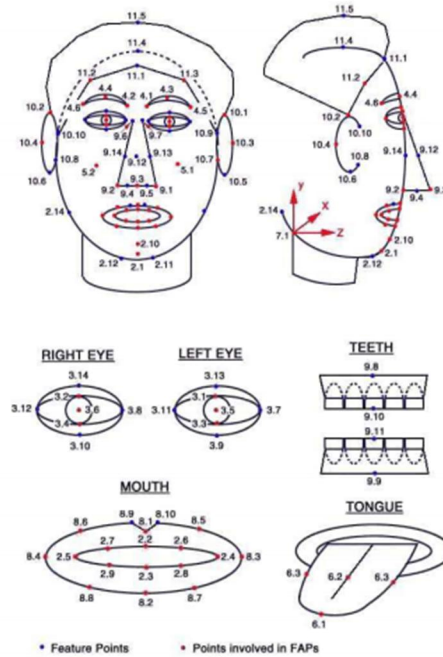


Figure 2.3 : *Modèle de visage mpeg-4 – Facial Animation*

La technologie de l'expression faciale est utilisée en automobile pour la détection de la somnolence du conducteur en temps réel. En effet, outre la fatigue, l'état émotionnel du conducteur constitue un facteur de risque. L'énervement en particulier peut rendre l'automobiliste plus téméraire ou moins attentif. En se basant sur l'analyse des expressions faciales, des chercheurs de l'EPFL, en collaboration avec PSA Peugeot Citroën, ont mis au point un détecteur d'émotions embarqué. Les tests réalisés dans un prototype ouvrent des pistes prometteuses. Aussi, plusieurs activités sportives utilisent des algorithmes d'expressions faciales et ont développé des techniques avancées d'analyses des données pour la détermination des capacités physiques des athlètes et la progression de leurs rendements [20].

2.3.3. Les méthodes de reconnaissance d'expression faciale

La littérature scientifique de la reconnaissance d'expression faciale pour les images statiques est plus ou moins comparable avec celle de la reconnaissance faciale. Les deux approches communes de l'extraction faciale sont les méthodes de caractérisation géométriques (locales) et d'apparence (globales) [21]. La représentation globale constitue des caractéristiques qui contiennent des informations relatives à l'image entière, tandis que la représentation locale caractérise des informations extraites des contours locaux [22], (e.g.,

Gabor wavelets utilisant les grid points) présente plus de performance que le global-based ones (e.g., based on eigenfaces, Fisher's discriminant analysis, etc.) [23].

Pour les images statiques, il existe le filtrage de Gabor [24,25]. Pour les séquences images des flux vidéo, il existe des méthodes de poursuite de caractéristique et différence d'images le modèle binaire local (MBL)[26], le modèle d'apparence active (MAA).

Plusieurs méthodes sont utilisées pour la reconnaissance de l'expression faciale telles que les réseaux de neurones (NN) [26, 27], l'entraînement de la machine à vecteurs de support (SVM) [28], l'analyse linéaire discriminante [29], l'analyse en composantes principales (ACP) [30], rule-based classifiers [31] Hidden Markov Models (HMM) [32]. D'autres méthodes utilisent un système de reconnaissance hybride [33, 34]. L'apprentissage profond pour d'autres, à savoir, A U-inspired Deep Networks (AUDN) [35], Deep Belief Networks [36] et le Convolutional Neural Network (CNN) [37] sont utilisés également en FER.

Les moments de Zernike (ZMs) sont aussi utilisés pour représenter le visage et l'expression faciale [38, 39]. Contrairement aux LBP's, les ZMs Locaux offrent une flexibilité en terme de taille de la description locale, avec moins de complexité de calcul. Similaires aux Moments de Zernike et Moments pseudo-Zernike ZMs/PZMs, la transformation radiale angulaire ART [40] et le descripteur de courbure (CSSD) [41] ont été adoptés par le MPEG-7 comme le descripteur basé région et le contour-based shape descriptor [42, 43].

La transformation radiale angulaire (ART) est utilisée dans plusieurs modèles de reconnaissance ; ses moments ont été appliqués pour la reconnaissance faciale et pour la description des modèles 2D et 3D des images [44].

Parmi le panorama des méthodes citées jusqu'à présent, nous allons présenter les plus significatives en reconnaissance de visages, à priori celles que nous allons utiliser dans notre travail, à savoir :

2.3.3.1. L'Analyse en Composante Principales (l'ACP)

Une Analyse en Composantes Principales (ACP) permet de définir, à partir d'un jeu de données d'apprentissage, un sous espace permettant de simultanément conserver l'information discriminante et supprimer les informations secondaires (non informatives). Cette méthode consiste à trouver une nouvelle base de l'espace des données dont tous les vecteurs sont orthogonaux entre eux. Le premier de ces vecteurs correspond à la direction de variance maximale des données d'apprentissage. Les autres composantes sont déterminées par

la contrainte d'orthogonalité entre les vecteurs tout en respectant une direction de variance maximum. Proposée en 1991 par Turk et Pentland [45], l'algorithme EigenFaces est une adaptation de l'ACP à la problématique de la reconnaissance des visages. L'espace de représentation des visages est construit en effectuant une ACP sur un ensemble de N images d'apprentissage.

On note $I = \{I_1, \dots, I_N\}$ la collection des N vecteurs de taille $M \times P$ représentant les N images d'apprentissage. L'image moyenne de cette base d'apprentissage est alors définie par :

$$\bar{I} = \frac{1}{N} \sum_{i=1}^N I_i \quad (2.1)$$

Les images I_i d'apprentissage sont alors normalisées en soustrayant l'image moyenne $\bar{I}$:

$$\Psi_i = I_i - \bar{I}, \forall_i \in 1..N \quad (2.2)$$

L'ACP permet de trouver un ensemble de K vecteurs orthogonaux U_i permettant de décrire, au mieux, l'ensemble des visages comme une combinaison linéaire de ces vecteurs :

$$\Psi_i = \sum_{j=1}^K \alpha_j U_j, \forall_i \in 1..N \quad (2.3)$$

La recherche de ces vecteurs orthogonaux revient à déterminer les vecteurs propres V_i de la matrice de covariance C_x des données d'apprentissage. Cette matrice est définie par :

$$C_x = \sum_{i=1}^N \Psi_i \Psi_i^t \quad (2.4)$$

En notant X, la matrice de taille $N \times (M \times P)$ contenant les N vecteurs d'apprentissage, la matrice de covariance peut être définie par :

$$C_x = XX^t \quad (2.5)$$

Les vecteurs propres de cette matrice de covariance C_x , correspondant aux composantes principales de visages et sont couramment appelés EigenFaces [46].



Figure 2.4 : Exemple d'EigenFaces

Une réduction de la dimension de l'espace représentatif des visages peut ensuite être obtenue en ne conservant qu'une sous-partie de ces vecteurs propres : Seuls les N' vecteurs

propres associés aux valeurs propres les plus élevées sont conservés pour constituer une base du nouvel espace des visages. Le choix de la valeur de N' peut être fait de manière à conserver un pourcentage donné de l'énergie globale ou bien de manière empirique en utilisant par exemple le critère de Kaiser (seules les vecteurs propres associés à une valeur propre supérieure à 1 sont conservés).

Le nouvel espace de représentation des visages E' est alors l'espace défini par les vecteurs v_i avec $i = 1..N'$. Chaque visage peut alors être caractérisé par sa projection dans cet espace. La comparaison entre deux visages V_1 et V_2 est alors faite en comparant les projections V_1' et V_2' de ses visages dans le nouvel espace E' . Le score de similarité entre ces deux visages est alors égal à la distance $d = \|V_1' - V_2'\|$.

Nous résumons les différentes étapes de l'eigenfaces comme suit :

L'apprentissage des visages propres s'effectue selon les étapes suivantes :

1. Collecte des N images faciales et construction de la matrice T de taille N , par concaténation des colonnes des images faciales. Prétraitement des images collectées.
2. Calcul du visage moyen en sommant les colonnes de la matrice T et en divisant le vecteur résultant par le nombre d'image d'entrée (N).
3. Soustraction du visage moyen de la matrice T pour obtenir la matrice A ; où chaque élément représente la variance des valeurs d'intensité de chaque pixel.
4. Calcul de la matrice C' , $C' = A^T \cdot A$.
5. Calcul des vecteurs propres de C' et les trier dans un ordre descendant selon les valeurs propres associées.
6. Calcul des vecteurs propres de la matrice de covariance C' et obtention des visages propres en multipliant les vecteurs propres de C' par la matrice A .
7. Choix des K meilleures valeurs propres et les vecteurs propres associés.
8. Détermination du poids des images d'entrée en projetant chaque image dans l'espace visage.
9. Chaque visage est maintenant représenté par un vecteur qui est utilisé pour reconstruire les images. Après on sauvegarde le visage moyen, les eigenfaces et la matrices de projection (de poids) des images. Les neuf étapes décrites transformeront une base de données d'images faciales en un ensemble de projections dans l'espace visage (face space).

10. L'étape de reconnaissance peut être résumée comme suit :

- Prétraitement de l'image d'entrée et soustraction du visage moyen.
- Détermination du poids de l'image d'entrée par la projection de celle-ci dans l'espace visage en multipliant le vecteur résultant de l'étape (1) par les eigenfaces de la base de données.
- Comparaison des résultats obtenus en utilisant des métriques telles que la distance Euclidienne. Nous allons décrire les différentes distances existantes dans la section suivante.

2.3.3.2. Les mesures de distance

Lorsqu'on souhaite comparer deux vecteurs de caractéristiques issus du module d'extraction de caractéristiques d'un système biométrique, on peut soit effectuer une mesure de similarité (ressemblance), soit une mesure de distance (divergence). La première catégorie de distances est constituée de distances Euclidiennes et sont définies à partir de la distance de Minkowski d'ordre p dans un espace euclidien (déterminant la dimension de l'espace euclidien).

Considérons deux vecteurs $X = (x_1, x_2, \dots, x_n)$ et $Y = (y_1, y_2, \dots, y_n)$, la distance de Minkowski d'ordre p notée L_p est définie par :

$$L_p = \sum_{i=1}^n |x_i - y_i|^p \quad (2.6)$$

Nous allons présenter quelques mesures de distance dans l'espace original des images puis dans l'espace de Mahalanobis.

1. Distances minkowski

A. Distance City Block (L1)

Pour $p = 1$, on obtient la distance City-Block (ou distance de Manhattan) :

$$L_1(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (2.7)$$

B. Distance Euclidienne (L2)

Pour $p = 2$, on obtient la distance euclidienne :

$$L_2(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \quad (2.8)$$

Les objets peuvent alors apparaître de façons très différentes selon la mesure de distance choisie (Figure 2.5).

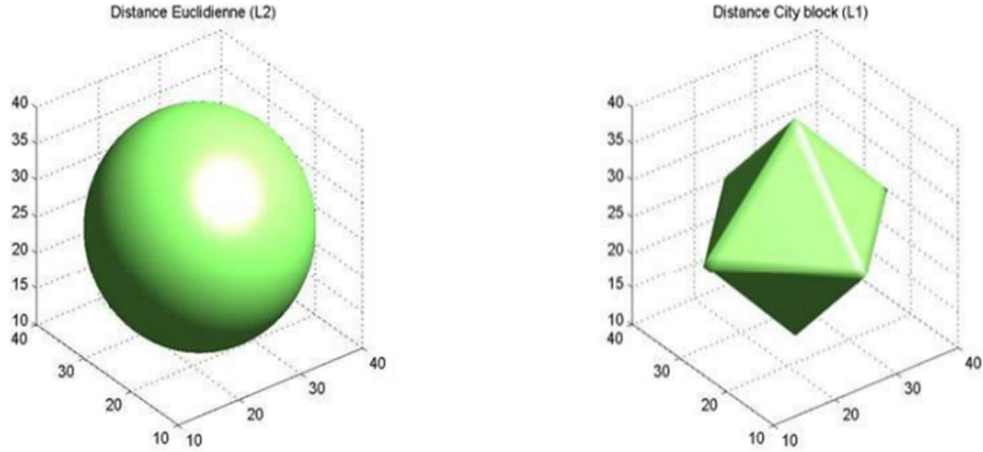


Figure 2.5 : Représentation d'une sphère avec la distance euclidienne et la distance City-Block

2. Distances dans l'Espace de Mahalanobis

A. De l'espace des images à l'espace de Mahalanobis

Avant de pouvoir effectuer des mesures de distance dans l'espace de Mahalanobis, il est essentiel de bien comprendre comment l'on passe de l'espace des images Im à l'espace de Mahalanobis $\mathcal{EMah}$.

En sortie de l'algorithme de l'ACP, nous obtenons des vecteurs propres associés à des valeurs propres (représentant la variance selon chaque dimension). Ces vecteurs propres définissent une rotation vers un espace dont la covariance entre les différentes dimensions est nulle. L'espace de Mahalanobis est un espace où la variance selon chaque dimension est égale à 1. On l'obtient à partir de l'espace des images Im en divisant chaque vecteur propre par son écart-type correspondant. Soit u et v deux vecteurs propres de Im , issus de l'algorithme PCA, et m et n deux vecteurs de $\mathcal{EMah}$. Soit λ_i les valeurs propres associées aux vecteurs u_i et v_i et σ_i l'écart-type, alors on définit λ_i . Les vecteurs u_i et v_i sont reliés aux vecteurs m_i et n_i à partir des relations suivantes :

$$m_i = \frac{u_i}{\sigma_i} = \frac{u_i}{\sqrt{\lambda_i}}$$

$$\text{et} \quad n_i = \frac{v_i}{\sigma_i} = \frac{v_i}{\sqrt{\lambda_i}} \quad (2.9)$$

B. Mahalanobis L1 (MahL1)

Cette distance est exactement la même que la distance City-Block sauf que les vecteurs sont projetés dans l'espace de Mahalanobis. Ainsi, pour des vecteurs propres u et v de projections respectives m et n sur l'espace de Mahalanobis, la distance Mahalanobis $L1$ est définie par :

$$Mah_{L1}(u, v) = \sum_{i=1}^N |m_i - n_i| \quad (2.10)$$

C. Mahalanobis L2 (MahL2)

Cette distance est identique à la distance euclidienne à part qu'elle est calculée dans l'espace de Mahalanobis. Ainsi, pour des vecteurs propres u et v de projections respectives m et n sur l'espace de Mahalanobis, la distance Mahalanobis $L2$ est définie par :

$$Mah_{L2}(u, v) = \sqrt{\sum_{i=1}^N |m_i - n_i|^2} \quad (2.11)$$

Par défaut, lorsqu'on parle de distance de Mahalanobis, c'est à cette distance que l'on doit se référer.

2.3.3.3. Machine à Vecteur de Support

C'est une technique qui a été proposée par V. Vapnik en 1995, elle est utilisée dans plusieurs domaines statistiques (classement, régression, fusion,... etc). Depuis son introduction dans le domaine de reconnaissance de formes, plusieurs travaux ont montré l'efficacité de cette technique, principalement en traitement d'images. L'algorithme s'avère très performant en matière de séparabilité des classes, une chose qui nous a motivés pour l'utiliser durant notre travail.

L'idée essentielle de cette approche consiste à projeter les données de l'espace d'entrée (appartenant à des classes différentes) non linéairement séparables, dans un espace de plus grande dimension appelé espace de caractéristiques, de façon à ce que les données deviennent linéairement séparables [47,48]. Dans cet espace, la technique de construction de l'hyperplan optimal est utilisée pour calculer la fonction de classement séparant les classes telle que :

- Les vecteurs appartenant aux différentes classes se trouvent de différents côtés de l'hyperplan.
- La plus petite distance entre les vecteurs et l'hyperplan (la marge) soit maximale.

Principe de l'algorithme

L'algorithme se base principalement sur 3 astuces pour obtenir de très bonnes performances tant en qualité de prédiction qu'en complexité de calcul.

- On cherche l'hyperplan comme solution d'un problème d'optimisation sous-contrainte. La fonction à optimiser intègre un terme de qualité de prédiction et un terme de complexité du modèle.
- Le passage à la recherche de surfaces séparatrices non linéaires est introduit en utilisant un noyau (kernel) qui code une transformation non linéaire des données.
- Numériquement, toutes les équations s'obtiennent en fonction de certains produits scalaires utilisant le noyau et certains points de la base de données (ce sont les Support Vectors).

Dans tout type de problème de classification à deux classes :

- Classification d'images : reconnaissance de visages, de chiffres manuscrits
- Interprétation textuelle : détection de Spam
- Aide au diagnostic biologique

De nouveaux centres d'intérêts se développent actuellement :

- Utilisation comme outil de sélection de variables
- Sparse SVM : utilisation de modèles parcimonieux pour la classification.

Contraintes :

- A priori, capable d'utiliser un grand nombre de variables puisque l'astuce du noyau « envoie » le problème dans l'espace des individus.
- Plus il y a d'individus dans la base de données et meilleure est la prédiction, mais plus les calculs sont longs...etc. [49]
- Les SVM peuvent être utilisés pour résoudre des problèmes de discrimination, ou de régression, c'est-à-dire prédire la valeur numérique d'une variable.

Un problème de discrimination qui se limite à deux classes est une **discrimination binaire**, c'est-à-dire $y \in \{-1, 1\}$, le vecteur d'entrée x étant dans un espace X muni d'un produit scalaire. Par exemple :

$$X = \mathbb{R}^N.$$

Discrimination linéaire et hyperplan séparateur

Le cas simple est le cas d'une fonction discriminante linéaire, obtenue par combinaison linéaire du vecteur d'entrée $x = (x_1, \dots, x_N)^T$ avec un vecteur de poids $w = (W_1, \dots, W_N)^T$

$$h(x) = w^T \cdot x + w_0 \quad 2.12$$

Il est alors décidé que x est de classe 1 si $h(x) \geq 0$ et de classe -1 sinon. C'est un **classifieur linéaire**. La frontière de décision $h(x) = 0$ est l'**hyper plan séparateur**. Le but d'un algorithme d'apprentissage supervisé est d'apprendre la fonction $h(x)$ par le biais d'un ensemble d'apprentissage :

$$\{(x_1, I_1), (x_2, I_2), \dots, (x_k, I_k), \dots (x_p, I_p)\} \in R^N$$

Où les I_k sont les labels, p est la taille de l'ensemble d'apprentissage, N la dimension des vecteurs d'entrée. Si le problème est linéairement séparable, on doit alors avoir :

$$I_k h(x_k) \geq 0 \quad 1 \leq k \leq p, \text{ autrement dit } I_k (w^T X_K + W_0) \geq 0$$

Choix d'une séparation

Choix de la séparation : hyperplan qui classe correctement les données (lorsque c'est possible) et qui se trouve "le plus loin possible de tous les exemples".

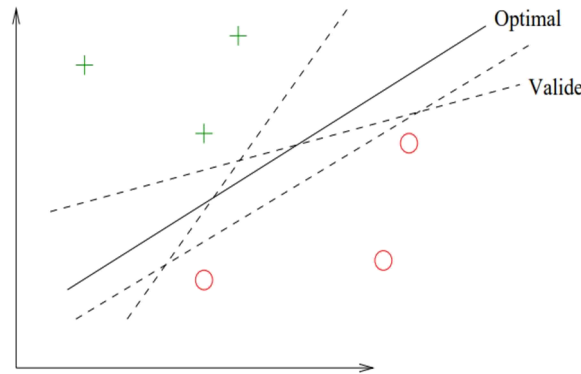


Figure 2.6 : Représentation de l'hyper plan optimal [50]

Maximisation de la marge

Marge = distance du point le plus proche à l'hyperplan.

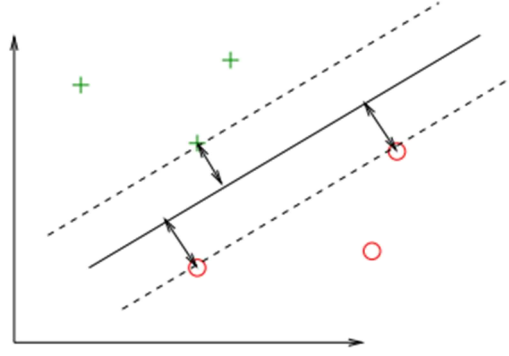


Figure 2.7 : Représentation géométrique de la maximisation de marge (SVM) [50]

La marge est la distance entre l'hyperplan et les échantillons les plus proches. Ces derniers sont appelés vecteurs supports. L'hyperplan qui maximise la marge est donné par :

$$\operatorname{argmax} \min\{\|x - x_k\| : x \in R^N, w^T x + w_0\} \quad 2.13$$

Il s'agit donc de trouver w et w_0 remplissant ces conditions, afin de déterminer l'équation de l'hyperplan séparateur :

$$h(x) = w^T x + w_0 = 0$$

Formulation

La marge est la plus petite distance entre les échantillons d'apprentissage et l'hyperplan séparateur qui satisfait la condition de séparabilité, à savoir : $h(x) = w^T x + w_0 = 0$

La distance d'un échantillon x_k à l'hyperplan est donnée par sa projection orthogonale sur le vecteur de poids :

$$\frac{l_k(w^T x_k + w_0)}{\|w\|}$$

L'hyperplan séparateur (w, w_0) de marge maximale est donc donné par :

$$\operatorname{argmax}_{w, w_0} \left\{ \frac{1}{\|w\|} \min_k [l_k(w^T x_k + w_0)] \right\}$$

Afin de faciliter l'optimisation, on choisit de normaliser w et w_0 , de telle manière que les échantillons à la marge x_{marge}^+ pour les vecteurs supports sur la frontière positive, et x_{marge}^- pour ceux situés sur la frontière opposée) satisfassent :

$$\begin{cases} w^T x_{marge}^+ + w_0 = 1 \\ w^T x_{marge}^- + w_0 = -1 \end{cases}$$

D'où pour tous les échantillons, $k=1, \dots, p$

$$l_k(w^T x_k + w_0) \geq 1$$

Cette normalisation est parfois appelée la forme canonique de l'hyper plan, ou hyperplan canonique.

Avec cette mise à échelle, la marge vaut désormais $\frac{1}{\|w\|}$, il s'agit donc de maximiser $\|w\|^{-1}$.

La formulation dite primaire a la forme suivante :

Minimiser $\frac{1}{2} \|w\|^2$ sous les contraintes $l_k(w^T x_k + w_0) \geq 1$

$$L(w, w_0, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{k=1}^p \alpha_k \{l_k(w^T x_k + w_0) - 1\} \quad (2.14)$$

En annulant les dérivées partielles du lagrangien, selon les conditions de Kuhn-Tucker, on obtient :

$$\begin{aligned} \sum_{k=1}^p \alpha_k l_k x_k &= w^* \\ \sum_{k=1}^p \alpha_k l_k &= 0 \end{aligned} \quad (2.15)$$

D'où pour tous les échantillons, $k=1, \dots, p$

$$\widetilde{L}(\alpha) = \sum_{k=1}^p \alpha_k - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j l_i l_j x_i^T x_j \quad (2.16)$$

En réinjectant ces valeurs dans l'équation (2.16), on obtient la *formulation duale* :

Maximiser $\widetilde{L}(\alpha) = \sum_{k=1}^p \alpha_k - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j l_i l_j x_i^T x_j$ sous les contraintes $\alpha_k \geq 0$, et

$$\sum_{k=1}^p \alpha_k l_k = 0 \quad (2.17)$$

Ce qui donne les multiplicateurs de Lagrange optimaux α_w^k .

Afin d'obtenir l'hyperplan solution, on remplace w par sa valeur optimale w^* , dans l'équation de l'hyperplan $h(x)$, ce qui donne :

$$h(x) = \sum_{k=1}^p \alpha_k^* l_k(x \cdot x_k) + w_0 \quad (2.18)$$

L'exemple le plus simple de fonction noyau est le noyau linéaire :

$$K(x_i, x_j) = x_i^T \cdot x_j \quad (2.19)$$

Des noyaux usuels employés avec les SVM sont :

$$K(x_i, x_j) = (x_i^T \cdot x_j + 1)^d \quad (2.20)$$

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

2.3.3.4. Description des formes

En deux dimensions, on peut définir deux notions de similarité entre formes, la similarité de région et la similarité de contour. Ceci est montré dans la figure 2.8. Les objets de la première ligne ont des similarités spatiales dans la distribution des pixels et sont similaires selon des critères régions. Alors qu'ils sont clairement différents selon leur contour. En considérant la similarité de contour, les objets de chaque colonne sont similaires. En prenant pour requête l'objet situé dans la première ligne et la deuxième colonne, la similitude région fournira les objets de la première ligne, alors que la similarité contour fournira les résultats de la deuxième colonne.

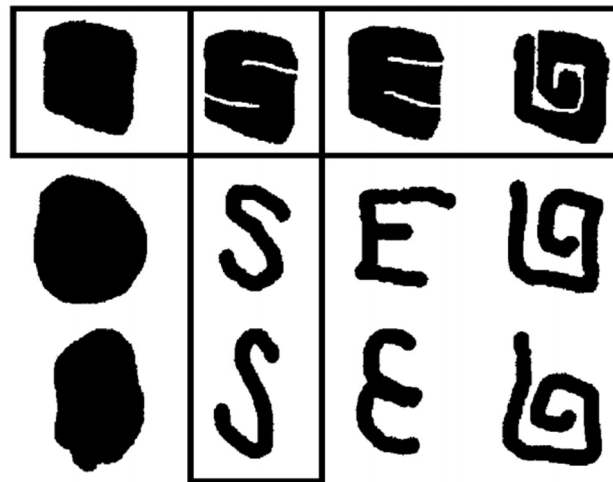


Figure 2.8 : Exemple de similarités contour et région

Des techniques de représentation et de description de formes basées contours et les méthodes basées régions peuvent être divisées en deux sous parties : les approches structurelles et les approches globales (tableau 2.1).

Ces sous-classes sont caractérisées si les descripteurs de forme sont calculés sur l'ensemble de la région ou du contour, ou sur un ensemble de primitives.

Parallèlement à cela, les méthodes globales permettent de calculer directement les descripteurs de forme sans une phase de sous-segmentation et avec une phase de comparaison numérique, ce qui permet de garantir un cout calculatoire d'indexation et de recherche faible. Les méthodes basées sur les régions traitent la forme dans son ensemble et utilisent l'information de tous les pixels de la région. Ces méthodes mesurent la distribution des pixels de la région et sont peu affectées par le bruit.

Pour ces différentes raisons, nous nous sommes principalement intéressés aux méthodes d'analyses de formes basées régions dans notre travail, et plus précisément à la transformation ART, Angular Radial Transform.

APPROCHES GLOBALES		APPROCHES REGIONS	
Structurelles	Globales	Structurelles	Globales
-Chain code (<i>Visa, 1996</i>) -Polygone (<i>Mehrorota et Gary, 1995</i>) -B-Splime (<i>Berretti et al.2000</i>) -Invariants (<i>Sonka et al.1999</i>)	- Excentricité Shape Signature (<i>Van Otterloo.1991</i>) - Distance de Haudroff -Descripteurs de Fourier (<i>Zhang et Lu, 2001</i>) -Descripteurs d'ondelettes (<i>Tieng et Boles, 1997</i>) -Curvature Scale Space (<i>Mokhtarian et al. 1996</i>) -Autoregressivité (<i>Sekita et al. 1992</i>) -Elastic Matching (<i>Del Bimbo et Pala, 1997</i>).	-Coque convexe (<i>Gonzalez et Woods ; 2001</i>) -Axes médians (<i>Irani et Ware, 2003</i>) -Géons (<i>Biederman , 1987</i>)	-Aire, Nombre d'Euler Moment géométriques (<i>Hu, 1962</i>). -Moments de Zernike (<i>Khotanzad et Hong, 1990</i>). -Pseudo Moments de Zernike (<i>Chong et al, 2003</i>). -Moments de Legendre (<i>Teague, 1980</i>). -Descripteurs de Fourier générique (<i>Zhang et lu, 2002</i>). -ART (<i>Kim et Kim , 1999</i>).

Tableau 2.1 : Classification des techniques de descriptions de formes 2D

On peut noter que la plupart des méthodes sont basées sur des approches globales. Les approches structurelles basées sur le contour ou basées sur la région, décomposent la forme en éléments de base et comparent les ensembles de primitives, ou les graphes les représentant, pour mesurer la similarité entre formes. En effet, de nombreuses méthodes structurelles décrivent la forme comme un graphe reliant les éléments de base et comparent deux objets en faisant de la comparaison de graphe.

Deux problèmes majeurs ont freinés ces méthodes : la segmentation et la comparaison. Il faut pour représenter convenablement une forme par un ensemble de primitives, avoir une phase de segmentation, qui extrait de la forme une représentation structurale. Il n'y a pas une définition formelle de la forme et l'on ne peut pas connaître le nombre de primitives requises pour correctement la décrire, ce qui entraîne un calcul complexe. D'autre part, le calcul de la mesure de similarité entre formes est non métrique, contrairement aux méthodes sur la forme globale, ce qui permet de faire de la comparaison de sous graphes, mais exige un coût de calcul élevé.

La Transformation Angulaire Radiale

Sur les images 2D, le comité MPEG-7 a proposé d'indexer les images en utilisant un descripteur de forme par approche région, comme la transformation radiale angulaire ou Angular Radial Transform (ART) [51, 52]. Ce descripteur de formes possède de nombreuses propriétés : petite taille, robuste aux bruits et aux variations d'échelle, invariance en rotation, possibilité de décrire des objets complexes.

Comme le descripteur basé-région fait usage de tous les pixels constituant la forme, il peut décrire efficacement diverses formes dans un seul descripteur, une chose qui nous a motivé pour l'utiliser pour l'extraction des caractéristiques des visages.

En effet, La transformation radiale angulaire (ART) est utilisée dans plusieurs modèles de reconnaissance, telle que la reconnaissance de logo [53] et les modèles multi vues 3D [54]. ART a été utilisé pour la première fois avec SVM pour la détection des visages [55] et depuis ses moments ont été appliqués et généralisés pour la description des modèles 2D et 3D [56] et pour la reconnaissance du visage et d'expression faciale, ayant également pour application la vidéo surveillance [57,58].

Les coefficients F_{nm} se calculent de la manière suivante :

$$F_{nm} = \iint_0^{2\pi} V_{nm}(\rho, \theta) f(\rho, \theta) d\rho d\theta \quad (2.15)$$

Où $f(\rho, \theta)$ est l'image projetée en coordonnées polaires, et V_{nm} est défini par :

$$V_{nm}(\rho, \theta) = \frac{1}{2\pi} \exp(jm\theta) R_n(\rho)$$
$$R_n(\rho) = \begin{cases} 1 & n = 0 \\ 2\cos(n\pi\rho) & n \neq 0 \end{cases} \quad (2.16)$$

Nous avons pour $n = \{0, 1, \dots, n\}$ et $m = \{0, 1, \dots, m\}$. Le standard MPEG-7 préconise l'utilisation de $n = 2$ et $m = 11$, ce qui donne un vecteur de dimension 36. Le vecteur descripteur se déduit des coefficients F_{nm} avec de simples opérations. L'invariance à la rotation, s'obtient en prenant le module des coefficients. L'invariance à l'échelle s'obtient en divisant les coefficients par F_{00} . La Figure 2.9 représente la forme binaire de la représentation ART.

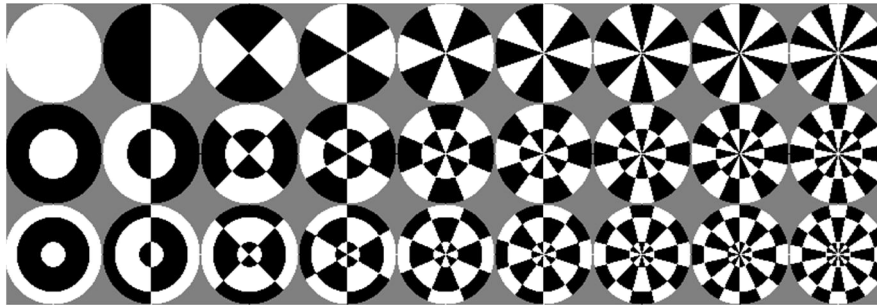


Figure 2.9 : La forme binaire de base de l'ART [52]

2.4. Le système de reconnaissance d'expression faciale

Un système de reconnaissance d'expressions faciales consiste en trois phases principales : détection de visage à étudier, extraction de composantes et enfin classification et reconnaissance de l'expression faciale.

2.4.1. La détection du visage

Elle s'applique sur une image dans laquelle on souhaite déceler la présence et la localisation d'un objet. Parmi les différentes techniques de détection de visage proposées ces dernières années figurent celles basées sur l'apparence. Les méthodes de cette catégorie reviennent à traiter le problème de la détection de visages comme un problème de classification.

Dans le but de déterminer si une image appartient à la classe des visages ou à celle des non visages, on utilise des techniques d'apprentissage automatique supervisé. Pour cela, un ensemble de données d'apprentissage est constitué d'une part des images représentant des visages et, d'autre part des images ne représentant pas des visages, permettant de fixer une règle de séparation entre les deux classes. Dans cette catégorie, la méthode développée par H.

Rowley, S. Baluja et T. Kanade est basée sur des réseaux de neurones. Ces techniques des plus efficaces et des plus adéquates pour une détection en temps-réel.

Détection des régions d'intérêt : Algorithme de Viola et Jones

La méthode de Viola et Jones est basée sur un apprentissage supervisé. Des exemples d'objets doivent être analysés à l'avance, pour être classifiés. Pendant l'analyse de l'image, des caractéristiques sont choisies par boosting, qui permet de classifier les caractéristiques, et de séparer les exemples positifs des exemples négatifs par cascade de décision.

Les étapes de cet algorithme sont :

1. Apprentissage : Il consiste à analyser un très grand nombre d'images positives et négatives et comprend :
 - Le calcul des caractéristiques pseudos Haar (positifs et négatifs).
 - L'entraînement de la cascade : pour chaque étage, on crée un classifieur fort, par ajout successif de classifieurs faibles entraînés sur la même caractéristique. On ajoute ces classifieurs faibles jusqu'à obtenir un classifieur fort possédant la performance souhaitée (taux de détection/faux positifs) pour l'étage.
2. Procédure de détection : C'est l'étape qui s'applique sur l'image dans laquelle on souhaite déceler la présence d'objet. On parcourt l'ensemble de l'image à toutes les positions et échelles, avec une sous fenêtre. On applique la cascade à chaque sous fenêtre :

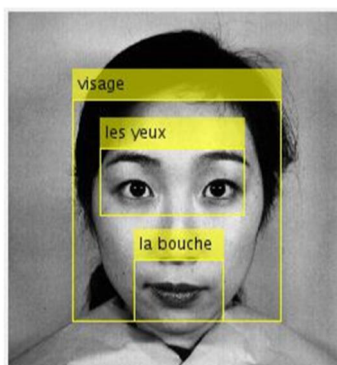


Figure 2.10 : Détection des régions d'intérêt

La méthode des cascades de classifieurs de Viola et Jones est basée sur une approche par recherche exhaustive sur l'ensemble de l'image qui teste la présence de l'objet dans une fenêtre à toutes les positions et à plusieurs échelles. Cette approche nécessite de nombreux calculs mais cette contrainte trouve une solution dans l'organisation de l'algorithme de détection en une cascade de classifieurs. Ces derniers appliqués séquentiellement prennent

une des décisions d'acceptation ; L'idée est que l'immense majorité des fenêtres testées négatives (c.-à-d. ne contiennent pas l'objet) seront rejetés avec le moins possible de calculs. Pour cela, les classifieurs les plus simples, donc les plus rapides, sont situés en début de cascade, et rejettent très rapidement la grande majorité des exemples négatifs.

Le Détecteur de Visages de Rowley

L'implémentation de ce détecteur diffère sur certains points des algorithmes publiés par l'auteur dans [59]. Dans ce détecteur, la classification est basée sur une série de perceptrons multicouches. Un perceptron multicouches est un réseau de neurones artificiels, organisé en couches.

Un neurone est un automate, dont la sortie est la comparaison d'un barycentre pondéré de plusieurs entrées à un seuil, via une fonction d'activation. Au sein d'une même couche, les entrées de chaque perceptron ne sont connectées qu'aux sorties de la couche précédente.

La dernière couche n'est constituée que d'un unique neurone, dont la polarité de la sortie définit l'appartenance ou non à la classe étudiée. Ce positionnement en couches (ou en bancs) des neurones permet de définir une séparation non linéaire entre classes. De manière plus précise que ce qui est publié dans [59], la version disponible en open source du détecteur de Rowley utilise 3 perceptrons multicouches successifs pour déterminer la présence ou l'absence d'un visage dans une fenêtre. En l'occurrence, la fonction d'activation utilisée est la fonction tangente hyperbolique (sigmoïde).

La 1ère couche de chaque perceptron est reliée aux descripteurs (l'intensité des pixels) sous forme de « rétines ». C'est-à-dire que chaque neurone de la 1ère couche est une combinaison linéaire de l'intensité de tous les pixels contenus dans une zone rectangulaire (champ réceptif) prédéfinie de la fenêtre étudiée. Ces champs réceptifs sont dédoublés, voire triplés, au sein d'un même perceptron. A savoir que chaque champ réceptif est représenté plusieurs fois (2 à 3 fois selon le perceptron) au niveau de la première couche du réseau de neurones. Etant donné que ces rétines sont munies d'un apprentissage différent, elles augmentent en principe la robustesse du détecteur.

Avant tout passage dans un perceptron, chaque fenêtre est sujette à une égalisation d'histogrammes. C'est une technique qui permet en principe d'être moins sujet aux problèmes d'ombrage et de faibles contrastes.

Un exemple de résultats de détection par l'algorithme de Rowley est disponible sur la Figure 2.11. Les détections positives selon Rowley sont encadrées en vert.



Figure 2.11 : *Quelques détections par l'algorithme de Rowley*

2.4.2. Extraction des composantes du visage

Quel que soit la reconnaissance effectuée (reconnaissance de visage ou d'expression faciale), il existe généralement deux approches pour traiter le problème.

L'approche globale : appelée aussi « image-based » est une méthode relativement récente qui analyse le visage dans sa globalité, elle se base principalement sur la mise en correspondance de modèles de visage.

L'approche par composantes : aussi appelée « feature-based » se base sur les différentes composantes du visage (yeux, nez, bouche . . .) indépendamment les uns des autres ou combinés.

Un système consiste à évaluer la possibilité de reconnaissance des six émotions universelles en analysant les déformations des caractéristiques faciales telles que les yeux, les sourcils, et la bouche. On considère que ces caractéristiques contiennent suffisamment les informations nécessaires.

Parmi les méthodes existantes figurent celles basées sur la segmentation, par masque pour le visage, par le modèle d'apprentissage et autres méthodes.

Méthode basée sur la segmentation, ce système est formé de quatre phases essentielles : la segmentation, extraction de données, analyse et classification.

1. Segmentation : le système va localiser les régions des yeux, des sourcils, et de la bouche ; un algorithme est utilisé pour avoir le squelette d'émotion.
2. Extraction de données : les squelettes sont extraits. En réalité, ce sont les contours des yeux, de la bouche, et des sourcils.
3. Analyse : A partir des squelettes, on choisit des distances pour caractériser les déformations.
4. Classification : À partir de ces distances caractéristiques, on utilise le modèle de croyance transférable pour la reconnaissance d'expressions faciale.

2.5. Les bases de données

Les recherches menées récemment ont débouchées sur une nouvelle base de données capable de combiner deux types d'informations : l'information bidimensionnelle ou de texture et l'information tridimensionnelle. Son usage reste purement scientifique [60].

La création d'une base de données destinée aux systèmes de reconnaissance faciale est une tâche complexe. Le grand nombre de facteurs qui permettent l'acquisition de données, exige l'application stricte d'un protocole afin que les images d'individus obtenues puissent être comparables entre elles. Ce qui explique que les bases d'images d'expressions faciales sont relativement peu nombreuses et présentent peu de diversités et de précision.

Les bases les plus complètes d'entre elles sont présentées dans la table 2.2 ci-après :

LA BASE DE DONNEE	NOMBRE TOTAL D'IMAGES
Radboud faces Database	67
Mug face database	575
MMI database	240
Pictures of Facial Affect (POFA)	110
JAFFE database	130
The Combined MIT/CMU Test Set with Ground for Frontal Face Detection	180

CBCL Face Database	472
Cohn-Kanade AU-Coded Facial Expression Database	456

Tableau 2.2 : *Les bases d'images d'expression faciales*

Dans notre cas, nous avons utilisé et comparé les résultats de deux bases de données, à savoir les bases JAFEE et Cohn-Kanade que nous détaillerons dans le chapitre 4.

2.6. Approche proposée

La base de tout système d'analyse et de reconnaissance d'expressions faciales est l'extraction des informations pertinentes qui peuvent décrire au mieux les expressions.

Nous avons opté pour des méthodes d'extraction de caractéristiques globales, à savoir, l'analyse en composantes principales en premier lieu et le descripteur basé région ART.

L'objectif principal derrière ce travail était d'abord, de proposer une nouvelle approche de fusion des deux méthodes sélectionnées, tout en montrant l'amélioration apportée par l'identification multi algorithmes. Parmi les divers niveaux de fusion existants, nous avons opté pour la fusion des caractéristiques, car les vecteurs caractéristiques issus de différents algorithmes et méthodes sont riches en information. De plus, la fusion à ce niveau peut s'appliquer à tous les systèmes.

Dans le but d'améliorer les performances de fusion, nous avons étudié les propriétés de la transformation angulaire radiale dans la description de forme. ART s'avère robuste aux variations de rotation et d'échelle. Il est généralement implémenté dans une forme cartésienne pour l'extraction des caractéristiques de l'image.

Nous avons ensuite proposé une forme différente d'ART que sa forme cartésienne, un développement mathématique polaire du descripteur ART est présenté, que nous avons baptisé (P-ART), que nous expliquons en détail dans le prochain chapitre.

2.7. Conclusion

Nous avons présenté dans ce chapitre un panorama succinct sur les méthodes de reconnaissance d'expression faciale, ainsi que les travaux relatifs à ces approches. Ceci nous

permet de déceler les limites (et/ou) les avantages de chaque méthode, quant à la robustesse et performance du système de reconnaissance.

La problématique alors soulevée, nous proposons, une nouvelle forme et une nouvelle fusion de deux méthodes pour améliorer les taux de reconnaissance d'expression faciale. Notre approche proposée est détaillée dans le chapitre suivant.

CHAPITRE III :

METHODES PROPOSEES

1.1. Introduction

Dans ce troisième chapitre, nous décrivons en détail notre approche pour le système de reconnaissance d'expression faciale, la méthode proposée consiste à fusionner les caractéristiques d'une méthode globale d'apparence et d'un descripteur basé région pour reconnaître l'expression faciale d'un visage donné.

Ce chapitre est divisé en deux parties : dans la première partie, nous allons présenter les différentes étapes du système, à savoir la phase du prétraitement des visages et notre technique d'extraction de caractéristiques, basée sur la transformation angulaire radiale(ART) et la réduction de dimension des données basée sur l'Analyse en Composantes Principales (ACP).

ART est généralement présenté dans la littérature scientifique en coordonnées cartésiennes. Dans la deuxième partie de ce chapitre, nous présentons notre deuxième contribution qui est une nouvelle forme d'ART. Nous présentons le développement mathématique polaire d'ART, qu'on a baptisé (P-ART) et sa fusion avec l'ACP pour la reconnaissance d'expression faciale.

1.2. L'algorithme proposé

Les images brutes permettent rarement de parvenir à une extraction directe des objets à analyser, elles sont souvent bruitées et présentent des variations d'éclairement, de contraste et de taille. L'objectif donc de la phase du prétraitement est d'obtenir une image avec des intensités normalisées, de tailles et de formes uniformes. La procédure de prétraitement vise aussi à éliminer les effets de luminance et éclaircissement et à redimensionner les images.

Les images traitées peuvent être représentées dans n'importe quelle taille, tout dépend de la résolution choisie durant la phase d'acquisition. Le problème se pose surtout dans le cas où nous traitons des images à hautes résolutions, par exemple, avec une image de résolution de un méga octet, il faut parcourir un million de pixel pour en extraire les caractéristiques ! Ce qui est trop lent même pour une machine puissante. Dans le module de prétraitement, les images sont souvent redimensionnées de [256x256] à [72x72], voir figure 3.1 (avant et après simple redimensionnement).

Dans notre cas, les images des bases de données JAFEE et CK, ont été recadrées afin de garder la région du visage uniquement (l'algorithme de prétraitement est expliqué ci-dessous (prétraitement), voir figure 3.4.

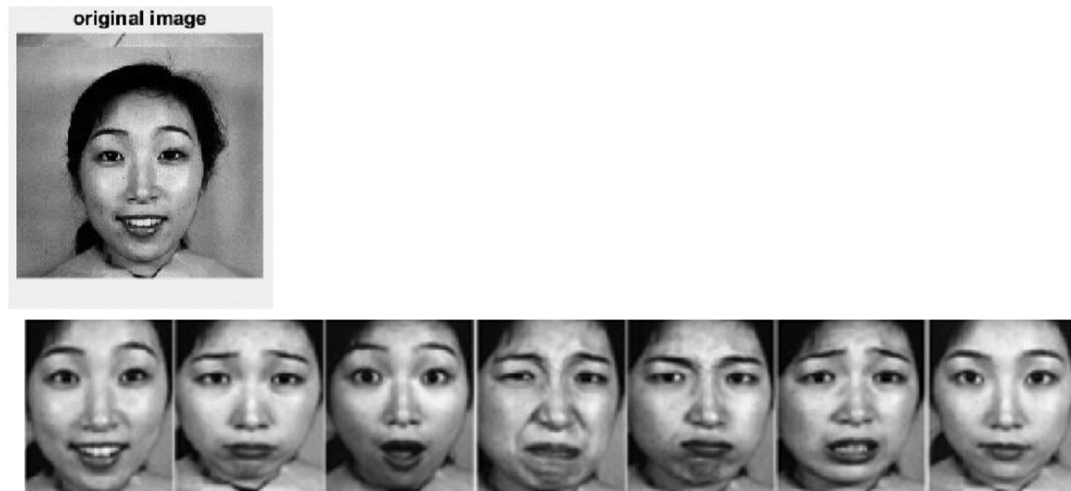


Figure 3.1 : *Exemple de prétraitement des visages de la base JAFFE avant (original image) et après simple rognage*

Le bloc de prétraitement est suivi par un bloc d'extraction de caractéristiques du visage, nous avons présenté dans l'état de l'art (chapitre 2) un panorama des méthodes existantes dans la littérature. Certains proposent une localisation globale (englobant les traits), d'autres essaient d'extraire plus précisément les contours, sauf que les modèles choisis sont très simples et les algorithmes nécessitent une phase de sélection manuelle des points.

Nous avons concentré notre travail sur les variations d'expression faciale, de ce fait, nous avons remarqué que ce type de variations induit dans certains cas à une variation de pose du visage. La plupart des méthodes citées dans la littérature ne sont pas forcément robustes aux variations de pose et rotation du visage (le roll). Cela nous a inspirés pour faire appel à un algorithme qui présente de bonnes performances en matière de rotation, d'échelle et robuste au bruit. Notre choix a porté sur le descripteur ART qui présente dans sa configuration radiale, une robustesse à la rotation (dans notre cas, le roll).

De plus, la plupart des approches et méthodes font recours à une phase d'apprentissage, jugée souvent lourde et volumineuse en termes d'implémentation et de calcul. A la différence de ART qui lui en tant que descripteur ne fait pas d'apprentissage sur les images introduites.

ART est généralement implémenté sous une forme cartésienne pour extraire les caractéristiques de l'image. Cette approche présente certains inconvénients. Nous proposons une nouvelle implémentation d'ART en utilisant ce que nous appelons une version polaire, nous étendons la transformation radiale angulaire (cartésienne) à Polar-ART.

Le nouvel ART est présenté différemment du descripteur cartésien conventionnel. Nous proposons un système de reconnaissance d'expression faciale, qui combine certaines caractéristiques de P-ART avec l'analyse en composantes principales (ACP). Ces caractéristiques sont ensuite apprises par un SVM pour déterminer l'expression faciale.

L'algorithme proposé pour la reconnaissance de l'expression faciale, est composé des étapes suivantes :

1. La phase du prétraitement (redimensionnement, binarisation, conversion en niveau de gris, filtrage, etc.),
2. La phase de l'extraction des caractéristiques locales en utilisant indépendamment : ART, ACP, ACP+ART,
3. La classification,
4. L'élaboration de la décision finale.

La figure 3.2 ci-dessous décrit les étapes du processus de la reconnaissance d'expression faciale :



Figure 3.2 : *Le processus de reconnaissance d'expression faciale*

1.3. Le système proposé pour la reconnaissance d'expression faciale

Un bloc diagramme est présenté dans la figure 3.3 et une description détaillée des différentes étapes est présentée dans les volets suivants du document.

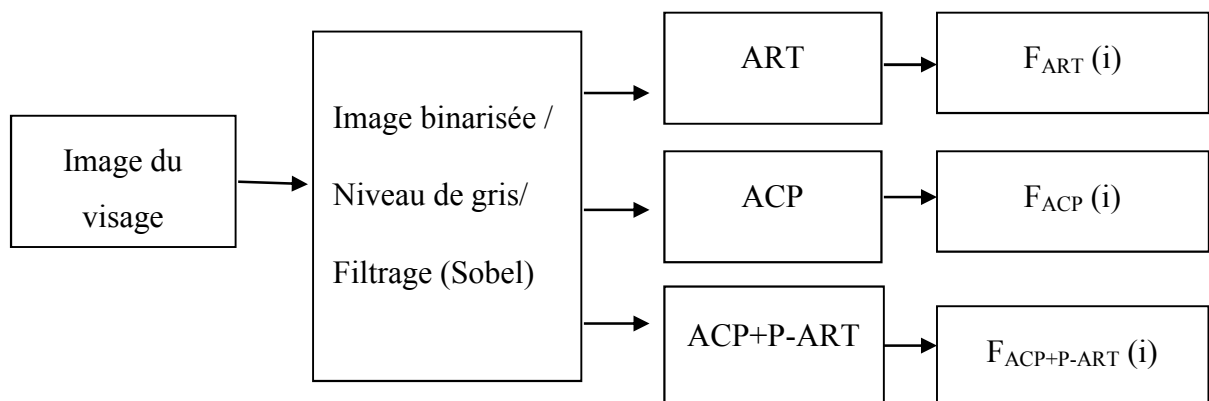


Figure 3.3 : *Le bloc diagramme du système de reconnaissance d'expression faciale*

1.3.1. Le prétraitement

Une fois le visage détecté dans l'image, le système de reconnaissance faciale commence d'abord par un bloc de prétraitement puisque l'image est souvent bruitée, mal éclairée, un contraste insuffisant ou l'image est de grande dimension. Dans notre cas, nous avons procédé de différentes manières et comparer à chaque fois les résultats d'extraction et de reconnaissance.

La procédure de prétraitement retenue de notre système FER, effectue les étapes suivantes pour convertir une image de la base de donnée JAFFE ou CK en une image à expression pure normalisée, pour l'extraction de caractéristiques:

1. Détecter manuellement les points caractéristiques du visage, y compris les yeux, le nez et la bouche;
2. Effectuer une rotation, pour aligner les coordonnées de l'œil;
3. Localiser et rogner la région de la face à l'aide d'un rectangle en fonction du modèle de face, comme indiqué sur la Fig.3.4

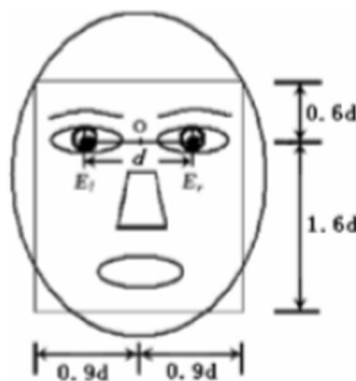


Figure 3.4 : Le prétraitement des images des bases de données utilisées (redimensionnement)

Il est vrai que les images en couleurs fournissent des informations plus riches que celles qui sont en niveau de gris, néanmoins, dans notre cas nous devons opter pour les méthodes qui sont à la fois efficaces et rapides.

Donc dans le cas de bases de données en couleur, toutes les images sont transformées en niveau de gris par la formule suivante :

$$I(x, y) = 0.2126 * R(x, y) + 0.712 * G(x, y) + 0.0722 * B(x, y) \quad 3.1$$

Ou $I(x, y)$ est l'intensité du pixel (x, y) de l'image en niveau de gris, $R(x, y), G(x, y), B(x, y)$ représentent respectivement, les composant Rouge, verte et Blue du pixel (x, y) .

Binarisation des visages

Le but de la binarisation est d'affecter un niveau uniforme aux pixels pertinents et d'éliminer les autres. Le seuillage consiste à affecter le niveau 255 aux pixels dont la valeur est supérieure à un seuil S et 0 le niveau aux autres. Le graphe de la transformation correspondante est le suivant (figure 3.5) :

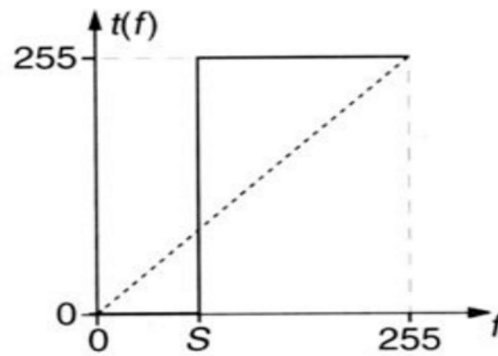


Figure 3.5 : La fonction de seuillage (la binarisation)

3.3.2. L'extraction des caractéristiques

Le processus d'extraction des caractéristiques convertit les données des pixels à des représentations et configuration spatiale du visage et ses composants. La représentation extraite est utilisé dans le processus de la classification.

L'extraction des caractéristiques réduit généralement les dimensions de l'espace d'entrée [32]. Dans ce travail on utilise un descripteur de forme région et une méthode holistique pour l'extraction de caractéristiques de l'image du visage. La fusion de l'information ou des différents ensembles de caractéristiques augmente la précision des systèmes biométriques [61]. Dans ce travail, l'information sur l'apparence (ACP) et l'information issue du descripteur région (ART) sont fusionnées pour obtenir une meilleure représentation.

Un descripteur de formes par approche région caractérise la distribution des pixels dans un objet ou une région 2D.

Il est basé sur l'étude des frontières et des pixels internes de la région à décrire. Il s'applique à un grand nombre d'objets, comme des objets complexes constituent de multiples régions discontinues ou des objets simples avec ou sans trou. Les techniques d'analyse de formes basées sur les moments appartiennent à la classe des descripteurs de formes par approche région.

3.3.2.1. Le Descripteur ART (Angular Radial Transform)

L'avantage de ce type de descripteurs est qu'ils sont tous invariants aux transformations de l'image qui correspondent à la rotation, le changement de l'échelle et la variation de l'éclairage. Ces propriétés les rendent parfaitement adaptés à la construction des vecteurs caractéristiques représentant l'expression ou la pose de la tête.

La transformation ART, Angular Radial Transform, proposée en(1999), est une transformation 2D complexe, définie sur un disque unité, en coordonnées polaires avec la formule suivante :

$$F_{nm} = \int_0^{2\pi} \int_0^1 V_{nm}(\rho, \theta) \cdot f(\rho, \theta) \rho d\rho d\theta \quad 3.2$$

Dans cette formule, F_{nm} est un coefficient d'ART d'ordre n et m , c'est une fonction image à deux variables (2D) définie en coordonnées polaires, et $V_{nm}(\rho, \theta)$ est une fonction de base d'ART qui est séparable dans la direction angulaire et radiale :

$$V_{nm}(\rho, \theta) = \frac{1}{2\pi} \exp(jm\theta) R_n(\rho) \quad 3.3$$

$$R_n(\rho) = \begin{cases} 1 & n = 0 \\ 2\cos(n\pi\rho) & n \neq 0 \end{cases}$$

Les indices n et m sont des indices respectivement radiaux et angulaires qui définissent l'ordre du coefficient F_{nm} .

Pour chaque image, il y aura $n \times m$ coefficients ART. Le problème est de savoir combien sont nécessaires pour la description d'une image. Les coefficients sont ensuite utilisés pour mettre en correspondance deux images. Pour déterminer le nombre de coefficients, il faut analyser les covariances entre les fonctions de base. Pour des descripteurs orthogonaux, il n'y a aucune covariance entre les fonctions de base. Le comité MPEG-7 propose de fixer n à 3 et m à 12. Ce sont les valeurs que nous utilisons dans notre étude. Ainsi, pour chaque image, il y aura $3 \times 12 = 36$ coefficients de l'ART.

Dans cette équation, au lieu de convertir l'image f en coordonnées polaires, les fonctions de base d'ART V_{nm} sont converties en coordonnées cartésiennes. Il faut noter que les coefficients d'ART calculés sont complexes.

L'équation 3.2 est reformulée en coordonnées cartésiennes par :

$$F_{nm} = \int \int V_{nm}(x, y) f(x, y) dx dy \quad 3.4$$

Normalisation des coefficients

Comme les coefficients de F_{nm} sont complexes, on se sert des modules des coefficients $\|F_{nm}\|$ qui sont normalisés par $\|F_{00}\|$. Ceci est appelé la normalisation de surface, puisque le premier coefficient F_{00} est égal à la somme des pixels non nuls dans l'image. Nous disposons ainsi de $n * m - 1$ coefficient, puisque F_{00} est utilisé pour la normalisation.

La principale différence entre les moments de Zernike et la transformation ART porte sur leurs fonctions de base qui sont utilisées pour extraire les coefficients de l'image qui décrivent la forme. Sauf qu'ils sont calculés tous deux dans un espace cartésien, ce qui peut conduire à des erreurs qui déforment la propriété ajoutée de la méthode.

Ces dernières années, quelques aspects importants des moments de Zernick ZM ont été étudiés, comme la question de la précision et des calculs rapides algorithmes utilisés [62,63].

En ce qui concerne le problème de la précision, il a été démontré qu'il y avait deux erreurs liées au calcul de ZM à partir de données numériques [64]: premièrement, l'erreur géométrique existe en raison de la nature circulaire du champ ; des fonctions d'image de Zernike.

La deuxième erreur numérique se produit lorsque l'implication implicite 2-D est approximative par une formule de cubage [62]. Il est intéressant de noter que l'erreur géométrique est inhérente et inévitable tant que nous avons utilisé des données cartésiennes classiques sur la base du calcul.

Pour cela, nous proposons un nouvel algorithme de calcul de l'ART qui protège fortement les propriétés intrinsèques de la méthode.

Deuxièmement, pour tester la performance de la méthode, nous proposons d'appliquer sur l'un des problèmes majeurs actuels, qui est la reconnaissance des expressions faciales.

Une comparaison avec ART calculé dans le domaine cartésien, Polar ART conserve sa robustesse à la rotation.

Puisque ART est un descripteur de forme, nous proposons également une combinaison avec l'ACP qui exploitera les informations d'apparence.

Nous montrerons que la combinaison de l'information de forme et de l'apparence basée sur l'ART polaire et l'ACP améliore les résultats de l'état de l'art.

Principe

ART est représenté généralement en coordonnées cartésiennes, les images numériques sont représentées par des pixels carrés (Figure 3.6).

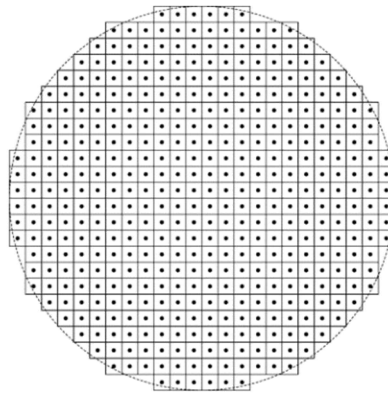


Figure 3.6 : *La Grille cartésienne de pixels pour calculer les moments ART*

Dans cette section, nous présentons un algorithme pour le calcul des ART en coordonnées polaires.

3.3.2.2.L'ART Polaire (P-ART) proposé

En calculant sur tous les angles $[0-2\pi]$, nous obtenons :

$$F_{mn} = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 V_{mn}^*(r, \theta) f(r, \theta) r dr d\theta \quad 3.5$$

Avec $V_{mn}^*(r, \theta)$, est le complexe conjugué de $V_{mn}(r, \theta)$.

$$V_{mn}(r, \theta) = R_m(r) \cdot W_n(\theta) \quad n, m \in N$$

$$R_m(r) = \begin{cases} 1 & m = 0 \\ 2\cos(m\pi r) & m > 0 \end{cases} \quad 3.6$$

$$W_n(\theta) = \frac{1}{2\pi} \exp(jn\theta)$$

La discrétisation (image $N \times N$) :

$$\widetilde{F_{mn}} = \frac{1}{2\pi} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} R_m(r_{xy}) \exp(-j \cdot n\theta_{xy}) \cdot f(x, y) \quad 3.7$$

$$r_{xy} = \sqrt{x^2 + y^2} \text{ et } \theta_{xy} = \arctang\left(\frac{y}{x}\right)$$

L'équation (3.2) révèle que les polynômes ART sont exprimés par les coordonnées polaires ρ, θ . Cela nous amène à penser que l'adoption de coordonnées polaires pourrait faciliter le calcul des coefficients d'ART.

Représentation de l'image en coordonnées polaires

Nous utilisons les critères suivants pour trouver un schéma de partition approprié, en décomposant l'image en piste, chaque piste en secteurs .L'image $N \times N$ est décrite comme suit :

- Le disque unité est uniformément divisé en $N / 2$ anneaux circulaires concentriques disjoints.
- Chaque anneau est divisé en secteurs disjoints ;
($8i + 4$), $i = 0 \dots, N / 2 - 1$, le contenu le plus interne du disque contient 4 secteurs.

(8 + 4 = 12) secteurs, (12 + 8) = 20 ... etc.

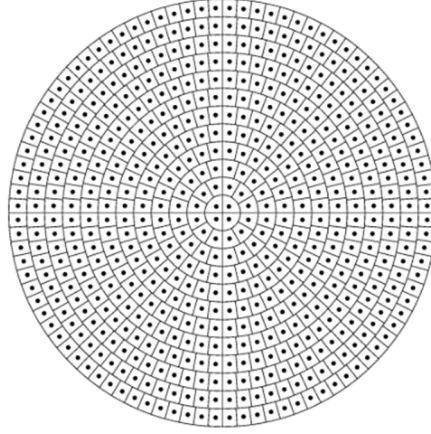


Figure 3.7 : Schématisation des pixels polaires d'ART

$$\widehat{F}_{mn} = \frac{1}{2\pi} \sum_i \sum_j \hat{f}(r_i, \theta_{i,j}) G_m(r_{ij}) H_n(\theta_{ij}) \quad 3.8$$

Avec :

$$\begin{aligned} G_m(r_i) &= \int_{\rho_i}^{r_{i+1}} R_m(r) r \cdot dr , \\ H_n(\theta_{ij}) &= \int_{\psi_{i,j}}^{\psi_{i,j+1}} \exp(-jn\theta) d\theta \end{aligned} \quad 3.9$$

$$\begin{aligned} \rho_i &= \frac{2i}{N} \quad \text{et} \quad \rho_{i+1} = \frac{(2i+1)}{N} \\ \left\{ \begin{array}{l} \psi_{i,j} = \theta_{i,j} - \frac{\pi}{(8i+4)} \\ \psi_{i,j+1} = \theta_{i,j} + \frac{\pi}{(8i+4)} \end{array} \right. \end{aligned} \quad 3.10$$

Après simplification, nous obtenons :

$$G_0(r_i) = \frac{\rho_{i+1}^2 + \rho_i^2}{2} \quad 3.11$$

$$G_m(r_i) = \left[\frac{2r_i}{m\pi} \sin(m\pi r_i) + \frac{2}{(m\pi)^2} \cos(m\pi r_i) \right]_{\rho}^{\varphi+1} \quad 3.12$$

$$H_n(\theta_{i,j}) = \frac{j}{n} [\exp(-jn\psi_{i,j+1}) - \exp(-jn\psi_{i,j})] \quad 3.12$$

Puisque les vecteurs résultants de la combinaison des descripteurs sont volumineux, nous définissons dans la section suivante l'une des techniques de sélection des variables qui permettent de réduire la dimension de ces vecteurs en choisissant les attributs les plus pertinents.

3.3.2.3.L'Analyse en Composantes Principales

L'ACP est l'un des algorithmes les plus efficaces qui a été largement utilisé pour l'extraction des caractéristiques et la réduction de dimensionnalité. L'analyse en composantes principales est utilisée pour projeter les images de visage de l'espace de l'image originale sur un sous-espace de visage, dans lequel la dimensionnalité est réduite tout en gardant l'information la plus importante. Tout d'abord, nous utilisons l'algorithme PCA pour réduire la dimension des vecteurs caractéristiques.

Ainsi, au lieu de stocker un vecteur image de taille N , on stocke un vecteur de taille $R \leq$ nombre d'images d'apprentissage P . Les étapes principales pour calculer la matrice de projection (U_{PCA}) de ACP sont :

- chaque image est représentée par un vecteur de taille N .

$$X^i = [x_1^i, x_2^i, \dots, x_N^i]^T \quad 3.14$$

- toutes les images sont centrées en soustrayant l'image moyenne de chaque vecteur image,

$$\overline{X^i} = X^i - m, \quad m = \frac{1}{p} \sum_{i=1}^p x X^i \quad 3.15$$

m : moyenne de toutes les images d'apprentissage. Ces vecteurs sont combinés, côte-à-côte, pour créer une matrice $\overline{X}$ de données d'apprentissage de taille $N \times P$ (où P est le nombre d'images d'apprentissage, N la taille du vecteur image). Le calcul de la matrice de covariance (C) qui est définie par la formule :

$$C = \overline{X} \overline{X}^T \quad 3.16$$

On calcule ensuite les vecteurs propres V et les valeurs propres D de la matrice C . Les valeurs propres de la matrice C représentent le taux de variation le long de l'axe du vecteur propre associé. La matrice de transformation de l'ACP est donc les R premiers vecteurs propres ordonnés par ordre décroissant des valeurs propres correspondantes (U_{ACP}).

Dans une première expérience, nous avons effectué une ACP sur tous les échantillons en considérant toutes les classes des expressions faciales. Ensuite, nous extrayons les principaux vecteurs propres, qui forment la base sur laquelle les caractéristiques sont extraites pour chaque échantillon d'image.

Dans une seconde expérience: considérer la construction de 7 sous-espaces PCA, chacun construit à partir d'un seul échantillon d'une classe d'expression faciale connue. Pour une image d'entrée, les entités sont extraites en considérant chacun des 7 sous-PCA, et sont utilisées dans un SVM en utilisant l'approche OVA (One Versus All). Le taux de reconnaissance obtenu est amélioré

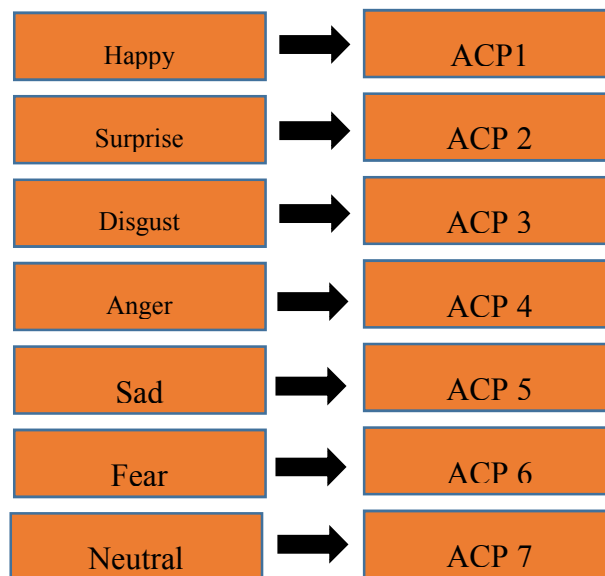


Figure 3.8 : *Le Block diagramme ACP*

3.3.2.4. Classification par SVM

Le SVM est une technique d'apprentissage statistique utilisée pour l'analyse des données et la reconnaissance des formes, proposé par Cortes et Vapnik [65]. L'algorithme SVM a été développé au cours des années 1990 à des fins industrielles [66, 67]. Il peut traiter nombreux problèmes tels que la classification, la régression et la fusion. Au début et dans sa forme de base, le SVM est utilisé comme une méthode de classification binaire basé sur un problème à deux classes. Le SVM binaire cherche à trouver l'hyperplan de séparation optimale entre les deux classes en maximisant la marge entre l'hyperplan et les deux classes qui sont étiquetées avec -1 et 1.

Pour le SVM non linéaire, la fonction de décision est une fonction non linéaire. Les données d'entrée sont reconstruites dans un espace de dimension élevée basé sur une fonction

de noyau (kernel) afin d'augmenter la précision de la classification. La précision du classificateur SVM dépend de la fonction kernel utilisée par celui-ci [68]. La fonction kernel la plus populaire est la fonction Gaussienne ou la fonction à base radiale (Radial Basis Function, RBF) [69, 70]. Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données. Le classificateur de marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables.

Le SVM est généralisé pour résoudre le problème multi-classes. Les algorithmes SVM multi-classes peuvent être divisés en deux catégories : One-Versus-All et One-Versus-One [70, 71]. Lorsque le nombre de classes des personnes dans notre système de reconnaissance est assez grand, nous utilisons la stratégie One-Versus-All basé sur le noyau RBF pour effectuer la vérification des images faciales entre les imposteurs et les clients. One-Versus-All est une méthode simple dans laquelle nous utilisons M classifieurs, un pour chaque classe. Les M classifieurs sont combinés pour prendre la décision finale.

Dans notre travail, les résultats ont été obtenus avec la fonction du noyau polynomial du cinquième ordre.

3.3.2.5. La fusion ART /ACP

La multi-biométrie tente de consolider les informations provenant de sources multiples selon trois niveaux distincts : i / niveau d'extraction des caractéristiques ; ii / niveau de score de correspondance ; et iii / niveau de décision. Étant donné que l'ensemble de caractéristiques contient des informations plus pertinentes sur les données biométriques brutes, la fusion au niveau des caractéristiques devrait fournir de meilleurs résultats de reconnaissance.

Pour améliorer la performance et la précision de la reconnaissance faciale, nous utilisons les concepts multi-algorithmiques [72], où une combinaison de différents algorithmes de reconnaissance d'expressions faciales est utilisée. Nous avons implémenté deux algorithmes de reconnaissance faciale bien connus, ils sont : i / Analyse de Composantes Principales, et ii / Transformation Radiale Angulaire.

Notre contribution, dans cet article est la fusion de deux types de caractéristiques d'un même visage, l'un lié à la forme (caractéristiques ART), l'autre lié à l'apparence globale (caractéristiques ACP) comme représenté sur la figure 3.9.

Soit $C_i = \{c_{i,1}, c_{i,2}, \dots, c_{i,p}\}$ et $Z_i = \{z_{i,1}, z_{i,2}, \dots, z_{i,q}\}$ représentent les vecteurs caractéristiques normalisés du visage i en utilisant ART et algorithmes ACP respectivement. Les vecteurs normalisés résultent de la transformation en valeurs de caractéristiques individuelles (via des schémas de normalisation tels que min-max, z-score, ...).

Le vecteur de caractéristique fusionné $X_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,d}\}$ Rd peut être obtenu en concaténant les deux vecteurs caractéristiques C_i et Z_i .

Dans MPEG-7, 12 fonctions angulaires et trois fonctions radiales sont utilisées ($n < 3$, $m < 12$), ces valeurs qui conduisent à une approximation acceptables de $f(\rho, \Theta)$ seront utilisées dans notre travail. Ainsi, pour les fonctions ART, 36 coefficients sont extraits pour chaque face. Pour l'ACP, 40 coefficients sont extraits qui correspondent aux variances les plus élevées dans les directions principales (suite aux expériences).

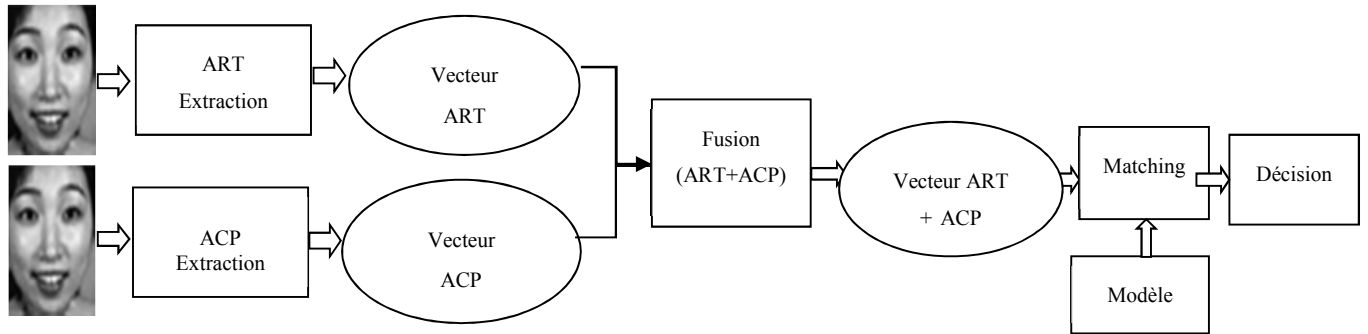


Figure 3.9 : *Extraction de caractéristiques par fusion ART + ACP*

3.4. Conclusion

Dans ce chapitre, nous avons introduit les moments d'ART dans la reconnaissance d'expression faciale et combiner deux méthodes, ART et l'ACP, au niveau de caractéristiques en se basant sur l'apparence et les régions d'intérêts.

Nous avons également présenté notre deuxième contribution, qui est le développement mathématique polaire d'ART, qu'on a baptisé (P-ART). Ce dernier présente une nouvelle forme de la transformation angulaire radiale.

Nous avons donc proposé deux algorithmes de fusion, pour la reconnaissance d'expression faciale. Le premier est basée sur l'ACP+ART et le deuxième basé sur ACP +P-ART.

CHAPITRE IV :

RESULTATS EXPERIMENTAUX

4.1. Introduction

Dans ce chapitre, nous allons aborder l'aspect pratique après avoir étudié les différentes faces de notre projet. A cet égard, nous avons proposé deux applications pratiques, à savoir, la fusion ART + ACP et le ART polaire que nous avons développé mathématiquement à partir de la représentation cartésienne de la transformation angulaire radiale. Les caractéristiques vont être tous deux fusionnés avec l'analyse en composantes principales.

L'entraînement de la machine à vecteurs de support (SVM) est utilisé pour reconnaître l'expression faciale pour une image de visage d'entrée. Enfin, les résultats expérimentaux montrent la performance du P-ART et de l'ACP. La fusion de ces deux techniques peut être meilleure que d'autres méthodes existantes de reconnaissance de l'expression faciale. Au cours de l'expérience, la base donnée de l'expression faciale féminine au Japon (JAFFE) et les bases de données Cohn-Kanade ont été utilisées.

4.2. Implémentation

Pour évaluer les fonctionnalités proposées de fusion ART et ACP, des expériences ont été réalisées sur deux bases de données : expression faciale féminine japonaise (JAFFE) [73] et Cohn-Kanade (CK) [74]. La première est une collection de 213 images de niveaux de gris différentes, prise à partir de dix personnes (femelles) exposées trois fois pour chaque expression. La base de donnée, Cohn-Kanade est composée de 486 séquences provenant de 97 sujets de taille 128×128 . Dans les deux bases de données, on peut observer que l'expression est forcée et non authentique, conduisant à des images fausses ou étiquetées confuses. C'est un problème majeur dans tous les types de bases de données collectées dans un environnement non naturel.

4.2.1. Base de données JAFFE

Les états émotionnels pris en considération sont ceux établis par Paul Ekman et la population concernée est composée de 10 sujets féminins japonais. A signaler que pour un état émotionnel donné, un sujet est pris trois fois. Quant à la base de données, elle est constituée de 213 vues dont chaque image est de taille 256×256 . La base a été conçue et assemblée par Michael Lyons, Miyuki Kamachi, et Jiro Gyoba. Les photos ont été prises au département de psychologie de l'université de Kyushu.

Résumons ces données : 10 Sujets ; 7 états émotionnels ; 213 vues. On rappelle les sept émotions évoquées précédemment :

Happiness/joie ; Sadness/tristesse ; Anger/colère ; Disgust/dégoût ; Fear/peur ; Surprise/surprise et Neutral/ neutre.



Figure 4.1 : *Echantillons de la base de données JAFFE*

4.2.2. Base de données Kohn-Canada

La base de données consiste en séquences d'expression de sujets, partant d'une expression neutre et modifiant leur visage jusqu'à une certaine expression. La base est composée de 486 séquences provenant de 97 sujets de taille 128×128 .

Dans notre travail, nous avons utilisé une base de minimum 10 sujets, chaque personne a au moins trois prises dans la même expression, qui revient à 30 sujets par expression.



Figure 4.2 : *Échantillons de la base de données Cohn-Canada*

Nous générons également des visages synthétiques en 3D avec à partir du logiciel FaceGen, téléchargeable gratuitement sur le net. Le logiciel vous permet des contrôles ajustables en fonction de l'âge, de la race, du sexe et autres paramètres.



Figure 4.3 : Échantillons de la base synthétique 3D du logiciel FaceGen

La raison derrière notre choix d'utiliser ce logiciel est de créer une expression faciale avec une précision meilleure par rapport aux bases de données existantes. Les faces générées sont utilisées dans notre test de comparaison (figure 4.3).

4.3. Résultats et analyses

Afin d'évaluer la performance de la méthode proposée, de nombreuses expériences ont été réalisées. Dans cette section, deux expériences différentes sont construites pour évaluer la performance de l'algorithme proposé. La première expérience est construite pour montrer la performance du descripteur ART sur les problèmes de FER et pour les comparer avec les autres algorithmes en ligne de la littérature. D'autre part, la seconde expérience est conçue pour tester le nouvel algorithme proposé de reconnaissance d'expression faciale sur la base de données JAFFE et la base de données CK. Les expériences comparatives avec les algorithmes d'expression faciale dans la littérature sont réalisées. De plus, les avantages de la combinaison P-ART et ACP sont présentés sur un problème de FER.

L'ART est un descripteur de région qui a été reconnu pour sa robustesse à l'échelle et à la rotation, mais qui n'est pas très utilisé dans la reconnaissance faciale. Lors de notre première expérimentation pour le FER, les taux n'étaient pas très élevés ; nous devons comprendre le principe et la sensibilité de ce descripteur pour mieux comprendre le problème. Nous avons donc testé ART sur une autre base de données, avec différents alignements

[72 × 72] et [64 × 64], nous avons en dernier, opté pour notre propre redimensionnement, tout comme il a été expliqué dans la section prétraitement du chapitre 3 (3.3.1).

Les résultats qui sont donnés par la base de données synthétique sont meilleurs et montrent ensuite que P-ART fonctionne bien (Tableau 4.4). Ensuite, il a été conclu que c'est la base de données utilisée principalement qui cause la chute du taux FER, et nous avons aligné à nouveau notre base de données à [64 × 64], des expériences ont été faites, et les résultats sont meilleurs.

La performance d'ART est évaluée. Les propriétés de l'ensemble de problèmes sont décrites ci-dessous :

1. ART est sensible à l'alignement des images dans la base de données.
2. Les expressions ne sont pas aussi évidentes que leurs étiquettes (cohérence de l'étiquette, l'expression n'est pas toujours correcte). Un exemple de ce cas est donné par la figure 4.4.

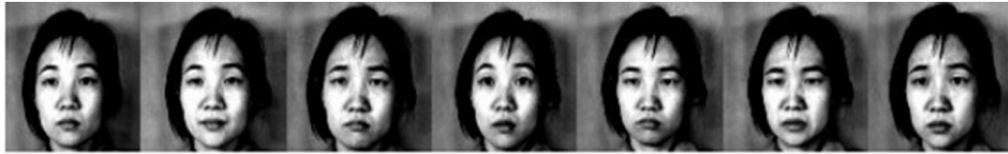


Figure 4.4 : *Echantillons de la base de données JAFFE (Problème de similarité)*

4.3.1. Résultats des taux de reconnaissance d'expression faciale RR en utilisant L'ACP

Le taux de reconnaissance moyen (RR) pour les échantillons de test JAFFE avec ACP est de 92,71% et 91,22% pour la base de données CK. Le tableau 1 montre la comparaison du taux de reconnaissance pour chaque expression avec la méthode ACP avec un ensemble d'apprentissage de 140 images et un ensemble de tests de 70 images. Le taux de précision du neutre et de la colère à l'aide de l'algorithme ACP est supérieur à celui des autres expressions pour 70 échantillons de test pour la base de données JAFFE. Le taux de reconnaissance du dégoût et de la peur avec l'algorithme PCA est plus élevé que les autres expressions pour 70 images de test pour la base de données CK. Le tableau 4.1 montre les résultats du système des 70 images testées en utilisant la méthode ACP.

FE	RR-PCA (JAFEE) %	RR-PCA (CK) %
Neutral	97.77	87.14
Anger	97.60	91.43
Disgust	92.17	94.29
Fear	84.00	94.29
Happy	83.00	90.00
Sad	83.00	90.00
Surprise	96.12	91.43
RR	92.71	91.22

Tableau 4.1 : Les taux de reconnaissance d'expression faciale donnés par L'ACP

4.3.2. Résultats des taux de reconnaissance d'expression faciale RR en utilisant ART

Le taux de reconnaissance moyen des échantillons de test en utilisant la projection ART est de 82,51% pour la base de données JAFEE et de 88,57% pour la base de données CK. Le tableau 4.2 montre la comparaison du taux de reconnaissance pour chaque expression avec la méthode ART (36 coefficients) avec un ensemble de 140 images et un ensemble de 70 images. Les taux donnés par l'expression surprise pour les bases de données JAFEE et CK est plus élevé que les autres expressions pour les 70 échantillons de test.

<i>FE</i>	<i>RR-ART % (JAFEE)</i>	<i>RR-ART % (CK)</i>
Neutral	80.33	81.43
Anger	86.00	88.57
Disgust	70.00	90.00
Fear	73.00	90.00
Happy	88.57	87.14
Sad	83.77	88.57
Surprise	96.00	94.29
RR	82.52	88.57

Tableau 4.2 : *Les taux de reconnaissance d'expression faciale donnés par ART*

4.3.3. Résultats des taux de reconnaissance d'expression faciale RR de la fusion ACP+ ART en utilisant les bases de données JAFEE et CK

Le tableau 4.3 montre les performances de reconnaissance du système proposé de fusion sur les bases de données JAFEE et CK respectivement. Nous observons ce qui suit:

1. Pour la base de données JAFEE, les expressions neutres, heureuses, de colère et de surprise sont facilement reconnaissables, tandis que la peur et les expressions tristes ne le sont pas.
2. Pour la base de données CK, la surprise, le dégoût, les expressions neutres et heureuses sont faciles à reconnaître, alors que la colère et les expressions tristes ne le sont pas.

<i>FE</i>	<i>RR-ACP+ART%</i> <i>(JAFEE)</i>	<i>RR-ACP+ART%</i> <i>(CK)</i>
Neutral	98.84	94.29
Anger	99.17	93.00
Disgust	97.33	99.10
Fear	93.66	95.00
Happy	97.17	97.14
Sad	92.33	92.00
Surprise	98.12	97.12
RR	96.66	95.37

Tableau 4.3 : Les taux de reconnaissance d'expression faciale donnés par la fusion ACP+ART

La figure 4.5 montre clairement l'augmentation du taux de reconnaissance pour les bases de données JAFEE et CK dans le cadre de l'approche par fusion.

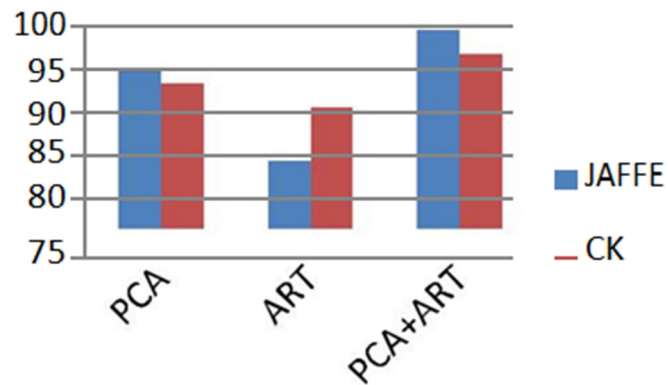


Figure 4.5 : La comparaison des taux de reconnaissance d'expression faciale ACP, ART et fusion de ACP+ART

4.3.4. Résultat obtenu en utilisant ART et P-ART (base de données JAFFE)

La figure 4.6 montre la comparaison des taux de reconnaissance des expressions faciales avec les méthodes ART et P-ART (36 coefficients). Nous concluons que le taux de reconnaissance de toutes les expressions faciales utilisant l'algorithme P-ART pour la base de données JAFFE est supérieur à l'algorithme ART. Une fonction de noyau polynomiale avec un ordre égal à cinq a été choisie pour la classification de SVM. Les taux de reconnaissance avec P-ART sont 92,42% et de 82,52% avec ART, en utilisant la base de données JAFFE.

Pour vérifier les performances de notre approche, nous avons décidé d'utiliser pour notre test une base de données synthétique que nous avons générée afin de remédier à des problèmes cruciaux de bases de données réelles (mauvais alignement, conditions de prise de vue différentes, mauvais classement, illumination ...).

Le taux d'apprentissage est de 69 % et 31% pour le test de données. Pour cela, nous avons utilisé les images en niveaux de gris telles qu'elles apparaissent dans la base de données, puis nous avons binarisé les images pour mieux encadrer les performances de notre approche. Les résultats de la reconnaissance faciale (taux de reconnaissance : RR) en utilisant P-ART sur une base de données synthétique sont présentés dans le tableau 4.4.

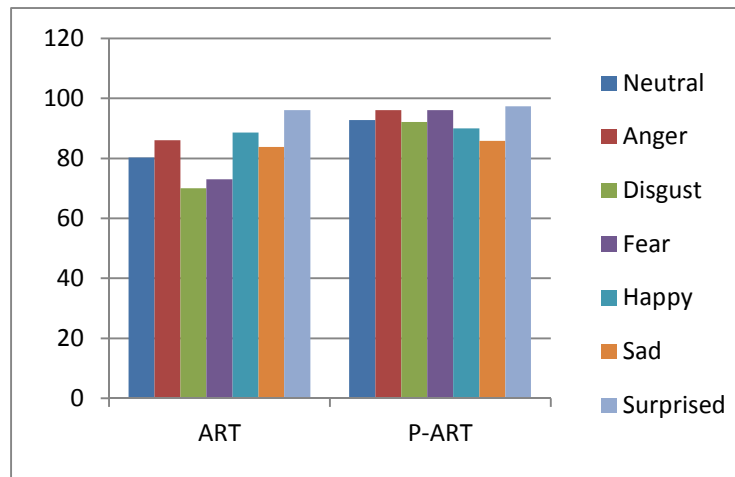


Figure 4.6 : Les taux de reconnaissance d'expression faciale pour (ART et P-ART)

FE	RR –base synthétique % [Image binarisées]	RR- base synthétique% [Image en niveau de gris]
Neutral	95,31	95,66
Anger	95,98	98,66
Disgust	92,12	89,17
Fear	98,88	97,33
Happy	92,33	92,33
Sad	88,11	89,33
Surprise	97,50	96,12
RR	94,31	94,08

Tableau 4.4 : Les taux de reconnaissance d'expression faciale donnés par P-ART en utilisant la base synthétique

4.3.5. Résultat obtenu avec ACP / P-ART indépendamment et avec la fusion (ACP + P-ART) (bases de données JAFFE et CK)

Le taux de reconnaissance moyen (RR) pour les échantillons JAFFE test, avec ACP est de 91,66% et 91,22% pour la base de données CK. Le tableau 4.5 révèle la comparaison du taux de reconnaissance pour chaque expression avec des méthodes ACP sur un ensemble d'apprentissage de 140 images et un ensemble de test de 70 images. Le taux de précision du neutre et de la colère en utilisant l'algorithme ACP est plus élevé que d'autres expressions pour 70 échantillons de test utilisant la base de données JAFFE.

Pour cela, 40 coefficients sont extraits qui correspondent aux variances les plus élevées dans les directions principales. Le taux de reconnaissance du dégoût et de la peur avec l'algorithme ACP est plus élevé que d'autres expressions pour 70 images de test pour la base de données CK.

FE	ACP % (JAFEE)	ACP % (CK)	P-ART% (JAFEE)	P-ART % (CK)	(ACP+ P-ART) (JAFEE)	(ACP+ P-ART) (CK)
Neutral	97.77	87.14	92.33	93.43	99.84	94.29
Anger	97.60	91.43	96.00	95.57	99.17	93.00
Disgust	94.17	94.29	90.00	92.00	97.66	100.00
Fear	87.00	94.29	96.00	95.00	96.66	95.00
Happy	84.33	90.00	89.57	90.14	91.84	93.14
Sad	84.66	90.00	85.77	88.57	94.66	93.66
Surprise	96.12	91.43	97.33	95.29	99.33	100.00
RR	91.66	91.22	92.42	92,85	97.02	95.58

Tableau 4.5 : Les taux de reconnaissance d'expression faciale donnés

Le tableau 4.5 révèle la comparaison du taux de reconnaissance en utilisant P-ART pour chaque expression pour 70 échantillons de test, (36 coefficients). Une fonction de noyau polynomiale avec un ordre égal à 5 a été choisie pour la classification de SVM. Les échantillons de test JAFEE de taux de reconnaissance moyen avec P-ART sont 92,42% pour la base de données JAFEE et 92,85% pour la base de données CK. Le taux de précision de l'expression surprise en utilisant l'algorithme ART pour les bases de données JAFEE et CK est plus élevé que les autres expressions pour 70 échantillons de test.

Ensuite, le vecteur de caractéristiques fusionné, est obtenu en concaténant les deux vecteurs caractéristiques (ACP+ P-ART).

Nous montrons les performances de reconnaissance du système proposé, basé sur la fusion, sur les bases de données JAFEE et CK respectivement. Nous observons ce qui suit.

1. Pour la base de données JAFEE, les expressions neutres, dégoût, colère et surprise sont facilement reconnues, tandis que les expressions heureuses, peur et tristesse ne le sont pas.

2. Pour la base de données CK, les expressions surprise et dégoût sont faciles à reconnaître alors que la colère, les expressions heureuses et tristes ne le sont pas.
3. Le tableau 4.5 montre clairement l'augmentation du taux de reconnaissance de toutes les expressions faciales dans l'approche de fusion (ACP + P-ART).

4.3.6. La comparaison des taux de reconnaissance d'expression faciale de plusieurs méthodes

Méthodes	Sujets	Images	Classes	RR (%)
Salient features vectors [75]	10	213	7	85,92
FEETS +PRNN [76]	10	210	7	85,2
Radial encoded jets [77]	10	213	7	89,67
LDNG [78]	10	213	7	90,6
BDBN [36]	10	213	7	93
Local curvet transform [79]	10	213	7	94,65
CSO-GA- PSO+SVM [80]	10	310	6	93,86
Gabor texture and CBP [81]	10	210	7	71,43
LBP+HOG [82]	10	213	7	87,6
ACP+ P- ART(Proposed)	10	213	7	97,02

Tableau 4.6 : Comparaison des taux de reconnaissance d'expression faciale entre plusieurs méthodes (JAFFE et CK /CK+)

Les taux de reconnaissance obtenus par les différentes méthodes de reconnaissance faciale, en particulier, celles qui combinent deux algorithmes, sont donnés dans le tableau 4.6. Nous avons choisi les algorithmes qui ont utilisé les mêmes bases de données, que nous avons utilisés (même nombre de sujets et de classes), afin de pouvoir comparer avec les résultats de notre approche proposée.

Les taux de reconnaissance d'expression faciale de notre approche proposée (la fusion ACP +P-ART), donne de meilleurs résultats, avec **97,02 %** pour la base de donnée JAFFE.

4.4. Conclusion

La multi-biométrie tente de consolider les informations provenant de sources multiples à trois niveaux distincts : niveau d'extraction de caractéristiques, niveau de correspondance de niveau, niveau de décision.

Étant donné que l'ensemble de caractéristiques, contient des informations plus pertinentes sur les données biométriques brutes, la fusion au niveau des caractéristiques devrait fournir de meilleurs résultats de reconnaissance. Pour améliorer la performance et la précision de la reconnaissance faciale, nous utilisons les concepts multi-algorithmiques, où une combinaison de différents algorithmes de reconnaissance d'expressions faciales est utilisée. Nous avons implémenté deux algorithmes de reconnaissance de visage, à savoir ACP et ART (P-ART).

Par conséquent, ce travail a présenté une nouvelle méthode de fusion utilisant un descripteur basé sur la région P-ART (version polaire) et une ACP holistique, pour l'extraction des vecteurs de caractéristiques et SVM avec une fonction de noyau polynomiale en tant que classifieur. La solution proposée conduit à un système capable de classer les sept émotions universelles de base, avec une précision moyenne de plus de 97% pour JAFFE et de 95,58% pour CK.

CONCLUSION GENERALE

L'aptitude à reconnaître les expressions faciales est cruciale pour le succès des interactions sociales.

Plusieurs autres domaines sont concernés par l'analyse des expressions faciales, notamment dans le traitement d'image, la robotique et la reconnaissance faciale en biométrie.

Les variations d'expression constituent des limitations importantes à la reconnaissance de visages en deux dimensions. Dans cette thèse, nous proposons d'augmenter la robustesse et performance des taux de reconnaissances faciales. Pour cela, nous proposons d'utiliser une fusion entre une méthode globale d'apparence, l'analyse en composantes principales (ACP) et un algorithme basé région, Angular Radial Transform (ART).

Ce travail présente également une nouvelle méthode de reconnaissance d'expression faciale (FER) utilisant un développement mathématique polaire basé sur la transformation radiale angulaire. La transformation radiale angulaire polaire, baptisée (**P -ART**) est combinée à l'analyse en composantes principales (ACP), pour améliorer la performance et la précision de la reconnaissance faciale (concept multi-algorithmiques). Le nouvel ART est un descripteur de forme différent que celui du descripteur cartésien conventionnel.

L'entraînement de la machine à vecteurs de support (SVM) est utilisé pour reconnaître l'expression faciale pour une image de visage d'entrée.

Par conséquent, ce travail a présenté une nouvelle méthode de fusion de caractéristiques utilisant un descripteur basé sur la région P-ART (version polaire) et une ACP holistique en tant que vecteurs de caractéristiques et SVM avec une fonction de noyau polynomiale en tant que classifieur. La solution proposée conduit à un système capable de classer les sept émotions universelles de base, avec une précision moyenne de plus de 97% pour la base de données JAFFE et de 95,58% pour la base CK. La fusion de ces deux techniques peut être alors meilleure que d'autres méthodes existantes de reconnaissance de l'expression faciale.

Il sera important de considérer d'autres bases de données pour assurer une meilleure évaluation des performances. D'éventuelles améliorations pourraient alors être envisagées au niveau des méthodes de caractérisation et de classification afin d'augmenter le taux de reconnaissance et également respecter les contraintes de temps réel.

REFERENCES BIBLIOGRAPHIQUES

- [1] S. Liu, M. Silverman. “A Practical Guide to Biometric Security Technology”, IEEE Computer Society, IT Pro-Security, Janvier-Février 2001.
- [2] Attalah bilal. “Conception d’un système de reconnaissance d’empreintes digitales par apprentissage”, thèse de magister, usto, 2012.
- [3] Benchenane Ibtissem, “Etude et mise au point d’un procédé biométrique multimodal pour la reconnaissance des individus”, thèse de doctorat en sciences, usto, 2015-2016.
- [4] Aizi kamel, “Identification Biométrique Multimodale à Distance”, thèse de doctorat LMD, usto, 2016.
- [5] L. Allano. “La Biométrie multimodale : stratégies de fusion de scores et mesures de dépendance appliquées aux bases de personnes virtuelles”, Thèse de doctorat, université d’Evry Val d’Essonne, 2009.
- [6] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-pie. “Image and Vision Computing”, 2010,28(5) :807–813.
- [7] Takeo Kanade. “Picture processing system by computer complex and recognition of human faces”, In doctoral dissertation, Kyoto University. pages 6, 15 et 19, November 1973.
- [8] B. Fasel and J. Luetttin. “Automatic facial expression analysis : a survey”, Pattern Recognition, 2003, Vol. 36, no. 1, pp. 259–275.
- [9] T. D. Bui, D. Heylen, and A. Nijholt. “Combination of facial movements on a 3D talking head,” in Computer Graphics International, 2004. Proceedings, pp. 284–290.
- [10] E. A. Boyle, A. H. Anderson, and A. Newlands. “The effects of visibility on dialogue and performance in a cooperative problem solving task” Language and speech,1994, Vol.37, no. 1, pp. 1–20.
- [11] G. M. Stephenson, K. Ayling, and D. R. Rutter. “The role of visual communication in social exchange,” British Journal of Social and Clinical Psychology, 1976, Vol. 15, no. 2, pp. 113–120.
- [12] A. Mehrabian, “Communication without words,” Psychological today, 1968, Vol. 2, pp. 53–55.

- [13] R. Cowie,, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, “Emotion recognition in375 human-computer interaction”, Signal Processing Magazine, IEEE, (2001) ,18(1) 32-80.
- [14] “Application: Facial Expression Recognition”. In: Machine Learning in Computer Vision. Computational Imaging and Vision , 2005, Vol.29. Springer, Dordrecht.
- [15] Medical Press, Sarah Worth, March 27, 2017. [University of South Florida](#) (accessed on 17 January 2018).
- [16]E. Jack, C.Blais, C.Scheepers, G. Schyns and R.Caldara, “Cultural Confusions Show that Facial Expressions Are Not Universal”, Current Biology, Canada, September 2009.
- [17] Shan, C.; Gong, S.; McOwan, P.W. Facial expression recognition based on local binary patterns: A comprehensive study. Image Vis. Comput. 2009, 27, 803–816.
- [18] Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. “The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression”, In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, USA, 13–18 June 2010, pp. 94–101.
- [19] Ekman P. and Friesen W. V. “Facial action coding system”, Consulting Psychologist Press, August 1978,18, no.11:881–905.
- [20] Khadoudja GHANEM. “Reconnaissance des Expressions Faciales à Base d’Informations Vidéo ; Estimation de l’Intensité des Expressions Faciales”, Doctorat en sciences, Université Mentouri de Constantine, 2010.
- [21] Tian, Y.L., Kanade, T. and Cohn, J.F. “Facial expression analysis”, Hand Book of Face Recognition, 2005, pp.247–275.
- [22] Seyed Mehdi Lajevardi, Zahir M. Hussain. “Automatic facial expression recognition: feature extraction and selection”, Springer, 2010.
- [23] B. Heisele, P. Ho, J. Wu, T. Poggio. “Face recognition: component-based versus global approaches”, Comput. Vis. Image Underst, 2003. 91, pp.6–21.

- [24] Hong-Bo Deng ,Lian – Wen Jin ,Li-Xin Zhen, Jian –Cheng Huang. “A New Facial Expression Recognition Method based on Local Gabor Filter Bank and PCA plus LDA”, 2005.
- [25] Jacob Whitehill ,Gwen Littlewort ,Ian Fasel, Marian Bartlett. “Toward Practical Smile Detection” , IEEE Transactions, 2009.
- [26] T. Ahonen, A. Hadid, M. Pietikainen. “Face recognition with local binary patterns, in: European Conference on Computer Vision” (ECCV), 2004.
- [27] Pooja Sharma. “Feature Based Method for “Human Facial Emotion Detection using optical Flow Based Analysis”, International Journal of Research in Computer Science 2011, ISSN 2249-8265, Vol.1, Issue 1, pp. 31-38.
- [28] Sander Koelstra, Student Member, “A Dynamic Texture-Based Approach to Recognition of Facial Actions and Their Temporal Models”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, Vol. 32, No. 11.
- [29] Martin Michael Schulze. “Facial expression recognition with support vector machines”, University of applied sciences Konstanz, Germany, 2002.
- [30] Hong-Bo Deng, Lian – Wen Jin ,Li-Xin Zhen, Jian –Cheng Huang. “A New Facial Expression Recognition Method based on Local Gabor Filter Bank and PCA plus LDA”, International Journal of Information Technology, 2005, Vol. 11, No. 11.
- [31] Amir Jamshidnezhad, Mdjan Nordin. “A Classifier Model based on the Features Quantitative Analysis for Facial Expression Recognition”, Proceeding of the International Conference on Advanced Science, Engineering and Information Technology, 2011.
- [32] Aleksic, P.S; Katsaggelos,A.K. “Automatic facial expression using facial animation parameters and multistream HMMs”, Information forensics and security, IEEE transactions , march 2006, Vol.1, issue 1.
- [33] Uçar, A., Demir, Y. & Güzeliş. “A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering”, C. Neural Comput & Applic, Springer, 2016.
- [34] Yu-Dong Zhang, Zhang-Jing Yang, Hui-Min Lu. “Facial Emotion Recognition Based on Biorthogonal Wavelet Entropy, Fuzzy Support Vector Machine, and Stratified Cross Validation”, IEEE Access, 2016, Vol. 4.

- [35] Shaoxin.Li, Shiguang. Shan Xilin.Chen. “AU-inspired Deep Networks for Facial Expression Feature Learning”. *Neurocomputing*, 2015, 159, pp.126-136.
- [36] Liu, P., Han, S., Meng, Z., & Tong, Y. “Facial expression recognition via a boosted deep belief network”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1805-1812.
- [37] KHORRAMI, Pooya, PAINE, Thomas, HUANG, Thomas. “Do deep neural networks learn facial action units when doing expression recognition?” In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 19-27.
- [38] A. Ono. “Face recognition with Zernike moments”, *Systems and Computers in Japan*, 2003, 34(10), 26-35.382.
- [39] C. Singh, N. Mittal, and E Walia. “Face recognition using Zernike and complex Zernike moment features”, *Pattern Recognition -383 and Image Analysis*, 2011, 21, pp. 71-81.
- [40] W.-Y. Kim and Y.-S. Kim. “A New Region-Based Shape Descriptor,” *ISO/IEC MPEG99/M5472*, Maui, Hawaii, 1999.
- [41] M. Bober, “Shape Descriptor Based on Curvature Scale Space,” *MPEG-7 Proposal P320*, Lancaster U.K., 1998.
- [42] J. M. Martinez, R. Koenen, and F. Pereira. “MPEG-7: The Generic Multimedia Content Description Standard,” *IEEE Multimedia (Part 1)*, 2002, 9 (2), 78–87.
- [43] J. M. Martinez, “Standards—MPEG-7: Overview MPEG-7 Description Tools,” *IEEE Multimedia (Part 2)*, 2002 (3), pp. 83–93.
- [44] Ricard, J., Coeurjolly, D., &Baskurt, A. “Generalizations of angular radial transform for 2D and 3D shape retrieval”, 2005, *Pattern Recognition Letters*, 26(14), 2174-2186.
- [45] Matthew A Turk and Alex P Pentland. “Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition*”, *Proceedings CVPR’91*, IEEE Computer Society Conference, IEEE, 1991, pp. 586–591.
- [46] Baback Moghaddam, Wasiuddin Wahid, and Alex Pentland. “Beyond eigenfaces : Probabilistic matching for face recognition”, In *Automatic Face and Gesture Recognition*, , *Proceedings. Third IEEE International Conference. IEEE*, 1998, pp. 30–35 (Cité pages vii et 17)

- [47] W.Zhao, R.Chellappa, P.J.Phillips. “ACM Computing Surveys”, December 2003, Vol. 35, No. 4, pp. 399–458.
- [48] G. Guo, S.Z. Li, K. Chan. “ Face Recognition by Support Vector Machines”, Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition, 26-30 March 2000, Grenoble, France, pp. 196-201.
- [49] Algorithmes de Support Vector. “Machines Sebastien Gadat”, Laboratoire de Statistique et Probabilités, UMR 5583 CNRS-UPS, www.lsp.ups-tlse.fr/gadat.
- [50] Olivier Bousquet. “Introduction aux “Support Vector Machines (SVM) ”, Centre de Mathématiques Appliquées Ecole Polytechnique, Palaiseau Orsay, 15 Novembre 2001.
- [51] S. Jeannin. “Mpeg-7 Visual part of experimentation Model Version 9.0”. Dans ISO/IECJTC1/SC29/ WG11/N3914, 55th Mpeg Meeting, Pisa, Janvier 2001.
- [52] W.-Y. Kim et Y.-S Kim. “A new region-based shape descriptor”, Dans TR 15-01, Pisa, Décembre 1999.
- [53] Omar Mohammed Wahdan, Khairuddin Omar and Mohammad. “Logo Recognition System Using Angular Radial Transform Descriptors”, Department of System Science and Management, Faculty of Information Science and Technology, University Kebangsaan Malaysia, Journal of Computer Science,2011, 7 (9), pp. 1416-1422.
- [54] D.-Y. Chen and M. Ouhyoung. “A 3d model alignment and retrieval system,” in International Computer Symposium , Hualien, R.O.C, 2002.
- [55] J. Fang and G. Qiu. “Human face detection using angular radial transform and support vector machines,” in International Conference on Image Processing, 2003, pp. I: 669–672.
- [56] Ricard, J., Coeurjolly, D., &Baskurt, A. “Generalizations of angular radial transform for 2D and 3D shape retrieval”, Pattern Recognition Letters, 2005, 26(14), 2174-2186.
- [57] S.H. Lee,S.Sharma, L. Sang, J. Park, Y.G.Park.“An Intelligent Video Security System using Object Tracking and Shape Recognition”, ACVIS LNCS Springer-Verlag, Berlin Heidelberg, 2011, Vol. 6915, pp. 471-482.
- [58] Fernandes, Jovan de Andrade Aragao, Maria Gessica dos Santos. “Geometrical Approaches for Facial Expression Recognition Using Support Vector Machines”, 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2017.

- [59] H.A. Rowley, S. Baluja, T. Kanade. "Neural network-based face detection", Transactions on Pattern Analysis and Machine Intelligence, IEEE, Jan. 1998, Vol.20, no.1, pp.23-38,
- [60] L.Serrir et W.Sebba. "Reconnaissance d'expressions faciales par apprentissage, Mémoire de Master ", Faculté d'informatique, USTHB, 2013.
- [61] A. A. Ross and R. Govindarajan. "Feature level fusion of hand and face biometrics", in Defense and Security, 2005, pp. 196-204.
- [62] S. X. Liao and M. Pawlak. "On the accuracy of Zernike moments for image analysis," IEEE Trans. Pattern Anal. Mach. Intell. 1998, Vol. 20, no. 12, pp. 1358–1364.
- [63] R. Mukundan and K. Ramakrishnan. "Moment Functions in Image Analysis: Theory and Applications". Singapore: World Scientific 1998.
- [64] M. Pawlak and S. X. Liao. "On the recovery of a function on a circular domain". IEEE Trans. Inf. Theory, 2002, Vol.48, no. 10, pp. 2736–2753.
- [65] C. Cortes and V. Vapnik. "Support-vector networks" Machine learning, 1995, Vol. 20, pp. 273-297,
- [66] M. Paci, L. Nanni, A. Lahti, K. Aalto-Setälä, J. Hyttinen, and S. Severi. "Non-binary coding for texture descriptors in sub-cellular and stem cell image classification", Current Bioinformatics, 2013, Vol.8, pp. 208-219.
- [67] B. E. Boser, I. M. Guyon, and V. N. Vapnik. "A training algorithm for optimal margin classifiers", in Proceedings of the fifth annual workshop on Computational learning theory, 1992, pp. 144-152.
- [68] S. Meshgini, A. Aghagolzadeh, and H. Seyedarabi. "Face recognition using Gaborbased direct linear discriminant analysis and support vector machine", Computers & Electrical Engineering, 2013, Vol. 39, pp. 727-745.
- [69] Y. Lei, M. Bennamoun, and A. A. El-Sallam, "An efficient 3D face recognition approach based on the fusion of novel local low-level features", Pattern Recognition, 2013, Vol. 46, pp. 24-37.
- [70] H. Chih-Wei and L. Chih-Jen. "A comparison of methods for multiclass support vector machines", Neural Networks, IEEE Transactions, 2002, Vol. 13, pp. 415-425.

- [71] L. Zhao, Y. Song, Y. Zhu, C. Zhang, and Y. Zheng. "Face recognition based on multiclass SVM", in Control and Decision Conference, CCDC '09. Chinese, 2009, pp. 5871-5873.
- [72] J.L. Wayman. "Biometric Data Fusion: Moving beyond multimodal biometrics". International workshop in recent advances in biometrics systems, ITT Kanpur, April 2005, pp 245-250.
- [73] Lyons, M., Akamatsu, S., Kamachi, M., Gyoba. J. "Coding facial expressions with Gabor wavelets, in automatic face and gesture recognition", In: Proceedings. Third IEEE International Conference on Japan, 1998, pp. 200–205.
- [74] Kanade T, Cohn JF, Yingli T. "Comprehensive database for facial expression analysis", In: IEEE 4th international conference on automatic face and gesture recognition. Pittsburgh, USA, 2000. pp 46–53.
- [75] M. Kyperountas, A. Tefas, I. Pitas. "Salient feature and reliable classifier selection for facial expression classification", Pattern Recognition, 2010, 43, pp. 972–986.
- [76] Wong, J. J., & Cho, S. Y. "A face emotion tree structure representation with probabilistic recursive neural network modeling", Neural Computing and Applications, 2010, 19(1), pp. 33-54.
- [77] Gu, W., Xiang, C., Venkatesh, Y. V., Huang, D., & Lin, H. "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis", Pattern Recognition, 2012, 45(1), 80-91.
- [78] Ramirez Rivera, A., Rojas Castillo, J., & Chae, O. "Local directional number pattern for face analysis: Face and expression recognition. Image Processing", IEEE Transactions on, 2013, 22(5), 1740-1752.
- [79] Ays, egurc, ar, Yakup Demir ,Cu`neytGu`zelis. "A new facial expression recognition based on curvelet transform and on line sequential extreme learning machine initialized with spherical clustering", Springer-Verlag London, 2014.
- [80] [T. V. Vivek](#); [Guddeti Ram Mohana Reddy](#). "A Hybrid Bioinspired Algorithm for Facial Emotion Recognition Using CSO-GA-PSO-SVM", Fifth International Conference on Communication Systems and Network Technologies (CSNT), 2015.

[81] Li, L. S., Ying, Z. L., & Huang, B. B. “Facial Expression Recognition Based on Gabor Texture Features and Centre Binary Pattern”, In *Applied Mechanics and Materials*, 2015, Vol. 742, pp. 257-260.

[82] Yanpeng Liu, Yibin Li, Xin Ma and Rui Song. “Facial Expression Recognition with Fusion Features Extracted from Salient Facial Areas”, School of Control Science and Engineering, Shandong University, Jinan 250061, China, 2017.

ANNEXE

PRODUCTION SCIENTIFIQUE DE L'AUTEUR

PUBLICATIONS (JOURNAUX)

- Publication dans le journal international de la biométrie IJBM, :
Imène Taleb*, Madani Ould Mammar and Abdelaziz Ouamri, ‘**New Face Expression Recognition using Polar Angular Radial Transform and Principal Component Analysis**’. **Int. J. Biometrics, Vol. 10, No. 2, 2018.**
Cet article présente une nouvelle méthode de reconnaissance d'expression faciale (FER) utilisant un développement mathématique polaire basé sur la transformation radiale angulaire (ART) que nous avons baptisé (P-ART), ce dernier est combiné à la méthode d'apparence l'ACP. Les caractéristiques des deux méthodes sont concaténées pour augmenter le taux de reconnaissance. Le vecteur de caractéristiques résultant est classifié par la machine à vecteurs de support (SVM) pour reconnaître l'expression faciale pour une image de visage d'entrée. Enfin, les résultats expérimentaux montrent la performance du P-ART et de l'ACP. La fusion de ces deux techniques peut être meilleure que d'autres méthodes existantes de reconnaissance de l'expression faciale. Au cours de l'expérience, la base de l'expression faciale féminine du Japon (JAFPE) et les bases de données Cohn-Kanade ont été utilisées.
- Publication dans : International Journal of International Journal of Scientific Research & Engineering Technology (IJSET), intitulée ‘**Improved Facial Expression Recognition Using Angular Radial Transform and Principal Component Analysis**’. Ce travail consiste à présenter la fusion de deux experts rarement fusionnés dans la littérature dans le but de reconnaître l'expression faciale, sur une personne bien déterminée (ACP+ART). L'Analyse en composante principale (ACP) est une méthode classique et la transformation angulaire radiale (ART) est moins classique, le taux de reconnaissance d'expression faciale donné par la fusion des caractéristiques de ces deux méthodes a donné un résultat meilleur que celui donné par l'implémentation de ces deux méthodes séparément.

COMMUNICATIONS (CONFERNCES ET WORKSHOPS)

1. La 6 ème Conférence Internationale en Automatique & Traitement de Signal (ATS-2018), Tanger au Maroc, du 28 au 30 mars 2018 au TECHNOPARK de Tanger. 'Improved Facial Expression Recognition Using Angular Radial Transform and Principal Component Analysis'.
2. The PHC-TASILI SCCIBOV workshop, Système Conjoint de Compression et d'Indexation Basé Objet pour la Vidéo, (SCCIBOV'2015), 2015, Sidi Bel Abbes, Algeria.
3. International conference on signal, image, vision and their applications (SIVA'15), 2015, Guelma, Algeria.
4. International symposium on programming and systems ISPS'15, 2015, USTHB, Algiers, Algeria.
5. International conference on telecommunications and TIC 5ict TELECOM 2015), 2015, Oran, Algeria.
6. International conference on image and signal processing and their applications (ISPA'15), 2015, Mostaganem, Algeria.
7. International conference on advanced communications systems and signal processing (ICOSIP'15), 2015, Tlemcen, Algeria.
8. Participation in the National Exhibition of National Outcomes of the National Research Programs at the Oran Convention Center 2014, organized by the DG-RSDT.
9. International Congress on Telecommunication and Application (ICTA'14), 2014, Bejaia, Algeria.
10. International Symposium, Concepts and Tools for Knowledge Management (KM) (ISKO-Maghreb), 2014, Algiers, Algeria.