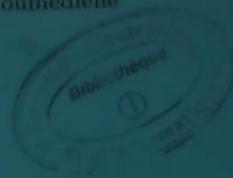


République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université des Sciences et de la Technologie Houari Boumediene



**Faculté d'Electronique et d'Informatique**

Département d'informatique

Projet de Fin d'Etudes pour l'obtention du diplôme de Master

Option : Systèmes Informatiques Intelligents

---

**Thème :**  
**Reconnaissance d'expressions faciales par**  
**apprentissage**

---

Sujet proposé par :

Pr Ahmed GUESSOUM  
Dr Saliha AOUAT

Présenté par :

Mlle Louiza SERIR  
M. Walid SEBA

Soutenu le : 19 Juin 2013

Devant le jury composé de :

M. S. LARABI : *Président*  
Mme. N. BAHA : *Membre*  
M. A. KHETTAR : *Membre*

N° Binôme : 103/2013

## Remerciements

Il existe des projets qui sont tellement soutenus de toute part de leur début à leur fin, que la réalisation de celui qui les a engendré devint presque négligeable, ce travail fait partie de ce genre de projet.

Je voudrais exprimer ma profonde gratitude à ceux, qui ont rendu à mes yeux, la traversée d'un orage pluvieux, aussi agréable qu'un voyage dans les cieux, en me donnant à chaque pas ce qu'il avait de mieux.

Je remercie Imène, aussi aimable que zen, et Lydia, la sereine qui s'est donné tant de peine.

Je remercie nos promoteurs, qui ont été le moteur, d'un travail à la hauteur, d'un engagement à grande ampleur.

Je remercie l'AUF, véritable organisme vivant, qui se développe et qui respire, en offrant un avenir, à ceux qui savent y recourir.

Je remercie mes amis, tous, un par un, sans exception, que je ne peux citer de nom, cela prendrait toute une section.

Je remercie ma famille, de la plus proche à la plus lointaine, c'est le même sang dans nos veines.

Je remercie, et là je fais une exception, trois dames que je citerais de nom, car elles le méritent amplement ; la première est ma binôme '*Louiza*', qui a supporté avec talent le phénomène que je suis, et qui a su être là à chaque fois que l'orage nous poursuit. La seconde est ma petite soeur, '*Asma*' tu es pour moi la lumière que perçoit, un aveugle pour la première fois, et ce à chaque fois, que je suis avec toi. La dernière c'est l'auteur de l'auteur, la main qui agite la plume, la mer où vient se réfugier le ruisseau.. Oui ! C'est ma '*Mère*'... Merci.

Walid

## CHAPITRE 0. REMERCIEMENTS

UN projet de fin d'études est le résultat d'un travail de longue haleine. C'est le fruit de multiples recherches, discussions, et applications au cours de diverses collaborations. L'organisation de ces idées qui constitue notre contribution modeste mais, on l'espère, originale, n'est pas possible sans un cadre de travail matériel et intellectuel favorable.

C'est pourquoi je tiens à remercier chaleureusement le responsable du Campus Numérique Francophone d'Alger M. Eddoud Abdelkader, de nous avoir si bien accueilli en cet établissement, nous ayant permis de découvrir le monde professionnel et pratique tout en nous offrant d'excellentes conditions de travail, ainsi que Mlle Berraoui Yasmina, et Mlle Touzouirt Nassera pour les discussions enrichissantes que nous avons échangées et les conseils avisés qui s'en dégageaient.

Je tiens à remercier les membres du jury d'avoir rapporté ce projet, de nous avoir honorés par l'évaluation de ce travail, pour le temps qu'ils auront bien voulu y consacrer. Qu'ils soient certains que nous considérons leur contribution comme un honneur.

Nous voudrions exprimer notre plus sincère reconnaissance au Pr. Ahmed GUES-SOUM, ainsi qu'à Mlle. AOUAT, nos promoteurs de ce projet, qui ont su s'y intéresser de façon remarquable, et chez qui nous avons trouvé soutien, disponibilité, patience et compréhension et par leurs conseils et leur aide précieuse, nous ont guidés tout au long de notre travail.

Qu'ils reçoivent toute notre gratitude pour nous avoir fait confiance dans nos choix, tout en nous apportant leur rigueur scientifique.

Je ne remercierais jamais assez ma très chère maman, qui m'a toujours soutenue dans tout ce que je fais, ainsi que mon père, l'homme le plus formidable qui existe sur terre, et sans oublier mes trois frères Amine, Sofiane et Radouane.

Je tiens à remercier aussi en particulier mon binôme, M. SEBA Walid, avec lequel j'ai appris beaucoup de choses, aussi à Mlle TOUMI Imène pour son grand soutien, Mlle YTAGHENE Lydia, sans oublier toutes les personnes qui, scientifiquement ou socialement ont contribué à la réussite de ce travail.

Je remercie également nos familles et amis respectifs pour leur soutien, leur aide et leur présence tout au long de ces mois de labeur.

Je termine en remerciant la faculté d'Electronique et Informatique d'avoir mis en œuvre cette expérience très enrichissante pour nous.

Louiza

## Table des matières

|   |    |
|---|----|
| <b>Remerciements</b>  | i  |
| <b>Résumé</b>   | ix |
| <b>Abstract</b>   | x  |
| <b>Introduction Générale</b>  | 1  |
| 0.1 Introduction . . . . .  | 1  |
| 0.2 Objectifs de travail . . . . .  | 2  |
| 0.3 Organisation du mémoire . . . . .   | 2  |
| <b>1 Les expressions faciales</b>   | 3  |
| 1.1 Introduction . . . . .  | 3  |
| 1.2 Définitions . . . . .   | 3  |
| 1.2.1 Expression faciale . . . . .  | 3  |
| 1.2.2 Émotion . . . . .   | 4  |
| 1.2.3 Mimique faciale . . . . .   | 5  |
| 1.3 Physiologie faciale . . . . .   | 5  |
| 1.4 Mesure des expressions faciales . . . . .   | 7  |
| 1.4.1 Interprétation humaine . . . . .  | 7  |
| 1.4.2 Interprétation par symbole . . . . .  | 7  |
| 1.4.2.1 Facial Action Coding System (FACS) . . . . .                                    | 7  |
| 1.4.2.2 Facial Action Coding System <sup>+</sup> ( <i>FACS</i> <sup>+</sup> ) . . . . . | 8  |
| 1.4.2.3 La norme MPEG-4 . . . . .   | 9  |
| 1.4.2.4 Candide . . . . .   | 10 |
| 1.4.2.5 Le modèle Sign Stream [1] . . . . .   | 11 |
| 1.5 Conclusion . . . . .  | 11 |
| <b>2 Analyses et classification des expressions faciales par apprentissage</b>          | 13 |
| 2.1 Introduction . . . . .  | 13 |
| 2.2 Historique . . . . .  | 13 |
| 2.3 Analyse humaine des expressions faciales . . . . .                                  | 15 |
| 2.4 Analyse automatique d'expressions faciales . . . . .                                | 15 |
| 2.4.1 Détection du visage . . . . .   | 16 |
| 2.4.2 Extraction des caractéristiques . . . . .   | 17 |



---

**TABLE DES MATIÈRES**

|          |   |           |
|----------|---|-----------|
| 2.4.3    | Classification des expressions . . . . .                        | 18        |
| 2.4.3.1  | Modèle spatio-temporel . . . . .                                | 19        |
| 2.4.3.2  | Modèle spatial . . . . .  | 19        |
| 2.5      | Conclusion . . . . .  | 19        |
| <b>3</b> | <b>Classification</b>   | <b>20</b> |
| 3.1      | Introduction . . . . .  | 20        |
| 3.2      | État de l'art sur les classificateurs . . . . .                 | 21        |
| 3.2.1    | La fonction discriminante . . . . .                             | 21        |
| 3.2.2    | La fonction de coût . . . . .                                   | 21        |
| 3.3      | Les classificateurs paramétriques . . . . .                     | 21        |
| 3.3.1    | Classificateur Euclidien . . . . .                              | 22        |
| 3.3.2    | Classificateur Quadratique . . . . .                            | 22        |
| 3.4      | Les Classificateurs Non Paramétriques . . . . .                 | 23        |
| 3.4.1    | Le Plus Proche Voisin . . . . .                                 | 24        |
| 3.4.2    | Les k plus Proches Voisins . . . . .                            | 24        |
| 3.5      | Classificateurs Neuronaux . . . . .                             | 25        |
| 3.5.1    | Les neurones biologiques . . . . .                              | 25        |
| 3.5.2    | Le Perceptron (1958) . . . . .                                  | 26        |
| 3.5.3    | Le modèle formel 1987 . . . . .                                 | 27        |
| 3.5.4    | La Carte Auto-Organisatrice 1990 . . . . .                      | 28        |
| 3.5.5    | Le Perceptron Multicouche . . . . .                             | 29        |
| 3.6      | Conclusion . . . . .  | 30        |
|          |   | 31        |
| <b>4</b> | <b>Conception</b>   | <b>32</b> |
| 4.1      | Introduction . . . . .  | 32        |
| 4.2      | Résumé de l'approche utilisée . . . . .                         | 32        |
| 4.3      | Présentation des bases d'apprentissage . . . . .                | 32        |
| 4.3.1    | La récolte des données . . . . .                                | 33        |
| 4.3.2    | Présentation de la base d'images . . . . .                      | 34        |
| 4.3.3    | Filtrage de la base d'images par la méthode des juges . . . . . | 35        |
| 4.3.4    | Post-sélection . . . . .  | 36        |
| 4.4      | Caractéristiques faciales (prétraitement des données) . . . . . | 36        |
| 4.4.1    | Les points d'intérêt faciaux . . . . .                          | 36        |
| 4.4.2    | Méthode d'extraction des points faciaux . . . . .               | 39        |
| 4.4.3    | Les distances entre les points d'intérêt . . . . .              | 41        |
| 4.4.4    | Les rapports de distance . . . . .                              | 43        |
| 4.5      | Apprentissage . . . . .   | 45        |
| 4.5.1    | Le choix du réseau . . . . .                                    | 46        |
| 4.5.2    | Les entrées de notre réseau (Input) . . . . .                   | 47        |
| 4.5.3    | Les sorties de notre réseau (Output) . . . . .                  | 47        |
| 4.6      | Conclusion . . . . .  | 48        |
| <b>5</b> | <b>Expérimentation</b>  | <b>48</b> |
| 5.1      | Introduction . . . . .  | 49        |
| 5.2      | Environnement logiciel et matériel . . . . .                    | 49        |
| 5.2.1    | Hardware . . . . .  | 49        |
|          |   | 49        |

---

**TABLE DES MATIÈRES**

|  |           |
|--|-----------|
| 5.2.2 Software . . . . .                         | 49        |
| 5.3 Présentation du logiciel développé . . . . . | 51        |
| 5.3.1 Résultat et Tests . . . . .                | 56        |
| 5.3.1.1 Résultats de l'apprentissage . . . . .   | 56        |
| 5.3.1.2 Résultats des simulations . . . . .      | 62        |
| 5.4 Conclusion . . . . .                         | 63        |
| <b>Conclusion Générale</b>                       | <b>64</b> |
| 5.5 Rappel des objectifs de travail . . . . .    | 64        |
| 5.6 Objectifs atteints . . . . .                 | 64        |
| 5.7 Perspectives . . . . .                       | 65        |
| <b>A</b>   | <b>66</b> |
| <b>Bibliographie</b>                             | <b>69</b> |

## Table des figures

|      |   |    |
|------|---|----|
| 1.1  | Quelques expressions faciales . . . . .   | 4  |
| 1.2  | Théorie de Cannon-Brad sur le déclenchement des émotions [2, 3] . . . . .                               | 4  |
| 1.3  | Déclencheurs d'expressions faciales . . . . .   | 5  |
| 1.4  | Les expressions faciales universelles définies par Paul Ekman [4] . . . . .                             | 6  |
| 1.5  | Les « <i>Action Unit</i> » et leurs muscles associés . . . . .  | 8  |
| 1.6  | Squelette du MPEG-4 avec les Facial Feature Points (FFP) associés . . . . .                             | 10 |
| 1.7  | Modèle candide avec les points inter et intra individu.[5] . . . . .                                    | 11 |
| 2.1  | Guillaume Duchenne de Boulogne en compagnie de son cobaye . . . . .                                     | 14 |
| 2.2  | Processus d'analyse et de classification d'expressions faciales . . . . .                               | 16 |
| 2.3  | Détection des visages . . . . .   | 17 |
| 2.4  | Détection de certaines zones d'intérêt du visage. . . . .   | 18 |
| 3.1  | Représentation d'objets appartenant à deux classes distinctes dans un espace à deux dimensions. . . . . | 21 |
| 3.2  | Frontière fournie par le classificateur Euclidien dans le cas d'un problème à deux classes . . . . .    | 23 |
| 3.3  | Frontière obtenue par le classificateur quadratique . . . . .   | 24 |
| 3.4  | Frontière fournie par le classificateur du plus proche voisin. . . . .                                  | 25 |
| 3.5  | Schéma d'un neurone biologique . . . . .  | 27 |
| 3.6  | Représentation d'un perceptron . . . . .  | 27 |
| 3.7  | Représentation d'un perceptron multicouche . . . . .  | 28 |
| 3.8  | Frontière de décision d'un perceptron . . . . .   | 28 |
| 3.9  | Modèle de neurone formel . . . . .  | 29 |
| 3.10 | Carte auto-organisatrice à deux dimensions . . . . .  | 29 |
| 3.11 | Perceptron multicouche avec une couche cachée . . . . .   | 30 |
| 3.12 | Fonctions sigmoïdales . . . . .   | 30 |
| 3.13 | Fonction de décision non linéaire obtenue à l'aide d'un perceptron multicouche                          | 31 |
| 4.1  | Quelques photos de la base d'images COHN-KANADE . . . . .   | 34 |
| 4.2  | Quelques photos de la base d'images JAFFE . . . . .   | 35 |
| 4.3  | Quelques photos de la base d'images MUG . . . . .   | 35 |
| 4.4  | Les unités d'action les plus importantes du système FACS avec leurs muscles associés . . . . .          | 37 |

---

**TABLE DES FIGURES**

|      |   |    |
|------|---|----|
| 4.5  | Liste des expressions faciales avec les unités d'action correspondantes et leur représentation visuelle . . . . . | 38 |
| 4.6  | Exemple d'un modèle de forme active ASM . . . . .   | 39 |
| 4.7  | Les points faciaux des sourcils . . . . .   | 40 |
| 4.8  | Les points faciaux des yeux . . . . .   | 40 |
| 4.9  | Les points faciaux de la bouche . . . . .   | 40 |
| 4.10 | Points d'intérêt du modèle de forme active . . . . .  | 42 |
| 4.11 | Adaptation du masque du modèle active d'apparence à une expression faciale . . . . .                              | 42 |
| 4.12 | Les points d'intérêts sur un visage ainsi que les distances choisies . . . . .                                    | 43 |
| 4.13 | Les points d'intérêt sur un visage ainsi que les distances choisies . . . . .                                     | 45 |
| 4.14 | Les différentes combinaisons de mouvements faciaux par rapport aux expressions étudiées . . . . .                 | 46 |
| 4.15 | Liste des rapports de distances . . . . .   | 46 |
| 4.16 | Tableau des entrées . . . . .   | 47 |
| 5.1  | Vue générale sur notre IHM . . . . .  | 52 |
| 5.2  | Importation de l'image . . . . .  | 52 |
| 5.3  | Image importée . . . . .  | 53 |
| 5.4  | Détection des points d'intérêt . . . . .  | 53 |
| 5.5  | Classification de l'expression faciale reconnue . . . . .   | 54 |
| 5.6  | Exemple sur l'expression faciale 'surprise' . . . . .   | 55 |
| 5.7  | Exemple sur l'expression faciale 'neutre' . . . . .   | 55 |
| 5.8  | Exemple sur l'expression faciale 'peur' . . . . .   | 56 |
| 5.9  | Résultat d'apprentissage sur les sept expressions de base . . . . .   | 57 |
| 5.10 | Performance ainsi que la régression de la colère . . . . .  | 57 |
| 5.11 | Performance ainsi que la régression du dégoût . . . . .   | 58 |
| 5.12 | Performance ainsi que la régression de la joie . . . . .  | 58 |
| 5.13 | Performance ainsi que la régression de l'expression neutre . . . . .  | 59 |
| 5.14 | Performance ainsi que la régression de la peur . . . . .  | 59 |
| 5.15 | Performance ainsi que la régression de la surprise . . . . .  | 60 |
| 5.16 | Performance ainsi que la régression de la tristesse . . . . .   | 60 |
| 5.17 | Performance et régression entre la surprise et la colère . . . . .  | 61 |
| 5.18 | Performance et régression entre le dégoût et la colère . . . . .  | 61 |
| 5.19 | Performance et régression des cinq expressions faciales . . . . .   | 62 |
| 5.20 | Représentation du taux de reconnaissance pour les différentes expressions .                                       | 63 |

## Résumé

**L**a reconnaissance d'émotions est une thématique importante aux frontières de la psychologie et de la vision par ordinateur. Ses applications sont aussi importantes que variées, allant de l'animation de visage réaliste à l'amélioration de la communication homme-machine, en passant par les robots répondants aux émotions des utilisateurs.

Dans cette optique, l'approche proposée se base sur une étude physiologique des muscles du visage responsables des expressions faciales. Nous commençons par la détection du visage puis passons à l'extraction d'un ensemble réduit et significatif de caractéristiques qui nous permettraient de déduire des données suffisantes pour en faire un apprentissage automatique. Celui-ci est fait à l'aide d'un réseau de neurones. L'implémentation ainsi réalisée permet de reconnaître certaines des expressions faciales qui existent et que nous avons étudiées dans ce projet.

**Mots-clés :** *Expressions faciales, Émotions, réseaux de neurones, Vision par ordinateur, FACS, Intelligence artificielle.*

## Abstract

The recognition of feelings is an important research area which lies at the border between Psychology and Computer Vision. Its applications are as important as varied, going from the animation of realistic faces to the improvement of human-machine communication, such as robots that respond to the user's feelings.

To this end, the proposed approach is based on a physiological study of the muscles that are responsible for facial expressions. Our solution starts with the detection of the face in a picture and then extracts a set of reduced though significant characteristics which would make it possible to infer sufficient information to enable a machine learning to be done. The latter is performed through the use of a neural network. Our implementation of this approach makes it possible to recognize some of facial expressions that we have selected for our study.

**Keywords :** *Facial expressions, Emotions, Neural networks, Vision by computer, FACS, Artificial intelligence.*

# Introduction Générale

## Sommaire

---

|     |                         |   |
|-----|-------------------------|---|
| 0.1 | Introduction            | 1 |
| 0.2 | Objectifs de travail    | 2 |
| 0.3 | Organisation du mémoire | 2 |

---

## 0.1 Introduction

Étudiée par de nombreux chercheurs depuis le 19ème siècle tel que Charles Darwin et Duchenne de Boulogne, l'expression faciale a joué un rôle majeur dans la recherche sur les émotions avec les travaux de Sylvan Tomkins dans les années 1960. Ses élèves Paul Ekman et Caroll Izard ont prouvé qu'il existait un nombre limité d'émotions de base auxquelles sont associées des expressions faciales automatiques, universelles et innées.

De nos jours, La reconnaissance des expressions faciales est un thème de recherche de prédilection dans différents domaines incluant la reconnaissance et l'animation des visages, l'interaction homme machine ou l'interaction sociale.

Auparavant réservée à la recherche psychologique, la reconnaissance d'émotions s'est ouverte à la recherche en vision par ordinateur avec l'arrivée d'outils comme la détection et le suivi de visages, ainsi qu'une puissance de calcul croissante. Les applications envisagées recouvrent plusieurs domaines de l'analyse d'images pour la robotique [6] et les interfaces hommes-machines (IHM) [7, 8, 9], à l'analyse de visages en médecine [10] ou pour des études psychologiques, en passant par la compression d'images de visages, la réalité augmentée ou l'animation de visages virtuels [11].

Pourtant le domaine de la reconnaissance d'expressions faciales n'est qu'à ses débuts, et il reste beaucoup à faire pour arriver à des résultats aussi satisfaisants que dans le domaine de la reconnaissance faciale par exemple, ce retard est dû à plusieurs raisons, parmi elles :

- Le manque de bases d'images précises et complètes qui serviraient de ressources pour une méthode de classification comme c'est souvent le cas dans la reconnaissance faciale.

## CHAPITRE 0. ABSTRACT

- Les techniques utilisées négligent souvent l'aspect physiologique et psychologique des expressions faciales.
- Les méthodes de reconnaissance d'expressions faciales négligent aussi dans leur recherche l'intérêt de l'interprétation humaine par souci d'objectivité, ce qui leur fait pourtant perdre une donnée précieuse.

### **0.2 Objectifs de travail**

Le but de notre projet est de proposer une méthode qui puisse combler les quelques lacunes citées plus haut, ceci en passant par les objectifs suivants :

- Création d'une base d'images unique à partir de toutes les autres bases disponibles dans le monde, qui sera aussi diversifiée que précise.
- Utiliser la détection humaine des expressions faciales par la méthode des juges, comme dans les études psychologiques, pour raffiner les données d'images au maximum.
- Utiliser une approche purement physiologique basée sur les études de Paul Ekman et son système de codification des muscles faciaux [12].
- Réduire les données d'apprentissage au maximum en utilisant des techniques d'extraction de caractéristiques pour palier au contraintes techniques imposant la limitation de l'espace mémoire requis ainsi que le temps de calcul.

### **0.3 Organisation du mémoire**

Nous présenterons dans la première partie un état de l'art détaillé sur les expressions faciales avec les définitions importantes et les techniques de mesure de ces expressions.

Nous exposerons dans la deuxième partie les différentes techniques d'analyse et de reconnaissance d'expressions faciales par apprentissage.

La troisième partie traitera les méthodes de classification et tout particulièrement les réseaux de neurones artificiels.

La quatrième partie sera consacrée à la conception et à la réalisation de notre approche, de la création de la base d'images à la technique d'apprentissage.

Enfin, dans la cinquième et dernière partie, nous exposerons nos résultats expérimentaux et notre méthode d'implémentation.

# Chapitre 1

## Les expressions faciales

### Sommaire

|       |                                 |    |
|-------|---------------------------------|----|
| 1.1   | Introduction                    | 3  |
| 1.2   | Définitions                     | 3  |
| 1.2.1 | Expression faciale              | 3  |
| 1.2.2 | Émotion                         | 4  |
| 1.2.3 | Mimique faciale                 | 5  |
| 1.3   | Physiologie faciale             | 5  |
| 1.4   | Mesure des expressions faciales | 7  |
| 1.4.1 | Interprétation humaine          | 7  |
| 1.4.2 | Interprétation par symbole      | 7  |
| 1.5   | Conclusion                      | 11 |

### 1.1 Introduction

Les expressions faciales jouent un rôle clé dans la plupart des aspects du comportement et de la communication humaine non-verbale.

Dans ce chapitre, nous présenterons quelques définitions concernant les expressions faciales, ainsi, écarter toutes ambiguïtés sur le sujet avant de passer aux particularités physiologiques du visage humain, et terminer par une présentation des différentes techniques de mesure des expressions faciales.

### 1.2 Définitions

Afin d'éviter toute confusion, commençons par définir ce qu'est une expression faciale, une émotion et une mimique faciale.

#### 1.2.1 Expression faciale

Une expression faciale est un changement dans le visage perceptible visuellement et qui est dû à l'activation (contraction), volontaire ou pas, de l'un ou de plusieurs des 44 muscles composant le visage. Ce changement peut être interprété par une émotion ou un signe (Figure 1.1). Des recherches ont montré qu'il existait plus de 250 000 expressions faciales possibles [12]. Ces expressions faciales peuvent être vides de sens comme elles

---

## CHAPITRE 1. LES EXPRESSIONS FACIALES

peuvent être significatives, auquel cas, elles sont soit dues à un déclencheur émotionnel, sinon pour véhiculer une information.



FIGURE 1.1 – Quelques expressions faciales

### 1.2.2 Émotion

Une émotion est un processus chimique corporel, activé par des stimuli internes ou externes ; elle se traduit souvent par le langage non verbal correspondant, comme la position du corps, de la tête et des expressions faciales qui varient d'intensité selon le degré de l'émotion et le contrôle de l'individu sur cette dernière (voir Figure 1.2) [2, 3] . Une émotion est l'un des déclencheurs principaux d'une expression faciale.

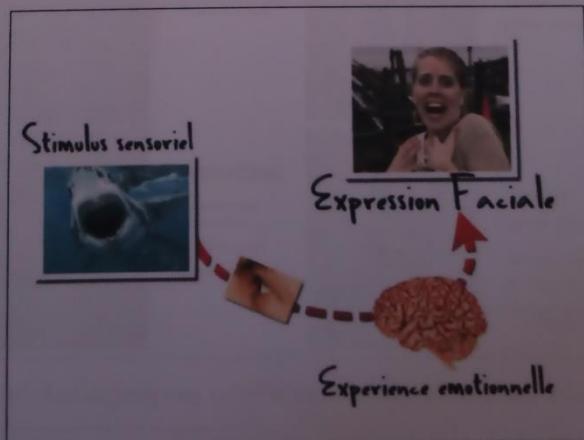


FIGURE 1.2 – Théorie de Cannon-Bard sur le déclenchement des émotions [2, 3]

Les muscles de la zone supérieure du visage<sup>2</sup> ont très peu d'influence sur les muscles de la zone inférieure du visage<sup>3</sup> et vice versa.

Les expressions qui donnent une indication sur l'émotion réelle de la personne sont dites « *spontanées* », alors que les expressions sans émotion sont dites expressions « *forcées* ». Les deux types d'expressions, bien que pouvant véhiculer le même sens, sont activées par des régions différentes du cerveau et n'activent pas les mêmes muscles.

Par exemple, il existe deux types de sourires : un sourire « *forcé* » qui est activé par le mouvement du zygomaticus majeur vers le haut, et un sourire « *spontané* » ou « *authentique* » qui active le zygomaticus majeur mais aussi l'orbicularis-oculi qui est un muscle autour des yeux et qu'on ne peut activer volontairement. Ce dernier sourire est appelé « *sourire de Duchenne* » en hommage à Guillaume Duchenne De Boulogne qui a découvert cette distinction.

Notons enfin qu'il existe un certain nombre d'émotions représentées par des expressions universelles et innées, reconnaissables par tous, indépendamment des cultures sociales (Figure 1.4). Ces expressions sont la joie, la tristesse, la colère, le dégoût, la surprise, la peur et bien évidemment, l'expression neutre [4].

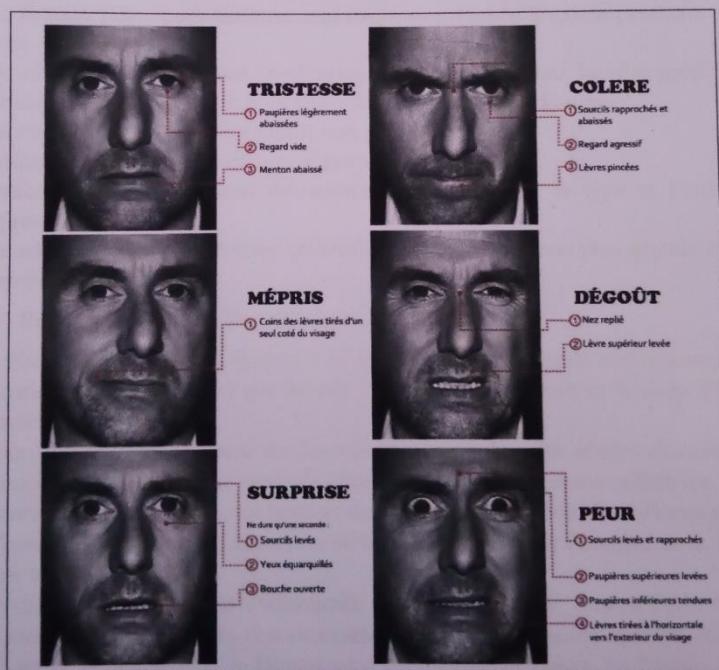


FIGURE 1.4 – Les expressions faciales universelles définies par Paul Ekman [4]

- 2. Des yeux et du front
- 3. Bouche et joues

## **1.4 Mesure des expressions faciales**

Il existe deux sortes d'interprétations ; humaine et celle par symbole, nous détaillerons chacun des types dans la partie ci-après.

### **1.4.1 Interprétation humaine**

L'interprétation humaine consiste en une classification des expressions en 7 catégories par des experts et non-experts telles que les a définies Paul Ekman [4]. Cette méthode bien que fiable, facile à appliquer et très répandue, présente des désavantages considérables, surtout au niveau du timing et du degré d'intensité des expressions. Par exemple, si l'expression est rapide, les experts humains peuvent la rater ou la confondre, ce qui est fréquent dans le cas des micro-expressions<sup>4</sup>.

### **1.4.2 Interprétation par symbole**

L'interprétation par symboles est une approche qui a l'avantage de séparer la reconnaissance et l'interprétation des expressions faciales, et d'être plus objective car elle se concentre sur des unités faciales spécifiques au lieu d'une interprétation globale du visage.

Selon Ekman [13] , pour mesurer une expression faciale il existe plusieurs critères à prendre en compte :

- une référence externe pour l'évaluation de certaines actions faciales spécifiques, car certaines expressions sont plus aisées à reconnaître que d'autres ;
- des expressions spontanées plutôt que posées ;
- l'utilisation d'experts et de non-experts pour la mesure ;
- l'évaluation de la précision des mesures à la fois pour le type et l'intensité des expressions.

Une approche par symbole est donc préférable pour garantir une plus grande objectivité des interprétations.

#### **1.4.2.1 Facial Action Coding System (FACS) :**

Le « *Facial Action Coding System* » (FACS) est un système de codification manuelle des expressions faciales, inventé par les psychologues Paul Ekman et Wallace Friesen à la fin des années 1970 [12].

C'est désormais le principal outil de description utilisé dans les études s'intéressant aux expressions faciales, comme par exemple dans le domaine des sciences affectives. Cette nomenclature s'inspire elle-même des travaux de l'anatomiste suédois, Carl-Herman Hjortsjo, publiés dans son ouvrage sur l'imitation faciale, « *Le visage de l'homme et le langage de l'imitation* » [14].

Les mouvements du visage sont décomposés en unités d'action (*Action Unit, AU*). Nous pouvons ainsi coder l'ensemble des expressions faciales existantes avec 64 AUs qui décrivent chacune le mouvement indépendant d'un muscle du visage. La Figure 1.5 illustre quelques AUs, la liste complète est donnée en Annexe.

4. Mouvement facial d'une durée inférieure à 5 milliseconde.



FIGURE 1.5 – Les « *Action Unit* » et leurs muscles associés

Le système FACS est un index d'expressions faciales et les intensités des AUs sont notées par des lettres de 'A' à 'E'. 'A' étant le degré le plus faible, et 'E' l'intensité maximale pour l'Action Unit. Les degrés sont les suivants :

- **A** : Trace
- **B** : Léger
- **C** : Marqué/Prononcé
- **D** : Sévère
- **E** : Maximum

#### 1.4.2.2 Facial Action Coding System<sup>+</sup> (*FACS*<sup>+</sup>) :

*FACS*<sup>+</sup> est une méthode qui vient combler les quelques inconvénients existants dans FACS [15] tels que :

1. **la Complexité :**

On estime qu'il faut 100 heures d'apprentissage pour maîtriser les principaux concepts de FACS [16].

2. **la Difficulté de manipulation par une machine :**

FACS a d'abord été créé pour des psychologues ; certaines mesures restent floues et difficilement évaluables par une machine.

3. **Le Manque de précision :**

Les transitions entre deux états d'un muscle sont représentées de manière linéaire, ce qui est une approximation de la réalité. En particulier, les mesures temporelles de l'activation des muscles faciaux (*onset*, *apex* et *offset*) ne sont pas mises en évidence.

*FACS<sup>+</sup>* utilise un modèle musculaire plus facilement codifiable par ordinateur mais la principale critique de cette approche [16] est qu'elle manque toujours de maturité comparée à FACS qui a nécessité plus d'une vingtaine d'années de développement, mais aussi parce que *FACS<sup>+</sup>* est difficilement interprétable par un être humain.

**1.4.2.3 La norme MPEG-4**

La norme de codage MPEG-4[17] est un modèle articulé du visage humain en 3D, développé par le groupe « *Face and Body AdHoc Group* » qui l'a construit sur un ensemble d'attributs faciaux « *Facial Feature Points* » (FFP). Les mouvements du modèle reposent sur des mesures effectuées sur les FFP appelés « *Facial Animation Parameter Units* » (FAPU). Ces distances permettent de définir les mouvements naturels et élémentaires du visage.

Cependant, il est impossible de décrire les mouvements faciaux de manière absolue à cause de la grande différence entre les attributs faciaux des individus. Nous pouvons cependant utiliser des mesures relatives qui ne changent pas quel que soit le visage choisi (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...)

Cependant, MPEG-4 n'est pas comme FACS un modèle basé sur les muscles du visage, ce qui peut créer une source d'incohérence en termes de réalisme. Par exemple, l'Action Unit 26 (AU 26) du modèle FACS est l'abaissement du menton ce qui provoque l'ouverture de la bouche et l'abaissement de la lèvre inférieure. Pourtant, dans MPEG-4, cette action (FAP 3 - open-jaw) ne s'accompagne pas d'un abaissement de la lèvre, ce qui est un non-sens d'un point de vue musculaire. Les points caractéristiques du visage MPEG-4 sont illustrés dans la Figure 1.6.

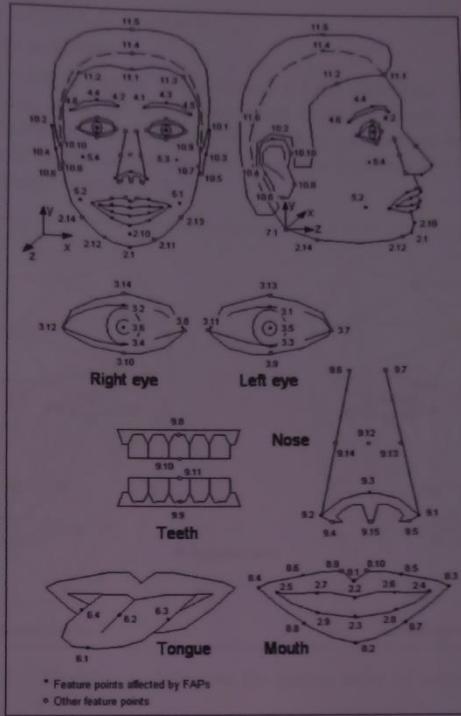


FIGURE 1.6 – Squelette du MPEG-4 avec les Facial Feature Points (FFP) associés

#### 1.4.2.4 Candide

Candide [5] est un modèle de fil de fer qui prend en compte deux paramètres :

- « *Shape Unit* » (unité de la forme) : il représente les différences interindividuelles qui changent d'une personne à une autre et qui sont :

1. hauteur de la tête,
  2. position verticale des sourcils,
  3. position verticale des yeux,
  4. largeur des yeux,
  5. hauteur des yeux,
  6. distance de séparation des yeux,
  7. profondeur des joues,
  8. profondeur du nez,
  9. position verticale du nez,
  10. degré de courbure du nez (s'il pointe vers le haut ou non),

11. position verticale de la bouche,
  12. largeur de la bouche.
- « *Animation Unit* » (unité d'animation) : Elle représente les différences intra-individu, ce paramètre est une composition entre les unités d'action de FACS et les mesures FAPU du modèle MPEG-4. Ces deux paramètres seront représentés sur un modèle de fil de fer comme une liste de points à mettre à jour selon l'individu et l'animation choisie (Figure 1.7).

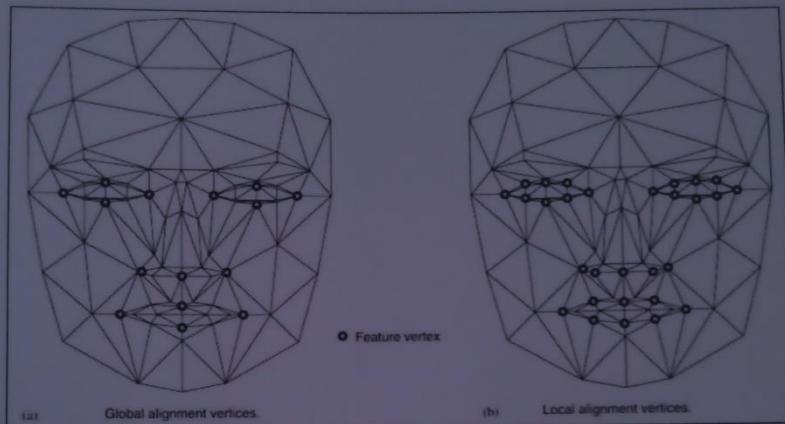


FIGURE 1.7 – Modèle candide avec les points inter et intra individu.[5]

#### 1.4.2.5 Le modèle Sign Stream [1]

SignStream [1], est un modèle développé par le « *National Center for Sign Language and Gesture Resources* » pour qu'il soit utilisé par les linguistes dans la codification du langage américain des signes. Un sous-ensemble du système de notation est dédié au codage des expressions du visage rencontrées en langage des signes.

Nous pouvons noter que le langage des signes donne une grande importance à des expressions faciales peu abordées dans les autres modèles, comme par exemple la présence de la langue entre les dents qui n'est pas représentable dans candide et n'incluant pas les mouvements de la langue.

### 1.5 Conclusion

Dans ce chapitre nous montrons d'un côté la grande diversité du domaine des expressions faciales et de l'autre le large éventail des techniques et des méthodes possibles.

Malgré toutes les tentatives de développer de meilleurs modèles, le modèle FACS reste le plus répandu et le plus fiable en terme de précision. Nous l'utiliserons plus tard pour créer un nouveau modèle d'interprétation des expressions faciales, mais aussi pour sélectionner les éléments de la base d'images. Nous nous baserons aussi sur l'approche de

---

## *CHAPITRE 1. LES EXPRESSIONS FACIALES*

l'interprétation humaine et réaliste afin d'en bénéficier des avantages.

Nous verrons dans le prochain chapitre toutes les étapes d'une analyse des expressions faciales humaines et automatique.

# Chapitre 2

## Analyses et classification des expressions faciales par apprentissage

### Sommaire

|       |  |    |
|-------|--|----|
| 2.1   | Introduction                               | 13 |
| 2.2   | Historique                                 | 13 |
| 2.3   | Analyse humaine des expressions faciales   | 15 |
| 2.4   | Analyse automatique d'expressions faciales | 15 |
| 2.4.1 | Détection du visage                        | 16 |
| 2.4.2 | Extraction des caractéristiques            | 17 |
| 2.4.3 | Classification des expressions             | 18 |
| 2.5   | Conclusion                                 | 19 |

### 2.1 Introduction

Nous présenterons dans ce chapitre, après un bref historique des recherches effectuées en psychologie sur les émotions, tous les détails concernant l'analyse des expressions faciales. D'abord d'un point de vue humain, pour ensuite passer à une analyse automatique par ordinateur, et enfin, nous expliquerons les méthodes de classification et d'interprétation des expressions faciales.

### 2.2 Historique

1860 : Guillaume Duchenne de Boulogne est un psychologue français (Figure ??), qui a entamé les premières recherches sérieuses sur les expressions faciales et le mouvement indépendant des muscles du visage. Il avait trouvé un cobaye souffrant d'insensibilité musculaire au visage ce qui lui a permis à l'aide de stimuli électriques d'activer les muscles faciaux indépendamment et sur des périodes suffisantes et ainsi offrir au monde les premières images de mimiques et d'expressions faciales, il a en particulier étudier les muscles du « faux » et du « vrai sourire » appelé aussi *sourire de Duchenne*.

### 1.2.3 Mimique faciale

Une mimique faciale est une combinaison de muscles faciaux contractés, qui n'est pas forcément chargée de sens. Par exemple, l'activation du muscle *zygomatique majeur*<sup>1</sup> avec les paupières fermées est une mimique faciale.

À partir des définitions précédentes nous pouvons conclure ce qui suit :

- Une expression faciale est une mimique faciale, alors que le contraire n'est pas forcément juste.
- Une émotion peut causer une expression faciale, mais une expression faciale peut avoir été causée par une émotion ou autres stimuli externes et internes (i.e elle peut être déclenchée par autre qu'une émotion, Figure 1.3).

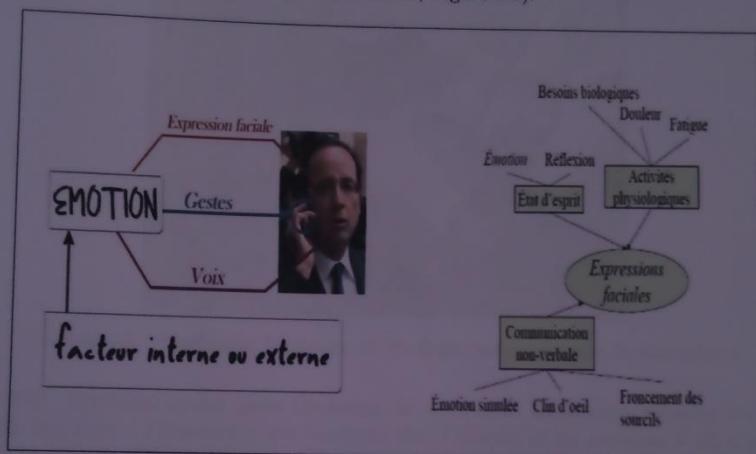


FIGURE 1.3 – Déclencheurs d'expressions faciales

## 1.3 Physiologie faciale

Avant de parler de la mesure des expressions faciales, nous devons décrire les particularités du visage humain et de son anatomie ainsi que le lien entre le visage et les émotions.

Le visage est la partie la plus importante du corps en termes de communication et d'interaction sociale. Il est composé de 44 muscles qui peuvent être activés séparément (selon les individus et leur degré d'entraînement).

L'activation d'un muscle facial passe par trois phases :

- la phase « **ATTACK** » où l'activité musculaire devient de plus en plus importante ;
- la phase « **SUSTAIN** » qui correspond à la période de soutien de l'activité musculaire ; et
- la phase « **RELAXATION** » où le muscle reprend sa position initiale.

1. Soulèvement des coins de la bouche

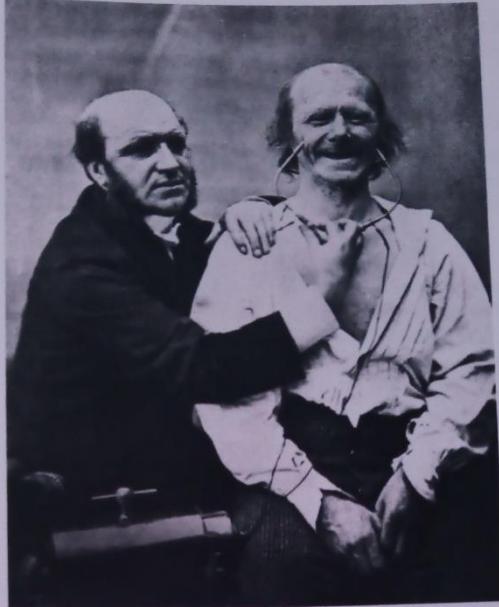


FIGURE 2.1 – Guillaume Duchenne de Boulogne en compagnie de son cobaye

1872 : Quelques années après Duchenne, le naturaliste britannique Charles Darwin publie son livre « *l'Expression des émotions chez l'homme et les animaux* » où il parle pour la première fois d'universalité des expressions faciales, son principal argument est la présence des mêmes expressions chez les nouveaux nés et les adultes; ainsi, même les enfants sourds et aveugles sourient. Cette théorie universaliste est, pour Darwin, la preuve que les expressions faciales sont un instinct de survie chez tous les êtres humains.

1960 : Paule EKMAN est un psychologue qui a consacré toute sa vie aux études sur les expressions faciales. Voulant tout d'abord, prouver la théorie de Darwin sur l'universalité des expressions faciales, il parcourut le monde pour récolter des centaines de clichés de différents peuples et ethnies et notamment des tribus africaines isolées du monde extérieur (indigènes). De cette recherche, il valida l'existence des six expressions universelles : la peur, la tristesse, la joie, la colère, le dégoût et la surprise. Il publia avec Wallace Friesen à la fin des années 1970, son nouveau système de mesure et d'analyse des expressions faciales FACS, [18]. Depuis, Paule EKMAN qui a été consultant dans de nombreuses recherches notamment pour la sécurité et la criminalité est, jusqu'à nos jours, considéré comme la référence mondiale en termes d'analyse des expressions faciales.

1968 : Mehrabian [19] révèle l'importance du non-verbal et plus spécialement des expressions faciales dans la communication. Sa recherche démontre que le contenu textuel ne

**CHAPITRE 2. ANALYSES ET CLASSIFICATION DES EXPRESSIONS FACIALES  
PAR APPRENTISSAGE**

contribue uniquement que de 7% dans l'interaction humaine, le reste étant du domaine du non verbal. Par conséquent, l'expression faciale est considérée depuis comme une modalité essentielle de la communication.

### 2.3 Analyse humaine des expressions faciales

Avant d'entamer l'analyse automatique, intéressons-nous à quelques notions sur l'analyse du visage et des expressions chez l'être humain qui ont servi à développer le modèle automatique. La plupart de ces informations sont dues à l'étude d'une maladie neurologique : *la prosopagnosie*.

Les prosopagnosiques ne reconnaissent pas ce qui fait l'identité d'un visage :

- Les humains mémorisent les visages de leurs semblables plus que les objets du quotidien.
- La reconnaissance des expressions et du visage sont des processus innés car les nouveaux nés préfèrent suivre du regard les visages plutôt qu'autre chose.
- La reconnaissance du visage et celle des expressions sont des processus indépendants et parallèles.
- Déetecter un visage utilise ce qui est commun à tous les visages, en ignorant donc les différences inter-individus.
- Le processus de reconnaissance des expressions se base sur les différences intra-individus [20].
- Reconnaître une expression fait appel à une analyse par composantes avec une indication sur la forme des composantes du visage. Ainsi, un ensemble de points représentants la position de chaque composante du visage suffit pour qu'un humain reconnaisse une expression.

### 2.4 Analyse automatique d'expressions faciales

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

- **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.
- **Approche par composantes** : aussi appelée « *feature-based* » se base sur les différentes composantes du visage (yeux, nez, bouche, ...) indépendamment les uns des autres et qui forment par la suite un tout (le visage).

Nous allons par la suite parcourir l'ensemble des étapes de l'analyse automatique des expressions faciales (Figure 2.2) tout en expliquant les deux techniques pour chacune d'elles.

**CHAPITRE 2. ANALYSES ET CLASSIFICATION DES EXPRESSIONS FACIALES  
PAR APPRENTISSAGE**

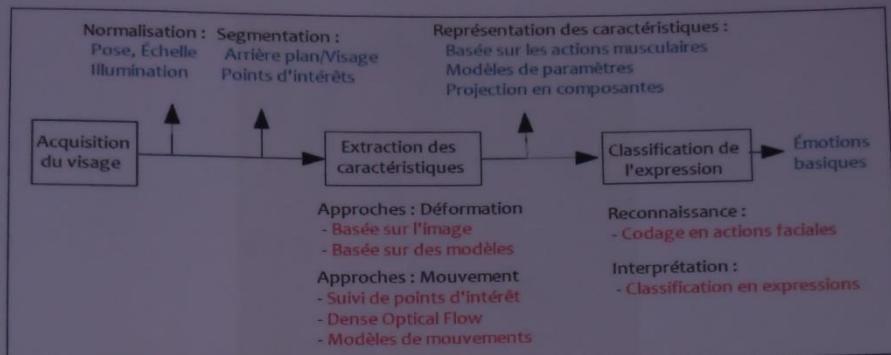


FIGURE 2.2 – Processus d'analyse et de classification d'expressions faciales

#### 2.4.1 Détection du visage

La détection du visage est un des problèmes les plus récurrents en analyse faciale qui a été traité par de multiples approches [21, 22]. Le plus souvent, la méthode utilisée est de prendre des sous-images de l'image initiale et de leur appliquer un détecteur peu robuste (généralement basé sur la couleur de la peau ou les contours du visage). Cette 1ère phase a pour but de trouver une région approximative de la position du visage sans dépasser les limites critiques en temps de calcul, La 2ème phase consiste en une analyse plus robuste de la région en question.

La détection peut être globale « *image-based* » en comparant des sous-images avec un modèle de visage construit automatiquement à partir d'un ensemble de visages d'apprentissage. Dans ce cas, une mesure d'erreur entre le modèle et le visage observé permet d'avoir une idée de la ressemblance. Elle peut aussi être par composantes « *feature-based* » dans ce cas, la détection se fait par identification des zones qui contiennent un ensemble de composantes faciales réparties spatialement d'une façon particulière à former un visage (Figure 2.3).

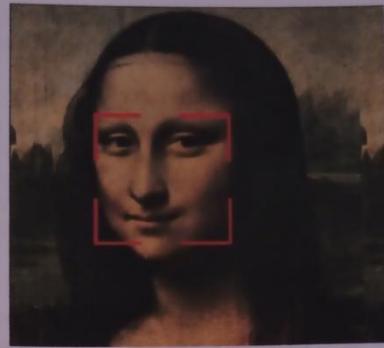


FIGURE 2.3 – Détection des visages

#### 2.4.2 Extraction des caractéristiques

La reconnaissance d'expressions et l'extraction de caractéristiques est l'étape la plus critique dans l'analyse des expressions. Il est possible aussi d'utiliser ici la reconnaissance d'expressions par approche globale en utilisant un modèle appris à partir d'un certain nombre d'exemples. Pourtant, cette méthode est peu utilisée en raison de déficits en terme d'apprentissage. En effet, le corpus doit être très diversifié et doit représenter toutes les combinaisons possibles des muscles faciaux ce qui n'est pas facilement applicable.

La 2ème alternative (*feature-based*) est donc la plus judicieuse. En appliquant un ensemble d'opérateurs spécialisés dans la détection et la mesure des caractéristiques d'un composant particulier du visage (Figure 2.4), nous pouvons le détecter et en extraire les données souhaitées.

Voici quelques caractéristiques des principales composantes du visage essentielles à l'analyse des expressions faciales :

##### 1. Sourcils :

Les sourcils (s'ils sont présents) sont généralement plus foncés que le reste du visage et sont plus longs que larges. Ainsi un opérateur basé sur le gradient de l'image peut être un opérateur adéquat pour cette partie du visage.

##### 2. Yeux :

Les yeux (s'ils ne sont pas fermés) sont formés d'une zone sphérique (l'iris) entourée d'une zone très claire (la sclérotique). Un détecteur de zones claires peut former la base d'un opérateur efficace de détection des yeux [23]. On peut ainsi faire varier un seuil de binarisation jusqu'à obtenir deux formes distinctes. En se basant sur des critères morpho-métriques on peut donc détecter la position des yeux.

La forme ronde de l'œil peut aussi servir à la détection, par exemple par une transformée de Hough.

##### 3. Paupières :

Les paupières ont deux particularités : Elles ont la couleur de la peau et sont à

CHAPITRE 2. ANALYSES ET CLASSIFICATION DES EXPRESSIONS FACIALES  
PAR APPRENTISSAGE

l'endroit des yeux, il est aussi possible de détecter les yeux en se basant sur le principe qu'ils clignent avec une certaine fréquence et tous les deux en même temps.

4. Bouche :

La bouche, étant liée à la mâchoire a une grande liberté de mouvement et peut donc avoir des formes très différentes surtout en terme d'expressions faciales. La seule possibilité de la localiser est de détecter les lèvres mais aussi les dents et la langue. La plupart du temps, la technique utilisée est celle des contours actifs, le problème avec cette technique est sa difficulté à être initialisée.

Une approche [24] serait de détecter la commissure de la bouche qu'elle soit ouverte ou fermée.

5. Rides :

Les rides du visage sont significatives dans quelques recherches sur l'analyse des expressions faciales. Elles peuvent être générées par l'âge ou l'activation de muscles comme pour la joie sur les joues ou le dégoût sur le nez. Ce relâchement de la peau est caractérisé par une différence de luminosité qui peut être repérée par un détecteur de contour dans les zones susceptibles d'avoir des rides.



FIGURE 2.4 – Détection de certaines zones d'intérêt du visage.

#### 2.4.3 Classification des expressions

La dernière étape de l'analyse est la classification. Nous pouvons ici distinguer la reconnaissance des expressions qui est le codage en action faciale, en utilisant un système de règles ou des dictionnaires d'expressions faciales pour retranscrire les actions faciales en catégories d'émotions [25, 26] et l'interprétation qui est la classification en terme

d'expression [27, 15]

On distingue souvent deux approches de classification : spatiale et spatio-temporelle :

#### 2.4.3.1 Modèle spatio-temporel

Dans les approches spatio-temporelles (variables sur le temps) on utilise couramment les modèles de Markov cachés (*HMMs, Hidden Markov Models*) de par leur efficacité et leur aspect dynamique qui ont déjà fait leurs preuves dans la reconnaissance vocale [28, 29].

#### 2.4.3.2 Modèle spatial

Dans les modèles spatiaux on utilise généralement un système d'apprentissage neuro-nal tel que des réseaux de neurones récurrents [30, 31], soit directement sur des images de visage, soit combiné avec des techniques d'extraction de caractéristiques comme l'analyse en composante principale [32], ou les filtres à base d'ondelettes de Gabor [33, 34, 35]. Ces dernières méthodes permettent une grande réduction dimensionnelle, ce qui simplifie la classification. L'inconvénient majeur des réseaux de neurones est leur apprentissage, qui est très important lorsqu'on tente de reconnaître une expression parmi les milliers de combinaisons que le système FACS offre. Plus récemment, plusieurs méthodes à base de réseaux bayésiens [36, 32], de machines à vecteurs de support (*SVM, Support Vector Machine*) [37, 38] offrent des résultats appréciables et généralement similaires.

Les résultats sont souvent combinés par une méthode de boosting comme **Adaboost**, diminutif d'*Adaptive Boosting* [39]. Un algorithme d'apprentissage pouvant être utilisé en conjonction avec d'autres algorithmes pour améliorer leurs performances. **AdaBoost** est adaptatif dans le sens que les classificateurs reconstruits sont modifiés au regard des cas mal classifiés par les précédents. **AdaBoost** est sensible au bruit et aux données aberrantes.

Cependant, dans la majorité des cas, il est moins sensible au problème de sur-apprentissage que la plupart des algorithmes d'apprentissage.

## 2.5 Conclusion

Dans ce chapitre, nous avons vu les différentes étapes de l'analyse des expressions faciales, de la détection du visage à la classification en passant par l'extraction de données selon des méthodes globales et locales. Il reste aujourd'hui difficile de faire une comparaison entre les différentes techniques, de par leur diversité et le manque de base de données appropriées.

Aucun système actuel ne peut reconnaître toutes les unités d'action (*46 action unit*) définie par Paule Ekman. Mais d'un autre côté, aucune recherche ne prouve que toutes les AU's sont toutes nécessaires pour reconnaître une expression faciale de base. Au contraire, il suffit d'un nombre réduit d'unités d'actions pour distinguer entre les principales émotions universelles.

Dans le chapitre suivant, nous nous intéresserons aux systèmes de classification en général et aux Réseaux de Neurones Artificiels en particulier.

# Chapitre 3

## Classification

### Sommaire

|       |                                       |    |
|-------|---------------------------------------|----|
| 3.1   | Introduction                          | 20 |
| 3.2   | État de l'art sur les classificateurs | 21 |
| 3.2.1 | La fonction discriminante             | 21 |
| 3.2.2 | La fonction de coût                   | 21 |
| 3.3   | Les classificateurs paramétriques     | 22 |
| 3.3.1 | Classificateur Euclidien              | 22 |
| 3.3.2 | Classificateur Quadratique            | 23 |
| 3.4   | Les Classificateurs Non Paramétriques | 24 |
| 3.4.1 | Le Plus Proche Voisin                 | 24 |
| 3.4.2 | Les k plus Proches Voisins            | 25 |
| 3.5   | Classificateurs Neuronaux             | 25 |
| 3.5.1 | Les neurones biologiques              | 26 |
| 3.5.2 | Le Perceptron (1958)                  | 27 |
| 3.5.3 | Le modèle formel 1987                 | 28 |
| 3.5.4 | La Carte Auto-Organisatrice 1990      | 29 |
| 3.5.5 | Le Perceptron Multicouche             | 30 |
| 3.6   | Conclusion                            | 31 |

### 3.1 Introduction

L'essentiel de ce chapitre est inspiré de la thèse de doctorat de Bernard GOSSELIN sur « *L'application des réseaux de neurones artificiels à la reconnaissance automatique des caractères manuscrits* » [40], et de Yves Rangoni sur « *Les réseaux de neurones dynamiques* » [41].

Nous parlerons, dans un premier temps, des classificateurs en général et détaillerons les classificateurs les plus utilisés par la suite.

### 3.2 État de l'art sur les classificateurs

Soit un ensemble d'objets exprimés par des vecteurs de caractéristiques  $X = [x_1, x_2, x_3 \dots x_c]$  et qui sont représentés chacun par un point dans un espace euclidien  $R_c$ , ils peuvent ainsi être regroupés selon des classes spécifiques.

Un classificateur a pour rôle de classer, parmi un ensemble fini de classes, auxquelles appartient un objet donné, il doit être capable de modéliser au mieux les frontières qui séparent les classes les unes des autres (Figure 3.1).

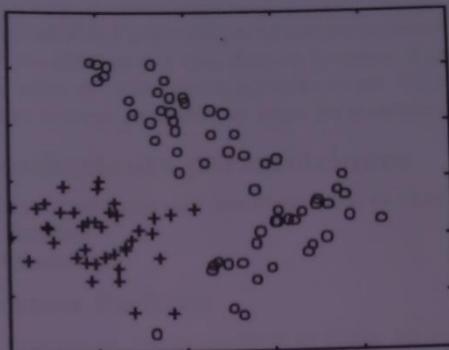


FIGURE 3.1 – Représentation d'objets appartenant à deux classes distinctes dans un espace à deux dimensions.

Une classification inclut généralement deux paramètres :

#### 3.2.1 La fonction discriminante

Cette modélisation fait appel à la notion de fonction discriminante qui permet d'exprimer le critère de classification afin de trouver les frontières optimales entre les classes, de la manière suivante :

"Assigner la classe  $w_i$  à l'objet représenté par le vecteur  $X$ , si et seulement si, la valeur de la fonction discriminante de la classe  $w_i$  est supérieure à celle de la fonction discriminante de n'importe quelle autre classe  $w_i'$ ".

Mathématiquement :

$$X \in \omega_I \Leftrightarrow \phi_I(X) \geq \phi_J(X) \forall J = 1, 2, 3, \dots, C; J \neq I \quad (3.1)$$

$\phi_I(X)$  : la fonction discriminante de  $\omega_I$

$C$  : le nombre total de classes.

#### 3.2.2 La fonction de coût

Le coût est simplement la fonction d'erreur définie par le nombre de classes incorrectes assignées à un objet donné. Une classification est donc dite optimale lorsque le coût

est le plus minimal, et parfaite lorsqu'il est nul. Elle peut être utilisée pour calculer les probabilités à posteriori ainsi [42] :

$$X \in \omega_i \leftrightarrow \pi(\omega_i | X) \geq \pi(\omega_j | X) \forall j = 1, 2, 3, \dots, c; j \neq i \quad (3.2)$$

$\pi(\omega_i | X)$  : La probabilité à posteriori de la classe  $\omega_i$ .

La classe attribuée à l'objet représenté par le vecteur  $X$  est alors celle dont la probabilité étant donné  $X$  est supérieure à la probabilité de n'importe quelle autre classe, étant donné  $X$ .

Le calcul exact des probabilités à posteriori est cependant rarement possible. Des modèles de classificateurs ont été développés à base d'autres fonctions discriminantes que la probabilité à posteriori ; celles qui sont définies entièrement par des paramètres calculables, celles qui ne dépendent d'aucun paramètre, et enfin, les classificateurs neuronaux.

### 3.3 Les classificateurs paramétriques

Les classificateurs paramétriques sont nombreux, nous en citerons deux :

- Classificateur Euclidien
- Classificateur Quadratique

#### 3.3.1 Classificateur Euclidien

Le classificateur euclidien est une façon simple de définir les frontières entre classes, il suffit tout d'abord de calculer les vecteurs de caractéristiques moyen à partir d'un nombre fini de prototypes de chaque classe :

$$\hat{M}_i = \frac{1}{N_i} \sum_{X_k \in \omega_i} (X_k) \quad (3.3)$$

Avec :

- $N_i$  : Le nombre total d'objets de classe  $\omega_i$  qui sont à disposition.
- $X_k$  : Les vecteurs de caractéristiques qui représentent ces objets.

Il s'agit ensuite de calculer la fonction de discrimination qui calcule la distance euclidienne entre l'objet à classer et les vecteurs de caractéristiques, l'objet sera assigné à la classe dont le vecteur est le plus proche :

$$\phi_i(X) = -\frac{1}{2}(X - M_i)^T(X - M_i) \quad (3.4)$$

Avec :

- $M_i$  : le vecteur de caractéristiques moyen des éléments qui appartiennent à la classe  $\omega_i$ .

Les frontières qui séparent les classes dans l'espace  $R_d$  sont ici linéaires (Figure 3.2) :

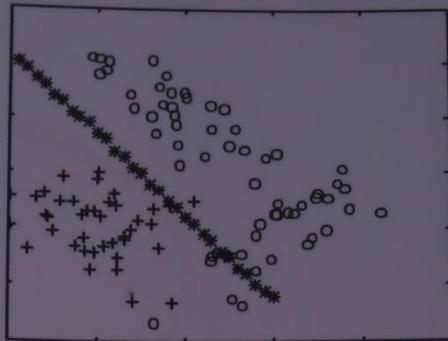


FIGURE 3.2 – Frontière fournie par le classificateur Euclidien dans le cas d'un problème à deux classes

### 3.3.2 Classificateur Quadratique

Comme pour le classificateur euclidien, les classificateurs quadratiques ont besoin de calculer les vecteurs de caractéristiques moyens, de plus, ils utilisent des matrices de covariance, une estimation non biaisée est donnée par [43] :

$$\hat{\Sigma}_i = \frac{1}{N_i - 1} \sum_{X_k \in \omega_i} (X_k - \hat{M}_i)(X_k - \hat{M}_i)^T \quad (3.5)$$

$\hat{\Sigma}_i$  : La matrice de covariance.

$\hat{M}_i$  : Vecteur de caractéristiques moyen.

Par la suite, la matrice de covariance est utilisée dans la fonction discriminante suivante :

$$\phi_i(X) = -\frac{1}{2}(X - M_i)^T \Sigma_i^{-1} (X - M_i) \quad (3.6)$$

Les frontières fournies par ce modèle sont quadratiques (Figure 3.3).

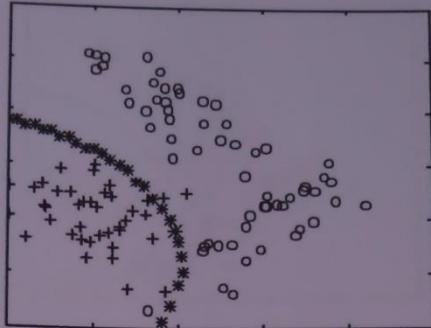


FIGURE 3.3 – Frontière obtenue par le classificateur quadratique

### 3.4 Les Classificateurs Non Paramétriques

Les classificateurs non paramétriques utilisent des techniques de voisinage telles que :

- Le plus proche voisin.
- Le k plus proche voisin.

#### 3.4.1 Le Plus Proche Voisin

La méthode du plus proche voisin est une extrapolation du classificateur euclidien qui remplace le vecteur de caractéristiques moyen par l'utilisation de tous les vecteurs disponibles. La distance euclidienne est calculée pour chacun d'entre eux avec l'objet à classifier, la classe assignée à l'objet est simplement celle du prototype le plus proche de ce dernier (Figure 3.4).

La fonction de discrimination est :

$$\phi_i(X) = \min_{X_k \in \omega_i} (X_k^T X - \frac{1}{2} X_k^T X_k) \quad (3.7)$$

Bien que cette méthode donne des classifications très élaborées, elle reste très coûteuse en terme de temps, de calcul et de consommation de mémoire.

Les frontières obtenues sont contrairement aux classificateurs paramétriques très complexes constituées d'une suite de polygones convexes ou pas, sur une seule ou plusieurs pièces.



FIGURE 3.4 – Frontière fournie par le classificateur du plus proche voisin.

### 3.4.2 Les k plus Proches Voisins

La méthode des  $k$  plus proches voisins vient corriger l'inconvénient majeur de la méthode du plus proche voisin. En effet, cette dernière, très sensible aux abords des frontières entre classes, peut mettre un objet dans une classe incorrecte car son plus proche voisin appartient à cette classe alors que la plupart de ses voisins ne le sont pas, afin de remédier à cela on se basera plutôt sur les  $k$  objets voisins qui vont définir la classe appropriée pour cet objet. La fonction discriminante d'une classe est alors simplement le nombre de prototypes de cette même classe qui se situent parmi les  $k$  plus proches voisins de l'objet à classifier :

$$\phi_i(X) = \sum_{X_j \in \omega_i} |X_j \in \Psi_k(X)| \quad (3.8)$$

## 3.5 Classificateurs Neuronaux

Haykin définit les réseaux de neurones artificiels ainsi [44] :

*« Un réseau de neurones est un processus distribué de manière massivement parallèle, qui a une propension naturelle à mémoriser des connaissances de façon expérimentale et de les rendre disponibles pour utilisation. Il ressemble au cerveau en deux points :*

- La connaissance est acquise au travers d'un processus d'apprentissage.*
- Les poids des connections entre les neurones sont utilisés pour mémoriser la connaissance. »*

La création des réseaux de neurones a largement été influencée par une nouvelle approche en informatique qui se base sur les concepts biologiques du cerveau humain. Ce dernier étant très différent d'un ordinateur, utilise un réseau complexe de structures appelé neurones, plus de 200 milliards, ces centres de calcul microscopiques massivement connectés entre eux (en moyenne dix mille connexions par neurone), se modifient en permanence, les neurones se restructurent selon les besoins du moment pour accomplir des

tâches très élaborées. Une autre particularité du cerveau serait qu'il utiliserait ses neurones d'une manière non linéaire et extrêmement parallèle, c'est vers cette optique que les réseaux de neurones artificiels ont été mis au point.

### 3.5.1 Les neurones biologiques

Du point de vue morphologique, les neurones existent sous des formes variées mais présentent une architecture commune qui se compose de trois régions principales : **le corps cellulaire** qui renferme le noyau, **les dendrites<sup>1</sup>** et **l'axone<sup>2</sup>** (Figure 3.5). Les dendrites ont pour fonction de transmettre au corps cellulaire les influx nerveux en provenance d'autres cellules, tandis que les axones transportent les influx en provenance du corps cellulaire vers d'autres cellules.

Du point de vue fonctionnel [45], le neurone doit être observé comme une entité polarisée car l'information est transmise dans un seul sens des dendrites vers l'axone. Le neurone reçoit par l'intermédiaire de ses dendrites des informations venant d'autres neurones. Toutes ces informations sont sommées au niveau du corps cellulaire. Un potentiel d'action<sup>3</sup> se forme pour être ensuite transmis le long de l'axone jusqu'aux terminaisons synaptiques.

L'influx nerveux (ou l'information) lors de son arrivée, va libérer des neurotransmetteurs<sup>4</sup> dans la fente synaptique grâce aux vésicules synaptiques fusionnant avec la membrane cellulaire.

Les neurotransmetteurs rendent possible le passage du flux nerveux entre les neurones car le signal électrique ne peut pas passer la synapse dans le cas précis d'une synapse chimique. Il existe deux types de neurotransmetteurs (*excitateur ou inhibiteur*), ils excitent ou inhibent le neurone auquel ils sont connectés et par conséquent peuvent autoriser ou interdire la propagation de l'information à d'autres neurones.

Sur la membrane dendritique au niveau post-synaptique, se trouvent des récepteurs pour les neurotransmetteurs. En fonction du type de neurotransmetteurs et du type de récepteurs, l'activité du neurone (excitation) augmente ou diminue pour propager ou non l'information. Les synapses possèdent une **mémoire** permettant d'ajuster leur fonctionnement, par exemple, la co-activation répétée entre deux neurones modifie les connexions synaptiques qui les relient (Règle de Hebb) [46].

Ces connexions se modifient au fur et à mesure des flux nerveux pour faciliter ou non leur passage. Cette plasticité est à l'origine des mécanismes d'apprentissage. La modification synaptique peut intervenir sous trois formes principales :

- Augmentation ou diminution du nombre de récepteurs sur le neurone post-synaptique.
- Augmentation ou diminution du nombre de neurotransmetteurs émis.
- Augmentation ou diminution de l'efficacité d'un neurotransmetteur pour la génération d'un potentiel d'action.

1. prolongements arborescents du corps cellulaire, relativement courts

2. Prolongement long et fibreux

3. Signal électrique

4. Médiateurs chimiques

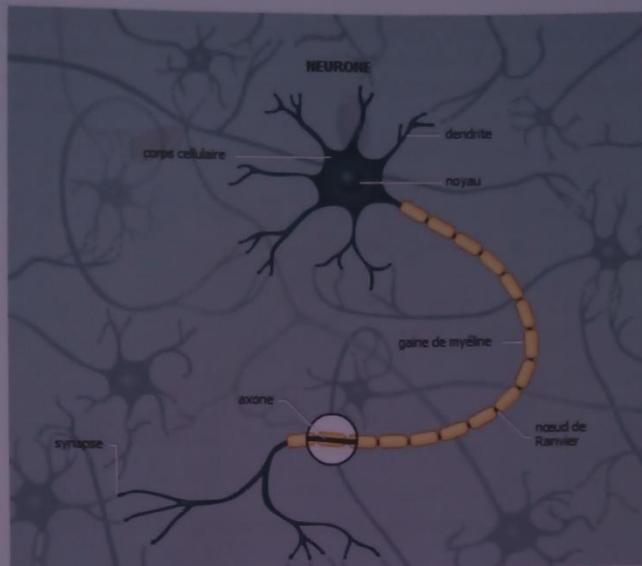


FIGURE 3.5 – Schéma d'un neurone biologique

### 3.5.2 Le Perceptron (1958)

Abordons tout d'abord le Perceptron qui est la forme la plus simple des réseaux de neurones, il ne permet la classification que de deux classes seulement car il consiste en un seul neurone qui possède un seuil ainsi qu'un vecteur de poids synaptique ajustable.

Après l'apprentissage et fixation des poids Figure 3.6, le neurone réalise une simple somme pondérée de ses entrées, compare une valeur de seuil et fournit une réponse binaire en sortie. Présenté initialement par Rosenblatt, en 1958, le perceptron permet de classifier correctement des objets appartenant à deux classes linéairement séparables.

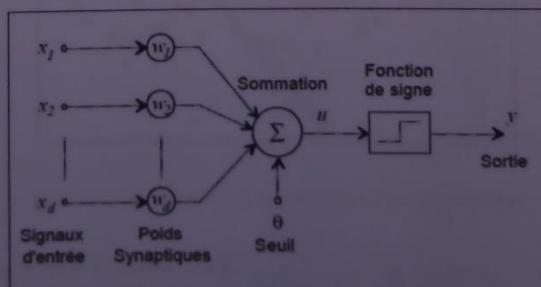


FIGURE 3.6 – Représentation d'un perceptron

Il est possible d'agrandir le champ d'action sur les classes par la mise en parallèle de plusieurs perceptrons (Figure 3.7), le perceptron ainsi obtenu comporte un neurone par classe, chacun de ceux-ci réalisant une fonction discriminante linéaire de la classe à laquelle il est associé.

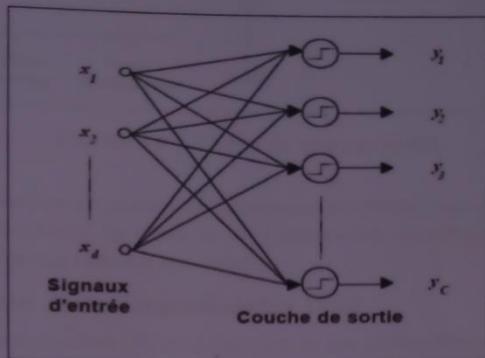


FIGURE 3.7 – Représentation d'un perceptron multicouche

Les frontières ainsi obtenues sont linéaires (Figure 3.8) comme pour la méthode euclidienne, toutefois, l'apprentissage discriminant que subit le perceptron lui confère un avantage certain vis-à-vis des autres classificateurs.

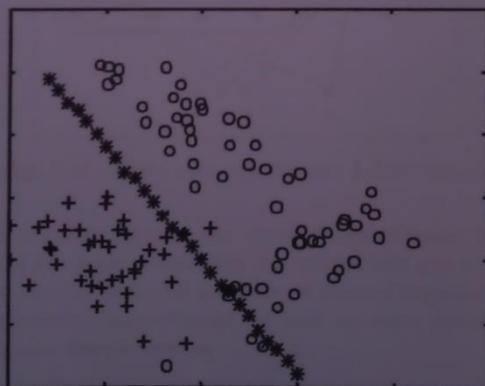


FIGURE 3.8 – Frontière de décision d'un perceptron

### 3.5.3 Le modèle formel 1987

La première étude systématique du neurone artificiel est due au neuropsychiatre McCulloch et au logicien Lippmann [47], qui, s'inspirant de leurs travaux sur les neurones biologiques, proposèrent en 1943 le modèle suivant (Figure 3.9) :

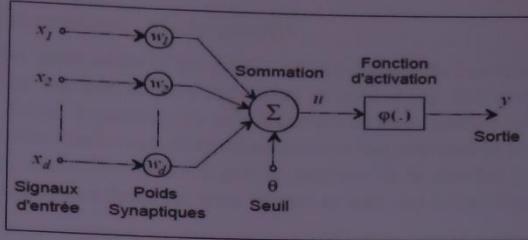


FIGURE 3.9 – Modèle de neurone formel

Ce neurone formel est un processeur élémentaire qui réalise une somme pondérée des signaux qui lui parviennent. La valeur de cette sommation est comparée à un seuil et la sortie du neurone est une fonction non-linéaire.

### 3.5.4 La Carte Auto-Organisatrice 1990

La carte auto-organisatrice [48] tient son nom de son architecture en forme de carte à deux dimensions où se situe l'entrée de tous les neurones (Figure 3.10). Chacune ayant des coordonnées ( $x, y$ ) indiquant sa position sur la carte, et un vecteur de poids synaptiques, appelé ainsi par analogie avec les synapses rencontrées dans le cerveau.

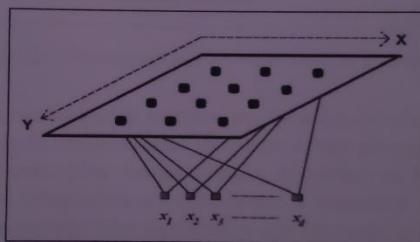


FIGURE 3.10 – Carte auto-organisatrice à deux dimensions

Tout réseau de neurones requiert une phase d'apprentissage, laquelle a pour but d'adapter les valeurs des poids synaptiques des cellules afin que le réseau soit capable de remplir la tâche que l'on désire lui attribuer en phase d'exploitation. Dans le cas de la carte auto-organisatrice, l'apprentissage est basé sur deux constatations biologiques essentielles, à savoir que, dans le cerveau :

- Chaque cellule nerveuse correspond à un stimulus spécifique ;
- il existe une région d'intense activité autour de la cellule la plus stimulée.

Le modèle d'entraînement de la carte consistera donc à :

- Sélectionner le neurone correspondant le mieux à un signal d'entrée donné ;
- Induire dans un voisinage de l'élu une région d'intense activité.

### 3.5.5 Le Perceptron Multicouche

Un perceptron multicouche (Figure 3.11) est une cascade de perceptrons non-linéaire avec une fonction continue d'allure sigmoïdale (Figure 3.12). Par exemple, quand un vecteur d'objet arrive, il communique avec les neurones de la première couche, les sorties des neurones de cette couche sont alors communiquées aux neurones de la couche suivante, et ainsi de suite, toutes les couches à part la première et la dernière (sortie du réseau) sont appelées **couches cachées** car leurs valeurs ne sont pas accessibles de l'extérieur.

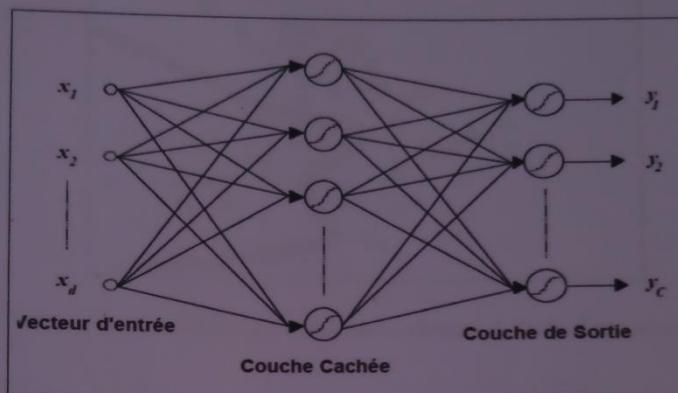


FIGURE 3.11 – Perceptron multicouche avec une couche cachée

L'apprentissage de ces réseaux de neurones multicouches est réalisé par l'intermédiaire de la retro-propagation du gradient, cette technique « *back-propagation* » permet de calculer le gradient de l'erreur pour chaque neurone du réseau, de la dernière couche vers la première. Les poids synaptiques qui contribuent à engendrer une erreur importante se verront modifiés de manière plus significative que les poids qui ont engendré une erreur marginale. Ce principe est à la base des réseaux de neurones multicouches comme le perceptron multicouche.

$$\phi(u) = \frac{1}{1 + \exp(-\beta u)}$$

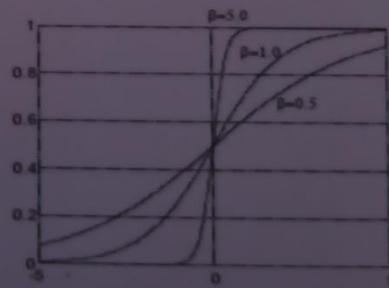
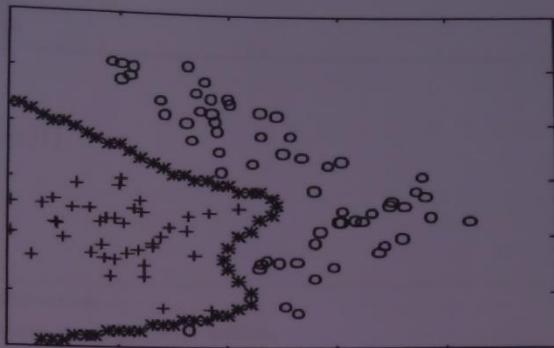


FIGURE 3.12 – Fonctions sigmoidales

L'apprentissage tend à augmenter la valeur de sortie du neurone qui correspond à la

classe correcte, tout en cherchant à diminuer la valeur des sorties des neurones associés aux classes incorrectes. Il en résulte donc un apprentissage discriminant. La Figure 3.13 ci-après illustre un exemple de frontière de décision non-linéaire générée par un perceptron multicouche.



**FIGURE 3.13 –** Fonction de décision non linéaire obtenue à l'aide d'un perceptron multicouche

### 3.6 Conclusion

Nous avons vu dans cette partie les différentes méthodes d'apprentissage et plus particulièrement les réseaux de neurones que nous utiliserons dans notre approche. Ces méthodes pour la plupart jeunes, sont de nos jours couramment utilisées et ne cessent d'évoluer, ce qui prouve leur efficacité et leur adaptabilité dans de nombreux domaines.

Dans le prochain chapitre, nous nous étalerons sur la conception de notre projet et les techniques utilisées partiellement citées dans ce chapitre.

## Conception

### Sommaire

|            |  |           |
|------------|--|-----------|
| <b>4.1</b> | <b>Introduction</b>  | <b>32</b> |
| <b>4.2</b> | <b>Résumé de l'approche utilisée</b>                         | <b>32</b> |
| <b>4.3</b> | <b>Présentation des bases d'apprentissage</b>                | <b>33</b> |
| 4.3.1      | La récolte des données                                       | 34        |
| 4.3.2      | Présentation de la base d'images                             | 35        |
| 4.3.3      | Filtrage de la base d'images par la méthode des juges        | 36        |
| 4.3.4      | Post-sélection   | 36        |
| <b>4.4</b> | <b>Caractéristiques faciales (prétraitement des données)</b> | <b>36</b> |
| 4.4.1      | Les points d'intérêt faciaux                                 | 39        |
| 4.4.2      | Méthode d'extraction des points faciaux                      | 41        |
| 4.4.3      | Les distances entre les points d'intérêt                     | 43        |
| 4.4.4      | Les rapports de distance                                     | 45        |
| <b>4.5</b> | <b>Apprentissage</b>   | <b>46</b> |
| 4.5.1      | Le choix du réseau   | 47        |
| 4.5.2      | Les entrées de notre réseau (Input)                          | 47        |
| 4.5.3      | Les sorties de notre réseau (Output)                         | 48        |
| <b>4.6</b> | <b>Conclusion</b>  | <b>48</b> |

### 4.1 Introduction

Après avoir fait un balayage assez complet sur les différentes techniques dont nous avons besoin dans notre travail, nous présenterons dans ce chapitre le fruit de l'analyse de toutes ces méthodes, en partant des bases d'images que nous avons utilisé jusqu'à la méthode de classification.

### 4.2 Résumé de l'approche utilisée

Voulant développer une approche inspirée de la biologie humaine, nous avons créé une base de données raffinée à partir de toutes celles existantes en nous référant au jugement humain. Ensuite, une méthode de détection de points faciaux nous a permis de réduire

les données de toute une image en seulement 18 points, ce qui est très ressemblant à la détection d'expressions faites par le cerveau humain qui n'a besoin que d'un ensemble réduit de points pour reconnaître une expression.

Les données extraites nous ont permis de calculer des rapports de distance entre ces points en nous basant sur le système de codification d'expressions faciales de Paule Ekman. Une fois ce travail effectué, nous avons utilisé une méthode dynamique pour choisir l'architecture de notre réseau de neurones et ainsi faire notre apprentissage.

### 4.3 Présentation des bases d'apprentissage

L'objectif de notre travail est de proposer une nouvelle méthode pour la reconnaissance des expressions faciales à partir d'une image. Pour pouvoir réaliser cela, il nous a fallu récolter des bases d'images portant sur le sujet. Les bases d'images d'expressions faciales sont relativement peu nombreuses et présentent peu de diversité et de précision par rapport aux bases d'images de reconnaissance faciale par exemple. Les plus complètes d'entre elles sont présentées dans le tableau 4.1 ci-dessous.

| Nom de la base d'images  | Nombre totale d'images |
|--|------------------------|
| Radboud Faces Database   | 67                     |
| Mug face database  | 575                    |
| MMI database   | 240                    |
| Pictures of Facial Affect (POFA)   | 110                    |
| JAFF database  | 130                    |
| The Combined MIT/CMU Test Set with Ground Truth for Frontal Face Detection | 180                    |
| CBCL Face Database   | 472                    |
| Cohn-Kanade AU-Coded Facial Expression Database                            | 456                    |

TABLE 4.1 – Les bases d'images d'expressions faciales

Toutes ces bases bien que parmi les plus complètes, manquent de précision au niveau de la répartition des expressions faciales et de la représentation de ces dernières. Ce manque de précision est dû à plusieurs raisons :

- Le domaine de la reconnaissance faciale est relativement jeune et la plupart des bases d'images sont toujours en cours de construction et correction.
- Les personnes utilisées pour représenter les expressions faciales sont généralement des étudiants ou des volontaires qui ne bénéficient guère de formations pour simuler les expressions correctement. Ces expressions sont donc difficiles à obtenir surtout pour des non-acteurs.
- Les expressions proposées, sont des expressions simulées, rarement identiques aux expressions spontanées. Afin de parfaire l'apprentissage, une base d'images complète est nécessaire. Or, ceci serait coûteux en temps, car il faudrait collecter

un ensemble de vidéos ou images de personnes exprimant leurs véritables émotions.

Pour pallier à ce problème, nous avons décidé de créer notre propre base d'images en combinant les différentes bases disponibles. Pour le faire, nous sommes passés par trois étapes présentées ci-après.

#### 4.3.1 La récolte des données

Nous avons contacté les plus importantes institutions de recherche qui possèdent une base d'images sur les expressions faciales. Le but étant de rassembler le plus grand nombre d'images à partir de toutes ces bases selon la répartition des expressions faciales disponibles pour ensuite en faire une base commune. Les principales bases d'images qui nous ont été accordées et que nous avons utilisées sont les suivantes :

##### 1. Cohn-Kanade AU-Coded Facial Expression Database

Cette base d'images, contient 456 séquences de 97 personnes. Chaque séquence commence avec une expression neutre et culmine avec l'expression ciblée qui peut être la colère, surprise ou la tristesse. Les étiquettes des émotions sont basées sur le code FACS.



FIGURE 4.1 – Quelques photos de la base d'images COHN-KANADE

##### 2. JAFF database

Cette base de données contient 213 images de 7 expressions faciales (6 expressions faciales de base + 1 neutre) faites par 10 modèles féminins japonais. Chaque image a été évaluée sur 6 adjectifs d'émotions par 60 sujets japonais. La base de données 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.



FIGURE 4.2 – Quelques photos de la base d'images JAFFE

### 3. MUG database

Toutes les séquences ont été capturées dans un environnement de laboratoire contrôlé à haute résolution. La collection se compose de deux parties :

- (a) La première partie décrit quatre vingt six sujets effectuant les six expressions de base en fonction des « *Prototypes d'émotion* » tels que définis dans le Guide de l'enquêteur dans le manuel FACS[12].
- (b) La deuxième partie contient les mêmes sujets enregistrés alors qu'ils regardaient une vidéo induisant des émotions.



FIGURE 4.3 – Quelques photos de la base d'images MUG

Cette base brute nous a servi à créer notre propre base d'images selon une sélection rigoureuse basée sur la méthode des juges humains.

#### 4.3.2 Présentation de la base d'images

Nous avons sélectionnée manuellement parmi les bases proposées, plus de 1000 images réparties, non équitablement, sur les sept expressions en incluant le neutre (joie, surprise, colère, peur, dégoût, tristesse et neutre). Cette présélection est basée sur les études de Paul Ekman sur les muscles faciaux. En effet, ce dernier a défini les apparences musculaires faciales selon la nature de l'émotion ressentie. Suite à cette sélection basée sur le système FACS, nous avons obtenu 950 images comportant les unités d'actions

appropriées pour chaque expression.

**Remarque :** Nous avons volontairement choisi un large éventail d'images variées pour en tirer les meilleurs résultats durant les prochaines étapes.

#### 4.3.3 Filtrage de la base d'images par la méthode des juges

Nous avons réuni un groupe de 22 juges comportant un personnel universitaire et des étudiants des deux sexes pour leurs exposer les images présélectionnées. Leur objectif était de définir instinctivement si l'image exposée représentait une expression parmi celles proposées et sélectionner celle qui y correspond. Leurs choix étaient inscrit dans des formulaires anonymes qui sont ensuite comptabilisés selon chaque expression.

Cette étape a réduit la base de 950 images à 650, réparties en moyenne sur 90 images par expression.

#### 4.3.4 Post-sélection

Cette dernière étape est une post-sélection destinée à la vérification des résultats obtenus. Pour ce faire, nous avons pris 30 % de la base pour l'exposer à cinq autres juges. Les résultats semblables à l'étape précédente à 80 %. La base d'images résultante de l'étape 4.3.3 sera notre base en entrée pour le prétraitement avant l'étape de l'apprentissage.

### 4.4 Caractéristiques faciales (prétraitement des données)

Suite à notre étude des différentes méthodes d'extraction de caractéristiques faciales, nous avons opté pour une approche basée modèle détaillée dans la partie 4.4.2. Ce choix, motivé principalement par le fait que cette méthode apporte un degré de précision satisfaisant par rapport à son temps de calcul relativement réduit améliorant grandement le rendu de notre apprentissage.

Pour choisir un nombre de caractéristiques aussi restreint que pertinent, nous nous sommes basés sur les études des muscles du visage qui proviennent, pour la plupart, des recherches de Paul Ekman qui est de nos jours la sommité la plus consultée dans ce domaine. Cette approche purement physiologique nous permettra de sélectionner de façon très réaliste les mouvements les plus importants dans les différentes expressions faciales.

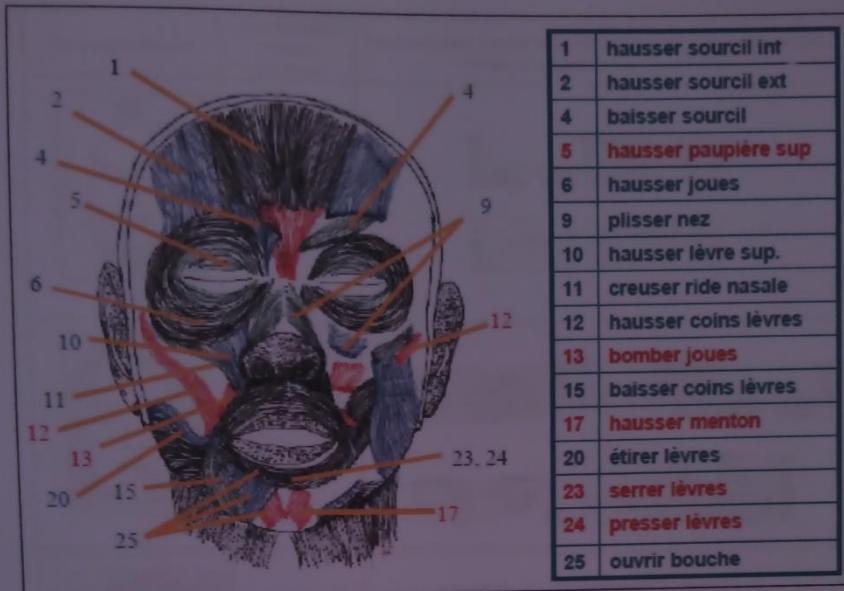


FIGURE 4.4 – Les unités d'action les plus importantes du système FACS avec leurs muscles associés

Dans leur « *Facial Action Coding System : A Technique for the Measurement of Facial Movement* » [12], Ekman et Friesen ont codifié tous les mouvements possibles des muscles du visage (voir l'Annexe). Ils ont aussi sélectionné un certain nombre de mouvements spécifiques à chaque expression (Tableau 4.5). Ces mouvements sont appelés « *Unités d'Actions* » (*Action Unit*, « *AU* »).

CHAPITRE 4. CONCEPTION

| Expressions faciales  | Combinaison d'Unités d'Actions | Représentation visuelle des unités d'action (AU) et de l'expression faciale résultante de leur combinaison |
|---|--------------------------------|--|
|    | 4+5+23+24                      |                          |
| Colère  |                                |  |
|    | 1+2+4+5+20+25                  |                          |
| Peur  |                                |  |
|    | 1+2+5+25                       |                          |
| Peur  |                                |  |
|  | 1+4+15                         |                        |
| Tristesse   |                                |  |
|  | 10+13+25                       |                        |
| Joie  |                                |  |
|  | 10+25                          |                        |
| Dégoût  |                                |  |

FIGURE 4.5 – Liste des expressions faciales avec les unités d'action correspondantes et leur représentation visuelle

Ces mouvements varient en intensité et en nombre selon le degré d'émotion ressenti par la personne. Nous en avons choisi les plus importantes qui distinguent chaque expression d'une autre. À partir de là et en se basant sur ces études, nous sommes passés par trois étapes afin d'extraire les caractéristiques faciales voulues.

#### 4.4.1 Les points d'intérêt faciaux

Afin de capter les unités d'action, il nous a fallu sélectionner leurs points saillants. Ces points du visage serviront de nœuds entre les distances. Ces dernières, seront sélectionnées selon les unités d'action et d'autres critères que nous détaillerons par la suite.

La méthode de modèle de la forme active (*ASM Active Shape Model*), configurée pour identifier un ensemble de points de repère sur l'image, détecte un nombre important de points significatifs, (Figure 4.6), ce qui nous a amené à nous concentrer essentiellement sur les trois parties les plus expressives en termes d'émotions : *les sourcils, les yeux et la bouche*.

Voici un exemple de points détectés par la méthode ASM :



FIGURE 4.6 – Exemple d'un modèle de forme active ASM

##### 1. sourcils :

Nous avons sélectionné deux points aux extrémités de chaque sourcil nous permettant de localiser tous les mouvements des muscles frontaux souvent les plus décisifs pour une expression faciale. (Figure 4.7).



FIGURE 4.7 – Les points faciaux des sourcils

2. Les yeux :

Il existe quatre points intéressants pour chaque œil. Dans la figure suivante, les deux premiers points (1, 2) et (5, 6) nous permettront de détecter le soulèvement de la paupière supérieure, tandis que les deux autres (3, 4) et (7, 8) nous serviront de repère pour calculer le reste des distances du fait qu'ils restent fixes dans le visage quel que soit l'expression affichée. (Figure 4.8).



FIGURE 4.8 – Les points faciaux des yeux

3. La bouche :

Étant donné la grande diversité des mouvements possibles avec la bouche et la complexité des unités d'action qui s'y associent, nous avons choisi six points saillants pour couvrir le plus grand nombre de mouvements de la mâchoire inférieure et des lèvres (Figure 4.9).



FIGURE 4.9 – Les points faciaux de la bouche

Il existe d'autres zones susceptibles d'être choisies mais que nous avons préféré écarter telles que : *le nez, les joues et le front.*

4. Le nez :

Bien que les contours du nez soient détectables, leur utilisation n'apporterait pas de résultats significatifs en termes d'expressions faciales considérant la complexité

du traitement engendrée en retour.

**5. Les joues :**

les muscles reliés aux joues sont primordiaux dans les expressions notamment pour le sourire. Néanmoins, notre choix ne s'est pas tournée vers eux pour deux principales raisons : les contours des joues sont difficilement identifiables, et diffèrent d'un individu à un autre.

Les muscles des joues sont directement reliés aux extrémités de la bouche, cette particularité rend inutile leur détection facilement identifiable par le mouvement des coins de la bouche.

**6. Le front :**

Bien que le front soit une zone importante, il est aussi difficile à aborder que les joues et peut être facilement remplacé par le mouvements des sourcils.

Ayant adopté cette approche que nous appelons *minimaliste*, nous avons obtenu un total de dix-huit points pour l'intégralité du visage, un nombre relativement réduit, notamment par le fait que nous ayons abandonné tous les points de contour du visage inutiles dans notre approche.

#### 4.4.2 Méthode d'extraction des points faciaux

Afin d'extraire les points d'intérêt choisis, nous avons opté pour une méthode d'extraction par **modèle de forme active**. Ce choix a été motivé par le fait que cette méthode soit parmi les plus efficaces en terme d'extraction de données faciales, de plus, elle couvre tous les points dont nous avons besoin et repose sur l'utilisation d'un calcul à partir de descriptions texturales hiérarchiques de visage d'une phase d'entraînement et combine des modèles de formes et de textures, pour finalement produire un masque de points sur le visage épousant de manière automatique la forme du visage selon une procédure d'adaptation itérative.



FIGURE 4.10 – Points d'intérêt du modèle de forme active

Nous avons utilisé des bibliothèques d'apprentissage *open source* [49] qui offrent des masques faciaux personnalisables. Le masque (modèle) de la figure 4.10 est le résultat d'un apprentissage à partir de la base d'images : « *The Muct Face Data base* », notre programme l'utilise pour extraire les dix huit points faciaux nous intéressant. (Figure 4.11).



FIGURE 4.11 – Adaptation du masque du modèle active d'apparence à une expression faciale

#### 4.4.3 Les distances entre les points d'intérêt

Afin de reconnaître une expression, la détection des différents mouvements des trois parties citées plus haut (*les sourcils, les yeux et la bouche*) est indispensable, c'est pourquoi nous choisirons des paires de points parmi ceux sélectionnés précédemment afin de calculer la distance euclidienne entre eux nous permettant ainsi de connaître le degré de rapprochement ou d'éloignement des paires de points sélectionnées.

Prenons deux points  $P_i$  et  $P_j$ , leur distance Euclidienne serait égale à :

$$dist_{eu}(P_i, P_j) = \sqrt{((P_i.y - P_j.y)^2 + (P_i.x - P_j.x)^2)} \quad (4.1)$$

Avec :

$P_i.x$   $P_i.y$  : La position du point  $P_i$  par rapport à la matrice de l'image en millimètre.

Les distances que nous avons choisies ont pour but de correspondre aux unités d'action spécifiques à chaque expression faciale et de représenter le système de codification FACS le plus fidélement possible (Figure 4.12).



FIGURE 4.12 – Les points d'intérêts sur un visage ainsi que les distances choisies

Le tableau suivant justifie le choix des treize distances présentées dans la figure ci-dessus ainsi que les unités d'action qui leur correspondent dans le système FACS (Tableau 4.13) :

## CHAPITRE 4. CONCEPTION

### Distance 1 :

Elle correspond au rapprochement des extrémités intérieures des sourcils. Ce qui est spécifique à l'unité d'action 4 qui est associée le plus souvent à la colère.



AU 4

### Distances 2 et 5 :

Correspond au rapprochement et à l'écartement des coins extérieurs des sourcils par rapport aux coins extérieurs des yeux. cette distance décrit les mouvements du muscle frontal extérieur représenté par l'unité d'action 2 et qui détermine souvent la nature de l'expression faciale (entre la surprise et la tristesse par exemple)



AU 2

### Distances 3 et 4 :

Décris les mouvements du muscle frontal centrale qui se contracte en levant les coins intérieurs des sourcils ou en les abaissant, elle représente l'unité d'action 1 qui est spécifique à la tristesse



AU 1

### Distances 6 et 7 :

Cette distance est dédiée au soulèvement de la paupière supérieure qui produit l'ouverture des yeux comme dans la surprise ou la peur, Ceci correspond à l'unité d'action 5 ou la fermeture partielle des yeux comme dans la joie ou la tristesse



AU 5

### Distances 8 et 11 :

Les muscles qui contrôlent les joues influencent le mouvement des coins des lèvres, il peuvent ce soulever pour ce rapprocher des coins extérieurs des yeux comme pour faire un sourire (unité d'action 12), ou s'éloigner par l'activation d'un autre muscle situé en dessous de la bouche qui tire les coins de cette dernière vers le bas comme dans le cas de la tristesse (unité d'action 15)



AU 12



AU 15

## Distance 9 et 10 :

Nous avons ajouté cette distance spécifiquement pour pallier à la difficulté de distinguer l'expression de dégoût qui ce confond le plus souvent avec la colère, elle représente l'étirement de la lèvre supérieur par le biais d'un muscle situer à chaque coin du nez qui est codifier par l'unité d'action 10



AU 10

## Distance 12 :

Cette distance représente l'élargissement de la bouche comme dans le cas de la peur. Cet étirement des lèvres est causé par la contraction du muscle horizontalement relié aux coins des lèvres (unité d'action 20)



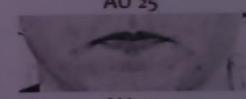
AU 20

## Distance 13 :

La dernière distance décrit l'action de plusieurs muscles qui produisent le pincement de la bouche ou son ouverture. Ceci est produit par l'abaissement de la mâchoire inférieure. Ces mouvements sont codifiés par les unités d'action : 23, 24 et 25 qui change pour chaque expression faciale



AU 25



AU 24



AU 23

FIGURE 4.13 – Les points d'intérêt sur un visage ainsi que les distances choisies

Nous avons au final 13 distances qui représentent les différents mouvements faciaux codifiés par le système FACS.

**4.4.4 Les rapports de distance**

La dernière étape consiste à calculer quelques rapports entre les distances précédemment choisies. Ces rapports représentent les caractéristiques extraites du prétraitement de chaque image.

Le nombre possible de rapports entre les 13 distances est trop grand pour pouvoir être

exploité exhaustivement (156 rapports). Il est donc préférable de choisir les rapports les plus pertinents, nous en avons donc sélectionné 25 en nous basant sur les différentes combinaisons de mouvements pour chaque expression faciale (Figure 4.14).

Ce nombre de rapports peut encore être réduit en supprimant les récurrences, ce qui nous amène finalement à 19 rapports significatifs.

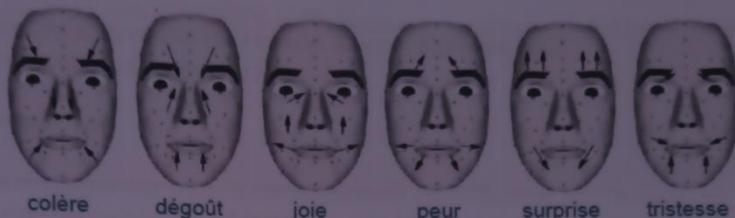


FIGURE 4.14 – Les différentes combinaisons de mouvements faciaux par rapport aux expressions étudiées

Le tableau suivant (Tableau 4) liste les rapports de distances ( $Dist_x / Dist_y$ ) choisies selon les différentes combinaisons de mouvements pour chaque expression faciale.

| Dist 2/ Dist3   | Dist5/ Dist4    | Dist7/ Dist11  | Dist 3/ Dist 6   | Dist 4/ Dist 7   |
|-----------------|-----------------|----------------|------------------|------------------|
| Dist 13/ Dist12 | Dist 6/ Dist8   | Dist6/ Dist13  | Dist 8/ Dist 1   | Dist 11/ Dist 13 |
| Dist3/ Dist1    | Dist4/ Dist1    | Dist 8/ Dist12 | Dist9/ Dist1     | Dist 10/ Dist 1  |
| Dist 12/ Dist1  | Dist 7/ Dist 13 | Dist 9/ Dist 8 | Dist 10/ Dist 11 |                  |

FIGURE 4.15 – Liste des rapports de distances

Avec  $Dist_x$  définie pour les points  $x$  tel illustré dans la Figure 4.13.

Ce nombre réduit de rapports augmentera considérablement l'efficacité de notre apprentissage par réseaux de neurones artificiels. En effet, les données d'entrée ont encore diminué après l'étape du prétraitement, car cette dernière ne positionnait pas toujours les points au bon endroit sur le visage. Ces erreurs sont dues à sa grande sensibilité au bruit, ombres, barbes, moustaches, lunettes, et mèches cachant les sourcils. Cette étape a réduit notre benchmark à moins de 400 images pertinentes.

## 4.5 Apprentissage

Après avoir extrait les caractéristiques de chaque visage de la base d'apprentissage, nous les introduirons comme paramètres d'entrée à notre réseau de neurones. Le choix a été porté sur ces derniers pour leur faible sensibilité aux bruits et leur grande capacité d'apprentissage, ceci devrait réduire considérablement les erreurs de classification.

### 4.5.1 Le choix du réseau

Pour pouvoir être sûr que notre architecture a bien était choisie, nous avons décidé de faire un balayage assez exhaustif des différentes combinaisons des paramètres de notre réseau de neurones. Pour cela, nous avons écrit un script en *Matlab* afin de varier les valeurs de ces différents paramètres :

- Nombre de couches cachées (1 ou 2).
- Nombre de neurones sur chacune de ces couches cachées.
- Les fonctions d'apprentissage.

Autrement dit :

- Notre programme teste automatiquement les différentes combinaisons de neurones en couches cachées, de 5 à 200 par pas de 10 pour la première couche du réseau, et de 5 à 100 par pas de 10 aussi pour la 2ème couche. Suite à cela, nous sélectionnerons les plages de neurones où les résultats sont les plus pertinents pour en réduire le pas. Nous répéterons ce procédé jusqu'à trouver le meilleur résultat d'apprentissage possible.
- Nous testerons de la même manière les 6 différentes fonctions d'apprentissage qui donnent habituellement les meilleurs résultats dans le domaine de la reconnaissance de formes. Ce sont : '*trainrp*', '*trainscg*', '*traincgp*', '*trainsgb*', '*trainlm*' et '*trainscg*'. Pour chaque fonction, nous relançons l'apprentissage de façon automatique jusqu'à 20 fois car les résultats varient d'une itération à une autre.
- Le résultat de chaque itération, est enregistré pour être comparé aux autres par rapport aux performances, aux régressions et au temps d'exécution. Nous affichons au final les meilleurs résultats de chaque fonction.

### 4.5.2 Les entrées de notre réseau (Input)

Les entrées du réseau sont les 19 rapports obtenus à partir de chaque image de la base sélectionnée, nous avons donc 19 valeurs réelles en entrée de notre réseau.

Voici un petit échantillon des entrées de notre réseau :

| R1     | R2     | R3     | R4     | R5     | R6...R11 | R12   | R13    | R14    | R15    | R16    | R17    | R18    | R19    |
|--------|--------|--------|--------|--------|----------|-------|--------|--------|--------|--------|--------|--------|--------|
| 0,7984 | 0,7111 | 1,8220 | 1,7449 | 1,2177 | .....    | 0,392 | 0,3751 | 0,6275 | 0,6256 | 1,176  | 1,0354 | 0,8543 | 0,9892 |
| 1,3043 | 1,0025 | 1,6089 | 1,4716 | 2,2041 | .....    | 0,550 | 0,5442 | 0,4988 | 0,5391 | 1,3637 | 1,6162 | 0,8645 | 0,9457 |
| 0,8908 | 0,8901 | 2,1958 | 2,1149 | 1,6141 | .....    | 0,598 | 0,6027 | 0,4261 | 0,4455 | 1,6849 | 0,9579 | 0,8363 | 0,8870 |
| 0,9379 | 0,6775 | 1,9569 | 2,1231 | 1,8864 | .....    | 0,667 | 0,6927 | 0,8437 | 0,8066 | 1,5530 | 1,2147 | 0,7899 | 0,7868 |

FIGURE 4.16 – Tableau des entrées

#### 4.5.3 Les sorties de notre réseau (Output)

La sortie de notre réseau comportera 7 neurones, qui représentent une des 7 expressions faciales à classifier, elle sera codifiée de la manière suivante (Tableau 4.2) :

| Expression | Codification en sortie |
|------------|------------------------|
| Colère     | 1 0 0 0 0 0 0          |
| Dégoût     | 0 1 0 0 0 0 0          |
| Peur       | 0 0 1 0 0 0 0          |
| Joie       | 0 0 0 1 0 0 0          |
| Neutre     | 0 0 0 0 1 0 0          |
| Tristesse  | 0 0 0 0 0 1 0          |
| Surprise   | 0 0 0 0 0 0 1          |

TABLE 4.2 – Sortie de notre réseau de neurones

Le réseau final comportera deux couches cachées et les fonctions d'apprentissage qui ont les courbes de régression les plus avantageuses, ainsi que le nombre de neurones appropriés à chaque couche (les résultats expérimentaux sont décrits dans le chapitre qui suit).

### 4.6 Conclusion

Dans ce chapitre, nous avons détaillé notre méthode de conception en partant de la construction de la base de données jusqu'au calcul des rapports entre les distances en terminant avec les caractéristiques de notre réseau de neurones.

Dans le prochain chapitre, nous exposerons les différents environnements utilisés ainsi que les résultats de notre implémentation et tests effectués.

# Chapitre 5

## Expérimentation

### Sommaire

|       |                                    |    |
|-------|------------------------------------|----|
| 5.1   | Introduction                       | 49 |
| 5.2   | Environnement logiciel et matériel | 49 |
| 5.2.1 | Hardware                           | 49 |
| 5.2.2 | Software                           | 49 |
| 5.3   | Présentation du logiciel développé | 51 |
| 5.3.1 | Résultat et Tests                  | 56 |
| 5.4   | Conclusion                         | 63 |

### 5.1 Introduction

Ce chapitre est consacré à la description de la dernière étape de notre étude. Nous commencerons par détailler l'environnement matériel et logiciel choisi, nous entamerons ensuite une présentation de l'interface développée et son fonctionnement, pour enfin terminer avec un exposé des résultats obtenus et une analyse des tests expérimentaux effectués.

### 5.2 Environnement logiciel et matériel

L'environnement sur lequel nous avons développé notre application est structuré comme suit :

#### 5.2.1 Hardware

Notre application a été développée sur un micro-ordinateur de la configuration suivante :

- Microprocesseur : IntelCore™2DuoCPU T5800@2.00GHz × 2.
- Mémoire vive (RAM) : 3 GB.
- Disque Dur : 250 GB.
- Résolution de l'écran : 1280 x 800 Pixels.

#### 5.2.2 Software

##### - Système d'Exploitation :

*GNU/Linuz* : Distribution *Ubuntu 12.04*, pour sa stabilité et sa gestion optimisée.

misée, la liberté, gratuité, modularité, sécurité, facilité d'utilisation des différentes ressources de la machine, simplicité d'implémentation, ainsi que la disponibilité des différents outils dont nous avions besoin.

- **Langage de programmation :**

**C++** : Langage de programmation; conçu par Bjarne Stroustrup au début des années 1980; permet, d'un côté, de maîtriser la représentation bas niveau des données manipulées (arithmétique de pointeurs, allocation manuelle de la mémoire, ...) tout en fournissant des outils (références, exceptions, classes, templates, ...), et d'un autre côté, de construire des structures de plus haut niveau. Il est actuellement le 3<sup>ème</sup> langage le plus utilisé au monde. Le langage **C++** est libre, impliquant une liberté d'utilisation sans besoin d'une autorisation ou obligation de paiement.

- **Traitements d'images :**

La bibliothèque **OpenCV** (version 2.4.5) : (*Open Computer Vision*) est une bibliothèque graphique libre, conçue pour le traitement d'images en temps réel et a été officiellement lancé en 1999. Le projet **OpenCV** a été développé initialement par *Intel* pour optimiser les applications gourmandes en temps / processeur. Cela faisait partie d'une série de projets tels que l'affichage d'un mur en 3 dimensions. Les principaux acteurs du projet sont l'équipe de développement de bibliothèque de chez *Intel* ainsi qu'un certain nombre d'experts dans l'optimisation de chez *Intel Russie*.

Les objectifs de base du projet étaient :

- Faire des recherches sur la vision par ordinateur en vue de fournir un logiciel libre et optimisé.
- Établir une infrastructure commune s'appuyant sur les développeurs pour obtenir un code plus lisible et transférable.
- Continuer à développer en rendant le code portable et permettre des performances optimisées gratuites avec une licence qui est libre de toutes contraintes commerciales.

- **Compilation de la partie traitement de l'image :**

**CMake** : Un système extensible, libre, qui gère le processus de génération de compilateur dans un système d'exploitation de manière indépendante. Contrairement à de nombreux systèmes multi-plateformes, **CMake** a été conçu pour être utilisé en conjonction avec l'environnement de compilation native. Les fichiers de configuration simples placés dans chaque répertoire source (appelé fichiers *CMakeLists.txt*) sont utilisés pour générer des fichiers de *build standard* (par exemple, fichiers *makefile* sur Unix et projets / espaces de travail sous Windows MSVC) qui sont utilisés de la manière habituelle. **CMake** peut générer un environnement natif de construction qui va compiler le code source, créer des bibliothèques, générer des *wrappers* et construire des exécutables dans des combinaisons arbitraires. **CMake** supporte les compilation dans le même répertoire ainsi que ceux en dehors de ce dernier, et peut donc prendre en charge plusieurs versions d'un arbre, d'un code source unique. **CMake** prend également en charge les versions des bibliothèques

statiques et dynamiques.

Une autre fonctionnalité intéressante de *CMake* est la génération d'un fichier cache conçu pour être utilisé avec un éditeur graphique. Par exemple, lorsque *CMake* est exécuté, il localise des fichiers inclus, bibliothèques et exécutables, et peut rencontrer des directives de compilation optionnelle. Ces informations sont collectées dans la mémoire cache, qui peut être modifiée par l'utilisateur avant la génération des fichiers sources compilés.

– **Environnement de Développement Intégré (EDI) et Apprentissage :**

**Matlab** : (*MATrix LABoratory*) est un environnement puissant, complet et facile à utiliser destiné aux calculs scientifiques. Il apporte aux ingénieurs, chercheurs et à tout scientifique un système interactif intégrant calcul numérique et visualisation. C'est un environnement performant, ouvert et programmable qui permet de remarquables gains de productivité et de créativité.

Dans notre travail, nous avons notamment eu recours à l'utilisation des logiciels suivants :

- **Gimp / Blender** : Pour la création des différents schémas et images utilisés dans les différents chapitres ainsi que dans l'interface.
- **LaTeX** : Pour la rédaction de notre document.
- **Prezie** : Pour la réalisation de la présentation.

### 5.3 Présentation du logiciel développé

Dans cette section, nous présenterons notre application avec tous les détails de son fonctionnement.

L'interface homme-machine développée Figure 5.1 intègre toutes les fonctionnalités implémentées auparavant, allant du traitement d'images jusqu'à la classification. Les étapes suivantes illustrent le mode d'utilisation de cette application en s'appuyant sur l'exemple suivant :

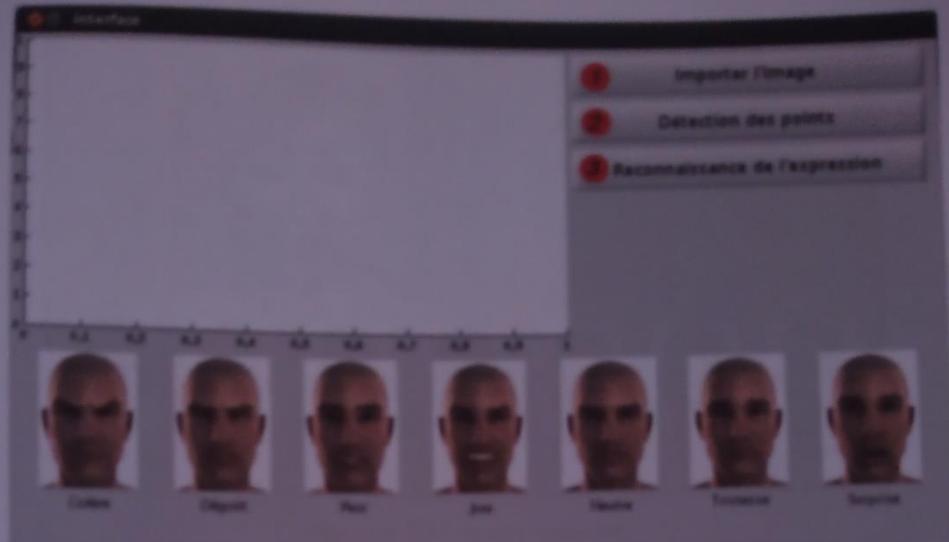


Figure 5.1 – Vue générale sur notre IHM

### 1. Importation de l'image test :

Le 1<sup>er</sup> bouton dans la figure 5.1 permet l'importation d'une image au format png (Figure 5.2) qui sera affichée dans un cadre de l'interface (Figure 5.3).

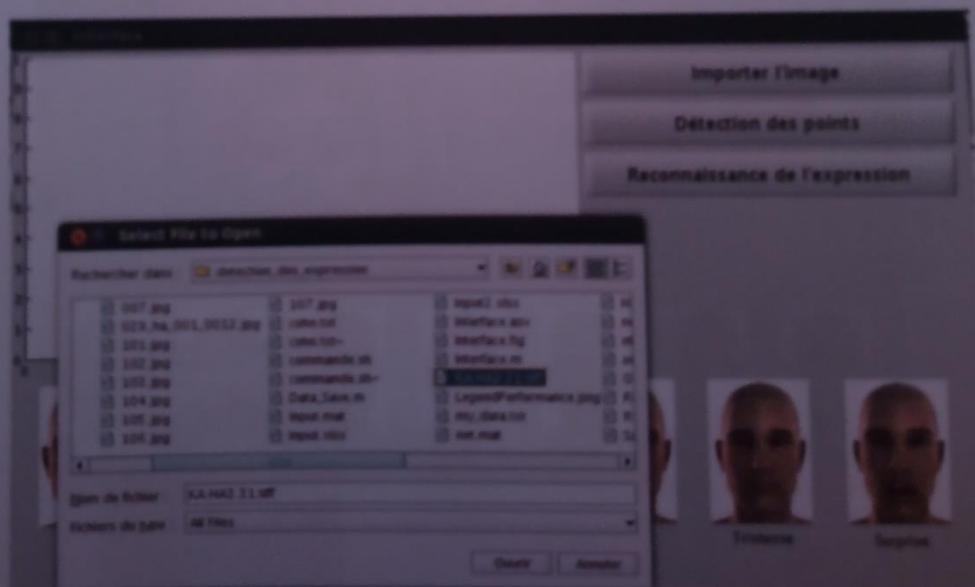


FIGURE 5.2 – Importation de l'image



FIGURE 5.3 – Image importée

## 2. Extraction de caractéristiques :

Le 2<sup>ème</sup> bouton sert à la détection des points du visage dans l'image importée, le résultat de ce traitement est affiché dans une nouvelle fenêtre (Figure 5.4).



FIGURE 5.4 – Détection des points d'intérêt

À la suite de la détection des points faciaux, le programme calcule les distances puis les rapports appropriés afin de les stocker dans un vecteur de test.

3. Classification :

Le 3<sup>ème</sup> bouton est utilisé pour la reconnaissance de l'expression à partir des données d'apprentissage précédemment créées, le résultat est affiché en pourcentage en dessous des figures de chaque expression (Figure 5.5), l'émotion la plus dominante est affichée d'une couleur différente (dans notre cas la joie).

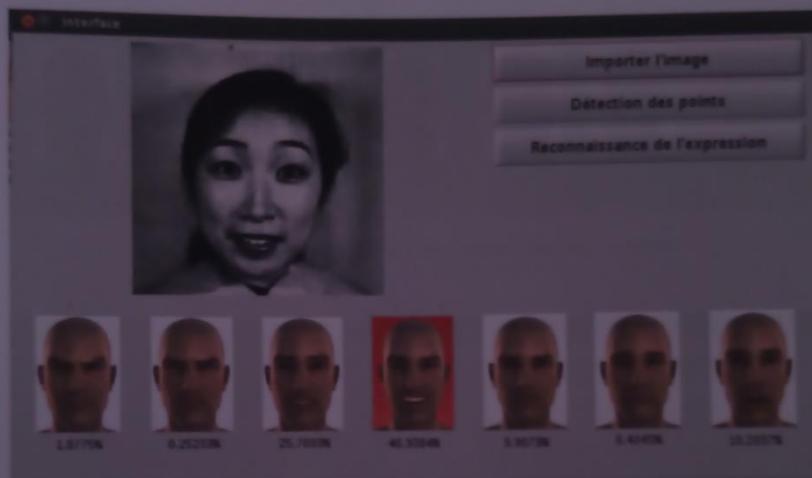


FIGURE 5.5 – Classification de l'expression faciale reconnue

Les figures suivantes (Figure 5.6, Figure 5.7) sont d'autres exemples qui donnent une bonne reconnaissance :

CHAPITRE 5. EXPÉRIMENTATION



FIGURE 5.6 – Exemple sur l'expression faciale 'surprise'



FIGURE 5.7 – Exemple sur l'expression faciale 'neutre'

À titre expérimental, nous avons testé quelques images éliminées précédemment durant le filtrage de la base de données. Nous avons donc constaté que les résultats étaient souvent médiocres comme dans la figure 5.8 qui a été classifiée comme étant de la joie alors qu'en réalité c'est de la peur :



FIGURE 5.8 – Exemple sur l'expression faciale "peur"

### 5.3.1 Résultat et Tests

Cette dernière partie porte sur les résultats obtenus lors de l'apprentissage et les tests effectués.

#### 5.3.1.1 Résultats de l'apprentissage

En utilisant notre programme d'apprentissage, nous avons usé de nos données de différentes manières afin d'avoir les résultats les plus satisfaisants et une analyse pertinente.

Les prochaines étapes illustrent les différentes combinaisons de tests effectués avec leurs courbes de régressions et de performances, les données de tests sont choisies de manière aléatoire dans notre réseau de neurones.

##### 1. Apprentissage sur les sept expressions de base :

Notre premier choix s'est porté sur l'utilisation de toute la base de données avec les sept expressions (joie, tristesse, colère, peur, surprise, neutre, dégoût). En sortie, le résultat obtenu est le suivant (Figure 5.9).

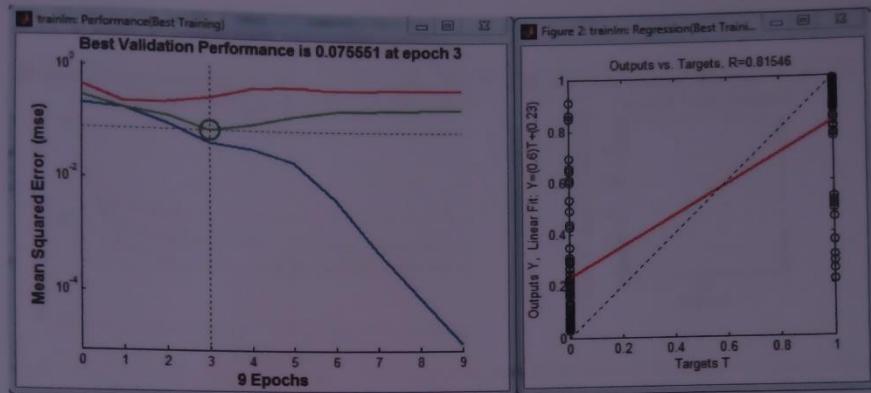


FIGURE 5.9 – Résultat d'apprentissage sur les sept expressions de base

Ces courbes montrent un résultat d'apprentissage relativement satisfaisant mais qui est loin d'être optimal, par souci d'optimisation nous avons effectué d'autres situations d'apprentissage pour arriver à situer la base de ce manque de précision.

## 2. Apprentissage sur chaque émotions séparément :

Cette étape consiste en une comparaison entre chaque expression et les autres, ce traitement a pour objectif de tester chaque expression séparément et d'identifier leurs performances.

les figures suivantes montrent les courbes de régression et de performance de chaque expression comparée avec toutes les autres.

### (a) Colère :

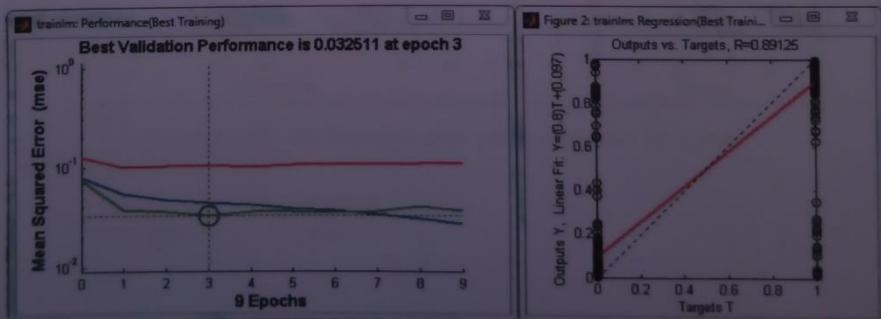


FIGURE 5.10 – Performance ainsi que la régression de la colère

$$\begin{aligned} \text{Régression totale} &= 0.89125. \\ \text{Performance} &= 0.032511. \end{aligned}$$

(b) Dégoût :

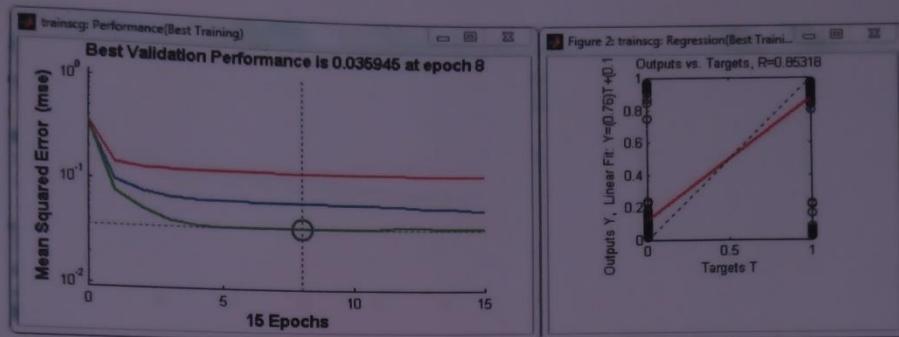


FIGURE 5.11 – Performance ainsi que la régression du dégoût

*Régression totale = 0.85318.**Performance = 0.035945.*

(c) Joie :

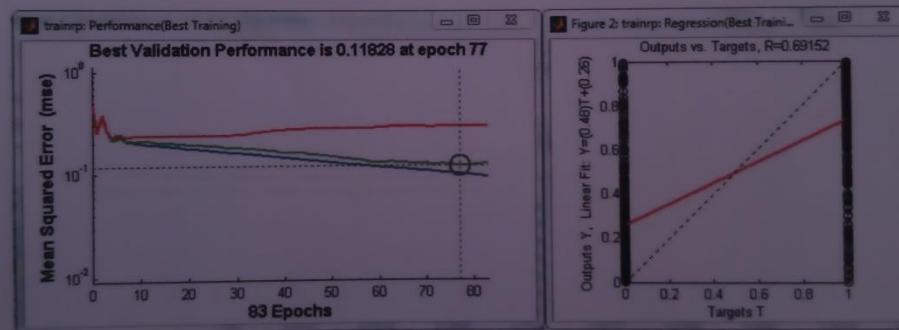


FIGURE 5.12 – Performance ainsi que la régression de la joie

*Régression totale = 0.69152.**Performance = 0.11828.*

(d) Neutre :

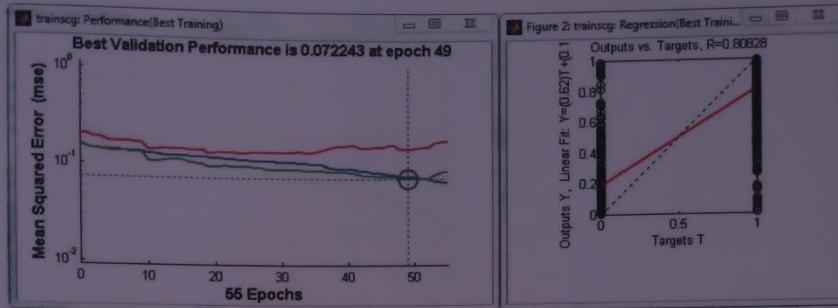


FIGURE 5.13 – Performance ainsi que la régression de l'expression neutre

Régression totale = 0.80828.  
Performance = 0.072243.

(e) Peur :

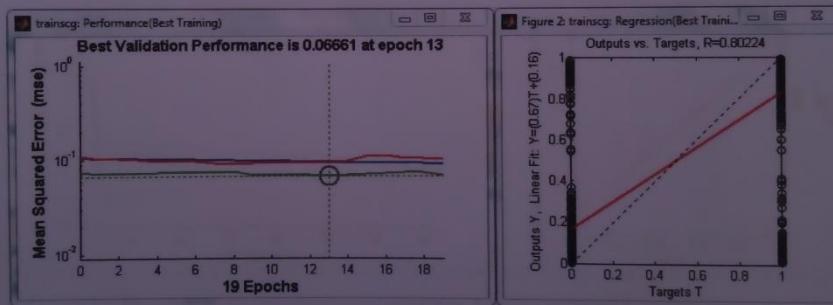


FIGURE 5.14 – Performance ainsi que la régression de la peur

Régression totale = 0.80224.  
Performance = 0.06661.

(f) Surprise :

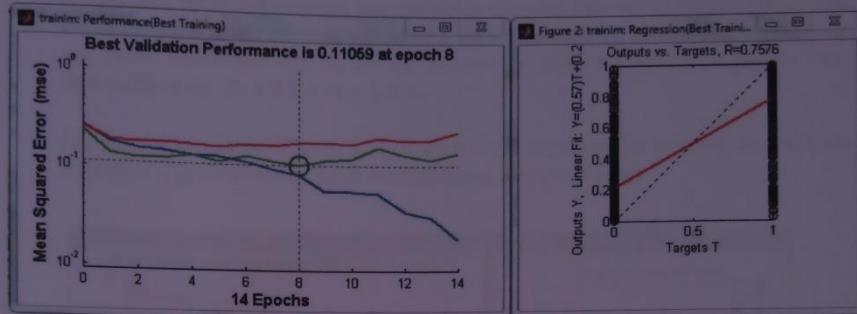


FIGURE 5.15 – Performance ainsi que la régression de la surprise

Régression totale = 0.7576.

Performance = 0.11059.

(g) Tristesse :

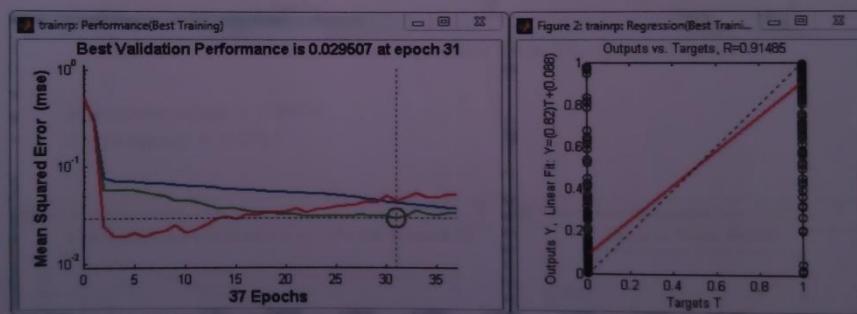


FIGURE 5.16 – Performance ainsi que la régression de la tristesse

Régression totale = 0.91485.

Performance = 0.029507.

À la suite de tout ces tests, nous pouvons conclure que les résultats sont relativement proches entre eux et qu'il n'existe pas de réelle amélioration en prenant chaque expression à part. Néanmoins, nous remarquons une certaine corrélation entre la performance d'apprentissage et le nombre d'images utilisées dans l'apprentissage pour chaque expression, cet indice nous a orienté vers la prochaine configuration.

### 3. Apprentissage par paire :

En partant de la supposition que certaines données affectaient négativement nos résultats, nous avons effectué une série d'apprentissages en prenant les expressions par paire. Autrement dit, les sorties de notre architecture seront en binaire avec une codification de « 0 1 » et « 1 0 ».

Les figures suivantes (Figure 5.17, Figure 5.19) illustrent les courbes de quelques exemples d'expressions de cette configuration d'apprentissage.

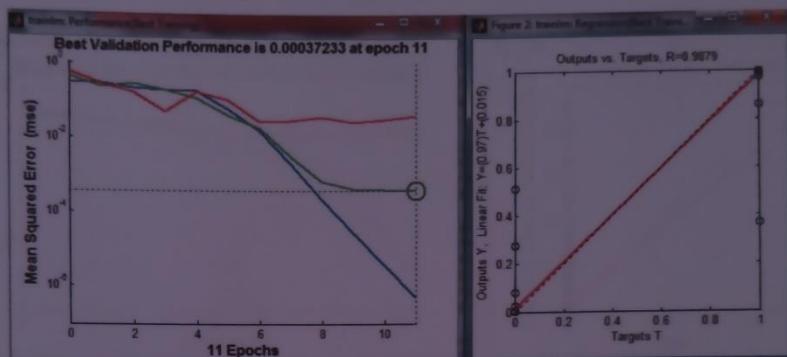


FIGURE 5.17 – Performance et régression entre la surprise et la colère

Régression totale = 0.9879.

Performance = 0.0003.

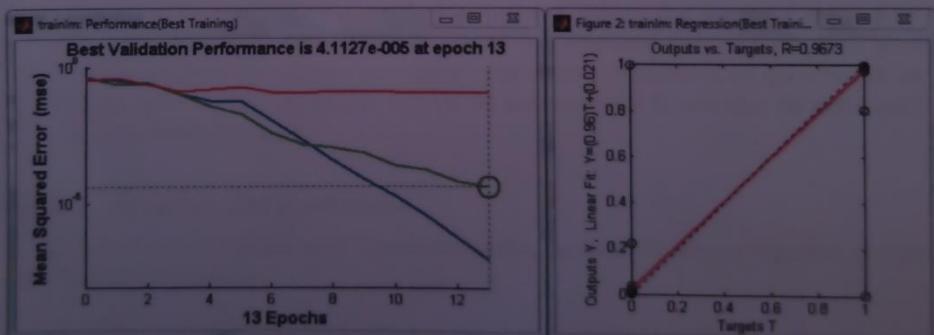


FIGURE 5.18 – Performance et régression entre le dégoût et la colère

Régression totale = 0.9673.

Performance = 0.00041127.

À la suite de ces tests très satisfaisants, nous pouvons conclure que nous trouvons une nette amélioration par rapport aux résultats précédents, ceci est de même

pour des expressions très proches comme la dégoût contre la colère (Figure 5.17), ce dernier test nous a confirmé la présence de certaines données qui dégradent l'apprentissage.

#### 4. Apprentissage avec cinq expressions :

Suite à une grande série de tests, nous sommes enfin arrivés à un résultat optimal (Figure 5.19), cet avancement a été la suite de l'élimination de deux expressions (joie , neutre).

Performance et régression des 5 expressions faciales

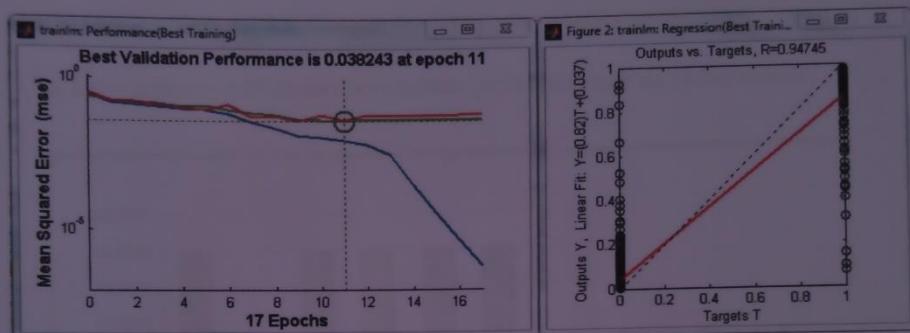


FIGURE 5.19 – Performance et régression des cinq expressions faciales

Régression totale = 0.94745.

Performance = 0.03822.

Ces résultats suivis d'une série de tests nous montrent que les deux expressions éliminées diminuent grandement la qualité de l'apprentissage, cela est dû à leurs grands nombres de données par rapport aux autres expressions, ce qui a faussé les écarts entre chaque expression, cela est dû notamment à la précision de notre outil de traitement d'images.

##### 5.3.1.2 Résultats des simulations

La base de tests utilisée pour l'expérimentation contient 70 images réparties sur les sept expressions étudiées.

Le tableau suivant 5.1 donne le résultat des tests sur les sept expressions faciales avec le nombre d'images reconnues.

| Expression | Nombre d'images reconnues / nombre total d'images | Taux de reconnaissance |
|------------|---|------------------------|
| Colère     | 8/10  | 80%                    |
| Dégoût     | 7/10  | 70%                    |
| Peur       | 8/10  | 80%                    |
| Joie       | 8/10  | 80%                    |
| Neutre     | 8/10  | 80%                    |
| Tristesse  | 8/10  | 80%                    |
| Surprise   | 7/10  | 70%                    |

TABLE 5.1 – Résultat des tests pour chaque expression

La figure suivante 5.20 illustre les résultats précédents dans un histogramme :

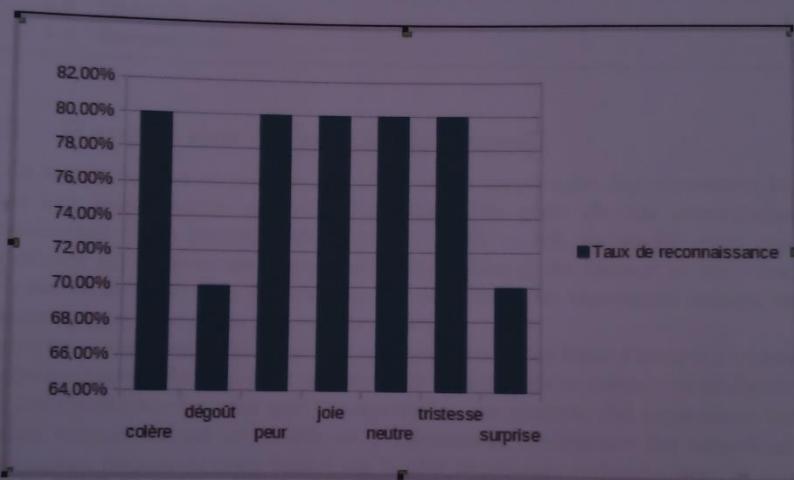


FIGURE 5.20 – Représentation du taux de reconnaissance pour les différentes expressions

#### 5.4 Conclusion

Dans ce chapitre, nous avons décrit en premier lieu l'environnement de développement matériel et logiciel de l'application, ensuite son interface et les différentes fonctionnalités qu'elle offre à l'utilisateur.

Enfin, nous avons présenté les difficultés que notre apprentissage a rencontré et les différentes combinaisons effectuées pour palier à ce problème.

D'autre part, les derniers résultats obtenus certifient la validité de l'approche proposée et démontre l'efficacité de notre conception.

# Conclusion Générale

## Sommaire

---

|   |    |
|---|----|
| 5.5 Rappel des objectifs de travail . . . . . | 64 |
| 5.6 Objectifs atteints . . . . .              | 64 |
| 5.7 Perspectives . . . . .                    | 65 |

---

### 5.5 Rappel des objectifs de travail

Ce travail est issu en premier lieu d'une étude approfondie des expressions faciales et de leur interprétation en terme d'émotions d'un point de vue psychologique et physiologique. Notre objectif était de créer à partir de cela un modèle aussi innovant que efficace qui combinerait la puissance d'apprentissage des réseaux de neurones avec une nouvelle interprétation des systèmes de codification des expressions faciales, ceci en utilisant la base d'images la plus précise possible.

Pour ce faire, nous avons commencé par rassembler toutes les bases d'images d'expressions faciales existantes afin d'en créer une nouvelle plus raffinée et précise, en utilisant aussi bien l'interprétation humaine que l'interprétation par symbole des expressions faciales. Ensuite, nous avons créé un modèle de codification et d'extraction des caractéristiques d'expressions faciales en nous basant sur la plus importante recherche dans ce domaine (*Facial Action Coding System*).

Enfin la méthode d'apprentissage et de classification a été grandement approfondie pour en tirer le plus de résultats et de conclusions que possible.

### 5.6 Objectifs atteints

À la fin de ce parcours passionnant, les plus importantes réalisations de ce travail sont :

- Une nouvelle base d'images d'expressions faciales qui regroupe ce qu'il y a de meilleur dans tous les autres, et qui sera mis à la disposition de tout chercheur voulant l'utiliser ou l'enrichir.
- Une nouvelle méthode d'extraction des caractéristiques d'expressions faciales, qui réduit toutes les données d'une image à quelques points d'intérêt significatifs, ces

derniers serviraient à eux seuls à calculer les distances, les rapports et à interpréter cette expression.

- Des résultats d'apprentissage prometteurs qui prouvent l'efficacité de la méthode de conception et qui classifie les différentes expressions de base d'une façon relativement précise.
- Nous pouvons Ajouter à tout cela la réalisation d'un script en Matlab qui modifie de façon automatique le nombre de neurones, les fonctions utilisées, et le nombre d'itérations pour chaque couche. À la fin de ce traitement, le programme nous donnera les courbes, les données et l'architecture des meilleurs résultats obtenus. Ce script sera très utile à toute personne voulant travailler avec les réseaux de neurones d'une façon rapide et optimale.

### 5.7 Perspectives

Bien que nous comptons continuer à améliorer ce travail à l'avenir, nous proposons à toute personne voulant se pencher sur ce sujet pour l'exploiter ou l'enrichir à explorer ces quelques pistes :

- La base d'images mériterait grandement d'être enrichie par les nouvelles bases en ce moment en construction, une autre approche serait de recréer une base d'images d'expressions faciales plus précise et qui se baserait sur le réalisme et la compétence des volontaires. Cette 2ème proposition que nous avons commencé à explorer mais que nous avons abandonné en prenant connaissance du temps et de la technicité que cela demanderait à réaliser de manière optimale, bien que plus coûteuse en terme de temps, pourra donner de meilleurs résultats d'apprentissage.
- Bien que la méthode de traitement d'images utilisée soit relativement efficace, il serait judicieux d'essayer d'autres techniques de reconnaissance de formes qui donneraient de meilleurs résultats et diminueraient le bruit.
- Les réseaux de neurones ne sont pas la seule méthode de classification possible, il existe d'autres techniques applicables à ce problème tel que les réseaux bayésiens qui pourrait donner des résultats intéressants.
- Cette étude a porté sur un nombre restreint d'expressions faciales dites *universelles*, l'étude pourrait s'étendre sur d'autres expressions moins communes comme le *méprit*, ou sur des systèmes de communication spécifiques comme les expressions faciales dans le langage des signes.

## Annexe

## A

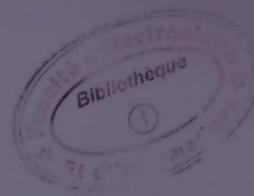
| AU | Nom de l'A.U   | Effet   | Muscles activés                             |
|----|--|---|---|
| 01 | Lever le front intérieur (Inner brow raiser)           | Relève la partie intérieure des sourcils  | Frontalis, pars medialis                    |
| 02 | Lever le front extérieur (Outer brow raiser)           | Relève la partie extérieure des sourcils  | Frontalis, pars lateralis                   |
| 04 | Abaïsser le front (Brow lower)                         | Baisse les sourcils et les rapproche l'un de l'autre  | Corrigator supercilii, dépressor supercilii |
| 05 | Lever la paupière supérieure (Upper lid raiser)        | Relève les paupières supérieures, dégageant ainsi une plus grande partie des globes oculaires           | Levator, palpebrae superioris               |
| 06 | Lever la joue (Cheek raiser)                           | Relève les joues, produit des plis cutanés dans la partie latérale extérieure des yeux et sous les yeux | Orbicularis oculi, pars orbitalis           |
| 07 | Lever la paupière inférieure (Lid tightener)           | Relève et tend les paupières inférieures  | Orbicularis oculi, pars palpebralis         |
| 09 | Plisser le nez (Nose wrinkle)                          | Plisse la peau du nez en la tirant vers le haut   | Levator labii, superioris alaqueae nasi     |
| 10 | Lever la lèvre supérieure (Upper lip raiser)           | Relève la lèvre supérieure en lui donnant une forme anguleuse   | Levator labii superioris                    |
| 11 | Approfondir le sillon nasogénien (Nasolabial deepener) | Accentue le sillon nasogénien   | Zygomaticus minor                           |
| 12 | Etirer le coin des lèvres (Lip corner puller)          | Tire les coins des lèvres de façon diagonale vers les os des joues (sourire)                            | Zygomaticus major                           |
| 13 | Bouffer les joues (Cheek puffer)                       | Augmente l'apparition du triangle infra orbital   | Levator anguli oris (a.k.a caninus)         |
| 14 | Fossettes (Dimpler)                                    | Fait apparaître les fossettes aux coins des lèvres  | Buccinator                                  |

ANNEXE A.

|    |  |  |  |
|----|--|--|--|
| 15 | Abaïsser le coin des lèvres (Lip corner depressor) | Abaïsse le coin des lèvres   | Depressor anguli oris (a.k.a triangularis)                               |
| 16 | Abaïsser la lèvre inférieure (Lower lip depressor) | Abaïsse la lèvre inférieure  | Depressor labii inferioris   |
| 17 | Lever le menton (Chin raiser)                      | Pousse la bosse du menton et la lèvre inférieure vers le haut                        | Mentalis   |
| 18 | Plisser les lèvres (Lip puckerer)                  | Rétracte les lèvres et les fait ressortir vers l'avant (moue)                        | Incisivii labii superioris, incisivii labii inferioris                   |
| 19 | Tirer la langue (Inner brow raiser)                | La langue sort de la bouche  | Risorius, platysma   |
| 20 | Extendre les lèvres (Lip stretcher)                | Etire le coin des lèvres vers l'extérieur  | Platysma   |
| 21 | Tendre le cou (Neck tightener)                     | Tend le cou  | Platysma   |
| 22 | Mettre les lèvres en entonnoir (Lip funneler)      | Projette les lèvres les lèvres vers l'avant et expose d'avantage leur partie colorée | Orbicularis oris   |
| 23 | Tendre les lèvres (Lip tightener)                  | Imprime une tension sur les lèvres et les fait paraître plus minces                  | Orbicularis oris   |
| 24 | Compresser les lèvres (Lip pressor)                | Pince les lèvres vers l'intérieur  | Orbicularis oris   |
| 25 | Séparer les lèvres (lips parted)                   | Ouvre la bouche et sépare très légèrement les lèvres                                 | Depressor labii inferioris ou détente du mentalis ou du orbicularis oris |
| 26 | Baisser la mâchoire (Jaw drop)                     | Ouvre la bouche au point où il y a un espace entre les dents                         | Masseter, détente du temporalis et pterygoid interne                     |
| 27 | S'étirer la bouche (Mouth stretch)                 | Bâillement   | Pterygoids, digastric  |
| 28 | Sucer les lèvres (Lip suck)                        | Pince les lèvres pour les sucer  | Orbicularis oris   |

TABLE A.1 – Liste non exhaustive de quelques unités d'action les plus importantes du système FACS

# Bibliographie



- [1] Carol NEIDLE. Signstream[tm] annotation : Conventions used for the american sign-languagelinguisticresearch project. Technical report, 11, Boston University, August 2002.
- [2] P. Bard. A diencephalic mechanism for the expression of rage with special reference to the central nervous system. *American journal of psychology*, 1928.
- [3] W. Cannon. The james-lange theory of emotions : A critical examination and an alternative theory. *American journal of psychology*, 1927.
- [4] W. V. FRIESEN Paul EKMAN. Unmasking the face. 1975.
- [5] Jörgen AHLBERG. Candide-3 - un updated parameterised face. Technical report, Dept. of Electrical Engineering, Linköping University, Sweden, January 2001.
- [6] S. Matsumoto N. Nishiuchi K. Masuda K. Kurihara, D. Sugiyama. *Computers and Industrial Engineering*, volume 56, chapter Facial emotion and gesture reproduction method for substitute robot of remote person. 2009.
- [7] V. Callaghan F. Doctor E. Leon, G. Clarke. *Affect-aware behaviour modelling and control inside an intelligent*.
- [8] A. Aldea J. Martinez-Miranda. Emotions in human and artificial intelligence. In *Computers in Human Behavior*, volume 21, 2005.
- [9] J. Klein R. W. Picard. Computer that recognise and respond to user emotion theoretical and practical. In *Interacting with Computers*, volume 14, 2002.
- [10] D. Wang C. Gysin P. Gidley B. Gantz J. Rubinstein S. Seiff L. Poon K. Lun Y. Ng P. Dulguerov, F. Marchal. *Review of objective topographic facial nerve evaluation methods*, volume 5. 1999.
- [11] R. Koenen. *Mpeg-4 Project Overview*. PhD thesis, International Organisation for Standardisation, 2000. ISO/IEC JTC1/SC29/WG11.
- [12] W.V. Friesen P. Ekman. Facial action coding system : A technique for the measurement of facial movement. In *Consulting Psychologists Press*. Palo Alto, 1978.
- [13] P. Ekman. *Methods for measuring facial actions*, in : K.Scherer, *Handbook of Methods in Nonverbal Behaviour Research*. PhD thesis, Cambridge University Press, Cambridge, 1982.
- [14] CH Hjortsjo. *Manniskans sikt och mimiska sprake*. 1969.

---

BIBLIOGRAPHIE

- [15] Alex PENTLAND Irfan ESSA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, chapter Coding, Analysis, Interpretation, and Recognition of Facial Expressions. 1997.
- [16] Paul EKMAN Terence J. SEJNOWSKI Marian Stewart BARTLETT, Joseph C. HAGER. *Measuring Facial Expressions by Computer Image Analysis*, volume 36 of *Psychophysiology*. 1999.
- [17] Working group on visual, 2001.
- [18] W.V. Friesen P. Ekman. *Personality Social Psychol*, chapter Constants across cultures in the face and emotion. 1971.
- [19] A. MEHRABIAN. *Communication without Words*, volume 2, chapter 4, page 53–56. PsychologyToday, 1968.
- [20] N. DAILEY Curtis PADGETT Ralph ADOLPHS W. Garrison, Mathew COTTERELL. *Computational, Geometric, and Process Perspectives on Facial Cognition : Contexts and Challenges*, chapter Is All Face Processing Holistic ? The viewfrom UCSD. M. Wenger and J. Townsend, 2000.
- [21] King N. NGAN Douglas CHAI. Locating facial region of a headand- shoulderscolor image. In *In Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition (FG'98)*, pages 124–129, Nara, Japan, 1998.
- [22] Narendra AHUJA Ming-Hsuan YANG, David J. KRIEGMAN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 1, chapter Detecting Faces in Images : A Survey, pages 34–58. 2002.
- [23] Christophe COLLET. *CapRe : un système de capture du regard dans un contexte d'Interaction Homme-Machine*. PhD thesis, LIMSI, 1998.
- [24] Patrice DELMAS. *Extraction des contours de lèvres d'un visage parlant par contours actifs - Application à la communication multimodale*. PhD thesis, INPG, Grenoble, 2000.
- [25] Y. Yacoob M. Black. Recognizing facial expressions in image sequences using local parameterized models of image motion. In *Internat. J. Comput. Vision*, volume 25. 1997.
- [26] L. Rothkrantz M. Pantic. Expert system for automatic analysis of facial expression. In *Image Vision Comput*, volume 18, J. 2000.
- [27] S. Akamatsu M. Lyons, J. Budynek. *IEEE Trans. Pattern Anal. Mach. Intell*, volume 12, chapter Automatic classification of single facial images. 1999.
- [28] J. Ohya T. Otsuka. Spotting segments displaying facial expression from image sequences using hmm, ieee. In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, Nara,Japan, 1998.
- [29] T.S. Huang I. Cohen, A. Garg. *Emotion Recognition from Facial Expressions using Multilevel HMM*. 2000.
- [30] L. Davis M. Rosenblum, Y. Yacoob. *IEEE Trans. Neural Networks*, volume 5, chapter Human expression recognition from motion using a radial basis function network architecture. 1996.

- [31] F. Hara H. Kobayashi. Dynamic recognition of basic facial expressions by discrete-time recurrent neural network. In *Proceedings of the International Joint Conference on Neural Networks*, 1993.
- [32] Q. Ji H. Gu, Y. Zhang. Taskoriented facial behavior recognition with selective sensing. In *Computer Vision and Image Understanding*, volume 100, 2005.
- [33] N. Tsapatsoulis S. Kollias W. Fellenz, J. Taylor. Comparing template-based, feature-based and supervised classification of facial expressions fromstatic images. In *Circuits, Systems, Communications and Computers*, 1999.
- [34] G. Cottrell M. Dailey. *PCA Gabor for expression recognition*. PhD thesis, Institution UCSD, 1999. Number CS-629.
- [35] G. K. Venayagamoorthy S. Bashyal. *Engineering Applications of Artificial Intelligence*, volume 21, chapter Recognition of facial expressions using Gabor wavelets and learning vector quantization. 2008.
- [36] Q. Ji Y. Zhang. *Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences*, volume 27. 2005.
- [37] Y. Sun I. Cohen T. Gevers T.S. Huang N. Sebe, M.S. Lew. Authentic facial expression analysis. In *Image and Vision Computing*, volume 25, 2007.
- [38] M. Pietikainen G. Zhao. Boosted multi-resolution spatiotemporal descriptors for facial expression recognition. In *Pattern Recognition Letters*, volume 30, 2009.
- [39] R. E.Schapir Y. Freund. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, (no. 55), 1997.
- [40] Bernard GOSELIN. *Application de réseaux de neurones artificiels à la reconnaissance automatique de caractères manuscrits*. PhD thesis, 1996.
- [41] Yves Rangoni. *Sur les Réseaux de neurones dynamique Perceptif*. 2007.
- [42] K. Fukunaga. *Introduction to Statistical Pattern Recognition* Academic Press. San Diego, 1990.
- [43] J.-P. Fénelon L. Lebart, A. Morineau. *Traitemet des données statistiques*. Paris, 1982.
- [44] S. Haykin. *Neural Networks*. PhD thesis, A comprehensive Foundation Macmillan College Publishing Company, New York, 1994.
- [45] Sofiane Boucenna. De la reconnaissance des expressions faciales à une perception visuelle partagée : une architecture sensori-motrice pour amorcer un référencement social d'objets, de lieux ou de comportements.
- [46] D. Hebb. *The organization of behavior : A neuro psychological theory*. LEA, 1949.
- [47] R. Lippmann. *An Introduction To Computing with Neural Nets*, volume 4 of pp 4-22. IEEE ASSP Magazine, Avril 1987.
- [48] T. Kohonen. *The Self-Organising Map proc. of the IEEE*, volume 78 of pp 1464-1480. Septembre 1990.
- [49] C.J. Taylor T.F. Cootes, G.J. Edwards. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 6, chapter Active appearance models, pages 681–685. June 2001.

---

BIBLIOGRAPHIE

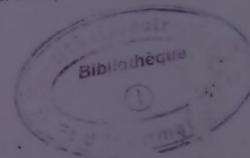
- [50] J. Lien Y. Wu T. Kanade J. Cohn, A. Zlochower. Automated face coding : a computer vision basedmethod of facial expression analysis. In *Seventh European Conference on Facial Expression Measurement and Meaning*. Salzburg, Austria, 1997.
- [51] MPEG Working Group on VISUAL. *International Standard on Coding of Audio-Visual Objects*, Part 2, 2001. Visual.

## Résumé

La reconnaissance d'émotions est une thématique importante aux frontières de la psychologie et de la vision par ordinateur. Ses applications sont aussi importantes que variées, allant de l'animation de visage réaliste à l'amélioration de la communication homme-machine, en passant par les robots répondants aux émotions des utilisateurs.

Dans cette optique, l'approche proposée se base sur une étude physiologique des muscles du visage responsables des expressions faciales. Nous commençons par la détection du visage puis passons à l'extraction d'un ensemble réduit et significatif de caractéristiques qui nous permettraient de déduire des données suffisantes pour en faire un apprentissage automatique. Celui-ci est fait à l'aide d'un réseau de neurones. L'implémentation ainsi réalisée permet de reconnaître certaines des expressions faciales qui existent et que nous avons étudiées dans ce projet.

*Mots-clés : Expressions faciales, Emotions, réseaux de neurones, Vision par ordinateur, FACS, Intelligence artificielle.*



**SII.103/13**