

Université L'arbi Ben M'Hidi Oum El Bouaghi
Faculté des sciences exactes et sciences de la nature et de la vie
Département d'informatique

Thèse de Doctorat 3ème Cycle LMD en vue de l'obtention du diplôme
De Docteur en informatique

THEME

MODELISATION DE L'ETAT AFFECTIF

DANS LES SEQUENCES VIDEO

Présentée par
Kenza Belhouchette

Juin 2018

Membres du jury:

Président	Pr. Nini Brahim
Directeur de thèse	Dr. BELHADEF Hacene
Co- Directeur de thèse	Dr. BERKANE Mohamed
Examineur	Dr. BOUZENADA Mourad
Examineur	Dr. GUERRAM Tahar

Dédicace

Je dédie cette thèse à mes chers parents, Mohammed et Djermane Yamina, qui ont toujours été présents pour moi, « Vous avez tout sacrifié pour moi, n'épargnant ni santé ni efforts. Vous m'avez donné un magnifique modèle de labeur et de persévérance. Je suis redevable d'une éducation dont je suis fière »

Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma considération pour les sacrifices que vous avez consentis pour mon instruction et mon bien être.

Je vous remercie pour tout le soutien et l'amour que vous me portez depuis mon enfance et j'espère que votre bénédiction m'accompagne toujours.

Que ce modeste travail soit l'exaucement de vos vœux tant formulés, le fruit de vos innombrables sacrifices, bien que je ne vous en acquitterai jamais assez.

Malheureusement après tous vos efforts, vous n'avez pas assisté au résultat. Vous n'avez pas vu que votre fille a réalisé votre rêve (mama).

Que Dieu vous accueille dans son vaste paradis,.....

Une dédicace très spéciale à mes parts de cœurs mes enfants, Younes et Eline.

Remerciement

Je souhaite adresser mes remerciements les plus sincères aux personnes qui m'ont apporté leur aide morale et physique pour la réalisation de cette thèse :

Mon directeur de thèse Mr. Hacene Belhadef, Docteur en informatique, pour avoir accepté de diriger cette thèse. Ses critiques associées à ses judicieux conseils ont largement contribué à l'aboutissement de ce travail. L'apport de ses orientations et ses remarques à la fois rigoureuses et objectives a été plus qu'indispensable à la concrétisation de cette thèse.

Docteur Mohamed Berkane de m'avoir dirigé en tant que co-encadreur. Il s'est toujours montré à l'écoute et très disponible tout au long de la réalisation de cette thèse. Ainsi que, je le remercie pour l'inspiration, l'aide et le temps qu'il a bien voulu me consacrer.

Je tiens à remercier avec plus grande gratitude Monsieur Brahim Nini, de l'honneur qu'il me fait d'avoir accepté de présider le jury de cette soutenance de thèse. Je remercie également Monsieur Tahar Guerram d'avoir accepté de se joindre à ce jury comme examinateur. Je tiens également à présenter mes remerciements à Monsieur M.BOUZENADA d'avoir accepté d'évaluer cette thèse.

J'adresse mes plus sincères remerciements à tous mes proches, spécialement mon mari Chibane Adel d'avoir été toujours mon soutien dans le courant de ma vie et non seulement durant cette thèse.

Ma tante Dalila et sa fille Assia de m'avoir encouragé, poussé et soutenus dans les plus difficiles moments.

Table des matières

Table des matières

Introduction générale.....	12
Introduction	13
Contributions.....	14
Organisation de la thèse	15
Chapitre 1 : L'émotion	16
1.1 Introduction	17
1.2 Vers une définition de l'émotion.....	17
1.2.1 L'émotion en philosophie.....	17
1.2.2 L'émotion en psychologie	18
1.3 Les composantes de l'émotion	19
1.3.1 La composante comportementale :	19
1.3.2 La composante physiologique	20
1.3.3 La composante cognitive/subjective.....	20
1.4 Emotions et approches conceptuelles.....	21
1.4.1 La théorie périphérique.....	21
1.4.2 Théorie centraliste :	21
1.4.3 La théorie bi-factorielle de Schachter :.....	22
1.5 Emotions de bases et émotions secondaires	23
1.6 La représentation des émotions	24
1.6.1 Représentation par le système MPEG4 :	24
1.6.2 Représentation par catégories :	25
1.6.3 Représentation par le système : Facial Action Coding System (FACS)	26
1.7 Conclusion.....	31
2.1 Introduction	33
2.2 Canaux de communication émotionnels :	34

2.3	La reconnaissance des émotions.....	34
2.4	Méthodes de reconnaissance des émotions par les expressions faciales	35
2.4.1	La détection du visage :.....	36
a.	Méthodes basées sur les caractéristiques invariantes	36
b.	Méthodes basées sur la mise en correspondance.....	38
c.	Méthode basée sur les apparences.....	38
d.	Méthode basée sur la connaissance :.....	41
2.4.2	L'extraction des caractéristiques :.....	41
a.	Analyse bas niveau.....	42
b.	Analyse intermédiaire.....	43
c.	Analyse haut niveau	43
2.4.3	La classification :.....	43
a.	Modèle de Markov Caché : HMM (Hidden Markov Model).....	44
b.	Les réseaux bayésiens	46
c.	Support Vector Machine :	47
2.5	Conclusion.....	49
Chapitre 3 : Les réseaux de neurones		50
3.1	Introduction	51
3.2	Neurones biologiques et neurones formels :	51
3.3	Modèle mathématique et architecture des réseaux de neurones.....	53
3.4	Apprentissage supervisé et non supervisé	53
4.1	Apprentissage supervisé.....	54
a.	Approches basées sur les instances (IBL)	54
b.	Approches basées sur des règles(RBL)	54
4.2	Apprentissage non supervisé.....	55
a.	La segmentation ou le regroupement (clustering) :	55
b.	Règles d'association:.....	56
3.5	Exemple de réseaux de neurones :.....	56

3.5.1	Perceptron et perceptron multi couches	56
a.	Algorithme de rétropropagation	57
5.2	Les réseaux de Hopfield	58
a.	Description	58
b.	Les mémoires auto-associatives	59
5.3	Les cartes autoorganisatrices de kohonen	60
a.	Architecture	60
b.	Algorithme de Kohonen :.....	61
5.4	Les réseaux de neurones convolutifs CNN):.....	63
4.1	Les réseaux de neurones et les émotions :.....	64
4.2	Conclusion.....	66
Chapitre 4 : Système neuronal pour la reconnaissance des émotions		67
4.1	Introduction	68
4.2	Système neuronal pour la reconnaissance d'émotion.....	69
4.2.1	La détection de visage :.....	70
4.2.2	Extraction des points stratégiques	73
4.2.3	Détection des unités d'action du système FACS :.....	76
4.2.4	Estimation de mouvement	77
4.2.5	Classification	81
4.2.6	RESULTATS ET DISSCUSSION :.....	86
a.	DATA bases :.....	86
b.	Présentation des résultats:.....	89
c.	Evaluation des résultats	93
d.	Discussion et comparaison	96
e.	Mesures d'évaluation	98
f.	Paramètres d'optimisation.....	99
4.3	Nouveau modèle pour le codage de l'émotion d'intérêt par les unités d'actions :.....	100
4.3.1	Définition de l'intérêt	100

4.3.2	L'intérêt et le e-learning	101
4.3.3	Codage proposé pour l'émotion d'intérêt.....	102
4.3.4	Approche proposée pour la détection de l'émotion d'intérêt	103
4.3.5	Résultats	104
a.	Ensemble de donnée.....	104
b.	Evaluation des résultats.....	105
c.	Paramètres d'optimisation.....	106
d.	Discussion	107
4.4	Conclusion.....	109
ANNEXE A		113
Références Bibliographiques.....		116

Table des figures :

Figure	Légende	Page
Figure 1.1	Schéma illustratif des composantes de l'émotion	13
Figure 1.2	Conceptions des émotions selon la théorie périphérique	14
Figure 1.3	James-Lange VS Cannon-Bard Vs Scharchter	16
Figure 1.4	Quelques unités d'action	21
Figure 1.5	Codage des émotions de base par les unités d'action	22
Figure 2.1	Canaux de communication émotionnelle et capteurs associés	27
Figure 2.2	Processus d'analyse des Expressions faciales	28
Figure 2.3	La détection du visage	29
Figure 2.4	Caractéristiques Haar dans l'algorithme original de Viola et Jones	33
Figure 2.5	L'image intégrale pour le calcul des caractéristiques	34
Figure 2.6	Extraction des caractéristiques faciales	35
Figure 2.7	Classification des émotions de base	37
Figure 3.1	Le cerveau humain	45
Figure 3.2	Neurone biologique et neurone formel	45
Figure 3.3	Perceptron multi couches	50
Figure 3.4	Le réseau de neurone Hopfield	52
Figure 3.5	Carte de Kohonen à deux dimensions	54
Figure 3.6	Notion de voisinage dans une Carte de Kohonen	54
Figure 3.7	Exemple d'architecture d'un CNN	57
Figure 4. 1	Organigramme de la méthode proposée	62
Figure 4.2	Détection des visages dans une image	63
Figure 4.3	La cascade de Viola Jones	66
Figure 4.4	Les points stratégiques dans le visage	68
Figure 4.5	Les points stratégiques manipulés par notre système	69
Figure 4.6	le mouvement de L'unité d'action : AU4	70
Figure 4.7	Recherche d'un bloc correspondant dans une zone de recherche (B_x) entre deux images I_1 et I_2	71
Figure 4. 8	La similarité entre les points	72
Figure 4. 9	Direction des mouvements des unités d'action	73
Figure 4.10	La fonction de voisinage	77
Figure 4. 11	Le nœud le plus proche de l'entrée à une carte de Kohonen	77

Figure 4. 12	Adaptation des poids	78
Figure 4.13	La carte de KOHONEN proposée pour la reconnaissance des émotions	78
Figure4.14	Exemple d'images d'expressions faciales à partir de la base de données Cohn-Kanade	79
Figure4.15	Exemple d'images d'expressions faciales à partir de la base de données CK+	80
Figure4.16	Exemple d'images d'expressions faciales à partir de la base de données JAFFE	82
Figure4.17	Les étapes de reconnaissance de l'émotion (MMI)	83
Figure 4.18	Les étapes de reconnaissance de l'émotion (Cohn-Kanade)	83
Figure 4.19	Détection du visage (Base MMI)	84
Figure 4.20	Détection du visage (Base CK+)	84
Figure 4.21	Extraction des points stratégiques	85
Figure 4.22	Estimation du mouvement des points stratégiques	85
Figure 4.23	Les AUs associées à l'émotion de la peur	86
Figure 4. 24	Résultats de comparaison du système proposé avec certaines méthodes de reconnaissance des émotions (unité%)	90
Figure 4.25	Les points stratégiques de l'émotion d'intérêt	96
Figure 4.26	Quelques images de la base de données LFW	98
Figure 4.27	La détection de l'intérêt	98

Introduction générale

Introduction

L'informatique affective est le concept que l'ordinateur soit intelligent à comprendre, reconnaître et interpréter l'émotion humaine. Savoir si une machine pourrait un jour ressentir des émotions est une question posée et discutée depuis longtemps, sans qu'une réponse définitive n'ait été apportée. L'expression de l'émotion constitue une dimension importante en communication. Elle permet de définir une relation particulière avec le sentiment d'un individu, de la mettre à jour, de telle sorte qu'une personne peut, de façon fiable, à partir de comportements verbaux ou non verbaux (tels que l'attitude corporelle, les expressions faciales) de déduire l'état affectif actuel d'une autre personne avec laquelle elle interagit.

Chacun de nous sait ce qu'est une émotion jusqu'à ce qu'on lui demande d'en donner une définition et à ce moment-là, il semble que plus personne n'en est capable. Malgré cela, on trouve dans la littérature plusieurs définitions. Nous avons choisi une, qui semble universelle et définissant l'émotion comme une réaction psychologique et physique face à une situation bien déterminée. L'état affectif est lui-même défini comme étant une réaction organique consciente provoquée par un phénomène interne ou externe.

Depuis une vingtaine d'années, la modélisation informatique des émotions constitue une thématique de plus en plus reconnue dans le domaine de l'interaction homme-machine. Les découvertes issues de la neurophysiologie et de la psychologie, qui établissent un lien ténu entre émotion, rationalité et prise de décision, ont accentué l'intérêt pour la prise en compte des émotions dans différents domaines d'application. Actuellement, il y a un vrai engouement autour des technologies permettant de reconnaître les actions réalisées par les humains à l'aide des systèmes avancés des caméras ou autres types de capteurs.

L'émotion est accompagnée de troubles physiologiques. Ces troubles peuvent s'exprimer à travers différents canaux de communication, ce qui explique la multi-modalité des expressions d'émotion. En effet, chaque état émotionnel éprouvé est interprété par un ensemble de comportements verbaux ou non verbaux dont les indicateurs sont soit les mots utilisés, soit le ton de la voix, les gestes et attitudes corporelles, soit finalement les expressions faciales.

Jusqu'à présent, et à notre connaissance, les travaux portant sur la prise en compte de l'expression de l'émotion dans l'interaction homme-machine se sont principalement focalisés

sur la mise en place de modèles de caractérisation automatique :D'une part, de l'expression vocale de l'émotion par analyse acoustique d'indicateurs essentiellement intonatifs à savoir ton, timbre, rythme de la voix, ...etc. D'autre part, l'expression textuelle par analyse de plusieurs indices de type lexical, morphologique, syntaxique, sémantique et dialogique. On trouve aussi des modèles basés sur l'expression physiologique en analysant certains indices d'ordres physiologiques et expressifs à savoir la pigmentation de la peau, le rythme respiratoire et d'autre basés sur l'expression faciale en analysante plusieurs indices d'ordre physiologique à savoir les traits du visage, le regard, le maintien de la posture.

La recherche actuelle sur l'analyse de l'expression faciale tourne autour de la reconnaissance de six expressions faciales globalement connues en utilisant des caractéristiques géométriques ou basées sur l'apparence. Ces expressions de base sont la joie, la tristesse, la surprise, la peur, la colère et le dégoût, telles que définies par Ekman en 1976.

Les méthodes basées sur l'apparence traitent la texture du visage en réduisant sa complexité et la rendant appropriée pour la comparaison avec des modèles. Choisir les bonnes méthodes de prétraitement et l'ampleur de la réduction de la dimensionnalité représentent des étapes cruciales dans cette classe de méthodes. Les caractéristiques géométriques proviennent de régions distinctes du visage, telles que les yeux, les narines et la bouche. L'ensemble choisi de points caractéristiques peut être suivi, et classé avec les réseaux de neurones (NN), les modèles de Markov cachés (HMM), ainsi que de nombreux autres algorithmes de classification.

Contributions

Comme première contribution de notre thèse, nous avons présenté un nouveau système de reconnaissance des émotions de base dans une séquence d'images. Cette dernière est constituée d'une image neutre et une deuxième qui contient l'émotion.

Notre approche de reconnaissance suit les étapes suivantes :

1. La détection de visage dans une séquence d'images.
2. L'extraction des caractéristiques faciales
3. La classification

Dans notre travail, nous avons atteint un taux de reconnaissance de 95%. Nous avons pu optimiser le temps de calcul et les ressources utilisées, avec la manipulation d'un vecteur de dix éléments représentant l'émotion.

Dans une deuxième contribution de cette thèse, nous avons présenté un système de représentation de l'émotion d'intérêt, vu son importance dans divers domaines, précisément le e-learning. Notre modèle proposé est basé sur des fondements psychologiques, avec une approche de reconnaissance de cette émotion.

Organisation de la thèse

Dans le cadre de cette thèse, nous présentons un système de reconnaissance des émotions de base, qui sont la joie, la tristesse, la peur, la colère, le dégoût et la surprise dans une séquence d'images. Cette reconnaissance est basée sur une approche de détection de mouvement.

Dans le premier chapitre intitulé définitions de l'émotion, nous présentons l'aspect psychologique de l'émotion, ces différentes approches conceptuelles, et les différents systèmes de représentation des émotions.

Le deuxième chapitre est réservé pour la reconnaissance des émotions, dans lequel nous présentons les différentes approches de reconnaissance des expressions faciales et les travaux existants dans ce domaine.

Le troisième chapitre introduit les réseaux de neurones, l'aspect formel et biologique. Les principes des différents réseaux de neurones existants, et l'interaction réseau de neurone/reconnaissance des émotions.

Le quatrième chapitre présente nos contributions, il s'agit d'une nouvelle approche de reconnaissance des émotions basée sur l'estimation du mouvement et la classification par une carte auto-organisatrice de KOHONEN, ainsi que la présentation d'une étude de cas de l'émotion d'intérêt.

Chapitre 1

L'émotion

1.1 Introduction

Qu'est-ce qu'une émotion ? Cette question n'est pas seulement une interrogation scientifique ; mais c'est aussi le titre d'un célèbre article sur l'émotion écrit en 1884: « *What is an emotion?* ». [James, 1884]

Il existe plus d'une centaine de définitions différentes du mot « émotion » et le seul point sur lequel tout le monde s'accorde c'est que le concept est difficile à définir. Une première difficulté vient probablement du fait qu'il s'agit d'un mot du langage populaire. De plus, la définition varie en fonction du point de vue théorique adopté par les différents auteurs.

L'émotion (du latin *motio* « action de mouvoir, mouvement ») est une expérience psychophysologique complexe de l'état d'esprit d'un individu lorsqu'il réagit aux influences biochimiques (interne) et environnementales (externe). Chez les humains, l'émotion inclut fondamentalement "un comportement physiologique, des comportements expressifs et une conscience" [Myers et David, 2004]. Dans le Dictionnaire de l'Académie française (cinquième édition, 1798), on trouve la définition suivante à l'entrée « émotion » :subs. fém. Altération, trouble, mouvement excité dans les humeurs, dans les esprits, dans l'âme. [DAF]

Dans ce chapitre nous présentons les différentes aires psychologiques de l'émotion, ainsi que les différentes manières de représentation des émotions dans le domaine de la reconnaissance émotionnelles.

1.2 Vers une définition de l'émotion

1.2.1 L'émotion en philosophie

En philosophie l'émotion peut se définir principalement comme une manifestation de la vie affective, généralement accompagnée d'un état de conscience agréable ou pénible. L'émotion est un trouble de durée variable, une rupture d'équilibre. Le trouble est parfois violent, et entraîne une augmentation des mouvements (colère, enthousiasme), ou, au contraire, un arrêt des mouvements (peur ou "coup de foudre" en amour). L'émotion agit

donc tantôt à la façon d'un excitant, tantôt à la façon d'un stupéfiant. Les répercussions sur le corps peuvent aller jusqu'à la syncope, mais se limite le plus souvent à des manifestations physiques minimales (rougissement, pâleur, ...). [Dico Philo,2016]

Kant : "L'émotion est le sentiment d'un plaisir ou d'un déplaisir actuel qui ne laisse pas le sujet parvenir à la réflexion. Dans l'émotion, l'esprit surpris par l'impression perd l'empire sur lui-même" (Anthropologie du point de vue pragmatique). [Kant, 1996]

Sartre : "Une émotion est une transformation du monde" [Jean- Paul, 1938]

Alain : L'émotion est un régime de mouvement qui s'établit dans le cœur sans la permission de la volonté, et qui change soudainement la couleur des pensées". [Alain, 1900].

1.2.2 L'émotion en psychologie

En psychologie, une émotion est une réaction à un stimulus affectif, environnemental ou psychologique. L'émotion ressentie par rapport à une situation est propre à chaque individu, à son passé et son histoire de vie, à ses capacités intellectuelles, et à son état psychologique. Les émotions impliquent des répercussions physiques du ressenti psychologique initial : la tristesse peut provoquer les larmes, la peur peut déclencher un cri, ou une perte urinaire parfois, et la joie peut générer un grand sourire, voire même des larmes. Une même situation implique des émotions différentes suivant l'individu concerné, le contexte et l'implication. L'émotion est à différencier de la sensation qui fait appel à une perception sensorielle, et du sentiment qui est un affectif ressenti, mais en l'absence de manifestation physique.

Toutefois, malgré ces difficultés, une définition semble s'être imposée. En 2013, David et al. considèrent que l'émotion est « un processus rapide, focalisé sur un événement et constitué de deux étapes : un mécanisme de déclenchement fondé sur la pertinence de l'événement (par exemple, l'événement est-il pertinent pour mes buts ?), et une réponse émotionnelle à plusieurs composantes (les tendances à l'action, les réactions du système nerveux autonome contrôlant par exemple le rythme cardiaque, les expressions et les sentiments) ». [David, 2013]

1.3 Les composantes de l'émotion

Une émotion est brève et est toujours déclenchée par un événement spécifique, mais, d'après cette définition, c'est aussi un phénomène dynamique qui présente de multiples composantes.

Kleinginna et al, ont analysé et classifié 92 définitions, en montrant qu'il y a peu de points communs entre elles et que la plupart sont trop vagues. Ils ont suggéré une définition opératoire que nous utilisons ici : « Les émotions sont le résultat de l'interaction de facteurs subjectifs et objectifs, réalisés par des systèmes neuronaux ou endocriniens, qui peuvent :

- Induire des expériences telles que des sentiments d'éveil, de plaisir ou de déplaisir ;
- Générer des processus cognitifs tels que des réorientations pertinentes sur le plan perceptif, des évaluations, des étiquetages ;
- Activer des ajustements physiologiques globaux ;
- Induire des comportements qui sont, le plus souvent, expressifs, dirigés vers un but et adaptatifs.

Dans cette définition, on peut noter la présence de trois composantes fondamentales de l'émotion. [Kleinginna P et Kleinginna A, 1981]

1.3.1 La composante comportementale :

Renvoie, quant à elle, à toutes les manifestations comportementales (fuite vs. approche) et expressives d'une émotion, qui sont donc dirigées vers l'extérieur. On peut ainsi citer les postures ou la tonalité de la voix. Toutefois, la composante comportementale la plus étudiée est sans doute l'expression faciale, considérée comme le canal majeur de communication émotionnelle tout simplement parce que le visage est une source considérable de « contact » à autrui.

1.3.2 La composante physiologique

Réfère à l'ensemble des manifestations physiologiques concomitantes à un événement émotionnel. On distingue souvent celles liées au système endocrinien, celles du système nerveux autonome, et celles de l'activité cérébrale.

1.3.3 La composante cognitive/subjective

Correspond aux changements d'état mental liés à l'émotion. En d'autres termes, cette composante désigne la composante subjective de l'expérience émotionnelle.

Pour illustrer l'interaction entre ces composantes, reprenons le célèbre exemple du psychologue William James :

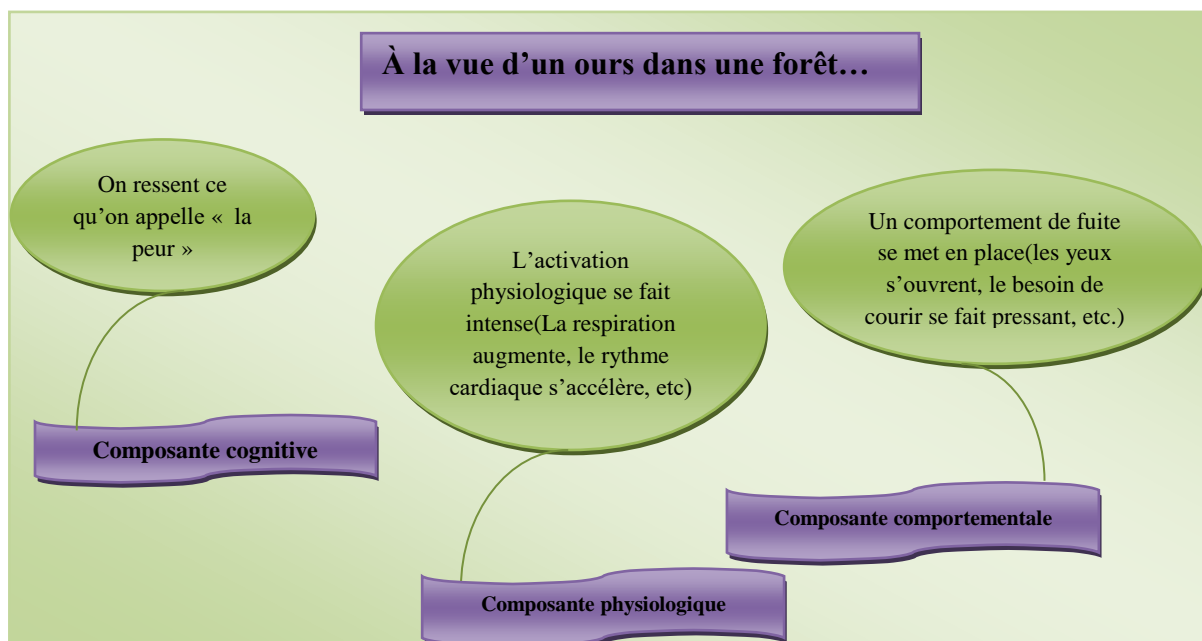


Figure 1.1 : Schéma illustratif des composantes de l'émotion

1.4 Emotions et approches conceptuelles

1.4.1 La théorie périphérique

James, concernant les émotions fortes : «*Ma thèse est que les modifications corporelles suivent directement la PERCEPTION du fait excitant ; et que notre conscience de ces modifications, à mesure qu'elles se produisent, EST l'émotion.* ». [James, 1884]

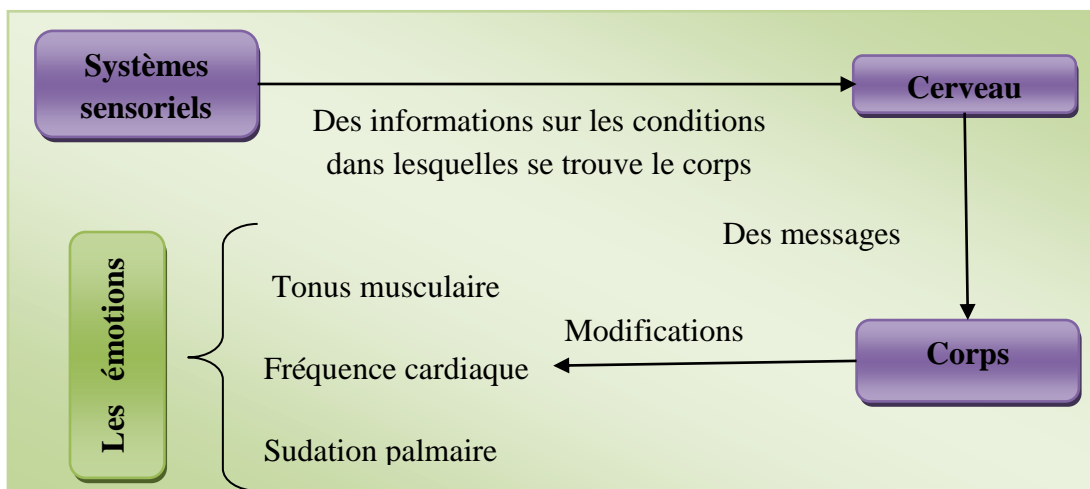


Figure 1.2 : Conceptions des émotions selon la théorie périphérique

1.4.2 Théorie centraliste :

En 1927, Walter CANNON proposa une nouvelle explication qui fut développée par Philip BARD. Selon la proposition dite cette fois « centraliste » de l'émotion, défendue par Cannon (1927) et Bard (1928), le déclenchement d'une émotion spécifique est déterminé par le traitement d'un stimulus au niveau du système nerveux central, le pattern d'activation périphérique n'étant ni spécifique ni causal.

Cette théorie met donc en avant l'importance du système nerveux central, et en particulier du thalamus, dans le déclenchement d'une émotion donnée. Ainsi, les changements

physiologiques ne sont pas conçus comme cause mais comme conséquence de l'émotion. Cannon a émis de nombreuses objections à la théorie de James-Lange et réalisé de nombreuses études empiriques visant à la mettre en défaut ; à titre d'exemples, il observe que des réactions viscérales semblables diffuses se produisent dans de nombreuses émotions, mais aussi dans des états non émotionnels (comme la digestion ou la fièvre) ; il rapporte encore que la suppression des afférences viscérales ne supprime pas les émotions. [Sofiane, 2011]

L'importance du débat James-Lange/Cannon-Bard pour la prise de conscience du rôle de la cognition dans l'émotion est très bien illustrée par les travaux menés par Schachter, qui figurent parmi l'une des contributions pionnières les plus influentes dans le champ des sciences affectives.

1.4.3 La théorie bi-factorielle de Schachter :

Selon la proposition dite « bi-factorielle » de Schachter (1964), une émotion est déterminée par une interaction entre deux composantes : une activation physiologique (arousal) : Le fait de dire « je suis en colère » correspond aux variations physiologiques et une cognition concernant la situation déclenchante de cette activation physiologique et c'est l'application des processus cognitifs, c-à-d. l'individu qui cherche une cause de ce qu'il ressent, ces causes en relation avec l'action physiologique sont les émotions.

Cette théorie est en accord avec Cannon-Bard car les changements physiologiques ne sont pas spécifiques à une émotion et elle est également en accord avec James-Lange car une activation physiologique est nécessaire pour la production d'une émotion. Schachter est l'un des pionniers dans le sens où cet auteur a une approche cognitive des émotions.

L'émotion est donc déterminée comme l'existence d'une cognition modulée par l'activation physiologique. La théorie bifactorielle utilise la dimension sociale (informations disponibles dans l'environnement). [Plutchik et Albert, 1962]

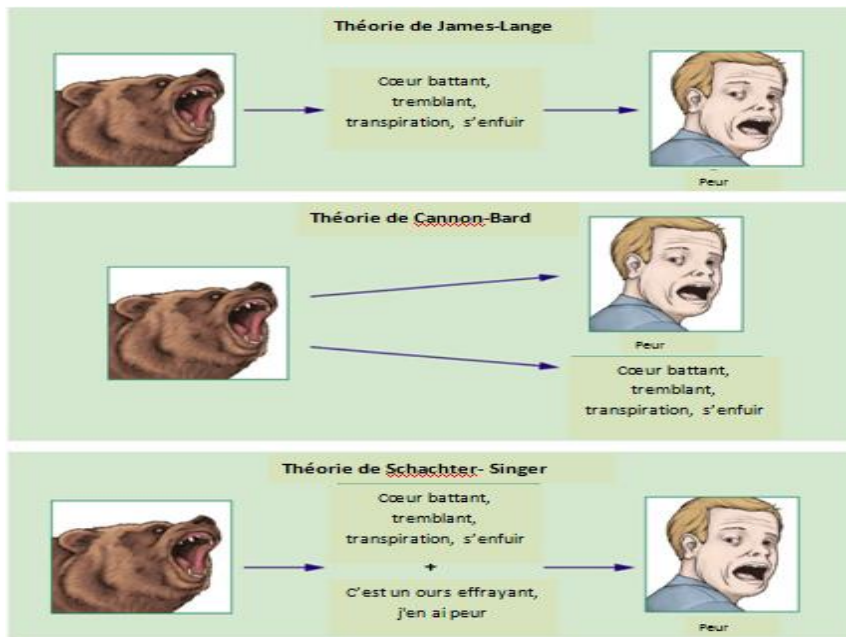


Figure 1.3 : James-Lange VS Cannon-Bard Vs Schachter

1.5 Emotions de bases et émotions secondaires

A partir du classement de photographies de visages, de dessins, ou encore d'analyse vocale exprimant des émotions, les psychologues ont distingué des émotions dites de base participant à la survie collective et individuelle des individus et d'autres dites secondaires considérées comme résultat d'un mélange des émotions de base. [Jacques, 2015]

Les émotions de base, dites également émotions primaires, sont universelles et innées. Elles s'expriment par des manifestations viscérales qui sont donc invisibles, mais également par des manifestations faciales que chacun est en capacité de reconnaître et qui transmettent à notre interlocuteur des ressentis internes. En effet, ces expressions seront les mêmes entre plusieurs personnes de culture, d'âge ou de sexe différents. Selon les courants psychologiques, il n'y a pas le même nombre d'émotions de base.

	WOODWORTH ET SCHLOSBERG (1964)	IZARD (1977)	SCHWARTZ ET SHAVEN (1987)	EKMAN ET FRIESENT (1995)
JOIE	X	X	X	X
SURPRISE	X	X	X	X
PEUR	X	X	X	X
COLERE	X	X	X	X
TRISTESSE	X	X	X	X
DEGOUT	X	X		X
MEPRIS	X	X		
DETRESSE	X	X		
INTERET		X		
CULPABILITE		X		
HONTE		X		
AMOUR	X	X	X	

Table 1.1 : Emotions de base selon chaque auteur

Les autres émotions (comme la fierté, la honte, la satisfaction,...) sont appelées émotions secondaires par opposition aux émotions de base. Cette distinction est due au fait qu'elles n'ont pas tout à fait le même fonctionnement. Contrairement aux émotions de base, elles ne sont pas systématiquement automatiques et spontanées. Leur déclenchement n'est pas forcément rapide et leur action peut durer dans le temps.

1.6 La représentation des émotions

Les représentations faciales relient les mouvements des muscles faciaux aux émotions exprimées. Elles représentent une sorte de dictionnaire utile à la reconnaissance des émotions. Dans cette section nous exposons trois catégories de représentations des expressions faciales émotionnelles.

1.6.1 Représentation par le système MPEG4 :

Le standard de compression MPEG-4 définit 66 paramètres d'animation faciale (FAPs) de bas niveau, issus de l'étude des actions faciales minimales, assez proches des actions des muscles. Les FAPs sont définis en fonction de deux caractéristiques : un ensemble de points clef permettant de définir la forme du visage (FPs) ainsi que des unités (Face

Animation Parameter Units- FAPUs) permettant de normaliser les déformations entre les sujets. L'information d'expression est alors contenue dans un ensemble de distances et angles normalisés entre les points caractéristiques. Par exemple le FAP numéro 3 (mâchoire ouverte) correspond au déplacement de FP2.1 (bas du menton) vers le bas exprimé en FAPU MNS (séparation entre la bouche et le nez). Ainsi un FAP numéro 3 à 0.5 indique que le menton a bougé vers le bas par rapport au visage neutre d'une distance équivalente à la moitié de la distance séparant la bouche du nez [Pandzic et Forchheimer, 2003]. Le logiciel face API [Machines, 2009] permet d'extraire ce type d'information.

Ce système est défini essentiellement pour compresser des données. C'est pourquoi, certains systèmes utilisent cette notation et ces informations de distances et angles afin de réduire la dimensionnalité de l'espace [Hupont et al, 2013].

Même si le visage neutre est pris en compte, on peut se demander si la représentation est suffisamment précise pour l'analyse comparative des expressions mélangées entre les sujets.

1.6.2 Représentation par catégories :

Une autre façon de décrire une expression faciale consiste à lui donner un label. Cette catégorisation a eu un franc succès lorsque la communauté d'analyse d'image s'est emparée des travaux d'Ekman indiquant l'universalité des expressions faciales correspondant aux six émotions de base (joie, sourire, peur, dégoût, tristesse et colère). Le label utilisé est d'ailleurs souvent l'émotion correspondante (joie) et non la déformation réelle observée (sourire, plissement des yeux). Dans ce type de représentation, la caractérisation d'une expression est sa sémantique et non la déformation réelle observée.

Cette représentation des expressions a atteint de très bons taux de reconnaissance sur des personnes connues (sujet présents dans la base d'apprentissage) comme le montre le challenge récent de Valstar et al, avec les meilleurs scores atteignant 100% pour la classification en catégories. Néanmoins, les taux de reconnaissance chutent fortement dès lors que le sujet n'est pas présent dans la base d'apprentissage (meilleur score à 75.2% - moyenne sur les 5 catégories). [Valstar et al 2011]

Ces résultats montrent que la généralisation à différentes morphologies n'est pas bien réalisée par les classificateurs. De nombreuses déformations différentes (notamment d'intensité variable) se retrouvent dans la même catégorie. C'est pourquoi cette représentation est peu à peu laissée.

1.6.3 Représentation par le système : Facial Action Coding System (FACS)

a. Définition

FACS est une méthode permettant de décrire les mouvements faciaux observables. Cette classification est le fruit des recherches des psychologues Ekman et Friesen dans les années 70 [Paul Ekman et Friesen, 1978]. Une première édition est parue en 1978 et une importante mise à jour a été effectuée en 2002. Le Facial Action Coding System est à ce jour la méthode la plus largement reconnue et utilisée pour coder les mouvements du visage humain [CNV, 2013]. Il consiste de manière approfondie à connaître l'anatomie et le fonctionnement de l'ensemble des muscles du visage. Il établit un lien entre les différents muscles de la face et les expressions menant à l'interprétation d'émotions permettant de cerner très précisément l'état d'esprit de la personne (joie, colère, peur, douleur, mépris...).

FACS permet de décomposer les mouvements faciaux en unité d'action, il définit 46 Action Units (AUs), qui sont autant de contraction ou de relaxation d'un ou plusieurs muscles, et dont l'association définit une expression faciale. La philosophie de base du système consistait à former des experts pour la reconnaissance et l'interprétation des AUs, mais désormais, le système est aussi utilisé pour automatiser la reconnaissance des AUs et donc des expressions, ainsi que pour la simulation graphique de visages. Parmi les AUs, certaines sont dites « fiables », ce qui signifie que leur mise en œuvre volontaire est difficilement réalisable. Pour qu'une AU soit fiable, il faut que moins de 25% de la population soit capable de réaliser volontairement et correctement l'AU au premier essai, et sans mobiliser d'autres AUs.

b. Description des Unités d'Action

Les "Action Units" sont divisées en 7 groupes suivant la partie concernée du visage:

AUs	Signification	Partie du visage concernée	
AU1	Remontée de la partie interne des sourcils.	Partie supérieure du visage, décrivent le mouvement des sourcils, des paupières et des joues	
AU 2	Remontée de la partie externe des sourcils.		
AU4	Abaissement et rapprochement des sourcils.		
AU 5	Ouverture entre la paupière supérieure et les sourcils		
AU 6	Remontée des joues		
AU 7	Tension de la paupière		
AU 10	Remontée de la partie supérieure de la lèvre		La partie inférieure du visage et ayant un mouvement vertical décrivant le mouvement vertical du nez, des lèvres, du menton et de la bouche.
AU 15	Abaissement des coins externes des lèvres		
AU 16	Ouverture de la lèvre inférieure		
AU 17	Élévation du menton		
AU 25	Ouverture de la bouche et séparation légère des lèvres		
AU 26	Ouverture de la mâchoire		
AU 27	Bâillement		
AU 9	Plissement de la peau du nez vers le haut.	La partie inférieure du visage et ayant un mouvement horizontal	
AU 20	Étirement externe des lèvres		
AU 14	Plissement externe des lèvres		
AU 11	Ouverture du nasolabial.	La partie inférieure du visage et ayant un mouvement oblique	
AU 12	Étirement du coin des lèvres		
AU 13	Étirement et rentrée des lèvres		
AU 18	Froncement central des lèvres	La partie inférieure du visage et ayant un mouvement orbital	
AU 22	Tension du cou		
AU 23	Tension refermante des lèvres		
AU 24	Lèvres pressées (pincement des lèvres)		
AU 28	Succion interne des lèvres		
Les AUs divers			Décrivent plusieurs parties différentes de la tête par exemple la langue ou la nuque
Les AUs du mouvement de la tête et des yeux		Décrivent les mouvements possibles de la tête (par exemple la tête tournée vers la droite AU 52), ainsi que les mouvements des yeux	

Table 1.2 : description des unités d'action définies par Ekman

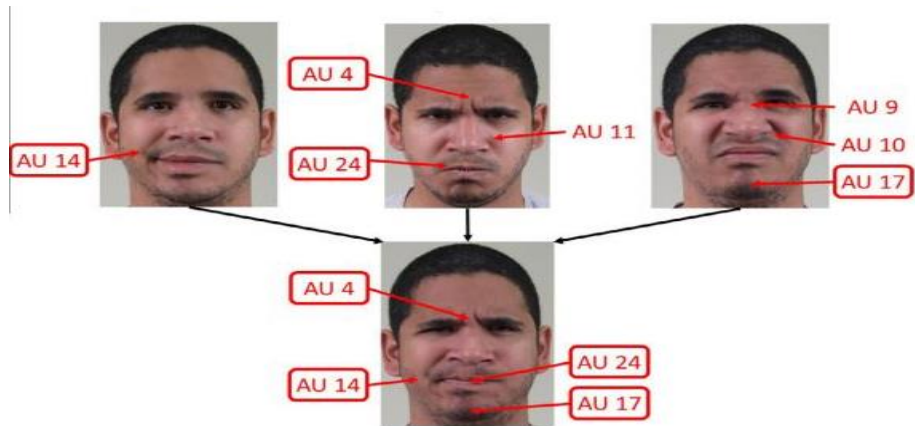


Figure 1.4 : Quelques unités d'action

c. Combinaison de plusieurs AUs :

A partir d'une série de code FACS, il est possible de reconstituer l'expression faciale codée, et donc d'y attribuer un sens. Par exemple, les expressions faciales émotionnelles ont fait l'objet de nombreuses études en lien avec le codage FACS. Ekman a défini six classes fondamentales d'expression faciale, à savoir la colère, le dégoût, la peur, le bonheur, la tristesse et la surprise, qui sont couramment utilisées par les chercheurs travaillant dans ce domaine.

Pour chaque expression faciale, il existe donc des prototypes de ces expressions faciales émotionnelles.

Voici les prototypes des 6 émotions de bases :

Emotion	Unité d'action
Joie	6+12
Tristesse	1+4+15
Surprise	1+2+5+26
Peur	1+2+4+5+20+26
Colère	4+5+7+23
Dégoût	9+15+16

Tableau 1.2 : Relation entre les AUs et les émotions

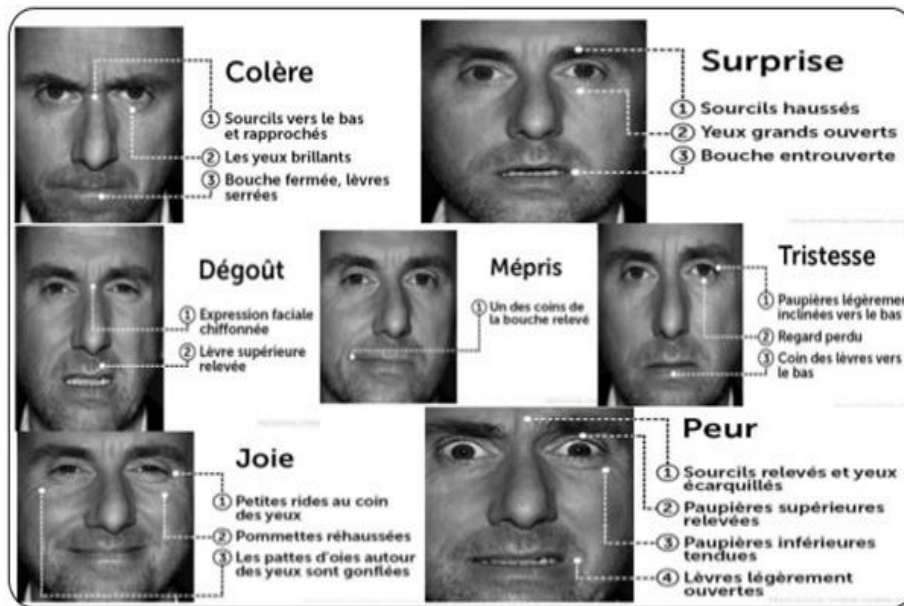


Figure 1.5 : Codage des émotions de base par les unités d'action

d. FACS et la reconnaissance des émotions :

On trouve dans la littérature des travaux portant sur l'utilisation des systèmes exploitant les composants du visage afin de modéliser et suivre les changements dans l'expression du visage. On note que le nombre de combinaisons possibles est de l'ordre de 7000, donc même si les UAs sont correctement reconnues, c'est difficile d'inférer à partir de ces combinaisons l'émotion prévue. Cependant, la reconnaissance précise des AUs est un progrès très important.

Wojdel et al., ont proposé une approche utilisant la logique floue pour définir les dépendances entre les AUs [Wojdel et al, 2003]. D'autres, proposent par exemple un système en deux couches. La première couche permet de réaliser des transitions lisses pour chaque canal d'expression. Six canaux d'expressions sont définis (signaux physiologiques, interactions sociales, émotions, direction du regard et position de lat tête). La seconde couche permet ensuite la combinaison de mouvements de ces différents canaux par la résolution de conflits entre les muscles. En effet, ces différents canaux peuvent porter sur des muscles identiques (par exemple, certains muscles des lèvres sont sollicités à la fois pour le sourire et pour la parole [Bui et al, 2004]. Tong et al. ont proposé une nouvelle approche qui prend en

compte systématiquement les relations entre les AUs. Plus précisément, ils ont utilisé un réseau bayésien dynamique (DBN) pour modéliser les relations entre les différents AUs, mais ils n'ont pas exploité ces AUs pour une reconnaissance de l'émotion ou autres applications d'interaction homme machine [Tong et al, 2007]. Ruan et al, ont exploité le système FACS pour construire un système de reconnaissance des émotions automatique [Ryan et al, 2009]. Yongmiana a proposé un travail qui explore l'utilisation de la technique de fusion d'information multi-sensorielle avec des réseaux bayésiens dynamiques (DBN) pour la modélisation et la compréhension des comportements temporels des expressions faciales dans les séquences d'images [Yongmian, 2005]. Valstar et Pantic ont utilisé (Support Vector Machines : SVM) basé sur des caractéristiques géométriques simples extraites de façon automatique à partir des points du visage. Une fois segmenté, les unités d'action sont déduites. Cependant, cette approche fournit un résultat acceptable pour un temps de calcul un peu élevé [Valstar et Pantic, 2007].

D'après l'ensemble des travaux ci-cités, nous remarquons que l'analyse des expressions faciales basées sur les unités d'actions fournit des résultats meilleurs et un temps de calcul acceptable.

Aussi le système FACS en tant qu'un système développé par les psychologues nous présente une description très proche de l'état affectif, plus précisément la description des émotions de base. L'automatisation de cet esprit psychologique fait l'objectif de notre approche.

Pour cette raison, nous avons adopté la stratégie basée sur les unités d'actions pour le développement de notre système de reconnaissance émotionnelle.

1.7 Conclusion :

Dans ce chapitre nous avons présenté la partie théorique de l'émotion. Ces différentes définitions données par des psychologues et des philosophes.

Nous avons aussi présenté les différents systèmes de représentation des émotions avec une description détaillée du système FACS utilisé dans notre travail.

Nous pouvons constater que les émotions peuvent être caractérisées par :

- les réactions expressives, comme le sourire, le froncement de sourcil, l'intonation de la voix, la posture ;
- les réactions physiologiques comme la fréquence cardiaque, le flux sanguin, la production des larmes ;
- les tendances à l'action et les réactions comportementales, comme l'attaque, l'évitement, la fuite, la recherche de support social ;
- les évaluations cognitives comme la pensée d'avoir été injustement traité par une autre personne ;
- et enfin par l'expérience subjective (ou sentiments subjectifs), c'est-à-dire ce qu'on pense ou on dit ressentir.

Ces descriptions ont amené à proposer plusieurs définitions de l'émotion et à définir des modèles de représentation des émotions qui ont facilité la tâche de l'automatisation dans le domaine de la reconnaissance des émotions.

Chapitre 2

La reconnaissance des émotions

2.1 Introduction

La reconnaissance des émotions faciales (Facial Expression Recognition FER) est un sujet important dans le domaine de la vision par ordinateur et de l'intelligence artificielle, en raison de son important potentiel, académique et commercial. Ce chapitre se concentre sur les études qui utilisent exclusivement des images faciales, car les émotions faciales sont des facteurs importants dans la communication humaine qui nous aident à comprendre l'intention des autres. En général, les gens déduisent les états émotionnels d'autres personnes, tels que la joie, la tristesse et la colère, en se basant sur des expressions faciales. [Kaulard et al, 2012]

Selon la théorie de rétroaction faciale, l'émotion est l'expérience des changements dans nos muscles faciaux. En d'autres termes, ce sont les changements dans nos muscles faciaux qui dirigent nos cerveaux et constituent la base de nos émotions. Comme il ya de nombreuses possibilités de configurations musculaires dans notre visage, il ya aussi un nombre apparemment très grand d'émotions.

L'analyse de des expressions faciales a joué un rôle majeur dans la recherche sur les émotions depuis les travaux de Tomkins dans les années 1960 [Tomkins, 1962]. Ce canal de communication occupe une place primordiale dans la communication non-verbale. Nos visages véhiculent des informations riches qui constituent des expressions de communication (attitude, opinion, humeur, etc.). Elles résultent d'un mouvement ou d'un positionnement des muscles du visage. Celui-ci dispose de 44 muscles différents et le nombre d'expressions faciales est supérieur à 250 000 [Abboud, 2004].

La reconnaissance des expressions faciales consiste à classer des déformations faciales en classes abstraites, basées sur l'information visuelle.

2.2 Canaux de communication émotionnels :

Depuis les travaux de Picard en 1997, de nombreux systèmes ont été développés pour la reconnaissance d'émotions. Ils sont basés sur divers canaux de communication (visage, voix, geste, réactions physiologiques et nerveuses) [Zeng, 2009]. Les systèmes de reconnaissance ont donc au fil des années, considéré les différents canaux de communication émotionnelle, en commençant par le visage et la voix, pour ensuite s'intéresser à la gestuelle et enfin aux réactions physiologiques, comme le montre la figure suivante (figure 2.1) :



Figure 2.1 : Canaux de communication émotionnelle et capteurs associés

2.3 La reconnaissance des émotions

Trois domaines principaux où les systèmes de reconnaissance d'émotions se sont développés : La reconnaissance d'émotions à partir de la parole (langage parlé), en utilisant des méthodes statistiques [Sanghamitra, 2016]. Il y a aussi ceux qui utilisent des caractéristiques spectrales telles que le coefficient cepstral de fréquence (MFCC) comme clairement expliqué par [Samal et al, 2013]. La reconnaissance d'émotions à partir de signaux du système nerveux (EEG). [Petrantonakis et Hadjileontiadis, 2009]. En troisième lieu, la

reconnaissance d'émotions à partir d'images fixes ou de vidéos (analyse d'expressions faciales).

Comme nous l'avons déjà signalé, dans le cadre de cette thèse, nous ne prenons en considération que la reconnaissance des émotions à partir des expressions faciales. L'expression faciale est en effet un élément clé dans l'étude et la compréhension des émotions [Ekman, 1993]. Plusieurs chercheurs ont privilégié le visage pour l'étude des états émotionnels. Darwin s'est particulièrement intéressé à l'étude des changements des expressions engendrées par les émotions [Darwin, 1872]. Tomkins, Izard et Ekman ont par ailleurs été parmi les premiers psychologues à s'intéresser aux expressions faciales ; notamment les expressions émotionnelles. Les expressions faciales remplissent plusieurs fonctions [Fasel, 2002]. Elles présentent :

- Des expressions spécifiques à l'état mental telles que : la réflexion, la conviction, l'émotion ressentie, etc.
- La communication non verbale (clins d'œil, grimaces pour passer un message).
- La communication verbale (mouvement de la bouche).
- Des états physiologiques (fatigue, peine...).

2.4 Méthodes de reconnaissance des émotions par les expressions faciales

Du point de vue physiologique, l'essentiel de l'information d'une expression est contenu dans la déformation des traits permanents principaux du visage, à savoir les yeux, le nez, les sourcils et la bouche. On se basant sur ces fondements physiologiques, tous processus de reconnaissance des émotions suit trois étapes fondamentales, en commençant par la détection du visage, ensuite l'extraction des caractéristiques faciales, et enfin une étape de classification pour identifier l'émotion (figure 2.2).

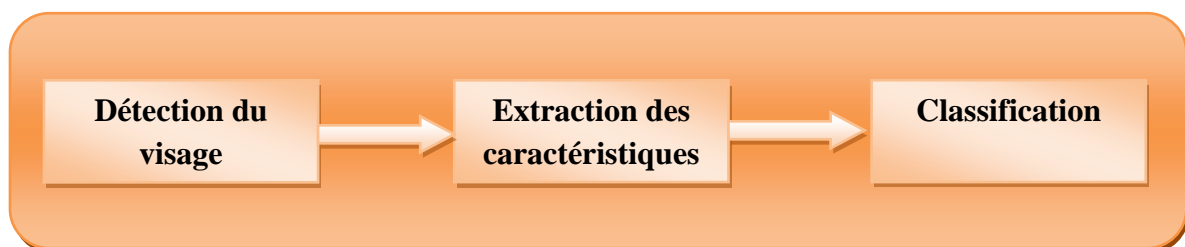


Figure 2.2: Processus d'analyse des Expressions faciales

2.4.1 La détection du visage :

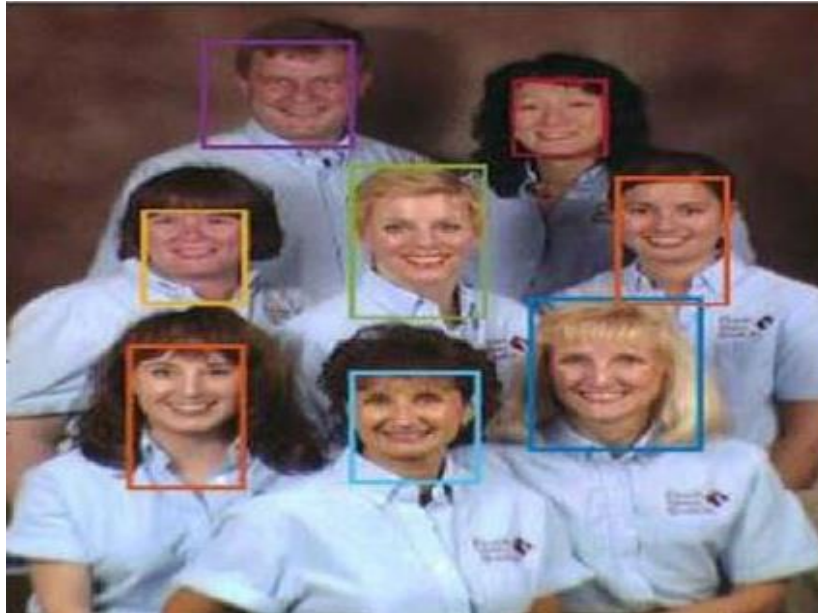


Figure 2.3. La détection du visage

Plusieurs méthodes de détection de visages ont été proposées au cours des dernières années. On peut les classer selon quatre catégories [Shalini et Neeta, 2016] décrites ci-dessous:

1. Méthodes basées sur les caractéristiques invariantes.
2. Méthodes basées sur la mise en correspondance.
3. Méthodes basées sur l'apparence.
4. Méthodes basées sur les connaissances acquises.

a. Méthodes basées sur les caractéristiques invariantes

Cette méthode a pour objectif de trouver les caractéristiques structurelles même si le visage est dans différentes positions, conditions lumineuses ou changement d'angle de vue. Le problème avec cette méthode est que la qualité des images peut être sévèrement diminuée à cause de l'illumination, le bruit ou l'occlusion ce qui empêche l'algorithme de fonctionner correctement. Cependant, Il existe plusieurs propriétés ou caractéristiques invariantes du visage dont les principales sont les suivantes :

✓ *Caractéristiques du visage*

Cette méthode utilise les plans d'arrêtes appelés "Canny detector" et des heuristiques pour supprimer tous les groupes d'arrêtes sauf celles qui représentent les contours du visage. Une ellipse est déduite comme frontière entre l'arrière-plan et le visage. Celle-ci est décrite comme étant formée des points de discontinuité dans la fonction de luminance (intensité) de l'image. Le principe de base consiste à reconnaître des objets dans une image à partir de modèle de contours connus aux préalables. Pour réaliser cette tâche, deux méthodes peuvent être appliquées : la transformée de Hough et la distance de Hausdorff.

✓ *Texture*

La texture de l'être humain est distinctive et peut être utilisée pour séparer les visages par rapport à d'autres objets. Augusteijin et Skufca en 1993 ont développé une méthode de détection de visages sur une image en se basant uniquement sur la texture. Le calcul de la texture se fait en utilisant les caractéristiques de second ordre sur des sous-images. Dans cette méthode, trois types de caractéristiques sont prises en considération : la peau, les cheveux et le reste des composants de visage.

✓ *Couleur de la peau*

Pour les visages, la couleur de la peau de l'être humain est une caractéristique spécifique, c'est pourquoi elle a été utilisée pour la détection des visages. En effet la couleur de la peau est différente selon les personnes et leur origine (Africain, Européen, Asiatique...). Dans ce contexte plusieurs études ont démontré que la plus grande différence s'étend largement entre intensité plutôt que la chrominance lumineuse [BENCHERT et al, 2005].

✓ *Caractéristiques multiples*

Il existe plusieurs méthodes qui combinent les différentes caractéristiques faciales pour la détection des visages. La majorité entre elles adoptent des propriétés globales : couleur de la peau, la forme, la taille du visage, pour trouver les candidats puis les vérifier localement en se basant sur les caractéristiques détaillées comme les yeux, les sourcilles, le nez et les cheveux. On trouve dans le contexte de reconnaissance des émotions, beaucoup de travaux utilisant ce type de méthode. Citons par exemple le travail de [Singh et al, 2003].

b. Méthodes basées sur la mise en correspondance

La détection des visages se fait à travers un apprentissage d'exemples standards de visages ou d'images frontales contenant des visages. La procédure se fait en corrélant les images d'entrées et les exemples enregistrés (gabarits). Le résultat est la décision finale soit de l'existence ou non d'un visage. On trouve deux types de corrélation suivant le type des gabarits :

✓ *Modèle prédéfini*

Les sous-gabarits souvent utilisés pour la détection frontale sont les yeux, le nez, la bouche et les contours du visage. Chaque sous-gabarit est défini par une segmentation des lignes. On commence par une comparaison entre les sous-images et les contours de gabarit pour détecter les visages candidats. Puis on met les sous gabarits comme les yeux et le nez sur les positions candidates. On peut conclure que la détection s'effectue en deux étapes : la première consiste en la détection des régions candidates, la deuxième est l'examen des détails pour déterminer les caractéristiques du visage. Cette méthode est utilisée par Sakai et Al. [Sakai et al, 1972].

✓ *Modèle déformable*

Cette approche est utilisée dans le but de modéliser les caractéristiques faciales qui s'adaptent élastiquement par rapport au modèle du visage présent. Dans cette méthode, les caractéristiques faciales sont décrites par des gabarits paramétrés. Une fonction est définie pour relier les contours, les sommets et les angles dans l'image d'entrée, pour faire correspondre les paramètres sur les gabarits. La meilleure adaptation du modèle élastique est de trouver la fonction énergétique en minimisant les paramètres. Cette méthode a été utilisée par [Kanade et al 2000]

c. Méthode basée sur les apparences

La différence entre cette méthode et celle de la mise en correspondances est que les modèles sont lus à partir des images d'apprentissage qui doivent être représentatives et faites à différentes positions du visage.

Généralement les méthodes d'apparence se basent sur des techniques d'analyse statistiques (pourcentage d'existence des modèles dans l'image) et d'apprentissage automatique pour trouver des caractéristiques significatives des visages et des non visage. Pour cette approche, il existe plusieurs méthodes où chacune d'entre elles se base sur une des caractéristiques du visage ou plus précisément une partie du visage qui peut être interprétée dans le cadre probabiliste.

✓ L'Analyse en Composantes Principales (PCA)

Cette méthode n'a besoin d'aucune connaissance préalable de l'image. Son principe de fonctionnement s'appuie sur la construction d'un sous espace vectoriel ne retenant que les meilleurs vecteurs propres, tout en gardant le maximum d'information utile non redondante, ce qui est très efficace pour réduire la dimension des données et le temps de calcul.

✓ Approches basées sur la distribution

Sung et Poggio ont développés la méthode basée sur la distribution pour la détection des visages. Elle démontre comment la distribution de l'image appartenant à une seule classe d'objet peut être classifiée comme exemple de classe positive ou négative.[Sung, Poggio,1998].

✓ Réseau de neurones

Le réseau de neurones est utilisé pour classifier les pixels de l'image, en tant que visage ou non-visage. Dans toute utilisation de réseaux de neurones, il faut définir une topologie du réseau. La topologie de réseau est déterminée par des tests successifs et il n'y a aucune méthode standard à suivre pour définir la meilleure. Un visage se distingue surtout par des yeux, un nez et une bouche. La topologie de base sera donc une unité finale qui fournit une réponse binaire ou probabiliste. On mettra derrière cette unité les couches cachées du réseau.

L'inconvénient de cette approche réside dans le temps de calcul qui ne permet pas souvent de faire des traitements en temps réel.

✓ Viola et Jones

La méthode de Viola et Jones est une méthode de détection de visages dans une image numérique, proposée par les chercheurs Paul Viola et Michael Jones. Elle fait partie des toutes premières méthodes capables de détecter efficacement et en temps réel des visages dans une image. La méthode de Viola et Jones est une approche basée sur l'apparence, qui consiste à parcourir l'ensemble de l'image en calculant un certain nombre de caractéristiques dans des zones rectangulaires qui se chevauchent. Elle a la particularité d'utiliser des caractéristiques très simples mais très nombreuses. Une première innovation de la méthode est l'introduction des images intégrales (Fig 2.5) qui permettent le calcul rapide de ces caractéristiques (Fig 2.4). Une deuxième innovation importante est la sélection de ces caractéristiques par boosting, en interprétant les caractéristiques comme des classifieurs (Fig 2.5). Enfin, la méthode propose une architecture pour combiner les classifieurs boostés en un processus en cascade, ce qui apporte un net gain en temps de détection.

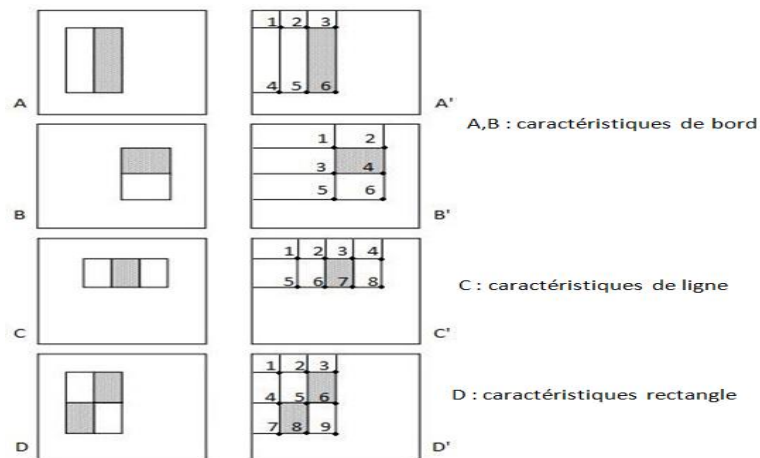


Figure 2.4. Caractéristiques Haar dans l'algorithme original de Viola et Jones

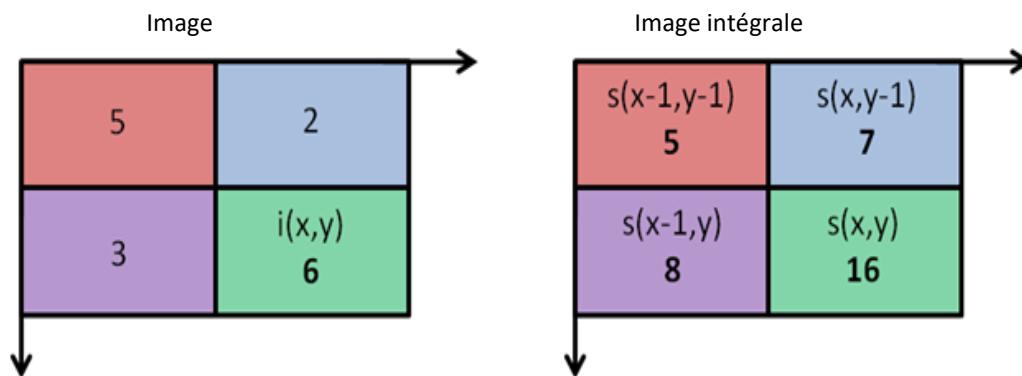


Figure 2.5. L'image intégrale pour le calcul des caractéristiques

Une variété de méthodes a été proposée dans la littérature pour la détection de visages, on se basant sur les méthodes d'apparence [Yang et al, 2004].

d. Méthode basée sur la connaissance :

Ces méthodes sont basées sur des règles strictes à partir des rapports entre les caractéristiques faciales. Elles s'intéressent aux parties caractéristiques du visage comme le nez, la bouche et les yeux. Ces méthodes sont conçues principalement pour la localisation de visage. L'inconvénient de ces méthodes est qu'elles n'arrivent pas à détecter le visage lorsque ce dernier se trouve dans un arrière-plan complexe. Le travail de Kotropoulos et Pitas applique ce type de méthode pour la reconnaissance [Kotropoulos et Pitas, 1997].

2.4.2 L'extraction des caractéristiques :

Après la détection du visage, l'étape suivante consiste en l'extraction des informations nécessaires sur l'expression faciale présentée dans la séquence d'images. Les caractéristiques faciales peuvent être classées en plusieurs catégories. En général, il ya deux types: les caractéristiques géométriques qui se concentrent sur l'extraction des formes, par exemple les yeux, les sourcils, le nez et la bouche, et les caractéristiques d'apparence qui représentent la texture de la peau, du visage, y compris les rides, les renflements et les sillons. Les

mouvements des traits du visage tels que les sourcils, les yeux et la bouche ont une forte relation avec l'information sur les expressions faciales.

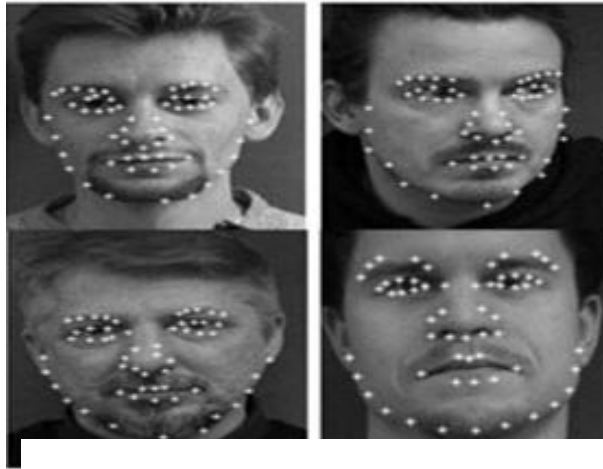


Figure2.6. : Extraction des caractéristiques faciales

Diverses techniques ont déjà été proposées pour l'extraction des caractéristiques faciales. Elles peuvent être classées en 3 grandes familles : analyse bas niveau, analyse intermédiaire et analyse haut niveau.

a. Analyse bas niveau

Les techniques de bas niveau supposent que les pixels de l'objet à segmenter possèdent des caractéristiques homogènes et différentes du fond. Or, la segmentation peut être effectuée par l'identification et la séparation des classes objets et fond. Pour cela, différentes solutions ont été proposées. Certaines utilisent un simple seuillage d'une grandeur colorimétrique, alors que d'autres mettent en œuvre des techniques de classification plus évoluées [Eveno, 2003].

Kapmann et Zhang ont exploité leurs travaux antérieurs qui sont basés sur la corrélation entre l'image originale et des modèles de coins d'œil et de la bouche, pour l'extraction des coins de l'œil afin de les représenter avec un modèle déformable [Kapmann et Zhang, 1998].

b. Analyse intermédiaire

A un niveau d'analyse intermédiaire, on cherche à détecter des caractéristiques indépendantes des conditions lumineuses et de l'orientation des visages. Les contours actifs (ou snakes) introduits par Kass et Witkin à la fin des années 80 sont des courbes qui peuvent se déformer progressivement de manière à s'approcher au plus près du contour d'un objet. Cette déformation est guidée par la minimisation d'une fonction d'énergie.[Kass et al, 1987].

c. Analyse haut niveau

Les méthodes précédentes d'extraction des caractéristiques faciales sont des processus à forme libre. Elles n'intègrent aucune connaissance a priori sur les formes admissibles. Contrairement aux méthodes de haut niveau qui sont basées sur des modèles caractéristiques des formes à segmenter, obtenus de manière heuristique ou statique. Ces modèles génériques sont déformés de manière à être adaptés aux contours de l'objet.

Plusieurs travaux ont utilisé les modèles actifs d'apparence (Active Appearance Models AAM) qui permettent de construire un modèle statique de l'objet à segmenter incluant à la fois la forme et les niveaux de gris. [Xiaoyi et al, 2006][Cheon et Kim, 2008].

2.4.3 La classification :

La classification représente la troisième étape dans le processus de reconnaissance des émotions. Elle classe les caractéristiques extraites du visage dans des classes d'émotions.



Figure2.7. : Classification des émotions de base

a. Modèle de Markov Caché : HMM (Hidden Markov Model)

✓ Description du modèle HMM

Le raisonnement derrière l'utilisation HMM, pour la reconnaissance d'émotion est que le HMM modèle le cerveau humain, comme l'émotion humaine est tout à fait complexe, un instinct humain contient la couche cachée aussi (comme l'esprit conscient secondaire).

Pour utiliser HMM, plusieurs choses telles que la topologie, vecteurs d'observation et les paramètres statistiques doivent être déterminés. Les caractéristiques de chaque prise de vue vidéo (image) doivent être transformées en vecteur d'observation. Une fois que le modèle de topologie et les vecteurs d'observation sont déterminés, la prochaine étape serait de former le modèle par des données de formation et de déterminer la transition d'état initial et les probabilités d'émission.

Les chaînes de Markov sont connues comme des modèles de commutation d'état. L'hypothèse implicite des modèles de commutation entre les différents états. Le comportement de commutation est dirigé par une matrice de probabilité de transition (TPM). Dans l'hypothèse d'un modèle à deux états, le TPM est de la forme:

$$\tau = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \quad (2.1)$$

Où p_{11} désigne la probabilité de rester dans le premier état de la période t à la période $t + 1$ et p_{12} est la probabilité de passage du premier au deuxième état. La deuxième rangée peut être interprétée de façon analogue. [TEOH et CHO, 2011].

L'estimation des paramètres pour le HMM peut se faire en utilisant l'algorithme de Baum-Welch : qui est un cas particulier de l'algorithme espérance-maximisation et utilise l'algorithme Forward-Backward. Il permet de ré-estimer les paramètres de manière itérative. On utilise aussi ICE (Iterated Conditional Estimation) : le vecteur d'observation est calculé à partir des caractéristiques de bas niveau, ensuite le vecteur d'observation sera décodé en la séquence la plus probable d'états cachés par programmation dynamique.

✓ Les HMM et l'émotion

Beaucoup de travaux dans le domaine de la reconnaissance des émotions se sont basés sur des classifieurs HMM. Citons le travail de Chumkamon et Hayashi et de Niu et al. [Chumkamon et Hayashi, 2013] [Niu et al, 2011].

✓ Les problèmes des HMMs :

Il ya trois problèmes canoniques associés aux HMMs:

1. Calculer la probabilité d'une séquence de sortie particulière, nécessite une sommation sur toutes les séquences possibles de l'état.
2. Calculer l'estimation de probabilité maximale des paramètres du HMM est un problème. Aucun algorithme traitable n'est connu pour le résoudre exactement, mais une probabilité maximale locale peut être dérivée de manière utilisant l'algorithme de Baum-Welch [Welch, 2003] ou l'algorithme Baldi-Chauvin. [Baldi et al ,1995]
3. Les HMMs comportent beaucoup de paramètres, ce qui nécessite beaucoup de données d'entraînement pour les estimer.

b. Les réseaux bayésiens

Une des grandes problématiques de notre époque est de traiter la grande quantité des données qui est mise à notre disposition (notamment grâce à l'informatique) pour en extraire de l'information. Il serait donc intéressant d'avoir un (ou plusieurs) modèle(s) effectuant le lien entre les observations et la réalité pour un objectif précis, et cela même lorsque les observations sont incomplètes et/ou imprécises. Les réseaux bayésiens permettent donc de transformer en modèle interprétable la connaissance contenue dans des données.

L'utilisation d'un réseau bayésien s'appelle « inférence ». Le réseau bayésien est alors véritablement une « machine à calculer des probabilités conditionnelles ». En fonction des informations observées, on calcule la probabilité des données non observées. Par exemple, en fonction des symptômes d'un malade, on calcule les probabilités des différentes pathologies compatibles avec ces symptômes. On peut aussi calculer la probabilité de symptômes non observés, et en déduire les examens complémentaires les plus intéressants.

✓ Définition des réseaux bayésiens :

Un réseau bayésien $B = (G, \Theta)$ est défini par : $G = (X, E)$, graphe dirigé sans circuit dont les sommets sont associés à un ensemble de variables aléatoires $X = \{X_1, \dots, X_n\}$, $\Theta = \{P(X_i / Pa(X_i))\}$, ensemble des probabilités de chaque nœud X_i lié conditionnellement à l'état de ses parents $Pa(X_i)$ dans G .

✓ Réseaux bayésiens et émotions :

Notre compréhension scientifique de l'émotion, de la personnalité, et leurs manifestations, tandis que vaste, est loin d'être complète. Cette connaissance incomplète motive l'utilisation des réseaux bayésiens comme un outil de modélisation approprié.

Dans ce cadre, le travail de Jack et Ball illustre cette idée, on a exploité les réseaux bayésiens dynamique : l'architecture utilise des modèles dynamiques d'émotions et de personnalité d'un utilisateur encodés en réseaux bayésiens, et générer un comportement approprié par un agent automatisé. Le système est capable de prédire les états émotionnels actuels et futurs basés sur une histoire d'interaction [Jack et Ball, 1998].

Sebe a proposé une topologie du réseau Bayésien permettant de reconnaître les émotions à partir d'expressions faciales [Sebe, 2005]. Datcu et Rothkrantz utilisent un réseau

Bayésien dynamique dont le modèle de fusion vise à déterminer l'émotion la plus probable du sujet en considérant celles déterminées dans les images précédentes [Datcu et Rothkrantz, 2008]. Cohen et al développent un système basé sur un algorithme de suivi des traits faciaux afin d'extraire les mouvements locaux des composantes faciales. Ces mouvements forment les entrées d'un réseau bayésien utilisé pour reconnaître les six expressions faciales, et la performance était de 86,45%. [Cohen et al, 2003]

Les réseaux bayésiens permettent la prise en compte du temps. Mais une première critique peut être adressée aux modèles probabilistes. S'il est vrai qu'ils permettent de prendre en compte les connaissances a priori dont on dispose sur la nature des dépendances entre les variables considérées, ce que l'on cherche à apprendre c'est une fonction de décision et non pas une loi de probabilité. La seconde critique est liée au problème de variabilité intra et inter utilisateur.

c. Support Vector Machine :

Les Support Vector Machines souvent traduit par l'appellation de Séparateur à Vaste Marge (SVM) sont une classe d'algorithmes d'apprentissage initialement définis pour la discrimination, c'est-à-dire la prévision d'une variable qualitative binaire.

Les SVM peuvent être utilisés pour résoudre des problèmes de discrimination, c'est-à-dire décider à quelle classe appartient un échantillon, ou de régression, c'est-à-dire prédire la valeur numérique d'une variable. La résolution de ces deux problèmes passe par la construction d'une fonction H qui à un vecteur d'entrée X fait correspondre une sortie Y :

$$y = h(x) \tag{2.2}$$

On se limite pour à un problème de discrimination à deux classes (discrimination binaire), c'est-à-dire $y \in \{-1, 1\}$

Dans le domaine de la reconnaissance des émotions, Lu et Evans ont proposé une méthode de transformation en ondelettes de Haar (HWT), et l'ont combiné avec SVM. [Lu et Evans, 2017]. Drume and Jalal ont combiné l'analyse en composantes principales (PCA) avec la machine à vecteurs de support (SVM)[Drume et Jalal, 2012]. Dans le travail de Buciu et al,

les auteurs ont rapporté un taux de performance de 90,34% lorsque les vecteurs caractéristiques ont été obtenus par représentation de Gabor avec une gamme de basses fréquences et la classification a été faite en utilisant un SVM à noyau quadratique.[Buciu et al, 2003]

2.5 Conclusion

Tous système de reconnaissance des émotions, suit les trois étapes fondamentales, qui sont : la détection du visage, l'extraction des caractéristiques faciales et la classification. Ils diffèrent entre eux par les démarches des approches proposées. Commenant par les différentes méthodes de détection de visage, basées soit sur les caractéristiques invariantes, la mise en correspondance, les apparences ou la connaissance.

Ensuite l'étape d'extraction des caractéristiques se fait par une analyse de niveau haut, bas ou intermédiaire.

Enfin, le type de classifieur joue un rôle très important dans la modification du taux de reconnaissance, et l'amélioration des résultats. Nous trouvons dans la littérature les Modèles de Markov Cachés, les réseaux bayésiens, les Support Vector Machin et les réseaux de neurone.

Dans cette thèse, nous nous concentrons sur les classifieurs basés sur les réseaux de neurones. Nous abordons ces derniers en détail dans le chapitre 3

Chapitre 3

Les réseaux de neurones

3.1 Introduction

Comment l'homme fait-il pour raisonner, parler, calculer, apprendre, ...?

Procéder d'abord à l'analyse logique des tâches relevant de la cognition humaine et tenter de les reconstituer par programme. C'est cette approche qui a été privilégiée par l'Intelligence Artificielle et la psychologie cognitive classique. Cette démarche est étiquetée sous le nom de *cognitivism*e. Puisque la pensée est produite par le cerveau ou en est une propriété, commencer par étudier comment celui-ci fonctionne.

Cela a mené à la définition et l'étude de réseaux de neurones formels qui sont des réseaux complexes d'unités de calcul élémentaires interconnectées. Les réseaux de neurones, fabriqués de structures cellulaires artificielles, constituent une approche permettant d'aborder les problèmes de perception, de mémoire, d'apprentissage et de raisonnement. Ils s'avèrent aussi des alternatives très prometteuses pour contourner certaines des limitations des ordinateurs classiques. Grâce à leur traitement parallèle de l'information et à leurs mécanismes inspirés des cellules nerveuses (neurones), ils infèrent des propriétés émergentes permettant de solutionner des problèmes jadis qualifiés de complexes.

Nous aborderons dans ce chapitre les principales architectures de réseaux de neurones que l'on retrouve dans la littérature. Il ne s'agit pas de les étudier toutes, car elles sont trop nombreuses, mais plutôt d'en comprendre les mécanismes fondamentaux.

3.2 Neurones biologiques et neurones formels :

La théorie de l'apprentissage est fortement influencée par la modélisation du cerveau. En effet, sous certains aspects, le cerveau peut être considéré comme une machine capable d'apprendre des tâches souvent très complexes. Il constitue une source d'inspiration pour la compréhension de l'apprentissage. Le cerveau humain est constitué de 100 milliards de neurones, avec un nombre moyen de 10 000 connexions par neurone. Aussi, un mm³ de cortex contient un milliard de connexions.



Figure 3.1. Le cerveau humain

Un neurone est une cellule capable de transmettre des informations à d'autres neurones au travers de ses différentes connexions (synapses). Il existe plusieurs types de neurones (pyramide, panier, Purkinje, etc.) avec des fonctionnements différents (sensoriel, moteur, inter-neurones, etc.) Les neurones sont inter-connectés et forment des réseaux.

On se basant sur la structure du neurone biologique, on s'est inspiré l'idée du neurone formel :

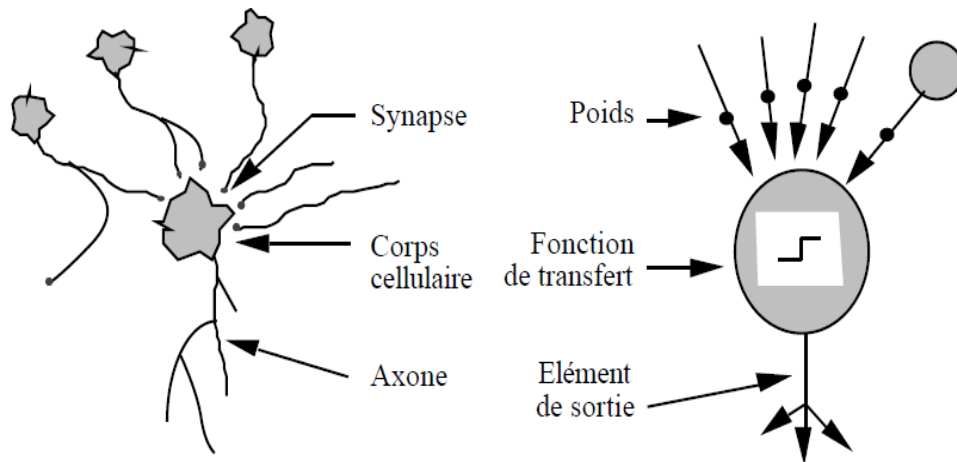


Figure 3.2. Neurone biologique et neurone formel

3.3 Modèle mathématique et architecture des réseaux de neurones

Le modèle mathématique d'un neurone artificiel est illustré dans l'équation 3.1. Un neurone est essentiellement constitué d'un intégrateur qui effectue la somme pondérée de ses entrées. Le résultat de cette somme est ensuite transformé par une fonction de transfert f qui produit la sortie du neurone. En suivant les notations :

Les R entrées du neurone correspondent au vecteur $P = [p_1, p_2, \dots, p_R]^T$, alors que $W = [w_{1,1}, w_{1,2}, \dots, w_{1,R}]^T$ représente le vecteur des poids du neurone. La sortie n de l'intégrateur est donnée par l'équation suivante :

$$n = \sum_{j=1}^R w_{1,j} p_j - b \quad (3.1)$$

Que l'on peut aussi écrire sous forme matricielle :

$$n = w^T P - b \quad (3.2)$$

Le résultat n de la somme pondérée s'appelle le niveau d'activation du neurone. Le biais b s'appelle aussi le seuil d'activation du neurone.

3.4 Apprentissage supervisé et non supervisé

En matière d'apprentissage automatisé, on oppose très fréquemment apprentissage supervisé et apprentissage non supervisé. Malgré que les deux types d'apprentissages relèvent de l'intelligence artificielle, une différence de traitement est claire :

4.1 Apprentissage supervisé

Dans le premier cas, un chercheur est là pour “guider” l’algorithme sur la voie de l’apprentissage en lui fournissant des exemples qu’il estime probants après les avoir préalablement étiquetés des résultats attendus. L’intelligence artificielle apprend alors de chaque exemple, avec pour but, d’être capable de généraliser son apprentissage à de nouveaux cas.

Il existe différentes approches pour apprendre à classer à partir d’un ensemble d’entraînement; on distingue notamment deux approches :

a. Approches basées sur les instances (IBL)

L’apprentissage basé sur les instances est dominé par les approches du type k Plus Proches Voisins (kPPV). Pour un exemple à classer ϵ , il s’agit de considérer l’ensemble des k exemples connus (appartenant à l’ensemble d’entraînement E) les plus proches de ϵ relativement à une mesure de (dis) similarité. La classe prédite pour ce nouvel exemple est alors la classe majoritaire, dans ce voisinage de k exemples. Les difficultés récurrentes, pour ce type de méthode, résident dans le choix du nombre k de voisins à considérer et dans la définition d’une mesure de (dis) similarité appropriée.

b. Approches basées sur des règles(RBL)

Les règles à générer sont de la forme :

Si < condition > alors Classe

Où la condition (aussi appelée complexe ou hypothèse) est une conjonction de sélecteurs. On dira qu’un complexe (resp. un sélecteur) couvre un exemple, si le complexe (resp. le sélecteur) est vrai pour cet exemple. Ainsi, le complexe vide (conjonction de zéro sélecteur) couvre tous les exemples de l’ensemble E. La classe prédite par la règle correspond à la classe majoritaire de l’ensemble des exemples couverts par la condition de la règle. Finalement, la disjonction des règles apprises permet de couvrir l’ensemble des données

d'entraînement. La prédiction de la classe d'un nouvel exemple ϵ est effectuée par la recherche d'une (ou plusieurs) règle(s) dont la condition sera couverte.

4.2 Apprentissage non supervisé

Dans le cas de l'apprentissage non supervisé, l'apprentissage par la machine se fait de façon totalement autonome. Des données sont alors communiquées à la machine sans lui fournir les exemples de résultats attendus en sortie. Différents algorithmes sont associés à l'apprentissage non supervisé

a. La segmentation ou le regroupement (clustering) :

Le principe du clustering est la construction des classes automatiquement en fonction des exemples disponibles. L'apprentissage non supervisé est très souvent synonyme de clustering. Beaucoup d'algorithmes appliquent le clustering, comme :

✓ *K-moyennes*

Une classe est représentée par son centre de gravité, un objet appartient à la classe dont le centre de gravité lui est le plus proche.

✓ *Méthodes à base de densité*

Le principe de base de ces méthodes est l'utilisation de la densité à la place de la distance. Donc les clusters sont des régions de l'espace qui ont une grande densité de points : un point est dense si le nombre de ses voisins dépasse un certain seuil et un point est voisin d'un autre point s'il est à une distance inférieure à une valeur fixée

✓ *Méthodes hiérarchiques*

C'est quand les clusters sont organisés sous la forme d'une structure d'arbre.

✓ *Méthodes à base de grille*

Ce sont des méthodes basées sur le découpage de l'espace des exemples suivant une grille, après initialisation. Toutes les opérations de clustering sont réalisées sur les cellules,

plutôt que sur les données. La construction des classes en rassemblant les cellules voisines en terme de distance.

b. Règles d'association:

C'est le fait d'analyser les relations entre les variables ou détecter des associations. Le principe est l'identification des items qui apparaissent souvent ensemble lors d'un évènement (découverte des corrélations entre attributs), et les règles sont du type

Si X alors Y ou (X->Y)

3.5 Exemple de réseaux de neurones :

3.5.1 Perceptron et perceptron multi couches

Le perceptron est un modèle de réseau de neurones avec algorithme d'apprentissage créé par Frank Rosenblatt en 1958. C'est un modèle utilisé dans l'entraînement supervisé qui met en correspondance des ensembles de données d'entrée dans un ensemble de sortie appropriée. Selon la topologie de réseau utilisé, qui est relatif au nombre des couches, le réseau sera ou non un multi-Layer Perceptron (MLP) [Christian, 2016]. Il est constitué de plusieurs couches de nœuds dans un graphe orienté, chaque couche étant entièrement connectée à la suivante. Pour cette raison, la taille du réseau joue un rôle important dans la définition de la complexité de la fonction apprise. Les trois principales choses à savoir sur les MLP sont la propagation vers l'avant, la propagation vers l'arrière et la façon de les former correctement.

Les premiers réseaux de neurones n'étaient pas capables de résoudre des problèmes non linéaires ; cette limitation fut supprimée au travers de la rétro-propagation du gradient de l'erreur dans les systèmes multicouches, proposé par Paul Werbos en 1974 et mis au point douze années plus tard, en 1986 par David Rumelhart .

Dans le perceptron multicouche à rétro-propagation, les neurones d'une couche sont reliés à la totalité des neurones des couches adjacentes. Ces liaisons sont soumises à un

coefficient altérant l'effet de l'information sur le neurone de destination. Ainsi, le poids de chacune de ces liaisons est l'élément clef du fonctionnement du réseau : la mise en place d'un Perceptron multicouche pour résoudre un problème passe donc par la détermination des meilleurs poids applicables à chacune des connexions inter-neuronales.

a. Algorithme de rétropropagation

L'algorithme de rétropropagation standard se résume à la série d'étapes suivantes :

1. Initialiser tous les poids à de petites valeurs aléatoires dans l'intervalle $[-0.5, 0.5]$.
2. Normaliser les données d'entraînement.
3. Permuter aléatoirement les données d'entraînement;
4. Pour chaque donnée d'entraînement n :
 - (a) Calculer les sorties observées en propageant les entrées vers l'avant.
 - (b) Ajuster les poids en rétropropageant l'erreur observée :

$$\begin{aligned} w_{ji}(n) &= w_{ji}(n-1) + \Delta w_{ji}(n) \\ &= w_{ji}(n-1) + \mu \delta_j(n) y_i(n) \end{aligned} \quad (3.3)$$

4. Répéter les étapes 3 et 4 jusqu'à un nombre maximum d'itérations ou jusqu'à ce que la racine de l'erreur quadratique moyenne (EQM) soit inférieure à un certain seuil.

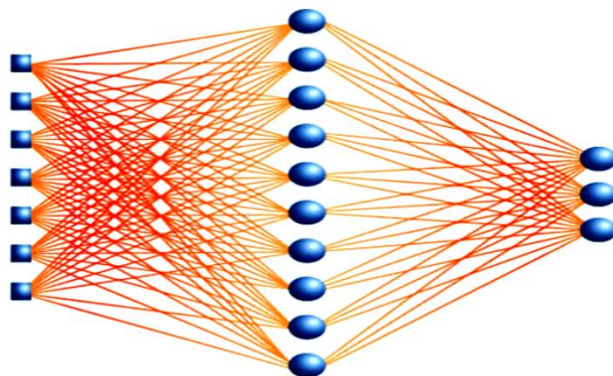


Figure 3.3: Perceptron multi couches

Dans le domaine de la reconnaissance des émotions, le perceptron multicouche est largement utilisé. Citons les travaux de [Danisman et al, 2013] et [Hayet et al, 2014].

5.2 Les réseaux de Hopfield

a. Description

Le modèle de Hopfield fut présenté en 1982. Ce modèle très simple est basé sur le principe des mémoires associatives. C'est d'ailleurs la raison pour laquelle ce type de réseau est dit associatif. Outre un intérêt pratique, ce réseau admet une analyse théorique précise et complète. Il a par ailleurs contribué à relancer les recherches sur les réseaux de neurones.

Les réseaux de Hopfield sont des réseaux à minimisation d'énergie : ils sont composés de neurones entièrement interconnectés ; ils n'ont pas une "entrée" et une "sortie", mais ils évoluent à partir d'un état initial .D'une façon générale, un réseau de Hopfield est donc défini par N neurones, et par une matrice de transfert, W, symétrique.

Dans le cas le plus simple, si on met des valeurs initiales (1 ou -1) dans tous les neurones, et si on applique à chacun des nœuds j successivement la transformation :

$$I_i = \sum_j w_{ij} V_j \quad (3.6)$$

w_{ij} Est le poids de la connexion du neurone i à j

V_j Est l'état du neurone j

Le calcul du nouvel état du neurone i se fait par :

$$V_i(t+1) = \begin{cases} 1 & \text{si } \sum_j w_{ij} V_j > 0 \\ -1 & \text{sinon} \end{cases} \quad (3.7)$$

Pour l'estimation des poids :

$$w_{ij} = 1/n \sum_{k=1}^p x_i^k x_j^k \quad (3.8)$$

Où w_{ij} est le poids de la connexion entre le neurone j et le neurone i, n est la dimension du vecteur d'entrée, p le nombre de motif d'entraînement et x_i^k et x_j^k sont respectivement la K ieme entrée des neurones i et j.

b. Les mémoires auto-associatives

Dans une mémoire informatique classique, une information est retrouvée à partir d'une clé arbitraire. Par opposition, une donnée entreposée dans une mémoire associative est accessible à partir d'informations qui lui sont associées.

La fonction d'une mémoire associative ressemble beaucoup à celle d'un filtre dont le but est de restituer une information en tenant compte de sa perturbation ou de son bruit. L'information doit alors se rapprocher d'une information apprise ou connue. Si les mémoires associatives restituent des informations qu'elles ont apprises à partir d'entrées incomplètes ou bruitées, il existe aussi des mémoires hétéro-associatives qui en plus peuvent associer plusieurs informations entre elles.

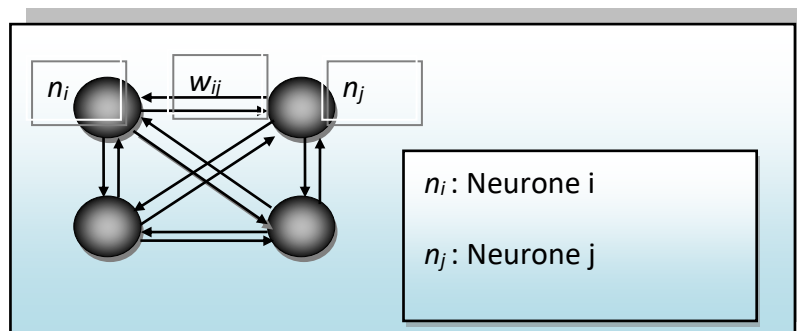


Figure 3.4: Le réseau de neurone Hopfield

Parmi les travaux de reconnaissances émotionnelles utilisant le réseau de neurone Hopfield, citons le travail de Gyanendra [Gyanendra et al, 2012]. Yoneyama et al ont utilisé Hopfield NN pour reconnaître quatre expressions faciales, et le résultat moyen de la classification est de 92,2%. [Yoneyama et al, 1997]

5.3 Les carte auto organisatrice de kohonen

Les cartes auto-organisatrices (Self-Organizing Maps (SOM)), furent inventées par Kohonen en 1984. Inspirée par le principe neuronal du cerveau des mammifères, une carte auto-organisatrice est un type de réseau de neurones artificiels dont l'apprentissage se déroule de manière non supervisée. Leur rôle principal est de faire une projection non linéaire des données de haute dimension sur un espace de faible dimension. Les cartes auto-organisatrices sont largement utilisées dans la classification de données.

a. Architecture

D'un point de vue architectural, les cartes auto-organisatrices de Kohonen sont constituées d'une grille (le plus souvent uni- ou bidimensionnelle). Dans chaque nœud de la grille se trouve un « neurone ». Chaque neurone est lié à un vecteur référent (figure 3.6), responsable d'une zone dans l'espace des données (appelé encore espace d'entrée).

Dans une carte auto-organisatrice, les vecteurs référents fournissent une représentation discrète de l'espace d'entrée. Ils sont positionnés de telle façon qu'ils conservent la forme topologique de l'espace d'entrée. En gardant les relations de voisinage dans la grille, ils permettent une indexation facile. Ceci s'avère utile dans divers domaines, comme la classification de textures, l'interpolation entre des données, la visualisation des données multidimensionnelle.

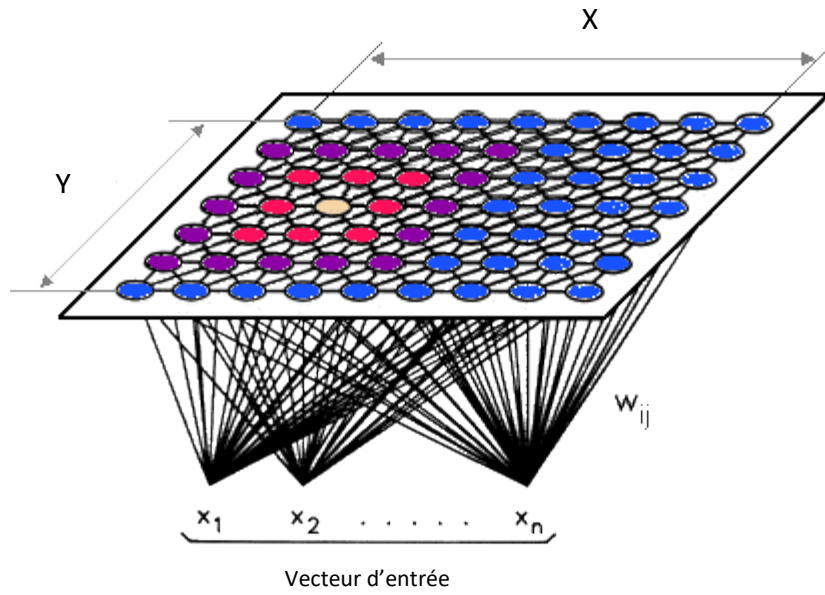


Figure 3.5: Carte de Kohonen à deux dimensions

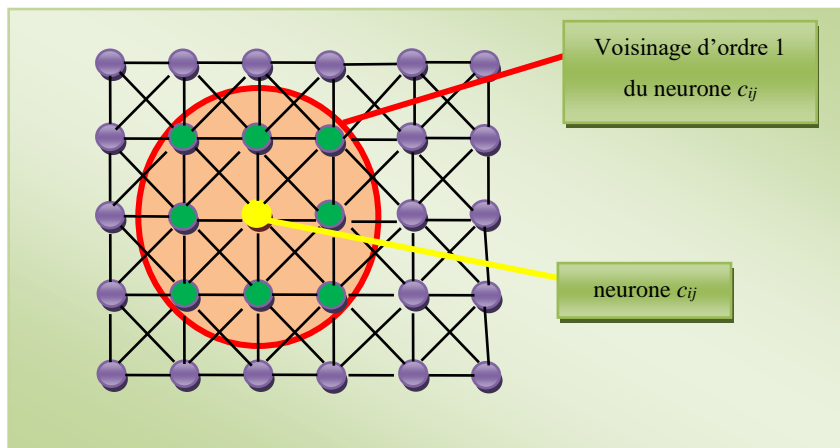


Figure 3.6: Notion de voisinage dans une Carte de Kohonen.

b. Algorithme de Kohonen :

Kohonen utilise un apprentissage non supervisé, à chaque entrée une réaction basée sur la notion de concurrence est apparue. Nous résumons le principe avec les étapes suivantes :

Etape 1 : Initialisation

La matrice poids est initialisée de manière aléatoire et un ensemble d'apprentissage est construit. Cet ensemble se compose de tous les stimuli qui représentent la population traitée autrement dit une représentation discrète de l'espace d'entrée.

Etape 2 : Modification des poids

Un élément de l'ensemble d'apprentissage est pris de façon aléatoire. Trouver le nœud dit « gagnant » (*winner*) : C'est le neurone dont le vecteur poids associé présente des valeurs les plus similaires au vecteur d'entrée.

On procède à la modification des poids du neurone gagnant, ainsi que ceux de son voisinage de manière à ce que les vecteurs associés (les vecteurs de poids) «se rapprochent d'avantage» du vecteur du stimulus d'entrée (équation 3.7). La modification des vecteurs associés aux neurones voisins se fait de manière différente selon la position des nœuds par rapport à neurone gagnant, les neurones les plus éloignées seront moins affectées.

$$\begin{aligned} w_i(t+1) &= w_i(t) + \alpha(t)(p_i - w_i(t)) && \text{si neurone vainqueur} \\ w_i(t+1) &= w_i(t) + h(r,t)(p_i - w_i(t)) && \text{si } i \in \text{voisinage} \\ w_i(t+1) &= w_i(t) && \text{si } i \notin \text{voisinage} \end{aligned} \quad (3.7)$$

$h(r,t) = \alpha(t) \cdot v(t)$: Le taux d'apprentissage. $v(t)$: la fonction de voisinage)

Décroître le coefficient d'apprentissage $\alpha(t)$, $\alpha(t)$ est un paramètre qui permet de contrôler l'importance des modifications appliquées aux vecteurs de poids.

Répéter l'étape 2 jusqu'à ce que le coefficient d'apprentissage soit nulle (convergence)

Etape 3 : Exploitation

Cette phase est entamée après que le réseau soit dans un état stable ou dite état de convergence. A chaque vecteur présenté à la couche d'entrée, on calcule le neurone gagnant pour désigner à quelle classe cet élément appartient.

5.4 Les réseaux de neurones convolutifs CNN):

Au cours des dernières années, il y a eu une percée dans les algorithmes d'apprentissage en profondeur(CNN) appliqués au domaine de la vision par ordinateur,

Ces algorithmes basés sur l'apprentissage en profondeur ont été utilisés pour l'extraction de caractéristiques, la classification et la reconnaissance. Le principal avantage d'un CNN est de supprimer complètement ou fortement la dépendance aux modèles basés sur la physique et / ou d'autres techniques de prétraitement en permettant un apprentissage «de bout en bout» directement à partir des images d'entrée [Walecki et al, 2017]. Pour ces raisons, CNN a obtenu des résultats très satisfaisant dans divers domaines, notamment la reconnaissance d'objets, la reconnaissance faciale, la compréhension de scènes.

Les réseaux de neurones sont désignés par l'acronyme CNN, de l'anglais « Convolutional Neural Network », ils comportent deux parties bien distinctes. En entrée, une image est fournie sous la forme d'une matrice de pixels. Elle a 2 dimensions pour une image en niveaux de gris. La couleur est représentée par une troisième dimension, de profondeur 3 pour représenter les couleurs fondamentales (Rouge, Vert, Bleu).

La première partie d'un CNN est la partie convolutive à proprement parler. Elle fonctionne comme un extracteur de caractéristiques des images. Une image est passée à travers une succession de filtres, ou noyaux de convolution, créant de nouvelles images appelées cartes de convolutions. Certains filtres intermédiaires réduisent la résolution de l'image par une opération de maximum local. Au final, les cartes de convolutions sont mises à plat et concaténées en un vecteur de caractéristiques, appelé code CNN.

Ce code CNN en sortie de la partie convolutive est ensuite branché en entrée d'une deuxième partie, constituée de couches entièrement connectées (perceptron multicouche). Le rôle de cette partie est de combiner les caractéristiques du code CNN pour classer l'image.

La sortie est une dernière couche comportant un neurone par catégorie. Les valeurs numériques obtenues sont généralement normalisées entre 0 et 1, de somme 1, pour produire une distribution de probabilité sur les catégories.

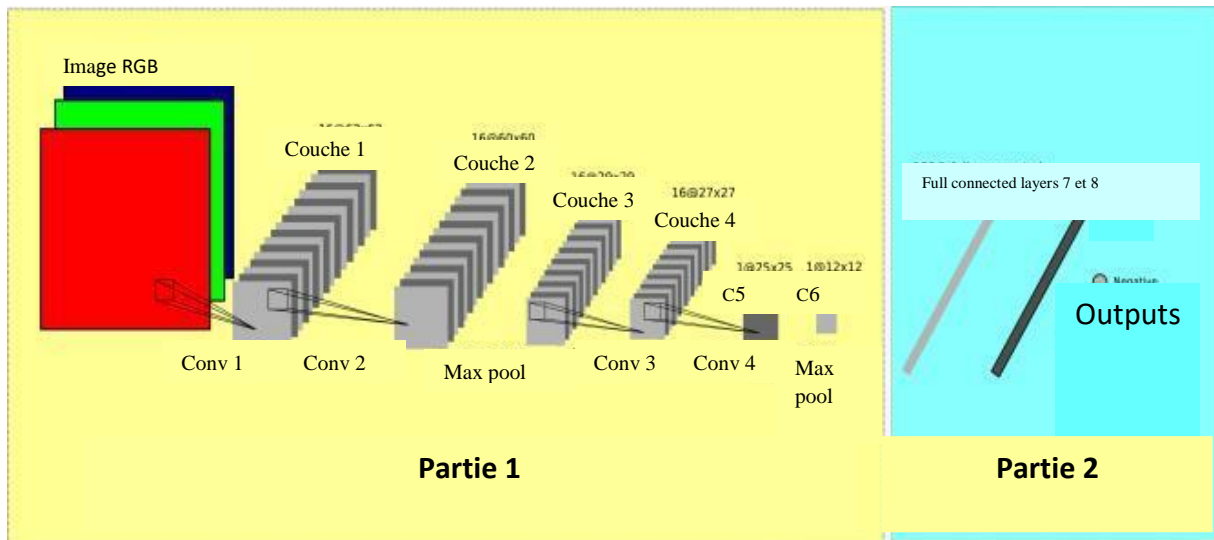


Figure 3.7. : Exemple d'architecture d'un CNN

4.1 Les réseaux de neurones et les émotions :

Les réseaux de neurones sont très efficaces pour la reconnaissance des expressions faciales. Par exemple, la reconnaissance des configurations de la bouche peut être détectée par un réseau de neurones. Le réseau aura alors été entraîné sur un ensemble d'images de bouches ayant des configurations différentes [PANTIC et Leon, 2000].

Garrison et al ont cherché des méthodes d'analyse automatique du visage les plus proches possibles de la réalité biologique. Ainsi, un réseau de neurones dit : auto-supervisé, c'est à dire dont la couche d'entrée et la couche de sortie sont identiques et égales à l'image d'un visage, effectue une analyse en composantes principales. Le nombre de composantes principales est donné par le nombre de neurones de la couche cachée. Chaque neurone de la couche cachée correspond aux valeurs propres de la décomposition [Garrison et al, 2000].

Liu, et al, ont proposé de construire une architecture profonde, inspirée de la théorie psychologique selon laquelle les expressions peuvent être décomposées en plusieurs unités d'action faciale (AUs).[Liu et al, 2015].

Récemment beaucoup d'études ont été faites pour la reconnaissance des émotions en se basant sur les CNN. Citons les travaux de Kim et al [Kim et al, 2017] et Multi-level AU

[Chu et al, 2017]. Breuer et Kimmel ont utilisé un CNN pour comprendre un modèle appris à l'aide de différents jeux de données FER (Facial Expression recognition) et ont démontré la capacité de ces réseaux pour la détection des émotions à travers les jeux de données[Breuer et Kimmel, 2017].Jung et al. utilisent deux types différents de CNN: le premier extrait les caractéristiques d'aspect temporel des séquences d'images, tandis que le second extrait les caractéristiques de la géométrie temporelle à partir des points faciaux temporels.[Jung et al, 2015].

4.2 Conclusion

Les réseaux de neurones (RNs) sont généralement utilisés pour leur faible sensibilité au bruit (robustesse au bruit) et leur capacité d'apprentissage. Ils ont connu un grand succès dans le traitement d'images. En raison de la capacité de conserver des informations sur les entrées passées, RN est capable d'apprendre les dépendances contextuelles avec les images, ce qui est avantageux par rapport à CNN (convolutional neural network). Pour cette raison, RNN est généralement combiné avec CNN afin d'atteindre une meilleure performance des tâches de vision telles que la segmentation de l'image [Visin et al, 2015].

Ils sont souvent difficiles à construire. Leur structure (nombre de couches cachées pour les perceptrons par exemple) influe beaucoup sur les résultats et il n'existe pas de méthode pour déterminer automatiquement cette structure. La phase d'apprentissage est difficile à mener puisque les exemples doivent être correctement choisis (en nombre et configuration). En plus, la plupart de ces méthodes sont testées uniquement sur des images utilisées au cours de l'apprentissage, c'est pourquoi on ne peut prévoir le comportement de ces méthodes dans le cas de sujets inconnus. En outre, la plupart de ces méthodes exigent une intervention manuelle.

Les réseaux de neurones servent aujourd'hui à toutes sortes d'applications dans divers domaines. Par exemple, on a développé un autopilote dans le domaine de l'aviation, ou encore un système de guidage pour automobile, on a conçu des systèmes de lecture automatique de chèques bancaires et d'adresses postales, on a produit des systèmes de traitement du signal pour différentes applications militaires, un système pour la synthèse de la parole, des réseaux sont aussi utilisés pour bâtir des systèmes de vision par ordinateur, pour faire des prévisions sur les marchés monétaires, pour évaluer le risque financier ou en assurance, pour différents processus manufacturiers, pour le diagnostic médical, pour l'exploration pétrolière ou gazière, en robotique, en télécommunication, et j'en passe ! Bref, les réseaux de neurones ont aujourd'hui un impact considérable et, il y a fort à parier, que leur importance ira grandissante dans le futur. Plus précisément dans le domaine de l'analyse de l'état affectif, les réseaux de neurones ont présenté des taux de reconnaissance émotionnelle très satisfaisants. Et ils ont prouvé leurs efficacités.

Chapitre 4

Systeme neuronal pour la reconnaissance des émotions

4.1 Introduction

Nous proposons dans cette thèse deux contributions pour la détection et la reconnaissance des émotions dans les séquences d'images. La première, c'est l'introduction d'un nouveau modèle qui donne une description claire, précise et objective de l'émotion d'intérêt. Chose qui n'a jamais été proposée dans le domaine de l'analyse de l'état affectif. Comme nos émotions sont la seule manière d'expression de nos sentiments, alors nous n'avons que la psychologie qui les décrit. Nous avons opté pour le fait de se baser sur des fondements psychologiques qui ont prouvé que les comportements de l'être humain envers son entourage suivent les émotions des personnes avec lesquelles il interagit. Nous avons développé un nouveau modèle pour interpréter les mouvements qui décrit l'émotion d'intérêt en unité d'action du système FACS en fonctions des descriptions des psychologues relatives à chaque émotion.

Notre deuxième contribution, s'articule sur la proposition d'un nouveau système de reconnaissance des émotions de base, y compris l'émotion d'intérêt. Nous nous sommes focalisés sur la reconnaissance dynamique des émotions, une tâche qui n'est pas facile, vu le changement des muscles faciaux en partant de l'image neutre et arrivant à l'image qui contient l'émotion. Donc, un traitement et un suivi de mouvement est présenté pour pouvoir reconnaître l'émotion.

Pour ce faire, nous avons fait une estimation des mouvements de visage, après l'avoir détecté et extrait les caractéristiques faciales, qui sont des points stratégiques dans le visage. Ces derniers ont été réduits à dix. Nous avons minimisé ce nombre par l'étude des muscles faciaux responsables des émotions de base et nous avons constaté que les dix points choisis sont diffusants pour la description des émotions. C'est un nombre assez petit pour décrire une émotion dans un visage, mais le choix de ces caractéristiques a influé positivement sur le temps d'exécution. L'estimation de mouvement est basée sur l'algorithme de bloc matching. Ensuite, nous classifions le vecteur du mouvement obtenu avec la carte auto organisatrice de KOHONEN. Les résultats sont très satisfaisants en termes de taux de reconnaissance (96%), et de la minimisation des données traitées

4.2 Système neuronal pour la reconnaissance d'émotion

Comme tout système de reconnaissance des émotions, notre approche suit les trois étapes fondamentales citées dans le chapitre 2.

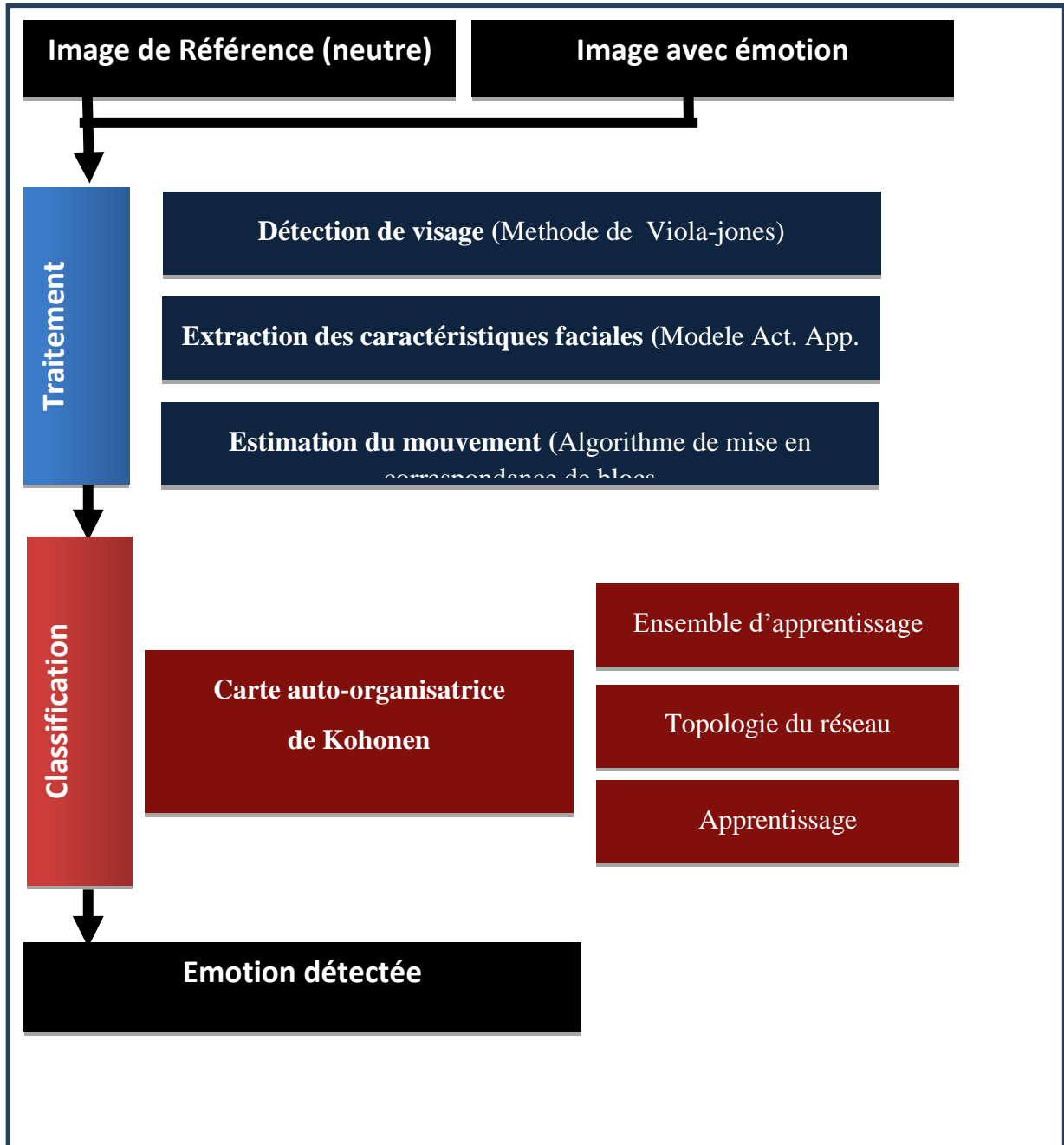


Figure 4. 1. : Organigramme de la méthode proposée

4.2.1 La détection du visage :

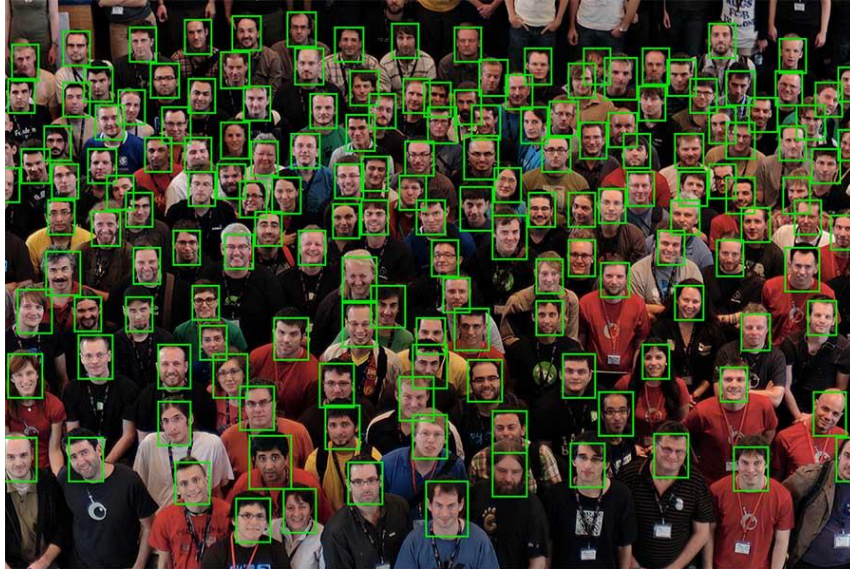


Figure 4.2. : Détection des visages dans une image

La détection du visage est un domaine de la vision par ordinateur consistant à détecter un visage humain dans une image numérique. C'est un cas spécifique de détection d'objet, où l'on cherche à détecter la présence et la localisation précise d'un ou plusieurs visages dans une image.

✓ Description

Afin de détecter les visages dans nos bases de données, nous avons utilisé la méthode de Viola-Jones caractérisée par sa rapidité et sa précision. Elle représente la première méthode travaillant en temps réel. L'algorithme utilisé se base sur trois étapes principales : l'image intégrale pour le calcul des caractéristiques, Adaboost pour la classification des caractéristiques et une cascade pour la détection des visages.

L'image intégrale est une image construite à partir de l'image d'origine. Elle contient en chacun de ses points la somme des pixels situés au-dessus et à gauche du pixel courant. L'image intégrale I_{int} au point (x,y) est défini à partir de l'image I (Eq4.1). A partir de l'image résultat, on utilise des fenêtres (masques) sous forme de zones rectangulaires

adjacentes ; les intensités de pixels de ces rectangles sont additionnées, formant des sommes dont la différence constitue une caractéristique.

$$I_{int}(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y') \quad (4.1)$$

Cette somme peut se calculer par récurrence, par:

$$S(x, y) = S(x, y - 1) + i(x, y) \quad (4.2)$$

$$ii(x, y) = ii(x - 1, y) + S(x, y) \quad (4.3)$$

Où $S(x, y)$ est la somme cumulée de la ligne x jusqu'à la colonne y . L'image intégrale peut donc se calculer avec un seul parcours de l'image d'origine.

Une fois que l'image intégrale est calculée, la somme des pixels à l'intérieur de n'importe quel rectangle ABCD à l'intérieur de l'image peut être évaluée en seulement 4 accès

$$i(x', y') = ii(A) + ii(C) - ii(B) - ii(D) \quad (4.4)$$

Le deuxième élément clé de la méthode de Viola et Jones est l'utilisation d'une méthode de boosting afin de classifier et sélectionner les meilleures caractéristiques. Son principe consiste à construire un classifieur « fort » $h(x)$ (Eq4.5) à partir d'une combinaison pondérée de classifieurs « faibles » h_j associés chaque caractéristique j en utilisant un apprentissage supervisé, un seuil est défini afin de séparer les exemples positifs des exemples négatifs. Le classifieur se réduit alors à un couple (caractéristique, seuil).

Algorithme :

- Pour chaque $(x_1, y_1), \dots, (x_n, y_n)$

Avec $y_i = 0$ ou 1 pour les négatifs et les positifs exemples respectivement.

- Initialisation des poids $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$
 Pour $y_i = 0, 1$ pour les négatifs et les positifs exemples respectivement
 m, l sont le nombre des négatifs et positifs exemples

- Pour $t = 1, \dots, T$

1. Normalisation des poids $w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$

w_t Est la probabilité de distribution

2. Pour chaque caractéristique, $j, y_j = h(x_j)$

L'erreur est donnée par $\varepsilon_j = \sum_i w_i |h_j(x_i) - y_i|$

3. Choisir le classifieur h_t qui minimise l'erreur.

4. Mise à jour des poids

$$w_{t+1,i} = w_{t,i} B_t^{1-\varepsilon_i}$$

Où $\varepsilon_i = 0$ si l'exemple x_i est bien classifié. 1 sinon

Et $B_t = \frac{\varepsilon_t}{1-\varepsilon_t}$

5. Le classifieur final est :

$$h(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{si non} \end{cases} \quad (4.5)$$

Avec $\alpha_t = \log 1/B_t$

La cascade (Eq4.6) est constituée d'une succession d'étages, chacune étant formée d'un classifieur fort appris par le boosting. L'apprentissage du classifieur de l'étage n est réalisé avec les exemples qui ont franchi l'étage n-1. Plus le nombre d'étages est important, plus on converge vers les vrais positifs. Ce nombre(K) est fixé par Viola et Jones à 32.

Le taux de détection de la cascade est donné par :

$$D = \prod_{i=1}^K d_i \tag{4.6}$$

d_i est le taux de détection minimal.

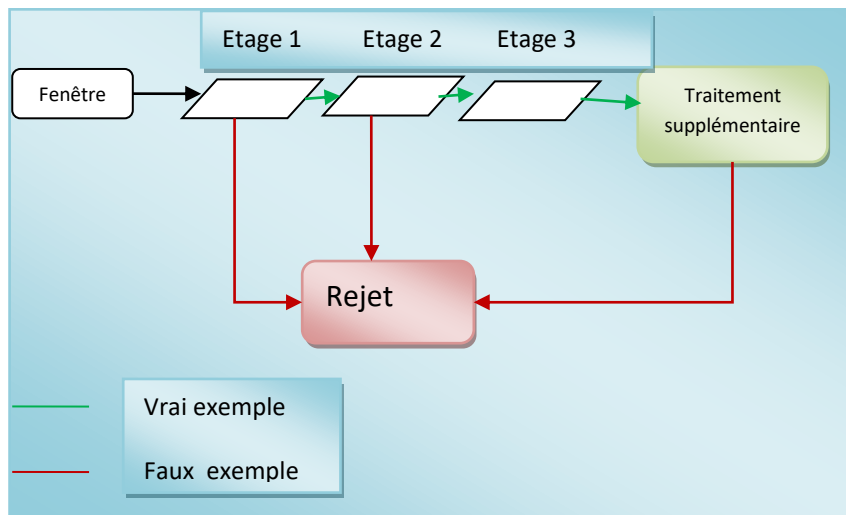


Figure 4.3. : La cascade de Viola Jones

4.2.2 Extraction des points stratégiques

Les méthodes d'extraction des informations locales au niveau des points sont plus intéressantes en terme de temps de calcul et en terme de mémoire utilisée puisque leurs vecteurs descripteurs sont moins volumineux que ceux des méthodes globales. Cependant, la robustesse de ces méthodes reste fortement liée à la précision de la position des points.

Les méthodes globales extraient quant à elles toutes les textures du visage, permettant ainsi de ne pas négliger des zones indispensables à la reconnaissance de l'expression. Elles

nécessitent cependant toujours une phase de prétraitement pour normaliser le visage afin d'éliminer les éventuelles erreurs causées par la pose de la tête et la différence de taille entre les visages. Ce que nous avons fait dans l' étape précédente,

Cela nous a motivé à utiliser la méthode d'extraction des points dans notre système de reconnaissance.

✓ Description du modèle d'apparence active :

Pour accomplir notre traitement, nous avons besoin des points qui délimitent les différentes caractéristiques faciales. Pour se faire, nous avons exploité les modèles d'apparence active (Active Appearance Model) (AAM) [Timothy et al, 2001] qui sont des modèles génératifs et paramétriques d'un certain phénomène visuel qui montrent à la fois des variations de forme et d'apparence. Nous l'avons choisi suite à sa puissance prouvée. Elle fournit une représentation du visage réaliste et naturelle. Il s'agit d'un algorithme qui permet de faire correspondre à une image une forme en utilisant un modèle statistique construit à partir de deux phases : une phase d'apprentissage qui combine un modèle de marquage et un autre de texture en se basant sur la technique d'Analyse en Composantes Principale (ACP) pour fournir une matrice d'expérience (Eq 4.7) et finalement produire un masque de points stratégiques sur le visage (Fig. 4.7) :

$$\begin{cases} s_i = \bar{s} + \Phi_s * b_{s_i} \\ g_i = \bar{g} + \Phi_g * b_{g_i} \end{cases} \quad (4.7)$$

Avec \bar{s} et \bar{g} les formes et textures moyennes, Φ_s et Φ_g les matrices constituant les matrices de vecteurs propres représentant les modes de variation $t_{b_{s_i}}$ et $t_{b_{g_i}}$ les vecteurs de composantes principales de forme et de texture.

L'apparence d'un visage et sa pose (position, orientation et échelle) dans l'image sont enfin représentés par le vecteur

$$P^T = c^T / t^T \quad (4.8)$$

Où t contient les paramètres d'une transformation globale rigide : Rotation (ν), Homothétie (s) et Translation (t_x, t_y).

c est le vecteur d'apparence de la classe de visages appris (il contrôle à la fois la forme et la texture d'un visage).



Figure4.4. : Les points stratégiques dans le visage

- ✓ Les points stratégiques manipulés par notre système :

Contrairement aux travaux de reconnaissance des émotions ayant extrait un nombre élevé de points pour pouvoir reconnaître l'expression faciale [Naoya et al, 2016] [Taner et al, 2014], ce qui consomme un temps de calcul relativement élevé exploitant des données consistantes. Nous avons étudié les muscles faciaux qui déterminent les émotions de base. Nous avons déterminé les points essentiels qui déclenchent le mouvement des ces muscles.

Notre système utilise seulement dix points que nous appelons « points stratégiques » sélectionnés parmi tout les points extraits : les deux coins de la bouche, le haut de la lèvre supérieure, le bas de la lèvre inférieure et deux points par sourcils et les deux points de la paupière supérieure. ; soient dix au total(Fig. 4.5)

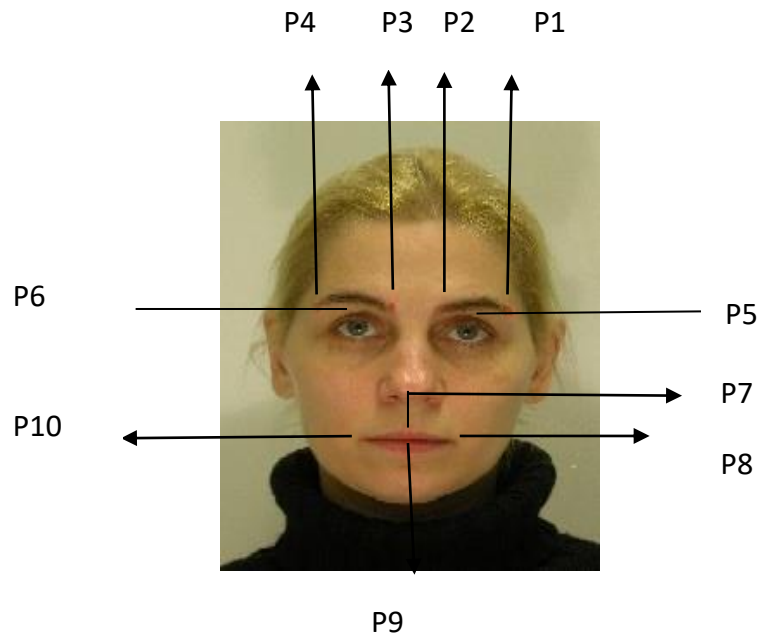


Figure 4.5. : Les points stratégiques manipulés par notre système

4.2.3 Détection des unités d'action du système FACS :

Les unités d'action présentées par le système FACS sont dérivées par des experts humains basés sur des expériences en psychologie. Des études utilisant l'approche FACS ont rapporté des précisions de classification allant de 70% à 95% sur six expressions faciales de base [Kotsia et al, 2008].

Bien qu'il n'est pas destiné à l'origine à la classification automatisée des expressions, FACS fournit l'un des ensembles de caractéristiques les plus fréquemment utilisés dans la littérature [Tsalakanidou et al, 2010]. FACS est un outil largement utilisé pour décrire les mouvements faciaux.

Comme nous l'avons déjà cité, chaque unité d'action reflète un mouvement facial. Donc pour savoir de quelle unité d'action s'agit-il, nous devons suivre le mouvement des points stratégiques extraits. par exemple, l'AU4 est relative à l'abaissement et rapprochement

des sourcils (figure 4.6). Ainsi, un suivi du mouvement des points P3 et P2 est obligatoire pour détecter une telle unité d'actions.



Figure 4.6. : le mouvement de AU4

De même, le tableau 4.1 résume tous les mouvements des autres points stratégiques:

AUs	Le mouvement	Points à suivre
AU1	Remontée de la partie interne des sourcils.	P3 P4
AU4	Abaissement et rapprochement des sourcils.	P3 P4
AU 5	Ouverture entre la paupière supérieure et les sourcils	P5 P6
AU 7	Tension de la paupière	P5 P6
AU 9	Plissement de la peau du nez vers le haut.	P7
AU 12	Étirement du coin des lèvres	P8 P10
AU 23	Tension refermante des lèvres	P8 P10
AU 26	Ouverture de la mâchoire	P7 P9

Table 4.1 : Interprétation des unités d'action en mouvements

4.2.4 Estimation de mouvement

- ✓ La méthode de Blocs matching

Le mouvement, dans le cadre du système visuel humain, est un sujet majeur de la vision par ordinateur et apporte une contribution importante à des sujets de recherche comme la compression et l'estimation vidéo, la reconstruction 3D et le suivi visuel. [Ramona et al, 2017].

Pour effectuer le suivi des points stratégiques extraits, nous exploitons la méthode de Blocs matching qui est l'un des algorithmes du modèles différentiels mis au point pour l'évaluation du mouvement dans les trames consécutives. La méthode est basée sur le fait que les zones de pixels se déplacent dans le cadre et peuvent être détectées dans le suivant. Par conséquent, la trame courante est divisée en blocs et une correspondance dans la trame suivante sera recherchée pour chacun d'eux.

Son principe repose sur la recherche de motifs de niveaux de gris se correspondant dans 2 images successives d'une séquence. Généralement, ces méthodes décomposent la première image en blocs recouvrant ou disjoints. Pour chacun des blocs X, on recherche un bloc correspondant dans la seconde image autour du centre du bloc initial dans une zone appelée fenêtre de recherche B_x (Fig 4.7). Cette recherche s'appuie sur une mesure de similarité ou dissimilarité. Le bloc conduisant au maximum de similarité (ou à un minimum de dissimilarité) indique le déplacement subi par le motif. La somme des valeurs absolues des différences (ang. Sum of Absolute Differences, SAD) et la somme des valeurs quadratiques des différences (ang. Sum of Quadratic Differences, SQD) sont des mesures très utilisées parce qu'elles sont cumulatives ; on peut ainsi suspendre prématurément l'accumulation dès qu'elles franchissent un seuil prédéfini, ce qui permet d'accélérer le calcul des correspondances.

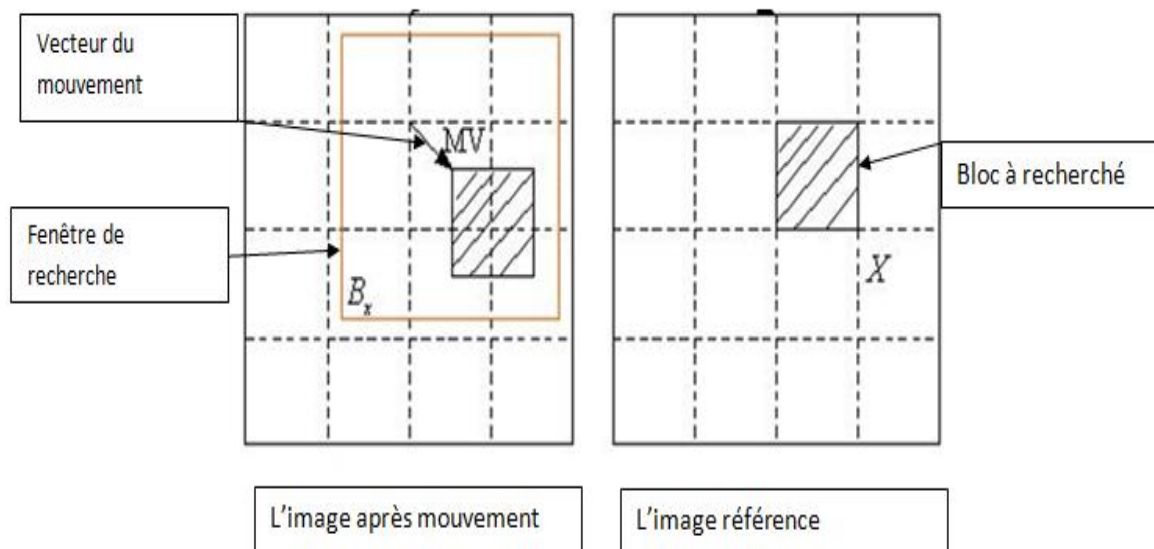


Figure 4.7 : Recherche d'un bloc correspondant dans une zone de recherche (B_x) entre deux images I_1 et I_2

Les caractéristiques associées aux méthodes de mise en correspondance de blocs sont :

- La taille du bloc X
- La taille de la fenêtre de recherche B_x
- La stratégie de recherche
- Le critère de similarité

La définition de ces paramètres dépend directement de l'application.

Nous avons utilisé un bloc de taille 3×3 , ce qui est suffisant pour localiser les points stratégiques.

La taille de la fenêtre de recherche est liée au déplacement maximal attendu. Nous l'avons fixé à 9×9 . Pour améliorer le résultat de recherche

Pour le critère de similarité, nous avons utilisé la SAD : Sum of Absolute Difference.

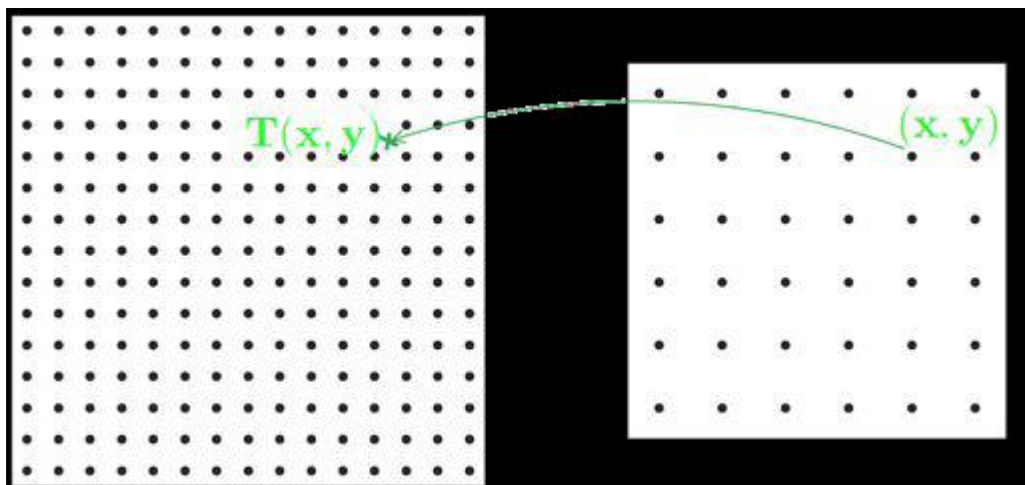


Figure 4.8 : La similarité entre les points

La SAD de deux blocs X et Y dans deux image I₁ et I₂: (X ∈ I₁, Y ∈ I₂) de N × N pixels est définie par:

$$SAD(X, Y) = \sum_{i=1}^N \sum_{j=2}^N |X(i, j) - Y(i, j)| \quad (4.9)$$

Pour un bloc source X donné, le bloc Y le plus similaire est celui qui minimise la SAD.

Pour la stratégie de recherche nous avons exploité l’algorithme de (Full Search Algorithm (FSA)), qui consiste à calculer de manière exhaustive l’erreur SAD pour tous les déplacements possibles dans un voisinage donné. Il offre la meilleure estimation de mouvement que nous pouvons obtenir avec un block-matching. Malgré qu’il est très couteux en temps de calcul, et comme nous ne manipulons qu’un nombre très réduit de points stratégiques, cette consommation de temps n’est pas prise en considération.

Le résultat obtenu est un vecteur de mouvement de 10 éléments, chaque élément correspond au mouvement d’un point stratégique. Ceci nous permet de déduire les unités d’action voulues (Fig. 4.9).

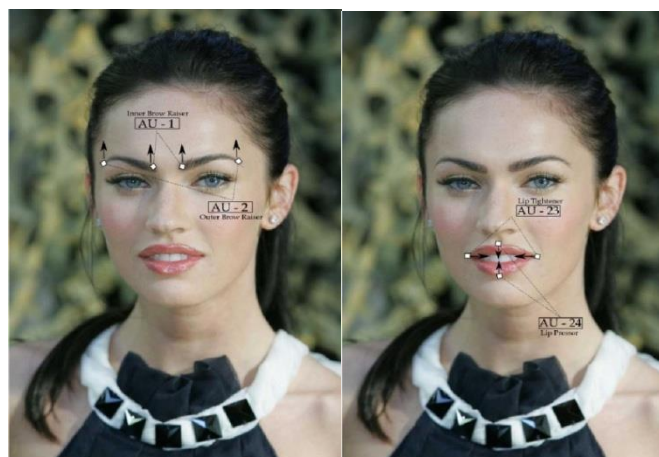


Figure 4. 9. : Direction des mouvements des unités d’action

4.2.5 Classification

La phase de classification permet de nous fournir l'émotion de la personne figurant dans la séquence d'image traitée. Elle reçoit, en entrée, le vecteur de mouvement obtenu dans l'étape précédente (représentant les unités d'actions).

Kohonen self-organizing map (KSOM) a une capacité supplémentaire à regrouper les données dans un ordre qui maintient la topologie des données d'entrée. En raison de cette propriété de KSOM, les données d'expressions faciales similaires se regroupent en zones plus proches ce qui rend la classification beaucoup mieux [Kohonen, 1990]. Cette propriété de KSOM nous a motivé à l'utiliser pour la détection des émotions : joie, surprise, colère, dégoût, tristesse, peur et intérêt.

On comprendra par auto-organisation, un procédé capable, bien que non supervisé, de trouver la solution à un problème d'optimisation au sens d'un certain critère. Sur cette base, Kohonen a proposé un réseau de neurones formel connu sous le nom de 'cartes auto-organisatrices de Kohonen' (SOM). Le but d'une SOM est de représenter un ensemble de données où chaque neurone se spécialise pour représenter un sous-ensemble bien particulier des données selon les points communs qui les rassemblent. Par conséquent, on obtient une classification en dimension multiple des données. Autrement dit, on peut simuler le fonctionnement d'une carte auto-organisatrice comme une carte topographique pour laquelle l'intensité des liaisons entre les neurones et les stimuli (élément de l'ensemble de données) donneraient l'information de relief. L'intérêt est de savoir, pour un ensemble de stimuli, quels neurones seront les vainqueurs (où se situent les pics aux sens du relief).

Suite à l'analyse de ces réponses, nous pouvons déduire un lien entre chaque stimulus et un neurone de la carte. Dans les versions actuelles, le vecteur poids d'un neurone de la carte de Kohonen permet de stocker un motif semblable à une « catégorie d'entrées » caractérisée par ce neurone après apprentissage.

Nous avons opté pour l'algorithme WTA (winner Take All). C'est un apprentissage compétitif qui consiste à modifier uniquement le vecteur des poids du neurone vainqueur. Ce neurone présente des poids fortement corrélés au stimulus présenté en entrée. Les valeurs de son vecteur poids sont alors adaptées en prenant en compte que la valeur du stimulus, sans

modifier les poids des neurones voisins. Ainsi, cet algorithme limite la propagation de l'apprentissage vers les voisins car seul le neurone sélectionné apprend au détriment des autres. Notre choix est justifié par l'efficacité d'un tel algorithme lorsqu'il existe une indépendance entre les classes à obtenir en résultat comme c'est le cas pour les émotions de base.

Dans notre cas, la carte de Kohonen est unidimensionnelle (linéaire). La couche d'entrée est constituée de 10 neurones au nombre des éléments du vecteur de mouvement de chaque points d'intérêt. Le nombre de neurones de la carte est égale au nombre d'émotions prévues ; dans notre cas, la taille est de 6. Chaque neurone de la carte est relié à tous les neurones de la couche d'entrée (Fig4.13).

Un réseau KSOM 2D est utilisé pour la classification du vecteur des caractéristiques obtenues avec une estimation du mouvement des points d'intérêt ;

$$X = [x_1, \dots, x_j]^T \in R^{10} \quad (4.10)$$

Pour chaque nœud j dans la structure du réseau 2D,

Vecteur de poids

$$W_j = [w_{j,1}, \dots, w_{j,10}]^T \in R^{10}$$

Matrice : $A \in R^{6 \times 10}$ et Vecteur de polarisation $b_j \in R^6$.

Pendant l'entraînement, chaque vecteur d'entrée X est comparé à tous les W_j pour trouver l'emplacement de la correspondance étroite :

$$\Delta X = X - W_j \quad (4.11)$$

Le nœud gagnant nommé EmotionWinner (EW) désigné par i , c'est celui qui minimise la distance entre le vecteur d'entrée X et le vecteur de poids W correspondant au nœud j .

$$i = \min \|X(n) - W_j\| \quad (4.12)$$

La distance euclidienne $d_{j,i}$ entre la EW et chaque neurone est donnée par :

$$h_{j,i}(n) = \exp\left(-\frac{d_{j,i}^2}{2\partial^2}\right) \quad (4.13)$$

$$d_{j,i}^2 = \|r_j - r_i\|^2 \quad (4.14)$$

Où $h_{j,i}(n)$ est la fonction de voisinage,

La règle de mise à jour des pondérations au nœud j et à l'itération n est donnée par :

$$w_j(n+1) = w_j(n) + \partial(n)h_{j,i}(n)[X(n) - W_j(n)] \quad (4.15)$$

$$\partial(n) = \partial_0 * \left(\frac{\partial_f}{\partial_0}\right)^{n/NI} \quad (4.16)$$

∂ est initialement considéré comme très grand et se diminue progressivement à chaque itération.

Où ∂ est le taux d'apprentissage à l'itération n , ∂_0 est le taux d'apprentissage initial, ∂_f est le taux d'apprentissage final et NI est le nombre d'itérations total nécessaires pour l'apprentissage.

La fonction de voisinage dans la carte est une fonction continue de forme gaussienne.

La décroissance de la taille du voisinage s'obtient par diminution de l'écart type ∂ . ∂ grand, beaucoup de neurones se rapprochent de $X(t)$, ∂ petit, l'adaptation reste très localisée.

La distance euclidienne entre EW i et le nœud j est donné par $d_{j,i}$

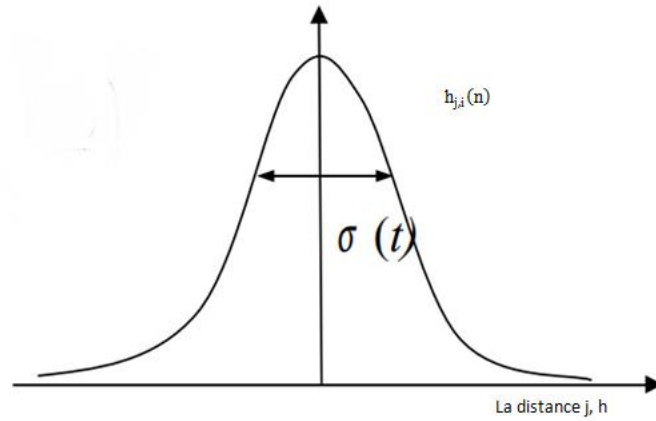


Figure 4. 10: La fonction de voisinage

✓ Interprétation de l'algorithme d'auto organisation :

La sélection du neurone le plus proche de l'exemple présenté X est : (Fig 4.11)

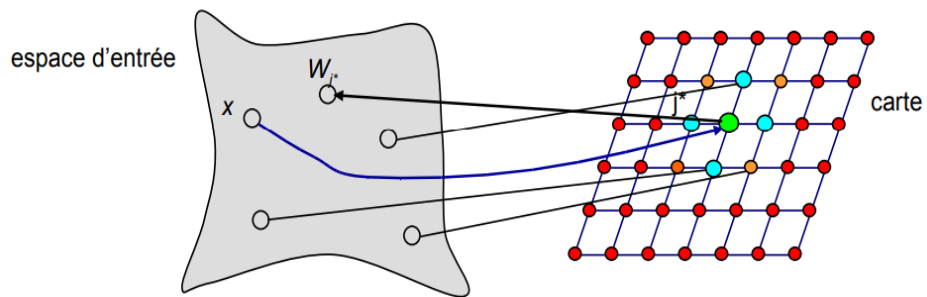


Figure 4.11 : Le nœud le plus proche de l'entrée à une carte de Kohonen

L'adaptation des poids peut être présentée par la figure 4.12.

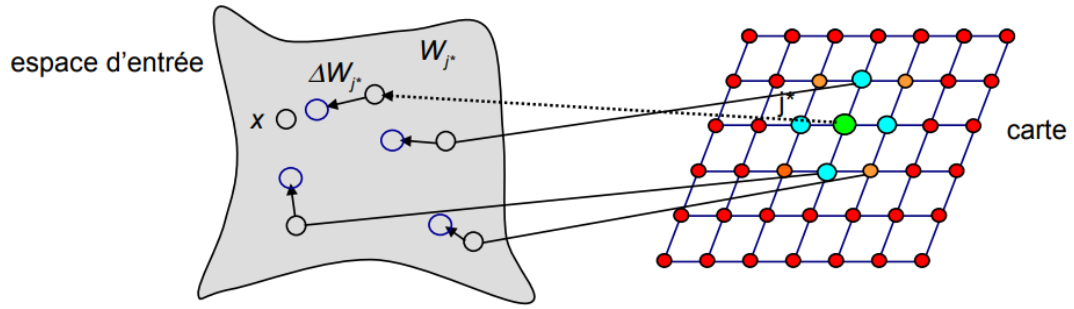


Figure 4. 12. : Adaptation des poids

La carte de Kohonen manipulée par notre système est décrite dans (Fig 4.13)

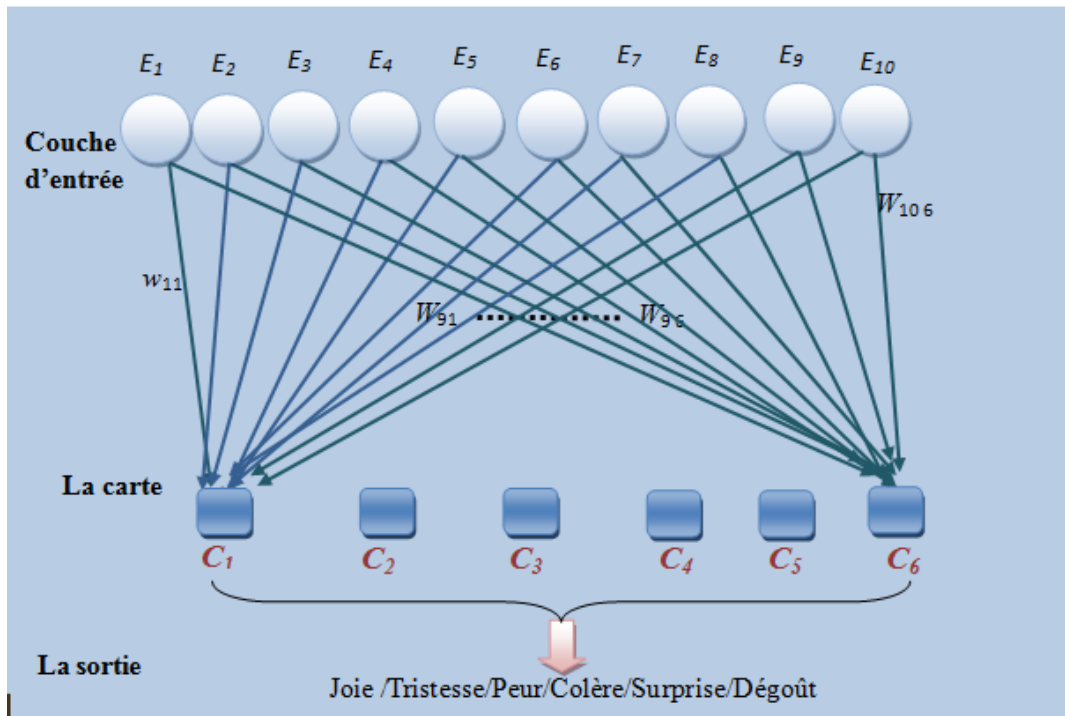


Figure 4.13. : Carte de KOHONEN proposée pour la reconnaissance des émotions

4.2.6 RESULTATS ET DISSCUSSION :

a. DATASET:

- ✓ LA base de données Cohen-Kanade

La base de données Cohn-Kanade comprend 2105 images de 182 sujets âgés de 18 à 30 ans [Kanade et al, 2000]. Les images ont la résolution de 640 * 480 ou 640 * 490 pixels. Environ 65% des sujets sont des femmes, 15% sont afro-américains et 3% sont asiatiques ou latinos. Toutes les séquences d'images commencent par une expression neutre et se terminent par le maximum de l'expression. Les images discriminantes d'une expression sont localisées dans les dernières trames des séquences. La figure 4.14 montre quelques exemples de six expressions faciales de base de la base de données Cohn-Kanade. La base de données Cohn-Kanade est couramment utilisée dans la littérature.



Figure4.14: Exemple d'images d'expressions faciales à partir de la base de données Cohn-Kanade

✓ CK+ data base

L'ensemble de données CK + [Lucey et al, 2010] est étendu à partir de l'ensemble de données CK [Kanade et al, 2000], qui a été construit en 2000; Il est disponible et considéré comme l'ensemble de données le plus largement utilisé pour évaluer la performance des systèmes de reconnaissance d'expression faciale. Cet ensemble de données comprend 593 séquences vidéo, dont sept expressions faciales de base réalisées par 123 participants. Les âges des sujets vont de 18 à 30 ans; 35% sont des hommes; 13% sont afro-américains; 80% sont euro-américains; et 6% sont des personnes d'autres races.

Dans les séquences vidéo, la taille de chaque image est de 640×490 ou 640×480 pixels. La plupart des images sont des images grises avec une précision de huit bits pour les valeurs de niveaux de gris. Les séquences vidéo contiennent des images de la phase neutre à la phase de crête des expressions faciales, et le débit vidéo est de 30 images par seconde. Dans notre étude, nous sélectionnons la première et la dernière image de la séquence. La rangée supérieure de la figure 4.10 montre des images de pic de six expressions de six sujets dans un ensemble de données CK +. Le tableau 4.2 montre le nombre de sujets pour chaque classe d'expression dans l'ensemble de données CK + dans notre expérience.



Figure 4.15 : Exemple d'images d'expressions faciales à partir de la base de données CK+

✓ MMI data set :

La base de données MMI [Valstar et al, 2010] est également utilisée. Cet ensemble de données est disponible au public et a été utilisé dans plusieurs publications, par exemple, [Pantic et al, 2005]. Elle comprend des sujets d'étudiants et de membres du personnel de recherche des deux sexes âgés de 19 à 62 ans. C'est une ressource sans cesse croissante pour la reconnaissance des émotions de base, dont le moteur de recherche en ligne facilitant la sélection des échantillons par les chercheurs en utilisant des critères différents.

Expression	CK+	MMI
Dégout	59	28
Colère	45	32
Peur	25	28
Joie	69	42
Tristesse	28	32
Surprise	83	41
Total	309	203

Table 4.2: Nombre de sujets pour chaque classe d'expression

✓ La base JAFF

La base de données JAFFE comprend 213 images d'expressions faciales en niveaux de gris concernant 10 femmes japonaises. Il y a sept différentes expressions faciales composées de joie, de tristesse, de peur, de surprise, de colère, de dégoût et de neutralité. Chaque sujet a deux à quatre images différentes avec une résolution de 256 * 256 pixels pour chaque expression. La figure 4. 16 montre quelques images comprenant sept expressions faciales de base de la base de données JAFFE.

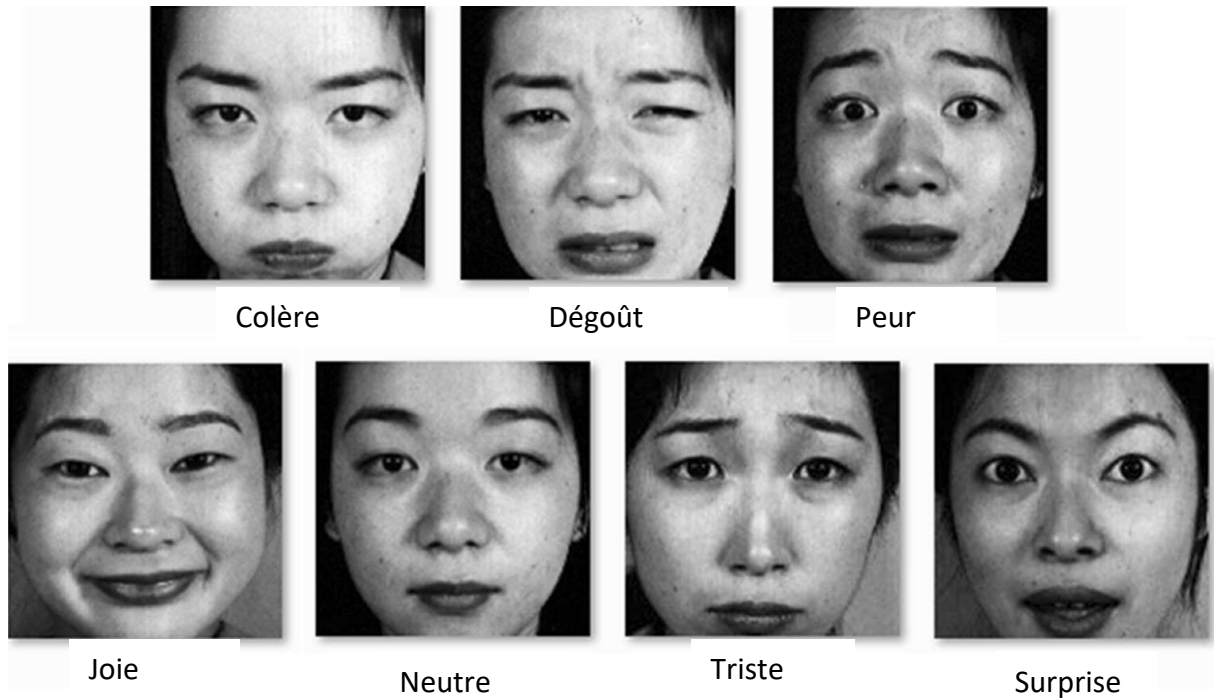


Figure4.16: Exemple d'images d'expressions faciales à partir de la base de données JAFFE

b. Présentation des résultats:

Notre système de reconnaissance des six émotions de base suit les étapes suivantes :

- Détection du visage
- extraction des points stratégiques
- estimation du mouvement de ces points pour établir un vecteur des AUs,
- et ensuite faire entrer ce dernier à la carte de Kohonen. Cette dernière étape fournit le neurone vainqueur après un certain nombre d'itérations, et selon le numéro du neurone vainqueur, nous pouvons conclure qu'il est associé à quelle émotion : comme il est montrée dans les figures 4.17 et 4.18 :

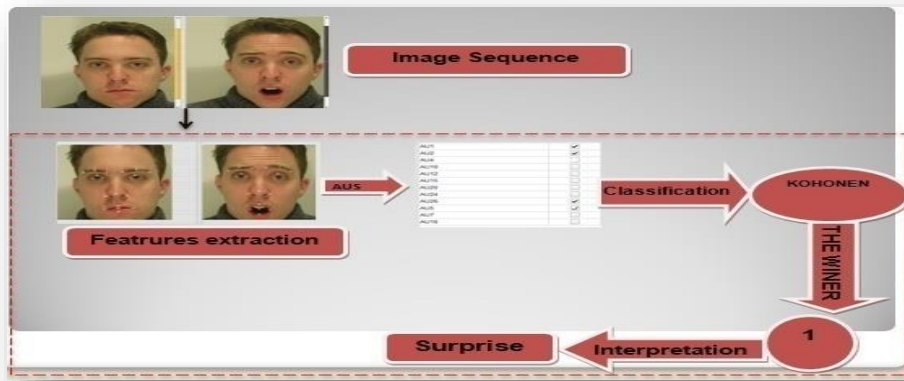


Figure4.17 : Etapes de reconnaissance de l'émotion (MMI)

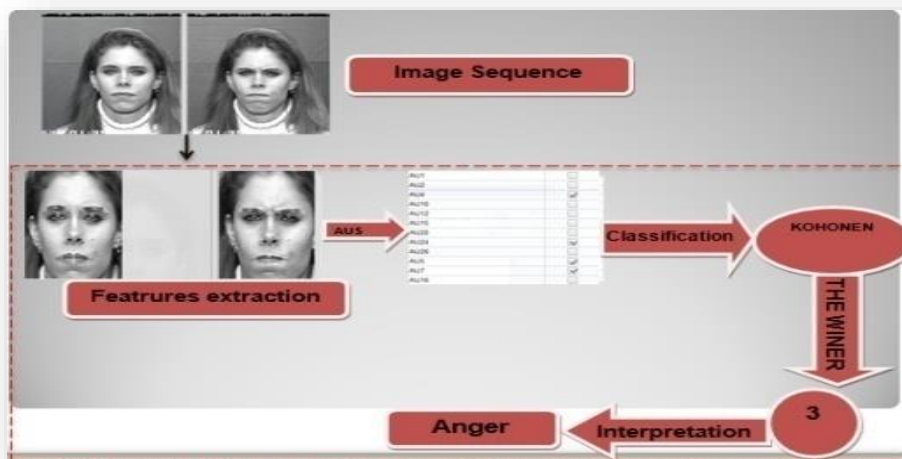


Figure4.18: Etapes de reconnaissance des émotions (Cohoen-Kanade)

✓ La détection du visage :

Les images manipulées sont des images qui ne contiennent que des visages, mais cette phase représente un prétraitement qui sert à localiser ou centrer le visage dans l'image. Pour l'amélioration des résultats.



Figure 4.19 : Détection du visage (Base MMI)



Figure 4.20 : Détection du visage (Base CK+)

✓ L'extraction des points stratégiques :

Comme nous l'avons cité ci-dessus les points stratégiques extraits sont dix comme le montre la figure 4.21



Figure 4.21 : Extraction des points stratégiques

✓ Estimation du mouvement

Les déplacements des points stratégiques sont montrés dans la figure 4.22



Figure 4.22: Estimation du mouvement des points stratégiques

✓ La déduction des AUs

Grace à l'estimation du mouvement, nous pouvons détecter les AUs (Fig 4.23)



Figure 4.23 : Les AUs associé à l'émotion de la peur

c. Evaluation des résultats

Nous avons suivi trois séries d'expériences pour évaluer la performance de notre méthode. Le premier ensemble d'expériences est conçu pour évaluer les performances en matière de détection d'AU basée sur l'image.

Plus précisément, nous fournissons des évaluations de performance quantitatives pour les bases de données posées et spontanées, et nous montrons que l'utilisation de fonctions dynamiques améliore les résultats de la détection d'AU par rapport à leurs homologues statiques (Table 4.7), ce qui améliore la reconnaissance des émotions. Les images que nous avons utilisées pour ce test proviennent de la base de données Cohn-Kanade et Cohen Kande + [Kanade et al, 2000]. Ce sont des images séquentielles consistant en un ensemble d'expressions faciales prototypiques (la joie, la colère, la peur, le dégoût, la tristesse et la surprise) qui ont été recueillies auprès d'un groupe d'étudiants en psychologie de différentes races ayant entre 18 et 30 ans. Chaque séquence commence à partir d'une expression neutre et se termine au pic d'une expression donnée. Nous avons sélectionné la première et la dernière image de chaque séquence.

La dernière série d'expériences explore la généralisation des résultats. À cette fin, nous effectuons une série d'expériences de bases de données croisées; nous avons fait

l'entraînement sur la base de données Cohn-Kanade et les tests sur la base de données MMI, (table4.3). Aussi, nous nous sommes exercés sur CK+ en faisant des tests sur la base de données JEFF (table 4.4).

Emotion	Joie	surprise	dégout	colère	Peur	Tristesse	Taux de reconnaissance
Joie	100	0	2	0	0	0	98.03
surprise	0	127	0	0	4	0	96.94
dégout	0	0	126	1	0	3	96.92
colère	130	0	0	0	3	0	97.74
peur	0	0	0	3	116	0	97.47
Tristesse	0	0	3	0	0	111	97.39

Table 4.3. : Taux de reconnaissance des émotions de base obtenus avec la base de données CK

Emotion	Joie	surprise	dégout	colère	Peur	Tristesse	Taux de reconnaissance
Joie	120	0	2	3	0	0	96
surprise	0	118	0	0	6	0	95.16
dégout	2	0	99	0	0	4	96.11
colère	0	0	0	113	6	0	94.95
peur	0	4	0	3	126	0	94.02
Tristesse	0	0	4	0	0	120	96.77

Table 4.4. : Taux de reconnaissance des émotions de base obtenus avec la base de données JAFFEE

Dans le cas où l'apprentissage et les tests sont effectués sur la même base de données, le taux de reconnaissance est relativement élevé, par rapport à un apprentissage avec une base de donnée et effectuer les tests sur une autre base (table 4.5)

Emotion	Joie	surprise	dégoût	colère	Peur	Tristesse	Taux de reconnaissance
Joie	141	0	0	1	0	0	99.29
surprise	0	139	0	0	4	0	97.20
dégoût	0	0	125	0	0	4	96.89
colère	0	0	0	130	4	0	97.01
peur	0	6	0	2	126	0	94.02
Tristesse	0	0	4	0	0	110	96.49

Table 4.5 : Taux de reconnaissance des émotions de base MMI

Emotion	Joie	surprise	dégoût	colère	Peur	Tristesse	Taux de reconnaissance
Joie	120	0	0	1	0	0	99.17
surprise	0	135	0	0	2	0	98.54
dégoût	0	0	119	0	0	3	98.34
colère	0	0	0	111	3	0	97.36
peur	0	2	0	1	136	0	97.84
Tristesse	0	0	3	0	0	119	97.54

Table 4.6 : Taux de reconnaissance des émotions de base CK

D'après les résultats de reconnaissances présentées dans les tableaux ci-dessus, nous remarquons toujours un conflit dans la détection de quelques émotions. Nous remarquons que les unités d'actions représentant l'émotion de Surprise sont incluses dans l'ensemble des AUs représentant l'émotion de Peur.

On trouve que pour le mouvement de la bouche d'une personne surprise est le même que celui de la bouche d'une personne qui a peur. Il s'agit de l'unité d'action Au26 où l'ouverture ou la séparation des lèvres constituent presque le même mouvement. La distance entre les deux lèvres réduira le conflit.

Pour les mouvements des sourcils et des paupières (Au1, Au5) aussi, ils sont les mêmes dans la description de ces deux émotions. C'est pourquoi quelques émotions de surprise ont été classé comme peur est vice versa.

Aussi pour les deux émotions de Dégout et de Tristesse, l'abaissement des coins externes des lèvres, est un mouvement commun, ce qui rend la forme des lèvres pas trop différentes dans les deux cas. Et donc un petit conflit peut apparaître. Il en est de même pour la Colère et la Peur, l'abaissement et le rapprochement des sourcils et l'ouverture entre la paupière supérieure et les sourcils est aussi un mouvement commun.

Ces erreurs de reconnaissance dues à la description de quelques mouvements communs sont presque négligeables grâce au principe d'auto-organisation offert par la carte de Kohonen. Ce Qui a influé sur le taux de reconnaissance qui est très satisfaisant par rapport à d'autres méthodes utilisant d'autre classifieurs (table4.8).

d. Discussion et comparaison

Beaucoup de travaux récents et antérieurs ont proposé des systèmes de reconnaissance des émotions. Chacun présente une architecture spécifique. Nous présentons dans cette section une étude comparative entre notre système de reconnaissance des émotions de base et quelques travaux cités dans la littérature:

Commençant par la détection des AUs, nous avons proposé une approche de détection des mouvements des muscles faciaux. Cette approche dynamique a amélioré le taux de détection.

Travail	Base de données	Caractéristiques faciales	Performance
[Lien et al,2000]	CK	Suivi des traits faciaux	HMM: 81%
[Tong et al, 2007]	CK et MMI	Ondelettes de Gabor	Adaboost: 91.2%
[Smith et al, 2001]	CK databases	Filtre de Gabor	combinaison: des AUs 93%
[Koelstra et Pantic,2008]	MMI et CK	Caractéristiques géométriques	SVM: 60%
Notre approche	MMI et CK	block matching	Dynamique 94%
Notre approche	CK+ et JEFF	block matching	Dynamique 96%

Table 4.7. : Comparaison de la détection des unités d'action

Dans notre système proposé pour la reconnaissance de six émotions de base,définies en tant que six émotions fondamentales par le psychologue Paul Ekman qui sont: (joie, tristesse, colère, dégoût, surprise, peur), nous avons manipulé des bases de données spéciales pour la reconnaissance émotionnelle.

Comme critères d'évaluation, nous avons minimisé : les données traitées dans l'apprentissage et le temps d'exécution.

Par rapport à la tâche de reconnaissance faciale, notre approche présente un gain considérable en termes de temps de calcul et de ressources utilisées. Toute image émotionnelle peut contenir plus de 20 000 pixels. Pour pouvoir manipuler toutes ces informations, toute application de reconnaissance d'émotions prendra un temps d'exécution

considérable. Cependant, notre approche manipule seulement un vecteur de 11 éléments qui représente toutes les informations dans l'image de l'émotion chose qui largement suffisant pour la reconnaissance.

En effet, considérons le travail d'Anima [Anima et al, 2014]. Dans ce travail, les auteurs présentent une approche de reconnaissance d'expression faciale basée sur le SOM. Ils utilisent en entrée un vecteur de 26 éléments, alors que notre approche utilise un vecteur d'entrée avec seulement 10 éléments, en utilisant la même base de données MMI.

Par rapport à d'autres travaux existants (table 4.8), notre approche présente un taux de reconnaissance très satisfaisant :

Auteur	traitement	Classifieur	Taux de reconnaissance
[Palo et al,2014]	Coefficients de prédiction linéaire	perception multicouche	56.20%
[André et al, 2017]	CNN	CNN	96.76
[Damir, et al., 2013]	ACP	les réseaux de neurones	70
[Md et al, 2017]	LBP	HMM	80
[Belhouchette, et al (notre système) ,2018]	Détection du mouvement	Catre de Kohonen	96%

Table4. 8. : Résultats de comparaison du système proposé avec certaines méthodes de reconnaissance des émotions.

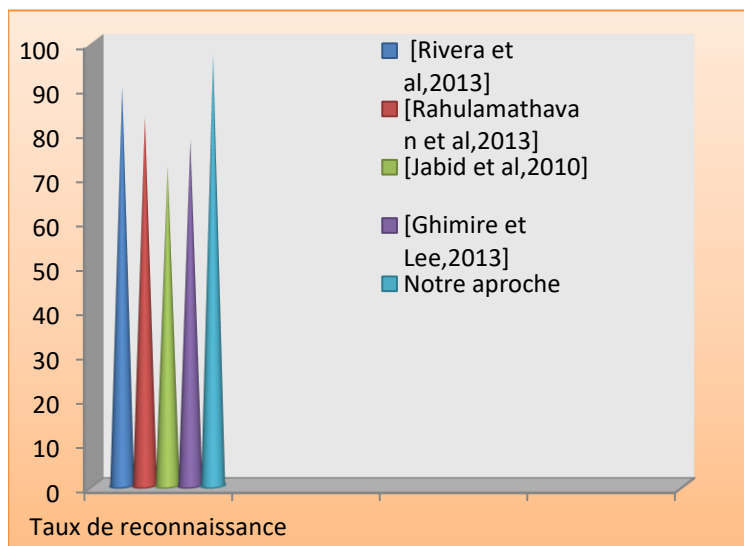


Figure 4. 24: Résultats de comparaison du système proposé avec certaines méthodes de reconnaissance des émotions (unité%)

e. Mesures d'évaluation

Nous avons appliqué deux métriques d'évaluation

- ✓ Le taux de bonne classification simplifié (tbc_s) :

Il s'agit de l'indicateur le plus naturel et le plus évident permettant d'évaluer les performances d'un système de classification. Cette valeur, simple à calculer, correspond au nombre d'éléments correctement identifiés par le système. La définition du taux de bonne classification sans la prise en compte du rejet est donnée par (Eq4.20):

$$tbc_s = \frac{\text{Nombre d'éléments correctement identifiés}}{\text{Nombre d'éléments total}} \tag{4.20}$$

On obtient le taux d'erreur par :

$$te_s = 1 - tbc_s \tag{4.21}$$

Lorsque le rejet d'une forme est possible, un troisième taux, dit "taux de rejet" (tr) est intégré. Il mesure le nombre d'éléments sur lesquels le système n'a pas pris de décision. Nous obtenons ainsi :

$$te_s = 1 - tbc_s - tr \tag{4.22}$$

Le taux de rejet est utilisé pour l'évaluation de la reconnaissance de l'émotion d'intérêt, parce que la personne soit intéressée soit non.

émotion	tbc_s	te_s
Joie	0.97	0.03
Colère	0.93	0.07
Dégoût	0.89	0.11
Surprise	0.95	0.05
Peur	0.92	0.08
tristesse	0.90	0.1
Intérêt	0.93	0.06

Table 4. 9. : Evaluation de la méthode de classification

✓ La précision (P)

La précision est la fraction des annotations automatiques de l'émotion qui sont correctement reconnues et est définie comme :

$$P = \frac{TP}{TP + FP} \quad (4.23)$$

Où TP est le nombre de vrais positifs dans l'ensemble de données, et FP est le nombre de faux positifs.

Ce qui suit constitue nos résultats obtenus :

Emotion	TP	FP	P
Joie	29	2	0.93
Colère	30	1	0.96
Dégoût	25	3	0.89
Surprise	40	1	0.97
Peur	26	2	0.92
Tristesse	28	2	0.93

Table4. 10. : La précision de notre approche

f. Paramètres d'optimisation

Notre approche propose un nouveau système de reconnaissance des émotions faciales. Il est important de noter que le paramètre principal par rapport à de nombreuses autres méthodes est la minimisation des points stratégiques extraits et l'introduction de description avec mouvements de l'émotion d'intérêt, chose qui n'a jamais été proposée par des recherches antérieures.

Pour estimer les points stratégiques nécessaires et suffisants pour la reconnaissance, nous avons exploité les fondements psychologiques qui déterminent les émotions. Nous avons traduit les mouvements faciaux de chaque émotion en un déplacement des points bien

déterminé. Nous avons donc amélioré le temps d'exécution, et minimiser la taille du vecteur entrant à la carte de KOHONEN.

4.3 Nouveau modèle pour le codage de l'émotion d'intérêt par les unités d'actions :

Nous commençons la présentation de notre travail, par la description du nouveau modèle proposé qui représente l'émotion d'intérêt.

4.3.1 Définition de l'intérêt

L'intérêt est décrit dans certaines théories de l'émotion discrète comme une émotion fondamentale qui attire l'attention. Les concepts d'intérêt peuvent être liés à la «faim envers l'information» : La recherche d'information, l'apprentissage et la compréhension. L'émotion d'intérêt se manifeste par la libération des opioïdes ou des endorphines au niveau du cerveau. Chez les personnes normales, l'émotion d'intérêt est continue dans la conscience jusqu'à ce qu'elle conduise à l'acquisition de nouvelles informations. Elle est déclenchée quand on sent l'engagement, la fascination et la curiosité. D'après Izard, l'intérêt renforce les liens entre les gens [Izard, 1991], quant à Tomkins, il a conclu qu'il n'y a pas de compétence humaine qui peut être atteinte en l'absence de l'intérêt [Tomkins et Silvan, 1962]. Cette émotion pourrait maintenir la santé du système nerveux central parce qu'elle engage la personne dans la construction active et l'organisation de la cognition

Donc, l'intérêt individuel est une prédisposition relativement durable à s'occuper de certains objets et événements et à s'engager dans certaines activités

L'intérêt est un état psychologique relativement transitoire qui comprend une attention focalisée, un fonctionnement cognitif, une persistance et une implication affective [Krapp et al, 1992].L'intérêt détermine la façon dont les individus assistent, choisissent et persistent à traiter certains types d'informations. Il a à la fois des composantes de motivation et de but, en particulier pour l'apprentissage, l'exploration et la recherche d'information (par exemple, [Silvia, 2008]). La curiosité pour de nouvelles possibilités élargit les expériences des individus et, par la suite, soutient le développement continu des connaissances et des compétences [Silvia,2005].

4.3.2 L'intérêt et le e-learning

Notre première contribution est la proposition d'un nouveau modèle qui code et décrit l'émotion d'intérêt, vu son objectivité dans divers domaines, plus particulièrement dans le domaine du e-learning. En effet, dans le contexte éducatif classique, l'observation et l'identification de l'état émotionnel de l'apprenant permettent à l'enseignant d'entreprendre des actions qui vont influencer la qualité de l'apprentissage et son exécution, chose qui n'est pas totalement prise en charge dans le contexte de l'enseignement à distance. En effet, la pertinence de l'apprentissage implique le suivi des apprenants et particulièrement la gestion des aspects intellectuels et socio-émotionnels [BRANGIER et al,2002] [NKAMBOU et al,2004]. Par exemple, si l'apprenant est frustré, le tuteur peut donner des indices pour faire avancer l'apprenant dans la construction de connaissances ou peut faire des commentaires empathiques propices au renforcement de la motivation. Si l'apprenant est non intéressé, le tuteur doit présenter des solutions plus attrayantes ou stimulantes pour pousser l'apprenant à travailler.

Prédire la réaction émotionnelle de l'apprenant dans un environnement d'apprentissage à distance n'est pas une tâche facile. Néanmoins, des tentatives théoriques ont été faites dans le but d'améliorer ces environnements en considérant et en contrôlant l'état émotionnel courant de l'apprenant.

Ces pourquoi les recherches récentes tentent d'intégrer les nouveaux moyens technologiques qui peuvent améliorer la qualité de ce type d'enseignement, par la reconnaissance des émotions de l'apprenant à distance. Par exemple, Kort et al ont proposé un modèle qui lie explicitement l'apprentissage et les états affectifs. Le modèle décrit le changement de l'état affectif durant des séances d'apprentissage.[Kort et al, 2001]

Récemment, Ayvaz et al ont développé un nouveau système de reconnaissance des émotions faciales (FERS), qui reconnaît les états émotionnels et les niveaux de motivation des élèves dans les systèmes d'apprentissage en ligne de type vidéo conférence. Le taux de reconnaissance était satisfaisant. Mais les émotions détectées ne donnent pas réellement une information claire pour l'enseignant. Tandis que savoir si l'étudiant à distance est intéressé ou pas est plus objectif. [Ayvaz et al, 2017]

Dans ce cadre, nous proposons un système de codage et de reconnaissance de l'émotion d'intérêt, qui peut nous fournir des informations concernant l'apprenant, si ce dernier est intéressé ou pas.

4.3.3 Codage proposé de l'émotion d'intérêt

Il existe en littérature des études approfondies sur les expressions faciales, mais il y a peu de recherches faisant le lien avec l'émotion d'intérêt. Néanmoins, Tomkins et Izard ont jugé que pour l'émotion d'intérêt, les sourcils soient légèrement abaissés ou élevés, les paupières inférieures soulevées, les lèvres séparées. Dans l'ensemble, il y a des indices préliminaires pour des expressions faciales exprimant l'intérêt [Silvia, 2001]. (Par exemple les sourcils soient légèrement abaissés veut dire que ... etc.)

Par conséquent, notre objectif est de trouver un modèle intégrant ces expressions en se basant sur les fondements psychologiques et les traduire en unités d'actions associées au système FACS. De ce fait, nous avons établi un lien entre les mouvements du visage exprimant l'émotion d'intérêt et les AUs (Table 4.11).

Partie du visage	Muscle concerné	mouvement associé	action unit
les sourcils	Dépresseur Supercilii	légèrement réduit	AU42
	muscle frontales	légèrement surélevée	AU1
Les yeux	muscle orbiculaire	paupières inférieures soulevées (pour aiguïser focus visuel)	AU5
La bouche	muscle abaisseur de la lèvre inférieure ou relaxation du muscle orbiculaire	lèvres entrouvertes	AU25
Les joues	muscle orbiculaire de l'œil (pars orbitalis)	joues soulevées	AU6

Table 4.11. Les unités d'action associées à l'intérêt

Cette combinaison des AUs nous fournit un modèle précis pour la reconnaissance de l'émotion d'intérêt.

4.3.4 Approche proposée pour la détection de l'émotion d'intérêt

Pour l'émotion d'intérêt, malheureusement, les données d'apprentissage et de tests sont très rares. Nous avons appliqué notre approche sur des images qui ont été recueillies en 2004. Un total de 99 étudiantes dont 64 femmes et 35 hommes.

Les mêmes étapes décrites ci-dessus ont été suivies pour la reconnaissance. Mais les caractéristiques extraites changent bien évidemment, ce qui implique un changement dans la carte de classification auto-organisatrice.

En effet, puisque le codage proposé limite les unités d'action associées à l'intérêt, et les changements des muscles sont bien décrits, nous ne manipulons que les points décrits dans la table 4.12.

action unit	Points d'intérêt
AU42/AU1	P2 P3
AU5	P1 P4
AU25	P7 P8
AU6	P5 P6

Table 4.12. : Les points stratégiques de l'émotion d'intérêt

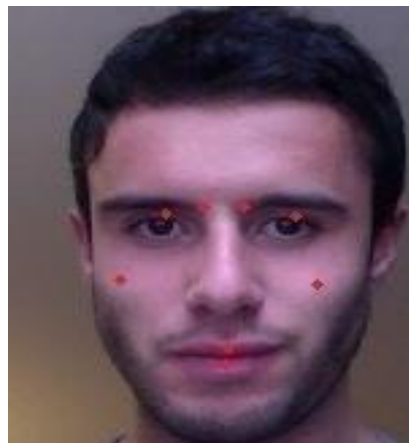


Figure 4.25 Les points stratégiques de l'émotion d'intérêt

Après avoir extrait le visage et les points stratégiques, nous effectuons une estimation du mouvement de ces points, en partant de l'image neutre à l'image intéressée. Avec l'algorithme de mise en correspondance de blocs.

Ensuite, le vecteur sortant est de 8 éléments et est entré à la carte de Kohonen, pour l'étape de la classification. La couche de sortie contient deux neurones, soit la personne est intéressée, soit elle ne l'est pas.

4.3.5 Résultats

a. Ensemble de donnée

La reconnaissance des émotions a grandement bénéficié des nombreuses bases de données qui ont été produites pour l'étudier. Mais très peu sont les bases de données qui expriment l'émotion d'intérêt.

La plupart des bases de données existantes ont été créées dans des conditions contrôlées pour faciliter l'étude de paramètres spécifiques sur le problème de la reconnaissance faciale. Ces paramètres incluent des variables telles que la position, la pose, l'éclairage, l'expression, l'arrière-plan, la qualité de la caméra, l'occlusion, l'âge et le sexe, ce qui limite la reconnaissance réelle de l'émotion d'intérêt.

✓ *LFW data set*

La motivation principale pour la base de données LFW [Kawulok et al, 2016] [Gary et al, 2007] (Labeled Faces in the Wild) pour la reconnaissance d'émotion est qu'elle fournit un grand ensemble d'images de visage dans des conditions non contrôlées. Parmi les images que contient cette base de données est que peu reflètent l'intérêt. Nous l'avons alors filtré pour obtenir nos données de test. (Fig 4.26)



Figure 4.26 : Quelques images de la base de données LFW

b. Evaluation des résultats



Figure 4.27 : La détection de l'intérêt

Emotion	Intéressé	Non Intéressé	Taux de reconnaissance
Intéressé	38	5	88.37
Non Intéressé	42	4	91.3

Table4.13. : Taux de reconnaissance de l'émotion d'intérêt

✓ *Mesure d'évaluation :*

Emotion	tbc_s	te_s
Intéressé	0.88	0.11
Non Intéressé	0.91	0.09

Table 4.14. : Evaluation de la méthode de classification

✓ *La précision (P)*

Suivant nos résultats obtenus :

Emotion	TP	FP	P
Intéressé	40	4	0.90
Non Intéressé	38	3	0.92

Table4.15. : La précision de notre approche

c. Paramètres d'optimisation

Notre approche propose un nouveau modèle pour le codage et la reconnaissance de l'émotion d'intérêt. Sur la base d'études de psychologie (Silvia, 2001) et d'autres recherches (Belhouchet et al. 2015). L'émotion d'intérêt est décrite mais aucun lien entre la description psychologique et l'informatique n'a été proposé.

Nous avons fait un lien entre les mouvements du visage qui exprime cette émotion et les unités d'action du système FACS. C'est le paramètre principal d'optimisation par rapport à d'autres travaux cité ci-dessus.

Les mêmes étapes que nous avons suivies dans la reconnaissance des émotions de base dans le chapitre précédent ont été suivies.

Les points stratégiques sont réduits à 8 et l'architecture de la carte auto-organisatrice de Kohonen est changée : 8 neurones en entrée et 2 en sortie. En effet, il y a deux classes, soit la personne est soit intéressée ou non.

d. Discussion

La détection de l'émotion d'intérêt est un grand problème posé dans divers domaines : pour la prédiction du comportement des clients envers une promotion par exemple, le domaine du e-learning,...etc. Son identification a fait l'objet de notre étude de thèse. Nous avons proposé un modèle géométrique pour la représentation de cette émotion en se basant sur des fondements psychologiques déjà cités.

La majorité des travaux antérieurs sur la reconnaissance de cette émotion se sont basés sur des tests de psychologie, comme le travail de Nummenmaa et al. Leur étude s'est intéressée à examiner comment les émotions sont ressenties lors d'un apprentissage à distance : l'intérêt des étudiants pour le sujet du cours et l'intérêt pour l'apprentissage sur le Web sont associés à des activités invisibles, collaboratives visibles, et non collaboratives [Nummenmaa et al, 2008].

Aussi dans le travail d'Ayvaz et al, le système d'information proposé détecte les états émotionnels des apprenants et donne un feedback à un éducateur sur leurs états émotionnels instantanés et pondérés basés sur les expressions faciales. Un taux de reconnaissance très satisfaisant a été atteint par cette étude [Ayvaz et al, 2017]. Mais la reconnaissance est limitée aux six émotions de base. Les auteurs Ainleyn et al ont étudié comment les facteurs d'intérêt individuels et situationnels contribuent dans l'intérêt à l'apprentissage. Les mesures traditionnelles d'auto-évaluation ont été combinées avec de nouvelles méthodes informatisées interactives d'enregistrement des réactions cognitives et affectives aux textes de science et de culture populaire, en surveillant leur développement en temps réel. Leur travail a permis de mieux comprendre comment l'intérêt influence l'apprentissage. [Ainleyn et al, 2002]

L'introduction d'une dynamique, basée sur l'estimation des mouvements faciaux, représente une vraie contribution dans ce domaine. C'est pour cette raison que notre approche a fait un changement dans le raisonnement envers la détection de l'émotion d'intérêt qui était réservée à la détection de l'état affectif général de la personne et de déduire si elle est intéressée ou non.

Le problème principal qui se pose au niveau de notre système de reconnaissance de l'émotion d'intérêt, c'est l'absence des bases des données riches. Parce qu'en générale toute

approche de reconnaissance basée sur l'apprentissage a besoin de grandes bases de données pour un meilleur entrainement, et donc de meilleures résultats.

4.4 Conclusion

Dans le présent chapitre, nous avons présenté notre système de reconnaissance des émotions de base. Le processus de reconnaissance consiste en premier d'effectuer une détection de visage avec la méthode de Viola Jones, ensuite une extraction des caractéristiques faciales qui sont les points stratégiques décrivant les émotions, puis l'estimation de mouvement de ces points qui nous fournit les unités d'action d'Ekman, et en la classification de ces dernières afin de nous donner l'émotion reconnue.

L'estimation de mouvement est réalisée par la méthode de mise en correspondance des blocs (Blocs matching), alors que pour la classification, nous avons exploité la carte auto-organisatrice de Kohonen. Un taux très satisfaisant est atteint, avec une optimisation dans le temps de calcul et les ressources utilisées.

Comme deuxième principale contribution, nous avons aussi présenté un nouveau système de reconnaissance de l'émotion de l'intérêt basé sur des fondements psychologiques.

CONCLUSION GENERALE

CONCLUSION GENERALE

La reconnaissance des émotions est l'un des domaines scientifiques les plus complexes. Ces dernières années, de plus en plus d'applications tentent de l'automatiser. Ces applications innovantes concernent plusieurs domaines comme l'aide aux enfants autistes, les jeux vidéo, l'interaction homme-machine. Les émotions sont véhiculées par plusieurs canaux. Nous traitons dans cette thèse les expressions émotionnelles faciales en s'intéressant particulièrement aux six émotions de base à savoir la joie, la colère, la peur, le dégoût, la tristesse et la surprise.

Dans cette thèse nous avons proposé comme première contribution un système de reconnaissance des émotions de base (joie, colère, dégoût, surprise, peur,) qui est basé sur l'estimation du mouvement et la carte auto-organisatrice de Kohonen. Avec l'exploitation de système FACS développé par les psychologues.

L'approche de reconnaissance des émotions proposée suit les trois étapes fondamentales d'un système de FER (Facial Emotion Recognition). Il s'agit de la détection de visage, l'extraction des caractéristiques faciales et la classification.

Nous avons détecté le visage avec la méthode de Viola Jones. Elle est basée sur des classifieurs en cascade en utilisant les caractéristiques (features) de types Haar. C'est une méthode supervisée, basée sur Adaboost Apprentissage de classifieurs faibles et apprentissage en cascade. Le résultat de l'extraction est parfait.

En suite l'extraction des points stratégiques dans le visage avec la méthode d'apparence active. Les points sont réduits à dix.

Pour pouvoir détecter les Aus nous avons effectué une estimation de mouvements des points stratégiques par la méthode de Bloc matching.

A ce stade, chaque séquence d'image est représentée par une combinaison des Aus, ces informations sont l'entrée de notre carte auto-organisatrice de KOHONEN. Pour effectuer la classification, et affirmer de quelle émotion s'agit-il.

Les résultats étaient très satisfaisants comme nous l'avons présenté, que ce soit en termes de taux de reconnaissance ou de temps de calcul.

La deuxième contribution de cette thèse c'était l'introduction d'un modèle bien précis décrivant l'émotion d'intérêt. Ce qui aide à la détection automatique de cette émotion, Qui est très importante dans divers domaine comme le e-learning.

Cependant, le manque de bases de données reflétant cette émotion a influé sur le jugement de reconnaissance. Nous souhaitons dans nos futurs travaux élaborer une base de données spécifique à cette émotion pour pouvoir améliorer notre modèle de description.

Aussi pour le raisonnement de reconnaissance des émotions d'émotions nos futurs travaux vont prendre l'axe d'actualité dans les réseaux de neurones, étant le Deep learning (Apprentissage profond), ou les réseaux de neurones conventionnels (CNN). C'est un nouveau raisonnement d'apprentissage et de classification.

ANNEXE

Une bibliothèque très riche à base de la méthode de Viola Jones a été conçue. C'est l'Open CV, nous l'avons exploité dans l'étape de détection du visage.

Open Source Computer Vision est une bibliothèque proposant un ensemble de plus de 2500 algorithmes de vision par ordinateur, accessibles au travers des API pour les langages C, C++, et Python. Elle est distribuée sous une licence BSD (libre) pour les plate-formes Windows, GNU/Linux, Android et MacOS.

Initialement écrite en C il y a 10 ans par des chercheurs de la société Intel, OpenCV est aujourd'hui développée, maintenue, documentée et utilisée par une communauté de plus de 40 000 membres actifs. C'est la bibliothèque de référence pour la vision par ordinateur, aussi bien dans le monde de la recherche que celui de l'industrie.

Les principaux modules accessibles au travers de son API C.

- ✓ les fonctionnalités de base (core) :

Cette bibliothèque permet de manipuler les structures de base, réaliser des opérations sur des matrices, dessiner sur des images, sauvegarder et charger des données dans des fichiers XML...

- ✓ Les fonctions et structures (imgproc)

Ce module est spécialisé dans les transformations d'images, au filtrage, à la détection de contours, de points d'intérêt.

- ✓ Descripteurs (features2d) :

Ce module concerne principalement l'extraction de descripteurs selon deux approches courantes (SURF et Star Detector).

- ✓ traitement de flux vidéo :

Ces fonctions servent à segmenter et suivre les objets en mouvement dans une vidéo.

- ✓ entrées-sorties et interface utilisateur (highgui) :

OpenCV intègre sa propre bibliothèque haut-niveau pour ouvrir, enregistrer et afficher des images et des flux vidéo. Celle-ci contient aussi un certain nombre de fonctions permettant de réaliser des interfaces graphiques très simples, mais largement suffisantes pour tester les programmes.

- ✓ calibration, estimation de pose et stéréovision (calib3d) :

Ce module contient des fonctions permettant de reconstruire une scène en 3D à partir d'images acquises avec plusieurs caméras simultanément.

- ✓ détection d'objets (objdetect) :

Cette bibliothèque permet de faire de la reconnaissance d'objets dans une image au moyen de l'algorithme Adaboost (Viola & Jones, 2001). C'est ce dernier module que nous avons manipulé pour la détection de visage dans notre système.

REFERENCES BIBLIOGRAPHIQUES

Références Bibliographiques

- [Ainleyn et al,2002] Ainleyn.M., Hidi.S ., Berndorff.D., « Interest, Learning, and the Psychological Processes That Mediate Their Relationship » Journal of Educational Psychology, Vol. 94, No. 3, 545–561.,(2002)
- [Alain,1900]
[André et al,2017] CAFE-PHILO à la Maison des Savoirs Emotion
André.T., Aguiar.E., Alberto.F ., Souza.D., Thiago.O., « Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order »,Pattern Recognition 61.,610–628.,(2017).
- [Anima et al, 2014] M.Anima ., B.Laxmidhar.,Venkatesh K.,«Emotion recognition from geometric facial features using self-organizing map»,Journal Pattern Recognition., V47.2014
- [Ayvaz et al,2017] Ayvaz.U., Gürüler.H., Osman.M., « Devrim USE OF FACIAL EMOTION RECOGNITION IN E-LEARNING SYSTEMS »,. ISSN: 2076-8184. Information Technologies and Learning Tools, Vol 60, №4.,(2017).
- [Baldi et al,1995] Baldi P., Chauvin Y., Hunkapiller T., McClure M., « Hidden Markov models of biological primary sequence information », Actes de Natural Academic Sciences, vol. 91- 3, pp. 1059-1063, 1995
- [BENCHERT et al,2005] BENCHERT., BOUALLEG., TEBIKH., « Segmentation de la couleur de peau par seuillage selon différents espaces couleurs », 3^{ème} journée internationale sur l’informatique graphique en TUNISIE .,(2005).
- [BRANGIER et al,2002] BRANGIER.E.,BOBILLIER.M.al., « Analyse psycho-ergonomique de l'interaction entre l'homme et les NTIC : introduction à une psychologie de l'environnement digitale », éditions INRS, Paris, vol 4, numéro 189, pp. 15-26.,(2002).
- [Breuer et Kimmel,2017] Breuer.R., Kimmel.R., « A deep learning perspective on the origin of facial expressions »,. arXiv:1705.01842.,(2017).
- [Buciu et al,2003] Buciu.I .,Kotropoulos.C ., Pitas.I., « ICA and Gabor representation for facial expression recognition ». In Proc. of IEEE Int. Conference on Image Processing, pages 855–858., (2003).
- [Bui et al, 2004] Bui.T. D., Heylen. D.,. and Nijholt. A., « Combination of facial movements on a 3D talking head »,in Computer Graphics International. Proceedings, pp. 284–290., (2004).
- [Cheon et Kim,2008] Cheon.Y et Kim.D., « A natural facial expression recognition using differential-aam and k-nns ».Multimedia, International

- Symposium on, 0:220–227.,(2008).
- [Cheng et Greiner,1999] Cheng.J., Greiner.R., « Comparing bayesian network classifiers », In Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-99), pages 101–108, San Francisco, CA., Morgan Kaufmann Publishers.,(1999).
- [Christian,2016] Christian .G., « cours de GIF-4101 : Apprentissage et reconnaissance », Faculté des sciences et de génie Département de génie électrique et de génie informatique, (2016).
- [Chu et al,2017] Chu.W.S., Torre.F.D., Cohn. J.F., « Learning spatial and temporal cues for multi-label facial action unit Detection », In Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA,; pp. 1–8. (30 May–3 June 2017)
- [Chumkamon et Hayashi,2013] Chumkamon.S ., Hayashi.E ., « Facial expression recognition using constrained local models and Hidden Markov models with consciousness-based architecture », IEEE/SICE International Symposium on System Integration (SII), pp 382–387.,(2013)
- [Cohen et al,2003] Cohen.I., Cozman.F. G., Sebe.N., Cirelo.M. C., Huang T. S.,” Learning Bayesian network classifiers for facial expression recognition using both labeled and unlabeled data”, IEEE conference on Computer Vision and Pattern Recognition (CVPR), 16-22 June Madison, Wisconsin, (2003).
- [Cozman et Cohen,2002] Cozman.F., Cohen.I., « Unlabeled data can degrade classification performance of generative classifiers », In Fifteenth International Florida Artificial Intelligence Society Conference, pages 327–331,(2002).
- [DAF] Dictionnaire de l'Académie française, neuvième édition Version informatisée.
- [Damir et al,2013] *F.Damir, M. Goran.*, «Emotion Recognition System by a Neural Network Based Facial Expression Analysis», Journal for Control, Measurement, Electronics, Computing and Communications.V54.(2013).
- [Danisman et al,2013] Danisman.T., Bilasco.IM ., Martinet.J ., Djeraba C ., »Intelligent pixels of interest selection with application to facial expression recognition using multilayer perceptron », Signal Process 93(6):1547–1556., (2013)
- [Darwin,1872] Darwin.C., « The expression of the emotions in man and animals » (1872)
- [Datcu et Rothkrantz,2008] Datcu.D ., Rothkrantz.L., »Automatic bi-modal emotion recognition system based on fusion of facial expressions and emotion extraction from speech. », 8th IEEE International Conference on Automatic Face and Gesture Recognition, 2008. FG '08, (2008).
- [David,2013] David Sander., « Vers une définition de l'émotion », (2013)

- [Drume et Jalal,2012] Drume.D ., Jalal.A.S., « A multi-level classification approach for facial emotion recognition », in: International Conference on Computational Intelligence and Computing Research, IEEE, Coimbatore, India, pp. 288–292.,(2012) .
- [Du et al2015] Du.Y., Wang.W.,Wang.L., « Hierarchical recurrent neural networkfor skeleton based action recognition », In Proc. IEEE Conference onComputer Vision and Pattern Recognition, pp. 1110-1118.,(2015).
- [Ekman et Friesen,1975] Ekman.P ., Friesen.W.V., « Unmasking the Face ». New Jersey: Prentice Hall., (1975).
- [Ekman,1993] Ekman.P., « Facial expression and emotion. American psychologist », 48(4) :384– 392, (1993).
- [Eveno,2003] Eveno.N., « Segmentation des l'èvres par un modèle déformable analytique. Spécialité signal image parole telecoms », LIS laboratory Grenoble France.,(2003).
- [Fasel,2002] Fasel.I., Bartlett.M., Movellan.J., « A comparison of gabor filter methods for automatic detection of facial landmarks », Fifth International Conference on automatic face and gesture recognition, pages 345–350, (2002).
- [Friedman et al 1997] Friedman.N., Geiger.D., Goldszmidt.M., « Bayesian network classifiers » Machine Learning, 29(2-3) :131–163, (1997).
- [Friedman et al,1997] Friedman.N., Geiger.D., Goldszmidt.M., « Bayesian network classifiers », Machine Learning, 29(2-3) :131–163., (1997).
- [Garrison et al,2000] Garrison.W., Mathew. N., DAILEY, Curtis PADGETT and Ralph. A. « Is All Face Processing Holistic ? The view from UCSD ». Chapitre à paraître dans Computational, Geometric, and Process Perspectives on Facial Cognition : Contexts and Challenges (M. Wenger and J. Townsend)., (2000).
- [Gary et al,2007] Gary.B., Huang, Ramesh.M., Berg.T., Learned-Miller.E., « Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments » University of Massachusetts, Amherst, Technical Report 07-49, October., (2007)
- [Geiger,1992] Geiger.D., « An entropy-based learning algorithm of bayesian conditional trees ». In Uncertainty in Artificial Intelligence : Proceedings of the Eighth Conference, pages 92–97, (UAI-1992),
- [Ghimire et Lee,2013] Ghimire.D., Lee. J., »Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines ». Sensors 13(6), 7714–

- 7734 ., (2013)
- [Greiner et al,1997] Greiner.R., Grove.A., Schuurmans.D., « Learning Bayesian nets that perform well », In Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97), pages 198–207, San Francisco, CA., (1997)
- [Gyanendra et al,2012] Gyanendra.P., Rakesh Kumar.R., Nagar.M., “Recognizing of Emotions using Neural Network Hopfield “., National Conference on Future Aspects of Artificial intelligence in Industrial Automation., India .,Proceedings published by International Journal of Computer Applications® (IJCA).,(2012).
- [Haussler et Baum,1989] Haussler.D., Baum.E ., « What size net gives valid generalization? », Neural Computation.;1(1):151-160.,(1989).
- [Hayet et al,2014] Hayet.B., Chtourou.M., Ben Amar.C., Chen.L., « Facial expression recognition based on a mlp neural network using constructive training algorithm », Springer Science+Business Media New York .,(2014).
- [Hupont et al,2013] I. Hupont, S. Baldassarri, et E. Cerezo., « Facial emotional classification : from a discrete perspective to a continuous emotional space, »Pattern Analysis & Applications.,(2013).
- [Izard,1991] Izard. C. E., “The Psychology of Emotions”., New York, NY: Plenum Press. (1991).
- [Jabid et al,2010] Jabid.T., Kabir.Md.H., Chae. O., « Robust facial expression recognition based on local directional pattern ». ,ETRI J. 32(5), 784–794 ., (2010).
- [Jack et Ball,1998] Jack .B., Ball.G., « Bayesian Networks for Modeling Emotional State and Personality », American Association for Artificial Intelligence., Pages10 Number MSR-TR-98-41.,(1998).
- [Jacques ,2015] Jacques Cosnier., « PSYCHOLOGIE des EMOTIONS et des SENTIMENTS », (2015)
- [Jacques ,2015] Jacques Cosnier., « PSYCHOLOGIE des EMOTIONS et des SENTIMENTS », (2015)
- [Jung et al,2015] Jung.H., Lee. S., Yim. J., Park.S., Kim. J., « Joint fine-tuning in deep neural networks for facial expression Recognition » . In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, ; pp. 2983–2991.,(7–12 December 2015)
- [Kanade et al 2000] Kanade.T., Cohn.JF., Yingli.T., « Comprehensive database for facial expression analysis », In: IEEE 4th international conference on automatic face and gesture recognition. Pittsburgh, PA, USA, (2000) . pp 46–53
- [Kant,1996] Kant., « Kant et la question de l’affectivité , lecture de la troisieme critique par Jérôme de Gramont ».Paris, librairie philosophique J.Vrin., (1996)
- [Kapmann et Zhang,1998] Kapmann .M., Zhang.L ., « Estimation of eye, eyebrow and nose features in videophone Sequences ». International

- Workshop on Very Low Bitrate Video Coding (VLBV 98), pages 101–104., (1998).
- [Kass et al,1987] Kass.M., Withins.A et Tersopolos.D., « Snakes : Actives contours models ». International Journal of computer vision, pages 321–331, (1987).
- [Kaulard et al, 2012] Kaulard, K., Cunningham, D.W., Bühlhoff, H.H., Wallraven, C. « The MPI facial expression Database, A validated database of emotional and conversational facial expressions ». PLoS ONE, 7. (2012)
- [Kawulok et al,2016] Kawulok.M., Celebi.E., Smolka.B., « Labeled Faces in the Wild: A Survey », In Advances in Face Detection and Facial Image Analysis, Springer, pages 189-248., (2016).
- [Kim et al,2017] Kim.D., Baddar.W., Jang.J., Ro.Y., « M.Multi-objective based Spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition ». IEEE Trans. Affect. Comput. (2017).
- [Kleinginna P et Kleinginna A ,1981] Kleinginna, P.R., Kleinginna, A.M. « A categorized list of emotion definitions with suggestions for a consensual definition ». Motiv. Emotion , 5 : 345-379. (1981)
- [Kotropoulos et Pitas,1997] Kotropoulos. C., Pitas.I., « Rule-based face detection in frontal Views ». In: 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP-97, vol. 4, pp. 2537–2540. IEEE (1997)
- [Koelstra et Pantic,2008] Koelstra. S., Pantic. M., « Non-rigid registration using free-form deformation for recognition of facial actions and their temporal dynamics », In Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG'08), 17-19 Sep., Amsterdam., (2008).
- [Kohonen,1990] Kohonen. T., « The self-organizing map », Proc.IEEE 78(9), 1464–1480., (1990).
- [Kort et al,2001] Kort.B., Reilly.R.,Picard.R.W., « An Affective Model of Interplay between Emotions and Learning: Reengineering Educational Pedagogy-Building a Learning Companion », IEEE International Conference on Advanced Learning Technologies, IEEE Computer Society, 0043., (2001).
- [Kotsia et al,2008] Kotsia.I., Buciu.I., Pitas.I., « An analysis of facial expression recognition under partial facial image occlusion », Image Vision Comput. 26 (7) ., 1052–1067.,(2008).
- [Krapp et al,1992] Andreas Krapp.S., Hidi.K., Renninger.A., »Interest, learning, and development », January (1992).
- [Lien et al,2000] Lien. J. J., Kanade. T., Cohn. J. F., Li. C. C., « Detection, Tracking, and Classification of Action Units in Facial Expressions », J. Robotics and Autonomous System., vol. 31, pp. 131-146.0(2000)

- [Liu et al,2015] Liu.,Shaoxin.L. , Shan.S., Xilin.C., « AU-inspired Deep Networks for Facial Expression Feature Learning », Lab of Intelligent Information Processing of Chinese Academy of Sciences(CAS),Institute of Computing Technology,CAS,Beijing100190,China.,(2015)
- [Lu et Evans,2017] Lu.S., Evans.F ., « Haar wavelet transform based facial emotion recognition », Adv. Comput. Sci. Res. 61 .,342–346., (2017).
- [Lucey et al,2010] Lucey. P.,Cohn. J. F., Kanade. T., Saragih. J., Ambadar. Z., Matthews. I., « The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression », in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPRW '10), pp. 94–101, San Francisco, Calif, USA, (June 2010).
- [Matthews et Baker,2004] Matthews.L., Baker.S., « Active Appearance Models Revisited »,International Journal of Computer Vision , Volume 60, Issue 2, pp 135–164.,(2004).
- [Md et al,2017] Md., Uddin .Z., Mohammed.M., Ahmad.A., Mansour.Z., Giancarlo.Z., Jim.T., « A facial expression recognition system using robust face features from depth videos and deep learning », Computers and Electrical Engineering 63.,114–125.,(2017).
- [Mikolov et al,2010] Mikolov.T., Karafiat.M.,Burget.L., Cernocky.J., Khudanpur.S., « Recurrent neural network based language model ». ,In Proc. INTERSPEECH, vol. 2, pp. 3, 2010.
- [Myers et David, 2004] Myers et David G., « Theories of Emotion. »Psychology: Seventh Edition, New York, NY: Worth Publishers, p. 500. (2004)
- [Naoya et al,2016] Naoya .Y., Maria .N., Caceres.,Fernando.D., Kenji.M., « Facial Expression Recognition System for User Preference Extraction », In Distributed Computing and Artificial Intelligence, 13th International Conference pp 453-461.,(2016).
- [Niu et al,2011] Niu.J.W., Zheng. X.H., Zhao.M., Fan.N., Ding.S.T.,» Landmark automatic identification from three dimensional (3D) data by using Hidden Markov Model (HMM) ». In: IEEE 18th International Conference on Industrial Engineering and Engineering Management, Changchun, China, 3-5 September, (2011).
- [NKAMBOU et al,2004] NKAMBOU.R. ,HERITIER.V.,« Reconnaissance émotionnelle par l’analyse des expressions faciales dans un tuteur intelligents affectif », La dimension cachée de e-learning : de la motivation à l’abondan..., (2004).
- [Nummenmaa et al,2008] Nummenmaa.M., Nummenmaa.L., « students’ emotions, interest and activities in a web-based learning environment Psychology ², University of Turku, Finland British Journal of

- Educational Psychology ., 78, 163–178 q 2008 The British Psychological Society., (2008).
- [Palo et al,2015] Palo, Mihir Narayana Mohanty and Mahesh Chandra., «Use of Different Features for Emotion Recognition Using MLP Network », Journal of Computational Vision and Robotics., Proceedings of ICCVR 2014.
<https://www.paulekman.com/paul-ekman/>
- [Pandzic et Forchheimer,2003] I. S. Pandzic et R. Forchheimer, « MPEG-4 facial animation : the standard, implementation and applications ». Wiley., (2003).
- [PANTIC et Leon,2000] PANTIC.M., Leon. J., « Expert system for automatic analysis of facial expressions”.; Image and Vision Computing Journal, 18(11) :881–905.,(2000)
- [Pantic et al,2005] Pantic.M., Valstar. M.F., Rademaker. R., Maat. L., « Web-based data base for facial expression analysis »,in:Proceedings of the IEEE International Conference on Multimedia and Expo,Amsterdam,The Netherlands.,pp.317–321.,(2005).
- [Paul Ekman,1978] [Paul Ekman,1978]
- [Plutchik et Albert ,1962] Robert Plutchik et Albert F. Ax., « A CRITIQUE OF DETERMINANTS OF EMOTIONAL STATE BY SCHACHTER AND SINGER », (1962)
- [Rahulamathavan et al,2013] Rahulamathavan.Y., Phan.R.C.-W., Chambers.J.A., Parish.D.J., « Facial expression recognition in the encrypted domain based on local fisher discriminant analysis », IEEE Trans. Affect. Comput. 4(1), 83–92 ., (2013).
- [Ramona et al,2017] Ramona.M., Brad.R., Brad.R., « A Comparative Study of Block Matching Optical Flow Algorithms », TEM Journal. Volume 6, Issue 4, Pages 760-770, ISSN 2217-8309, DOI: 10.18421/TEM64-16., (November 2017).
- [Rivera et al,2013] Rivera.A.R., Castillo.J.R., Chae.O., « Local directional number pattern for face analysis: face and expression recognition », IEEE Trans. Image Process. 22(5), 1740–1752., (2013).
- [Ryan.a ., ET al, 2009] Ryan.a ., ET al., « Automated Facial Expression Recognition System », Security Technology., 43rd A International Carnahan Conference., 2009.
- [Sakai et al,1972] Sakai. T., Sakai., Nagao. M, and Kanade.T., « Computer analysis and classification of photographs of human faces », in Proc. First USA—Japan Computer Conference, (1972),
- [Samal et al,2013] Samal.A., Parida.D ., Satapathy.M., Mohanty.M., « On the Use of MFCC Feature Vector Clustering for Efficient Text Dependent Speaker Recognition »,Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) pp 305-312., (2013)

- [Sanghamitra,2016] Sanghamitra Mohanty . « Language Independent Emotion Recognition in Speech Signals » International Journal of Advanced Research in Computer Science and Software Engineering. Volume 6, Issue 10, October 2016
- [Sebe,2005] Sebe.N., Bakker.,E., Cohen.I.,Gevers.T., Huang.T., « Bimodal emotion recognition », In Proc. 5th International Conference on Methods and Techniques in Behavioral Research.,(2005).
- [Shalini et Neeta,2016] Shalini.Y, Neeta.n., « A novel approach for face detection using hybrid skin color model » Reliable Intell Environ 2:145–158.,(2016).
- [Silvia,2001] Silvia. P. J., « Interest and interests: The psychology of constructive capriciousness » ., Review of General Psychology, 5, 270-290., (2001).
- [Silvia,2005] Silvia.Paul J., « What Is Interesting? Exploring the Appraisal Structure of Interest »,5,89-102. Made available courtesy of American Psychological Association.(2005)
- [Silvia,2008] Silvia.Paul J., « Interest—The Curious Emotion’’,SAGE journals., Volume: 17 issue: 1, page(s): 57-60 .,(2008)
- [Singh et al,2003] Singh. S., Chauhan.D., Vatsa. M., Singh., « A robust skin color based face detection algorithm ». Tamkang J. Sci. Eng. 6(4), 227–234 .,(2003).
- [Smith et al,2001] Smith E., Bartlett. M. S., Movellan. J. R., « Computer Recognition of Facial Actions: A Study of Co-Articulation Effects, » In Proc. 8th Ann. Joint Symposium Neural Computation., (2001).
- [Sofiane,2001] 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 », Computer science. Université de Cergy Pontoise, (2011).
- [Sung et Tomaso,1998] Sung.K and Tomaso.P., « Example-based learning for view-based human face detection ». IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 :39–51.,(1998).
- [Sungsoo et al,2009] Sungsoo.P., Daijin.K., Patter., « Subtle facial expression recognition using motion magnification » ., Recognition Letters 30.0,708–716., (2009).

- [Taner et al,2014] Taner.M., Kristin.S., Benli., « Facial expression recognition based on anatomy » ., Computer Vision and Image Understanding 119., 1–14.,(2014).
- [TEOH et CHO,2011] TEOH.T., CHO.S., « Human Emotional States Modeling by Hidden Markov Model », Seventh International Conference on Natural Computation.,(2011).
- [Timothy et al,2001] Timothy.F., Cootes., Gareth.J., Edwards., . Christopher J., Taylor., « Active Appearance Models »
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 23, NO. 6, JUNE .,(2001).
- [Tomkins,1962] Tomkins.S.S., « Affect Imagery Consciousness », Volume I, The Positive Affect. Springer, London : Tavistoc, 1962.
- [Tong et al, 2007] Tong. Y. et al., « Facial Action Unit Recognition by Exploiting Their Dynamic and Semantic Relationships », IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 29, NO. 10, 'OCTOBER 2007)
- [Tsalakanidou et al,2010] Tsalakanidou. F., Malassiotis. S., « Real-time 2D+3D facial action and expression Recognition », Pattern Recogn. 43 (5) .,1763–1775., (2010).
- [Valstar et Pantic,2007] Valstar., M.F. et Pantic., M., « Combined Support Vector Machines and Hidden Markov Models for Modeling Facial Action Temporal Dynamics » . Springer-Verlag Berlin Heidelberg 2007. pp. 118–127., (2007).
- [Valstar et al,2010] Valstar.M.F., Pantic. M., « Induced disgust,happiness and surprise: an addition to the MMI facial expression data base », in:Proceedings of the International Conference on Language Resources and Evaluation,Workshop on EMOTION, ,pp.65–70.,(2010)
- [Valstar et al 2011] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer., « The first facial expression recognition and analysis challenge, » in Automatic Face & Gesture Recognition and Work- shops (FG 2011), 2011 IEEE International Conference on, pp. 921–926., (2011).
- [Viola et Jones,2004] Viola.P., Jones .M.J., « Robust Real Time Face Detection ». International Journal of Cumputer Vision., 57(2): p.137-154.,(2004).
- [Visin et al,2015] Visin.F., Kastner.k.,Courville.A., Bengio.Y, et al., « A Recurrent Neural Network for Object Segmentation », arXiv preprint arXiv:1511.07053.,(2015).
- [Walecki et al,2017] Walecki.R., Rudovic. O., Pavlovic. V., Schuller. B., Pantic. M., « Deep structured learning for facial action unit intensity

- estimation » ., In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26; pp. 3405–3414. ,(July 2017)
- [Welch,2003] Welch.L., «Hidden Markov Models and the Baum-Welch Algorithm », IEEE Information Theory Society Newsletter, 53(4)., (2003).
- [Wojdel et al, 2003] Wojdel, A. Rothkrantz M., and. Wojdel J. C., « Fuzzy-logical implementation of cooccurrence rules for combining AUs, »Proc. CGIM 2003., (2003).
- [Xiaoyi et al,2006] Xiaoyi.F., Baohua.L., Zhen.L et Jiling.Z., « Automatic facial expression recognition with aam-based feature extraction and svm classifier ».Advances in Artificial Intelligence, (2006).
- [Yang et al,2004] Yang.J., Zhang. D., Frangi. A.F., Yang.J., « Two-dimensional PCA: a new approach to appearance-based face representation and recognition ». IEEE Trans. Pattern Anal. Mach. Intell. 26(1), 131–137 (2004).
- [Yoneyama et al,1997] Yoneyama.M ., Iwano.Y., Y., Ohtake.A., K. Shirai., « Facial expressions recognition using discrete hopfield neural networks, »Journal of IEEE, pp. 117-120., (1997).
- [Yongmian,2005] Yongmian Zhang ., « Qiang Ji Active and dynamic information fusion for facial expression understanding from image sequences ». IEEE Transactions on Pattern Analysis and Machine Intelligence., Volume: 27, Issue: 5., (May 2005)
- [Zeng,2009] Zeng.Z., Pantic.M ., Roisman.I, and T.S. Huang. « A survey of affect recognition methods: Audio, visual, and spontaneous expressions ». IEEE transactions on pattern analysis and machine intelligence, 31(1) :39–58, (2009).

WEBOGRAPHIE

WEBOGRAPHIE

- | | | |
|-----------------|---|-----------------------------|
| [CNV,2013] | http://www.la-communication-non-verbale.com/2013/03/facial-action-coding-system-6734.html | Consulté le :
06/04/2016 |
| [Dico Philo] | http://la-philosophie.com/emotion-philosophie | Consulté le :
01/02/2015 |
| [james, 1884] | https://www.brainpickings.org/2016/01/11/what-is-an-emotion-william-james/ | Consulté le :
07/06/2017 |
| [Machines,2009] | Machines. S., « Faceapi, »URL : http ://www.seeingmachines.com/product/faceapi. (2009). | Consulté le :
06/05/2017 |