

*A mon D efunt P ere,
A ma m ere, ma S œur et mes deux fr eres
A mon Mari « Salah »*

Remerciements

Tout d'abord, je remercie Pr. Alice Caplier (Co_encadrante) de m'avoir proposé ce sujet de thèse, je la remercie également pour la pertinence de ses conseils, sa rigueur scientifique et son soutien sans faille. Elle a su me guider sans être trop directive tout le long de ces quatre années.

Je remercie également mon Directeur de thèse Dr. MK. Kholadi pour son soutien et ses encouragements.

Si la rédaction d'une thèse n'est pas une sinécure, il en va de même pour sa lecture approfondie. Je remercie donc chaudement les membres du jury : Professeur Mohamed Benmohamed (Président), Professeur Hamid Seridi, Dr. Nadir Farah et Dr. Salim Chikhi.

Enfin, et surtout, je tiens à remercier Mr. Mihoubi Salah (Mon époux) pour m'avoir supporté (dans tous les sens du terme) durant la réalisation de cette thèse. Sans son aide je ne serais jamais arrivée la.

Abstract

Facial expressions recognition is a basic task in the human machine communication. During this thesis, we developed a facial expression classification system based mainly on transient features. In order to carry out this system, we considered two previous works developed in the laboratory. These two works relate to the detection of facial permanent features which are eyes, eyebrows and lips. This automatic detection allowed the localization of characteristic points of these features which are the corners and the limits of each feature. From these characteristic points, we detected facial regions where transient features can appear. A study on the presence or the absence of this type of features as well as other characteristics of this type of features gave a primary classification of facial expressions without providing much effort. This study is completed by a study on permanent features in order to refine obtained results.

This categorical classification proved its limit in the recognition of other facial expressions, this leads to develop another system which propose a dimensional classification of facial expressions in terms of positivity / negativity or pleasure / not pleasure.

An ideal facial expression system allows quantifying facial expressions. Another aim of this thesis is the development of a quantification system which allows estimating intensity of known as well as unknown facial expression.

The three suggested systems are based in addition to the transient features, on different deformations of permanent features. To measure the deformation of these features, five characteristic distances are computed from characteristic points located from detected permanent features.

The Transferable Belief Model is used in the three systems in order to fuse all available data coming from different sensors. This is why we have used this specified model. Another reason of the choice of this model is its ability to model the doubt between the different considered classes.

Another objective of this thesis is the recognition of facial expressions based on video information. When considering facial expressions databases, several actors present different expressions with various intensities in video sequences. Each video count more than 200 frames. Several data is extracted from each frame this leads to an important mass of data. To analyze and explore these data we use a data mining technique in order to extract a new temporal knowledge. The developed dynamic recognition system takes into the count the temporal deformations of face features. It induces temporal rules which describe facial expressions in a dynamic context. These rules can be added to existing ones proposed by M-PEG4 in a static context.

Keywords Transient features, permanent features, facial expressions, catégorial classification, dimensionnel classification, Transferable Belief Model, data mining.

Laboratory Misc
Nouvelle ville Constantine.

Table des Matières

Table des matières	1
Liste des figures	5
Liste des tableaux	7
Liste des tableaux	8
Introduction	
1 Etude préliminaire et Extraction de Données	
1.1 Introduction	12
1.2 Expression faciale et émotion	12
1.2.1 L'émotion	12
1.2.2 L'expression faciale	12
1.3 Représentation de l'expression faciale	13
1.3.1 Système de codification des actions faciales FACS	13
1.3.2 MPEG_4	14
1.3.3 Candide	15
1.4 Détection du visage	16
1.4.1 Le détecteur de visage de Rowley	17
1.5 Extraction des composants du visage	18
1.5.1 Détection des yeux et des sourcils	21
1.5.1.1 Détection de l'iris	21
1.5.1.2 Modèles paramétriques pour les yeux et les sourcils	22
1.5.1.3 Segmentation des yeux	22
1.5.1.4 Segmentation des sourcils	23
1.5.2 Détection des lèvres	24
1.5.2.1 Analyse comparative de la couleur des lèvres et de la peau	24
1.5.2.2 L'algorithme du jumping Snake	25
1.5.2.3 Détection des points caractéristiques	26
1.5.2.4 Déformation du modèle	29
1.6 Conclusion	29
2 Classification catégorielle des expressions faciales	32
2.1 Introduction	33
2.2 Etat de l'art	33
2.2.1 Approche basée modèles	33
2.2.2 Approche géométrique	34
2.2.3 Approches basées règles	36
2.2.4 Approches basées Réseaux de neurones	37
2.2.5 Approches basées flux optique	38
	40

2.3	Notre contribution	44
2.4	Méthode proposée	47
2.4.1	Segmentation	48
2.4.2	Extraction de données	50
2.4.3	Analyse	52
2.4.4	Classification	55
2.4.4.1	Principe de la théorie de l'évidence	55
2.4.4.2	Définition des fonctions de masse	56
2.4.4.3	Combinaison des masses d'évidence	57
2.4.4.4	Processus de décision	58
2.4.4.5	Application de la TBM dans le contexte de classification catégorielle	58
2.4.5	Post traitement	60
2.5	Résultats obtenus	63
2.6	Comparaison avec autre système	64
2.7	Conclusion	65
3	Classification Dimensionnelle des Expressions Faciales	
3.1	Introduction	66
3.2	Etat de l'art	68
3.3	Notre contribution	70
3.4	Méthode proposée	71
3.4.1	Descriptions des différentes sources	72
3.4.1.1	Source géométrique : Distances faciales	72
3.4.1.1.1	Description de l'expression neutre	72
3.4.1.1.2	Description des expressions positives ou négatives	72
3.4.1.2	Source géométrique : Angle nasolabial, Phase d'apprentissage	74
3.4.1.3	Source probabiliste : Présence ou absence des traits transitoires , phase d'apprentissage	74
3.5	Théorie de l'évidence dans le contexte de classification dimensionnelle des expressions faciales	77
3.5.1	Calcul des masses d'évidence locales	78
3.5.1.1	Source géométrique : distances faciales	78
3.5.1.2	Source géométrique : Angle nasolabial	79
3.5.1.3	Source probabiliste : présence ou absence des traits transitoires	80
3.5.2	Combinaison des trois sources : Approche globale	81
3.6	Résultats expérimentaux	82
3.7	Conclusion	83
4	Estimation de l'Intensité des Expressions Faciales	
4.1	Introduction	84
4.2	Etat de l'art	85
4.2.1	Approches locales	85

4.2.2	Approches globales	86
4.3	Notre contribution	88
4.4	Méthode proposée	92
4.5	Application de la théorie de l'évidence dans le contexte d'estimation de l'intensité des expressions faciales	93
4.5.1	Définition des états symboliques	93
4.5.2	Processus de modélisation	94
4.5.3	Définition des seuils	94
4.5.4	Définitions des intensités des expressions	95
4.5.4.1	Expression de la joie	95
4.5.4.2	Expression de la surprise	96
4.5.4.3	Expression du dégoût	96
4.5.4.4	Expression de la colère	97
4.5.4.5	Expression de la tristesse	97
4.5.4.6	Expression de la peur	98
4.5.5	Règles logiques entre les états symboliques et les expressions faciales	99
4.5.6	Fusion de données	99
4.5.7	Décision	100
4.6	Post traitement en cas de conflit	100
4.7	Résultats obtenus	102
4.7.1	Résultats sur la base [HAM base]	102
4.7.2	Résultats sur la base [EE base]	103
4.8	Estimation de l'intensité d'une expression inconnue	106
4.8.1	Application de la TBM dans le contexte de l'estimation d'une intensité d'expression inconnue	107
4.8.2	Résultats obtenus sur la base Dafex	108
4.9	Conclusion	112

5 Classification des Expressions Faciales sur la base d'Informations Vidéo

5.1	Introduction	113
5.2	Suivi dans les séquences vidéos	114
5.2.1	L'algorithme de Lucas-Kanade	115
5.2.2	Résultats de l'algorithme de Lucas-Kanade	118
5.3	Introduction au Data Mining	118
5.3.1	Définition du Data Mining	118
5.3.2	Origine et émergence du concept de datamining	118
5.3.3	Raisons du développement	119
5.3.4	Exemples d'applications	119
5.3.5	Principe du Data Mining	120
5.3.6	Algorithmes	121
5.3.6.1	Méthodes supervisées	121

5.3.6.2	Méthodes non supervisées	122
5.3.6.3	Méthodes semi supervisées	122
5.3.7	Principales tâches de l'apprentissage	123
5.3.8	Modèles du data mining	123
5.3.8.1	Modèles descriptifs	124
5.3.8.2	Modèles prédictifs	124
5.4	Notre contribution	125
5.4.1	Algèbre d'intervalle d'Allen	126
5.5	La démarche dataminig dans le contexte de classification des expressions faciales	127
5.5.1	Acquisition des données d'entrées	127
5.5.2	Sélection des données entrées	127
5.5.3	Prétraitement	128
5.5.3.1	Algorithme de transformation	128
5.5.4	Extraction des règles	129
5.5.4.1	Principe de l'algorithme « Une règle basée classification »	130
5.5.4.2	Procédure d'apprentissage	130
5.5.5	Interprétation et évaluation des résultats	134
5.5.6	Déploiement	137
5.5.6.1	Test de la base [HAM base]	137
5.5.6.2	Test de la base Cohn et Kanade	138
5.6	Conclusion	139
	Conclusion et perspectives	140
	Références bibliographiques	142
	Publications et Communications	151
	Résumé	152

Liste des Figures

- Figure 1.1. Processus d'analyse des Expressions faciales
- Figure 1.2 Générateurs de l'expression faciale et de l'émotion ;
- Figure 1.3. Les six expressions faciales universelles et le neutre.
- Figure 1.4. Modèle du visage MPEG-4 – Définition des points facials
- Figure 1.5. Version 3 du modèle Candide
- Figure 1.6 Principe de fonctionnement d'un perceptron multicouche dans le détecteur mis au point par Rowley et al.[ROW98] de manière isolée. Les « Receptive Fields = Champs Récepteurs» correspondent à ce que nous appelons rétines.
- Figure 1.7 Prétraitement d'une fenêtre avant classification par un perceptron. La répartition de l'intensité des pixels dans les fenêtres étudiées (b) est représentée de manière linéaire (c), en ne considérant que les pixels de la zone d'intérêt, en blanc sur (a). A partir de cette représentation, on effectue une égalisation d'histogramme sur les images originales (d), puis on normalise afin de rehausser les contrastes (e).
- Figure 1.8 Quelques détections par l'algorithme de Rowley. Les détections positives selon Rowley sont encadrées en vert.
- Figure 1.9 Modèle pour l'œil et le sourcil et points caractéristiques Pi
- Figure 1.10. à gauche : détection des coins C1 et C2 des yeux; à droite : initialisation du modèle de l'œil.
- Figure 1.11. Modèle choisi pour la bouche
- Figure 1.12. Initialisation du jumping snake. La position du germe S1 est calculée à partir de S0 (initialisé à l'intérieur du cadre noir) et des points (gros ronds) associés aux flux moyens les plus élevés.
- Figure 1.13. depuis le germe S0, le snake s'accroît par ajout de points à droite et à gauche suite à la maximisation de R_{top} à travers chaque nouveau segment.
- Figure 1.14. le maximum de ϕ_{total}^k donne la position du coin sur Lmini. Les courbes en pointillés sont les cubiques associées aux différents points k testés le long de Lmini.
- Figure 1.15 : Quelques exemples de contours des yeux, sourcils et lèvres détectés.
- Figure 2.1. Un exemple de rectangles entourant les régions faciales d'intérêts [Yac96].
- Figure 2.2. Suivi de point caractéristiques [Coh98].
- Figure 2.3. Exemple de doute entre Surprise et Peur.
- Figure 2.4. Démarche proposée.
- Figure 2.5. Les points caractéristiques du visage et les distances biométriques.
- Figure 2.6. Détection des traits permanents et localisation des régions d'intérêts.
- Figure 2.7. (a):Visage à l'état neutre sans la présence de traits transitoires ;(b): Visage avec expression (de joie) avec la présence de traits transitoire sur la zone nasolabiale.
- Figure 2.8. Détection et calcul de l'angle nasolabial.
- Figure 2.9. Angle nasolabial formé de gauche à droite: Colère (72.6°), Dégout (71.2°) et Joie (43.4°).(Eebase, H_Caplier databases)
- Figure 2.10. Variation de l'angle nasolabial formé avec les Cinq expressions faciales.
- Figure 2.11. Fusion d'information
- Figure 2.12. Exemple de processus de décision de plausibilité et croyances
- Figure 2.13. Modèle proposé pour la présence et absence des TTS.
- Figure 2.14. Modèle proposé pour l'angle nasolabial calculé.

Figure 2.15. Différentes descriptions de la colère de gauche à droite : (1) Les yeux sont à peine ouverts et la bouche est ouverte ; (2,3) les yeux sont normalement ouverts et les lèvres sont serrées ; (4) les yeux sont grand ouverts et la bouche également.

Figure 3.1. Dimension bipolaire : valence et activation proposé par Russell [RUS94].

Figure 3.2. Table des unités d'actions de la partie inférieure du visage [EKM].

Figure 3.3. (a): Visage sans traits transitoires (b): visage avec traits transitoires sur les régions nasolabiales.

Figure 3.4. TTs formés avec les expressions négatives

Figure 3.5. Modèle de la masse d'évidence basique pour chaque distance D_i

Figure 3.6. Modèle pour l'angle nasolabial

Figure 3.7. Modèle proposé concernant la présence ou l'absence des TTs.

Figure 4.1 les différents degrés d'intensité d'une unité d'action

Figure 4.2 Modèle d'intensité choisie

Figure 4.3. Les points caractéristiques du visage et les distances biométriques.

Figure 4.4 Unités d'actions de la partie inférieure du visage (Ekman)

Figure 4.5 Unités d'actions de la partie supérieure du visage (Ekman)

Figure 4.6 Modèle proposé pour l'estimation de l'intensité

Figure 4.7 Evolution des distances dans le cas de la joie

Figure 4.8 Evolution des distances dans le cas de la surprise

Figure 4.9. Evolution des distances dans le cas de Dégout

Figure 4.10 Evolution des distances dans le cas de la Colère

Figure 4.11 Evolution des distances dans le cas de la Tristesse.

Figure 4.12 Evolution des distances dans le cas de la Peur

Figure 4.13 Exemple d'expression de surprise avec conflit

Figure 4.14. Exemples des intensités des six expressions faciales avec leurs masses d'évidence associées

Figure 4.15. Quelques exemples des masses d'évidence associées aux différentes intensités estimées pour des images d'autres bases d'image.

Figure 4.16. Images d'un acteur montrant six expressions Colère, Dégout, Peur, Joie, Tristesse et Surprise respectivement avec les trois intensités max, moy et min respectivement pour chaque expression

Figure 4.17 expression de Surprise avec deux distances changées.

Figure 4.18 deux descriptions différentes de la colère avec intensité max.

Figure 4.19 deux descriptions différentes de la peur avec intensité max.

Figure 4.20 Les images (b,d) intensité moyenne deux distances changées D_1 , D_2 , Les images (a,c) intensité max trois distances changées D_1 , D_2 , D_4 , et D_3 .

Figure 4.21 distances Changées avec intensités différentes pour la même expression (a,c) int. moy(D_1, D_2, D_5); (b,d) int; max D_1, D_2, D_4

Figure 5.1 Un voisinage dans l'image I_t peut être retrouvé dans l'image I_{t+1} par une translation de vecteur d

Figure. 5.2 : Suivi d'un point caractéristique par l'algorithme de Lucas Kanade

Figure 5.3 Les relations de base entre les intervalles temporelles [ALL83].

Figure 5.4. Les points caractéristiques du visage et les distances biométriques.

Liste des Tableaux

- Table 2.1. Comparaison des différentes méthodes selon les différentes approches
- Table 2.2. Présence or absence de TTS sur chaque région d'intérêt et pour chaque expression.
- Table 2.3. Nouvelle description des six expressions faciales.
- Table 2.4. Différences potentielles entre les six expressions universelles.
- Table 2.5. Classification des expressions faciales basée sur les traits transitoires avec un post traitement sur la base des traits permanents.
- Table 2.6. Comparaison des Approches
- Table 3.1. Résumé des méthodes de classification en expressions positives et négatives par rapport à la méthode proposée.
- Table 3.2. Relations entre l'activation des Aus et l'évolution des distances D3 et D5.
- Table 3.3. Classification d'expressions positives et négatives basée sur les rides nasolabiales.
- Table 3.4. Présence ou absence des TTS sur chaque région faciale et pour chaque expression
- Table 3.5. Règles logiques des états associés aux distances caractéristiques pour chaque classe d'expression.
- Table 3.6. Taux de classification obtenus sur différentes bases d'images.
- Table 4.1 Comparaison des méthodes de l'état de l'art.
- Table 4.2 Descriptions des six expressions faciales et les déformations pertinentes du visage.
- Table 4.3 Etats des variables associées aux distances considérées pour chaque intensité de la joie
- Table 4.4. Etats des variables associées aux distances considérées pour chaque intensité de la surprise
- Table 4.5 Etats des variables associées aux distances considérées pour chaque intensité du dégoût
- Table 4.6 Etats des variables associées aux distances Considérées pour chaque intensité de Colère
- Table 4.7 Etats des variables associées aux distances considérées pour chaque intensité de la tristesse
- Table 4.8. Etats des variables associées aux distances considérées pour chaque intensité de la Peur
- Table 4.9 Règles logiques pour D1 et D3 (la joie)
- Table 4.10 Exemple de combinaison des deux distances
- Table 4.11. Exemple de conflit (erreur)
- Table 4.12 classification des intensités de la base [HAM base] avant post traitement
- Table 4.13 classification des intensités de la base [HAM base] après post traitement
- Table 4.14 classification des intensités de la base [EE base] après post traitement
- Table 4.15 Nouvelles expressions reconnues suite à la quantification des six expressions universelles et les réactions correspondantes (Rules Expert system)
- Table 4.16. Les différents états pris par une distance D_i et les expressions correspondants avec intensités
- Table 4.17 Classification des intensités basé sur la TBM
- Table 4.18. Taux de classification de l'intensité pour les bases Dafex et Hammal_caplier
- Table 5.1 Distances changées pour tous les acteurs de la base Dafex et quelques exemples de la base cohn et kanade pour quatre expressions faciales.
- Table 5.2 Règles déduites
- Table 5.3 Règles les plus intéressantes
- Table 5.4 Descriptions statique et dynamique des expressions faciales
- Table 5.5 Les taux de classification de la base Hammal-caplier.
- Table 5.6 Taux de classification de la base Cohn et Kanade.

Introduction

L'interaction homme-machine a longtemps confiné ses recherches au développement de techniques fondées sur l'usage du triplet écran-clavier-souris. Aujourd'hui, elle s'oriente vers de nouveaux paradigmes : l'utilisateur doit pouvoir évoluer sans entraves dans son milieu naturel ; les doigts, la main, le visage ou les objets familiers sont envisagés comme autant de dispositifs d'entrée/sortie ; la frontière entre les mondes électronique et physique tend à s'estomper. Ces nouvelles formes d'interaction nécessitent le plus souvent la capture du comportement observable de l'utilisateur et de son environnement. Elles s'appuient pour cela sur des techniques de perception artificielle, et notamment de vision par ordinateur. L'interaction Homme-Machine (IHM) est une discipline qui évolue rapidement. Les générations futures d'environnement Homme-Machine deviendront multimodales en intégrant de nouvelles informations, provenant de la prise en compte du comportement dynamique, de la parole et/ou des expressions faciales, de manière à rendre l'utilisation des machines la plus intuitive et naturelle possible.

Le visage étant la partie la plus expressive et communicative d'un être humain, il représente un centre d'intérêt majeur dans les recherches actuelles concernant l'amélioration de l'IHM pour l'établissement d'un dialogue entre les deux entités.

Une expression faciale est une manifestation visible sur un visage de l'état d'esprit (émotion, réflexion), de l'activité cognitive, de l'activité physiologique (fatigue, douleur), de la personnalité et de la psychopathologie d'une personne. Des travaux de recherche en psychologie ont démontré que les expressions faciales jouent un rôle prépondérant dans la coordination de la conversation humaine, et ont un impact plus important sur l'auditeur que le contenu textuel du message exprimé. Mehrabian [MEH96] remarque que la contribution du contenu textuel d'un message verbal en « face à face » à son impact global se limite à 7% alors que les signaux conversationnels (accentuation de mots, ponctuation, marqueurs d'une question, indicateurs d'une recherche de mots, etc.) et l'expression faciale du locuteur contribuent respectivement à 38 % et 55 % de l'impact global du message exprimé. Par conséquent, l'expression faciale peut être considérée comme une modalité essentielle de la communication humaine.

L'essentiel de l'information d'une expression faciale est contenue dans la déformation des traits permanents principaux du visage, caractérisée par un changement, perceptible visuellement. Cette déformation est due à l'activation volontaire ou non de l'un ou de plusieurs des 44 muscles composant le visage. Il émet en permanence des signes dont le décodage, non seulement renseigne sur l'état émotionnel de la personne, mais aussi éclaire sur ce qui est dit.

Aujourd'hui, l'analyse assistée par ordinateur du visage et de ses expressions est un domaine émergent. Les applications sont nombreuses.

En **Interaction Homme-Machine**, on cherche à avoir une idée de l'état émotionnel de l'utilisateur pour la conception d'interfaces plus ergonomiques et présentant un meilleur retour d'informations (*feedback*), la mesure de la direction du regard de l'utilisateur pourrait être un moyen efficace d'effectuer certaines tâches dans une interface graphique (comme la sélection d'une fenêtre ou d'une zone de saisie). D'une manière générale, les expressions du visage prennent une place importante dans le processus de communication humain et forment à elles seules un langage co-verbal. Les expressions du visage (au sens de mouvements musculaires) accompagnent le langage parlé, aussi bien en termes de mouvements physiques nécessaires à la parole (mouvements des lèvres), qu'en termes d'indicateur émotionnel accompagnant le langage parlé. Elles expriment une part non-négligeable du sens dans une communication orale. En particulier dans la Langue des Signes où les expressions du visage font partie intégrante du langage gestuel. Bien que beaucoup de travaux dans ce domaine tentent de classer les expressions de l'utilisateur en émotions universelles, certains se focalisent sur des composantes particulières du visage qui servent à l'interaction (suivi du mouvement des yeux pour la sélection par exemple). On ne cherche pas ici à avoir une description fine des expressions et des mouvements musculaires sous-jacents, mais plutôt à avoir une idée du mouvement de certaines composantes ou à avoir une idée d'un état émotionnel.

Dans le domaine de la **linguistique**, des **sciences comportementales**, de la **psychologie** ou de la **Langue des Signes**, on s'intéresse à une description détaillée des expressions en vue d'en fournir un sens. La classification en émotions universelles n'est généralement pas suffisante. Dans certains cas, les niveaux de détail d'analyse sont très fins, puisque certains cherchent à détecter les expressions spontanées des expressions forcées. Les deux types d'expression sont générés par deux zones distinctes du cerveau [BRU83] (c'est la base d'un détecteur de mensonges basé sur les expressions).

La **compression de données** s'intéresse à la description des expressions du visage. Le principe est de coder les expressions des visages présents dans une séquence vidéo et donc de réduire la quantité d'informations à transmettre. Le mouvement d'un personnage peut alors être reconstruit à partir d'un ensemble réduit de paramètres. Encore une fois, on cherche à ce que la reconstruction à partir de ce codage soit la plus réaliste possible d'un point de vue visuel.

En **animation**, on cherche à animer des personnages virtuels qui doivent paraître le plus réaliste possible. On ajoute alors aux mouvements des muscles faciaux nécessaires à la parole, un ensemble de mouvements faciaux qui traduisent un état émotionnel. L'analyse de la formation des expressions est donc nécessaire aussi bien pour la description que pour la synthèse. On s'attarde surtout à reconstruire une expression qui semble réaliste d'un point de vue visuel et qui est porteuse d'un sens (d'une émotion), en accentuant éventuellement les expressions (pour des personnages caricaturés par exemple).

Plusieurs autres domaines sont également concernés par l'analyse des expressions faciales notamment dans l'enseignement à distance (transmission au professeur de l'état des étudiants sous forme d'information de haut niveau), logiciels d'apprentissage, Système de mesure objective ou l'interprétation de l'observateur n'entre pas en ligne de compte, dispositif dans les voitures qui avertit le conducteur en cas de perte de vigilance, dispositif dans les avions (information supplémentaire pour la boîte noire), la reconnaissance de visages invariante à l'expression, jeux et shopping on-line, évaluation de la douleur chez les malades et les enfants, l'évaluation de nerfs faciales en médecine, la compréhension de l'image et dans la robotique.

Il ne s'agit plus pour l'homme de s'adapter à la machine, mais bien à la machine d'adopter les modes de communication humains. Les premières avancées dans ce sens, dans les années 90, furent les logiciels grand public de dictée automatique ou de reconnaissance de caractères. Les puissances de calcul actuelles permettent d'envisager des analyses encore plus complexes portant désormais sur le corps humain lui-même. Etant donnés les efforts de recherche et les progrès récents dans ce domaine, on peut d'ores et déjà s'attendre à ce que les ordinateurs obéissent bientôt «au doigt et à l'œil», au sens propre comme au figuré.

La zone du corps la plus chargée de sens est sans conteste le visage. En particulier, nous verrons, dans le premier chapitre de ce mémoire, qu'il est possible d'extraire une très grande quantité d'information depuis cette zone. Notre objectif est de faire sortir depuis cette zone toute information utile à la compréhension comportementale. Nous montrerons que plusieurs connaissances peuvent être découvertes depuis les informations extraites et vont de

la reconnaissance de l'expression faciale, passant par la détection de l'humeur d'un sujet (bonne ou mauvaise) à l'estimation de l'intensité de l'expression reconnue.

Dans le second chapitre, nous détaillerons la méthode de classification catégorielle des expressions faciales. Cette méthode va se baser principalement sur l'étude de la présence des traits transitoires sur le visage. Nous montrerons que ce type de trait peut arriver à donner des performances aussi importantes que celles d'une étude basée principalement sur les déformations des traits permanents.

Dans le chapitre trois, nous détaillerons la méthode de classification dimensionnelle des expressions faciales, la dimension qui nous intéresse dans cette étude est la dimension de valence c'est-à-dire la dimension qui nous permet de savoir si une personne est de bonne ou mauvaise humeur. Cette méthode portera sur l'étude des traits permanents aussi bien que les traits transitoires. La théorie de l'évidence est utilisée encore une fois afin de fusionner l'ensemble des informations issues des différentes sources car cette méthode est bien adaptée à de telles circonstances.

Dans le quatrième chapitre nous présenterons une autre méthode qui procède à l'estimation de l'intensité d'une expression connue ou inconnue afin de mettre en œuvre un système idéal d'analyse d'expressions faciales.

Dans chaque chapitre nous présenterons des états de l'art afin de motiver chacune de nos contributions. Nous résumerons les forces et les limites des méthodes présentées dans la littérature. Partant de ces constatations, nous esquisserons les grandes lignes de nos algorithmes.

Dans le dernier chapitre, nous aborderons une nouvelle technique utilisée pour la première fois dans l'analyse des expressions faciales. Cette technique procédera à la classification des expressions sur la base d'informations vidéo. Vu l'importance de la masse des données extraites depuis une séquence vidéo relative à la production d'une expression faciale, nous avons fait appel à une technique de Data Mining. Cette méthode procède à une extraction de connaissances depuis les données vidéo. Ces nouvelles connaissances seront sous forme de règles qui vont permettre de décrire une expression faciale dans un contexte dynamique.

Enfin, nous résumerons les principales contributions de ce travail de thèse et nous conclurons quant aux améliorations envisageables et aux perspectives pour la poursuite du projet.

Chapitre 1

Etude Préliminaire et Extraction Des Données de Base

1.1. Introduction

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

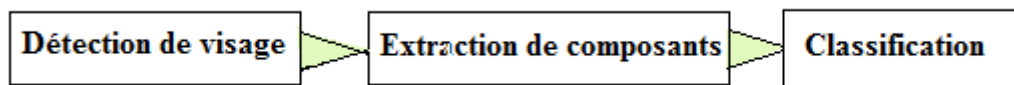


Figure 1.1. Processus d'analyse des Expressions faciales

Avant de présenter les différents traitements considérés dans chaque phase, nous allons définir la différence entre expression faciale et émotion ensuite nous donnerons les différentes représentations des expressions faciales.

1.2. Expression faciale et émotion

1.2.1 L'Emotion

Expressions et émotions sont très liées et parfois confondues, l'émotion est un des générateurs des expressions faciales. L'émotion se traduit via de nombreux canaux comme la position du corps, la voix et les expressions faciales. Une émotion implique généralement une expression faciale correspondante (dont l'intensité peut être plus ou moins contrôlée selon les individus), mais l'inverse n'est pas vrai : il est possible de mimer une expression représentant une émotion sans pour autant ressentir cette émotion. Alors que les expressions dépendent des individus et des cultures, on distingue généralement un nombre limité d'émotions universellement reconnues.

1.2.2. L'Expression Faciale

L'expression faciale est une mimique faciale chargée de sens. Le sens peut être l'expression d'une émotion, un indice sémantique ou une intonation dans la Langue des Signes. L'interprétation d'un ensemble de mouvements musculaires en expression est dépendante du contexte d'application. Dans le cas d'une application en interaction Homme-Machine où l'on désire connaître une indication sur l'état émotionnel d'un individu, on cherchera à classifier les mesures en terme d'émotions.

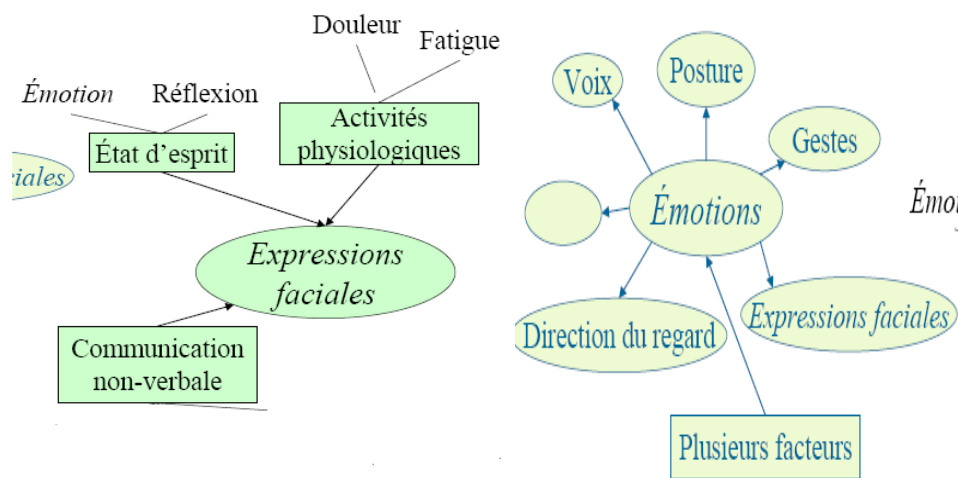


Figure 1.2 Générateurs de l'expression faciale et de l'émotion ;

1.3. Représentation de l'Expression Faciale

Le visage est une zone importante du corps humain composée de plusieurs muscles. L'électromyographie (EMG) est une technique permettant de mesurer l'activité musculaire au cours du temps. Les muscles de la zone supérieure du visage n'ont que peu d'influence sur les muscles de la zone inférieure et vice-versa. Il est donc possible de découper l'analyse en deux zones.

Les travaux fondamentaux dans le domaine de l'analyse des expressions du visage sont dus principalement à Charles Darwin, Guillaume Duchenne De Boulogne au XIXème siècle et plus récemment Paul Ekman.

Charles Darwin, est le premier à traiter de l'universalité des expressions du visage et à proposer une théorie évolutionniste sur la formation des expressions. L'argument principal est que les expressions des enfants et des nouveaux nés existent aussi chez les adultes. D'après lui, l'expression des émotions est un processus nécessaire à la survie. Ainsi, les expressions

non-verbales sont aussi importantes que les interactions verbales dans le processus de communication humain.

Au XIX^{ème} siècle, Guillaume Duchenne de Boulogne est le premier à localiser individuellement les différents muscles faciaux par activation électrique. Il est un des premiers à livrer à la communauté scientifique un ensemble de photographies montrant l'activation des différents muscles faciaux.

Paul Ekman, psychologue, s'intéresse à partir du milieu des années 1960 aux expressions et émotions humaines. Il établit qu'il existe un nombre limité d'expressions reconnues par tous, indépendamment de la culture. Il met donc en évidence l'universalité de certaines émotions innées qui correspondent aux sept émotions suivantes : la **neutralité**, la **joie**, la **tristesse**, la **surprise**, la **peur**, la **colère** et le **dégoût**.



Figure 1.3. Les six expressions faciales universelles et le neutre.

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

1.3.1. Système de Codification des Actions Faciales FACS

En 1978, Ekman et Friesen présentent un système de codification *manuelle* des expressions du visage [EKM78]. Leurs travaux d'observation leur permettent de décomposer tous les mouvements *visibles* du visage en termes de 46 unités d'*Actions* qui décrivent les mouvements *élémentaires* des muscles. N'importe quelle mimique observée peut donc être représentée sous la forme d'une combinaison d'unités d'*Actions*. Ce système de codage est connu sous le nom de Facial Action Coding System (FACS). FACS s'est imposé depuis comme un outil puissant de description des mimiques du visage, utilisé par de nombreux psychologues.

Bien que FACS soit un système de description bénéficiant d'une grande maturité (environ vingt années de développement), il souffre cependant de quelques inconvénients :

Complexité : on estime qu'il faut 100 heures d'apprentissage pour en maîtriser les principaux concepts.

Difficulté de manipulation par une machine : FACS a d'abord été créé pour des psychologues, Certaines mesures restent floues et difficilement évaluables par une machine.

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

1.3.2. MPEG4

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

Les *Facial Animation Parameter Units* (FAPU) permettent de définir des mouvements élémentaires du visage ayant un aspect naturel. En effet, il est difficile de définir les mouvements élémentaires des muscles de manière absolue : le déplacement absolu des muscles d'une personne à l'autre change, mais leur déplacement relatifs à certaines mesures pertinentes sont constantes. C'est ce qui permet d'animer des visages de manière réaliste et peut permettre de donner des expressions humaines à des personnages non-humains. Comme exemples de FAPU, on peut citer *la largeur de la bouche, la distance de séparation entre la bouche et le nez, la distance de séparation entre les yeux et le nez*, etc. Par exemple, l'étirement du coin de la lèvre gauche (Facial Animation Parameter 6 stretch_1_cornerlip) est défini comme le déplacement vers la droite du coin de la lèvre gauche d'une distance égale à la longueur de la bouche. Les FAPUs sont donc des mesures qui permettent de décrire des mouvements élémentaires et donc des animations.

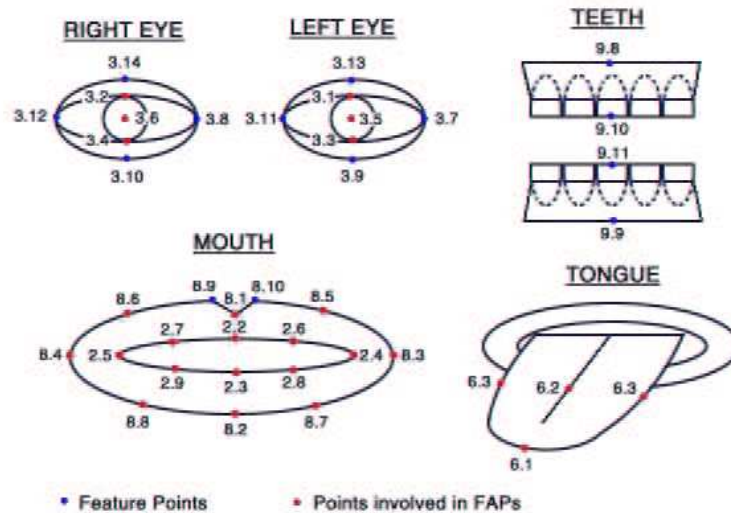


Figure 1.4. Modèle du visage MPEG-4 – Définition des points faciaux

1.3.3. Candide

Candide [AHL01] est un modèle du visage. Il est composé d'un modèle en fil de fer représentant un visage générique et d'un ensemble de paramètres :

a- **Paramètres de forme** (*Shape Units*) : ces paramètres permettent d'adapter le modèle générique à un individu particulier. Ils représentent les différences inter-individus et sont au nombre de 12 :

1. Hauteur de la tête,
2. Position verticale des sourcils,
3. Position verticale des yeux,
4. Largeur des yeux,
5. Hauteur des yeux,
6. Distance de séparation des yeux,
7. Profondeur des joues,
8. Profondeur du nez,
9. Position verticale du nez,
10. Degré de courbure du nez (s'il pointe vers le haut ou non),
11. Position verticale de la bouche,
12. Largeur de la bouche.

b- **Paramètres d'animation** (*Animation Units*) : ces paramètres représentent les différences intra-individus, *i.e.* les différentes actions faciales. Ils sont composés d'un sous-ensemble de

FACS et d'un sous-ensemble des FAPs de MPEG-4. Les FAPs sont définis par rapport à leur FAPU correspondant. Ces paramètres, qu'ils soient d'animation ou de forme, sont représentés sous forme d'une liste de points du modèle de fil de fer à mettre à jour. Candide permet de voir clairement la différence entre les AUs de FACS et les FAPs de MPEG-4 : les AUs de FACS sont exprimées de manière absolue, à la différence des FAPs qui sont exprimés par rapport à des mesures du visage (les FAPUs).

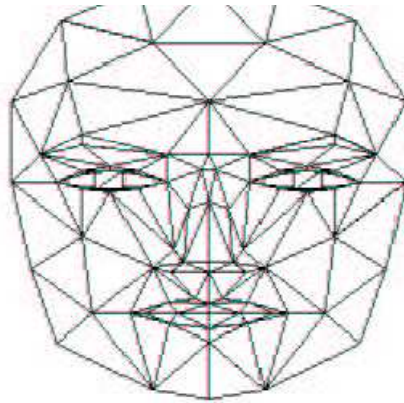


Figure 1.5. Version 3 du modèle Candide

Quelque soit la représentation utilisée dans un processus d'analyse des expressions faciale, le but est la reconnaissance de l'expression faciale étudiée.

Pour reconnaître une expression faciale, on doit suivre le schéma de la figure 1.1.

1.4. Détection du Visage

Le problème de l'extraction automatique du visage dans une image a été largement traité ces dernières années [EVE03]. Un état de l'art détaillé se trouve dans [MIN02].

Parmi les publications récentes de détecteurs comme [GAR04], on retrouve notamment des comparaisons par rapport aux performances de Viola & Jones, OpenCV, SNoW, et du détecteur de Rowley [FAS02]. Leurs performances sont semble-t-il bien supérieures à celles d'algorithmes plus anciens comme Eigenfaces et Fisherfaces. De plus, ils comptent parmi les premiers algorithmes à utiliser des bases de tests et d'apprentissage standardisées, ce qui permet d'avoir une base de résultats comparables.

1.4.1 Le Détecteur de Visages de Rowley

L'implémentation de ce détecteur diffère sur certains points des algorithmes publiés par l'auteur dans [ROW98]. Dans ce détecteur, la classification est basée sur une série de perceptrons multicouches. Un perceptron multicouches est un réseau de neurones artificiels, organisé en couches. Un neurone est un automate, dont la sortie est la comparaison d'un barycentre pondéré de plusieurs entrées à un seuil, via une fonction d'activation. Au sein d'une même couche, les entrées de chaque perceptron ne sont connectées qu'aux sorties de la couche précédente. La dernière couche n'est constituée que d'un unique neurone, dont la polarité de la sortie définit l'appartenance ou non à la classe étudiée. Ce positionnement en couches (ou en bancs) des neurones permet de définir une séparation non linéaire entre classes.

De manière plus précise que ce qui est publié dans [ROW98], la version disponible en opensource du détecteur de Rowley utilise 3 perceptrons multicouches successifs pour déterminer la présence ou l'absence d'un visage dans une fenêtre. En l'occurrence, la fonction d'activation utilisée est la fonction tangente hyperbolique (sigmoïde). La 1ère couche de chaque perceptron est reliée aux descripteurs (l'intensité des pixels) sous forme de « rétines ». C'est-à-dire que chaque neurone de la 1ère couche est une combinaison linéaire de l'intensité de tous les pixels contenus dans une zone rectangulaire (champ réceptif) prédéfinie de la fenêtre étudiée (Figure 1.6). Ces champs réceptifs sont dédoublés, voire triplés, au sein d'un même perceptron. A savoir que chaque champ réceptif est représenté plusieurs fois (2 à 3 fois selon le perceptron) au niveau de la première couche du réseau de neurones. Etant donné que ces rétines sont munies d'un apprentissage différent, elles augmentent en principe la robustesse du détecteur.

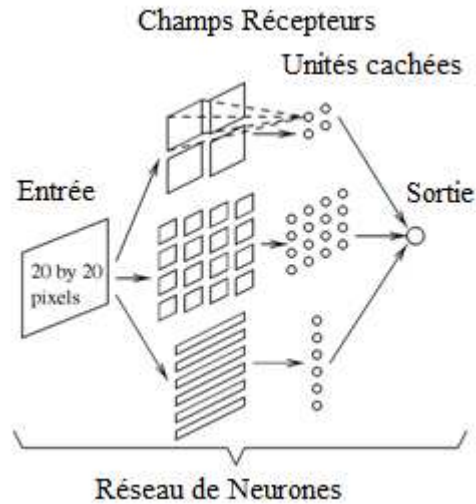


Figure 1.6 Principe de fonctionnement d'un perceptron multicouche dans le détecteur mis au point par Rowley et al.[ROW98] de manière isolée. Les « Receptive Fields = Champs Récepteurs » correspondent à ce que nous appelons rétines.

Avant tout passage dans un perceptron, chaque fenêtre est sujette à une égalisation d'histogrammes. C'est une technique qui permet en principe d'être moins sujet aux problèmes d'ombrage et de faibles contrastes (Figure 1.7).

Le 1er perceptron employé par ce détecteur possède une structure qui à notre connaissance n'est pas publiée. Il correspond à une fenêtre 30x30, étudiée tous les 10x10 pixels sur l'image d'origine redimensionnée. Sa fonction est de « trouver » des visages, c'est-à-dire qu'il n'est en principe pas extrêmement discriminant : en théorie il permet de trouver tous les visages, même si c'est au prix d'un nombre de faux-positifs élevé. En raison de la faible occurrence géographique de la classification par ce perceptron, il est sensé être peu sensible au bruit ainsi qu'aux décalages en x et en y. Avec ce perceptron, la taille totale de la fenêtre considérée est de 30x30 pixels. L'ensemble des rétines est constitué de 9 rétines de 10x10 pixels, 6 rétines de 30x5 pixels et 6 rétines de 5x30 pixels.

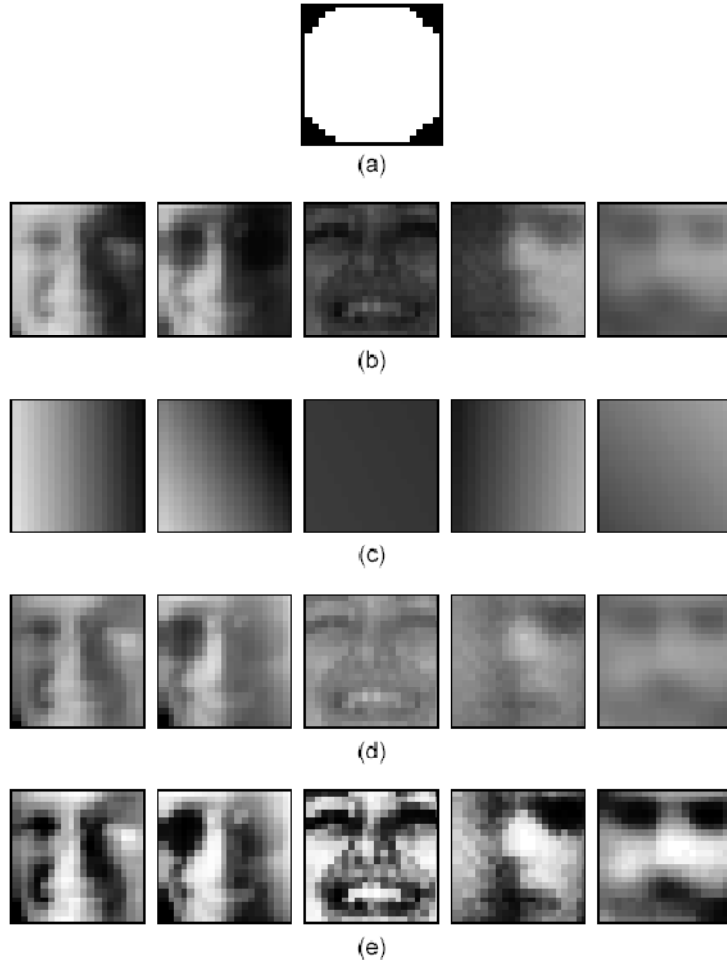


Figure 1.7 Prétraitement d'une fenêtre avant classification par un perceptron. La répartition de l'intensité des pixels dans les fenêtres étudiées (b) est représentée de manière linéaire (c), en ne considérant que les pixels de la zone d'intérêt, en blanc sur (a). A partir de cette représentation, on effectue une égalisation d'histogramme sur les images originales (d), puis on normalise afin de rehausser les contrastes (e).

Chacune de ces rétines est doublée. La 1ère couche est suivie par une couche de 10 neurones, elle-même reliée à la dernière couche d'un neurone, correspondant à la sortie du réseau.

Le 2ème perceptron n'est exécuté que dans le voisinage d'une détection offerte par le 1^{er} perceptron. Son but est de localiser les visages, c'est-à-dire qu'il doit être relativement discriminant : peu de positifs, assez peu de faux-positifs. Pour ce deuxième perceptron (ainsi que pour le troisième), la taille totale de la rétine considérée est de 20x20 pixels. L'ensemble des rétines est constitué de 4 rétines de 10x10 pixels, 16 rétines de 5x5 pixels, et 6 rétines de 5x20 pixels se superposant.

Chaque rétine est doublée. La 1ère couche est directement reliée à la 2ème et dernière couche, constituée d'un seul neurone (sortie).

Le 3ème et dernier perceptron n'est exécuté que sur des fenêtres classifiées comme positives par le perceptron précédent. Sa structure est identique à celle du perceptron

précédent, à ceci près que chaque rétine est non pas doublée, mais triplée. Ce perceptron a pour but d'éviter la localisation de faux-positifs suite à l'étape précédente. Un exemple de résultats de détection par l'algorithme de Rowley est disponible sur la Figure 1.8.



Figure 1.8 Quelques détections par l'algorithme de Rowley. Les détections positives selon Rowley sont encadrées en vert.

1.5. Extraction des Composants du Visage

1.5.1. Extraction des Yeux et des Sourcils

Afin d'obtenir des contours de suffisamment bonne qualité pour pouvoir être utilisé dans le cadre d'une reconnaissance d'expressions faciales Les travaux existants souffrent de deux limitations principales : certains proposent une localisation globale grossière de ces traits par l'extraction d'une boîte englobant ces traits [PAN00a], [HAR00] ; d'autres essaient d'extraire plus précisément les contours mais les modèles choisis sont trop simples et peu réalistes [TIA00a], [TIA00b], [TIA00c] et les algorithmes nécessitent une phase de sélection manuelle de points dans la première image. La méthode proposée s'efforce de pallier ces problèmes.

La zone de recherche de chaque iris est limitée aux parties hautes gauche et droite de la boîte englobant le visage. Les dimensions de chaque boîte de recherche de l'iris ont été déterminées suite à une phase d'apprentissage. Sur chaque image de la base ORL [ORL base], une boîte englobant chaque œil a été sélectionnée à la main et les dimensions respectives de ces boîtes ont été étudiées. Il a été déduit les relations suivantes entre les dimensions de la boîte englobant le visage et celles de la boîte englobant chaque œil :

$$Hauteur_{visage} = 4 * Hauteur_{oeil} ; L'arg eur_{visage} = 2.5 * L'arg eur_{oeil} \quad \text{Eq 1.1}$$

1.5.1.1. Détection de l'Iris

Le contour de l'iris représente la frontière entre le blanc de l'œil et la zone sombre associée l'iris. De ce fait, on recherche le contour de l'iris comme un cercle constitué de points de gradient de luminance maximum. Chaque cercle contour de l'iris maximise la quantité :

$$E = \sum_{p \in C} \nabla I(p) \cdot \vec{n}(p) \quad \text{Eq 1.2}$$

où I est la luminance du point p , $n(p)$ est la normale au contour au point p et C est un cercle. Plusieurs cercles sont testés dans la zone de recherche de chaque iris et le cercle qui maximise E est sélectionné. A l'heure actuelle, le rayon du cercle recherché est supposé connu pour des raisons de rapidité de calcul mais il est tout à fait possible de faire également varier la valeur du rayon lors du processus de maximisation de E .

La zone de recherche de chaque iris est limitée aux parties hautes gauche et droite de la boîte englobant le visage. Les dimensions de chaque boîte de recherche de l'iris ont été déterminées suite à une phase d'apprentissage. Sur chaque image de la base ORL [ORL base], une boîte englobant chaque œil a été sélectionnée à la main et les dimensions respectives de ces boîtes ont été étudiées. On en a déduit les relations suivantes entre les dimensions de la boîte englobant le visage et celles de la boîte englobant chaque œil :

$$Hauteur_{visage} = 4 * Hauteur_{oeil} ; L arg eur_{visage} = 2.5 * L arg eur_{oeil} \quad \text{Eq 1.3}$$

1.5.1.2 Modèles Paramétriques pour les Yeux et les Sourcils

Dans beaucoup de travaux, les yeux sont modélisés par deux paraboles. Une étude précise d'un certain nombre de cas montre qu'une parabole n'est pas la courbe la plus adaptée car elle induit une contrainte de symétrie verticale qui n'est pas toujours satisfaite ni pour les yeux ni pour les sourcils. Dans cette méthode, une courbe de Bézier à trois points de contrôle est utilisée. La Figure 1.9 montre le modèle choisi pour chaque œil : pour le contour inférieur, une parabole passant par les trois points P1, P2, P4 est considérée et pour le contour supérieur, une courbe de Bézier à trois points de contrôle P1, P2, P3 est définie.

Pour les sourcils, le modèle usuellement considéré est très simple puisqu'il s'agit de deux lignes brisées passant par les deux coins et un point milieu. Nous proposons de définir,

comme modèle adapté et réaliste, une courbe de Bézier définie par les trois points de contrôle P5, P6, P7 (voir Figure 1.9).

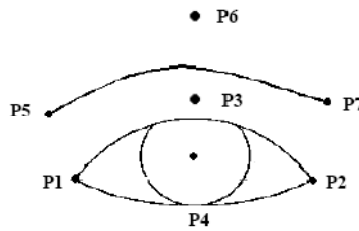


Figure 1.9. Modèle pour l'œil et le sourcil et points caractéristiques P_i

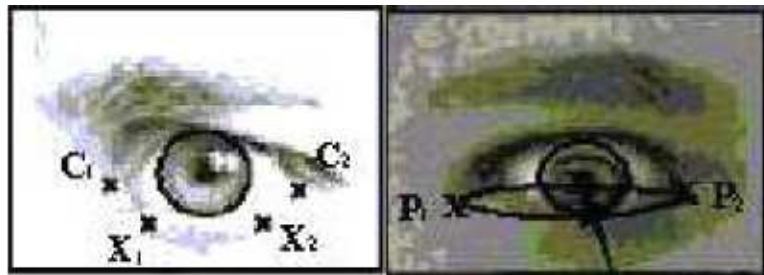


Figure 1.10. à gauche : détection des coins C1 et C2 des yeux; à droite : initialisation du modèle de l'œil.

1.5.1.3 Segmentation des Yeux

Le modèle étant défini, ce dernier est placé sur l'image à traiter grâce à l'extraction automatique des points caractéristiques (P1, P2, P3 et P4) et par la déformation de ce modèle conformément à des informations de maximum de gradient de luminance. En effet, les contours des yeux sont à la limite entre le blanc des yeux (zone claire) et la peau (zone plus sombre).

Un processus de suivi de points de gradient de luminance maximum est utilisé pour détecter les coins des yeux. La partie gauche de la Figure 1.10 donne une illustration de la méthode : partant des points X1 et X2, deux points de gradient de luminance maximum situés près de la verticale des limites du cercle détecté pour l'iris, un processus de suivi, vers la gauche, des points de gradient de luminance maximum conduit à la détection du coin C1. Le même processus de suivi vers la droite conduit à la détection du coin C2. Le chemin suivi entre X1 et C1 (resp. X2 et C2) ne contient que des points de gradient de luminance maximum.

Les points P1 et P2 du modèle sont placés sur les deux coins détectés C1 et C2. Le point P4 de la parabole est aligné avec le point le plus bas du cercle détecté pour l'iris et le point P3 coïncide avec le centre de ce même cercle. La partie droite de la Figure 1.10 présente un exemple d'initialisation du modèle associé à l'œil.

L'idée pour la déformation du modèle initial reste la même : chaque contour est un ensemble de points de gradient de luminance maximum. La courbe sélectionnée est celle qui maximise le flux du gradient de luminance à travers le contour.

Les points (P1, P2 et P4) sont détectés avec suffisamment de précision pour conduire à une parabole qui n'a plus besoin d'être ajustée. A l'opposé, puisque le point de contrôle P3 de la courbe de Bézier est initialisée au niveau du centre de l'iris, on sait que cette courbe doit être déformée. En particulier, il faut que le point P3 se déplace vers le haut (les points P1 et P2 restant fixes) jusqu'à ce que le flux du gradient de luminance à travers le contour soit maximum.

1.5.1.4. Segmentation des Sourcils

P5 et P7 sont les coins de chaque sourcil. Les abscisses x_5 et x_7 de ces deux points correspondent aux maximums droit et gauche de la quantité:

$$H(x) = \sum_{y=1}^{N_y} [255 - I(x, y)] \quad \text{Eq 1.4}$$

et les ordonnées y_5 et y_7 correspondent aux maximums droit et gauche de la quantité :

$$V(y) = \sum_{x=1}^{N_x} [255 - I(x, y)] \quad \text{Eq 1.5}$$

où $I(x,y)$ est la luminance au pixel (x,y) et (N_x, N_y) représentent les dimensions de la boîte englobant chaque sourcil (cette boîte étant limitée à la partie du visage située au dessus des yeux déjà détectés).

Le troisième point de contrôle P6 est déduit des positions de P5 et P7 de la manière suivante : $\{x_6=(x_5+x_7)/2 ; y_6=y_7\}$ De nombreux tests sur différentes images ont montré que la détection de ces points pouvaient conduire à des positionnements grossiers si bien que P5 et P7 doivent être ajustés lors de la phase d'ajustement du modèle aux données. Cet ajustement utilise la maximisation du flux de gradient de luminance, chaque point de contrôle étant déplacé tour à tour.

1.5.2. Détection des Lèvres (de la bouche)

Associé à l'algorithme de segmentation des yeux et des sourcils citée, un autre algorithme est utilisée également pour la détection des lèvres.

Plusieurs modèles paramétriques ont déjà été proposés pour modéliser le contour des lèvres. Des auteurs ont proposé de modéliser les lèvres par deux paraboles [TIA00c], d'autres ont proposé de modéliser le contour supérieur des lèvres à l'aide de deux paraboles au lieu d'une [COI95] ou encore d'utiliser des quartiques [HEN94]. Un gain en précision a été obtenu par rapport à la première idée, néanmoins tous ces modèles sont encore limités par leur trop grande rigidité, en particulier dans le cas d'une bouche non symétrique.

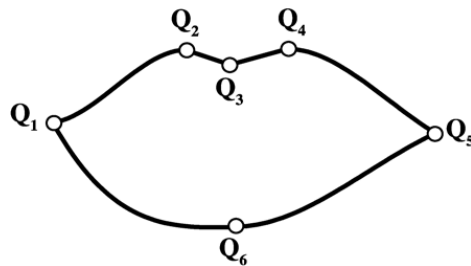


Figure 1.11. modèle choisi pour la bouche

Le choix d'un modèle adapté pour modéliser les lèvres est très délicat car la forme des lèvres est très variable. L'utilisation d'un modèle a priori lors de la phase de segmentation permet une régularisation des contours recherchés. Mais si le modèle choisi n'est pas bien adapté, le résultat de la segmentation ne sera pas de bonne qualité. Le modèle proposé dans cette méthode est composé de 5 courbes indépendantes, chacune d'entre elles décrivant une partie du contour labial. Entre Q2 et Q4, l'arc de Cupidon est décrit par une ligne brisée tandis que les autres portions du contour sont décrites par des courbes polynomiales cubiques γ_i (voir Figure 1.11). De plus, chaque cubique doit avoir une dérivée nulle au point Q2, Q4 ou Q6. Par exemple, γ_1 (cubique entre Q1 et Q2) doit avoir une dérivée nulle en Q2.

La détection de points caractéristiques sur la bouche en vue d'initialiser le modèle est plus complexe et elle se fait en utilisant conjointement une information discriminante combinant la luminance et la chrominance ainsi que la convergence d'un nouveau type de snake nommé « jumping snake ».

1.5.2.1. Analyse Comparative de la Couleur des Lèvres et de la Peau

L'objectif ici est de sélectionner l'espace couleur qui permettra la meilleure distinction possible entre les pixels lèvres et les pixels peau. Il est à souligner que la seule information de luminance n'est pas suffisante à cause essentiellement des nombreuses ombres pouvant apparaître sur le visage, ombres liées à la proéminence de certains segments faciaux (nez, lèvre inférieure...) ou à la présence de plissements de la peau (rides).

Dans l'espace RVB, les pixels peau et lèvres ont des distributions différentes. Certes pour les deux types de pixels, le rouge prédomine. Cependant, il y a plus de vert que de bleu dans le mélange de couleurs associé à la peau alors que ces deux composantes sont équivalentes pour les lèvres [EVE01]. La peau semble plus jaune que les lèvres du fait de la différence entre le rouge et le vert qui est plus marquée pour les lèvres que pour la peau. Hulbert et Poggio [HUL98] proposent de définir une pseudo-teinte h qui met en évidence cette différence :

$$h(x, y) = \frac{R(x, y)}{G(x, y) + R(x, y)} \quad \text{Eq 1.6}$$

où $R(x,y)$ et $G(x,y)$ sont respectivement les composantes rouge et verte au pixel (x,y) . Contrairement à la teinte usuelle, la pseudo-teinte est une fonction bijective. Elle est plus élevée pour les lèvres que pour la peau [EVE01].

La luminance est également une information intéressante à considérer. En général, la lumière provient du dessus du personnage ce qui fait que le contour supérieur des lèvres est très bien éclairé alors que le contour inférieur se retrouve dans l'ombre. Afin de tenir compte de la couleur et de la luminance, on utilise l'information hybride $R_{top}(x,y)$, définie dans [EVE02]. Elle est calculée de la manière suivante :

$$\bar{R}_{top}(x, y) = \bar{\nabla} [h_N(x, y) - I_N(x, y)] \quad \text{Eq 1.7}$$

où $h_N(x,y)$ et $I_N(x,y)$ sont respectivement la pseudo teinte et la luminance au pixel (x,y) , normalisées entre 0 et 1. $\bar{\nabla}$ représente l'opérateur gradient. Cette information hybride permet de faire ressortir la frontière supérieure des lèvres beaucoup mieux que le gradient de luminance ou de pseudo-teinte seul.

1.5.2.2. L'Algorithme du Jumping Snake

Les snakes [KAS88] ont été avantageusement utilisés en segmentation d'images. Cependant, ni le problème difficile du réglage des paramètres du snake, ni celui de sa haute dépendance à la position initiale n'ont pu être résolus. La méthode proposée permet de s'affranchir de ces deux difficultés. Pour détecter des points caractéristiques du contour supérieur de la lèvre, un nouveau type de contour actif est définie, il est désigné sous le nom de jumping snake car sa convergence fait intervenir successivement des phases de croissance et des phases de sauts [EVE03]. Il est initialisé avec un germe noté S_0 positionné

automatiquement dans la zone comprise entre le nez et la bouche près de l'axe de symétrie verticale grâce à la connaissance de la position des yeux et une étude des positions respectives moyennes des yeux et de la bouche effectuée sur la base ORL (voir Figure 12).

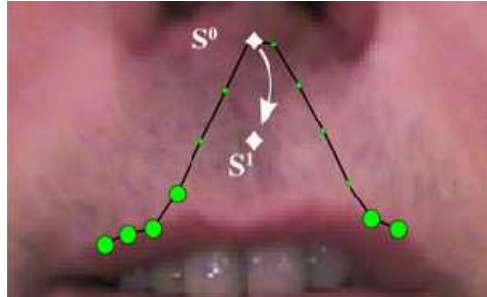


Figure 1.12. Initialisation du jumping snake. La position du germe $S1$ est calculée à partir de $S0$ (initialisé à l'intérieur du cadre noir) et des points (gros ronds) associés aux flux moyens les plus élevés.

Il est important que le germe initial soit positionné au dessus de la bouche et en dessous du nez mais ce positionnement ne requiert pas une très grande précision (par exemple, toutes les positions initiales à l'intérieur du rectangle noir défini de manière expérimentale sur la Figure 1.12 sont possibles) si bien que la position précise du nez n'a pas à être connue. Le jumping snake grandit depuis le germe jusqu'à atteindre un nombre prédéfini de points. Cette phase de croissance est semblable à ce qui est proposé par Berger et Mohr [BER90], en ce sens que le jumping snake est initialisé avec un seul point et qu'il s'accroît progressivement jusqu'à son point final. Dans une seconde phase, le germe «saute» vers une nouvelle position plus proche du contour recherché (voir Figure 1.12). Le processus s'arrête lorsque la taille autorisée pour le saut devient plus petite qu'un pixel (ce qui requiert en moyenne 4 à 5 itérations). Durant la phase de croissance du jumping snake, les meilleurs points à ajouter à droite et à gauche, notés respectivement $M_{-(i+1)}$ et M_{i+1} , sont les points qui maximisent le flux moyen de R_{top} à travers les segments $M_{-(i+1)}M_{-i}$ et M_iM_{i+1} (voir Figure 1.13).

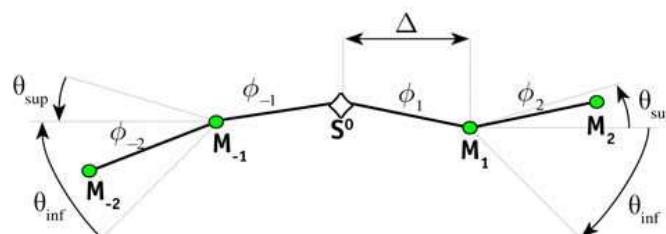


Figure 1.13. depuis le germe $S0$, le snake s'accroît par ajout de points à droite et à gauche suite à la maximisation de R_{top} à travers chaque nouveau segment.

Les deux flux moyens sont définis par :

$$\phi_{i+1} = \frac{\int_{M_i}^{M_{i+1}} \vec{R}_{top} \cdot d\vec{n}}{\|M_i M_{i+1}\|} \quad \phi_{-i-1} = \frac{\int_{M_{-i}}^{M_{-i-1}} \vec{R}_{top} \cdot d\vec{n}}{\|M_{-i} M_{-i-1}\|} \quad \text{Eq 1.8}$$

Où dn est la normale au segment. Les maximisations ϕ_{-i-1} et ϕ_{i+1} s'obtiennent en comparant les valeurs obtenues pour un petit ensemble de points candidats. Une méthode de maximisation de type descente de gradient n'a pas été retenue car, le nombre de points candidat à tester étant faible, (8 points environ pour $(\theta_{inf}, \theta_{sup}) = (-\frac{\pi}{3}, \frac{\pi}{5})$) la méthode d'optimisation exhaustive proposée permet d'obtenir une solution beaucoup plus rapidement. Lorsque le jumping snake a atteint un nombre prédéfini de points noté $2N+1$, la phase de croissance s'arrête. Commence alors la phase de saut de l'algorithme. Soient $\{M_{-N}, \dots, M_{-1}, S^0, M_1, \dots, M_N\}$ l'ensemble des points du snake et soient $\{\phi_{-N}, \dots, \phi_{-1}, \phi_1, \dots, \phi_N\}$ l'ensemble des flux moyens à travers les $2N$ segments. Le nouveau germe S^1 doit être proche d'une zone de fort gradient, à savoir d'une zone de flux moyen élevé. On considère que S^1 est le barycentre de S^0 et des points situés dans les zones de gradients les plus élevés (cf. les gros points sur la Figure 1.13). Soit $\{i_1, \dots, i_N\}$ les indices des N plus grands flux moyens, l'ordonnée du nouveau germe S^1 est calculée par la relation :

$$y_{S^1} = \frac{1}{2} \left(y_{S^0} + \frac{\sum_{k=1}^N \phi_{i_k} y(i_k)}{\sum_{k=1}^N \phi_{i_k}} \right) \quad \text{Eq 1.9}$$

où $y(i_k)$ est l'ordonnée du point M_{i_k} . L'abscisse x_{S^1} du germe est maintenue constante. Le processus complet est alors itéré : un nouveau jumping snake s'accroît à partir de ce nouveau germe jusqu'à ce qu'il atteigne la longueur prédéfinie et une nouvelle phase de saut se produit. La convergence du jumping snake est atteinte lorsque l'amplitude du saut associé au germe devient inférieure au pixel. L'algorithme du jumping snake fait intervenir 4 paramètres : $(\theta_{inf}, \theta_{sup})$ définissant le secteur angulaire d'évolution possible pour chaque branche du jumping snake ; Δ définissant la distance horizontale entre deux points consécutifs du snake et N réglant le nombre de points total de chaque branche du snake. Pour tous les résultats présentés dans cet article, ces paramètres ont été fixés à :

$(\theta_{\text{inf}}, \theta_{\text{sup}}, N, \Delta) = (-\frac{\pi}{3}, \frac{\pi}{5}, 6, 5)$ Pour le choix du secteur angulaire, il est impératif que

$|\theta_{\text{inf}}| > |\theta_{\text{sup}}|$ afin de faire évoluer le jumping snake vers le bas en direction de la bouche. Le choix de N et de Δ résulte d'un compromis entre précision du contour et rapidité de convergence. Une étude complète de l'influence de chacun de ces 4 paramètres est donnée dans [EVE03].

1.5.2.3. Détection des Points Caractéristiques.

On considère 6 points principaux (voir Figure 1.11) : les coins droit et gauche de la bouche (Q1 et Q5), le point central le plus bas de la lèvre inférieure (Q6) et les trois points de l'arc de Cupidon (Q2, Q3 et Q4).

Deux points secondaires situés à l'intérieur de la bouche sont également considérés : Q7 et Q8. Ils sont utilisés pour détecter automatiquement la position du point central bas Q6.

Les trois points supérieurs sont situés sur le contour résultant de la convergence du jumping snake : Q2 et Q4 sont les points les plus hauts de part et d'autre du germe final. Q3 est le point le plus bas du contour situé entre Q2 et Q4.

Les points Q6, Q7 et Q8 sont détectés par analyse de $\nabla_y(h)$, gradient 1D de la pseudo-teinte le long de l'axe vertical passant par Q3. La pseudo-teinte est plus forte pour les lèvres que pour la peau, la langue ou les dents. Le maximum de $\nabla_y(h)$ au-dessous du contour supérieur donne la position de Q7. Q6 et Q8 sont les minima de $\nabla_y(h)$ en-dessous et au-dessus de Q7 respectivement. Ceci suppose que le visage est aligné sur la verticale et donc que les lèvres sont horizontales.

1.5.2.4. Déformation du Modèle

Pour la déformation des cubiques associées aux lèvres, la stratégie est différente. Lorsqu'un être humain recherche les commissures, il utilise implicitement la connaissance qu'il a de la forme globale de la bouche : il suit les 10 contours supérieur et inférieur des lèvres et même en cas de contours devenant peu marqués, il les prolonge et place les commissures à l'intersection (éventuellement interpolée) de ces deux contours. A priori, une cubique est définie de manière unique par la connaissance de 4 paramètres. Dans le cas

présent, chaque cubique passe par l'un des points caractéristiques Q2, Q4 ou Q6 et doit y avoir une dérivée nulle. Ces considérations permettent de réduire le nombre de paramètres à estimer pour chaque cubique de 4 à 2. Il manque donc deux points supplémentaires afin d'obtenir une définition unique de chaque cubique. Les points manquants sont sélectionnés parmi les points les plus fiables des contours à savoir près de Q2, Q4 ou Q6. Pour les deux cubiques associées à la lèvre supérieure, les points manquants sont choisis sur le jumping snake. En ce qui concerne les contours de la lèvre inférieure, seul le point Q6 est connu. Pour obtenir des points supplémentaires, on fait croître un snake à partir du germe Q6. La croissance s'arrête après l'ajout de quelques points (voir les points blancs de la Figure 1.14).

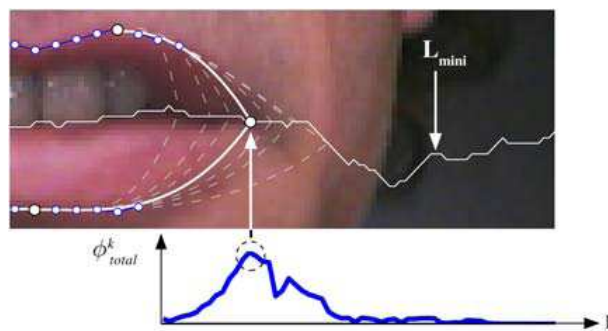


Figure 1.14. le maximum de ϕ_{total}^k donne la position du coin sur L_{mini} . Les courbes en pointillés sont les cubiques associées aux différents points k testés le long de L_{mini} .

Maintenant qu'il y a suffisamment de points sur chaque partie du contour des lèvres, il est possible de calculer les cubiques γ_i passant par ces points et de trouver les commissures en tant que points d'intersection de deux courbes. Afin de rendre les résultats plus robustes, on suppose de plus que les positions des coins (Q1, Q5) sont connues. Ceci permet d'ajouter une contrainte supplémentaire qui réduit encore d'une unité le nombre de paramètres à estimer pour chaque cubique. De là, l'estimation du paramètre manquant est faite très rapidement par minimisation au sens des moindres carrés et les résultats obtenus sont moins sensibles à la position des points supplémentaires choisis. En d'autres termes, pour une position donnée des commissures, il correspond un unique couple de courbes rapide à calculer. Donc finalement, l'ajustement des cubiques se fait en recherchant les coins qui conduisent aux contours les plus précis. Bien entendu, un test exhaustif de l'ensemble des points de l'image en tant que coins potentiels est hors de question. En remarquant que les coins de la bouche sont situés dans des zones sombres de l'image (ce qui est très majoritairement le cas car les commissures se trouvent dans une zone de replis en retrait des deux lèvres), la zone de recherche est considérablement réduite. Ainsi, une recherche du pixel de luminance minimale sur chaque

colonne située entre les frontières supérieure et inférieure est effectuée ce qui conduit à la construction d'une ligne de minima de luminance notée L_{mini} (voir Figure 1.14). Les coins sont alors supposés être sur cette ligne. A chaque coin (droit ou gauche) correspond un unique couple de cubiques, l'une pour la moitié de la lèvre supérieure et l'autre pour la moitié de la lèvre inférieure (cf. les courbes en pointillés sur la Figure 16). L'ajustement du modèle consiste à trouver les coins qui conduisent aux cubiques les plus proches des contours réels qui sont par hypothèse ceux qui satisfont au critère d'optimalité du maximum de flux de R_{top} . Si les cubiques estimées pour le contour supérieur γ_1 et γ_2 coïncident parfaitement avec les contours réels, elles sont orthogonales au champ de gradient. De même, les courbes γ_3 et γ_4 pour le contour inférieur doivent être normales en tout point au champ de gradient $\bar{\nabla}[h_N]$.

$\phi_{\text{top},i}$ et $\phi_{\text{low},i}$ sont calculées, flux moyens à travers les courbes du haut et du bas respectivement par les relations :

$$\phi_{\text{top},i} = \frac{\int_{\gamma_i} \bar{R}_{\text{top}} \cdot d\bar{n}}{\int_{\gamma_i} ds} \quad i \in \{1, 2\} \quad \phi_{\text{low},i} = \frac{\int_{\gamma_i} \bar{\nabla}[h_N] \cdot d\bar{n}}{\int_{\gamma_i} ds} \quad i \in \{3, 4\} \quad \text{Eq 1.10}$$

Où dn et ds représentent la normale en chaque point de contour et l'abscisse curviligne. On considère n positions possibles le long de L_{mini} pour chaque point Q1 et Q5. La meilleure position donne une valeur élevée pour $\phi_{\text{top},i}$ et une valeur fortement négative pour $\phi_{\text{low},i}$. De chaque côté, la quantité suivante est maximisée:

$$\phi_{\text{total}}^k = \phi_{\text{top},\text{normalisé}}^k - \phi_{\text{low},\text{normalisé}}^k, \quad k \in \{1, \dots, n\} \quad \text{Eq 1.11}$$

Avec :

$$\phi_{\text{top},\text{normalisé}}^k = \frac{\phi_{\text{top}}^k - \min_{j \in \{1, \dots, n\}} \{\phi_{\text{top}}^j\}}{\max_{j \in \{1, \dots, n\}} \{\phi_{\text{top}}^j\} - \min_{j \in \{1, \dots, n\}} \{\phi_{\text{top}}^j\}} \quad \text{Eq 1.12}$$

ϕ_{top}^k et ϕ_{low}^k étant associés au k ème point testé. $\phi_{\text{top},\text{normalisé}}^k$ et $\phi_{\text{low},\text{normalisé}}^k$ sont des valeurs normalisées sur l'ensemble des points testés. Dès lors que ϕ_{total}^k est grand, la position du coin est fiable car les courbes estimées coïncident bien (visuellement) avec les

contours réels ce qui montre qu'ajustement du modèle et détection des commissures sont réalisés en une seule opération. La Figure 1.14 montre l'évolution de ϕ^k_{total} pour différentes valeurs de k. Le maximum de ϕ^k_{total} donne la position du coin le long de L_{mini} .

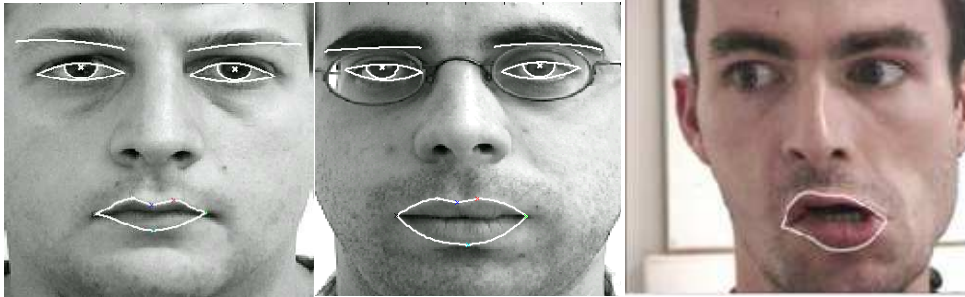


Figure 1.15 : Quelques exemples de contours des yeux, sourcils et lèvres détectés.

1.6. Conclusion

Dans ce chapitre nous avons présenté les différentes méthodes utilisées pour détecter en premier lieu le visage ensuite les contours des yeux, des sourcils ainsi que les lèvres. Ces travaux sont réalisés dans le cadre de la recherche d'un projet développé au sein du laboratoire GIPSA par des chercheurs dans le domaine. Différents modèles sont utilisés pour modéliser le mouvement de chacun des traits permanents du visage. Cette détection a permis la localisation des points caractéristiques du visage. Ces points sont la base des travaux qui seront réalisés au cours de cette thèse. Ils permettront le calcul de certaines distances faciales ainsi que la localisation des régions faciales ou des traits transitoires pourront apparaître.

Chapitre 2

Classification Catégorielle des Expressions Faciales

2.1. Introduction

Le processus de reconnaissance du visage est un processus dédié chez l'humain. En effet, il semblerait que le modèle utilisé par les humains pour reconnaître une expression puisse se résumer à une indication sur la forme des composantes du visage. Ainsi un ensemble de points représentant la position de chaque composante du visage suffit pour qu'un humain reconnaisse une expression. Le processus de reconnaissance des expressions, se base sur les différences intra-individus. « *facial expression identification requires finding something **common** across individuals, while face identification requires finding something **different** » ». On distingue deux types d'analyses du visage effectués par le cerveau humain : l'une dite globale (*holistic*) où le visage est traité comme un tout et l'autre dite par composantes (*feature-based*) où le visage est vu comme un ensemble de composantes (yeux, nez, bouche, etc.).*

La décomposition des méthodes de reconnaissance automatique des expressions en deux approches (par composantes et globale) est relativement naïve. Il reste un certain nombre de méthodes qui exploitent les avantages des deux approches présentées. D'une manière générale, toutes les méthodes globales, peuvent être appliquées de manière locale sur des composantes du visage en particulier, puisqu'elles ont été conçues dans un but de généralité.

On présente ici un ensemble de travaux sur l'analyse des expressions du visage afin de mettre notre travail dans le contexte. On pourra se référer aux états de l'art existants pour plus de détails (PAN00) [FAS03]).

2.2. Etat de l'Art

Différentes Approches de Reconnaissances Automatiques d'Expressions Faciales ont été proposées dans la littérature. Elles peuvent être divisées globalement en trois approches principales qui sont : L'approche basée Modèles, l'approche basée règles et l'approche basée

géométrie. Ces trois types d'approches utilisent différentes techniques. Les réseaux neurones et le flux optique sont les techniques les plus utilisées dans chacune des trois approches. Afin d'analyser les points forts et points faibles de ces techniques nous avons complété les trois approches par deux autres approches, une approche qui regroupe les méthodes qui utilisent les réseaux neurones et une autre qui regroupe les méthodes qui utilisent le flot optique.

2.2.1 Approche Basée Modèles

Cette approche consiste à voir le visage comme un tout. L'analyse consiste alors à mesurer la ressemblance du visage observé à un modèle (connu ou appris). Les méthodes de cette approche font généralement appel à des méthodes de mise en correspondance de modèles. L'intérêt de ces méthodes est qu'elles peuvent être appliquées de manière plus locale. Ainsi, la plupart des méthodes présentées dans cette section, valable pour le visage, sont généralement aussi valable pour n'importe quel objet et plus particulièrement pour les composantes du visage.

Afin de réaliser une classification catégorielle selon l'une des six expressions universelles définies par Ekman [EKM82] auxquelles s'ajoute l'expression neutre, Edwards *et al.* [EDW98a] utilisent le modèle actif d'apparence (AAM) pour reconnaître l'identité d'un individu observé de manière robuste par rapport à l'expression faciale ainsi que l'illumination et la pose. Afin d'atteindre ce but, ils utilisent la distance de Mahalanobis comme critère de similarité, une analyse discriminante linéaire (ADL) est appliquée pour maximiser la séparation des classes. Afin d'évaluer les performances de cette méthode, 2 X 200 images correspondantes aux six expressions faciales universelles, présentées par 25 sujets ont été testé. Les taux de reconnaissance réalisés pour les six expressions en plus de la neutralité étaient de 74%. Les auteurs ont expliqué ce taux relativement bas par la limitation des classificateurs linéaires. Par contre les résultats ne sont pas connus dans le cas des sujets inconnus.

Hong *et al.* [HON98] partent du principe que deux personnes qui se ressemblent affichent la même expression de manière similaire. Un graphe étiqueté est attribué à l'image de test puis la personne connue la plus proche est déterminée à l'aide d'une méthode de mise en correspondance de graphes élastiques. La galerie personnalisée propre à cette personne est alors utilisée pour reconnaître l'expression faciale de l'image de test. La méthode a été testée sur des images de 25 sujets. Les taux de reconnaissance étaient de 89% dans le cas de visages familiers et de 73% dans le cas de visages inconnus.

Un graphe étiqueté par des réponses de filtres de Gabor est par ailleurs utilisé par Lyons *et al.* [LYO99] et Bartlett *et al.*[BAR03]. L'ensemble des graphes construits sur un ensemble d'apprentissage est ensuite soumis à une ACP puis analysé à l'aide d'une analyse discriminante linéaire (ADL) afin de séparer les vecteurs dans des classes ayant des attributs faciaux différents. Le graphe étiqueté de l'image testée sera alors projeté sur les vecteurs discriminants de chaque classe afin de déterminer son éventuelle appartenance à cette classe. Pour tester leur méthode, Lyons et al ont utilisé un ensemble de 193 images avec différentes expressions affichées par 9 japonaises. Les taux de reconnaissances étaient de 92% pour les sujets familiers et de 75% pour les sujets inconnus.

Huang et Huang[HUA97] calculent 10 paramètres d'actions (APs). La différence entre le model utilisé et l'expression étudiée génère les (APs). Les deux premières valeurs propres sont utilisées pour représenter les variations des (APs), Les auteurs utilisent un classificateur de distance minimale pour classifier les deux principaux paramètres d'actions parmi les 90 images d'apprentissage dans une des six classes universelles d'expressions. Les Trois meilleures classes qui correspondent sont sélectionnés. Le score le plus élevé des trois corrélations détermine la classification finale de l'expression étudiée. La méthode proposée a été testé sur 90 autres images pour les mêmes sujets. Les taux de reconnaissances réalisées étaient de 84,5%. Les résultats sont inconnus dans le cas de sujets inconnus.

Taylor et Cootes [COO94] ont introduit le concept d'*Active Shape Model* et d'*Active Appearance Model* qui consiste à modéliser le visage en prenant en compte à la fois les informations de forme et les informations d'apparence. Un ensemble de points de contrôle est placé manuellement sur un ensemble de visages d'apprentissage.

De ces points, on déduit un arrangement spatial et on mémorise l'information de couleur (ou de niveaux de gris) de cette forme. En effectuant une analyse en composantes principales sur les données d'apprentissage (aussi bien sur les informations de forme que d'apparence), on peut ainsi recomposer un visage.

Ahlberg [AHL01] utilise une méthode inspirée des *Active Shape Models* à la différence près qu'il utilise un modèle de visage générique et non construit à partir des données (Candide). Ainsi, le système d'Ahlberg possède une bonne robustesse à l'occultation sans que le modèle soit difficile à construire.

Avantages des méthodes de cette approche : Les avantages principaux de ces méthodes sont qu'elles sont génériques et peuvent donc s'adapter à beaucoup de problèmes. Ces méthodes sont généralement moins sensibles au bruit que les méthodes classiques, puisqu'il

existe un modèle sous-jacent de comparaison. En plus elles sont robustes par rapport à l'occultation pour des visages de la base d'apprentissage et même sur des visages non connus (*hors* de la base d'apprentissage).

Inconvénients des méthodes de cette approche : Pour que ces méthodes soient robustes aux changements de pose et/ou d'illumination, il faut intégrer dans la base d'apprentissage des visages ayant différentes poses et/ou illuminations. La construction de la base d'apprentissage (corpus) devient alors difficile.

De plus, la taille de la matrice d'apprentissage est fixée et les images à décomposer doivent être éventuellement redimensionnées. En plus La construction du modèle est très longue et nécessite un corpus intelligent (adapté au problème). En plus, l'extraction initiale de la forme des visages est effectuée manuellement.

Du point de vue de l'analyse des expressions du visage, ces méthodes souffrent d'un autre inconvénient : elles ne donnent que des configurations et non des mesures. Elles peuvent être vues comme des méthodes de classification, des méthodes qui indiquent que l'observation se trouve dans un ensemble préétabli de configurations. Il est par exemple difficile de concevoir un réseau de neurones qui indique quel est le degré d'ouverture, en pixels, de la bouche ; il est plus facile de construire un réseau de neurones qui décide si la bouche est ouverte ou fermée. Ces méthodes sont donc bien adaptées à l'extraction d'informations sur des composantes dont certains états sont difficiles à quantifier par des opérateurs classiques (le gonflement de la joue par exemple) ou à l'extraction d'informations sur des composantes n'ayant qu'un nombre restreint d'états.

2.2.2. Approche Géométrique

Les années récentes ont vu l'utilisation croissante de l'analyse géométrique pour représenter l'information faciale à cause de sa capacité de quantification car elle donne des informations sous forme de mesures (chose qui n'était pas possible avec les méthodes de l'approche basée modèles). Dans cette approche, les mouvements faciaux sont quantifiés en mesurant les déplacements géométriques des points caractéristiques des traits permanents du visage entre deux images, une image expressive et une autre à l'état neutre.

Lien et al [LIE98] proposent une méthode hybride basée sur : premièrement suivi des points faciaux (points autour des contours des yeux, des sourcils, du nez et de la bouche manuellement localisés sur la première image de la séquence vidéo), deuxièmement, le flux

optique et troisièmement détection des rides afin d'extraire des informations sur l'expression. La classification des expressions est basée sur le système de codage des actions faciales (FACS) [EKM78]. Les chaînes de Markov HMM sont utilisées pour la discrimination entre les unités d'actions. Un lien entre les états du HMM représente la transition inhérente possible d'un état facial à un autre. Une unité d'action est identifiée si la chaîne de Markov correspondante a la plus grande probabilité parmi tous les HMMs. L'inconvénient principal de la méthode est le nombre important de HMMs exigé pour détecter un grand nombre d'AUs ou de combinaison d'AUs impliqué dans la classification faciale d'expressions.

Tian et al [TIA01] utilisent deux réseaux de neurones séparés afin de reconnaître 6 unités d'actions de la partie supérieure du visage et 10 de la partie inférieure, leur méthode est basée sur les traits permanents du visage ainsi que les traits transitoires. Les traits faciaux sont initialisés manuellement sur la première image et suivis dans le reste des images de la séquence vidéo. La reconnaissance des expressions faciales est réalisée par la combinaison des unités d'actions des deux parties supérieure et inférieure du visage.

Pardas et al [PAR02] et Tsapatsoulis et al [TSA00] proposent une description des six expressions en utilisant l'ensemble de définition des paramètres faciaux de MPEG-4. Ils utilisent tous les FAPs et proposent une classification basée sur un système d'inférence floue.

Cohen et al [COH03] 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. La représentation basée composantes requiert une détection efficace et fiable des composantes faciales pour faire face à la variation de l'illumination, du mouvement principal significatif et de la rotation.

Inconvénients des méthodes de cette approche : Le problème majeur des méthodes de cette approche est la détection non exacte des points caractéristiques du visage sur lesquels des mesures géométriques seront calculées à cause des variations de l'illumination, de l'occlusion et de beaucoup d'autres facteurs.

2.2.3. Approche Basée Règles

Une seule parmi les méthodes rencontrées dans la littérature est basée sur des règles. Elle a été proposée par Pantic et Rothkrantz [PAN00a]. La méthode réalise une codification automatique en unités d'action. Un multi détecteur automatique de composantes du visage est

appliqué. Depuis les contours des composantes, des modèles de composantes sont extraits. La différence est ensuite calculée entre le modèle détecté et le modèle correspondant du visage à l'état neutre. Basé sur les connaissances acquises de FACS, les règles de production classifient le modèle détecté dans la classe AUs appropriée. La méthode proposée a été testée sur un ensemble de 496 vues. Les taux de reconnaissance étaient de 92% pour les unités d'actions de la partie supérieure du visage et 86% pour celles du bas du visage. La classification finale dans une des six catégories correspondantes aux six expressions universelles est réalisée à l'aide de la description linguistique donnée par Eckman [EKM82]. La performance globale du système a été testée sur l'ensemble de 265 vues représentant six expressions faciales affichées par 8 sujets. Le taux de reconnaissance réalisé était de 91%.

Inconvénients des méthodes de cette approche : Le problème majeur de ces méthodes est qu'aucune méthode n'arrive à reconnaître les 45 unités d'actions définies dans FACS.

2.2.4 Approches Basées Réseaux de Neurones

Un réseau de neurones peut être vu comme une fonction ayant un certain nombre d'entrées et un certain nombre de sorties. Le principe de l'apprentissage est de donner en entrée du réseau un certain nombre d'exemples et de fixer la sortie à la valeur désirée. Une méthode d'apprentissage permet alors au neurone de s'adapter au mieux pour qu'il affiche la même sortie quand on lui donnera des données *proches* des données d'apprentissage.

Hara et Kobayashi [KOB97] appliquent un réseau de neurone de 234 X 50 X 6. Les unités d'entrée du réseau correspondent au nombre de certaines données extraites de l'image à étudier, tandis que les unités de sortie du réseau correspondent à une catégorie d'expressions. Le RN a été entraîné sur 90 images présentant les six expressions faciales, affichées par 15 sujets, et il a été testé sur 90 expressions affichées par 15 autres sujets. La moyenne des taux de reconnaissance est de 85%.

Padgett et Cottrell [PAD96] utilisent un RN qui a comme entrée Sept blocs de pixels de 32 X 32. La couche cachée du RN contient 10 nœuds et emploie une fonction sinusoïdale d'activation. La couche de sortie contient Sept unités, chacune correspond à une catégorie des expressions faciales. Les auteurs utilisent des images avec les six expressions plus le neutre de 12 sujets. Ils ont entraîné le RN sur les images de 11 sujets et l'ont testé sur celles du 12ème sujet. La moyenne de taux de reconnaissance réalisée était de 86%.

Zhang et al [ZHA98] emploient un réseau 680 X 7 X7. Les entrées consistent à des positions géométriques de 34 points caractéristiques faciales et à 18 coefficients de Gabor prélevés à chaque point. Chaque unité de sortie donne une évaluation de la probabilité de l'expression examinée appartenant à la catégorie associée. Le RN réalise une réduction non linéaire de la dimensionnalité des entrées et prend une décision statistique sur la catégorie de l'expression étudiée. Le RN a été entraîné et testé sur un ensemble de 213 images présentant différentes expressions affichées par neuf japonaises. Le taux de reconnaissance réalisé était de 90,1%. Les performances du réseau n'ont pas été testées pour des sujets inconnus.

Zhao et Kearney [ZHA96] ont utilisé un réseau de 10 X10 X3. 94 images présentant six expressions faciales sélectionnées de la base réunie par Eckman et Friesen [EKM75] ont été utilisées pour tester leur algorithme. Sur chaque image, 10 distances ont été mesurées manuellement. La différence entre chaque distance de l'état expressif et l'état neutre a été normalisée. Ensuite chaque distance normalisée est associée à un des huit intervalles de la déviation appropriée depuis la moyenne correspondante. Ces intervalles forment les entrées du RN. Les sorties du réseau correspondent à l'expression associée. Le réseau a été entraîné et testé sur 94 images avec un taux 100% de reconnaissance. Les résultats ne sont pas connus dans le cas de sujets inconnus.

Avantages des méthodes de cette approche: Les réseaux de neurones sont généralement utilisés pour leur faible sensibilité au bruit (robustesse au bruit) et leur capacité d'apprentissage.

Inconvénients des méthodes de cette approche: Malheureusement, les réseaux de neurones, 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.

2.2.5 Approches Basées Flux Optique

L'information précise du mouvement peut être obtenue en calculant le flux optique, qui représente la direction et l'importance d'un mouvement. Plusieurs travaux dans l'analyse des expressions faciales se sont concentrés sur l'analyse du flux optique de l'action faciale où le flux optique est employé pour modéliser les activités musculaires ou bien pour estimer les déplacements des points caractéristiques.

Yacoob et Davis [Yac96] ont proposé une représentation du mouvement facial basé sur le flux optique pour identifier les six expressions faciales universelles. Cette approche est divisée en trois étapes: premièrement, des régions rectangulaires encadrant les composantes faciales du visage (yeux, sourcils, nez) sont supposées données sur la première image de la séquence vidéo et sont ensuite suivies tout le long de la séquence; deuxièmement une évaluation du flux optique de ces composantes définit la représentation du niveau moyen qui décrit les changements faciaux observés sur chaque image par rapport à la première image (mouvement rigide et non rigide); troisièmement cette représentation du niveau moyen est classifié dans un des six expressions faciales en utilisant un système basé règles qui combine les actions de base des composantes [Bas78] et les règles de sélections de mouvement décrit dans [Ekm78].

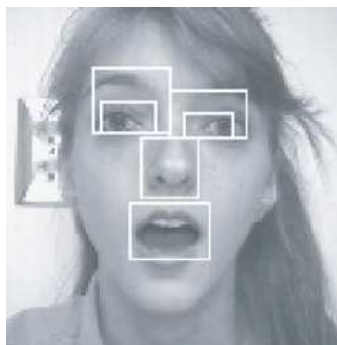


Figure 2.1. Un exemple de rectangles entourant les régions faciales d'intérêts [Yac96].

Rosenblum et al [Ros96] ont étendu le système ci-dessus en ajoutant une fonction radiale de réseau neurone pour chacune des six expressions faciales. Ces fonctions permettent l'établissement des corrélations entre les types de mouvement faciaux et les expressions faciales. Cependant, le système proposé a été testé uniquement avec la joie et la surprise. Pour améliorer la précision du modèle et afin d'être robuste par rapport au mouvement de la tête, Black et Yacoob [Bla97] ont présenté une approche d'un modèle local paramétré du mouvement de l'image pour l'analyse faciale d'expression. Les mouvements rigides de la tête

sont représentés par un modèle planaire afin de récupérer l'information sur le mouvement de la tête. Le mouvement des composantes faciales est déterminé relativement au visage. Les mouvements non rigides des composantes faciales (yeux, sourcils et bouche) sont représentés par un modèle affine. Un ensemble de paramètres estimés à partir des modèles par un schéma de régression [Bla93] est employé pour définir les attributs à mi-niveau qui décrivent le mouvement des composantes faciales. Un ensemble de règles est alors défini pour combiner les attributs à mi-niveau entre le début et la fin des expressions afin de les reconnaître.

Essa et Pentland [Ess97] ont proposé la combinaison d'un modèle dynamique physique et l'énergie de mouvement pour la classification d'expressions faciales. Le mouvement est estimé à partir du flux optique et est raffiné par le modèle physique dans une évaluation récursive. Un modèle physique de visage est appliqué pour modéliser l'activation des muscles et un idéal mouvement 2D est calculé pour les cinq expressions étudiées (colère, dégoût, joie, surprise et les expressions avec sourcils levés), (il est difficile de simuler les autres expressions par les sujets). Chaque modèle est délimité en moyennant le mouvement produit par deux sujets pour chaque expression. La classification faciale d'expressions est basée sur la distance euclidienne entre le modèle instruit de l'énergie de mouvement et de l'énergie de mouvement de l'image observée.

Cohn et al [Coh98] proposent un système automatique de classification basé sur la modélisation des unités d'actions AUs. Le déplacement de 36 points caractéristiques localisés manuellement autour des yeux, sourcils, nez et bouche (voir le schéma 3,9) sont estimés en utilisant le flux optique. Les matrices de variances –covariances sont employées pour la classification des AUs. Les auteurs utilisent deux fonctions discriminantes pour trois AUs dans la région des sourcils, deux fonctions discriminantes pour trois AUs dans la région de l'œil et cinq fonctions discriminantes pour neuf AUs dans les régions du nez et de la bouche. Cependant, le flux optique est calculé pour chaque Pixel dans une région d'intérêt spécifique. Cette approche ne permet pas toujours de distinguer entre le flux provoqué par le mouvement des composantes faciales et celui causé par le bruit, menant ainsi à de fausses détections. Par exemple, dans le cas de la surprise, la détection de la région bouche correspondant à un état ouvert peut être causé par la variation de luminance.

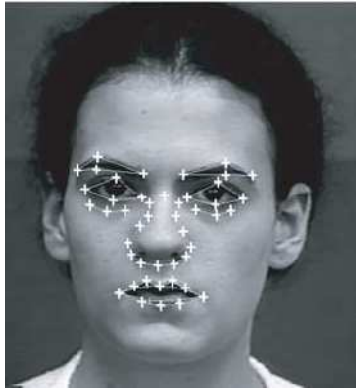


Figure 2.2. Suivi de point caractéristiques [Coh98].

Inconvénients des méthodes de cette approche : Le problème avec ce type de méthodes est également l'intervention manuelle par exemple pour la localisation des régions initiales entourant les composantes faciales. Par ailleurs, des évaluations du flux optiques sont facilement perturbées par les mouvements non rigides. Elles sont également sensibles à l'inexactitude de l'enregistrement d'image et aux discontinuités du mouvement.

Un résumé des différentes méthodes évoquées dans cet état de l'art est proposé dans la table 2.1.

Méthodes de l'état de l'art	Images	Catégories	Performance
Approche basée Modèles			
Analyse linéaire discriminante en utilisant la distance de Mahalanobis [EDW98]	2 X 200 images de 25 sujets	Six expressions universelles & le neutre	74%
Graphes Elastiques Matching avec les galeries personnalisées [HON98]	Images statiques tirées de séquence naturelles 25 subjects	Six expressions universelles	89% -> sujets familiers; 73% -> sujets inconnus.
Graphes étiquetés, ACP, ADL [LYO99] et [BAR03]	193 images 9 sujets	Six expressions universelles	92% -> sujets familiers; 75% -> sujets inconnus.
ACP avec Les Paramètres d'Actions, [HUA97]	images Statiques 9 sujets	Six expressions universelles	84,5% -> sujets familiers
Approche basée Réseaux de Neurones			
Réseau de Neurone Hopfield avec flux optique	images statiques	Tristesse, surprise, colère et joie	92,2%
Réseaux de neurones avec Back-propagation. ([KOB97], [PAD96], [ZHA98], [ZHA96])	Images statiques 9 à 15 sujets	Six expressions universelles	85% jusqu'à 100%

Approche basée Règles			
Méthode basée règles [PAN00a]	Images statiques 8 sujets	Reconnaissance d'unités d'actions	92% -> Aus sup 86% -> Aus Inf
Approche basée flux optique			
Flux optique + système à base de règles [YAC96]	32 sujets: Jo(37), Co(24) su(30), dé(13) Pe(7), Tr(5)	Six expressions universelles	88%
Flux optique + réseau de neurone [ROS96]	32 subjects	Joie et Surprise	88%
Modèle local paramétré [BLA97]	40 sujets: su(35), Co(20) Tr(8), Jo(61) Pe(6), dé(15)	Six expressions universelles	88%
Modèle dynamique physique et énergie de mouvement [ESS97]	7 sujets: su(10), rai_eye(10) Jo, Co(10)	Colère , dégoût, joie, surprise	98%
Modélisation des unités d'actions AUs et fonctions discriminantes [COH98]	100 sujets	8 AUs +7 AUs Combinaison	88%
Approche géométrique			
HMMs[LIE98]	85 sujets de la base Cohn et kanade	3AUs sup 6 Aus inf	85% 88%
Inférence floue [TSA00]		6 expressions	81,16%
Réseaux de neurones [TIA01]	14 AUs sup 32 AUS inf	6AUs sup 10 Aus inf	96,4% 96,7%
HMM [Par02]		6 expressions	84%
Réseau bayésien [Coh03a]	5 sujets	6 expressions	86,45%

Table 2.1. Comparaison des différentes méthodes selon les différentes approches

Comme nous venons de le montrer, toutes les méthodes relatives à chaque approche présentent des avantages tout autant que des inconvénients, Notre objectif est de tirer profit des avantages existants et de proposer de nouvelles solutions afin d'éviter les inconvénients sous jacents. La plupart de ces méthodes nécessitent une intervention manuelle au début du traitement. Elles proposent une classification d'une expression étudiée dans une seule catégorie universelle tel que postulé par Ekman: joie, Surprise, Colère, Dégout, Tristesse et Peur, hors ceci n'est pas réel car l'être humain n'est pas binaire. Pour passer d'une expression à une autre le visage passe par des expressions transitoires mélangées, par conséquent, la classification d'une expression dans une catégorie simple d'émotion n'est pas réaliste et,

idéalement, le système de classification devrait pouvoir identifier un tel mélange intermédiaire des expressions. Enfin la plupart de ces méthodes sont basées principalement sur les déformations des traits permanents du visage (Yeux, Sourcils et Bouche) car ils estiment que ces traits portent suffisamment d'informations utiles pour la classification des expressions faciales. Très peu de méthodes étudient la présence des traits transitoires (rides) sur certaines régions du visage dans un post traitement, afin de discriminer entre deux expressions.

Pour toutes ces raisons, nous proposons dans ce chapitre une méthode géométrique de classification des expressions faciales basée principalement sur les traits transitoires pouvant apparaître sur n'importe quelle région du visage en utilisant la théorie de l'évidence (Nous expliqueront plus loin les raisons du choix de cette méthode). Notre challenge est de prouver qu'une méthode basée traits transitoires peut être aussi performante qu'une méthode basée traits permanents.

2.3. Notre Contribution

L'expression faciale est due à l'activation d'un ou de plusieurs muscles du visage, ce qui produit une déformation des traits permanents du visage (Yeux, sourcils et bouche) ainsi que la formation de certains traits transitoires (connus sous le nom « rides d'expressions ») sur certaines région faciales.

Les rides correspondent à un relâchement ou contraction de la peau. Ce relâchement ou contraction peut être dû à l'usure de la peau (dû à l'âge) ou à l'activation d'un muscle. Il est donc nécessaire, non seulement de *détecter* la présence ou l'absence de rides, mais aussi de les *quantifier* ceci pour distinguer les rides permanentes de celles engendrées par les muscles.

En plus de la configuration des différentes composantes (yeux, sourcils, bouche, etc.) du visage, il est important de pouvoir caractériser les rides d'expression.

Les rides les plus importantes du visage sont les rides du haut du nez (froncement du nez), du front (relèvement des sourcils), du coin des yeux (plissement des yeux) et les rides naso-labiales qui interviennent généralement lors du sourire.

Les rides apparaissent sur les images sous forme d'une forte différence de luminosité. Utiliser un détecteur de contours (détectant aussi son orientation) sur des zones pertinentes de l'image (front, nez, coin des yeux, ...) permet d'extraire l'information sur les rides.

Certains travaux rencontrés dans la littérature ce sont intéressés à définir quelle est la partie la plus indicatrice et plus pertinente pour reconnaître une expression faciale [BOU75], [BAS78], [GOU00] : La partie supérieure du visage, la partie inférieure du visage ou bien tout le visage. L'intérêt de ce travail, porte sur les parties du visage ou les traits transitoires peuvent apparaître.

Plusieurs études se sont concentrées sur les traits transitoires mais pas pour la reconnaissance des expressions. La présence ou l'absence des TTs sur un visage peut être déterminée par l'analyse des composants des contours [KWO94], [LIE00] ou par l'analyse des images propres (eigen-image) [KIR90], [TUR90]. Terzopoulos et Waters [TER94] détectent les TTs nasolabial pour l'animation de visage, mais avec des marqueurs artificiels. Kwon et Lobo [KWO94] détectent les TTs en utilisant des snakes pour classifier des images de personnes dans différentes catégories d'âge. Ying-Li Tian et Takeo Kanade [LIE00] détectent les contours horizontaux, verticaux et diagonaux en utilisant un modèle de visage complexe, ensuite ils emploient le détecteur de contour de « Canny » pour mesurer la densité et l'orientation des TTs [LIE01].

Nous estimons qu'il est plus facile et plus rapide de détecter la présence ou l'absence des TTs sur un visage que de calculer des distances depuis certains point caractéristiques, les comparer à d'autres (ceux de l'état neutre) et enfin déduire si ces distances croissent ou décroissent. En plus, il est souvent difficile de détecter les points caractéristiques d'une façon très précise pour pouvoir donner des distances exactes, par ailleurs, il est plus simple de savoir si des TTs existent ou n'existent pas sur n'importe quelle zone faciale. L'idée est de montrer que ce type de trait véhicule lui aussi suffisamment d'informations pour identifier une expression étudiée.

D'un autre côté, l'imperfection des informations fait appel à plusieurs concepts. Le premier, généralement bien maîtrisé, concerne l'imprécision des informations. L'incertitude est un second concept à différencier de l'imprécision par le fait qu'il ne fait pas référence au contenu de l'information mais à sa "qualité". Enfin, l'incomplétude peut se rencontrer dans de nombreuses applications.

Modéliser des connaissances aussi hétérogènes devient alors rapidement un problème crucial. La communauté scientifique s'attarde d'ailleurs de plus en plus sur ce type de problématique (KDD : Knowledge Discovery in Databases). Le problème se complique encore lorsqu'il s'agit de fusionner ces informations dans un but de prendre la meilleure décision au sens d'un critère.

Différentes théories permettant la fusion de données sont disponibles pour modéliser dans un même formalisme mathématique les informations provenant d'une ou plusieurs sources. La plus répandue et la plus ancienne d'entre elles est la théorie des probabilités. Néanmoins, cette dernière présente des insuffisances lorsque la connaissance est incomplète, incertaine ou imprécise. Différentes théories sont alors apparues parmi lesquelles la théorie des possibilités et la théorie de l'évidence, ces dernières ont fait récemment l'objet de recherches approfondies.

Une synthèse de ces trois approches a fait l'objet d'une étude proposée afin de choisir celle qui semble la plus appropriée dans le domaine de fusion de données dans un contexte d'analyse des expressions faciales [RAM]. Cette étude a permis de choisir la théorie de l'évidence comme meilleur outil de fusion et comme seul outil (parmi les méthodes de fusion) qui permet de modéliser le doute.

La théorie de l'évidence [GIR05], [DEN06], [MER06], [RAM07], est bien adaptée à ce type de problèmes. Ce modèle facilite l'intégration de connaissances a priori et peut traiter les données incertaines et imprécises issues des algorithmes automatiques de segmentation. Il est bien adapté quand il est question de manque de données. En outre il peut modéliser le doute intrinsèque qui peut se produire entre les expressions faciales dans le processus de classification (voir la figure 2.3). La classification d'une expression faciale dans une catégorie simple d'émotion n'est pas réaliste parce que les expressions humaines sont variables selon l'individu.



Figure 2.3. Exemple de doute entre Surprise et Peur

2.4 Méthode Proposée

L'approche proposée consiste principalement en cinq étapes: segmentation, extraction de données, analyse de données, classification et enfin post traitement.

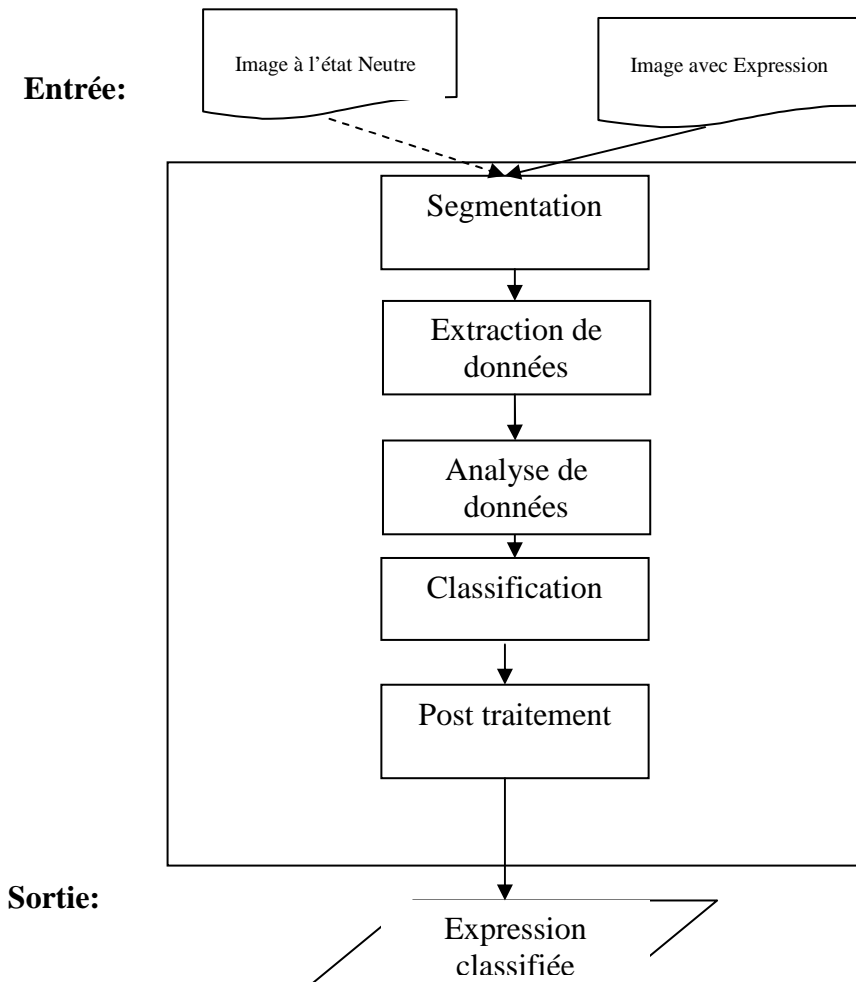


Figure 2.4. Démarche proposée

Dans l'étape de segmentation, une image faciale avec expression est présentée au système pour une éventuelle détection de contours de tous les traits transitoires ainsi que les traits permanents. Une image de référence (qui représente l'image du visage à l'état neutre) est segmentée une seule et unique fois pour chaque sujet afin de permettre certaines comparaisons qui facilitent l'analyse des expressions faciales.

Dans l'étape d'extraction de données, et pour chaque région de ride un rapport qui correspond au nombre des Pixels du contour du visage expressif par le nombre des Pixels de contour du visage neutre, est calculé, afin de savoir si des rides existent ou n'existent pas sur

une zone du visage. Dans la phase de post traitement, l'angle des rides nasolabial (s'ils existent) est calculé ainsi que quelques distances qui correspondent au degré d'ouverture ou de fermeture des yeux et le degré de froncement ou élévation des sourcils.

Dans l'étape d'analyse de données, nous associons à chaque région faciale de ride un état " présent " ou " absent ", puis nous caractérisons chaque région de ride par une combinaison d'expressions associées à la présence des rides sur ces zones. Ce processus est effectué après une phase d'apprentissage ou nous déterminons quel type de ride existe avec un type d'expression.

Dans l'étape de classification, le modèle de croyance connu également sous le nom de Théorie de Dempster-Shefer (TBM) est appliqué pour identifier l'expression faciale.

En conclusion, l'étape de post traitement, raffine le résultat obtenu dans l'étape de classification en réduisant le doute entre les expressions possibles.

2.4.1 Segmentation

L'étape de segmentation consiste à détecter le visage en premier lieu ensuite détecter les contours des traits permanents du visage qui sont : les yeux, les sourcils et la bouche. Le visage est détecté par la méthode proposée par [ROW98], Les yeux et les sourcils sont détectés par la méthode présentée dans le chapitre I [HAM07] et enfin la bouche est détectée par la méthode présentée dans le même chapitre I [EVE04].

La détection de ces traits permanents a permis de localiser 18 points caractéristiques du visage qui sont : les deux Coins de la bouche, le sommet et le plus bas point des courbes représentant les lèvres supérieur et inférieur, les deux coins des yeux, le sommet et le plus bas point des courbes représentant les deux paupières supérieure et inférieure des yeux, les deux centres des iris et enfin les coins des sourcils.

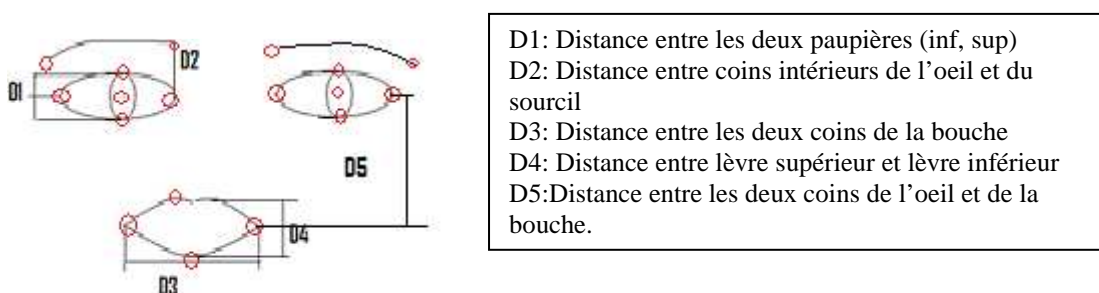


Figure 2.5. Les points caractéristiques du visage et les distances biométriques.

La détection de ces 18 points permet de calculer certaines distances faciales ainsi que la localisation des régions d'intérêts ou certains traits transitoires peuvent apparaître.

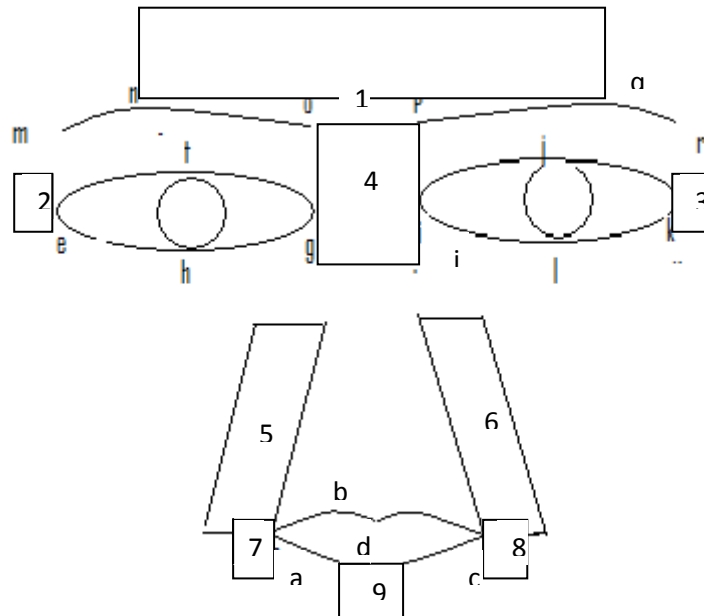


Figure 2.6. Détection des traits permanents et localisation des régions d'intérêts.

Avant de positionner et calculer la taille de toutes les zones, un facteur considérant les dimensions du visage est estimé :

$$\text{Fact} = ((k-e)-(c-a))/3$$

Eq 2.1

Ce facteur est calculé afin de donner la bonne taille aux zones de rides quelle que soit les dimensions de l'image étudiée.

- a- Zone d'intérêt du front (1): Cette zone est située juste au dessus des sourcils, elle est délimitée par les colonnes des deux points (n,q) qui représentent le sommet de la courbe modélisant les sourcils gauche et droit en largeur et par la ligne du point (n) moins (fact*2) et la ligne de ce même point en hauteur.
- b- Zones d'intérêts des coins extérieurs des yeux (2,3): La zone gauche est délimitée par la colonne du coin gauche de l'œil gauche (e) et cette même colonne plus (fact*2)/3 en largeur et la ligne de ce même point +/- (fact*2/3) en longueur. La taille et la position de La zone droite sont calculées de la même façon.

- c- Zone d'intérêt entre les yeux (4): cette zone est située entre les deux yeux sa largeur correspond à la distance entre les deux coins intérieurs des yeux (g,i) et sa hauteur correspond à la ligne du coin intérieur du sourcil (o ou p) et la ligne du coin de l'œil(g ou i) plus fact.
- d- Zone nasolabiale (5,6) : la zone gauche est délimitée par la colonne du coin intérieur de l'œil gauche(g) moins (fact+(fatc/2)) et la colonne même en largeur et par la ligne du point (g)+ (fact+(fatc/2)) et la ligne du coin gauche de la bouche (a). La taille et la position de La zone droite sont calculées de la même façon.
- e- Zones des coins de la bouche (7,8): la taille et la position de ces zones sont calculées de la même façon que les zones des coins extérieurs des yeux mais en partant du coin de la bouche(a,c).
- f- La zone du menton (9): Est délimitée par la colonne du point inférieur de la lèvre inférieur(d) +/-fact en largeur et par la ligne de ce point et cette même ligne plus fact*2 en hauteur.

Les traits transitoires peuvent également apparaître sur deux autres régions du visage, au dessus et au dessous des paupières supérieure et inférieure, hors ces deux zones ne seront pas considérées dans cette étude pour les raisons suivantes : les rides au dessus des paupières sont toujours confondues avec le contour des paupières ou bien avec les sourcils, et les rides au dessous des paupières sont toujours associées aux rides nasolabiales, donc l'étude de ces derniers est largement suffisante pour tirer le maximum d'informations depuis cette zone du visage.

Une fois les régions d'intérêts localisées, le détecteur de contour de « Canny » [LYO99] ainsi qu'une étape de seuillage [YON97] sont appliqués afin de déterminer la présence ou l'absence des traits transitoires sur chacune de ces régions.

2.4.2. Extraction de Données

Dans cette étape, seules les informations les plus pertinentes dans la découverte de nouvelles connaissances sont extraites depuis les résultats obtenus lors de la phase de segmentation.

Deux types de données sont extraits : des données issues des traits permanents et des données issues des traits transitoires. Concernant les données issues des traits permanents,

cinq distances biométriques sont calculées depuis les 18 points caractéristiques du visage (Figure 2.5). Ces distances représentent : Le degré de froncement ou d'élévation des sourcils, le degré d'ouverture ou fermeture des yeux, le degré d'ouverture ou fermeture (Horizontale et verticale) de la bouche et la distance entre les yeux et la bouche. Ces distances sont normalisées par rapport à la distance fixe entre les deux iris. L'objectif de cette normalisation est de rendre l'analyse indépendante par rapport à la variabilité des dimensions des visages et à la position des visages par rapport à la caméra.

Quant aux données qui peuvent être extraites des traits transitoires, on calcule le nombre de pixels de contour dans chaque région faciale d'intérêt avec expression et on le compare à celui calculé dans la même région du visage à l'état neutre, si la différence dépasse un seuil (Thigh), les rides sont supposées présentes et un état « Présent » est assigné à cette zone, si la différence est inférieure à un seuil (Tlow), les rides sont supposées absentes et un état « Absent » est assigné à cette zone, si la différence est entre (Thigh) et (Tlow), on suppose qu'il existe un doute dans la présence de ce type de traits sur la zone considérée et l'état « Présent OU Absent » lui est assigné.

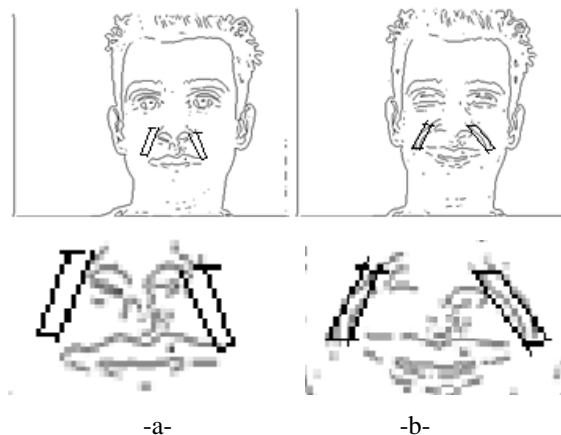


Figure 2.7. (a): Visage à l'état neutre sans la présence de traits transitoires ;
 (b): Visage avec expression (de joie) avec la présence de traits transitoire sur la zone nasolabiale.

Une autre information véhiculée par les traits transitoires concerne les traits transitoires de la zone nasolabiale, où l'angle formé entre la droite approximant ces rides et la ligne horizontale joignant les deux coins de la bouche est considérée. Une fois cet angle est calculé, il sera normalisé par rapport à l'angle 90° . Cet angle représente l'angle maximal qui peut être atteint avec une expression de colère ou de dégoût.

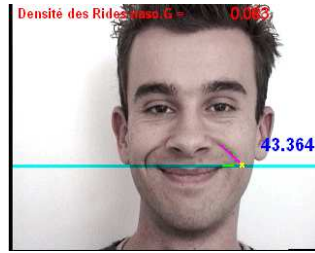


Figure 2.8. Détection et calcul de l'angle nasolabial.

2.4.3. Analyse

Différents types de traits transitoires peuvent apparaître lors de la production d'une expression faciale quelconque. Dans cette phase d'analyse, l'objectif est d'étudier la corrélation entre traits transitoires et expressions faciales, il est donc question d'associer la présence (ou absence) de certaines rides faciales à certaines expressions faciales. Pour mener cette étude à terme, nous avons effectué un apprentissage sur un ensemble d'images issues de différentes bases d'images faciales. Les images considérées sont : 90 images de la base Hammal_Caplier [HAM_Base], 462 images de la base EeBase [Ee_Base] et enfin 144 images de la base DAFex [DAF_Base].

En utilisant le détecteur de contours de « Canny », tous les traits transitoires sont détectés sur les différentes régions d'intérêts du visage pour chacune des expressions universelles. Des taux correspondants à la présence des rides sur chacune des zones et pour chacune des expressions sont calculés, ensuite les résultats numériques obtenus sont transformés en résultats logiques. La table 2.2 résume les résultats finaux :

Traits Transitoires	Joie	Surprise	Dégout	Colère	Tristesse	Peur
Menton	(0)	(0)	(0)	(1U0)	(1U0)	(1U0)
Coins bouche	(0)	(0)	(0)	(0)	(1U0)	(1U0)
Zones Nasolabiales	(1U0)	(0)	(1U0)	(1U0)	(1U0)	(1U0)
Coins des yeux	(1U0)	(0)	(1U0)	(1U0)	(0)	(0)
Région Nasal	(0)	(0)	(1U0)	(1U0)	(1U0)	(1U0)
Front	(0)	(1U0)	(0)	(1U0)	(1U0)	(1U0)

Table 2.2. Présence or absence de TTS sur chaque région d'intérêt et pour chaque expression faciale.

La table 2.2 donne les règles logiques concernant la présence ou l'absence des TTs sur chaque région d'intérêt et pour chaque expression faciale, elle présente en ligne les différentes

régions d'intérêts où les TTs peuvent apparaître, et en colonne les différentes expressions faciales considérées. La valeur «1» signifie qu'il peut y avoir des TTs sur une région considérée pour une expression donnée, la valeur «0» signifie qu'il ne peut y avoir de TTs sur une région considérée pour une expression donnée. Par exemple, dans le cas de la surprise, il ne peut y avoir de TTs que sur le front. Cette étude a permis la généralisation de ces règles logiques à travers toutes les bases d'images faciales.

Cette table permet également de caractériser chaque région d'intérêt par une combinaison d'expressions faciales avec lesquelles ces TTs peuvent apparaître. Par exemple, la présence des rides sur le menton veut dire qu'il peut s'agir de la colère, de la tristesse ou de la peur. En d'autres termes cette étude a permis d'associer l'apparition de certaines rides sur le visage à un certain ensemble d'expressions faciales et d'associer l'absence de certaines rides à d'autres expressions.

Une autre information concernant les TTS autre que leur présence ou absence, est l'angle nasolabial calculé dans la section (2.4.2).

Selon Ekman [EKM02], L'angle nasolabial formé avec la colère ou le dégoût est dû à l'activation des unités d'actions AU9 ou AU10, par contre l'angle formé avec la joie est dû à l'activation de l'unité d'action AU12, par conséquent l'angle formé avec la colère ou le dégoût est plus grand que celui formé avec la joie.



Figure 2.9. Angle nasolabial formé de gauche à droite: Colère (72.6°), Dégoût (71.2°) et Joie (43.4°).(Eebase, H_Caplier databases)

Afin d'extraire de nouvelles informations, nous avons mesuré l'angle nasolabial formé de 144 images issues des deux bases Eebase [EE_Base] et Dafex[DAF_Base] et présentant les cinq expressions faciales avec lesquelles l'angle nasolabial peut apparaître qui sont : Colère, Peur, Dégoût, Tristesse et Joie.

Les angles calculés sont ensuite normalisés par rapport à l'angle 90° (Angle qui peut être atteint lors de l'activation de l'unité d'action AU9 ou AU10 par exemple).

Pour chaque expression nous avons estimé les seuils correspondant aux valeurs minimales et maximales. Les limites minimales correspondent à la moyenne des valeurs minimales et les limites maximales correspondent aux valeurs maximales. La Figure 2.10 présente la variation des l'angle nasolabial formé avec les cinq expressions.

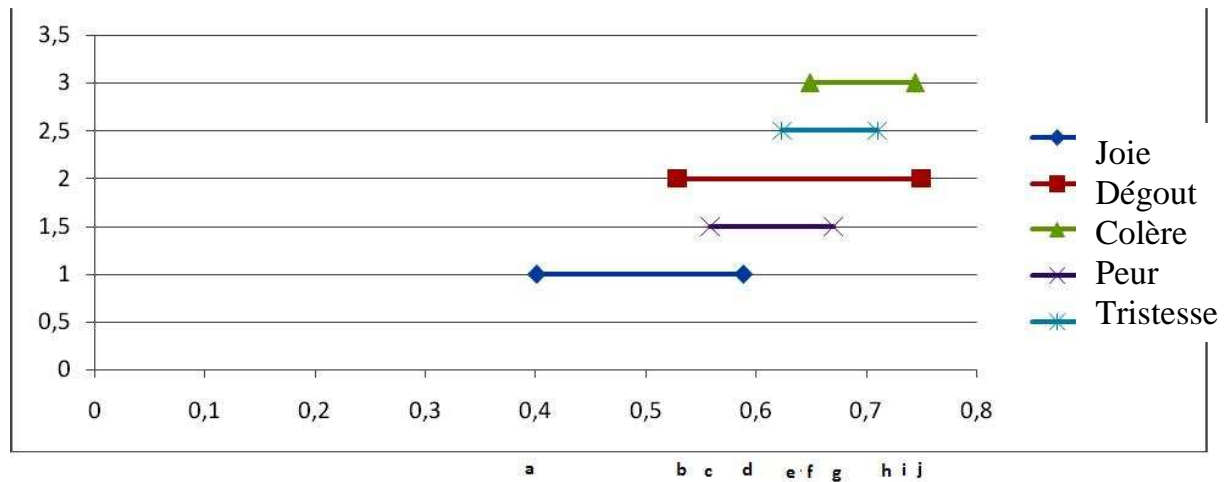


Figure 2.10. Variation de l'angle nasolabial formé avec les Cinq expressions faciales.

a : Seuil minimal de l'angle formé lors de la joie.

b : Seuil minimal de l'angle formé lors du dégoût.

c : Seuil minimal de l'angle formé lors de la Peur.

d : Seuil maximal de l'angle formé lors de la joie.

e : Seuil minimal de l'angle formé lors de la tristesse.

f : Seuil minimal de l'angle formé lors de la colère.

g : Seuil maximal de l'angle formé lors de la Peur.

h : Seuil maximal de l'angle formé lors de la tristesse.

i : Seuil maximal de l'angle formé lors de la colère.

j : Seuil maximal de l'angle formé lors du dégoût.

On peut constater que l'angle formé avec l'expression de joie est toujours plus petit que celui formé avec les autres expressions et d'un autre coté on peut également constater que l'angle formé avec les autres expressions est souvent équivalent (dans le même intervalle).

Comme nous sommes en présence de plusieurs régions d'intérêts et de différents types de données, il est indispensable de fusionner toutes ces informations afin de prendre une décision sur la catégorie de l'expression faciale. Pour pouvoir réaliser ceci, nous avons utilisé le modèle de croyance ou théorie de l'évidence.

2.4.4. Classification

2.4.4.1 Principe de la Théorie de l'Evidence (Modèle de Croyance)

Les concepts de base de la théorie de l'évidence trouvent leur origine au XVII^e siècle déjà, dans les travaux du mathématicien suisse Jacob Bernoulli (1654-1705), puis dans ceux de Johann Heinrich Lambert (1728-1777). Cependant, ce n'est qu'en 1967 qu'Arthur Dempster posa les fondements de ce que l'on appelle la théorie de l'évidence, ou encore la théorie de Dempster et Shafer, Shafer ayant été le collaborateur de Dempster.

Une dizaine d'années plus tard, la théorie mathématique de l'évidence vit le jour par la publication de Glenn Shafer en 1976. Celui-ci reconnut en 1979 que les travaux de Bernoulli et Lambert peuvent être considérés comme précurseurs de sa théorie de l'évidence.

Le Modèle des Croyances Transférables (TBM : Transferable Belief Model) est un cadre formel générique développé par Ph. Smets [SME94] pour la représentation et la combinaison des connaissances. La TBM est basée sur la définition de fonctions de croyance fournies par des sources d'information pouvant être complémentaires, redondantes et éventuellement non-indépendantes. Cette théorie propose un ensemble d'opérateurs permettant de combiner ces fonctions. Elle est donc naturellement employée dans le cadre de la fusion d'informations pour améliorer l'analyse et l'interprétation de données issues de sources d'informations multiples.

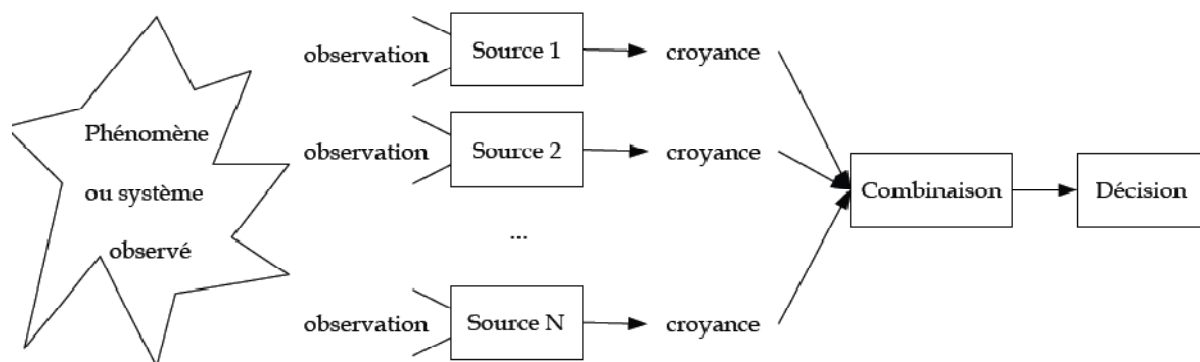


Figure 2.11 Fusion d'information

Des sources de croyance émettent une opinion pondérée sur l'état du système observé. Les différentes sources sont combinées afin de synthétiser la compréhension du système et améliorer la décision sur l'état du système après combinaison.

La théorie de Dempster-Shafer (DS) a été introduite par Dempster et formalisée par Shafer [DEM68, SHA76]. Elle représente à la fois l'imprécision et l'incertitude à l'aide de fonctions de masse m , de plausibilité Pl et de croyance Bel . Cette théorie se décompose en trois étapes : la définition des fonctions de masse, la combinaison d'informations et la décision.

2.4.4.2. Définition des Fonctions de Masse

L'ensemble des hypothèses pour une source (typiquement une classe dans un problème de classification multi source) est défini sur l'espace $\Omega = \{A_1, A_2, \dots, A_k, \dots, A_N\}$ appelé cadre ou espace de discernement où A_k désigne une hypothèse en faveur de laquelle une décision peut être prise.

Les fonctions de masse sont définies sur tous les sous ensembles de l'espace Ω et non seulement sur les singletons comme dans les probabilités.

Une fonction de masse m est définie comme une fonction de 2^Ω dans $[0,1]$. En général on impose $m(\emptyset) = 0$ et une normalisation de la forme :

$$\sum_{A \subseteq \Omega} m(A) = 1 \quad \text{Eq 2.2}$$

Une fonction de croyance Bel est une fonction totalement croissante de 2^Ω dans $[0,1]$ définie par :

$$\forall A_1 \in 2^\Omega, \dots, A_K \in 2^\Omega, \quad Bel(\cup_{i=1 \dots K} A_i) \geq \sum_{I \subseteq \{1 \dots K\}, I \neq \emptyset} (-1)^{|I|+1} Bel(\cap_{i \in I} A_i) \quad \text{Eq 2.3}$$

où $|I|$ désigne le cardinal de I et $Bel(\emptyset) = 0, Bel(\Omega) = 1$.

Etant donné une fonction de masse m , la fonction Bel définie par :

$$\forall A \in 2^\Omega, Bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B) \quad \text{Eq 2.4}$$

est une fonction de croyance.

Inversement, à partir d'une fonction de croyance Bel, on peut définir une fonction de masse m par :

$$\forall A \in 2^\Omega, m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B) \quad \text{Eq 2.5}$$

Une fonction de Plausibilité Pl est également une fonction de 2^Ω dans $[0,1]$ définie par :

$$\forall A \in 2^\Omega, Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad \text{Eq 2.6}$$

La plausibilité mesure la confiance maximum que l'on peut avoir en A. La possibilité d'affecter des masses aux hypothèses composées et donc de travailler sur 2^Ω plutôt que sur Ω constitue un des avantages de cette théorie. Elle permet une modélisation très riche et très souple, en particulier de l'ambiguïté ou de l'hésitation entre classes.

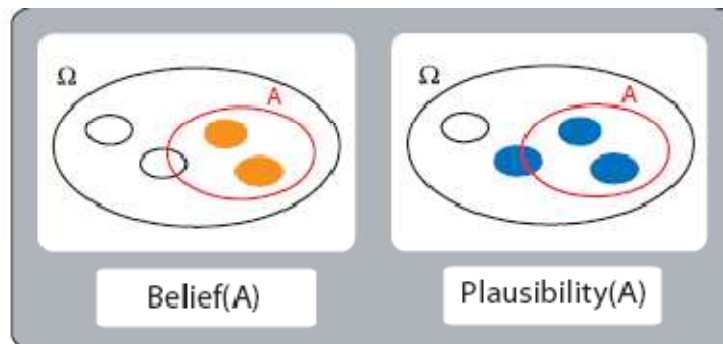


Figure 2.12. Exemple de processus de décision de plausibilité et croyances

2.4.4.3. Combinaison Des Masses d'Evidence

En présence de plusieurs capteurs ou de plusieurs informations provenant d'un même capteur, il devient intéressant de combiner les connaissances de chaque source pour en extraire une connaissance globale afin d'améliorer la prise de décision. Dans la théorie de DS, les masses sont combinées par la somme orthogonale de Dempster.

Soit m_j la fonction de masse associée à la source j , pour un sous-ensemble A de Ω on obtient :

$$(m_1 \oplus \dots \oplus m_l)(A) = \frac{\sum_{B_1 \cap \dots \cap B_l = A} m_1(B_1) \dots m_l(B_l)}{1 - \sum_{B_1 \cap \dots \cap B_l = \emptyset} m_1(B_1) \dots m_l(B_l)} \quad \text{Eq 2.7}$$

Ce type de combinaison qui n'est pas idempotente suppose l'indépendance cognitive des sources plutôt que l'indépendance statistique.

Le mode de combinaison disjonctif est aussi possible en remplaçant l'intersection dans la formule (1) par une opération ensembliste :

$$(m_1 \oplus_{\cup} \dots \oplus_{\cup} m_l)(A) = \sum_{B_1 \cup \dots \cup B_l = A} m_1(B_1) \dots m_l(B_l) \quad \text{Eq 2.8}$$

2.4.4.4. Processus de Décision

Contrairement à la théorie Bayésienne où le critère de décision est très souvent le maximum de vraisemblance, la théorie de l'évidence propose de nombreuses règles. Les plus utilisées sont le maximum de crédibilité, le maximum de plausibilité, les règles basées sur l'intervalle de confiance, le maximum de probabilité pignistique [SME00] et la décision par maximum de vraisemblance.

2.4.4.5 Application de la TBM dans le Contexte de Classification Catégorielle des Expressions Faciales

Dans notre cas, le cadre de discernement est représenté par l'ensemble $\Omega = \{\text{Joie, Colère, Dégout, Tristesse, Peur, Surprise, Neutre}\}$.

Neuf régions faciales sont concernées par la présence ou l'absence des traits transitoires, c'est pourquoi un modèle est utilisé afin de calculer la masse d'évidence concernant la présence ou l'absence des TTs sur chaque région. Le modèle est présenté sur la figure 2.13:

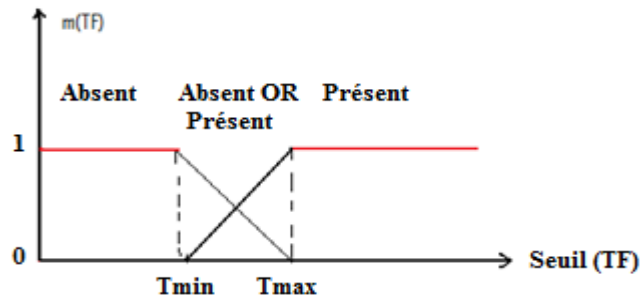


Figure 2.13. Modèle proposé pour la présence et absence des TTS.

Si le rapport calculé à partir du nombre de pixels est inférieur au seuil T_{min} (trouvé par apprentissage), les TTs sont supposés absent avec une masse d'évidence égale à « 1 », si le rapport est supérieur au seuil T_{max} (trouvé par apprentissage), les TTs sont supposés présent avec une masse d'évidence égale à « 1 », sinon, il y a un doute dans la présence ou l'absence des TTS avec une masse d'évidence entre $[0,1]$ calculable depuis le modèle proposé.

Les Rides nasolabiales (s'ils existent) sont caractérisées par une autre caractéristique qui est l'angle nasolabial ; Un autre modèle est également proposé afin de calculer la masse d'évidence concernant le seuil atteint par cet angle et la déduction de l'expression correspondante. Le modèle est présenté sur la figure 2.14:

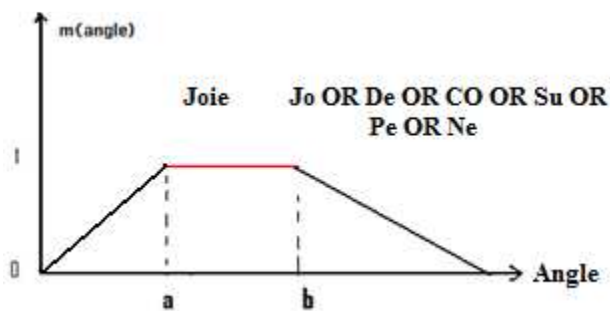


Figure 2.14. Modèle proposé pour l'angle nasolabial calculé.

a, b sont les seuils utilisés dans la figure 2.10, Si l'angle nasolabial calculé se trouve dans l'intervalle $[a,b]$, l'expression étudiée est l'expression de joie avec une masse d'évidence égale à « 1 », sinon elle peut être n'importe quelle autre expression c.a.d : Dégout, colère, Tristesse, Peur, Surprise ou Neutre avec une masse d'évidence entre $[0,1]$ calculable depuis le modèle proposé.

Pour être plus explicite, on suppose que nous avons trois régions faciales sur lesquelles des TTS apparaissent, la région nasale et les deux régions nasolabiales. Les états et masses d'évidence associés à ces faits sont :

$$- m(\text{région_nasal})(\text{présent})=1$$

$$- m(\text{région_naso})(\text{présent})=1$$

La caractérisation de chaque région par une combinaison des expressions faciales correspondantes en utilisant la table 2.2 est comme suit :

$$- m(\text{region_nasal})=m(\text{Dégout OR Colère OR Tristesse OR Peur})=1$$

$$- m(\text{region_naso})=m(\text{Joie OR Dégout OR colère OR tristesse OR Peur})=1$$

Le processus de fusion donne:

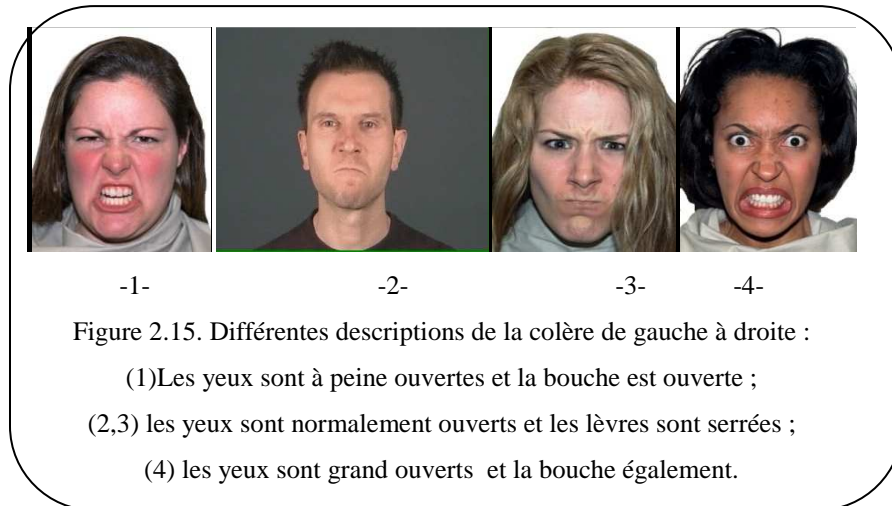
$$\Rightarrow m(\text{region_nasal};\text{région_naso})(\text{Dégout OR Colère OR Tristesse OR Peur}) = m(\text{Dégout OR colère OR tristesse OR peur}). m(\text{joie OR Dégout OR Colère OR Tristesse OR Peur}) = 1$$

Ceci veut dire que si des TTs sont présents sur la région nasale et les régions nasolabiales, l'expression correspondante pourrait être une des expressions suivantes : colère, Dégout, Tristesse, ou Peur.

Le résultat de la fusion de toutes les données disponible donne alors la classe de l'expression étudiée ou bien un doute entre un ensemble d'expressions. Dans le cas ou aucun de ces types d'information est détecté, toute cette partie d'étude sera ignorée. Le résultat sera le doute entre l'ensemble des éléments du cadre de discernement. Dans les deux cas, afin de réduire ou même éliminer ce doute une dernière étape est proposée.

2.4.5 .Post Traitement

Dans cette étape un autre type de trait est considéré. Les yeux, les sourcils et la bouche représentent des traits permanents du visage. Cette caractéristique de permanence permet de classifier une expression étudiée quelque soit les déformations. Afin de discriminer les différentes expressions résultantes, nous avons étudié les différentes descriptions proposées dans la littérature [TEK99], [HAM05], [CAR97], [EIB89], [PAR00], [TSA00]. Dans certaines descriptions la colère est définie par des lèvres serrées et dans d'autres les dents sont serrées et les lèvres étirées (la bouche est ouverte) (figure 2.15).



Dans le but d'unifier ces différentes descriptions, nous avons proposé une nouvelle description qui prend en considération toutes ces descriptions. Par exemple la colère est définie de deux façons : lèvres serrées ou lèvres étirées. Dans ce qui suit nous proposons la nouvelle description des différentes expressions faciales.

	Distance entre paupières	Distance entre œil et sourcil	Distance entre les coins de la bouche	Distance entre lèvre supérieure et lèvre inférieure	Distance les coins de l'œil et de la bouche
Joie	accroît ou décroît	accroît, ne change pas ou décroît	accroît	ne change pas ou accroît	décroît
Surprise	accroît	accroît	ne change pas ou décroît	accroît	ne change pas ou accroît
Dégout	décroît	décroît	accroît, ne change pas ou décroît	accroît	accroît, ne change pas ou décroît
Colère	décroît ou accroît	décroît	ne change pas ou décroît	accroît, ne change pas ou décroît	ne change pas ou accroît
Tristesse	décroît	accroît	ne change pas ou accroît	ne change pas ou accroît	ne change pas ou décroît
Peur	ne change pas ou accroît	ne change pas ou accroît	accroît, ne change pas ou décroît	ne change pas ou accroît	accroît, ne change pas ou décroît

Table 2.3. Nouvelle description des six expressions faciales.

Ensuite chaque deux expressions sont comparées en termes de distances afin de déduire les différences potentielles entre expressions. Par exemple, en comparant la joie avec la surprise et la colère, on constate qu'avec la joie la bouche est ouverte horizontalement, alors qu'avec la surprise ou la colère la bouche est ouverte verticalement. En comparant la joie avec les autres expressions, nous pouvons constater que les distances peuvent évoluer dans le même sens (Décroissent ou accroissent), c'est pourquoi le doute entre la joie et ce qui reste comme expressions ne peut être enlevé.

Expressions	Différences
Joie /surprise	Joie: D3 ↗ ; Surprise: D3 ↘ ou =
Joie /Colère	Joie: D3↗ ; Colère: D3= ↘ ou =
Surprise / Dégout	Surprise: D1 ↗ ; Dégout: D1 ↘ Surprise: D2↗ ; Dégout: D2 ↘
Surprise / Colère	Surprise: D2 ↗ ; Colère: D2 ↘
Surprise / Tristesse	Surprise: D1 ↗ ; Tristesse: D1 ↘
Dégout / Tristesse	Dégout: D2 ↘ ; Tristesse: D2 ↗
Dégout / Peur	Dégout: D1 ↘ ; Peur: D1 ↗ ou = Dégout: D2 ↘ ; Peur: D2 ↗ ou =
Colère / STristesse	Colère: D2 ↘ ; Tristesse: D2 ↗
Colère / Peur	Colère: D2 ↘ ; Peur: D2 ↗ ou =
Tristesse / Peur	Tristesse: D1 ↘ ; Peur: D1 ↗ ou =

Table 2.4. Différences potentielles entre les six expressions universelles.

Cette table présente les différences potentielles entre les six expressions universelles:

- Distance entre les coins de la bouche (D3 peut décroître= ↘ ou accroître=↗),
- Distance entre paupières (D1) et
- Distance entre les coins de l'œil et du sourcil (D2).

Cette table est utilisée pour réduire ou éliminer le doute entre expressions résultantes.

On reprend maintenant l'exemple cité dans la section classification, le résultat de la fusion était Colère OU Dégout OU Tristesse OU Peur.

Le post traitement permet de réduire le doute entre ces quatre expressions. En effet en utilisant la table des différences potentielles, on peut facilement exclure la tristesse car la différence entre

colère et tristesse est dans la distance entre les coins de l'œil et du sourcil, cette distance décroît avec la colère et accroît avec la tristesse.

On compare ensuite la colère avec la peur, on trouve que la différence entre ces deux expressions est également dans la distance entre l'œil et le sourcil, cette distance décroît avec la colère et accroît avec la peur. Il ne reste plus que les deux expressions de Dégout et colère, suivant la table des différences, on peut constater que toutes les distances considérées peuvent évoluer dans le même sens pour ces deux expressions, donc on ne peut les différencier et le résultats final sera donc Dégout OU Colère au lieu de Colère OU Dégout OU Tristesse OU Peur.

2.5. Résultats Obtenus

Afin d'évaluer l'approche proposée, nous avons testé toutes les images de la base Dafex [DAF base] (78 images avec TTs). Les résultats de classification obtenus sont présentés sur la table 2.5:

EXPERT / SYSTEM	Joie_ moy	Joie_ max	Dég moy	Dég max	Col moy	Col_ max	Tri moy	Tri max	Peur_ moy	Peur_ max	Sur Moy	Sur max
Joy	50%	87,5 %										
Dégout												
Colère												
Tristesse							100	100				
Peur									66,67%	42,86%		
Surprise												
Joie OR Deg	37,5%											
Col OR Deg			100	100	100	85,7%						
Peu OR Sur									33,33%	57,14%	100	100
Erreur	12,5%	12,5 %				14,3%						
Total Reconnu	87,5	87,5	100	100	100	85,7	100	100	100	100	100	100
Total	100	100	100	100	100	100	100	100	100	100	100	100

Table 2.5. Classification des expressions faciales basée sur les trait transitoires avec un post traitement sur la base des traits permanents.

A partir de cette table, on peut constater que la joie peut être confondue avec le dégoût surtout quand l'intensité de l'expression est moyenne ou minimale, le dégoût peut être

confondu avec la colère et la surprise avec la peur. La tristesse est la seule expression qui est reconnue sans aucun doute.

Nous avons également testé une autre base (EEbase) afin de discriminer la colère et le dégoût, un taux important a été trouvé concernant la reconnaissance de la colère sans aucun doute. Ceci peut être expliqué par le fait que cette expression de colère peut être exprimée de différentes façons comme nous l'avons montré avant.

2.6. Comparaison avec autre Système

Afin d'évaluer les performances de notre approche (basée sur les traits transitoires), nous l'avons comparé à une autre approche basée sur les traits permanents [Ham05]

	Hammal et al Approche/ CKE base d'images	Notre approche/ Dafex base d'images
Joie	64,51%	50% Si intensité = Moyenne 87,5% Si intensité = maximale
Joie OU Dégout	32.27%	37.5% Si intensité =Moyenne
Surprise OU Peur	84%	100%

Table 2.6. Comparaison des Approches

La table 2.6 résume les résultats obtenus par chacune des deux approches. Les deux systèmes utilisent la théorie de l'évidence comme outil de fusion des données utilisées, Notre approche se base sur des données issues principalement sur les traits transitoires quand à l'approche présenté dans [HAM05] est basée principalement sur les traits permanents. Comme cette dernière approche n'a été testée que sur trois expressions faciales qui sont : La Joie, Le dégoût et la Surprise, nous avons considéré la différence en ce qui concerne ces trois expressions.

Pour les résultats de la joie, les taux obtenus sont comparables. Quand l'intensité est maximale, les meilleures performances sont données par notre système de classification. D'un autre coté les deux systèmes introduisent des sources de doute qui peut exister entre la joie et le dégoût et un autre doute entre la surprise et la peur. Les taux obtenus pour ces deux confusions sont pratiquement les mêmes.

Ces résultats ont apporté l'évidence convergente pour la similitude des taux de classification en considérant soit les traits permanents soit les traits transitoires.

Une autre source de confusion est présentée par notre système de classification: le doute entre le dégoût et la colère, cette source n'apparaît pas avec l'autre système, puisque la colère n'est pas évaluée dans le système référencé.

2.7. Conclusion

La reconnaissance automatique des expressions faciales basée sur les traits transitoires est un nouveau axe de recherche. Les résultats obtenus ont prouvé que cette approche peut être une base admise pour la reconnaissance des expressions faciales. Même si souvent le doute peut exister entre deux expressions, nous estimons qu'il est préférable de préserver le doute que de prendre le risque de donner une mauvaise classification. C'est pour cette raison que nous avons préféré utilisé la Théorie de l'évidence car cette dernière modélise parfaitement le doute en plus c'est un outil de fusion très puissant quand il s'agit de données issues de plusieurs sources.

La comparaison présentée indique parfaitement que la reconnaissance basée traits transitoires réalise des résultats aussi bien que la reconnaissance basée traits permanents. La combinaison des deux approches pourrait donner dans le futur des résultats optimaux.

Dans le chapitre suivant un autre type de classification des expressions faciales est présenté. Comme il n'est pas toujours possible de reconnaître toutes les expressions faciales (plus de 200 000 expressions) et que souvent une expression est confondue avec une autre expression, nous proposerons une classification dimensionnelle d'expressions, la dimension explorée est la dimension de valence.

Chapitre 3

Classification Dimensionnelle des Expressions Faciales

3.1 Introduction

La classification des expressions faciales peut être réalisée généralement de deux façons : classification dimensionnelle ou classification catégorielle. La classification catégorielle repose sur quelques catégories émotionnelles de base tel que peur, colère, joie, tristesse, etc. qui peuvent se combiner pour rendre compte de la grande variété d'expressions. Il y a dans ce cas peu de confusion dans une tâche de catégorisation d'expressions émotionnelles (Ekman & Frazen). La majorité des auteurs est actuellement en faveur de cette dernière hypothèse. Le chapitre II a fait l'objet de la présentation d'une approche dans cette hypothèse.

La classification dimensionnelle propose de représenter les émotions dans un espace multidimensionnel. Les dimensions peuvent être un axe de plaisir et de déplaisir, d'éveil ou d'ennui, de nervosité, de puissance, de maîtrise de soi ou de bien d'autres dimensions. De nos jours à l'instar de Russel, la plupart des chercheurs qui s'inscrivent dans cette approche s'accordent les deux premières dimensions : La valence et l'activation ou l'arousal. Russel estime que l'émotion est mieux caractérisée par une approche dimensionnelle bipolaire, plutôt qu'avec un petit nombre de catégories d'émotions discrètes. La valence permet de distinguer les émotions positives, agréables comme la joie, des émotions négatives, désagréables comme la colère. L'activation représente le niveau d'activation corporelle qui transparait par un nombre de réactions physiologiques comme l'accélération du cœur et la transpiration. Par exemple, la tristesse a une activation (excitation) faible, tandis que la surprise a un niveau d'activation (excitation) élevé.

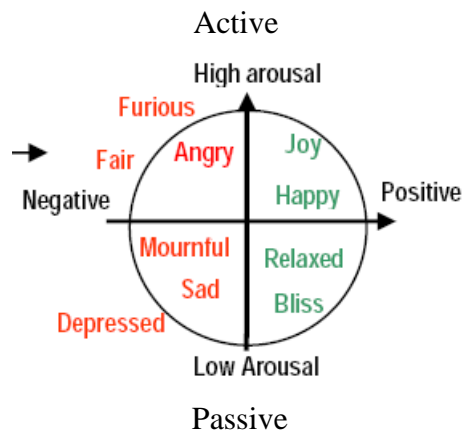


Figure 3.1. Dimension bipolaire : valence et activation proposé par Russell [RUS94].

A partir du moment où les recherches dans ce domaine ont commencé, la plupart des approches proposées dans la littérature s'intéressaient à classer une expression donnée dans l'une des six émotions universelles postulée par Ekman à savoir: la joie, dégoût, colère, tristesse, surprise et de crainte. Durant les années 90, Ekman élargi cette liste des émotions de base [EKM99]. Les émotions nouvellement incluses sont : Amusement, Mépris, Contentement, Embarras, Excitation, Culpabilité, Fier de sa réussite, Soulagement, Satisfaction, Plaisir sensoriel et Honte.

Une émotion donnée de cet ensemble peut être visuellement exprimée par différentes expressions faciales (différentes déformations des traits du visage). Ainsi, une telle approche discrète souffre de la rigidité (avec son one-to-one mapping). Un autre problème est que une expression faciale peut être complexe (composé de plusieurs catégories) et le choix d'une seule catégorie peut-être trop restrictive. L'approche dimensionnelle élimine ces restrictions, donnant un éventail de messages affectifs.

Dans ce chapitre, nous nous sommes intéressés à la dimension de valence afin de savoir si une personne est de bonne humeur ou de mauvaise humeur. Par conséquent, nous considérons le problème de la reconnaissance des expressions positives (Attractions, Contentement, Excitation, Satisfaction, Plaisir sensoriel, Fier de sa réussite, Heureux, Intérêt) et la reconnaissance des expressions négatives (Outrage, Embarras, Culpabilité, Tristesse, Colère, Peur, Ennui, Amertume, Anxiété et Terreur). Pour des raisons de non disponibilité de bases d'images contenant ces différentes expressions positives et négatives, notre contribution a été évaluée uniquement sur les six expressions faciales universelles, plus le neutre. Ces expressions sont classées en trois catégories, à savoir positive, négative ou neutre. Colère,

Dégoût, Tristesse, Peur et Surprise négative sont considérées comme des expressions négatives, et Joie et Surprise positive sont considérées comme des expressions positives. En effet, la Surprise est une expression particulière, car elle peut être positive ou négative en fonction de la situation.

La motivation de cette contribution est que l'étude de l'espace de valence est très importante tant dans la recherche affective (psychologie [MEH96], psychiatrie, science du comportement, neurosciences) et dans l'animation de personnages virtuels [GEB05], [KSH02], [BEC06].

En outre, dans de nombreuses interactions homme-machine, on ne s'intéresse pas toujours à reconnaître l'expression spécifique du visage. En effet on s'intéresse souvent à détecter l'état d'esprit de l'interlocuteur dans le but d'induire des réactions spécifiques. En raison de l'importance de l'état affectif dans l'interaction homme machine, la détection automatique de l'humeur a suscité l'intérêt de nombreux chercheurs. Ici, nous examinons brièvement des travaux antérieurs afin de mettre notre travail dans le contexte.

3.2. Etat de l'Art

Des résultats concrets sur la reconnaissance des expressions faciales ont montré que les taux de classification sont significativement plus élevés lorsqu'on considère des catégories d'expressions plus générales comme les catégories « Positive », « Négative » et « Neutre » que lorsqu'on considère des catégories basiques (fines) comme « Joie », « Colère », « Dégout » etc....

Dans [COH02], Les auteurs ont proposé une méthode de reconnaissance des émotions positives et négatives depuis des séquences vidéo. Ils introduisent des classificateurs du type (Tree-Augmented-Naïve Bayes) qui apprennent les dépendances entre les composants du visage et fournissent un algorithme pour trouver la meilleure structure des classificateurs. Le system peut prévoir avec une précision de 77% si une personne affiche une expression positive ou négative. Ce programme a été testé sur une petite base d'images collectées par Chen [CHE00], elle compte cinq sujets.

Dans [SEB02], les auteurs présentent une méthode de classification des expressions depuis une séquence vidéo en utilisant un classificateur naïf de Bayes (naive Bayes classifier). L'hypothèse communément admise est que le modèle de distribution est gaussien. Toutefois, ils ont réussi à utiliser l'hypothèse de distribution de Cauchy. Ils obtiennent des taux de 81%

(pour les expressions positives) et 79% (pour les expressions négatives) de classification correcte. L'essai de leur algorithme a été effectué sur une base de données de cinq sujets.

Dans [MIU02], Miura propose une méthode qui classe une expression faciale comme positive, négative ou neutre depuis des séquences d'images. La vitesse de filtrage de l'espace (space-filtering velocity) est premièrement employée pour détecter la vitesse de déplacement des objets. Ensuite l'estimation du flux optique est utilisée pour détecter le déplacement des objets et suivre les régions de la bouche et des yeux. Leur programme a été testé sur une base d'images de huit sujets. Ils obtiennent 60.9% pour une classification correcte au début de leur expérience. Par contre, Les estimations du flux optique sont facilement brouillées par les déplacements non rigides et la variation de l'éclairage, et sont sensibles à l'inexactitude d'enregistrement de l'image et de la discontinuité du mouvement.

D'autres travaux se sont intéressés à la classification des unités d'actions (AUS) [RUC93], [FRA93], [SAY95] et [SAY]. Par exemple, dans [SAY01], les auteurs codent un ensemble d'unités d'actions comme positives ou négatives. La fiabilité des combinaisons des unités positives et négatives est bonne à excellente. Le problème avec ces méthodes est que plusieurs unités d'actions ne sont pas codées et plusieurs expressions faciales sont produites par l'activation d'unités d'actions non codées.

Une des difficultés rencontrée en travaillant dans la reconnaissance des expressions faciales est le manque de bases d'images référentielles universelles afin de comparer les différents algorithmes, c'est pourquoi nos études ont été menées sur différentes bases. Comme dans la classification catégorielle, une limite des méthodes de classification dimensionnelle est qu'elles ne permettent pas de modéliser le doute qui peut exister entre les différentes classes. En effet, on ne peut être toujours sûr s'il s'agit d'une expression positive, négative ou neutre. Récemment, la théorie de l'évidence a été introduite comme un outil adéquat pour la modélisation du doute dans l'analyse des expressions faciales [HAM05].

Dans ce qui suit, nous allons effectuer des investigations afin d'évaluer cette méthode quant à la modélisation du doute entre les différentes classes proposées, ainsi que de la capacité de fusion des données issues de différentes sources afin de prendre une décision sur la classe de l'expression étudiée. En plus, les travaux précédents utilisent spécialement des informations issues des traits permanents, nous proposons ici l'exploitation de tout type de déformation visuelle sur le visage étudiée : traits permanents et traits transitoires.

La table (3.1) résume les caractéristiques des différentes méthodes présentées dans l'état de l'art et l'apport de notre contribution dans la classification dimensionnelle.

Les méthodes de l'état de l'art	Type d'images	Bases d'images	Catégories	Performances
[COH02] : TAN	Séquence vidéos	Cinq sujets	Positive, négative, neutre surprise	77%
[SEB02] : Filtre Gaussien et filtre de Cauchy	Séquence vidéos	Cinq sujets	Positive, négative	79% et 81%
[MIU02] : Flux optique	Séquence vidéos	Huit sujets	Positive, négative, neutre	60,9%
[SAY]: codification des AUs		Quelques Aus	Aus Positives, Aus négatives	
Notre Méthode Théorie de l'évidence	Images fixes	Caplier_H, EEbase, Dafex, Chen[CHE00] Cohn et kanade	Positive, négative, neutre, OU Doute entre les classes	96,66% et 95,15%

Table 3.1. Résumé des méthodes de classification en expressions positives et négatives par rapport à la méthode proposée.

3.3. Notre Contribution

La base de tout système d'analyse d'expressions faciales est l'extraction des informations pertinentes qui peuvent décrire au mieux les phénomènes physiques. Afin d'extraire ces informations pertinentes, une détection du visage suivie d'une détection des traits permanents suivie d'une détection des traits transitoires est effectuée. Les différentes méthodes utilisées pour la détection de ces trois composants sont les mêmes méthodes présentées dans le chapitre I et le chapitre II.

L'idée est de définir les différentes sources susceptibles de nous procurer ce type d'informations. Chaque source est spécialisée dans un type d'information. La théorie de Dempster_Shefer est appliquée d'une façon locale sur chaque source afin de fusionner toutes les informations propre à un type. Une fonction de croyance de base lui est assignée. Enfin,

cette même méthode est appliquée d'une façon globale afin de fusionner les informations issues de toutes les sources et une fonction de croyance globale est estimée pour donner la classe de l'expression étudiée.

Dans notre cas, les classes considérées sont au nombre de trois et sont : « Positive », « Négative » et « Neutre ». Une source coïncide avec l'un des type d'information suivants: (1) Distances Faciales, (2) Valeurs de l'Angle nasolabial et (3) fonction de probabilité concernant la présence des traits transitoires.

3.4. Méthode Proposée

La méthode de classification dimensionnelle proposée dans ce chapitre diffère de la méthode de classification catégorielle présentée dans le chapitre II du point de vue séquençement dans les étapes. En effet, la méthode de classification catégorielle sépare l'étude des traits transitoires de l'étude des traits permanents. L'étape de classification étudie uniquement les traits transitoires, ce n'est que dans l'étape de post traitement (pas toujours effectuée) que les traits permanents seront étudiés.

La méthode de classification dimensionnelle consiste à la considération de n'importe quel type d'information en même temps, une fusion de toutes les informations issues que ce soit des traits transitoires ou des traits permanents est effectuée.

Trois sources de données sont considérées, la première source correspond au type numérique « Valeur distance ». Cette source comprend cinq distances faciales calculées et normalisées. La deuxième source correspond au type numérique « Valeur Angle ». Cette source (si elle existe) comprend une seule valeur calculée et normalisée. La dernière source correspond au type logique « Présence ou absence des TTs ». Cette source (si elle existe) comprend neuf régions pour chaque région est associée une valeur entre 0 et 1 qui va nous renseigner sur la présence des TTS (valeur=1), absence des TTS (valeur=0) ou qu'il ya un doute sur la présence des TTs sur une région (valeur= 0 OU 1). Pour chaque source, une fusion de toutes les données qui peuvent être extraites est appliquée en utilisant la théorie de Dempster_Shefer. A la fin, cette même méthode est appliquée afin de fusionner les résultats issus des trois sources.

3.4.1. Description des Différentes Sources

3.4.1.1. Source Géométrique : Distances Faciales

Cette source est basée sur des distances entre les points caractéristiques des traits permanents du visage. Cinq distances (les mêmes distances utilisées dans le chapitre II) sont estimées et normalisée par rapport à la distance entre les deux centres de l'iris. Ces distances sont ensuite comparées à celles du visage à l'état neutre. Notre objectif est de caractériser chaque classe d'expression par une combinaison spécifique de l'évolution des distances faciales. Une variable d'état V_i est associée à chaque distance D_i . Les différents états possibles de V_i sont: "C +" si la distance D_i accroît, "C-" si la distance D_i décroît et "S" si la distance D_i ne change pas.

3.4.1.1.1. Description de l'Expression Neutre: Pour vérifier si le visage étudié est à l'état neutre, on calcule les distances entre les paupières (D_1), entre les coins intérieures des yeux et des sourcils (D_2), entre les coins de la bouche (D_3), entre les lèvres (D_4), et entre les coins des yeux et de la bouche (D_5) (Figure 2.5). Si ces distances sont similaires à celles calculées depuis l'image de référence (visage à l'état neutre), le visage est considéré comme neutre. Sinon, l'expression est considérée comme positive, négative ou dans le doute entre les deux classes.

3.4.1.1.2. Description des Expressions Positives et Négatives: Le système de codification des actions faciales (FACS) [EKM78] est largement connu pour être le système le plus objectif disponible pour la description d'une expression faciale. Il peut aussi être utilisé pour discriminer entre les expressions positives et les expressions négatives. En étudiant les déformations qui se produisent sur un visage expressif, et en tenant compte de la table des unités d'actions (AUs) déduite par Ekman et al, nous pouvons constater que la partie inférieure du visage est la partie la plus instructive et la plus révélatrice sur la nature de l'expression : positive ou négative. Par conséquent, nous focaliserons notre étude sur l'analyse des unités d'actions relatives à la partie inférieure du visage (figure.3.2.).



















Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
					
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
					
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
					
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

Figure 3.2. Table des unités d'actions de la partie inférieure du visage [EKM78].

Notre objectif est de déduire la relation entre l'activation de certaines Aus et les distances changées. En d'autres termes, on voudrait connaître les conséquences de l'activation de certaines AUs sur les distances faciales.

Dans le cas des expressions négatives, selon [EKM82], [EKM86], [EKM80b], [ROZ94], [SOU96], [GOS95] and [VRA93], les AUs : AU9 (rides du nez), AU10 (élévation de la lèvre supérieure), AU14, AU15, AU20, et AU1+ AU4 sont censés être activés au cours d'une émotion négative. L'activation d'un ou d'une combinaison de ces Aus produit des changements sur certaines distances faciales qui sont D3 (distance entre les coins de la bouche) et D5 (la distance entre les coins de la bouche et des yeux). Par exemple, lorsque AU9 ou AU10 sont activés, D3 ne change pas mais D5 décroît (cf. Figure 3.2).

Les émotions positives résultent de l'activation de AU12 ou AU13, par conséquent la distance D3 accroît et la distance D5 décroît (cf. Table 3.2 et Figure 3.2).

Dans les autres cas, nous sommes dans le doute entre expressions positives et négatives.

Aus Activées	Distances changées	Classe d'Expression
AU9,AU10	D3 ne change pas; D5 décroît	E-
AU15,AU16	D3 ne change pas; D5 accroît	E-
AU20	D3 accroît; D5 accroît	E-
AU17, AU18, AU22, AU23, AU26, AU27	D3 décroît	E-
AU12, AU13	D3 accroît ; D5 décroît	E+
AU12, AU13	D3 accroît ; D5 ne change pas	E+ ou E-

Table 3.2. Relations entre l'activation des Aus et l'évolution des distances D3 et D5.

3.4.1.2. Source Géométrique: Angle Nasolabial : Phase d'Apprentissage

Dans cette phase, l'étude menée sur les rides nasolabiales dans la classification catégorielle est reportée ici pour la classification dimensionnelle. Les résultats de cette étude étant représentés sur la figure (2.10) du chapitre II.

Depuis cette figure nous pouvons déduire la table (3.3). Cette table présente le moyen que l'on peut utiliser afin de séparer la classe des expressions positives de celle des expressions négatives en se basant sur la seule information qui est l'angle nasolabial (s'il existe).

Intervalles Considérés	Expressions Positives	Exp. Positives OU Exp. Négatives	Expressions Négatives
[a,b]	X		
[b,d]		X	
[d,j]			X

Table 3.3. Classification d'expressions positives et négatives basée sur les rides nasolabiales.

A partir de la table (3.3), on peut conclure que : Si l'angle normé appartient à l'intervalle [a,b], la classe de l'expression étudiée est la classe « Positive », si l'angle appartient à l'intervalle [b,d], la classe est la classe « Négative Ou Positive » sinon la classe est « Négative ».

3.4.1.3. Source Probabiliste: Présence ou Absence des Traits Transitoires : Phase d'Apprentissage

Cette source est basée sur la présence ou l'absence des traits transitoires sur les régions d'intérêts du visage. Le but de cet apprentissage est d'associer la présence de certains traits transitoires à une classe d'expressions : classe positive ou classe négative. Les TTS peuvent apparaître sur une seule ou sur plusieurs régions faciales. Dans cette étude six régions sont considérées. La région nasale, les deux régions nasolabiales, les deux régions sur les deux coins de la bouche et sur le menton.(Figure 3.3).

La présence ou l'absence des traits transitoires est établie par la méthode présentée au chapitre II.

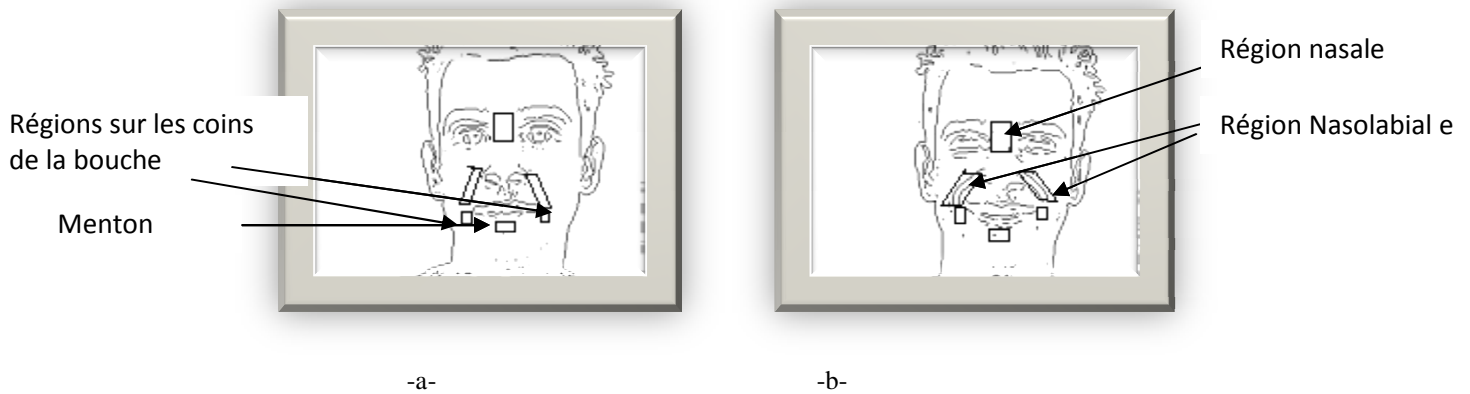


Figure 3.3. (a): Visage sans traits transitoires (b): visage avec traits transitoires sur les régions nasolabiales.

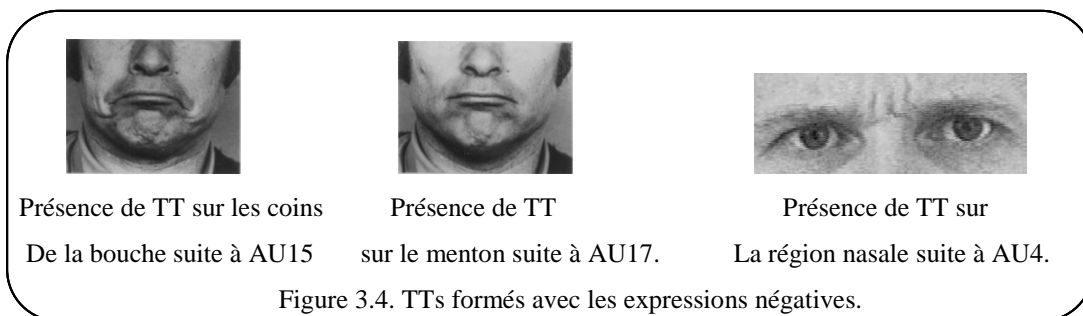
Afin de déterminer quels sont les traits transitoires qui apparaissent uniquement avec les expressions positives ou bien uniquement avec les expressions négatives, nous avons testé trois bases d'images. Chaque base est divisée en deux parties : base d'apprentissage et base de test : Hammal_Caplier base d'apprentissage [HAM base] (13 sujets, quatre expressions : *Joie*, *Dégout*, *surprise* et *Neutre*); Hammal_Caplier base de test (8 sujets, quatre expressions : *joie*, *Dégout*, *surprise* et *Neutre*); EEbase base d'apprentissage [EE base] (22 sujets, Six expressions universelles); EEbase base de test [EE base] (20 sujets, six expressions universelles); Dafex base d'apprentissage [Dafex base] (4 sujets, six expressions universelles) et enfin, Dafex base de test [Dafex base] (4 sujets, six expressions universelles). Nous avons détecté tous les TTs sur les différentes régions d'intérêts et nous avons résumé dans la table (3.4) les pourcentages de présence des TTS sur chaque région d'intérêt et pour chaque expression faciale.

Traits Transitoires	Expressions Positives		Expressions Négatives				
	Joie	SURPRISE POSITIVE	SURPRISE NEGATIVE	Dégout	Colère	Tristesse	Peur
Menton	/	/	/	/	39%	50%	21%
Coins de la bouche	/	/	/	/	/	28,5%	21%
Régions Nasolabiales	94%	/	/	94%	57%	21%	10%
Coins des yeux	40%	/	/	17%	7%	/	/
Région Nasale	/	/	/	77%	71%	53,5%	18%
Front	/	35%	25%	/	3,5%	28,5%	60%

Table 3.4. Présence ou absence des TTS sur chaque région faciale et pour chaque expression

Les lignes présentent les différentes régions d'intérêts et les colonnes présentent l'ensemble des expressions positives et négatives. Chaque case présente le pourcentage des sujets qui présentent des TTs sur une région considérée avec une expression donnée. Par exemple, 94% des sujets avec une expression de joie présentent des TTs sur les deux régions nasolabiales et 40% présentent des TTs sur les coins des yeux pour la même expression. En étudiant les résultats collectés dans cette table, on peut constater que quelques traits transitoires sont associés aux expressions négatives (voir Figure 3.4). Seulement trois régions d'intérêts sont concernées :

- a- La région des deux coins de la bouche : Des TTs peuvent apparaître sur cette région avec les expressions de Tristesse ou Peur (Expressions Négatives) suite à l'activation de AU15.
- b- Sur la région du menton : Des TTs peuvent apparaître sur cette région avec les expressions de Colère, Tristesse ou de Peur (Expressions Négatives) suite à l'activation de AU17.
- c- Sur la région nasale : Des TTs peuvent apparaître sur cette région avec les expressions de Dégout, Colère, Tristesse ou de Peur (Expressions Négatives) suite à l'activation de AU4.



Afin de prouver la dépendance entre la présence des TTs sur ces trois régions faciales et les expressions négatives, nous avons utilisé la statistique de Pearson (test de X^2) comme un critère de dépendance.

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Eq 3.1

O_{ij} est le nombre d'images observables avec l'expression E_i présentant des TTs sur la région R_j et E_{ij} est la valeur théorique correspondante (p : est le nombre des expressions considérées ; q : est le nombre des régions d'intérêts considérées).

Avec la supposition que les classe des expressions est indépendante de la présence des TTs sur les régions spécifiques du visage, et en choisissant un seuil d'erreur ($\alpha=0.05$), la statistique calculée est supérieure à la valeur observé sur la table de Pearson relative à la loi du X^2 . Ceci veut dire qu'il y a une forte dépendance entre ces deux variables. Cette mesure assure une généralisation de la règle déduite. Nous pouvons enfin conclure que si les TTs apparaissent sur l'une des trois régions d'intérêts : Sur les coins de la bouche, sur le menton ou sur la région nasale d'un visage, alors il ne peut s'agir que d'une expression négative. Dans le cas où les TTs sont absents sur ces trois régions, l'expression étudiée peut être positive, négative ou neutre.

Comme nous avons plusieurs source de données à considérer et pour chaque source, plusieurs données doivent être fusionnées, nous avons choisi d'utiliser la théorie de Dempster_Shefer (présentée dans le chapitre précédent) pour les mêmes raisons citées dans le chapitre II.

3.5. Théorie de l'Evidence dans le Contexte de Classification Dimensionnelle des Expressions Faciales

Comme nous l'avons indiqué avant, trois sources de données sont considérées dans cette classification: source géométrique (Distances faciales), source géométrique (Angle nasolabial) et la source probabiliste (Présence ou absence des TTs sur certaines régions faciales).

Le cadre de discernement global est $\Omega = \{E+, E-, E_n\}$ / $E+$ = expression positive; $E-$ = expression négative et E_n = expression neutre.

La théorie de l'évidence s'avère être tout à fait efficace dans la classification d'information issues de plusieurs sources même des sources de différent type. Afin de calculer l'évidence globale, chacune des sources est traitée séparément. Pour chaque source, une masse d'évidence locale est calculée.

3.5.1. Calcul des Masses d'Evidence Locales :

Pour calculer la masse d'évidence correspondante à chacune des sources (masse d'évidence locale), un modèle est proposé pour chaque donnée de chacune de ces sources.

3.5.1.1. Source Géométrique : Distances Faciales

La théorie de Dempster-Shafer est appliquée au niveau de chaque source. Le cadre de discernement associé à cette source est : $\Omega = \Omega = \{E+, E-, E_n\}$ (Chaque expression devrait être positive, négative ou neutre).

Suivant la section (3.4.1.1) deux distances sont considérées : D_3 la distance entre les deux coins de la bouche et D_5 la distance entre les coins de l'œil et de la bouche.

Une masse d'évidence de base (BBA) est assignée à chaque distance. Une variable V_i est associée à chaque distance D_i , si la distance D_i décroît l'état « C - » est assigné à V_i , si la distance D_i accroît, l'état « C + » est assigné à la variable V_i , si la distance D_i ne change pas, l'état « S » est assigné à la variable V_i .

Quand une expression faciale est positive, D_3 accroît ($V_3=C+$) et D_5 décroît ($V_5=C-$). Quand l'expression est négative, D_3 décroît et D_5 également ($V_3=C-$ et $V_5=C+$) ou bien D_3 décroît et D_5 ne change pas ($V_3=C-$ and $V_5=S$). Si l'expression est neutre, aucune distance ne change ($V_3=S$ et $V_5=S$). Un modèle est défini pour chaque distance indépendamment de l'expression.

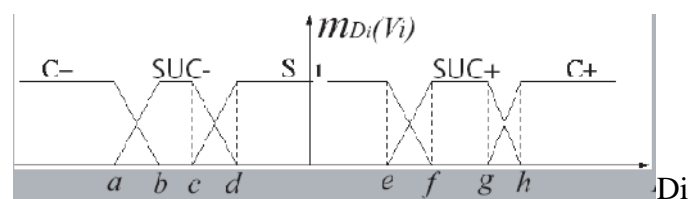


Figure 3.5. Modèle de la masse d'évidence basique pour chaque distance D_i

Les seuils (a, b, c, d, e, f, g, h) de chaque modèle sont estimés suite à une analyse statistique effectuée sur la base d'images [HAM base]. Cette base d'images comporte 21 sujets, elle a été divisée en un ensemble d'apprentissage appelé HCEL et un ensemble de test appelé HCE. L'ensemble d'apprentissage est ensuite divisé en plusieurs séquences expressives noté HCELe et en séquences neutre notés HCELn.

Les formules suivantes sont appliquées afin de trouver chacun de ces seuils.

$$\begin{aligned}
a &= \text{meanHCELe}(\min(D_i)_{1 \leq i \leq 5}) \\
h &= \text{meanHCELe}(\max(D_i)_{1 \leq i \leq 5}) \\
d &= \text{meanHCELn}(\min(D_i)_{1 \leq i \leq 5}) \\
e &= \text{meanHCELn}(\max(D_i)_{1 \leq i \leq 5})
\end{aligned}$$

$$\begin{aligned}
\text{Medianmin} &= \text{medianHCELe}(\min(D_i)_{1 \leq i \leq 5}) \\
\text{Medianmax} &= \text{medianHCELe}(\max(D_i)_{1 \leq i \leq 5}) \\
b &= a + \text{Medianmin} \\
c &= h - \text{Medianmax}
\end{aligned}$$

Après avoir calculé les seuils, la masse d'évidence $m_{D_i}(V_i)$ associée à chaque proposition concernant la distance D_i est estimée par la fonction représentée sur la figure (3.5).

Dans le but de formuler l'assignement de l'évidence de base en termes de combinaison d'expressions faciales, La table de règles logique des distances est utilisée. La table 3.5 donne un exemple de règles logiques relative à la distance D_3 reliant entre l'état pris par la variable d'état associée à cette distance et la classe d'expression correspondante.

D3->V3	E+	En	E-
C+	1U0	0	1U0
S	0	1	1U0
C-	0	0	1U0

Table 3.5. Règles logiques des états associés aux distances caractéristiques pour chaque classe d'expression.

La table (3.5) peut être interprétée comme suit: si D_3 accroît, les classes correspondantes peuvent être : E+ ou bien E-, si cette distance décroît, les classes correspondantes peuvent être En ou bien E-.

La règle de Dempster basée sur la somme orthogonale est ensuite utilisée afin de fusionner les deux données correspondantes à cette source (D_3 et D_5). Par conséquent, la masse d'évidence locale associée à la source géométrique « Distances » est calculée.

3.5.1.2. Source Géométrique : Angle Nasolabial

Selon la section (3.4.1.2), dans le cas de présence des rides nasolabiales, l'angle formé par ce type de rides est calculé ensuite comparé aux seuils (calculée dans la même section).

Si l'angle calculé appartient à l'intervalle $[a,b]$, l'expression étudiée est considérée positive, si l'angle est dans l'intervalle $[b,d]$, le résultat de classification comporte un doute entre positive et négative et si l'angle est dans l'intervalle $[d,j]$, l'expression est considérée négative.

C'est pourquoi, le cadre de discernement correspondant à cette source est défini par : $\Omega_2 = \{E+, E-\}$, il représente un sous ensemble du cadre de discernement global Ω . Le modèle proposé pour cette source est présenté sur la figure (3.6).

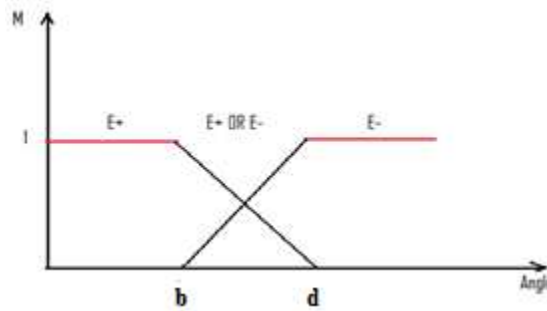


Figure 3.6. Modèle pour l'angle nasolabial

Le modèle proposé permet de calculer la masse d'évidence locale pour cette source. Suivant ce modèle, si l'angle nasolabial est inférieur à « b », l'expression étudiée est positive avec une masse d'évidence égale à « 1 », si l'angle nasolabial est supérieur à « d », l'expression étudiée est négative avec une masse d'évidence égale à « 1 », sinon il existe un doute entre les classes positive et négative et la masse d'évidence est calculée de puis la figure 3.6..

3.5.1.3. Source Probabiliste : Présence ou Absence des Traits Transitoires

La troisième source correspond à la présence de certains traits transitoires sur certaines régions faciales. Selon les conclusions de la section (3.4.1.3), si des TTs apparaissent sur l'une ou plus d'une des trois régions (Région nasale, Région du menton ou sur les coins de la bouche), l'expression étudiée est négative, sinon l'expression peut être positive, négative ou neutre.

Par conséquent, le cadre de discernement associé à cette source est définie par : $\Omega_3 = \Omega = \{E+, E-, E_n\}$.

Le modèle proposé pour chaque région d'intérêt est représenté sur la figure (3.7) :

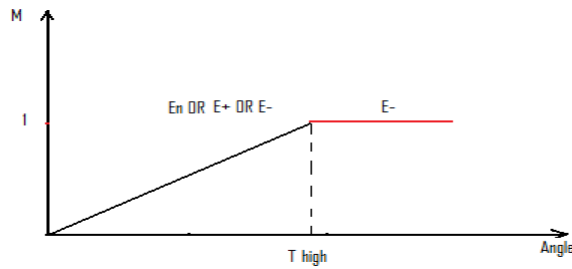


Figure 3.7. Modèle proposé concernant la présence ou l'absence des TTs.

Suivant ce modèle, si des TTs sont supposés présents sur une région donnée ($\text{Seuil} > T_{\text{high}}$), l'expression étudiée est négative avec une masse d'évidence égale à « 1 », sinon elle pourrait être positive, négative ou neutre avec une masse d'évidence calculée depuis le modèle représenté sur la figure (3.7).

La règle de Dempster_Shefer est utilisée au niveau de cette source afin de calculer la masse d'évidence locale.

3.5.2. Combinaison des Trois Sources: Approche Globale

Pour pouvoir appliquer l'approche globale de la théorie de l'évidence, les cadres de discernement associés à chacune des sources que l'on veut combiner doivent être compatibles ou égaux [SME00]. Dans notre cas :

le cadre de discernement de la première source est égale à $\Omega_1 = \{E+, E-, E_n\}$;

le cadre de discernement de la seconde source est égale à $\Omega_2 = \{E+, E-\}$ et enfin

le cadre de discernement de la troisième source est égale à $\Omega_3 = \{E+, E-, E_n\}$.

Comme nous pouvons le constater le premier et le troisième cadre de discernement sont égaux, le deuxième cadre de discernement est un sous ensemble des deux autres donc il n'est pas différent mais plutôt compatible aux deux autres cadres de discernement. Par conséquent l'approche globale peut s'appliquer sans aucune contrainte.

Puisque chacune des trois sources est analysée séparément des autres sources, l'approche globale permet de fusionner les résultats issus des trois approches locales en utilisant toujours la règle de Dempster.

Le résultat de cette fusion sera la classe qui possède la masse d'évidence maximale.

3.6. Résultats Expérimentaux

Afin d'évaluer les performances de l'approche proposée, plusieurs bases d'images (Images qui n'ont pas été utilisées dans la phase d'apprentissage) ont été testées. Ces bases sont : Hammal_Caplier [HAM base], EEbase [EE base], Dafex [Dafex base], JAFFE [Jaffe base], La base d'images positive/ négative [POS base] et enfin la base d'images de Cohn et Kanade [COH_base].

La table (3.6) présente les taux de classification obtenus. Les colonnes représentent les expressions classées par un expert et les lignes les classes d'expressions reconnues par le système.

	Neutre	Exp. Neg.	Exp. Pos.
Neutre	81,82%	0,93%	0,59%
Exp.Neg.	18,18%	<u>96,66%</u>	
Exp.Pos.		1,43%	<u>95,15%</u>
Neg OR Pos		0,98%	4,26%
Totale Reconnu		97,64%	99,41%
Totale	100%	100%	100%

Table 3.6. Taux de classification obtenus sur différentes bases d'images.

Les taux de classification des expressions positives et les taux de classification des expressions négatives sont bons. Parfois la classification donne un résultat qui contient le doute entre les deux classes, ceci est dû spécialement à l'activation des mêmes muscles faciaux soit avec des expressions positives ou négatives. Comme c'est le cas entre la joie et le dégoût. Cette confusion a été trouvée également dans différents autres systèmes d'analyse d'expressions faciales comme dans [HAM05].

3.7. Conclusion

Dans ce chapitre, nous avons présenté une nouvelle approche de classification dimensionnelle des expressions faciales. Cette approche étudie la dimension valence, c'est-à-dire la dimension où il est question de reconnaître si une expression est positive ou négative. Cette dimension est très intéressante dans l'étude comportementale et dans le monde virtuel. Avant la classification d'une expression étudiée, deux images sont présentées au système pour une éventuelle segmentation. Des informations de différents types sont extraites depuis les traits permanents ainsi que les traits transitoires du visage. Afin de mieux analyser ces données extraites, nous avons divisé ces données selon leurs types. Trois types de données sont utilisés. Chaque type est associé à une source : Source géométrique : Distances faciales, source géométrique : angle nasolabial (s'il existe) et enfin source probabiliste : Présence ou absence de certains traits transitoires. Un type spécifique de la théorie de l'évidence a été évalué dans ce chapitre ; Ce type correspond à celui qui utilise la fonction d'évidence globale résultat de l'approche globale. Cette approche est appliquée afin de fusionner des résultats issus de plusieurs approches locales. Chaque approche locale est appliquée sur un ensemble de données à fusionner de même type.

Selon les résultats obtenus, nous avons réussi à montrer la capacité de cette méthode dans la fusion de données issues de plusieurs sources et de différents types.

Dans le prochain chapitre nous allons évaluer cette même méthode dans la quantification des expressions faciales.

Chapitre 4

Estimation de l'intensité des Expressions Faciales

4.1. Introduction

De nos jours, la technologie occupe une place très importante dans notre société, elle ne cesse d'évoluer cependant les utilisateurs n'ont plus de temps pour s'adapter de plus en plus à la complexité des machines, c'est pourquoi la machine doit s'adapter à l'utilisateur en lui proposant une interface conviviale et ergonomique.

Afin de faciliter la communication homme machine, il est nécessaire d'équiper la machine d'un système émotionnel. Selon [LIE98], un système émotionnel doit pouvoir non seulement reconnaître l'expression faciale mais aussi la quantifier. Edwards [EDW98b] souligne l'importance des modèles quantitatifs des émotions et est étonné que peu de chercheurs soient intéressés par le calcul des intensités des expressions faciales. Il est pourtant intéressant d'avoir une idée de l'intensité d'une expression relativement à l'individu.

Dans le cadre de modélisation d'un modèle effective de dialogue, il est essentiel d'associer l'intensité à une expression faciale, car selon son degré, elle n'influencera pas le dialogue de la même manière. Ainsi une personne légèrement irritée ne se comportera pas d'une manière violente autant qu'une personne furieuse envers son interlocuteur.

Comme en psychologie, la psycho-analyse, la biologie (Darwin), la philosophie (Descartes), la médecine, la télé-formation, la simulation des personnes en réalité virtuelle, la commande de la vigilance pour un conducteur, les jeux interactifs ou dans les vidéoconférences, l'identification des expressions faciales avec leurs intensités est impliquée dans le procédé de décision pour définir la réaction correspondante par rapport au comportement de l'interlocuteur, que ce soit une machine ou un être humain.

Des chercheurs dans le domaine des expressions faciales sont influencés par Ekman, Friesen et Izard de sorte qu'ils travaillent généralement sur les six expressions universelles (la joie, dégoût, surprise, tristesse, colère, peur). Cependant avec l'étude de l'intensité de chaque expression, nous pouvons obtenir des classes secondaires d'expressions. Par exemple, pour la colère nous pouvons déduire: fureur, colère ou ennui et pour la peur: inquiétude, peur (crainte) ou terreur.

D'un autre coté L'intensité d'une expression faciale peut être d'intérêt pour d'autres raisons. Par exemple, dans [EKM80a] Ekman a constaté que l'intensité de l'action principale du

muscle zygomatique peut être en lien avec l'identification de l'expression faciale. Ce qui signifie qu'en estimant l'intensité, nous pouvons identifier l'expression faciale. En plus, la vitesse du début de sourire par rapport à l'intensité semble également différer nettement entre les sourires posés et spontanés [COH04]. Ce qui veut dire qu'on connaissant l'intensité d'une expression, nous pourrions discriminer entre expressions posées et expressions spontanées.

Pour toutes ses raisons, nous proposons dans ce chapitre une approche qui quantifie n'importe quelle expression faciale.

4.2. Etat de l'Art

Peu de chercheurs se sont intéressés à l'estimation de l'intensité d'une expression faciale. Les quelques systèmes développés pour l'estimation de l'intensité des expressions faciales rencontrés dans la littérature peuvent être divisés globalement en deux approches :

4.2.1 Approches Locales

Les méthodes dans cette approche suivent la position de quelques composants faciaux comme les yeux et la bouche, et supposent que le mouvement relatif de ces composants est lié à l'intensité de l'expression [HON98],[LIE98J],[WAN98]. En utilisant des méthodes de cette approche le nombre des entrées est considérablement réduit ce qui provoque une réduction du temps de calcul et minimisation de la complexité. Dans ce cas l'efficacité du traqueur devient plus importante.

Dans [LIE98J] les auteurs quantifient l'intensité des unités d'actions. Par contre ils ne se sont pas intéressés au problème de discrimination entre les différentes intensités mais plutôt ils ont utilisé l'intensité afin de reconnaître les unités d'actions. Dans [BAR99] les auteurs ont testé leurs algorithmes sur les expressions faciales qui changent systématiquement d'intensité comme il a été fait avec FACS, bien qu'ils n'aient pas réussi à séparer entre chaque niveau d'intensité et un autre (comme avec FACS), leurs travaux ont comme même apporté un peu de succès.

Dans [TIA00a] les auteurs ont réussi à comparer la codification de la variation de l'intensité manuelle et automatique, ces chercheurs ont utilisé le filtre de Gabor et les réseaux neurones afin de discriminer l'intensité de la fermeture des yeux. Une moyenne de reconnaissance de 83% est obtenue pour 112 images de 12 sujets.

Dans [PAN00a] les auteurs proposent un système qui reconnaît 30 actions faciales et leurs combinaisons ainsi que leur intensité en appliquant des systèmes d'intelligence artificielle et des systèmes non intelligents. Une approche hybride sous forme de combinaison de différentes

techniques de traitement d'images est appliquée afin d'extraire des informations faciales à partir d'images statiques. Ensuite, un système expert basé règles est appliqué afin de coder et quantifier les unités d'actions. Finalement un autre système expert est appliqué pour ajuster les résultats. Les études de validation sur le prototype réalisent des taux de reconnaissance de 90%. Il a été prouvé que ce taux dévie par une moyenne de 8% par rapport aux taux obtenus par FACS.

4.2.2. Approches Globales

Les méthodes de cette approche tiennent compte de toute l'information du visage [CHA99], [KIM97], [LIS98]. Cette information est préservée et elle permet au classificateur de découvrir les données pertinentes parmi les données recueillies. Cependant, l'étape de normalisation implique habituellement l'image entière et cela prend généralement beaucoup de temps. Le traitement de tous les Pixels dans l'image est coûteux et un grand espace mémoire est exigé. Ces méthodes sont relativement peu sensibles au mouvement subtil et le résultat peut être facilement affecté par le changement de l'éclairage.

Dans [ESS97] les auteurs ont représenté la variation de l'intensité de la joie en utilisant le flux optique.

Kimura et Yachida [KIM97] ont quantifié l'intensité par rapport à l'émotion, A. Loizides et al [LOI99] ont fait appel aux algorithmes génétiques pour donner un score entre 0 et 100 pour joie-tristesse, colère-calme et peur-relaxe, ils ont utilisé 25 marqueurs et 25 distances.

Dans [LEE03] les auteurs estiment l'intensité de trois expressions qui sont : Joie, Colère et Tristesse en temps réel. Ils entraînent d'abord le système afin de trouver les liens entre les positions des points faciaux et les intensités. Les données sont normalisées, ensuite traitées en utilisant un traçage isométrique (Isomap) afin d'extraire les paramètres de l'intensité. La modélisation de l'intensité se fait en utilisant des réseaux neurones en cascade (CNN) et des machines à vecteur de support (SVM). Les points faciaux sont suivies et normalisés en temps réel et l'intensité est estimée par apprentissage du modèle de l'intensité. Les taux de quantification sont de 95.0%, 76.7% et 70.6% pour la joie, la colère et la tristesse respectivement.

Dans [LIE98] les auteurs développent un système qui reconnaît automatiquement des unités d'actions en utilisant les modèles de Markov cachés (HMM) et estiment leurs intensités. Trois modules sont utilisés afin d'extraire des informations faciales : suivi de points caractéristiques faciaux, suivi de flux dense avec une analyse en composantes principales (ACP) et enfin la détection des rides. Après reconnaissance de la séquence de l'expression en entrée, l'intensité est estimée d'une image de la séquence en utilisant la propriété de corrélation de l'ACP. En effet la distance minimale entre deux points (Vecteurs de poids) dans l'espace propre possède la corrélation

maximale. La somme carrée des différences (SSD) est utilisée pour trouver la meilleure image de la séquence qui correspond exactement en intensité depuis n'importe quelle séquence utilisée dans l'apprentissage qui présente la même expression du test. L'intensité est quantifiée grâce à un entier qui prend ces valeurs entre 0 et 1 pour le minimum et maximum d'intensité respectivement.

Dans [KIM97] les auteurs quantifient la variation de l'intensité de trois expressions qui sont : Joie, Colère et surprise. Ils utilisent des modèles élastiques qui suivent le mouvement des contours du visage. Ensuite les vecteurs du mouvement des nœuds sont représentés dans un espace propre, et l'estimation est réalisée en projetant les images d'entrées sur l'espace des émotions.

Méthodes de l'état de l'art	Images	Catégories/Classes	Performances
SVM + Réseaux neurones			95.0%, 76.7% et 70.6% pour Joie, colère et tristesse
Flux dense avec analyse composantes principales (PCA), et la somme des carrées de différence (SSD)	Séquences d'expressions	Entier avec des valeurs entre 0 et 1, pour l'intensité min et max respectivement	
Modèles élastiques (Elastic Net Model)			Joie, colère et surprise
Filtre de Gabor et réseau de neurones artificiel	Séquences d'images de 12 sujets	Fermeture de l'œil AU41, AU42, et AU43 (comme FACS)	83%
Systèmes d'intelligence artificielle et systèmes non intelligents	Images statiques	5 unités d'actions	90% dévié par la moyenne de 8% par rapport à la quantification manuelle)
Méthode proposée			
théorie de l'évidence	Images statiques: Hammal_C (21suj) EEbase (42 suj) et Dafex (8 suj).	Trois niveaux: Min, moy et max + Doute entre intensités	93,58%, 92,31%, 91,02%

Table 4.1 Comparaison des méthodes de l'état de l'art.

4.3. Notre Contribution:

Le visage est une zone importante du corps humain qui possède une trentaine de muscles, l'électromyographie (EMG) est une technique permettant de mesurer l'activité musculaire au cours du temps. Cette technique manuelle, ne pouvait fournir un système complet de mesure en effet, une électrode de surface mesure n'importe quelle activité musculaire dans son secteur, éliminant les distinctions qui peuvent être faites visuellement. Une solution pourrait être d'utiliser une électrode d'aiguille qui mesure seulement l'activité du muscle en lequel elle est insérée, ceci a donné naissance aux « marqueurs » placés sur le visage. Jusqu'ici, les systèmes basés Marqueurs sont les seules capable de coder toutes les activités et les intensités des unités d'actions [VAL99].

Les actions faciales sont décrites par leur situation et leur intensité. Ceci est le rôle de FACS proposé par Ekman et Friesen [EKM78] et a été considéré comme une base pour la description des

expressions faciales. FACS utilise 46 unités d’actions pour décrire les actions faciales par rapport à leur location et leur intensité, cette dernière est mesurée par trois ou cinq niveaux d’intensité. FACS est la méthode la plus utilisée pour mesurer et décrire les comportements du visage. Elle a été conçue pour déterminer comment la contraction de chaque muscle change l’apparence du visage. Ekman a proposé un modèle pour donner un score à quelques unités d’actions. L’intensité d’une unité peut être marquée sur un plan en cinq points d’échelle ordinaire (A, B, C, D, E) comme l’indique la Figure 4.1. FACS utilise des conventions ou des règles pour fixer des seuils pour la notation de l’intensité de l’unité.



Figure 4.1 les différents degrés d’intensité d’une unité d’action

L’échelle de pointage A-B-C-D-E n’est pas un intervalle à échelle égale, les niveaux C et D, forment un intervalle large qui présentent les changements d’apparence les plus fréquents, et la plupart des variations des unités d’actions coïncident avec ces niveaux.

A, B et E sont définis comme étant très restreints. Les niveaux A et B sont souvent confondus, la séparation entre D et E est difficile à déterminer, en plus de toutes ces différences la combinaison de deux ou plusieurs AUs modifient l’intensité de l’AU. Pour toutes ces raisons, nous avons proposé un modèle qui ressemble à celui proposé par Ekman avec la différence d’utilisation de trois niveaux au lieu de cinq, les niveaux sont d’intervalles égaux afin de faciliter la détection des limites entre niveaux.

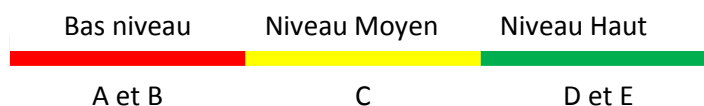


Figure 4.2 Modèle d’intensité choisie

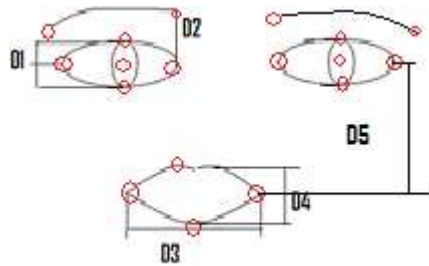
- « bas niveau » remplace A et B
- «niveau moyen» remplace C
- «haut niveau »remplace D et E.

Avec FACS, seules six unités d’actions peuvent être quantifiés : AUs 25-27 et AUs 41-43. Le problème est que la quantification de ces six unités ne suffit pas pour quantifier toutes les expressions faciales, c’est pourquoi nous avons remplacé la quantification des unités d’actions par la

quantification des degrés de fermeture_Ouverture des yeux, la fermeture_Ouverture de la bouche et le degré de froncement ou élévation des sourcils.

Une nécessité s'impose alors qui est la formulation des déformations qui sont en termes d'unités d'action en termes de distances faciales calculées depuis les points caractéristiques du visage.

Les distances considérées sont les mêmes utilisées dans le chapitre I de ce mémoire ; La figure (4.3) rappelle ces distances :



D1: Distance entre les deux paupières (inf, sup)
 D2: Distance entre coins intérieurs de l'oeil et du sourcil
 D3: Distance entre les deux coins de la bouche
 D4: Distance entre lèvre supérieur et lèvre inférieur
 D5: Distance entre les deux coins de l'oeil et de la bouche.

Figure 4.3. Les points caractéristiques du visage et les distances biométriques.

Afin de réaliser cette conversion, nous considérons la table des unités d'actions d'Ekman (Figure 4.4 et Figure 4.5). Dans la région de la bouche, les AUs 12, 13, 14 correspondent à l'ouverture horizontale. En termes de distance, D3 remplace ces trois AUs.

Les AUs 25, 26, 27 correspondent à l'ouverture verticale de la bouche. En termes de distance, D4 remplace ces trois AUs.

Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

Figure 4.4 Unités d'actions de la partie inférieure du visage(Ekman)

Dans la région de l'œil, les AUs, 1, 2 et 4 représentent le mouvement des sourcils. En termes de distance, D2 remplace ces trois AUs.

Les AUs 41, 42, 43 ou 45 représentent la variation moyenne de l'intensité de la fermeture des yeux. En termes de distance, D1 remplace ces trois AUs.

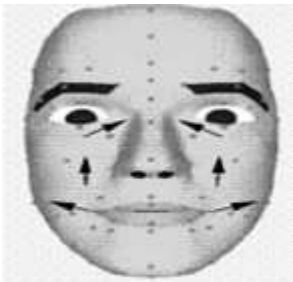
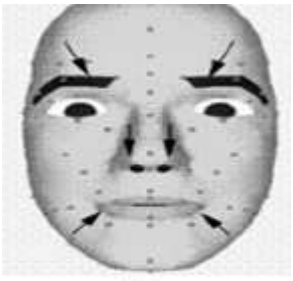
Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46

Figure 4.5 Unités d'actions de la partie supérieure du visage(Ekman)

Pour pouvoir relier les déformations de la partie supérieure du visage avec celles de la partie inférieure du visage, une dernière distance est ajoutée (D5) cette distance représente le mouvement de la bouche par rapport à l'œil.

Afin de définir quelles sont les distances qui caractérisent au mieux l'intensité de l'expression nous étudions la description des expressions faciales (table 4.2).

Lors de la production d'une expression faciale, il apparaît sur le visage un ensemble de déformations au niveau des traits permanents du visage .

L'expression faciale	Les déformations survenues sur le visage
 <p>joie</p>	<ul style="list-style-type: none"> • Les yeux sont légèrement plissés, c'est-à-dire que la paupière inférieure couvre en partie l'œil. • La bouche est ouverte, c'est un mouvement horizontal. Les lèvres sont donc étirées, toujours dans un mouvement latéral • La personne peut montrer les dents si elle le désire, donc on parle d'un mouvement vertical de la bouche
 <p>colère</p>	<ul style="list-style-type: none"> • La paupière recouvre une partie de l'œil donc les yeux seront presque fermés. • Quant à la bouche, elle reste fermée mais assez serrée, sinon, elle s'ouvre verticalement • Les sourcils ont tendance à se rejoindre, ils sont froncés, plissés. De plus, leur partie intérieure est abaissée légèrement

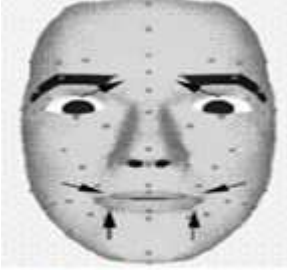
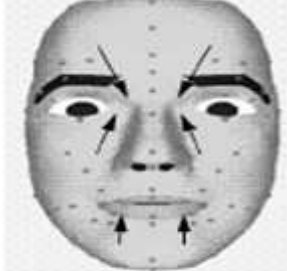
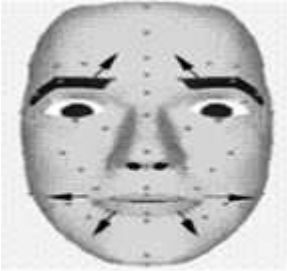
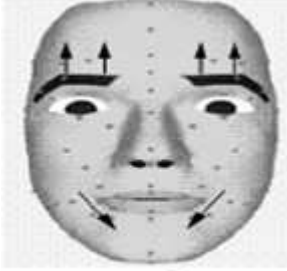
 <p>tristesse</p>	<ul style="list-style-type: none"> • Les coins intérieurs des sourcils sont légèrement élevés pour donner cette forme / \ ou encore ㄩ. • Les paupières recouvrent une partie du champ de vision • La bouche est serrée mais elle descend légèrement, les coins seront étirés vers le bas.
 <p>dégoût</p>	<ul style="list-style-type: none"> • Les coins intérieurs des sourcils sont légèrement abaissés. • La bouche est fermée mais on peut remarquer que la lèvre supérieure remonte • la réduction du champ de vision : l'œil est à demi-ouvert.
 <p>peur</p>	<ul style="list-style-type: none"> • Les yeux sont grands ouverts, écarquillés. • Ce mouvement des yeux a pour effet le redressement des sourcils. • La bouche est ouverte mais cela reste néanmoins un mouvement horizontal. Les lèvres sont donc étirées, toujours dans un mouvement latéral • La paupière est entièrement levée. La pupille est visible dans sa totalité. Le champ de vision est au maximum. La personne semble fixer quelque chose comme si elle ne pouvait s'en détacher.
 <p>surprise</p>	<ul style="list-style-type: none"> • Les yeux sont grands ouverts. • La bouche est ouverte verticalement. • Les sourcils sont soulevés

Table 4.2 Descriptions des six expressions faciales et les déformations pertinentes survenues sur le visage.

L'intensité d'une expression faciale est en fonction de l'intensité des changements apparus sur le visage. Les déformations faciales représentant l'information pertinente dans l'estimation de l'intensité de l'expression faciale sont ensuite fusionnées afin de donner une décision sur l'intensité finale de l'expression étudiée.

4.4. Méthode Proposée

La méthode proposée s'applique sur des images statiques qui présentent une des six expressions faciales connues. L'expression faciale est reconnue en utilisant la méthode de classification catégorielle présentée dans le chapitre II. Elle est basée principalement sur le degré des déformations faciales géométriques de certaines composantes faciales. Une image de référence (image avec expression neutre) est utilisée pour chaque sujet afin de la comparer avec l'image d'expression et déduire les différentes déformations. Après l'étape d'extraction des composantes permanentes du visage, et la localisation des points caractéristiques faciales de l'image étudiée (méthode présentée dans le chapitre I), des distances biométriques sont alors calculées. Un modèle est utilisé afin de modéliser le degré de déformation de chaque trait permanent. Un ensemble de distances qui caractérisent au mieux l'intensité de l'expression est considéré pour chaque expression (Joie, Dégout, Surprise, Peur, Tristesse et colère). Une méthode de fusion de données qui est une fois encore la théorie de Dempster-fisher est ensuite appliquée afin de donner la décision sur l'intensité de l'expression. En cas de conflit une étape de post traitement est alors ajoutée pour résoudre le conflit.

4.5. Application de la Théorie de l'Evidence dans le Contexte d'Estimation de l'Intensité des Expressions Faciales

Comme nous l'avons définie dans le chapitre II, l'application de la théorie de l'évidence nécessite la définition d'un cadre de discernement qui rassemble tous les cas possible.

Dans notre application : $\Omega = \{E_{i_min}, E_{i_Moy}, E_{i_max}\}$; $i=1..6$; 6 est le nombre des expressions universelles (Joie, Dégout, Tristesse, Colère, Peur, Surprise).

- L'hypothèse **Ei_min** correspond à l'intensité faible de l'expression Ei
- L'hypothèse **Ei_moy** correspond à l'intensité moyenne.

- L'hypothèse **Ei_max** correspond à l'intensité maximale.

2^Ω correspond à l'ensemble des intensités élémentaires d'une expression ou à une combinaison de deux intensités d'expression, tel que :

$$2^\Omega = \{Ei_min, Ei_moy, Ei_max, (Ei_min \cup Ei_moy), (Ei_moy \cup Ei_max)\}$$

A est l'un de ses éléments, sachant que cette définition prend en considération n'importe quel type d'expressions **Ei**. nous supposons que les hypothèses impossible tel que : **Ei_min** \cup **Ei_max** sont enlevés de 2^Ω .

Ei_min \cup **Ei_moy** indique le doute entre l'expression avec intensité minimale et l'expression avec intensité moyenne.

4.5.1 Définition des États Symboliques :

Une variable d'état **Vi** ($1 \leq i \leq 5$) est associée à chaque caractéristique de la distance **Di** afin de convertir la valeur numérique de la distance à un état symbolique. L'analyse de chaque variable montre que **Vi** peut prendre trois États,

$$\Omega' = \{min, moy, max\}; 2^{\Omega'} = \{min, moy, max, minUmoy, moyUmax\}$$

Où **minUmoy** indique le doute entre **min** et **moy**, **moyUmax** indique le doute entre **moy** et **max**.

Nous supposons également que les symboles impossibles (par exemple **minUmax**) sont enlevés de $2^{\Omega'}$.

4.5.2. Le Processus de Modélisation

L'objectif du processus de modélisation est d'associer un état symbolique **Vi** à chaque distance **Di** et à lui assigner une masse d'évidence. Pour mener à bien cette conversion, nous définissons un modèle pour chaque distance en utilisant les états de $2^{\Omega'}$.

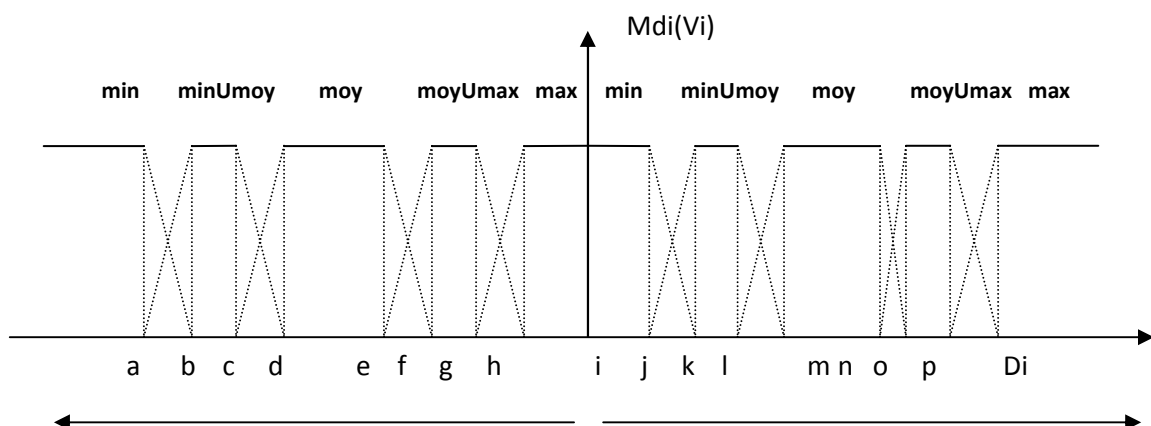


Figure 4.6 Modèle proposé pour l'estimation de l'intensité

Un seul modèle est défini pour chaque distance indépendamment de l'expression faciale et de son intensité. On sera donc dans l'un de ces deux cas possibles :

- Si la distance calculée accroît par rapport à l'état neutre, on prendra en considération la partie droite du modèle c.à.d. à partir du seuil **i** jusqu'au seuil **p**.
- Si la distance calculée décroît, on prendra en considération la moitié gauche du modèle c.à.d. du seuil **a** jusqu'au seuil **h** (Figure 4.6).

L'évidence $m_{DI}(\mathbf{Vi})$ est obtenue par la fonction illustrée sur la figure 4.6.

4.5.3. Définition des Seuils

Les Seuils (**a**, **b**, ..., **p**) de chaque modèle sont définis par l'analyse statistique effectuée sur la base (Hammal_Caplier). Cette base d'images comporte 21 sujets, elle a été divisée en un ensemble d'apprentissage appelé **HCE_L** et un ensemble de test appelé **HCE**.

L'ensemble d'apprentissage est ensuite divisé en plusieurs séquences expressives noté **HCE_{L_e}** et en séquences neutre notés **HCE_{L_n}**.

- Le seuil minimum **a** est une moyenne des valeurs minimales des distances de la base de données de **HCE_{L_e}**.
- De même, le seuil maximum **p** est obtenu à partir des valeurs maximales.
- Les seuils du milieu **h** et **i** sont définis respectivement en tant que la moyenne du minimum et maximum des distances caractéristiques de la base de données **HCE_{L_n}**.
- Le seuil **b** est la médiane des valeurs caractéristiques de distances pour les expressions faciales assignées à l'état le plus élevé **min**.
- **g** est la médiane des valeurs caractéristiques de distances pour les expressions faciales assignées à l'état inférieur **S**.
- Le seuil intermédiaire **d** est calculé comme étant la moyenne de différence entre la limite des seuils **a** et **h** divisé par **trois** (conformément à la supposition de section 2,1) augmentée de la valeur du seuil **a**.
- De même, le seuil **e** est la moyenne de la différence entre la limite des seuils **a** et **h** divisé par trois et réduit par la valeur du seuil **h**.

- Les seuils **c** et **f** sont calculés comme la moyenne des seuils **b** et **d** respectivement **e** et **g**.
Les seuils de la partie droite du modèle proposé sont calculés de la même façon.

4.5.4. Définition des Intensités des Expressions

L'étude de la table 4.2 a permis de définir un ensemble de distances qui caractérisent l'intensité de chaque expression faciale.

4.5.3.1. Expression de la Joie E1

Les changements les plus importants apparaissant sur un visage avec l'expression de joie sont les suivants : les coins de la bouche sont tirés vers les oreilles et les yeux deviennent légèrement fermés ainsi les distances les plus importantes considérées dans l'évaluation de l'intensité de l'expression de joie sont **D1** et **D3**.



Table 4.3 Etats des variables associées aux distances considérées pour chaque intensité de la joie

Figure 4.7 Evolution des distances dans le cas de la joie

4.5.3.2 Expression de la Surprise E2

Les changements les plus importants apparaissant sur le visage avec une expression de surprise sont : la bouche s'ouvre verticalement, les yeux sont grand ouverts et les sourcils sont élevés. Ainsi les distances considérées dans l'évaluation de l'intensité de l'expression de surprise sont **D1**, **D2** et **D4** (figure 4.8)



Table 4.4. Etats des variables associées aux distances considérées pour chaque intensité de la surprise

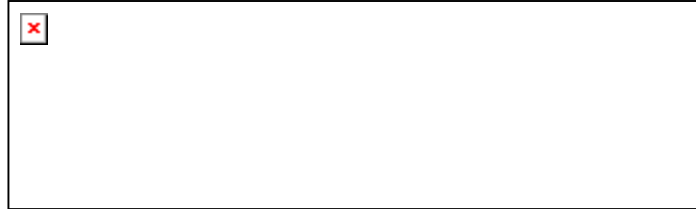


Figure 4.8 Evolution des distances dans le cas de la surprise

4.5.3.3. Expression du Dégoût E3 :

Les changements les plus importants apparaissant sur le visage avec une expression de dégoût sont : la bouche est ouverte et les yeux deviennent légèrement fermés. Ainsi les distances les plus importantes considérées dans l'évaluation de l'intensité de l'expression de dégoût sont **D1** et **D4** Figure 4.9.



Table 4.5 Etats des variables associées aux distances considérées pour chaque intensité du dégout



Figure 4.9. Evolution des distances dans le cas de Dégout

4.5.3.4. Expression de Colère

La colère est une expression qui possède différentes descriptions (table 4.2), dans une description, les yeux sont grand ouverts et la bouche également, dans une autre description les yeux sont presque fermés et les lèvres serrées par contre la distance entre les coins intérieurs des sourcils et les coins intérieurs des yeux est minimisée. Donc les distances considérées dans l'estimation de l'intensité de la colère sont : D2 et D4.

	V2	V4
E4_min	Max	Min
E4_moy	MoyUMax	MoyUMin
E4_max	Min	Max

Table 4.6 Etats des variables associées aux distances Considérées pour chaque intensité de Colère

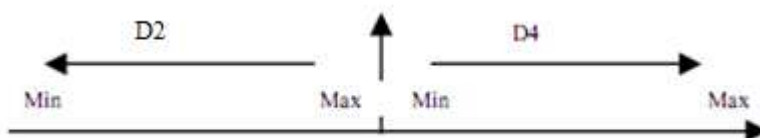


Figure 4.10 Evolution des distances dans le cas de la Colère.

4.5.3.5. Expression de Tristesse E5

Les changements les plus importants apparaissant sur le visage avec une expression de tristesse sont : les coins intérieurs des sourcils sont élevés, les paupières sont relâchées donc les yeux deviennent légèrement fermés et parfois, les coins de la bouche sont étirés vers le bas. Ainsi les distances les plus importantes considérées dans l'évaluation de l'intensité de l'expression de tristesse sont **D1** et **D2** Figure 4.11.

	V1	V2
E5_min	Min	Min
E5_moy	MinUMoy	Moy
E5_max	Max	Max

Table 4.7 Etats des variables associées aux distances considérées pour chaque intensité de la tristesse

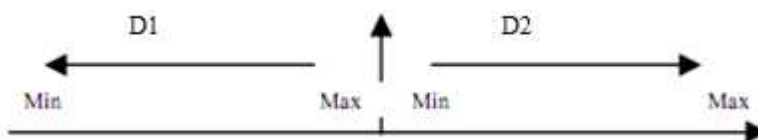


Figure 4.11 Evolution des distances dans le cas de la Tristesse.

4.5.3.6. Expression de la Peur E6

Les changements les plus importants apparaissant sur le visage avec une expression de Peur sont : la bouche s'ouvre verticalement, les yeux sont ouverts et les sourcils sont élevés. Ainsi les

distances considérées dans l'évaluation de l'intensité de l'expression de peur sont **D1**, **D2** et **D4** (figure 4.12)

	V1	V2	V4
E6min	Min	Min	Min
E6moy	Moy	MoyUMin	MoyUMin
E6max	Max	Max	Max

Table 4.8. Etats des variables associées aux distances considérées pour chaque intensité de la Peur

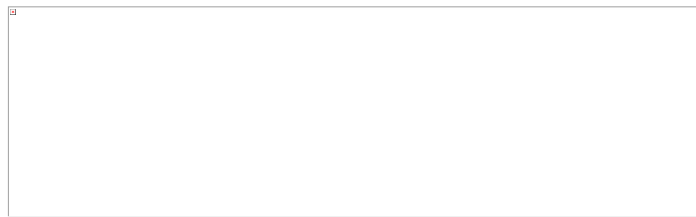


Figure 4.12 Evolution des distances dans le cas de la Peur

4.5.5. Règles Logiques entre les États Symboliques et les Intensités d'Expressions

Dès que les états symboliques seront assignés à chaque distance caractéristique, Il est essentiel de raffiner le processus en formulant les états de distances en termes d'intensité d'expressions. Afin de réaliser cette tâche, les tables de règles logiques déduites depuis les tables des états des variables associées aux distances considérées pour chaque intensité de chaque expression faciale sont utilisées.

La table 4.9 donne les règles logiques associées aux distances **D1** et **D3** pour l'expression de joie.

Vi	State Value	E1_min	E1_moy	E1_max
V1	Min	0	0	1
	Moy	0	1	0
	Max	1U0	1U0	0
V3	min	1U0	1U0	0
	Moy	0	1	0
	max	0	0	1

Table 4.9 Règles logiques pour D1 et D3 (la joie)

Si un état (min, moy, max) est atteint lors d'une expression, la valeur « 1 » est assignée à l'intensité correspondante, sinon la valeur « 0 » lui est alors assignée.

Par exemple, si **V1 = max** alors l'expression atteinte correspond à **E1_minUE1_moy**, cela signifie que:

$$\begin{aligned}
m_{D1}(\max) &= m_{D1}(E1_minUE1_moy) \\
m_{D1}(\text{moy}) &= m_{D1}(E1_moy) \\
m_{D1}(\min) &= m_{D1}(E1_max) \\
m_{D3}(\max) &= m_{D3}(E1_max) \\
m_{D3}(\text{moy}) &= m_{D3}(E1_moy) \\
m_{D3}(\min) &= m_{D3}(E1_minUE1_moy)
\end{aligned}$$

4.5.6. Fusion de Données

En vue d'estimer l'intensité d'une expression faciale, l'information disponible m_{D_i} est combinée pour être intégrée à la loi d'association de Dempster [Dem67] (combinaison conjonctive). Par exemple, on considère les deux distances caractéristiques D_i, D_j pour lesquelles nous associons deux masses d'évidence élémentaires m_{D_i} et m_{D_j} définies sur le même espace de discernement 2^Ω . La masse d'évidence globale (C.A.D des deux distances combinées) $m_{D_{ij}}$ est donnée par le biais de la combinaison conjonctive (somme orthogonale) :

$$\begin{aligned}
m_{D_{ij}}(A) &= (m_{D_i} \oplus m_{D_j})(A) \\
&= \sum_{B \cap C = A} m_{D_i}(B) m_{D_j}(C)
\end{aligned}$$

Où A, B et C désignent des propositions et $B \cap C$ désigne la conjonction (intersection) entre les propositions B et C .

Pour être plus explicite, nous considérons les masses suivantes :

$$m_{D1}(\minUE1_moy) = m_{D1}(E1_maxUE1_moy) = 0,15$$

$$m_{D1}(\text{moy}) = m_{D1}(E1_moy) = 0,85$$

$$m_{D3}(\text{moy}) = m_{D3}(E1_moy) = 1$$

La jointure des deux distances sera comme suit :

D1/D3	E1_minUE1_moy	E1_moy
E1_moy	E1_moy	E1_moy

$$\begin{aligned}
m_{D13}(E1_moy) &= \\
&= m_{D1}(E1_minUE1_moy).m_{D3}(E1_moy) + m_{D1}(E1_moy). \\
&= m_{D3}(E1_moy) = 1
\end{aligned}$$

Table 4.10 Exemple de combinaison des deux distances

4.5.7. Décision

La décision est la dernière étape du processus de classification de l'intensité d'une expression autant que min, moy ou max. Elle consiste à faire un choix entre les trois intensités ou bien de considérer le doute possible entre ces intensités. Faire un choix, c'est prendre un risque, sauf si le résultat de la combinaison est parfaitement fiable ($E_i = 1$). Ici, la proposition retenue est celle avec la valeur maximale de la masse d'évidence qui représente l'élément de preuve.

4.6. Post-Traitement en Cas de Conflit

Parfois, l'ensemble vide apparaît dans le tableau de combinaisons des distances. Il correspond à des situations où des valeurs des distances caractéristiques conduisant à la configuration d'états symboliques ne correspondant à aucune des définitions d'une expression. Cela doit être lié au fait que Ω n'est pas vraiment exhaustive.

En réalité, tout le monde exprime ses émotions différemment. En cas de la surprise, parfois, une personne ouvre les yeux mais n'ouvre pas la bouche (voir Figure 4.13).



Figure 4.13 Exemple d'expression de surprise avec conflit

$$m_{D1}(\text{moy}) = 1 = m_{D1}(E2_moy)$$

$$m_{D2}(\text{moy}) = 1 = m_{D2}(E2_moy)$$

$$m_{D4}(\text{min}) = 1 = m_{D4}(E2_min)$$

D1/D2	E2_moy
E2_moy	E2_moy

$$m_{D12}(E2_moy) = m_{D1}(E2_moy) \cdot m_{D2}(E2_moy)$$

D1/D4	E2_moy
E2_min	∅

Table 4.11. Exemple de conflit (erreur)

Dans ce cas, une résolution de conflit s'impose. Selon le guide de l'investigateur de FACS, le nombre d'AUs activé est employé pour estimer l'intensité d'une expression. « Si seulement 2 AUs parmi 4 sont activés quand une expression est exprimée, nous pouvons conclure que l'intensité n'est pas maximale ».

De la même manière, puisque deux ou trois distances seulement sont considérées pour chaque expression, quand une distance n'est pas à sa limite (maximum ou minimum), l'expression n'est pas à sa limite non plus (maximum ou minimum) ainsi c'est une intensité moyenne.

4.7. Résultats Expérimentaux

Dans le but d'évaluer les performances de la méthode proposée, plusieurs bases d'images ont été testées.

4.7.1. Résultats sur la Base [HAM base]

Nous avons considéré 10 sujets de la base [Ham base] qui n'ont pas servi dans la phase d'apprentissage, ces sujets ont enregistré des vidéos des expressions de Joie, Dégout et Surprise avec différentes intensités. Chaque vidéo comprend au minimum 100 images. 10 images avec une intensité minimale qui correspondent aux premières images de la vidéo où un expert humain arrive à distinguer les premiers changements qui apparaissent sur le visage, 10 images avec une intensité maximale qui correspondent à l'apex de l'intensité de l'expression faciale et enfin 20 images avec une intensité moyenne qui correspondent à des images prises depuis la séquence vidéo et qui sont situées entre les images avec intensité minimale et les images avec intensité maximale.

Les résultats de classification de l'intensité des expressions de cette base avant la phase de post traitement sont résumés dans la table 4.12.

Exp.	Reconnues	Doute min/moy	Doute moy/max	Erreur
E1_min	70%	30%	0%	0%
E1_moy	95%	0%	5%	0%
E1_max	100%	0%	0%	0%
E2_min	52,89%	47,11%	0%	0%
E2_moy	81,25%	0%	0%	18,75%
E2_max	66,66%	0%	0%	33,33
E3_min	32,5%	42,5%	0%	25%
E3_moy	53,85%	7,69%	0%	38,5%

E3_max	62,5%	0%	12,5%	25%
--------	--------------	----	-------	-----

Table 4.12 classification des intensités de la base [HAM base] avant le post traitement

Les lignes correspondent aux trois intensités des trois expressions étudiées, et les colonnes présentent les taux de classification de l'intensité.

Nous pouvons constater que les taux de reconnaissances des intensités telles quelles sont étiquetées dans la base par un expert sont importants spécialement les taux de reconnaissance de l'intensité maximale. Ceci peut être expliqué par le fait que l'intensité maximale est l'intensité la plus facile à simuler. Les mauvais taux sont assignés à l'intensité minimale, ceci peut être expliqué par le fait que l'intensité minimale est l'intensité la plus difficile à simuler car elle doit être exprimée d'une façon suffisamment petite pour qu'elle soit étiquetée autant que minimale et suffisamment grande pour qu'elle soit reconnue. C'est pourquoi les taux de reconnaissance de cette intensité baissent en faveur du doute entre intensité minimale et intensité moyenne. Comme nous pouvons le constater des taux importants sont assignés au doute entre les intensités. Nous estimons qu'il est préférable de garder le doute que de donner une mauvaise classification.

Des taux d'erreur sont également jugés importants, ceci est dû à la variation dans l'expression des émotions selon les différents individus.

La phase de post traitement finie par résoudre ces cas d'erreur ce qui donne une amélioration des taux de la reconnaissance. Les résultats de classification après application de cette phase sont regroupés dans la table 4.13:

Exp.	Reconnue	Doute min/moy	Doute moy/max	Erreur
E1_min	70%	30%	0%	0%
E1_moy	95%	0%	5%	0%
E1_max	100%	0%	0%	0%
E2_min	52,89%	47,11%	0%	0%
E2_moy	100%	0%	0%	0%
E2_max	100%	0%	0%	0%
E3_min	57,5%	42,5%	0%	0%
E3_moy	92,11%	7,69%	0%	0%
E3_max	87,5%	0%	12,5%	0%

Table 4.13 classification des intensités de la base [HAM base] après post traitement

Finalement nous pouvons constater que les meilleurs taux sont donnés pour la reconnaissance de l'expression de Joie. Ceci peut être expliqué par l'universalité et l'unification de la description de cette expression.

4.7.2. Résultats sur la Base EEbase

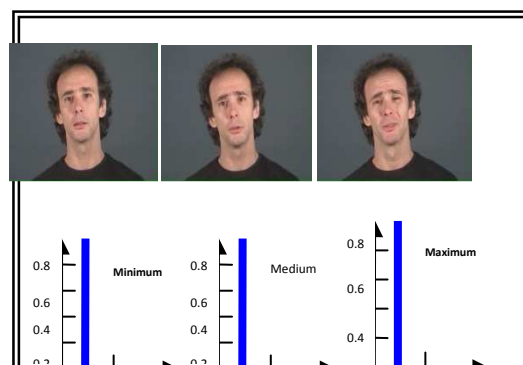
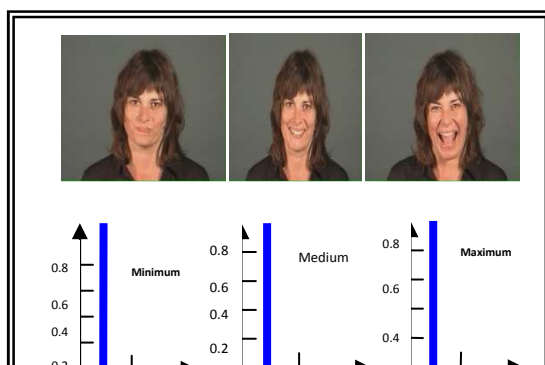
Une autre base d'images a été testée vu que la première base ne comportait pas des images avec les expressions de colère, tristesse et peur. La base [EE base] comporte 42 sujets, ces sujets expriment six expressions faciales (Joie, Dégout, Colère, Peur, tristesse et surprise) et chaque expression avec différentes intensités.

Les résultats de classification de l'intensité des six expressions de tous les sujets de la base sont résumés dans la table 4.14:

Exp.	Reconnue	Doute min/moy	Doute moy/max	Erreur
E1_min	50%	50%	0%	0%
E1_moy	70%	0%	30%	0%
E1_max	100%	0%	0%	0%
E2_min	Images non disponibles			
E2_moy	Images non disponibles			
E2_max	100%	0%	0%	0%
E3_min	Images non disponibles			
E3_moy	100%	0%	0%	0%
E3_max	90%	0%	10%	0%
E4_min	Images non disponibles			
E4_moy	66,7%	2.5%	5.1%	25,07%
E4_max	97.44%	0%	2.5%	0%
E5_min	Images non disponibles			
E5_moy	70.73%	12.2%	9.75%	6,91%
E5_max	79.45%	0	15.38%	5.1%
E6_min	Images non disponibles			
E6_moy	59.45%	2.7%	18.9%	18.9%
E6_max	79.5%	0	15.38%	5.1%

Table 4.14 classification des intensités de la base [EE base] après post traitement

Les remarques sur les résultats obtenus sur cette base sont les mêmes remarques constatées avec la base [HAM base].



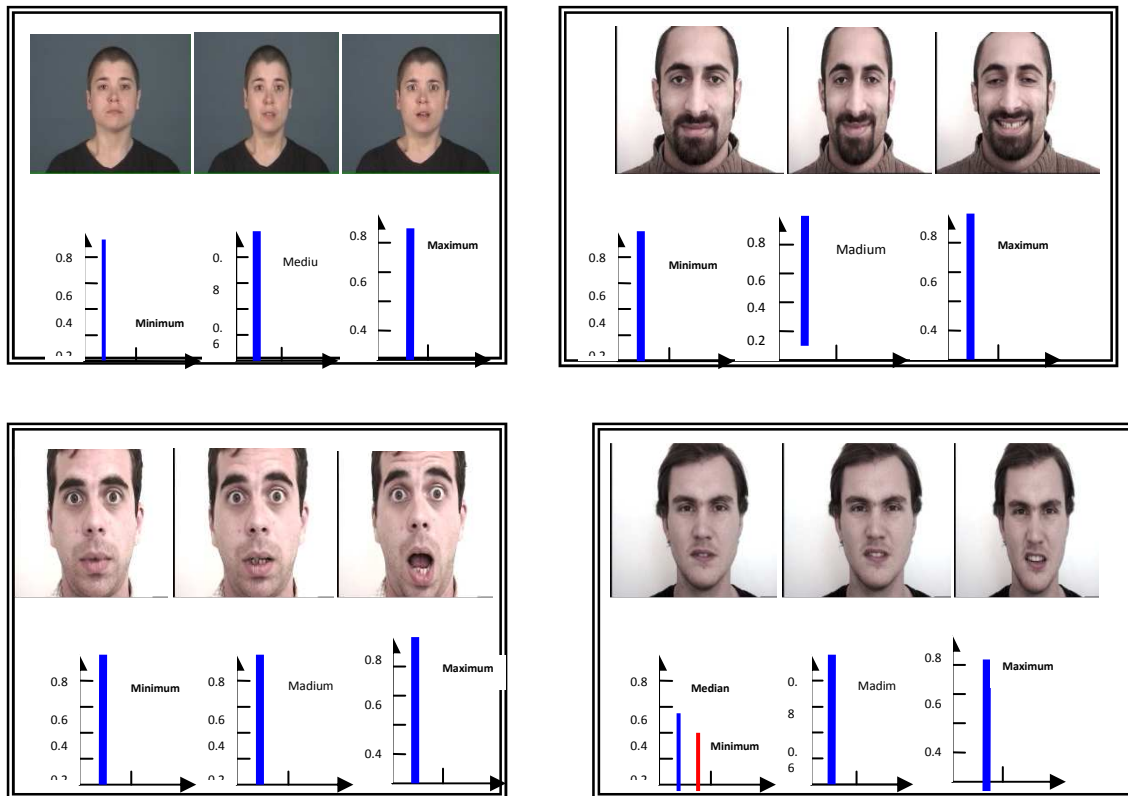


Figure 4.14. Exemples des intensités des six expressions faciales avec leurs masses d'évidence associées

Un autre résultat très intéressant concernant cette étude est la reconnaissance de nouvelles expressions faciales jusqu'ici non reconnues. Ces expressions présentent des sous classes des six expressions universelles. En effet en associant trois intensités (minimale, moyenne et maximale) pour chaque expression universelle, on déduit des sous classes de ces expressions. A chaque intensité d'expression est associée une réaction correspondante. La table 4.15 présente les sous classes reconnues des expressions faciales ainsi que les réactions correspondantes. Cette table présente un outil de base dans un processus de proposition d'un système expert.

Expressions Universelles	Intensités	Nouvelles expressions reconnues	Réactions Correspondantes
Joie	min	Bien être	Réagir positivement
	moy	Joie	Hurler avec le rire
	max	Bonheur	Sauter
Dégout	min	ennui	Faire des grimaces
	moy	Dégout	Utiliser des mots de dégoût
	max	Amertume	Vomir
Surprise	min	étonnement	Poser des questions
	moy	Surprise	Faire des grimaces
	max	Stupéfaction	Geler
Colère	min	ennui	Se Disputer
	moy	Colère	Crier
	max	Rage	Casser et tuer
Tristesse	min	Trouble	S'insulter

	Moy	Tristesse	pleurer
	max	Abatement	Depression / se suicider
Peur	min	anxiété	Faire des grimaces
	moy	Peur	Crier et pleurer
	max	Terreur	Se cacher et disparaître

Table 4.15 Nouvelles expressions reconnues suite à la quantification des six expressions universelles et les réactions correspondantes (Règles d'un système Expert)

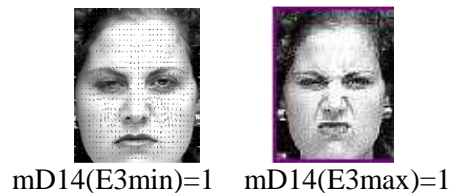
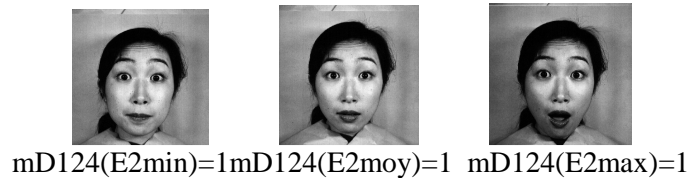


Figure 4.15. Quelques exemples des masses d'évidence associées aux différentes intensités estimées pour des images d'autres bases d'image.

4.8. Estimation de l'Intensité d'une Expression Faciale Inconnue

Un système d'analyse d'expressions faciales doit donc pouvoir non seulement reconnaître l'expression mais aussi estimer son intensité. La méthode présentée dans la section précédente quantifie une expression reconnue, hors il n'est pas toujours possible de reconnaître l'expression étudiée car seul l'ensemble des expressions universelles qui sont la joie, colère, dégoût, tristesse, Peur, Surprise est reconnu. Si le visage étudié présente une autre expression, le système de quantification développé ne peut être appliqué. D'un autre côté, il est parfois important de connaître l'intensité du bien être d'une personne ou son mal être sans trop s'intéresser à l'émotion. Enfin dans pas mal de cas une expression faciale est reconnue avec un doute par rapport à une autre expression.

Puisque l'intensité doit être estimée quelque soit l'expression, nous avons développé une autre méthode qui quantifie l'intensité d'une expression faciale inconnue. La méthode suit la même démarche que la méthode qui estime l'intensité d'une expression reconnue. La différence réside dans

le fait qu'avec cette méthode, seules les distances changeantes sont prises en considération dans le processus de fusion.

4.8.1. Application de la Théorie de l'Evidence dans le Contexte de l'Estimation d'Expressions Inconnues

Dans ce cas d'étude, le cadre de discernement Ω est égal $\{E_{\min}, E_{\text{moy}}, E_{\max}\}$ tel que :

- E_{\min} est l'intensité minimale de n'importe quelle expression ;
- E_{moy} est l'intensité moyenne ;
- E_{\max} est l'intensité maximale.

Une variable d'état symbolique est associée à chaque distance changée, le modèle présenté dans la section (4.5.2) sur la figure 4.6 est également utilisé dans cette méthode afin d'associer un état symbolique à chaque distance D_i et lui assigne une masse d'évidence de base.

Comme l'expression n'est pas reconnue, aucune distance n'est définie à priori pour caractériser au mieux l'expression étudiée. C'est pourquoi une simple formulation des états associés aux distances changées en termes d'expression avec intensité est donnée dans la table 4.16 suivante :

Expression	E_{\min}	E_{moy}	E_{\max}	$E_{\min} \cup E_{\text{moy}}$	$E_{\text{moy}} \cup E_{\max}$
V_i	min	moy	max	$\min \cup \text{moy}$	$\text{moy} \cup \max$

Table 4.16. Les différents états pris par une distance D_i et les expressions correspondantes avec intensités

A partir de cette table, nous pouvons déduire que si la variable d'état associée à une distance changée D_i est égale à « min », l'intensité de l'expression étudiée correspond à une expression avec une intensité minimale et avec une masse d'évidence égale.

De la même façon nous aurons :

$$V = \text{moy} ; mD(\text{moy}) = mD(E_{\text{moy}});$$

$$V = \text{max} ; mD(\text{max}) = mD(E_{\max});$$

$$V = \min \cup \text{moy} ; mD(\min \cup \text{moy}) = mD(E_{\min} \cup E_{\text{moy}});$$

$$V = \text{moy} \cup \max ; mD(\text{moy} \cup \max) = mD(E_{\text{moy}} \cup E_{\max}).$$

Pour être plus explicite, nous considérons les deux distances D_1 et D_2 de façon que :

$$V_1 = \text{moy} \text{ et } V_2 = \text{moy} \cup \max \Rightarrow mD_1(\text{moy}) = mD_1(E_{\text{moy}}),$$

$$mD_2(\text{moy} \cup \max) = mD_2(E_{\text{moy}} \cup E_{\max})$$

La fusion des deux données et donnée par la somme orthogonale comme suit:

$D_1 \setminus D_2$	$E_{\text{moy}} \cup E_{\max}$
E_{moy}	E_{moy}

L'intensité résultante est donc une intensité moyenne.

Les cas de conflit sont traités de la même façon que dans la section (4.6).

4.8.2. Résultats Obtenus sur la Base Dafex

Les performances du système de classification proposé sont évaluées sur tous les sujets de la base Dafex. Avant de donner les résultats finaux, et pour être plus explicite nous donnons un exemple des données extraites ainsi que les états associés pour un acteur de la base présentant les six expressions faciales avec trois intensités pour chaque expression.



Figure 4.16. Images d'un acteur montrant six expressions Colère, Dégout, Peur, Joie, Tristesse et Surprise respectivement avec les trois intensités max, moy et min respectivement pour chaque expression.

Les données extraites depuis les distances changées après assignation des états symboliques sont présentés dans la table 4.17:

Image	D1	D2	D3	D4	D5	Classification par TBM	Images étiquetées par un expert
A4joie	MAX		MOY U MAX	MAX	MAX	MAX	joie_MAX
A4Colère		MAX		MOY U MAX		MAX	Colère_MAX
A4dégout	MAX	MOY U MAX		MAX		MAX	Dégout_MAX
A4SUR	MAX	MAX		MAX		MAX	Sur_MAX
A4triste	MAX	MAX			MAX	MAX	Tristesse-MAX
A4peur	MOY U MAX	MAX		MOY U MAX		MAX	Peur_MAX
A4joie	MOY		MOY	MOY	MOY	MOY	Joie_MOY
A4Colère	MIN U MOY			MOY		MOY	Colère_MOY
A4dégout	MIN U MOY			MOY		MOY	Dégout_MOY
A4SUR	MOY	MAX		MOY		ERROR	Sur_MOY
A4triste	MOY	MOY			MOY	MOY	Tristesse- MOY

A4peur	MOY			MOY		MOY	Peur_ MOY
A4joie	MIN		MIN	MIN U MOY		MIN	Joie_MIN
A4Colère		MIN		MIN U MOY		MIN	Colère_ MIN
A4dégout	MIN			MIN		MIN	Dégout_ MIN
A4SUR	MOY	MOY		MOY		MOY	Sur_ MIN
A4Triste	MIN	MIN				MIN	Tristesse- MIN
A4Peur	MIN	MIN		MOY		ERROR	Peur_ MIN

Table 4.17 Classification des intensités basée sur la TBM

Les lignes présentent les différentes expressions avec différentes intensités, les cinq premières colonnes présentent les cinq distances faciales et les dernières colonnes présentent la classification effectuée par le système proposé et la classification par l'expert respectivement.

On peut constater que pour la plupart des cas la classification est faite correctement.

	Int.Min	Int.Moy	Int.Max
Intensité reconnue	56/78 71,79%	72/78 92,31%	59/78 75,64%
MinUMoy	17/78 21,79%		
MoyUMax			12/78 15,38%
Erreur	5/78 6,41%	6/78 7,69%	7/78 8,97%
Total Reconnu	93,58%	92,31%	91,02%
TOTAL	100%	100%	100%

Table 4.18. Taux de classification de l'intensité pour les bases Dafex et Hammal_caplier

Les lignes présentent les taux de classification réalisés en utilisant la méthode proposée et les colonnes présentent les intensités telles qu'elles sont étiquetées dans la base.

On peut constater que les taux de classification sont en général très proches, ils sont également plus importants que ceux obtenus avec la méthode précédente. Des taux assez importants sont assignés au doute entre intensités et enfin les taux d'erreur nous renseignent sur le taux de classification des intensités dans des classes autres que celles données par un expert.

Une autre constatation concerne le nombre de distances concernées dans le processus de décision quant à la quantification des expressions. Ce nombre est de deux ou trois distances comme il a été prouvé dans la méthode précédente. Les distances concernées dans cette étude pour chaque expression correspondent à un sous ensemble ou bien au même ensemble considéré dans la méthode précédente.

Il est évident que les taux obtenus avec l'intensité d'une expression inconnue sont meilleurs que ceux d'une expression connue. Le fait est si on considère par exemple l'image de la figure 4.17 (expression inconnue), uniquement deux distances changent tel que :

$$V1 = \min U \text{ moy et } mD1(\min U \text{ moy}) = mD1(E_{\min U} E_{\text{moy}}) = 1;$$

$$\begin{aligned}
V_2 = \text{moy} &\Rightarrow mD_2(\text{moy}) = mD_2(\text{Emoy}) = 1; \\
&\Rightarrow mD_1(\text{Emin} \cup \text{Emoy}) \oplus mD_2(\text{Emoy}) = mD_{12}(\text{Emoy}) = 1 \\
&\Rightarrow \text{Intensité} = \text{moy}.
\end{aligned}$$

Par contre, si on sait qu'il s'agit de la surprise (connue), trois distances sont prises en considération dans le processus de fusion D_1, D_2 et D_4 , et comme $V_4 = \min \Rightarrow$ la fusion nous donne erreur.



D'autres remarques importantes ont été constatées suite à cette étude, ces remarques confirment le fait que la méthode d'estimation de l'intensité d'une expression inconnue est plus significative que celle d'une expression connue.

Il a été noté que certaines émotions sont exprimées différemment d'un sujet à un autre (figure 4.18), mais malgré ça l'intensité a été donnée correctement car on ne considère pas certaines distances spécifiques pour chaque expression et en suivant l'évolution correspondante pour chaque expression (comme c'est le cas dans la première méthode), mais en considère la distance changée quelque soit son évolution (décroît ou accroît).



Par exemple sur la figure 4.18, on voit deux images avec la colère exprimée différemment. Dans l'image à gauche, la bouche est grande ouverte et dans l'image à droite les lèvres sont serrées, dans les deux cas l'intensité est max et c'est correct.



Figure 4.19 deux descriptions différentes de la peur avec intensité max.

Une autre raison est illustré sur la figure 4.19, ou nous constatons que parfois quand l'intensité diminue, le nombre des distances changées décroît également on ne peut donc considérer des distances spécifiques à chaque intensité puisque nous avons défini un modèle unique pour toutes les intensités.

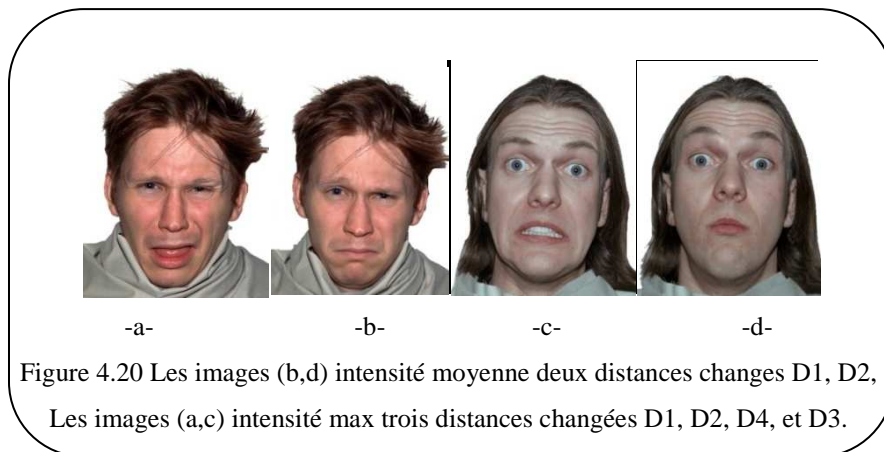


Figure 4.20 Les images (b,d) intensité moyenne deux distances changes D1, D2,
Les images (a,c) intensité max trois distances changées D1, D2, D4, et D3.

On Remarque également que d'une intensité à une autre les distances ne sont pas toujours les mêmes qui changent figure 4.21.

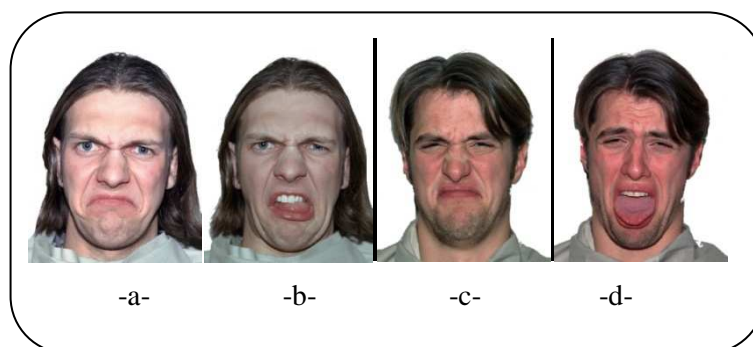


Figure 4.21 distances Changées avec intensités différentes pour la même expression
(a,c) int. moy(D1,D2,D5); (b,d) int; max D1,D2,D4

En conclusion on peut dire que plusieurs raisons nous incite à utiliser la méthode d'estimation de l'intensité d'une expression inconnue, c'est dans ce contexte que nous avons développé cette deuxième méthode.

4.9. Conclusion

Dans ce chapitre nous avons développé une nouvelle méthode afin d'estimer l'intensité d'une expression faciale. Pour ce faire, nous avons considéré les différentes déformations géométriques des traits permanent du visage qui sont les yeux, les sourcils et la bouche. Un certain nombre de distances spécifiques est considéré pour chaque expression. Un modèle est proposé pour chaque distance pour modéliser le degré d'évolution de cette distance. Toutes les distances changées sont ensuite fusionnées afin de donner une décision finale sur l'intensité de l'expression étudiée en utilisant la théorie de l'évidence si bien adaptée quand il s'agit d'informations issues de différentes sources. Une étape de post traitement est proposée en cas de conflit. Comme il n'est pas toujours évident de reconnaître l'expression faciale étudiée, nous avons développé une autre méthode qui estime l'intensité d'une expression faciale inconnue. La démarche proposée de cette méthode est la même que celle de la méthode précédente avec la seule différence de considérer uniquement les distances changées au lieu de considérer certaines distances spécifiques (même si certaines ne changent pas). Les résultats obtenus ont prouvé l'efficacité de la deuxième méthode par rapport à la première. Dans les deux cas, la théorie de l'évidence est employée vu que cette dernière est bien adaptée quand il s'agit de fusion de données non complète et issues de différentes sources dans le but de la réalisation d'une classification.

Dans le chapitre suivant, nous allons évaluer une nouvelle méthode de classification basée non sur des données statiques, mais sur des données dynamique

Chapitre 5

Classification des Expressions Faciales A Base d'Informations Vidéo En utilisant une Méthode de Data Mining

5.1. Introduction

La classification catégorielle des expressions faciales a déjà fait l'objet d'une étude présentée dans le chapitre II, cependant l'étude portait spécialement sur des images fixes en utilisant la théorie de l'évidence comme méthode d'analyse. Dans ce chapitre, l'étude porte sur des séquences vidéo en utilisant pour la première fois une technique qui n'a jamais été utilisée dans le domaine de l'analyse des expressions faciales, cette technique est une technique de Data Mining. Un état de l'art sur les méthodes de reconnaissance des expressions faciales a été présenté au chapitre II, donc nous nous limitons à présenter ici une simple synthèse des approches de suivi dans une séquence vidéo ensuite une introduction au Data Mining afin de prouver la nécessité de faire appel aux méthodes du Data Mining.

La technique de Data Mining utilisée est basée principalement sur l'extraction de connaissances depuis des séries temporelles. Ces connaissances sont définies sous forme de règles. L'idée est de fournir une nouvelle description des six expressions faciales dans un contexte dynamique en utilisant ces règles découvertes. Ensuite les utiliser pour classer de nouveaux exemples. L'objectif visé est de présenter à l'utilisateur des informations plus synthétiques ou des données particulièrement pertinentes dans le contexte qui l'intéresse. On parle d'extraction de *connaissances* car à partir d'un volume initial de données inintelligibles et inexploitable en l'état, le processus de fouille fourni des informations explicatives permettant à l'utilisateur d'appréhender plus précisément les phénomènes à l'origine de la production de ces données : on parle donc de *connaissances* ou de *pépites* (en analogie avec le terme *Data Mining*). Il ne s'agit pas d'obtenir un modèle exact d'évolution des données, mais plutôt d'y déceler des informations rares et particulièrement intéressantes.

Il existe deux domaines principaux de Data Mining, un domaine centré sur les données commerciales et l'autre centré sur les données scientifiques. La plupart des méthodes de Data Mining sont orientées commerce. Dans ce chapitre nous abordons et pour la première fois une de ces méthodes dans le domaine de l'analyse des expressions faciales.

L'Extraction de Connaissances à partir de Données (ECD), communément appelée Data Mining, est un domaine aujourd'hui très en vogue, pour ne pas dire à la mode. On la définit comme *"un processus non-trivial d'identification de structures inconnues, valides et potentiellement exploitables dans les bases de données [Fay96]"*. Cette définition est une des premières qui traite explicitement de l'ECD (Knowledge Discovery in Databases), par la suite plusieurs tentatives de redéfinition sont apparues pour mieux préciser le domaine mais aucune ne s'est réellement imposée. En tous les cas, à la lecture des différents documents qui traitent de l'ECD, on peut se dire que, finalement, cela fait plus de 30 ans qu'on le pratique avec ce qu'on appelle l'analyse de données et les statistiques exploratoires. Et on n'aurait pas complètement tort.

En réalité, ce n'est pas aussi simple, l'ECD possède des particularités qui sont loin d'être négligeables:

(1) des techniques d'analyse qui ne sont pas dans la culture des statisticiens, en provenance de l'apprentissage automatique (Intelligence artificielle) et des bases de données ;
(2) l'extraction de connaissances est intégrée dans le schéma organisationnel de l'entreprise. Ainsi, les données ne sont plus issues d'enquêtes ou de sondages mais proviennent d'entrepôts construits sciemment pour une exploitation aux fins d'analyse, le DATAWAREHOUSE.
(3) enfin, dernier élément important, le traitement des données sort de plus en plus des sentiers battus en traitant, non seulement des fichiers plats "individus x variables", mais également des données sous forme non structurée, le texte, depuis un bon moment déjà, mais aussi les images et la vidéo. Cette orientation attribue une place primordiale à l'appréhension et la préparation des données.

5.2. Suivi dans les Séquences Vidéos

L'estimation de mouvement est un problème fondamental en traitement d'image appliqué à des séquences vidéo. Dans ce domaine, de très nombreuses méthodes ont été (et continuent d'être) proposées. On peut distinguer trois approches principales.

Tout d'abord, les méthodes différentielles (differential methods) s'appuient sur l'équation de contrainte du mouvement apparent issue d'un développement de Taylor de l'équation 5.1.

Parmi les différentes variantes proposées, certaines sont basées sur des dérivées du premier ordre avec ou sans contrainte de régularisation sur le champ de vecteurs vitesse [LUC84][Hor81]. Il est également possible d'utiliser des dérivées d'ordre supérieur. De plus,

il est possible de réduire la sensibilité des calculs numériques en utilisant des contraintes de régularisation locales [URA88] ou globales [NAG87].

Ensuite, les méthodes de mise en correspondance (block-matching techniques) tentent d'estimer le mouvement d'une région de l'image courante en minimisant la distance avec une région candidate de l'image suivante. En général, cette mesure de similarité est obtenue par une somme des différences inter-pixels au carré (Sum-Squared Difference - SSD). Comme il est évident qu'un test exhaustif de toutes les régions possibles est très coûteux en temps de calcul, de nombreux algorithmes «rapides» ont été proposés. Anandan préconise une approche de type multi-résolution en utilisant une décomposition pyramidale de l'image [ANA89]. Le déplacement est estimé itérativement en commençant par le niveau de résolution le plus grossier. Dans [KOG81], Koga propose une méthode de recherche rapide (logarithmic search) permettant d'estimer le déplacement d'un bloc en suivant la direction de moindre déformation. Dans le même esprit, on peut également citer la technique de recherche en trois étapes utilisée dans le codeur vidéo H.263 [ITU95].

Enfin, les méthodes fréquentielles utilisent des bancs de filtres passe-bande permettant de décomposer le signal d'entrée selon l'échelle, la vitesse et l'orientation. Dans [HEE88], Heeger analyse l'énergie à la sortie de filtres de Gabor pour estimer les vitesses. Dans [FLE90], Fleet et Jepson préconisent d'utiliser plutôt la phase des signaux de sortie car elle est beaucoup plus stable que l'amplitude.

Pour notre application, le suivi doit tout d'abord être précis. Nous avons donc écarté les méthodes de mise en correspondance car elles ne permettent d'estimer que des déplacements entiers (ou demi-entiers pour les techniques mises en œuvre dans les codeurs MPEG-1). De plus, les algorithmes de recherche rapide qu'elles utilisent conduisent fréquemment à des minima locaux. D'après l'étude très détaillée menée dans [BAR94], les techniques les plus fiables et les plus précises sont la méthode différentielle du premier ordre de Lucas et Kanade et la méthode de phase de Fleet et Jepson. Cependant, il est à noter que la méthode proposée par Fleet et Jepson est beaucoup plus lente car elle nécessite un grand nombre de filtrages. Finalement, comme notre algorithme doit être rapide, nous avons donc opté pour la méthode de Lucas et Kanade dont le principe général est exposé dans la partie suivante.

5.3.1 L'algorithme de Lucas-Kanade

La méthode d'estimation de mouvement que nous utilisons est basée sur l'algorithme de flux optique développé par *Lucas et Kanade* dans [LUC84]. Dans cette méthode, on suppose

que le voisinage du point suivi dans l'image I_t se retrouve dans l'image suivante I_{t+1} par une translation :

$$I_t(x - d(x)) = I_{t+1}(x) \quad (\text{eq 5.1})$$

Où $d(x)$ est le vecteur déplacement du pixel de coordonnée x (x est un vecteur). La figure 5.1 illustre cette égalité dans le cas d'un signal mono-dimensionnel.

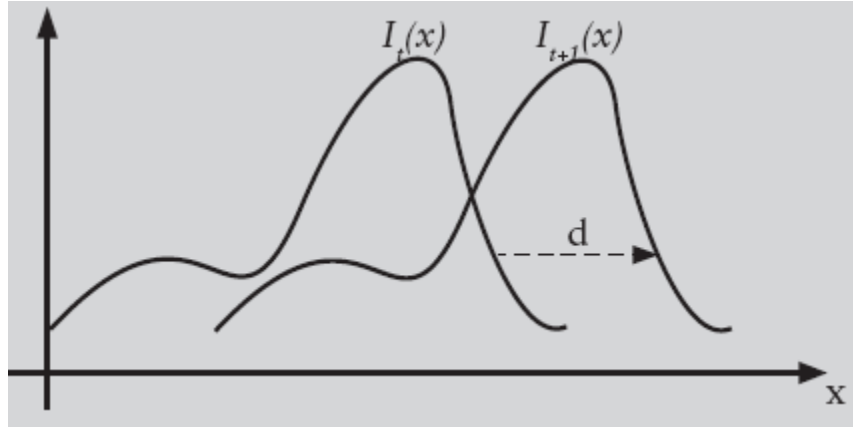


Figure 5.1 Un voisinage dans l'image I_t peut être retrouvé dans l'image I_{t+1} par une translation de vecteur d

Le but est de retrouver dans l'image suivante la région la plus ressemblante à R . On note $I_t(x)$ et $I_{t+1}(x)$ les valeurs des niveaux de gris dans ces 2 images. Pour cela, il faut minimiser une fonction coût égale à la somme des différences inter-pixels au carré :

$$\mathcal{E}(d(x)) = \sum_{x \in R} [I_t(x - d(x)) - I_{t+1}(x)]^2 w(x) \quad (\text{eq 5.2})$$

Où $w(x)$ est une fonction de pondération. En général, $w(x)$ est constante et vaut 1. Mais elle peut également prendre une forme gaussienne si on veut donner plus d'importance au centre de la fenêtre. La minimisation de la fonction \mathcal{E} est réalisée de manière itérative. On note $d^i(x)$ la valeur du déplacement total calculée au début de l'itération i . Le déplacement final $d^{i+1}(x)$ peut alors s'exprimer de la manière suivante :

$$d^{i+1}(x) = d^i(x) + \Delta d^i(x) \quad (\text{eq 5.3})$$

Où $d^i(x)$ est le déplacement incrémental à déterminer avec une précision sub-pixel. Dans toute la suite de cette section, on suppose que le voisinage considéré ne subit pas de déformation. Par conséquent, la valeur du déplacement est la même pour tous les pixels de R . L'équation précédente peut donc être écrite plus simplement de la manière suivante :

$$d^{i+1} = d^i + \Delta d^i \quad (\text{eq 5.4})$$

Ainsi, on peut écrire :

$$I_t(x - d^{i+1}) = I_t(x - (d^i + \Delta d^i)) \quad (\text{eq 5.5})$$

En utilisant un développement de Taylor au premier ordre, cette équation devient :

$$I_t(x - d^{i+1}) \approx I_t(x - d^i) - g^T \Delta d^i \quad (\text{eq 5.6})$$

Où g est le vecteur gradient :

$$g = \begin{pmatrix} \left(\frac{\partial I_t}{\partial x} \right)_{x=d^i} \\ \left(\frac{\partial I_t}{\partial y} \right)_{x=d^i} \end{pmatrix} \quad (\text{eq 5.7})$$

En tenant compte de cette linéarisation, l'expression de la fonction coût \mathcal{E} (équation 5.2) s'écrit :

$$\begin{aligned} \mathcal{E}(d) &\approx \sum_{x \in R} [I_t(x - d^i) - g^T \Delta d^i - I_{t+1}(x)]^2 w(x) \\ &= \sum_{x \in R} [h - g^T \Delta d^i]^2 w(x) \end{aligned} \quad (\text{eq 5.8})$$

Où $h = I(x - d^i) - I_{t+1}(x)$.

Il s'agit d'obtenir la valeur de Δd^i qui minimise \mathcal{E} . On dérive donc $\mathcal{E}(d)$ par rapport à Δd^i :

$$\frac{\partial \mathcal{E}}{\partial \Delta d^i} = \sum_{x \in R} (h - g^T \Delta d^i) g w(x) \quad (\text{eq 5.9})$$

Or, $(g^T \Delta d^i) g = (g g^T) \Delta d^i$. donc, l'équation (3.28) peut s'écrire :

$$\frac{\partial \mathcal{E}}{\partial \Delta d^i} = 2 \left(\sum_{x \in R} h g w(x) \right) - 2 \left(\sum_{x \in R} g g^T w(x) \right) \Delta d^i \quad (\text{eq 5.10})$$

L'annulation de cette dérivée conduit à l'égalité suivante :

$G \Delta d^i = e$	(eq 5.11)
--------------------	-----------

Avec :

$$\begin{cases} G = \sum_{x \in R} g g^T w(x) \\ e = \sum_{x \in R} (I_t(x - d^i) - I_{t+1}(x)) g w(x) \end{cases} \quad (\text{eq 5.12})$$

L'équation (5.11) est la relation fondamentale de l'algorithme de **Lucas-Kanade**. Pour tout couple d'images adjacentes, la matrice G peut être calculée à partir de la première image en calculant le gradient de la luminance sur le voisinage du point à suivre. D'autre part, le vecteur e est obtenu en multipliant ce gradient par la différence entre les 2 fenêtres d'observation. Finalement, le déplacement incrémental recherché Δd^i est la solution du système (5.11).

5.2.2 Résultats de l'Algorithme de Lucas-Kanade

Au début du processus, on fixe $d^0 = [0,0]^T$. Puis, le déplacement total est calculé en plusieurs itérations. A chaque fois, l'équation (5.11) permet de déterminer le déplacement incrémental à effectuer. La figure 5.2 illustre le déroulement de l'algorithme en présentant les positions successives d'un point suivi. Lorsque le déplacement calculé Δd^i devient inférieur à un seuil ou que le nombre d'itérations dépasse une certaine limite, le processus s'arrête.



Figure. 5.2 : Suivi d'un point caractéristique par l'algorithme de Lucas Kanade

A partir de la position d'un point dans l'image à l'instant t (O), l'algorithme de Lucas-Kanade estime la position correspondante dans l'image suivante (□). Les positions intermédiaires calculées à chaque itération sont symbolisées par des petits points blancs.

5.3. Introduction au Data Mining

5.3.1 Définition du Data Mining

Le data-mining est un processus de **découverte** de règle, relations, corrélations et/ou dépendances à travers une grande quantité de données, grâce à des méthodes statistiques, mathématiques et de reconnaissances de formes.

5.3.2. Origine et Emergence du Concept de Data Mining

Historiquement, le Data Mining est très jeune. Le concept apparaît en 1989 sous un premier nom de KDD (Knowledge Discovery in Databases) Extraction de Connaissances à

partir des Données ECD, ce n'est qu'en 1991 qu'apparaisse pour la première fois le terme de DataMining ou « minage / fouille des données ».

Comme l'expliquent fort bien Michael Berry et Gordon Linoff, ce concept – tel qu'on l'entend aujourd'hui, et surtout tel qu'on l'applique dans les services marketing – est étroitement lié au concept du « *one-to-one relationship* ». C'est à dire la personnalisation des rapports entre l'entreprise et sa clientèle.

5.3.3. Raisons du Développement

« Data Mining », dans sa forme et sa compréhension actuelle, comme champ à la fois scientifique et industriel émerge non pas par hasard mais parce que c'est le résultat de la combinaison de nombreux facteurs à la fois technologiques, économiques et même sociopolitiques. On peut voir le « Data Mining » comme une nécessité imposée par le besoin économique des entreprises de valoriser les données qu'elles accumulent dans leurs bases, le développement de la technologie de l'information qui a induit un faible coût de stockage de données, la saisie automatique de transaction (code bar, click, données de localisation GPS). En effet, le développement des capacités de stockage, l'augmentation de la puissance de calculs des ordinateurs et les vitesses de transmission des réseaux rendent l'extraction de la connaissance à partir de grandes bases de données possible.

5.3.4. Exemples d'Applications

Les techniques de « Data Mining » ont été employées avec beaucoup de succès dans de grands secteurs d'application : la gestion de la relation client (GRC) ou « customer relationship management », gestion des connaissances « knowledge management », indexation de documents, les finances (minimisation de risque financiers), les assurances (détection de fraudes), les banques (prêts), les grandes surfaces (Organisation de rayonnage), e-commerce, Internet (spam, optimisation des sites web, sécurité et détection d'intrusion ou d'anomalies en général), la Bioinformatique (Analyse du génome, mise au point de médicaments), prévisions météorologiques, reconnaissance de la parole etc...

Aucun domaine d'application n'est a priori exclu car dès que nous sommes en présence de données empiriques, le « Data Mining » peut rendre de nombreux services.

5.3.5. Principe du Data Mining

Plus qu'un domaine ou une théorie floue, le Data Mining est avant tout un cadre précisant la démarche à suivre pour exploiter les données, quelles que soient leur formes, en vue d'en extraire de la connaissance. Plusieurs méthodes existent aujourd'hui pour mener à bien cette fouille, de méthodes purement calculatoires (Agrawal and Srikant, 1994) à des méthodes beaucoup plus visuelles (Blanchard, 2005). Toutes ces méthodes partagent en général les mêmes étapes (Fayyad et al., 1996; Blanchard, 2005) qui sont :

1. Acquisition des données ;
2. Sélection de données ;
3. Prétraitement: La plupart du temps, on travaille sur des données brutes présentant donc certaines imperfections. Il s'agit donc de les rendre propres à la « consommation » par le processus d'ECD. Il concerne les points suivants :
 - Mise en forme des données entrées selon leur type (numérique, symbolique, image, texte, son) ;
 - Nettoyage des données (détecter semi- automatiquement et corriger les valeurs bizarres) ;
 - Imputation de valeurs manquantes ;
 - Sélection des données,
 - Élimination des doublons,
 - Élimination des valeurs aberrantes,
 - Transformation des variables,
 - Création de nouvelles variables,
 - Encodage des données (selon l'algorithme d'apprentissage) ,
 - Normalisation des variables réelles (soustraire moyenne, diviser par écart type..),
 - Optionnellement discrétisation de variables numériques choisies ;
4. Identifier le type de problèmes (Segmentation, Classification, etc...) et choisir un algorithme. Il est temps alors de « faire parler » les chiffres. Modèles et typologies sont alors mis en œuvre, afin de produire des réponses au(x) problème(s) posé(s). Cette phase est souvent décrite comme le cœur de la démarche de DataMining. C'est elle, en tout cas, qui a bénéficié d'une grande partie de l'intérêt porté à la discipline ;

5. Evaluer les performances de l'algorithme : Cette phase d'évaluation permet de choisir la meilleure des solutions ;
6. Déployer l'application : (La mise en production) cette phase est le prolongement concret de l'étude qui vient d'être menée. Elle met en application les résultats proposés, mais ne doit jamais être oubliée une fois effectuée. Il faut encore prendre du recul sur l'action engagée, et la décortiquer une fois qu'elle a produit des effets. Ces résultats, positifs et négatifs, permettront d'améliorer les futurs modèles, et à ce titre devront être réinjectés dans le dispositif. Ce qui boucle le cercle vertueux du Data Mining.

5.3.6. Algorithmes :

Résoudre une problématique avec un processus de Data Mining impose généralement l'utilisation d'un grand nombre de méthodes et algorithmes différents. On peut distinguer 3 grandes familles d'algorithmes :

5.3.6.1 Méthodes Supervisées

L'objectif de ce type de méthodes est à partir d'un ensemble d'observations $\{x_1, \dots, x_n\} \in X^d$ et de mesures $\{y_i\} \in Y$, on cherche à estimer les dépendances entre l'ensemble X et Y.

Exemple : on cherche à estimer les liens entre les habitudes alimentaires et le risque d'infarctus. « xi » est un patient décrit par « d » caractéristiques concernant son régime et « yi » une catégorie (risque, pas risque). On parle d'apprentissage *supervisé* car les yi permettent de guider le processus d'estimation.

Leur raison d'être est d'expliquer et/ou de prévoir un ou plusieurs phénomènes observables et effectivement mesurés. Concrètement, elles vont s'intéresser à une ou plusieurs variables de la [base de données](#) définies comme étant les *cibles* de l'analyse.

Différentes techniques existent dans cette branche parmi lesquelles , nous citons : les techniques à base d'arbres de décision, techniques statistiques de régressions linéaires et non linéaires au sens large ([Régression linéaire](#), [Régression linéaire multiple](#), [Régression logistique](#) binaire ou multinomiale, [Analyse discriminante linéaire](#) ou quadratique, ...), techniques à base de [Réseau de neurones](#) (perceptron mono ou multicouches avec ou sans rétro propagation des erreurs, [réseaux à fonction radiale de base](#)..), techniques à base d'[algorithme génétique](#), techniques à base d'[Inférence bayésienne](#) ([Réseau bayésien](#)), [raisonnement par cas](#), [filtrage collaboratif](#), [exploration Hypercubique](#) etc..

Ce type de méthodes est utilisé principalement dans la détection de fraude, marketing téléphonique, changement d'opérateurs téléphonique etc...

5.3.6.2 Méthodes Non-Supervisées

Comme seules les observations $\{x_1, \dots, x_n\} \in X^d$ sont disponibles, l'objectif est de décrire comment les données sont organisées et d'en extraire des sous-ensemble homogènes.

Elles permettent de travailler sur un ensemble de [données](#) dans lequel aucune des [données](#) ou des variables à disposition n'a d'importance particulière par rapport aux autres.

Différentes techniques sont utilisées dans cette branche telles que : les techniques à base de [Réseau de neurones](#) : carte de Kohonen ..., techniques utilisées classiquement dans le monde des statistiques : [classification ascendante hiérarchique](#), [k-means](#) et les nuées dynamiques ([Recherche des plus proches voisins](#)), les [classification mixtes](#) (Birch...), les [classifications relationnelles](#), techniques dites de **recherche d'associations** et analyse des liens: algorithmes [a priori](#), [Carma](#), ...

Ces méthodes sont utilisées principalement pour dégager d'un ensemble d'individus des groupes homogènes ([typologie](#)), pour construire des normes de comportements et donc des déviations par rapport à ces normes (détection de fraudes nouvelles ou inconnues à la carte bancaire, à l'[assurance maladie](#)...), pour réaliser de la compression d'informations (compression d'image), l'identification de document similaires. Elles sont spécialement appliquées dans l'identification de segments de marchés et étude du panier de la ménagère, et sont également appliquées à des problèmes d'analyse de parcours de navigation sur les sites web.

5.3.6.3. Méthodes Semi-supervisées

Parmi les observations $\{x_1, \dots, x_n\} \in X^d$, seulement un petit nombre d'entre elles ont un label $\{y_i\}$. L'objectif est le même que pour l'apprentissage supervisé mais on aimerait tirer profit des observations non labélisées.

Différentes techniques sont utilisées dans cette branche parmi lesquelles nous citons : les méthodes bayésiennes, Séparateur à Vastes Marges, etc...

Ces méthodes sont utilisées dans la discrimination de pages Web, le nombre d'exemples peut être très grand mais leur associer un label est coûteux.

5.3.7 Principales Tâches de l'Apprentissage

- La classification : (binaire ou multi classe) consiste à examiner les caractéristiques d'un objet et lui attribuer une classe, la classe est un champ particulier à valeurs discrètes.
- L'estimation : consiste à estimer la valeur d'un champ à partir des caractéristiques d'un objet. Le champ à estimer est un champ à valeurs continues. L'estimation peut être utilisée dans un but de classification. Il suffit d'attribuer une classe particulière pour un intervalle de valeurs du champ estimé.
- La prédiction : consiste à estimer une valeur future. En général, les valeurs connues sont historiées. On cherche à prédire la valeur future d'un champ. Cette tâche est proche des précédentes. Les méthodes de classification et d'estimation peuvent être utilisées en prédiction.
- Extraction de règles d'association : consiste à déterminer les valeurs qui sont associées. L'exemple type est la détermination des articles (le poisson et le citron ; la baguette et le camembert et les lentilles, ...) qui se retrouvent ensemble sur un même ticket de supermarché. Cette tâche peut être effectuée pour identifier des opportunités de vente croisée et concevoir des groupements attractifs de produit. C'est une des tâches qui nécessite de très grands jeux de données pour être effective.
- La segmentation : consiste à former des groupes (clusters) homogènes à l'intérieur d'une population. Pour cette tâche, il n'y a pas de classe à expliquer ou de valeur à prédire définie a priori, il s'agit de créer des groupes homogènes dans la population (l'ensemble des enregistrements). Il appartient ensuite à un expert du domaine de déterminer l'intérêt et la signification des groupes ainsi constitués. Cette tâche est souvent effectuée avant les précédentes pour construire des groupes sur lesquels on applique des tâches de classification ou d'estimation.

5.3.8 Modèles du Data Mining

Deux modèles sont associés au Data Mining, les modèles descriptifs et les modèles prédictifs.

5.3.8.1. Modèles Descriptifs

Ces modèles suivent une démarche exploratoire ou on espère découvrir des relations jusque là inconnues dans les données. Cette démarche s'appuie donc sur un apprentissage non supervisé.

5.3.8.2 Modèles Prédicatifs

Ces modèles suivent une démarche prédictive ou on sait ce qu'on veut prédire. Cette démarche s'appuie donc sur un apprentissage supervisé (ou semi-supervisé). C'est dans la modélisation prédictive qu'il y a le plus à gagner.

Chaque type de modèles est utilisé dans le but de réaliser certaines tâches(classification, clustering, règles d'association..).

Une question importante, dans le domaine du « Data Mining », est de pouvoir répondre du choix de l'outil approprié en regard du problème à résoudre. Selon le type de problème, il existe de nombreuses méthodes de « Data Mining » concurrentes. Un consensus général semble se dégager pour reconnaître qu'aucune méthode ne surpasse les autres car elles ont toutes leurs forces et leurs faiblesses spécifiques. En tout état de cause, un fait important communément admis est : « Il n'existe pas de méthode supérieure à toutes les autres ». Il semble plus avantageux de faire coopérer des méthodes pour avoir des résultats optimaux. Par conséquent, à tout jeu de données et tout problème correspond une ou plusieurs méthodes. Le choix se fera en fonction :

- de la tâche à résoudre,
- de la nature et de la disponibilité des données,
- des connaissances et des compétences disponibles,
- de la finalité du modèle construit. Pour cela, les critères suivants sont importants : complexité de la construction du modèle, complexité de son utilisation, ses performances, sa pérennité, et, plus généralement, de l'environnement ou le domaine.

5.4 Notre Contribution

L'objectif de ce travail est de classifier une expression faciale affichée à travers une séquence vidéo. Une séquence vidéo correspond à un enregistrement de plusieurs images par unité de temps (généralement par seconde), elle est donc composée de plusieurs images, le traitement de toutes ses images se fait par une détection des points caractéristiques du visage de la première image, ensuite ses points seront suivis tout le long de la séquence d'images. Le résultat de ce suivi est un ensemble de **séries chronologiques** qui fournissent des informations supplémentaires susceptibles de rendre la classification plus robuste et plus rapide. Chaque série contient des centaines voir des milliers (dépend de la durée de l'enregistrement) de valeurs.

MPEG-4 a proposé une description linguistique des six expressions faciales dans un contexte statique. L'objectif est de découvrir de nouvelles connaissances depuis le comportement dynamique de séries temporelles afin de procurer une nouvelle description de ces expressions dans un contexte dynamique. En effet, en introduisant le « temps », la description des expressions faciales doit être enrichie par de nouvelles règles qui considèrent ce nouveau critère.

La notion d' « **Ordre** » est une caractéristique de base introduite par le temps. L'idée est d'exploiter cette caractéristique afin de découvrir des rapports temporels entre les différentes séries chronologiques associées aux différentes modifications qui interviennent sur le visage. Ainsi dans l'ordre résultant des événements, on s'intéresse à **l'ordre de changement des distances caractéristiques**. En considérant cet ordre, des règles concernant les relations temporelles pour chaque expression peuvent être dérivées.

Il existe plusieurs modèles de logique temporelles. Dans les logiques temporelles avec datation, on distingue les logiques temporelles avec datation par points et par intervalles. La datation par points correspond à la représentation du temps par une succession de dates ponctuelles, alors que la datation par intervalles représente le temps par des intervalles entre deux dates. Il existe plusieurs logiques par intervalles dont celle d'Allen (1983), de Mac Dermott (1982) et de Shoham (1987). La logique par points implique des relations d'ordre sur les éléments alors que la logique par intervalles implique des relations d'ordre sur les intervalles disjoints et les relations de type ensembliste d'inclusion et de chevauchement.

5.4.1 Algèbre d'Intervalles d'Allen

Allen [ALL83] a proposé un cadre d'algèbre d'intervalles pour représenter l'information temporelle hiérarchique et probablement indéfinie et incomplète. Cette représentation diffère de la représentation basée sur (timestamps) des horodateurs, car elle permet de définir des relations relatives et à différents niveaux de granularité. Les événements sont représentés par des intervalles de temps (contrairement aux points de temps). Il y a treize relations de base entre les intervalles de temps. Les relations de base sont disjointes et exhaustives (Figure 5.3).

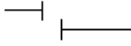

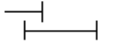
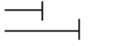
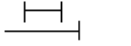
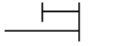

Relation	Symbol	Inverse	Meaning
X before Y	b	bi	
X meets Y	m	mi	
X overlaps Y	o	oi	
X starts Y	s	si	
X during Y	d	di	
X finishes Y	f	fi	
X equals Y		eq	

Figure 5.3 Les relations de base entre les intervalles temporelles [ALL83].

Dans ce travail nous ne sommes pas intéressés à diviser la série chronologique en intervalles et marquer chaque intervalle pour rechercher des similitudes entre ces intervalles, mais nous avons considéré que chaque série chronologique constitue un seul et unique intervalle et on cherche des similitudes entre les différentes séries qui représentent eux aussi des intervalles à part. Par exemple, en cas de l'expression de joie, chaque série sera donc qualifiée par la « série 1 (D3) » précède la « série 2 (D1) », en cas de colère la « série 3 (D4) » recouvre la « série 4 (D1) » etc.....

Avant d'appliquer la démarche du Data Mining dans le contexte de classification dynamique des expressions faciales, nous présenterons la méthode de suivi des points caractéristiques.

5.5 La Démarche Data Mining dans le Contexte de Classification Dynamique des Expressions Faciales :

Comme nous l'avons définie dans la section (5.2.5), le procédé du Data Mining est composé de six étapes : (1) acquisition des données d'entrées, (2) sélection de données, (3) pré traitement, (4) extraction de règles, (5) interprétation et évaluation des résultats, (6) déploiement.

5.5.1 Acquisition des Données d'Entrées

Les données d'entrées sont acquises depuis des bases d'images émotionnelles. Ces bases d'images comptent plusieurs acteurs (de 20 à plus de 1000) et chaque acteur affiche six expressions faciales, dans pas mal de fois, chaque expression est donnée avec différentes intensités. Pour chaque expression, chaque acteur est enregistré afin de produire des séquences vidéo. Enfin chaque vidéo produit plusieurs images généralement de 10 à 400 images.

5.5.2 Sélection des Données Entrées

Chaque image faciale de chaque vidéo propre à un acteur est caractérisée par cinq distances biométriques. Ces distances sont les mêmes distances définies dans le chapitre II. Elles représentent les distances entre les 18 points caractéristiques localisés depuis les traits permanents sur la première image de chaque vidéo. Ces points sont détectés par les méthodes présentés dans le chapitre I. Ces points sont ensuite suivis (Trackés) afin de calculer les nouvelles distances de la suite des images de la séquence vidéo en utilisant l'algorithme de Lucas et Kanade [LUC83].

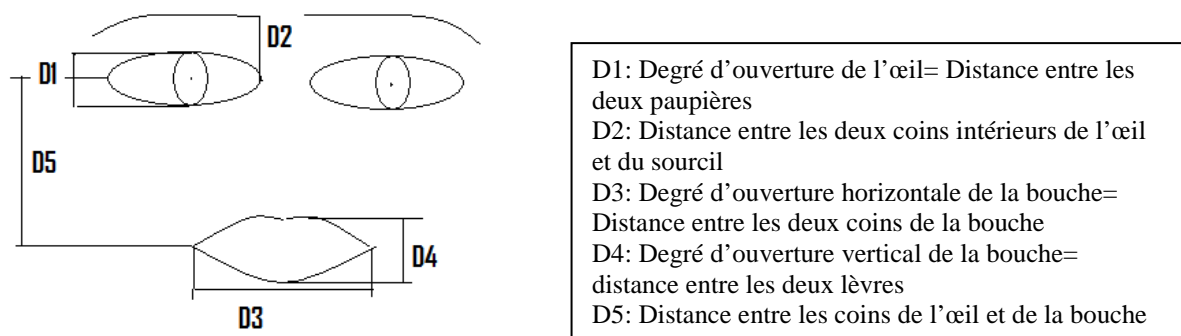


Figure 5.4. Les points caractéristiques du visage et les distances biométriques.

Toutes ces mesures sont calculées pour chaque acteur, pour chaque expression faciale et enfin pour chaque intensité d'expression, elles sont collectées à travers le temps c'est-à-dire chaque seconde. Le résultat est une multitude de séries chronologiques qui constitue une masse importante de données. Ces données présentent une source importante de connaissances potentielles qui est très intéressante pour un utilisateur.

5.5.3 Pré Traitement

Cette étape comprend des opérations de réduction de dimension, nettoyage de données (comme élimination du bruit, et des valeurs aberrantes...), transformation de données...

Dans cette étude, le type de données est quantitatif car les données consistent à un ensemble de séries chronologiques. Dans notre cas, il est préférable de changer le type des données quantitatif en type de données qualitatif afin de ne plus traiter des données numériques chronologiques en utilisant la logique temporelle d'Allen. Afin de réaliser cette transformation de type, nous considérons l'algorithme(1) suivant :

5.5.3.1. Algorithme de Transformation

Pour chaque expression faciale E_i : $i=1..k$

Pour chaque acteur S_j : $j=1..N$

1. Trouver la première distance qui change (Attribut) (Décroit ou accroît);
2. Trouver la seconde distance qui change (Attribut);
3. Trouver la troisième distance qui change (Attribut);

E_i est une expression faciale et k est le nombre d'expressions faciales; S_j est un sujet ou acteur et N est le nombre d'acteurs dans la base d'images.

Après application de cet algorithme sur l'ensemble des données recueillies, le nombre de ces données est réduit suite à cette transformation de type.

Par exemple, pour un acteur qui enregistre une vidéo d'expression de 200 images, on aurait calculé cinq distances pour chaque image donc ($5 \times 200 = 1000$) distances pour toute la vidéo. En appliquant l'algorithme (1), trois variables qualitatives remplacent les 1000 variables quantitatives. Chaque variable peut prendre une des trois valeurs suivantes : "1", "2"

et “3”. Uniquement les trois premiers ordres sont considérés car en général, trois distances changent lors de la production d’une expression sur un visage, (parfois moins et parfois plus).

Avant transformation						Après Transformation					
S1	D1	D2	D3	D4	D5	S1	D1	D2	D3	D4	D5
Image1							3	/	1	2	/
Image2											
.....						
Image 200											

1= La première distance qui change est D3;

2=la deuxième distance qui change est D4 et

3=la troisième distance qui change est D1.

5.5.4 Extraction de Règles (connaissances)

Pour extraire des règles ou des connaissances, une méthode doit être appliquée. Il existe des méthodes d’apprentissage très robustes qui couvrent les exemples d’une manière efficace et compacte. Parmi ces méthodes il y a la méthode « Induction de règles » appelé également « Modèle de règles basées classification »

La méthode « Règles d’induction » est une méthode qui a été introduite par Hunt [HUN62] dans son concept de système d’apprentissage « CLS » en 1962, ensuite elle a été étendue pour manipuler des données numériques ; Elle a été utilisée également par Ross Quinlan [QUI79] pour son système ‘Dichotomiser Itératif’ (ID3) en 1979. Depuis lors elle a été largement copiée et est la base d’une variété de Règles d’inductions commerciales [QUI86], [QUI87].

Quelques paradigmes principaux de règles d’induction sont les "Règles d’Association". L’association peut être entre deux ou plus de deux attributs, le but est de trouver les associations qui se produisent raisonnablement souvent, plus souvent que la chance suggérerait. La méthode de règles d’association est une méthode de recherche populaire pour découvrir des relations intéressantes entre variables dans de grandes bases de données. Elle est considérée comme potentiellement utile pour la découverte de connaissances puisque ses règles sont facilement comprises par l’être humain. Selon Ross, les règles d’association les plus simples qui impliquent un attribut seulement dans la partie condition, fonctionnent souvent bien dans la pratique avec des données réelles. L’idée de l’algorithme « OneR »

(règle à un-attribut) est de trouver l'attribut qui peut être utilisé pour classifier une nouvelle observation avec un minimum d'erreur de prévision.

5.5.4.1 Principe de L'Algorithme « Une Règle Basée Classification »

En considérant un ensemble d'exemples S , où chaque exemple est composé de paramètres observés et une classification correcte, le problème est de trouver le meilleur ensemble de règles RS_{best} ou le taux d'erreur sur les nouvelles observations $Err_{\text{true}}(RS_{\text{best}})$ est minimale.

Algorithme (2):

Pour chaque Attribut A :

Pour chaque valeur V de l'attribut considéré, créer une règle:

1. compter combien de fois chaque classe apparaît
2. Trouver la classe la plus fréquente, c
3. Construire une règle "Si $A=V$ alors $C=c$ "

Calculer le taux d'erreur de cette règle

Sélectionnez l'attribut dont les règles produisent le plus bas taux d'erreur

Remarque : Si A Alors B et C est une règle :

Exactitude de la règle (Rule's accuracy) = Le nombre de cas pour lesquels la partie Si est vraie (Condition A).

5.5.4.2. Procédure d'Apprentissage

En se donnant un bon ensemble d'apprentissage, la plupart des cas reflétant la réalité sont représentés ce qui permet de produire un ensemble exhaustif (ou presque) de règles. C'est pourquoi, différents exemples issues de deux bases d'images (Dafex [Dafex base] et Cohn et Kanade [COH base]) sont considérés dans cette phase. 10 acteurs présentant l'expression de joie, 6 acteurs présentant l'expression de dégoût, 13 acteurs présentant l'expression de colère, 18 acteurs présentant l'expression de tristesse issus de la base Cohn et Kanade, ainsi que 8 acteurs présentant les six expressions faciales issus de la base Dafex sont considérés.

Quatre expressions faciales sont étudiées : Joie, Dégoût, Colère et Tristesse. Cinq distances faciales sont calculées pour chaque image de chaque vidéo correspondante à chaque acteur pour chaque expression faciale. L'Algorithme (1) est appliqué afin de convertir le type

de données quantitatives en données qualitatives. Les résultats finaux pour tous les acteurs et les quatre expressions faciales sont regroupé dans le tableau suivant :

	D1+	D1-	D2+	D2-	D3+	D3-	D4+	D4-	D5+	D5-
S1-E1		3			1		2			
S2-E1		3			1		2			
S3-E1					1		2			3
S4-E1					1		2			3
S5-E1		3			1		2			
S6-E1		3			1		2			
S7-E1					1					2
S8-E1					1		2			3
S9-E1		3			1		2			
S10-E1		3			1		2			
S11-E1		3			1		2			
S12-E1		3			1		2			
S13-E1					1		2			3
S14-E1		3			1		2			
S15-E1		3			1		2			
S16-E1		3			1		2			
S17-E1		3			1		2			
S18-E1		3			1		2			
S1-E2		2		3			1			
S2-E2		2			3		1			
S3-E2		3			2		1			
S4-E2		2		3			1			
S5-E2		2					1			
S6-E2		2					1			3
S7-E2		2					1		3	
S8-E2		3		1						2
S9-E2		3		1						2
S10-E2		3		1						2
S11-E2		3		1						2
S12-E2		3		1						2
S13-E2		3		1						2
S14-E2		3		1						2
S1-E3		2				1		3		
S2-E3		3				2		1		
S3-E3		3				1		2		
S4-E3					2			1	3	
S5-E3	1					2		3		
S6-E3	1					2				
S7-E3		2					3	1		
S8-E3		2		3				1		
S9-E3		3		1				2		
S10-E3		3		1				2		
S11-E3		3		1				2		
S12-E3		3		2				1		
S13-E3		3		1				2		
S14-E3		3		2				1		
S15-E3		3		2				1		
S16-E3				3		1		2		
S17-E3				3		1		2		
S18-E3				3		1		2		
S19-E3		3		1				2		
S20-E3				1				2		
S21-E3		3		1				2		

S1-E4			1							
S2-E4		1		2	3					
S3-E4		1								
S4-E4			1							
S5-E4		1		2						
S6-E4		1								
S7-E4		2	1							
S8-E4		1		2						
S9-E4				3				2		1
S10-E4			2						1	
S11-E4			2						1	
S12-E4									1	
S13-E4			2						1	
S14-E4				2					1	
S15-E4				3				2		1
S16-E4								2	1	
S17-E4				3				2		1
S18-E4		2	1							
S19-E4			2						1	
S20-E4									1	
S21-E4		1								2
S22-E4		1	3							2
S23-E4					2				1	
S24-E4					2				1	
S25-E4			2						1	
S26-E4		2							1	

Table 5.1 Distances changées pour tous les acteurs de la base Dafex et quelques exemples de la base cohn et kanade pour quatre expressions faciales.

Les classes à considérer sont $\{E1,E2,E3,E4\}$ /E1= Joie; E2=Dégout; E3=Colère; E4=Tristesse.

Les colonnes présentent le sens de déformation des distances faciales (accroit= +, décroît=-) ; Les lignes présentent les acteurs considérées dans la phase d'apprentissage pour les quatre expressions faciales. L'ordre des distances changées est donné pour chaque acteur et pour chaque expression.

Par exemple la première distance qui a changé pour l'acteur 1 et l'expression joie (E1) est D3. Cette distance accroît. La seconde distance qui change est D4 elle accroît également. La dernière distance qui change est D1 elle décroît.

Cette table présente l'ensemble de données sur lesquelles l'algorithme (2) va être appliqué afin d'extraire des règles.

L'ensemble des attributs considérés sont égales à $\{D1+,D1-,D2-,D3+,D3-,D4+,D4-,D5+,D5-\}$; Chaque attribut peut prendre une des trois valeurs $\{1,2,3\}$ / tel que : 1= Premier ordre, 2= Second ordre, 3= Troisième ordre.

Les règles déduites sont regroupées dans la table (5.2):

Attributs	Vals	Classes qui appaissent	Classes Frequentes	Règles déduites	Erreur	Exactitude	Commentaire
D1+	1	E3	E3	If D1+=1 then Exp=E3	0/2	2/2=100%	
Total					0/2 Enreg		
D1-	1	E4	E4	If D1-=1 then Exp=E4	0/7	7/7=100%	
	2	E2,E3	E2	If D1-=2 then Exp=E2	6/9 Enreg	6/9=33,33 %	33,33% exp=E3 33;3% exp=E4
	3	E1,E2,E3	E1	If D1-=3 then Exp=E1	13/32 Enreg	13/32=40,62%	34,38% exp=E3 25% exp=E2
Total					19/48 Enreg		
D2+	1	E4	E4	If D2+=1 then exp=E4	0/4 Enreg	4/4=100%	
	2	E4	E4	If D2+=2 then exp=E4	0/5 Enreg	5/5=100%	
	3	E4	E4	If D2+=3 then exp=E4	0/1 Enreg	1/1=100%	
Total					0/10 Enreg		
D2-	1	E2,E3	E2,E3	If D2-=1 then exp=E2 If D2-=1 then exp=E3	7/14 Enreg	7/14=50%	7/14=50% exp=E3
	2	E3,E4	E4	If D2-=2 then exp=E4	3/7 Enreg	4/7=57,14 %	42,86% Exp=E3
	3	E2,E3, E4	E3	If D2-=3 then exp=E3	5/9 Enreg	4/9=44,44 %	22,22% exp=E2 33,33% exp=E4
Total					15/30 Enreg		
D3+	1	E1	E1	If D3+=1 then Exp=E1	0/18 Enreg	18/18=100 %	
	2	E2,E3, E4	E4	If D3+=2 then exp=E4 If D3+=2 then exp=E3	2/4 Enreg	2/4=50%	25% exp=E2 25% exp=E3
	3	E2,E4	E2,E4	If D3+=3 then exp =E2	1/2 Enreg	1/2=50%	50% exp=E4
Total					3/24 Enreg		
D3-	1	E3	E3	If D3-=1 then Exp=E3	0/5 Enreg	5/5=100%	
	2	E3	E3	If D3-=2 then exp=E3	0/3 Enreg	3/3=100%	
	3	/	/	/	/	/	
Total					0/8 Enreg		
D4+	1	E2	E2	If D4+=1 then Exp=E2	0/7 Enreg	7/7=100%	
	2	E1	E1	If D4+=2 then exp=E1	0/17 Enreg	17/17=100 %	
	3	E3	E3	If D4+=3 then exp =E3	0/1 Enreg	1/1=100%	
Total					0/25 Enreg		
D4-	1	E3	E3	If D4-=1 then Exp=E3	0/7 Enreg	7/7=100%	
	2	E3,E4	E3	If D4-=2 then exp=E3	4/15 Enreg	11/15=73,33%	26,67% exp=E4
	3	E3	E3	If D4-=3 then exp=E3	0/2 Enreg	2/2=100%	
Total					4/24Enreg		
D5+	1	E4	E4	If D5+=1 then exp=E4	0/11	11/11=100 %	
Total					0/11		
D5-	1	E4	E4	If D5-=1 then exp=E4	0/3	3/3=100%	

	2	E2,E1, E4	E2	If D5=2 then exp=E2	3/8 Enreg	5/8=62,5%	12,5% exp=E1 25% exp=E4
	3	E1,E2	E1	If D5=3 then exp =E1	1/5 Enreg	1/5=80%	20% exp=E2
Total					2/16 Enreg		

Table 5.2 Règles déduites

5.5.5 Interprétation et Evaluation des Résultats

Pour sélectionner les règles les plus intéressantes depuis l'ensemble des règles déduites, des informations et des connaissances complémentaires doivent être incorporées :

- Il existe cinq classes, donc il faut au minimum cinq règles, chaque règle décrit une classe.
- Si une expression est reconnue sans aucun doute en utilisant une règle premier ordre, les règles du deuxième et troisième ordre sont ignorées mêmes si ces règles ont un minimum d'erreur.
- Pour chaque attribut on s'intéresse aux règles qui impliquent le maximum d'instances et qui ont le maximum d'exactitude.
- Si une règle reconnaît une expression avec un doute en utilisant une règle premier ordre, les règles du deuxième ordre (si elles existent) sont alors utilisées afin de réduire ou éliminer ce doute.
- Si une expression est reconnue comme (E_i) avec un doute avec une expression (E_j) en utilisant une règle « Second ordre », tel que (E_j) peut être reconnue en utilisant une règle « premier ordre » sans aucun doute donc l'expression étudiée est reconnue autant que E_i sans aucun doute.
- En appliquant l'algorithme (2), S'il existe plus qu'une règle associée à une classe E_i, l'algorithme retient la règle qui possède le minimum d'erreur. Ceci est vrai lorsqu'il s'agit d'un même attribut.
- Si plusieurs règles sont associées à une classe E_i avec différents attributs dans la partie condition, alors toutes les règles sont retenues même s'il existe un doute ou bien le taux d'erreur est supérieur car ceci veut dire qu'il y a différentes façons d'exprimer une émotion, ce qui reflète vraiment la réalité. Par exemple une personne peut exprimer le dégoût avec une déformation au niveau des sourcils, et une autre personne exprime le dégoût avec une déformation au niveau des lèvres.
- Enfin si aucune des règles retenues n'est appliquée, l'expression étudiée est classifiée autant qu'expression inconnue (E_{un}). La table 5.3 résume les règles les plus intéressantes correspondantes aux critères citées.

Attributs	Règles	Erreur	Exactitude	Instances	expressions Reconnues
D1+	(1)If D1+=1 then Exp=E3	0/2	100%	2	E3
D1-	(2)If D1-=1 then Exp=E4	19/48	100%	7	E4
D2+	(3)If D2+=1 then exp=E4	0/10	100%	4	E4
	(4)If D2+=2 then exp=E4		100%	5	E4
D2-	(5)If D2-=1 then exp=E2	15/30	50%	14	E2: 50%; E3: 50%
	(6)If D2-=1 then exp=E3				
D3+	(7)If D3+=1 then Exp=E1	3/24	100%	18	E1
D3-	(8)If D3-=1 then Exp=E3	0/8	100%	5	E3
D4+	(9)If D4+=1 then Exp=E2	0/25	100%	7	E2
	(10)If D4+=2 then exp=E1		100%	17	E1
D4-	(11)If D4-=1 then Exp=E3	4/24	100%	7	E3
	(12)If D4-=2 then exp=E3		73,33%	11	E3:73,33%,E4:26,67%
D5+	(13)If D5+=1 then exp=E4	0/11	100%	11	E4
D5-	(14)If D5-=1 then exp=E4	0/3	100%	3	E4
	(15)If D5-=2 then exp=E2	2/16	62,5%	8	E2:62,5%;E1:12,5%, E4:25% (E1,E4 reconnus en premier ordre) =>E2:100%

Table 5.3 Les Règles les plus intéressantes

Les règles pour chaque classe sont ensuite regroupées pour éliminer les règles redondantes :

- Il existe deux règles pour reconnaître E1. La règle (7) utilise le “premier ordre”. La règle (10) utilise le “second ordre”. Puisque les règles “premier ordre” sont prioritaires par rapport aux règles “second ordre”, uniquement les règles “premier ordre” sont retenues.
- Il existe trois règles pour reconnaître E2. Deux règles (5,9) utilisent le premier ordre et la règle (15) utilise le “second ordre”. Comme nous avons besoin des différentes descriptions de chaque expression, les deux règles sont retenues car les deux règles utilisent différents attributs. La règle (5) avec “premier ordre”, reconnaît l’expression étudiée avec un doute, automatiquement la règle “second ordre est utilisée afin de réduire ou éliminer le doute. C’est pourquoi, la règle (15) est retenue.
- Il existe cinq règles pour reconnaître E3. Quatre règles (1, 6, 8, 11) utilisent le “premier ordre” est une règle (12) utilise le “second ordre”. Les quatre règles “premier ordre” sont toutes retenues car elles utilisent des attributs différents. La règle (6) du “premier ordre”, reconnaît l’expression étudiée avec un doute, une règle second ordre est recherchée c’est pourquoi la règle (12) est retenue.
- Il existe six règles pour reconnaître E4. Quatre règles (2, 3, 13, 14) utilisent le “premier ordre” et deux règles (4, 15) utilisent le “second ordre”. Les quatre règles

sont retenues puisqu'elles utilisent des attributs différents. Les règles (4,15) utilisent le mêmes attributs que (3, 14) , ces règles sont considérés redondants ainsi elles sont éliminées.

Finalement, les règles retenues sont pour chaque expression :

E1(7),

E2(5,9,15),

E3(1,6,8,11,12),

E4(2,3,13,14)

Ceci veut dire que moins que la moitié des règles induites sont sélectionnées (27 règles induites et Seulement 13 règles sont retenues).

Depuis les résultats obtenus, on peut déduire que l'expression de joie ne possède pas de descriptions différentes, cependant les autres expressions présentent différentes descriptions. Les descriptions déduites ici sont les plus rencontrées (selon les cultures, les nations et les continents).

A partir de règles retenues, nous avons obtenu de nouvelles descriptions des expressions faciales, ces descriptions possèdent une sémantique temporelle. La table 5.4 présente la description statique complétée par la description dynamique des six expressions universelles.

Expressions	Description des Expressions faciales dans un contexte statique	Description des Expressions faciales dans un contexte dynamique
Joie	Les yeux sont légèrement plissés, la bouche est ouverte, en un mouvement horizontal. Les lèvres sont donc étirées, toujours dans un mouvement latéral	Le premier changement apparu est l'étirement des coins des lèvres vers les Oreilles, ensuite les yeux se plissent.
Dégout	Les coins intérieurs des sourcils sont légèrement abaissés. La bouche est fermée mais on peut remarquer que la lèvre supérieure remonte la réduction du champ de vision : l'œil est à demi-ouvert.	Le dégoût peut commencer par une déformation au niveau de la lèvre supérieure qui monte, ou bien Les coins intérieurs des sourcils sont légèrement abaissés suivit de la lèvre qui monte.
Colère	La paupière recouvre une partie de l'œil donc les yeux sont presque fermés. Les sourcils ont tendance à se rejoindre, ils sont froncés, plissés. Les lèvres se serrent.	La colère commence par des yeux qui s'ouvrent, ou bien Les lèvres qui se serrent ou bien La bouche qui rétrécit ou bien Les sourcils se rejoignent et les lèvres se serrent
Tristesse	Les coins intérieurs des sourcils sont légèrement froncés pour donner cette forme / \ ou encore) \ .les paupières recouvrent une partie du champ de vision et les coins de la bouche seront étirés vers le bas.	La tristesse peut commencer par les paupières qui recouvrent une partie du champ de vision ou bien Les coins intérieurs des sourcils sont légèrement froncés pour donner cette forme / \ ou bien les coins de la bouche seront étirés vers le bas

Table 5.4 Descriptions statique et dynamique des expressions faciales

En conclusion, on peut déduire que l'exploitation du facteur temps associé à la notion d'« Ordre » est très importante dans l'analyse dynamique car ce facteur apporte quelque chose de nouveau dans l'analyse des expressions faciales, il peut attacher des informations sur la sémantique qui peut décrire une expression faciale.

5.5.6 Déploiement

La dernière étape consiste à exploiter directement les connaissances extraites, et les incorporer dans un nouveau système ou bien simplement en documentant les connaissances découvertes. Dans notre cas on incorpore les connaissances extraites dans un nouveau système afin d'évaluer l'efficacité des nouvelles règles retenues. Plusieurs exemples issus de deux bases de données sont testés.

5.5.6.1 Test de la Base Hammal_caplier

La base comporte 21 sujets et elle présente quatre expressions: Joie, Dégout, Surprise et Neutre. Chaque sujet a enregistré une vidéo de plus de 200 images pour chaque expression. Les résultats obtenus sont résumés dans le tableau suivant :

	Joie	Dégout
Joie	100%	
Dégout		90,48%
Dégout Or Colère		4,76%
Unconnue		1/21 4,76%
Total Reconnu	100%	95,24%
Total	100%	100%

Table 5.5 Les taux de classification de la base Hammal-caplier.

5.5.6.2 Test de la Base Cohn et Kanade

L'ensemble des exemples qui n'ont pas été utilisés dans la phase d'apprentissage sont testés. Les exemples comptent 40 sujets avec une expression de dégoût, 19 sujets avec une expression de colère, 74 sujets avec une expression de joie et enfin 40 sujets avec une expression de tristesse.

	Joie 97->84-10	Dégoût 97->46-6	Colère 97->32-13	Tristesse 97->58-18
Joie 97-13	74/74(Règle 7) 100%			
Dégoût		8/40(Règle9) 29/40(Règle 5) 92,5%		
Colère			1/19(Règle 8) 7/19(Règle 11) 8/19(Règle 5) 84,24%	
Dégoût Or Colère		7,5%	2/19 10,5%	
Tristesse				32/40(Règle 13) 6/40(Règle 3) 1/40(Règle 14) 1/40(Règle 2) 100%
Inconnue			1/19 5,26%	
Total Reconnu	100%	100%	94,74%	100%
Total	100%	100%	100%	100%

Table 5.6 Taux de classification de la base Cohn et Kanade.

Les tables 5.5 et 5.6 présentent en colonnes les expressions classées par un expert et en lignes les expressions classées par notre système. La reconnaissance de l'expression de joie est basée sur une seule règle, par contre la colère, le dégoût et la tristesse sont basées sur plusieurs règles. Pour chacune de ces trois expressions, le nombre de sujets qui vérifient une règle est associé à cette règle.

On peut observer que les taux obtenus pour le doute entre Dégoût et Colère est dû à l'existence d'une règle « premier ordre » pour classer une expression avec un doute et l'absence de règles « deuxième ordre » qui peuvent être utilisées afin d'éliminer ce doute.

On observe également que la classification d'une expression telle que « Expression Inconnue » est due à l'absence des règles "premier et deuxième ordre" qui peuvent être utilisées. Ceci reflète bien la réalité des choses vu, qu'il n'existe pas que ces quatre expressions étudiées, en effet il peut bien s'agir d'autres expressions telles que peur, surprise ou autres.

L'expression de joie est reconnue sans aucun doute et obtient par conséquent un taux maximal qui est de 100%. Les taux de reconnaissance des deux expressions de colère et dégoût sont également maximisés. En général tous les taux trouvés sont réellement importants.

5.6 Conclusion

Dans ce chapitre, nous avons introduit l'utilisation des méthodes de Data Mining dans la classification des expressions faciales. Quatre expressions faciales sont étudiées : Joie, Dégoût, Colère et Tristesse. La méthode la plus facile qui est la méthode de « La règle d'association » , "The one association rule" est évaluée.

Une nouvelle découverte de connaissance est reportée, la connaissance correspond à la déduction de nouvelles règles qui décrivent une expression faciale dans un contexte dynamique. Ces règles sont très efficace et exactes (généralement 100%). Afin de prouver l'efficacité de ces règles, une autre base d'images a été testée. Les taux obtenus sont très importants par comparaison avec les résultats obtenus avec d'autres méthodes.

Conclusion et Perspectives

Durant cette thèse nous nous sommes intéressés à l'analyse des expressions faciales, une telle analyse a fait l'objet de plusieurs travaux réalisés par des chercheurs dans différents domaines (psychologie, psychosociologie, neurophysiologie, psycho comportementaliste, physiologie, Informatique...). L'objectif de cette thèse est de suivre des chemins peu ou pas explorés jusqu'à ce jour, en considérant de nouveaux paramètres et de nouvelles données. Un autre objectif était de tester de nouvelles techniques ou méthodes qui jusque la n'ont pas été testé dans le domaine de l'analyse des expressions faciales. Dans ce travail nous avons proposé différentes contributions pour la mise en place d'un système automatique de classification d'expressions faciales. Nous avons donc développé un système de reconnaissance des expressions faciales basé principalement sur les traits transitoires connus également sous le nom « Rides d'Expressions ». Afin de réaliser ce système, nous avons considéré les deux travaux effectués dans le laboratoire qui a proposé ce sujet de thèse. Ces deux travaux concernent la détection des traits permanents du visage qui sont les yeux, les sourcils et les lèvres. Cette détection automatique a permis la localisation des points caractéristiques de ces traits qui sont les coins et les limites de chaque trait. A partir de ces points caractéristiques, nous avons détecté des zones faciales ou des traits transitoires peuvent apparaître. Une étude sur la présence ou l'absence de ce type de trait a donné une classification primaire des expressions faciales sans fournir beaucoup d'effort. Cette étude est complétée par une étude sur les traits permanents afin de finaliser les résultats. Par ce travail, nous avons prouvé qu'une classification basée principalement sur les traits transitoires est aussi performante qu'une classification basée principalement sur les traits permanents.

Cette classification catégorielle a prouvé ces limites dans la reconnaissance d'expressions faciales autres que les expressions universelles qui sont la joie, le dégoût, la colère, la tristesse, la surprise et la peur, ceci nous a poussé à développer un autre système qui fait la classification dimensionnelle des expressions faciales dans le sens positivité /négativité ou plaisir / non plaisir des expressions.

Un système d'expressions faciales idéal permet également de quantifier une expression faciale, ceci est l'objet du deuxième objectif de cette thèse qui consiste à un

développement d'un système qui permet l'estimation de l'intensité des expressions faciales connues ou inconnues.

Les trois systèmes proposés sont basés en plus des traits transitoires, sur les déformations des traits permanents. Pour mesurer la déformation de ces traits, cinq distances caractéristiques sont définies depuis les points caractéristiques des traits permanents détectés.

Le Modèle de Croyance Transférable est utilisé dans les trois systèmes proposés afin de fusionner l'ensemble des données disponible dans chaque étude. Ces données sont issues de différentes sources, et sont de différents types ce qui explique notre choix d'utilisation de ce modèle préconisé dans ce type de scénarios. En plus ce modèle permet la modélisation du doute qui peut exister entre les différentes classes considérées : expressions catégorielles, classe positives et négatives ou bien les classes d'intensités.

Un autre objectif de cette thèse est la reconnaissance des expressions faciales dans une séquence vidéo. Dans une base d'images d'expressions faciales, plusieurs acteurs présentent différentes expressions faciales avec différentes intensités sous forme de séquences vidéo, chaque vidéo peut contenir jusqu'à 400 images. L'extraction des données de toutes ces images en vue d'une éventuelle analyse nous conduit à une masse importante de données à explorer. Cette contrainte nous a conduit à faire appel aux techniques de datamining afin d'extraire de nouvelles connaissances temporelles qui peuvent compléter la description statique des expressions faciales proposée par la norme M-PEG4.

Le système de reconnaissance développé tient compte de l'évolution au cours du temps des déformations des traits du visage. Il permet d'induire des règles temporelles qui décrivent les déformations des traits permanents. Ces règles peuvent donc s'ajouter aux règles déjà existantes (proposées par MPEG-4) pour décrire une expression faciale dans un contexte dynamique.

Comme perspectives, nous estimons qu'aucun travail n'est complètement fini, le notre non plus. Nous envisageons donc de continuer à explorer surtout d'autres techniques de datamining pour la découverte de nouvelles connaissances. Ces nouvelles connaissances permettront une meilleure analyse des expressions faciales en utilisant les informations issues des séquences vidéo. Une autre perspective est d'explorer le nouvel univers des expressions spontanées. En effet une expression donnée n'est pas structurellement identique selon qu'elle est spontanée ou posée et cette dissociation entraîne des différences de perception.

Références Bibliographiques

- [AHL01] J. Ahlberg, CANDIDE-3 -- an updated parameterized face, Report No. LiTH-ISY-R-2326, Dept. of Electrical Engineering, Linköping University, Sweden, 2001.
- [ALL83] James F. Allen: Maintaining knowledge about temporal intervals. In: Communications of the ACM. 26/11/1983. ACM Press. S. 832-843, ISSN 0001-0782 Bruce D. Lucas
- [ANA89] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. In International Journal on Computer Vision, vol. 2, pp. 283-310, 1989.
- [BAR94] L. Barron, S. S. Beauchemin, and D. J. Fleet. Performance of optical flow techniques. In International Journal on Computer Vision, vol. 12, pp. 43-77, 1994.
- [BAR99] M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski,. "Measuring facial expressions by computer image analysis". *Psychophysiology*, 36:253–264, 1999.
- [BAR03] M.S. Bartlett, G. Littlewort, I. Fasel, J. R. Movellan, Face Detection, Facial Expression Recognition: Development and Applications to Human Computer Interaction, In IEEE workshop on Computer Vision and Pattern Recognition for Human Computer Interaction, Madison, U.S.A., June, 2003.
- [BAS78] Bassili J. N. Facial motion in the perception of faces and of emotional expression. *Experimental Psychology - Human Perception and Performance*, 4 no.3:373–379, 1978.
- [BEC06] Becker, C., Leßmann, N., Kopp, S. and Wachsmuth, I. 2006. Connecting feelings and thoughts - modeling the interaction of emotion and cognition in embodied agents. In Proceedings of the Seventh International Conference on Cognitive Modeling (ICCM-
- [BER90] M.O. Berger, R. Mohr „Towards Autonomy in Active Contour Models“ In Proc. ICPR'90, June 1990, pp.847-851.
- [BLA93] Black M.J and Anandan P. A framework for the robust estimation of optical flow. *Proc. Computer Vision, ICCV*, pages 231–236, Berlin, Germany, 1993.
- [BLA97] Black M.J. and Yacoob Y. Recognizing facial expression in image sequences using local parametrized models of image motion. *Trans. Computer Vision*, 25 no.1:23–48, 1997.
- [BOU75] Boucher J.D. and Ekman P. Facial areas and emotional information. *Journal of Communication*, 25 no. 2:21–29, 1975.
- [BRU83] Raymond Bruyer "Le visage et l'expression faciale : approche neuropsychologique" ISBN 2870091753 : 9782870091753. pp 167-174 P. éditeur : Mardaga, 1983.

- [BUI02] Bui TD, Heylen D, Poel M, Nijholt A, « ParleE: An adaptive plan-based event appraisal model of emotions » Proceedings KI, 25th German Conference on Artificial Intelligence, eds,2002.
- [CAR97] J.M Carroll, J.Russell,” Facial Expression in hollywood’s Portrayal of Emotion”,J.Personality and social psychology, vol.72, p.164-176, 1997.
- [CHA99] N. P. Chandrasiri, M.C. Park and T. Naemura and H. Harashima, "Personal Facial Expression Space based on Multi-dimensional Scaling for the Recognition Improvement", Proc.IEEE ISSPA'99, pp. 943-946, Brisbane, Australia, August 1999
- [CHE00] L. S. Chen.: “Joint Processing of Audio-visual Information for the Recognition of Emotional Expressions in Human-computer Interaction”. PhD thesis, University of Illinois at Urbana-Champaign, 2000;
- [COH98] Cohn J.F., Zlochower A. J., Lien J. J and Kanade T. Feature point tracking by optical flow discriminates subtles differences in facial expression. IEEE International Conference on Automatic Face and Gesture Recognition, pages 396–401, April Nara, Japan, 1998.
- [COH02] I. Cohen, N. Sebe, A. Garg, M.S. Lew, T.S. Huang, “Facial Expression Recognition from Video Sequences » IEEE International Conference on Multimedia and Expo (ICME'02), vol II, pp. 121-124, Lausanne, Switzerland, August 2002
- [COH03] Cohen I., Cozman F. G., Sebe N., Cirelo M. C. and 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.
- [COHO4]: Cohn, J. F., & Schmidt, K. L. The timing of facial motion in posed and spontaneous smiles. International Journal of Wavelets, Multiresolution and Information Processing, 2, 1-12 (2004).
- [COH base]: Kanade,T., Cohn,J. F., & Tian, Y. “Comprehensive database for facial expression analysis”; Proceedings of the fourth IEEE International Conference on Automatic Face and Gesture recognition(FG'00), Grenoble, France 46-53.
- [COI95] T. Coianiz, L Torresani , B. Caprile “2D Deformable Models for Visual Speech Analysis” In NATO Advanced Study Institute : Speech reading by Man and Machine, 1995, pp.391-398.
- [COO94] T. Cootes, C. Taylor, A. Lanitis, Multi-resolution search using active shape models, 12th International Conference on Pattern Recognition, Vol. 1, IEEE CS Press, Los Alamitos, CA, 1994, pp. 610–612.
- [Dafex base]: Dafex Database: <http://tcc.itc.it/research.i3p/dafex/index.html>
- [DEM67] A. Dempster, “Upper and miner probability inferences based on a sample from a finite univariate”. Biometrika, 54: 515-528, 1967.

- [DEM68] A. Dempster. "A generalization of Bayesian inference". Journal of the royal statistical society, Vol30, ages 205-245, 1968.
- [DEN06] T. Denoeux and Ph. Smets. "Classification using belief functions : The relationship between the case-based and model-based approaches". IEEE Trans. on Systems, Man and Cybernetics, 2006.
- [EDW98a] G.J. Edwards, T.F. Cootes, and C.J. Taylor, "Face Recognition Using Active Appearance Models" Proc. European Conf. Computer Vision, vol. 2, pp. 581-695, 1998.
- [EDW98b] Edwards, K, "The face of time: Temporal cues in facial expression of emotion". Psychological Science, 9, 270-276, 1998.
- [EE base]: EEBase Database: <http://www.cs.cmu.edu/afs/cs/project/cil/ftp/html/v-images.html>
- [EIB89] I.Eibel-Eihesfeldt, Human Ethology. New work: Aldine de Gruvter,1989. Handbook of Cognition and Emotion. Sussex, U.K.: John Wiley & Sons, Ltd.
- [EKM75] P. Ekman and W.V. Friesen, Unmasking the Face. New Jersey: Prentice Hall, 1975.
- [EKM78] Ekman P. and Friesen W. V. Facial action coding system. Consulting Psychologist Press, 18, no.11:881-905, August 1978.
- [EKM 80a]: Ekman, P. Facial asymmetry. Science, 209, 833-834. (1980).
- [EKM 80b] Ekman, P. Friesen, W. V. & Ancoli, S. "Facial signs of emotional experience". Journal of Personality and Social Psychology. 39, 1125-1134. (1980)
- [EKM82] P. Ekman, Emotion in the Human Face. Cambridge Univ. Press, 1982.
- [EKM86] Ekman, P. & Friesen, W.V. "A new pan-cultural facial expression of emotion". Motivation and Emotion, 10, 159-168. (1986).
- [EKM99]: Paul Ekman (1999). Basic Emotions. In T. Dalgleish and M. Power (Eds.).
- [EKM02] P.Ekman, Wallace V. Friesen, Joseph C. Hager :Facial Action Coding System : A Technique for the Measurement of Facial Movement. Psychology, Physiological. ISBN 0-931835-01-1 1. 1978, 2002
- [ESS97] I. Essa and A. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions". IEEE Trans. on Pattern Analysis and Machine Intell., 19(7):757-763, 1997.
- [EVE01] Eveno N., Caplier A., Coulon P. Y. A new color transformation for lip segmentation. Proc. IEEE MSSP'01, September Cannes, France, 2001.
- [EVE02] N. Eveno, A. Caplier, P.Y. Coulon. A parametric model for realistic lip segmentation. International Conference on Control, Automation, Robotics and Vision (ICARV'02), Singapore, December 2002.

- [EVE03] N. Eveno, A. Caplier, P.Y. Coulon. “Jumping snakes and parametric model for lip segmentation”. International Conference on Image Processing, Barcelone, Espagne, Septembre 2003.
- [EVE04] N. Eveno, A. Caplier, and P.Y. Coulon “Automatic and Accurate Lip Tracking”. To appear in IEEE Trans. On Circuits and Systems for Video Technology, May 2004.
- [FAS02] I. R. Fasel, J. R. Movellan, “A Comparison of Face Detection Algorithms”, Proceedings of ICANN, vol. 2415, p. 143, 2002
- [FAS03] Beat FASEL and Juergen LUETTIN. « Automatic Facial Expression Analysis : A Survey ». Pattern Recognition, 36(1) :259–275, 2003.
- [FAY96] Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth: Knowledge Discovery and Data Mining: Towards a Unifying Framework. KDD 1996: 82-88
- [FRA93] Frank, M.G., Ekman, P., Friesen, W.V. (1993). “Behavioral markers and recognizability of the smile of enjoyment”. Journal of Personality and Social Psychology 64, 83–93.
- [FLE90] D.J. Fleet and A.D. Jepson. Computation of component image velocity from local phase information. In International Journal on Computer Vision, vol. 5, pp. 77-104, 1990.
- [GAR04] C. Garcia, M. Delakis, “Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 11, pp. 1408-1423, Nov. 2004
- [GEB05] Gebhard, P. 2005. ALMA - A Layered Model of Affect In Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05) (Utrecht, 2005).
- [GIR05] V. Girondel, A. Caplier, L. Bonnaud, and M. Rombaut. “Belief theory-based classifiers comparison for static human body postures recognition in video”. Int. Jour. of Signal Processing, 2(1) :29–33, 2005.
- [GOS95] Gosselin, P., Kirouac, G., & Dore, F.Y. “Components and recognition of facial expression in the communication of emotion by actors”. Journal of Personality and Social Psychology, 68, 83-96. (1995).
- [GOU00] Gouta K. and Miyamoto M. Facial areas and emotional information. Japanese journal of psychology, 71 no. 3:211–218, 2000.
- [HAM05]: Hammal Z., Caplier A. and Rombaut M. :A Fusion Process Based on Belief Theory Classification of Facial Basic Emotions. Proc. the 8th International Conference on Information fusion, (ISIF), Philadelphia, PA, USA, 2005.
- [HAM07] Z. Hammal, L. Couvreur, A. Caplier, M. Rombaut “Facial Expression Classification: An approach based on the fusion of facial deformations using the Transferable Belief Model” Int. Jour. Approximate Reasoning, doi: 10.1016/j.ijar.2007.02.003, 2007.

- [HAM base] Hammal_Caplier data base :
http://www.lis.inpg.fr/pages_perso/caplier/english/emotionnelle.html.en/emotionnelle_2.html.en.html
- [HAR00] A. Haro, M. Flickner, and I. Essa "Detecting and tracking eyes by using their physiological properties, dynamics and appearance". Proc. IEEE CVPR, South Carolina, June 2000.
- [HEE88] D. J. Heeger. Optical flow using spatiotemporal filters. In International Journal on Computer Vision, vol. 1, pp. 279-302, 1988.
- [HEN94] M.E. Hennecke, K.V Prasad, D.G. Storck "Using Deformable Templates to Infer Visual Speech Dynamics". In Proc. 28th Annual Asilomar Conference on Signals, Systems and computers, 1994, pp. 578-582.
- [HON98] H. Hong, H. Neven and C. Y. d. Mahurg, "Online Facial Expression Recognition based on Personalized Gallery", Proc. int . Conf. on Automatic Face and Gesture Recognition, pp. 354-359, Nara, Japan, 1998.
- [HOR81] B.K.P Horn and B.G. Schunck. Determining optical flow. AI 17, pp. 185-204, 1981.
- [HUA97] C.L. Huang and Y.M. Huang, "Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification," J. Visual Comm. and Image Representation, vol. 8, no. 3, pp. 278-290, 1997.
- [HUL98] A. Hulbert, T. Poggio "Synthesizing a Color Algorithm from Examples" Science, Vol.239, 1998, pp.482-485.
- [HUN62] Hunt, E.B. (1962). Concept learning: An information processing problem. New York: Wiley.
- [ITU95] ITU-T SG15 WP15/1, Draft Recommendation H.263 (video coding for low bitrate communications), Document LBC-95-251, October 1995.
- [Jaffe base]: JAFFE database: http://www.kasrl.org/jaffe_download.html
- [KAS88] M. Kass, A. Witkins, D. Tersopoulos „Snakes : Actives Contours Models“, International Journal of computer vision, 1(4), January 1988, pp.321-331.
- [KIM97] S. Kimura and M. Yachida, "Facial Expression Recognition and its Degree Estimation", Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 295-300, 1997
- [KIR90] M.Kirby et L.Sirovich, « Application of the K-L Procedure for the characterization of Human Faces », IEEE Trans. Pattern analysis and machine intelligence, vol 12, N°1, p.103-108, jan. 1990.
- [KOB97] H. Kobayashi and F. Hara, "Facial Interaction between Animated 3D Face Robot and Human Beings," Proc. Int'l Conf. Systems, Man, Cybernetics,, pp. 3,732-3,737, 1997.

- [KOG81] T. Koga. Motion compensated interframe coding for video conferencing. National telecommunication conference, New Orleans, November 1981.
- [KSH02] Kshirsagar, S. 2002. A multilayer personality model. In Proceedings of the 2nd international symposium on Smart graphics SMARTGRAPH '02 (Hawthorne, New York, 2002). ACM Press.
- [KWO94] Y. Kwon and N.Lobo, « Age classification from facial images » proc. IEEE Conf. Computer vision and Pattern Recognition, p. 762-767 , 1994
- [LIE98] Lien J.J, Kanade T., Cohn J.F. and Li C. Subtly different facial expression recognition and expression intensity estimation. Proc. Computer Vision and Pattern Recognition (CVPR), pages 853–859, June 23-25 Santa Barbara, CA, 1998.
- [LIE00] J.Lien, T.Kanade,J.F Cohn , « Detection tracking, and classification of Action Units in Facial expression », J.Robotics and autonomous system, vol.31, p. 131-146, 2000.
- [LIE01] Y.Li Tian, T.Kanade et J.F.Cohn « Recognizing action units for facial expression analysis”,IEEE Trans. Pattern analysis and machine intelligence, vol 23, N°2, p.97-114, feb.2001.
- [LEE03] Ka Keung Lee “Real-time Estimation of Facial Expression Intensity”, proceedings of the 2003 IEEE Internatioid Conference 00 Robotics &Automation Taipei, TaiWBn, September 14-19, 2003.
- [LIS98] C. Lisetti and D. Rumelhart, "Facial Expression Recognition Using a Neural Network", Proc. the 11 th Int. Conf. Facial Expression Recognition, AAAI Press, 1998.
- [LUC84] B.D. Lucas. Generalized Image Matching by the Method of Differences. Carnegie Mellon University, Technical Report CMU-CS-85-160, Ph.D. dissertation, July 1984.
- [LYO99] M.J. LYONS, J. BUDYNEK, S. AKAMATSU, Automatic Classification of Single Facial Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(12), p. 1 357–1 362, December 1999
- [MEH96] Mehrabian, A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychology, 14 (1996), 261-292.
- [MER06] D. Mercier. “information Fusion for automatic recognition of postal addresses with belief functions theory” PhD thesis, University of Technologie of Compiegne, December 2006.
- [MIN02] Ming-Hsuan Yang, David J. Kriegman et Narendra Ahuja : Detecting faces in images : A survey. Dans IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 24(1), pages 34–58, 2002.
- [MIU02] Kazuyuki Miura: “Discrimination of Facial Expressions using Velocity Estimation from Image Sequence”. Proceedings of The First International Workshop on Kansei.

- [MPEG-4] Audio and video object coding, MPEG-4 ISO/IEC 14496-1.
- [NAG87] N.-H. Nagel. On the estimation of optical flow : relations between different approaches and some new results. *AI* 33, pp. 299-324, 1987.
- [ORL base] ORL face database http://www.uk.research.att.com:pub/data/att_faces.zip
- [PAD96] C. Padgett and G.W. Cottrell, "Representing Face Images for Emotion Classification," *Proc. Conf. Advances in Neural Information Processing Systems*, pp. 894-900, 1996.
- [PAN00a] Pantic M. and Rothkrantz L. J. M. Expert system for automatic analysis of facial expressions. *Image and Vision Computing Journal*, 18, no.11:881–905, August 2000.
- [PAN00b] M. Pantic, and L. Rothkrantz. "Automatic Analysis of Facial Expressions : the State of the Art". *IEEE PAMI*, Vol.22, N°12, pp. 1424-1445, December 2000.
- [PAR00] M. Pardàs. Extraction and Tracking of the Eyelids, *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2357-2360, Istanbul, Turkey, June 2000.
- [PAR02] Pardas M., Bonafonte A. Facial animation parameters extraction and expression detection using hmm. *Signal Processing: Image Communication*, 17:675–688, 2002.
- [POS base]: I. Cohen and colleagues :Facial Expression Recognition from Video Sequences. *Computer Vision and image understanding* , volume 91, Issue 1-2, ISSN;1077 3142 pages: 160-187, 2003.
- [QUI79] Quinlan, J.R. (1979). Discovering rules by induction from large collections of examples. In D. Michie (Ed.), *Expert systems in the micro electronic age*. Edinburgh University Press.
- [QUI86] J. Ross Quinlan, "Induction of Decision Trees", *Machine Learning*, 1, 1986, 81-106.
- [QUI87] J. R. Quinlan. Generating production rules from decision trees. In *IJCAI 87: Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pages 304-307, Los Altos, 1987. Morgan Kaufmann.
- [RAM07] E. Ramasso, C. Panagiotakis, M. Rombaut, and D. Pellerin. "Human action recognition in videos based on the transferable belief model - application to athletics jumps". *Pattern Analysis and Applications Journal*, 2007.
- [ROS96] Rosenblum M., Yacoob Y. and Davis L.S. Human expression recognition from motion using a radial basis function network architecture. *IEEE Trans. Neural Networks*, 7 no.5:1121–1137, 1996.
- [ROW98] H.A. Rowley, S. Baluja, T. Kanade, "Neural network-based face detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.20, no.1, pp.23-38, Jan. 1998

- [ROZ94] Rozin, P., Lowery, L., & Ebert, R.. “Varieties of disgust faces and the structure of disgust”. *Journal of Personality and Social Psychology*, 66, 870-881. (1994)
- [RUC93] Ruch, W. (1993). “Extraversion, alcohol, and enjoyment. *Personality and Individual Differences*”, 16, 89–102.
- [RUS94]: Russell, J.A. Is there universal recognition of emotion from facial expression? *Psychological Bulletin*, 115, 1 (Jan. 1994), 102-141.
- [SAY95] Sayette, M.A., & Hufford, M.R. (1995). “Urge and affect: A facial coding analysis of smokers. *Experimental and Clinical psychopharmacology*”, 3, 417–423.
- [SAY01] Michael A. Sayette, Jeffrey F. Cohn, Joan M. Wertz, Michael A. Perrott, and Dominic J. Parrott “A Psychometric Evaluation of the Facial Action Coding System for Assessing Spontaneous Expression” *Journal of Nonverbal Behavior*, 25, pp.167-186.
- [SEB02] N. Sebe, I. Cohen, A. Garg, M.S. Lew, T.S. Huang, “Emotion Recognition using a Cauchy Naive Bayes Classifier” *International Conference on Pattern Recognition (ICPR'02)*, vol I, pp. 17-20, Quebec City, Canada, August 2002.
- [SHA76] G. Shafer. “A Mathematical Theory of Evidence”. Vo.12702. Princeton,1976.
- [SME94] P. Smets, R. Kennes, “The Transferable Belief Model”. *Artificial Intelligence*, 66 (2): 191-234, 1994.
- [SME00]: Philippe Smets, *Data Fusion in the Transferable Belief Model*. Proceedings of the Third International Conference on Information Fusion, PS21- PS33 vol.1, 2000.
- [SOU96] Soussignan, R. & Schaal, B. “Children’s responsiveness to odors: Influences of hedonic valence of odor, gender, age, and social presence”. *Developmental Psychology*, 32, 367-379. (1996).
- [TEK99] M. Tekalp M. “Face and 2d mesh animation in mpeg-4”. Tutorial Issue on the MPEG-4 Standard, *Image Communication Journal*, Elsevier, 1999.
- [TER94] D.Terzopoulos et K.Waters, « Analysis of facial images using physical and anatomical models », *proc. IEEE Intl’Conf. Computer vision and Pattern Recognition*, p. 762-767 , 1994
- [TIA00a] Y.-L. Tian, T. Kanade, and J. Cohn, “Eye-state action unit detection by Gabor wavelets”. In *Proc. of Int. Conf. on Multi-modal Interfaces*, 143–150, Sept 2000.
- [TIA00b] Y. Tian, T. Kanade., and J. Cohn “Dual state Parametric Eye Tracking”. *Proc. of the 4th IEEE Inter. Conf. on Automatic Face and Gesture Recognition* , March 2000, pp. 110 – 115.
- [TIA00c] Y. Tian, T. Kanade, J. Cohn “Robust Lip Tracking by Combining Shape, Color and Motion”. *Proc ACCV’00*, 2000.

- [TIA01] Tian Y., Kanade T. and Cohn J.F. Recognizing action units for facial expression analysis. *Trans. IEEE Pattern Analysis and Machine Intelligence*, 23 no.2:97–115, February 2001.
- [TSA00] Tsapatsoulis N., Karpouzis K., Stamou G., Piat F. and Kollias S. A fuzzy system for emotion classification based on the mpeg-4 facial definition parameter set. *Proc. 10th European Signal Processing Conference*, September 5-8 Tampere, Finland, 2000.
- [TUR90] M.Turk et A.Pentland, « Face recognition using eigenfaces », *Proc. IEEE Conf. Computer vision*, p. 727-732 , 1990.
- [URA88] S. Uras, F. Girosi, A. Verri and V. Torre. A computational approach to motion perception. In *Biol. Cybern.* 60, pp 79-97, 1988.
- [VAL99] S. Valente. *Analyse, Synthèse et Animatin de Clones dans un Contexte de Téléreunion Virtuelle*. Thèse de doctorat, Ecole Polytechnique Fédérale de Lausanne, Institut Eurécom, France, 1999.
- [VRA93] Vrana, S.R. “The psychophysiology of disgust: Differentiating negative emotional contexts with facial EMG”. *Psychophysiology*, 30, 279-286. (1993).
- [WAN98] M. Wang, Y. Iwai and M. Yachida, "Recognizing Degree of Continuous Facial Expression Change", *Proc. Fourteenth Int. Conf. On Pattern Recognition*, Vol. 2, pp. 1188-1190,1998.
- [Yac96] Yacoob Y. and Davis L.S. Recognizing human facial expressions from long image sequences using optical flow. *IEEE Trans. Pattern Analysis and Machine Intelligence*,18 no.6:636–642, June 1996.
- [YON97] M. Yoneyama, Y. Iwano, A. Ohtake, and K. Shirai, “Facial Expressions Recognition Using Discrete Hopfield Neural Networks, ° *Proc. Int'l Conf. Information Processing*, vol. 3, pp. 117-120, 1997.
- [ZHA98] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, “Comparison between Geometry-Based and Gabor Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron,° *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 454-459, 1998.
- [ZHA96] J. Zhao and G. Kearney, “Classifying Facial Emotions by Backpropagation Neural Networks with Fuzzy Inputs” *Proc. Conf. Neural Information Processing*, vol. 1, pp. 454-457, 1996.

Publications et Communications

[Ghanem et al] K.Ghanem, A.Caplier , MK Kholadi “ Contribution of facial transient features in facial expression analysis : Classification and quantification“ Journal of theoretical and Applied information technology ISSN 1817-3195 pp 69-78 Volume 15 n°1 . May 2010.

[Ghanem et al] K.Ghanem, A.Caplier , MK Kholadi “ Intensity estimation of unknown expression based on a study of facial permanent features deformations” International Symposium on Modeling and implementation of complex systems. Pp 77-87 May 2010.

[Ghanem et al] K.Ghanem, A.Caplier “Estimation of Facial Expression Intensity Based on the Belief Theory” VISAPP 2008: Proc. the Third International Conference on Computer Vision Theory and Applications, Funchal, Madeira,Portugal, - Volume 1, 452-460,2008.

[Ghanem et al] K.Ghanem, A.Caplier , MK Kholadi “Estimation of Anger, Sadness and fear expressions intensity based on the Belief theory” ACIT 2008: Proc. 9th International Arab Conference on information Technology, Hammamet, Tunisia, december,2008.

[Ghanem et al] K.Ghanem, A.Caplier “Classification of facial Expressions based on Transient Features“, 3rd Workshop on Emotion and Computing, Kaiserslautern, Germany, September 2008.

Résumé

La reconnaissance des expressions faciales est une tâche très importante dans les systèmes de communication homme machine. Au cours de cette thèse, nous avons développé un système de reconnaissance des expressions faciales basé principalement sur les traits transitoires connus également sous le nom « Rides d'Expressions ». Afin de réaliser ce système, nous avons considéré les deux travaux effectués dans le laboratoire qui a proposé ce sujet de thèse. Ces deux travaux concernent la détection des traits permanents du visage qui sont les yeux les sourcils et les lèvres. Cette détection automatique a permis la localisation des points caractéristiques de ces traits qui sont les coins et les limites de chaque trait. A partir de ces points caractéristiques, nous avons détecté des zones faciales ou des traits transitoires peuvent apparaître. Une étude sur la présence ou l'absence de ce type de trait a donné une classification primaire des expressions faciales sans fournir beaucoup d'effort. Cette étude est complétée par une étude sur les traits permanents afin de finaliser les résultats.

Cette classification catégorielle a prouvé ces limites dans la reconnaissance d'expressions faciales autres que les expressions universelles qui sont la joie, le dégoût, la colère, la tristesse, la surprise et la peur, ceci nous a poussé à développer un autre système qui fait la classification dimensionnelle des expressions faciales dans le sens positivité /négativité ou plaisir / non plaisir des expressions.

Un système d'expressions faciales idéal permet également de quantifier une expression faciale, ceci est l'objet du deuxième objectif de cette thèse qui consiste à un développement d'un système qui permet l'estimation de l'intensité des expressions faciales connues ou inconnues.

Les trois systèmes proposés sont basés en plus des traits transitoires, sur les déformations des traits permanents. Pour mesurer la déformation de ces traits, cinq distances caractéristiques sont définies depuis les points caractéristiques des traits permanents détectés.

Le Modèle de Croyance Transférable est utilisé dans les trois systèmes proposés afin de fusionner l'ensemble des données disponible dans chaque étude. Ces données sont issues de différentes sources, et sont de différents types ce qui explique notre choix d'utilisation de ce modèle préconisé dans ce type de scénarios. En plus ce modèle permet la modélisation du doute qui peut exister entre les différentes classes considérées : expressions catégorielles, classe positives et négatives ou bien les classes d'intensités.

Un autre objectif de cette thèse est la reconnaissance des expressions faciales dans une séquence vidéo. Dans une base d'images d'expressions faciales, plusieurs acteurs présentent différentes expressions faciales avec différentes intensités sous forme de séquences vidéo, chaque vidéo peut contenir jusqu'à 400 images. L'extraction des données de toutes ces images en vue d'une éventuelle analyse nous conduit à une masse importante de données à explorer. Cette contrainte nous a conduit à faire appel aux techniques de datamining afin d'extraire de nouvelles connaissances temporelles qui peuvent compléter la description statique des expressions faciales proposée par la norme M-PEG4.

Le système de reconnaissance développé tient compte de l'évolution au cours du temps des déformations des traits du visage. Il permet d'induire des règles temporelles qui décrivent les déformations des traits permanents. Ces règles peuvent donc s'ajouter aux règles déjà existantes (proposées par MPEG-4) pour décrire une expression faciale dans un contexte dynamique.

Mots-clefs: Traits transitoires, traits permanents, expressions faciales, classification catégorielle, classification dimensionnelle, Modèle de Croyance Transférable, datamining.

Laboratoire MISC

Nouvelle ville Constantine

التعرف على انفعالات الوجه عملية جد مهمة على مستوى أنظمة التعامل أنسان/ آلة. خلال هذه الرسالة طورنا نظام تعرف على الأنفعالات الوجهية مبني أساسا على قسّمات الوجه العابرة أو الوقتية المعروفة تحت اسم القسّمات الأنفعالية. من أجل انجاز هذا النظام اعتمدنا على نتائج الدراسات المنجزة في المخبر الذي طرح موضوع هذه الرسالة. الأعمال تخص تحديد أطراف قسّمات الوجه الدائمة : العينان, الحاجبان و الشفتان. هذا التحديد الأتوماتيكي سمح بتحديد نقاط مميزة لقسّمات الوجه. ابتداء من هذه النقاط تمكنا من تحديد مناطق خاصة على الوجه أين يمكن أن تظهر القسّمات العابرة. الدراسة التي تمت على أساس ظهور أو عدم ظهور هذا النوع من القسّمات اضافة الى بعض خصائص هذه القسّمات أثبتت امكانية التعرف على الأنفعالات الوجهية دون بذل جهد. و جاءت دراسة بسيطة للقسّمات الدائمة كدراسة تكميلية للدراسة السابقة من أجل تحسين النتائج المتحصل عليها.

ثم بينت هذه الدراسة حدودها من حيث أنها لا تتعرف إلا على سبعة انفعالات وجهية و المعروفة عالميا و هي: الفرح, الغضب, الحزن, الخوف, الأشمزاز و الأندهاش. لهذا السبب قمنا بتطوير نظام آخر يتعرف على الأنفعالات الموجبة أي التي تبين الحالة الحسنة للإنسان و الأنفعالات السالبة أي التي تبين الحالة السيئة للإنسان.

من معايير تقييم أي نظام تحليل الأنفعالات الوجهية هو امكانية تقييم حدة او قوة هذه الأنفعالات, لهذا السبب قمنا بتطوير نظام يقيم قوة الأنفعالات الوجهية المعروفة و غير المعروفة.

الأنظمة المقترحة الثلاثة و تعتمد كلها اضافة الى القسّمات العابرة على القسّمات الدائمة. لقياس مدى التشوهات و الحركات البارزة أثناء ظهور انفعالات على الوجه, اعتمدنا على خمسة مسافات مأخوذة مابين النقاط المميزة للقسّمات الوجهية الدائمة التي تم تعيينها من قبل.

استعمل نموذج الاعتقاد القابل للتحويل في الأنظمة المقترحة الثلاثة من أجل ضم مجموعة المعلومات الموجودة و الصادرة من مختلف الجهات. هذه المعلومات مختلفة النوع مما يبرر اختيارنا لهذا النموذج بالأضافة الى أن هذا النموذج يسمح بتمثيل الشك بين مختلف الأنفعالات الوجهية و كذلك مختلف التقييمات الأنفعالية.

آخر هدف لهذه الرسالة هو التعرف على الأنفعالات الوجهية بناء على معلومات مصدرها الفيديو. نجد في خزينة الصور الناتجة عن مجموعة من الفيديوهات و المسجلة لمجموعة من الأشخاص الذين يعملون في المخابر و الذين يظهرون انفعالات مختلفة على وجوههم و بدرجات مختلفة من الحدة كميات هائلة من المعلومات التي يجب تحليلها. هذه الكميات من المعلومات يصعب تحليلها بواسطة طرق كلاسيكية مما جعلنا نستعمل تقنية مميزة من تقنيات "استخراج المعارف" و هي تقنية تسمح باستخراج معارف مرتبطة بالزمن أو الوقت. هذه المعارف تصف كيفية ظهور الأنفعالات الوجهية المعروفة, و هي على شكل قواعد ديناميكية يمكن اضافتها للقواعد المقترحة من طرف MPEG-4 لوصف الأنفعالات بطريقة ستاتيكية.