Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

BADJI MOKHTAR UNIVERSITY-ANNABA -

UNIVERSITE BADJI MOKHTAR-ANNABA-

جامعة باجي مختار

- عنابة -

Faculté des Sciences de l'Ingéniorat
Département d'Informatique

Année : 2014-2015

THESE

Présentée en vue de l'obtention du diplôme de ***DOCTORAT en Informatique***

Reconnaissance et interprétation des expressions faciales

Option :

Informatique

Par

Benmohamed Abderrahim

Devant le Jury

Président:	Suici-Meslati Labiba	Prof	Université de Annaba
Rapporteur:	Ramdani Messaoud	MCA,	Université de Annaba
Examinateurs:	Bouden Toufik	Prof	Université de Jijel
	Seridi Hamid	Prof	Université de Guelma
	Bouchrika Imed	MCA	Université de S. Ahras
	Alimi Adel (invité)	Prof	Université de Sfax

Année Universitaire : 2014-2015

Remerciement :

Je tiens à remercier ALLAH qui m'a donné la force pour accomplir ce travail.

J'exprime également mes remerciements et ma gratitude à mon encadreur Mr Ramdani Messaoud pour l'opportunité qu'il m'a offert de travailler sous sa supervision.

Mes remerciements s'adressent aussi aux membres de jury.

Je saisie cette opportunité pour remercier :

Mes familles et mes amis qui m'ont épaulé tout au long de travail, je les remercie du fond du cœur.

Abstract

Face detection, person identification and recognition of facial expression have potential applications in various aspects of life from day to day. A system that performs these operations will find many applications domain, e.g. Identification of criminals, authentication in secure systems, and recently in emotional user feedback through media. Most of the current works are based on the identification.

We propose in this work two emotion analysis systems based, respectively, on models of auto-associative neural networks, and fuzzy entropy for feature selection.

The proposed systems are able to detect and identify the user through his facial expressions to recognize his emotional state. They are then composed of two main modules.

The experiments were conducted to verify the feasibility of each proposed system. Their objective is the validation of the face detection and recognition of the user and his emotions through facial expressions with FABO and Jaffe databases.

The results show that the proposed emotion analysis systems provide a good performance with high recognition rates.

Résumé

La reconnaissance des visages et la reconnaissance de l'expression faciale ont des applications potentielles dans différents aspects de la vie de jour en jour. Un système qui exécute ces opérations trouvera de nombreux domaines d'applications, par exemple : identification des criminels, l'authentification dans les systèmes sécurisés, et récemment en affective feedback avec l'utilisateur axée sur l'usage des médias. La plupart des travaux à ce jour sont basé sur l'identification.

Nous proposons dans ce travail deux systèmes d'analyse de l'émotion cognitive basés respectivement, sur des modèles de réseaux neuronaux auto-associatifs, et sur l'entropie floue pour sélection des caractéristiques. Les systèmes proposés sont capables de détecter et d'identifier l'utilisateur à travers ses expressions faciales afin de reconnaître son état émotionnel.

Les expériences ont été menées afin de vérifier la faisabilité de chaque système proposé. Leur objectif est la validation de la détection des visages et la reconnaissance de l'utilisateur et ses émotions à travers ses expressions faciales avec les bases de données FABO et JAFFE.

Les résultats expérimentaux montrent que chaque système proposé fournit une bonne performance avec un bon taux de reconnaissance.

Résumé

ملخص

التعرف على الأشخاص من خلال الوجه و كذلك التعرف على تعبيرات الوجه لها تطبيقات عديدة في مختلف جوانب الحياة تزداد من يوم لآخر، هذه الأنظمة يمكن تطبيقها في مجالات عدة، على سبيل المثال: التعرف على المجرمين، الأنظمة الأمنية، ومؤخرا في ردود الفعل العاطفية لمستخدم الكمبيوتر باستعمال وسائل الكترونية. معظم الأبحاث الحالية تحديد تتركز في التعرف على الوجه.

نقترح في هذا العمل نظامين للتعرف على العاطفة يتركزان على نماذج الشبكات العصبية الألية و الانتروبي الغامضة من أجل اختيار المتغيرات.

النظامين المقترحين بإمكانهما التعرف على الأشخاص و كذلك حالتهم النفسية من خلال تعابير الوجه.

أجريت التجارب للتحقق من جدوى كل من النظامين المقترحين على كل من قاعدة البيانات **فابو** و **جافي** ، الهدف منها هو اختبار الكشف على الوجه وكذلك التعرف على المستخدم ومشاعره من خلال تعابير الوجه.

النتائج الجيدة المتحصل عليها تترجم صلابة كلا النظامين.

Sommaire

Chapitre 3 : L'émotion

Chapitre 4 : Présentation du système

Chapitre 5 : Expérimentations

Introduction générale :

Depuis longtemps, les interfaces homme-machine pour les produits et services ont continué d'évoluer pour mieux s'adapter au besoin des utilisateurs. La conception la plus importante qui permet la réussite de l'humain est de l'aider à effectuer rapidement et facilement ses tâches et d'offrir la souplesse qui permet aux utilisateurs de manier facilement les différents appareils.

Au cours des dernières années, une énorme attention a été accordée à l'amélioration de tous les aspects de l'interaction entre l'homme et la machine en développant des interfaces intelligentes. Dans ce contexte, de nombreuses études ont été proposées dans différents domaines tels que la vision par ordinateur, l'ingénierie, la psychologie et les neurosciences. Ces études visent à améliorer les interfaces-ordinateur et réformer également les actions que l'ordinateur peut exécuter en fonction de la rétroaction de l'utilisateur. Par conséquent, les ordinateurs doivent être capables d'interpréter le comportement de l'utilisateur afin de satisfaire ses demandes (Neji, Benammar, Wali, & Alimi, 2013). En ce sens, l'auteur dans (Picard R. W., 1999.) dit: «si nous voulons des ordinateurs qu'ils soient véritablement intelligents et interagissent naturellement avec nous, nous devons leur donner la capacité de reconnaître, comprendre, et même d'avoir et d'exprimer des émotions". Ainsi, pour faciliter l'interaction homme-machine, la machine doit être capable de reconnaître, de comprendre et de gérer les émotions. De la même façon, les auteurs dans (Minsky, 1985) ont posé la question de la nécessité d'émotions pour la synthèse de l'intelligence et disent: "la question n'est pas de savoir si les machines intelligentes peuvent avoir des émotions, mais si les machines peuvent être intelligentes sans émotions." Par ailleurs, Christine Lisetti[1], explique le terme« affective » liés à l'ordinateur en disant " il est à prendre en compte le rôle des émotions dans la cognition pour améliorer l'interaction homme-machine ". Toutes ces études mettent en évidence l'importance de l'interaction efficace entre l'homme et la machine.

En fait, bien que les nouvelles technologies soient présentes dans notre vie quotidienne, ils ne fournissent pas une interface adéquate qui les rend plus abordables pour les utilisateurs. Par conséquent, l'informatique affective en améliorant l'interaction homme-ordinateur, permet aux ordinateurs d'être plus adaptés à l'homme et non pas l'inverse. Dans ce contexte, notre recherche se concentre sur la reconnaissance des émotions à partir d'expressions faciales afin de fournir un système d'analyse de l'émotion cognitive.

[1] Professeur à l'université de Floride

Reconnaître les émotions à partir d'expressions faciales a gagné récemment une attention accrue, parce que la reconnaissance automatique de l'émotion peut aider les gens à développer et concevoir de nombreuses applications sur la communication homme-machine. Les systèmes de reconnaissance des émotions à partir d'expressions faciales qui ont été utilisés comprennent plusieurs états émotionnels tels que la joie, la peur, la tristesse, le dégoût, la colère, la surprise et neutre. Au cours des dernières années, le visage et la reconnaissance des émotions a été un domaine de recherche actif en raison de la disponibilité des systèmes informatiques rapides et les exigences de sécurité accrues. Cette recherche a conduit à la mise au point d'algorithmes de contrôle d'accès et des systèmes de vérification de l'identité. L'objectif a toujours été de parvenir à une bonne analyse de l'image indépendamment de l'état de la personne portant des lunettes, chapeau, barbe, certaines coupes de cheveux, etc.

Dans ce travail, nous proposons deux systèmes de reconnaissance de l'émotion à partir des expressions faciales. Premièrement, les deux systèmes commencent par la détection de visage dans une séquence vidéo. La détection des visages est la phase la plus sensible, car une bonne détection conduit à une bonne reconnaissance si nous utilisons des caractéristiques pertinentes. Dans les méthodes basées sur les connaissances (Jingru & W.Y., 1999) (Leonardis & Bischof, 2000), la variation de la luminance rend difficile la mise en place des règles de recherche de la peau, que ce soit dans l'espace HSV ou RGB. Même avec les méthodes basées sur les caractéristiques invariantes (Canny), il est pénible à extraire d'une image une primitive pertinente dans de telles circonstances. D'autre part, les méthodes modélistes (Lanitis, Taylor, & Cootes., 1994) (Lanitis, Taylor, 1995) sont robustes aux changements d'éclairage, mais pas aux changements de l'échelle. Enfin les méthodes globales (Turk & Pentland, 1991) (Sung, 1996) (Rowley, Baluja, & Kanade, 1998) nécessitent un temps d'apprentissage trop long. Dans notre cas, nous devons détecter les visages dans divers conditions d'éclairage, quelle que soit la distance entre l'utilisateur et la caméra, et ceci dans le temps le plus court possible. Pour ce faire, nous proposons une méthode qui combine les informations sur la couleur et la forme avec la recherche des yeux qui permet d'ajuster au mieux l'ellipse entourant le visage. Nous combinons ces deux méthodes, car la première permet de détecter le visage dans différents positions ou de détecter une autre chose qui ressemble à la couleur de la peau, par exemple, le cou. Pour cela, la détection des yeux rassure que l'objet détecté est un visage. Nous avons également mis en place la méthode proposée dans (Yu-Tzu, Ruei-Yan, L.Yu-Chih, & C.Greg., 2012) avec une légère

amélioration pour l'élimination de la luminance. Par la suite, une détection des lèvres, basée sur les modèles géométriques (Frank & Chao-Fa, 2004), est réalisée.

Pour la reconnaissance des visages et des émotions, nous combinons deux espaces de paramètres, en fait, chacun traite l'ensemble des primitives à sa façon, le premier système combine les moments de Zernike avec les EAR-LBP et utilise une technique floue de sélection des paramètres pertinents, tandis que le deuxième système utilise le même paramétrage mais dans un contexte multi-échelle, et emploie une technique de réduction des paramètres basée sur les mémoires auto-associatives.

Cette thèse est composée de cinq chapitres. Le premier chapitre récapitule les méthodes de détection des visages tandis que le deuxième est consacré pour les approches de reconnaissance. Dans le troisième chapitre, nous mettons l'emphase sur le concept de l'émotion, et enfin le quatrième et cinquième chapitres, développent respectivement la partie conceptuelle, et la partie expérimentale de notre approche.

Chapitre 1

Détection des visages

Introduction

Toutes les applications automatiques imitant la perception humaine et destinées à reconnaitre des êtres humains, à identifier leurs états émotionnels, ou simplement segmenter des visages, doivent passer par une phase initiale et primordiale, à savoir, la détection du visage humain.

Une image, statique ou vidéo, peut contenir un ou plusieurs visages, ou n'en posséder plus. Une détection implique une localisation précise permettant aux phases postérieures de traitement de se focaliser sur la bonne partie de l'image, alors de fonctionner avec plus de précision et efficacité, évitant ainsi de tomber dans les problèmes de fausses reconnaissances dites respectivement, la fausse positive, et la fausse négative , que nous verrons en détails dans les chapitres subséquents.

Il est vrai qu'une détection de visage est une tâche très facile pour l'être humain, néanmoins, en ce qui concerne une machine intelligente cela devient de plus en plus difficile, que ce soit pour les images photographiques ou les vidéos.

Une localisation implique une segmentation, une extraction des primitives, puis une vérification. Et tout cela se déroule en tenant compte des différentes conditions d'échelles, de luminances, et d'orientations.

Les techniques de détection des visages se répartirent en deux grandes familles d'approches, notamment, les approches structurelles et les approches globales. La première catégorie utilise les informations concernant le visage dans lesquels des caractéristiques de bas niveau sont extraites, tandis que la deuxième catégorie classe la détection du visage humain parmi les problèmes généraux de reconnaissance de formes.

Dans ce chapitre, nous présentons brièvement les différentes techniques utilisées. Nous mettons l'emphase sur l'aspect théorique, et nous essayons de mentionner quelques résultats obtenus.

1 Les approches structurelles (basée-caractéristiques)

1.1 Analyse de bas niveau

Les contours

l'extraction de contours est amplement utilisée dans la détection des visages (Brunnelli & Poggio., 1993) (Choi, Kim, & Rhee, 1999.) (Herpers, Kattner, Rodax, & Sommer, 1995) (Low & Ibrahim, 1997). Elle consiste en une détection des traits de visage sur lesquels une analyse ultérieure est élaborée. Tous les traits détectés sont labélisés et comparés avec un modèle de visage pour vérifier si la détection est correcte ou pas.

Plusieurs types de détecteur sont appliqués, notamment, le filtre de Sobel, le Laplacien, le filtre de Marr-Hildreth…etc. Dans ce paragraphe nous allons présenter brièvement deux méthodes qui utilisent respectivement, le filtre de Sobel et le filtre Laplacien.

Le filtre de Sobel est aussi employé pour la détection des contours de visage, Ceci est effectué à l'aide des deux matrices de convolution suivante :

$$\partial_x = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{bmatrix} \quad (1.1) \qquad \text{et} \qquad \partial_Y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (1.2)$$

∂_x et ∂_y servent à calculer les deux composantes du gradient de l'image. La magnitude du gradient est calculée au niveau de chaque pixel afin de définir une nouvelle image appelée l'image magnitude du gradient, qui, à son tour, passe par une phase de binarisation. Enfin un algorithme de détection des contours ovales est lancé sur l'image binaire pour tracer le visage.

La méthode du gradient détecte les contours en se basant sur le maximum et le minimum de la première fonction dérivative de l'image.

Une deuxième méthode, appelée le Laplacien, consiste à retrouver les passages par zéro de la seconde fonction dérivative de l'image. Le Laplacien d'une image sert à détecter les changements rapides d'intensités inter-régions.

Généralement l'image doit passer d'abord par des filtres de lissage - avant le Laplacien- afin de lui rendre moins sensible au bruit.

Le Laplacien d'une image est défini ainsi:

$$L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (1.3)$$

Comme l'image est en forme de pixel, nous pouvons utiliser des matrices de convolution discrètes qui permettent de donner une approximation de la seconde dérivative, cette convolution est définie à l'aide des matrices suivantes :

$$\partial_x = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (1.4) \qquad \text{et} \qquad \partial_Y = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -1 \end{bmatrix} \quad (1.5)$$

Niveau de gris

Le niveau de gris représente aussi une caractéristique pertinente pour la détection des visages. Les yeux, les lèvres et les sourcils se montrent toujours plus sombres par rapport aux autres caractéristiques faciales. Cette idée est exploitée dans plusieurs approches dont nous présentons ici le travail de Yang et al. (Yang & Huang, 1994) et celui de Sobotka et al. (Sobottka & Pitas, 1996).

Les auteurs dans (Yang & Huang, 1994) ont développé un système automatique de détection des visages sans tenir compte de leur taille, nombre, et emplacement dans l'image. L'idée principale consiste à transformer l'image principale en une image mosaïque, puis à une certaine résolution, un ensemble de règles est appliqué afin de trouver les composants de visage. L'image mosaïque est obtenue en diminuant la résolution de l'image originale, elle est composée d'un ensemble de cadres dont chacun est de taille de n × n pixels, et le niveau de gris de chaque cadre est défini par la moyenne des intensités des pixels y sont inclus.

Sobotka et al dans (Sobottka & Pitas, 1996) proposent une approche qui consiste à trouver le visage suite à une détection de ses composants, à savoir, les sourcils, les yeux, le nez, et la bouche. Ceux-ci sont déterminés en cherchant les minimas et les maximas des projections horizontales et verticales des niveaux de gris, ils sont notés respectivement, par y-projection et x-projection.

y-projection est définie par la moyenne des niveaux de gris dans chaque ligne, les minimas significatifs, déterminés par le gradient, correspondent aux cheveux, sourcils, les yeux, le nez, et la bouche. Ensuite les x-projection, qui représentent la moyenne verticale des niveaux de gris, de chaque minima sont définis. Plus de détails sont donnés dans la figure 1.1.

La détection des composantes faciales se fait par un ensemble de règles récapitulées dans le tableau 1.1.

Sourcils, yeux	nez	La bouche
Deux minima significatifs	Deux minima significatifs	Deux maxima significatifs
La partie centrale ou supérieure du visage	La partie centrale de la tête	La partie centrale ou inférieure du visage
Un maximum significatif entre deux minimas	Un maximum significatif entre deux minimas	Un minimum significatif entre deux maximas
La distance entre minimas correspond à la largeur de la tête	Petite distance entre minimas	La distance entre maximas correspond à la largeur de la tête
Niveaux de gris similaire		

Table1.1. Description des caractéristiques faciales.

La couleur

C'est un excellent facteur de discrimination entre objets, deux formes qui possèdent le même niveau de gris peuvent apparaitre totalement différentes dans l'espace de couleurs. Le processus de détection des visages consiste en une projection de l'image dans un espace de couleurs, puis, une détection des skins pixels est élaborée à l'aide de règles heuristiques.

Les auteurs dans (Peer & Solina, 1999) (Solina, Peer, Batagelj, & Juvan, 2002) utilisent les règles suivantes pour classer les skins pixels dans un espace RGB (Red, Green, Blue) :

$$P(x,y) est\ un\ skin\ pixel\ si \begin{cases} R > 95\ et\ G > 40\ et\ B > 20 \\ \max\{R,G,B\} - \min\{R,G,B\} > 15 \\ |R-G| > 15\ et\ R > G\ et\ R > B \end{cases} \quad (1.6)$$

Autrement, Sobotka et al (Sobotka & Pitas, 1996), utilise l'espace de couleur HSV (Hue-Saturation-Value), représenté sous forme d'un Hexagone comme illustré dans la figure 1.2.

Comme montré dans la figure 2, H est un angle, la pureté de la couleur est définie par la saturation S qui varie entre 0 et 1, et le contraste de la couleur est déterminé par le composant V qui varie aussi entre 0 et 1.

Un pixel est classé comme spin pixel s'il vérifie les conditions suivantes :

$$\begin{cases} 0.23 \leq S \leq 0.68 \\ 0^{\circ} \leq H \leq 50^{\circ} \end{cases} \quad (1.7)$$

Cependant, une autre technique basée sur l'espace de couleurs (YC_bC_r) est proposée dans (Hsu, Abdel-Mottaleb, & Jain, 2002)]. Etant donné, Y, C_b, et C_r les centres $\overline{C_b}\ et\ \overline{C_r}$, et les déviations Wc_b et Wc_r sont déterminés par l'ensemble des formules ci-dessous :

$$Wc_i(Y) = \begin{cases} W_{Lc_i} + \frac{(Y - Y_{\min})(Wc_i - W_{Lc_i})}{K_L - Y_{min}} ; Y < K_L \\ W_{Hc_i} + \frac{(Y_{\max} - Y)(Wc_i - W_{Hc_i})}{K_L - Y_{min}} ; K_h < Y \\ Wc_i \qquad\qquad ; autrement \end{cases} \quad (1.8)$$

$$\overline{C_b}(Y) = \begin{cases} 108 + \frac{10(K_L - Y)}{K_L - Y_{min}} ; \ Y < K_L \\ 108 + \frac{10(Y - K_h)}{Y_{max} - K_h} ; \ K_h < Y \\ 108 \qquad\qquad ; autrement \end{cases} \quad (1.9)$$

$$\overline{C_r}(Y) = \begin{cases} 154 + \frac{10(K_L - Y)}{K_L - Y_{min}} ; \ Y < K_L \\ 154 + \frac{22(Y - K_h)}{Y_{max} - K_h} ; \ K_h < Y \\ 108 \qquad\qquad ; autrement \end{cases} \quad (1.10)$$

Où : i est b ou r, $c_b = 46{,}97,\ W_{Lc_b} = 23, W_{Hc_b} = 14, Wc_r = 38{,}76,\ W_{Lc_r} = 20, W_{Hc_r} = 10,$ $K_L = 125, K_h = 188, Y_{min} = 16, Y_{max} = 235.$

Un pixel classé skin pixel doit vérifier les conditions suivantes :

$$\overline{C_b}(Y) - \propto Wc_b(Y) < C_b < \overline{C_b}(Y) + \propto Wc_b(Y) \quad (1.11)$$

$$\overline{C_r}(Y) - \propto Wc_r(Y) < C_r < \overline{C_r}(Y) + \propto Wc_r(Y) \quad (1.12)$$

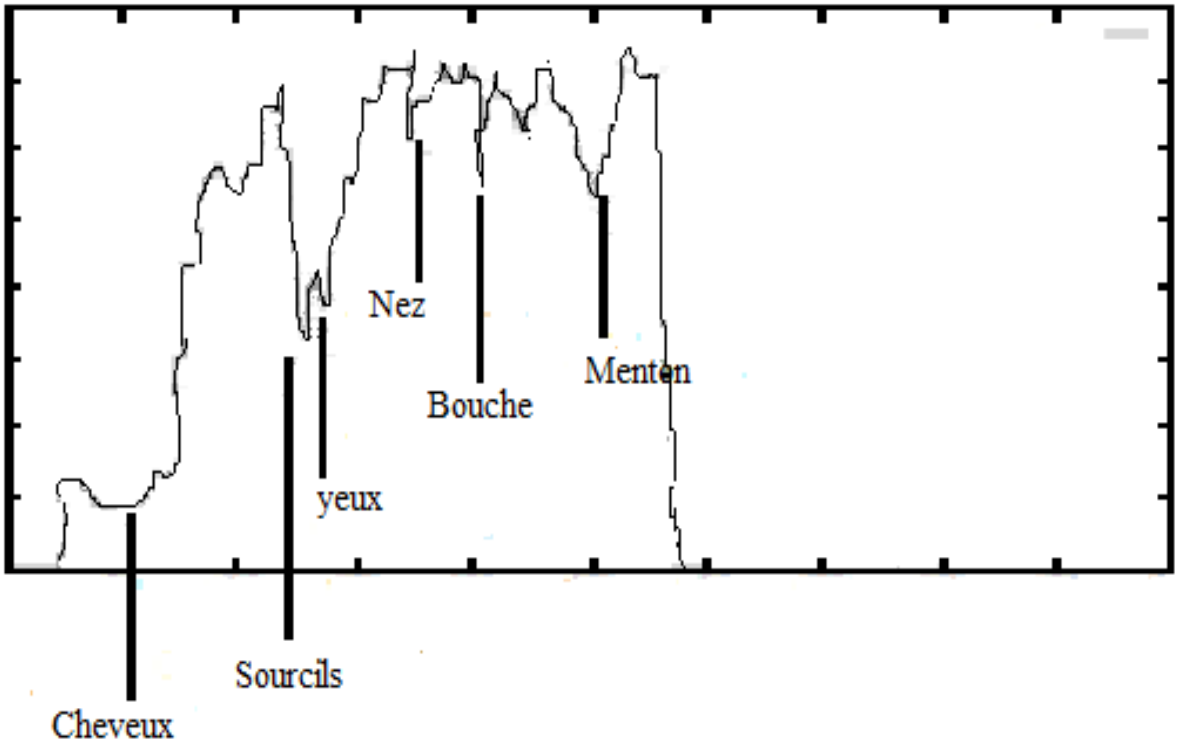

Figure 1.1. Projection horizontale des niveaux de gris.

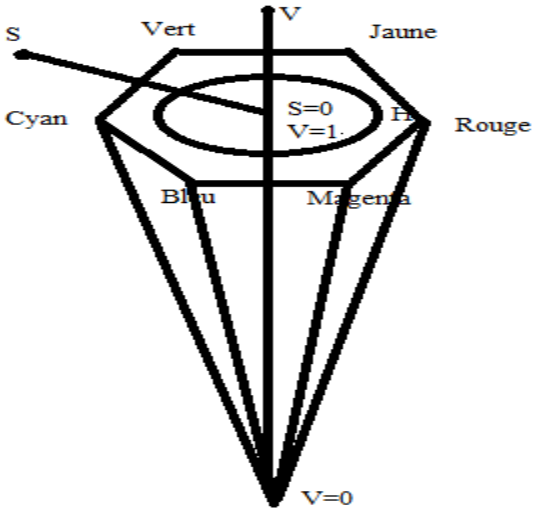

Figure 1.2. Hue-Saturation-Value (HSV) espace de couleurs.

Le mouvement

L'information sur le mouvement est employée dans le cas d'utilisation d'une séquence vidéo. La segmentation se fait par l'analyse de différence entre trames, qui sert à extraire efficacement le premier-plan sans tenir compte du contenu de l'arrière-plan.

Les auteurs dans (Beek, Reinders, Sankur, & Lubbe., 1992) (Graf, Cosatto, Gibson, Petajan, & Kocheisen, 1996) (Turk & Pentland, 1991), font l'extraction du visage et des parties du corps humain en utilisant un seuil de différence de trame accumulé. D'autres approches telles que dans (Low & Ibrahim, 1997) (Luthon & Lievin, 1997) (Low B. K. Computer Extraction of Human Faces, 1998) (Crowley & Berard, 1997) utilisent aussi la différence de trame afin de localiser les caractéristiques faciales. Par exemple Berard (Crowley & Berard, 1997) estime une existence des yeux en mesurant le déplacement horizontal et vertical entre deux régions adjacentes.

Une autre méthode plus fiable, pour mesurer le mouvement, est basée sur le déplacement des contours. McKenna *et al*. (McKenna, Gong, & Liddell, 1995) utilisent un filtre Gaussien pour détecter les mouvements de visage et du corps humain. Pour ce faire, une convolution de l'image grisée $I(x,y)$ par le moyen d'un opérateur temporel du second ordre, noté, $m(x,y,t)$, qui est défini par le filtre Gaussien $G(x,y,t)$ comme suit :

$$G(x,y,t) = \mu\left(\frac{a}{\pi}\right)^{\frac{3}{2}} e^{-a(x^2+y^2+\mu^2t^2)} \qquad (1.13)$$

$$m(x,y,t) = -\left(\nabla^2 + \frac{1}{\mu^2}\frac{\partial^2}{\partial t^2}\right) G(x,y,t) \quad (1.14)$$

Où : μ : est un facteur d'échelle de temps, a est la largeur du filtre.

Ensuite l'opérateur temporel m est convolué avec les trames consécutives de la séquence vidéo comme montré ci-dessous :

$$S(x,y,t) = m(x,y,t) \otimes I(x,y,t) \quad (1.15)$$

Le résultat de cette convolution $S(x,y,t)$ contient des passages par zéro indiquant des déplacements de contours dans $I(x,y,t)$. Ces passages par zéro sont repérés et regroupés afin de localiser les mouvements des objets.

Autres mesures :

D'autre chercheurs ont préféré d'employer des primitives visuelles de bas niveau, Reisfeld et Yeshurun (Reisfeld & Yeshurun, 1998) ont utilisé, dans leur approche, un opérateur de symétrie basé sur les pixels de bordure. Une mesure de symétrie consiste à attribuer une magnitude pour chaque position d'un pixel de l'image en fonction de son voisinage. La magnitude de symétrie,$M_\sigma(p)$, pour le pixel p est définie ainsi :

$$M_\sigma(p) = \sum_{(i,j)\in\Gamma(p)} C(i,j) \quad (1.16)$$

Où : $C(i,j)$ est la contribution des pixels voisins de P dans un ensemble $\Gamma(p)$ qui sont définis comme suit :

$$C(i,j) = D_\sigma(i,j)P(i,j)r_i r_j \ (1.17)$$

$$\Gamma(p) = \left[(i,j)\middle|\frac{p_i + p_j}{2} = p\right] \quad (1.18)$$

$D(i,j)$, est une fonction distance, $P(i,j)$ est une fonction de phase.

$$D_\sigma(i,j) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{\|p_i - p_j\|}{2\sigma}} \quad (1.19)$$

$$P(i,j) = \big(1 - \cos\big(\theta_i + \theta_j - 2\alpha_{ij}\big)\big)\big(1 - \cos\big(\theta_i - \theta_j\big)\big) \quad (1.20)$$

$$r_k = \log(1 + \|\nabla_{pk}\|) \quad (1.21)$$

$$\theta_k = arctan\left[\frac{\frac{\partial}{\partial y}pk}{\frac{\partial}{\partial x}pk}\right] \quad (1.22)$$

Où : P_k est un point de coordonnées (x_k, y_k) et ∇_{pk}est le gradient des intensité des points p_k.

Lin et al (Lin & Lin, 1996) proposent plusieurs approches pour l'extraction des caractéristiques faciales dans des images en niveau de gris. Ils utilisent certaines propriétés photométriques rarement exploitées comme primitives de base. Notamment, ils appliquent de nouvelles mesures sur la symétrie radiale d'orientation de gradient, un mécanisme d'inhibition pour extraire les traits du visage internes et un mécanisme simple pour les traits externes.

1.2 Analyse des caractéristiques

Les approches d'analyse de bas-niveau sont très limitées et parfois susceptibles d'être ambigus. Par exemple, la recherche de visage basée sur la couleur de la peau peut détecter d'autres objets de teint similaire. Ce problème peut être résolu en employant des caractéristiques de haut-niveau dont la géométrie du visage est l'une des techniques les plus utilisées.

Recherche de caractéristiques

Cette technique commence par une détection des caractéristiques les plus remarquables de visage, qui servent à en extraire d'autres moins importantes en utilisant des mesures anthropométriques. L'anthropométrie est la technique qui concerne la mesure des particularités dimensionnelles d'un homme. Elle est particulièrement utilisée en ergonomie. Jeng et al, (Jeng, Liao, Han, Chern, & Liu, 1998) proposent un système qui commence par une binarisation de l'image, par la suite, il tente de localiser les yeux, et à partir de chaque paire détectée, une recherche du nez, des lèvres, et des sourcils est déclenchée. Ce système, appliqué sur une base de 114 images prise dans des conditions d'imagerie contrôlées avec des personnes placées dans de différentes directions et un fond encombré, qui atteint un taux de détection de 86%.

Analyse de constellation ***:***

Ces méthodes servent à contourner les problèmes liés à la position de la tête et à la complexité de l'arrière-plan. Elles consistent à regrouper les caractéristiques faciales, dans ce qu'on appelle une constellation de visage, en utilisant les techniques de modélisation telles que l'analyse statistique.

Burl et al, (Burl, Leung, & Perona, 1995) (Burl & Perona, 1996) localise le visage par le moyen d'une analyse statistique de la forme. Pour ce faire, un ensemble de détecteurs locaux, à savoir les filtres multi-orientation et le Gaussien multi-échelle, sont appliqués sur l'image afin de détecter les emplacements candidats des caractéristiques faciales telles que les yeux, les lèvres, et le nez. Ces détecteurs forment donc, une constellation de visage. Les auteurs, dans la partie classification, se recoururent à la théorie statistique de formes développée par Kendall (Kendall., 1989).

Par ailleurs, D. Maio et al (Maio & Maltoni, 2000), proposent une approche qui commence par une détection approximative des zones pouvant contenir des visages, puis lance un processus itératif de vérification. Leur algorithme commence par une extraction des bordures

en employant la technique de l'image directionnelle, Figure3, par la suite cherche les blobs de forme elliptique par le moyen de la transformée généralisée de Hough (Schubert, 2000). Cette approche est très robuste face aux différentes variations d'illumination et d'échelle, même dans le cas de petites rotations de la tête.

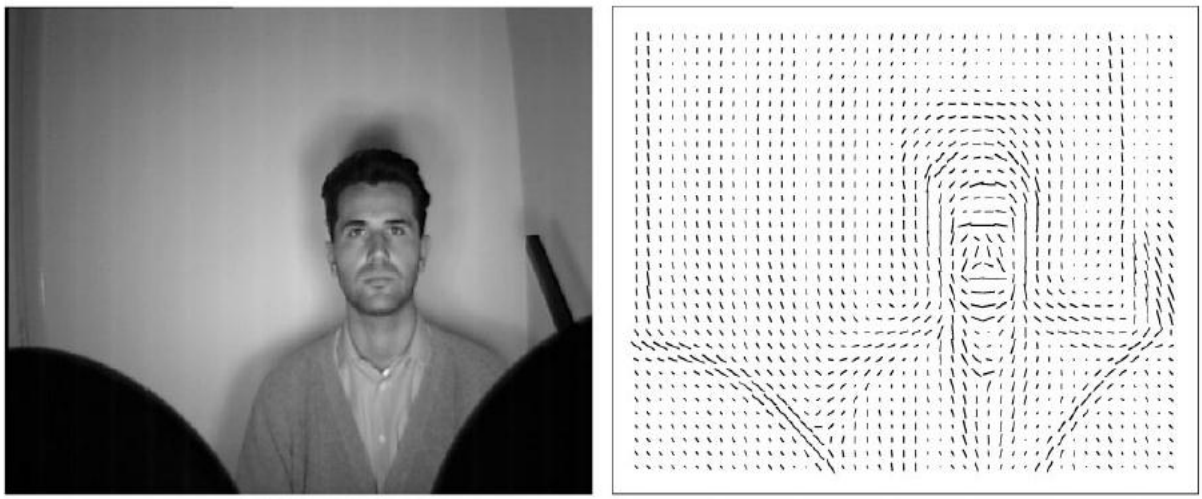

Figure 1.3. Extraction de l'image directionnelle

1.3 Les modèles de forme active

Ces modèles définissent l'état physique actuel de la forme. Un modèle de forme active interagit avec les caractéristiques locales de l'image, notamment, les bordures et la luminosité, le contour se déforme graduellement selon la forme des composants de l'image. Dans les travaux de recherche récents, on peut distinguer trois types du modèle de forme active, à savoir, les contours actifs nommés snake elements (snakels) et introduits par Kass et al (Kass, Witkin, & Terzopoulos, 1987), les modèles déformables mis en place par Yuille et al (Yuille, Hallinan, & Cohen, 1992), et les modèles de point distribué définis par Cootes et al (Cootes & Taylor., 1992) (Cootes, Taylor, Cooper, & Graham, 1995).

Les contours actifs (snake)

un contour actif est une courbe déformable qui dépend de ses forces intérieures et extérieures. Les forces intérieures imposent les contraintes de lissage de contour et celle extérieures, attirent le contour vers les composants de l'image. Le snake doit être, dans un premier temps, initialisé, ensuite un traçage de contour est effectué en plaçant le snake de l'image f_t dans f_{t+1}.

L'initialisation de snake consiste en un échantillonnage de contour en m nœuds $v_i = (x_i, y_i), i = 0, \dots, M-1$, on les appelle également Snakels (snake elements), et on peut

définir leur position initiale de différentes façons, par exemple, selon le nombre de point de contour ou la distance euclidienne entre deux snakels successifs. Figure 4 en donne l'exemple. L'écart entre deux snakels successifs dépend de la distance par rapport au centre. Si le nombre des points de contour entre deux snakels est fixe, le problème est résolu, mais lorsque le contour s'étend diagonalement, dans ce cas, la distance augmente.

Un contour actif est basé sur une minimisation de l'énergie de snake, elle est définie ainsi : $E_{snake} = \sum_{i=0}^{M-1} E_{int}(v_i) + E_{ext}(v_i)$ (1.23)

Où : $E_{int}(v_i)\ et\ E_{ext}(v_i)$ dénotent l'énergie intérieure et extérieure de snake v_i, elles définissent son comportement dynamique. Une minimisation de l'énergie de snake détermine sa position optimale.

L'énergie intérieure assure que le contour soit résistant aux étirements et garantit un comportement régulier de snake, elle est définie comme suit :

$$E_{int}(v_i) = w_1 \times \left|\frac{dv_i}{ds}\right|^2 + w_2 \times \left|\frac{d^2v_i}{ds^2}\right|^2 \quad (1.24)$$

Où : $v_i = (x_i, y_i), i = 0, \dots, M-1,$

Les dérivations du premier et second ordre sont définies approximativement par :

$$\left|\frac{dv_i}{ds}\right|^2 \approx 0.5\ (|v_i - v_{i-1}|^2 + |v_i - v_{i+1}|^2) \quad (1.25)$$

$$\left|\frac{d^2v_i}{ds^2}\right|^2 \approx |v_{i-1} - 2.v_i + v_{i+1}|^2 \quad (1.26)$$

Le choix des poids $w_1\ et\ w_2$ est critique, par ce qu'ils régularisent la tension et la rigidité de snake (Leymarie & Levine., 1993,).

Par ailleurs, l'énergie extérieure est une force qui conduit le contour actif ou le snake vers les caractéristiques significatives de visage. En général, l'information sur les bordures est employée dans le guidage de snake. Dans ce contexte, les auteurs dans (Yuille, Hallinan, & Cohen, 1992), utilisent la couleur de peau, c'est-à-dire si la couleur d'un pixel est une couleur de peau, et ceci est en dehors de snake, alors le snakel est tiré à l'extérieur et vice versa.

L'énergie extérieure est définie comme suit :

$$E_{ext}(v_i) = -\sum_{(x,y)\in N_{int}(v_i)} 1 - s(x,y) + \sum_{(x,y)\in N_{ext}(v_i)} s(x,y) \quad (1.27)$$

Où : $s(x, y)$ indique la couleur de la peau.

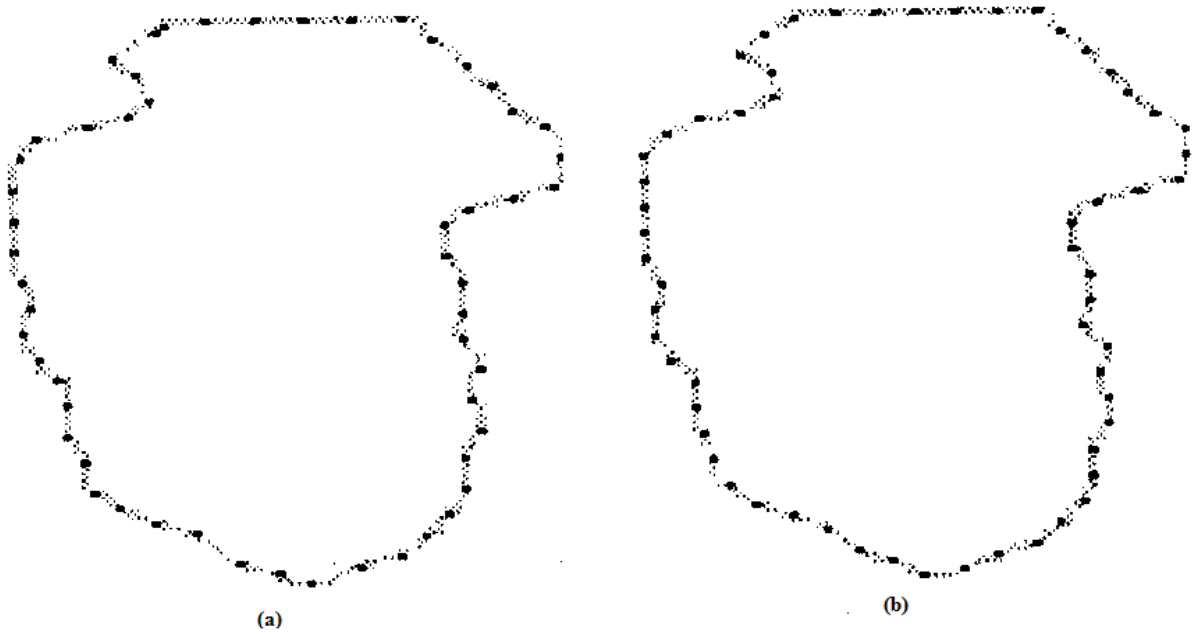

Figure 1.4. Initialisation de snake : (a) nombre constant de point contour, (b) Distance Euclidienne constante

L'implémentation des Snakes est difficile à réaliser, notamment, lorsque l'image est de faible contraste ou ombrée. Une idée proposée dans (Yuille, Hallinan, & Cohen, 1992) consiste à utiliser les snakes en leur incorporant l'information sur les yeux afin d'améliorer la qualité de détection. Les auteurs définissent un modèle déformable des yeux basé sur leurs caractéristiques proéminentes. Une fois le modèle, qui fonctionne avec le même principe des snakes, initialisé à proximité d'une caractéristique des yeux, il se déforme vers les frontières de caractéristiques optimales. Le mécanisme de déformation implique une minimisation de la descente du gradient d'une combinaison d'énergie externe en raison de la vallée, bord, pic, et luminosité de l'image.

$$E = E_v + E_e + E_p + E_i + E_{interne} \quad (1.28)$$

$$E_{internal} = \frac{k_1}{2}(x_e - x_c)^2 + \frac{k_2}{2}(p_1 + \frac{1}{2}\{r + b\})^2 + + \frac{k_2}{2}(p_2 + \frac{1}{2}\{r + b\})^2 + + \frac{k_3}{2}(b - 2r)^2 \quad (1.29)$$

Les coefficients $k_1,\ k_2,\ k_3$ contrôlent la déformation du modèle.

Dans les applications de modèle déformable, certaines mesures doivent être tenues en compte :

1- L'évolution de déformation de modèle dépend de sa position initiale, par exemple, si le modèle est placé au-dessus des yeux, il peut s'attirer vers les sourcils.
2- Le temps d'exécution est également très élevé en raison de la mise en œuvre séquentielle des processus de minimisation.
3- Les coefficients de contrôle sont heuristiques et difficiles à mettre en œuvre.

Modèle de points distribués

Cette approche permet de décrire les formes en se basant sur les statistiques (Cootes, Taylor, Cooper, & Graham, 1995), Le PDM est différent par rapport aux autres modèles actifs de forme. Le contour est doté d'un ensemble de points labélisés, dont les variations sont paramétrées dans une phase initiale d'apprentissage menée sur des objets de différentes tailles et positions. En se basant sur l'analyse en composante principale, les caractéristiques d'un ensemble d'apprentissage sont construites comme un modèle linéaire. Ce dernier comprend la moyenne de toutes les caractéristiques et les principaux modes de variations de chaque point.

$x = \bar{x} + Pv$ (30)

Où x représente un point PDM, $\bar{x}$ est la moyenne des caractéristiques dans un ensemble d'apprentissage pour ce point.

$P = [P_1, P_2, \dots, P_t]$ est une matrice qui contient « t » axe principal de variation.

v , est un vecteur poids défini pour chaque mode.

Lanitis et al (Lanitis, Taylor, & Cootes, 1995) est le premier à mettre en œuvre un PDM pour un visage. Le modèle décrit globalement un visage, inclues ses caractéristiques faciales tels que les sourcils, les yeux, et le nez. Pour ce faire, il utilise 152 points plantés manuellement dans 160 images. Le modèle peut rapprocher jusqu'à 95% des formes de visages dans l'ensemble d'apprentissage.

2 Les approches globales

Les approches structurelles souffrent énormément des problèmes de position et d'orientation de la tête, même aussi aux conditions environnementales, ce qui rend parfois la détection des visages quasiment impossible. En outre, d'autres études ont tendances à des systèmes aptes à

détecter plusieurs visages dans la même scène, et cela se fait dans des conditions critiques ; des milieux pleins d'obstacles et à forte intensité. Cette exigence a inspiré un nouveau domaine de recherche dans lequel la détection de visage est traitée comme un problème de reconnaissance de formes, car c'est un problème d'apprendre à reconnaître un motif de visage à partir d'exemples.

Erik, dans (Hjelmas., 2001) a plus au moins divisé les approches globales dans les méthodes de sous-espaces linéaires, réseaux de neurones, et les approches statistiques.

2.1 Les méthodes de sous-espaces linéaires

La recherche d'un visage dans une image comprend une recherche d'un sous espace, contenant le visage, dans l'espace globale de l'image. Les procédés utilisés pour tel objectif sont appelés les méthodes statistiques. Parmi lesquelles nous citons, l'analyse en composante principale, l'analyse discriminante linéaire, l'analyse factorielle…etc.

Dans cette partie nous allons décrire l'aspect théorique des méthodes basées sur l'analyse en composante principale, et donner ainsi des résultats mentionnés dans quelques approches.

Les auteurs dans (Sirovic & Kirby, 1987) (Kirby & Srovich., 1990), essaient de retenir l'information pertinente pour représenter le visage. Leur approche, en quelque sorte, capture la variation dans une collection d'images. En termes mathématiques, elle cherche les composantes principales dans une collection de visages, ou les vecteurs propres de la matrice de covariance de l'ensemble des images de visages. Ces vecteurs sont appelés « Eigenvector » dont chacun représente une image spectrale appelée « Eigenface ». Les M meilleures « Eigenfaces » représentent un espace de dimension M de toutes les images.

Soit l'image bidimensionnelle $I(x, y)$ à la taille de $N \times N$. Cette image peut être représentée par un vecteur avec une taille de N^2, de façon qu'une image de taille 256 x 256 soit définie par un vecteur de dimension de 65536. Imaginons la taille d'un espace contenant une centaine d'image ?

L'espace, de grande envergure, contenant l'ensemble de toutes les images est réduit en un sous-espace de très faible dimension. Le rôle de l'analyse en composante principale est de trouver, dans l'espace image, les vecteurs qui contribuent fortement à la représentation des visages dans l'espace image.

Soit l'ensemble des images, $\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_M$, utilisé dans l'apprentissage, le visage moyen est donné par $\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_n$ (1.31)

La différence entre chaque visage et le moyen est définie par $\Phi_i = \Gamma_i - \Psi$ (1.32)

Sur l'ensemble Φ , une analyse en composante principale est appliquée pour en extraire les vecteurs ortho-normaux u_n . Chaque vecteur u_k est défini par :

$\lambda_k = \frac{1}{M} \sum_{n=1}^{M} (u_k^T \Phi_n)^2$ (33)

Où : λ_k est le maximum et $u_l u_k = \delta_{lk} = \begin{cases} 1 \; si \; l = k \\ 0 \quad sinon \end{cases}$ (34)

u_k et λ_k sont respectivement, les vecteurs propres et les valeurs propres de la matrice de variances-covariances définie par : $C = \sum_{n=1}^{M} \Phi_n \Phi_n^T$ (35)

Le nombre de vecteurs propres trouvés M, représente la nouvelle taille de l'espace des images. Notez bien que M<<N².

Cette technique a été amplement appliquée, Pentland *et al.* (Pentland, Moghaddam, & Starne., 1994) (Pentland & Choudhury, 2000) l'ont utilisé pour rechercher les caractéristiques faciales. Ils ont obtenu un taux de 94% dans les tests effectués sur une base de visages contenant 7562 images. Autrement, Moghaddam et Pentland (Moghaddam & Pentland, 1997), ont appliqué l'ACP dans un cadre probabiliste et ils ont mentionné un taux de détection de 95% dans les expérimentations menées sur une base de 7000 images. Par ailleurs, Jebara et Pentland (Jebara & Pentland., 1997), ont utilisé l'ACP dans un système de détection des visages basé sur les couleurs et les mouvements.

2.2 Les réseaux de neurones

Les approches neuronales sont largement utilisées pour la détection des visages. Leurs concepts dépassent la notion d'un simple perceptron multi couches, elles impliquent les architectures modulaires, les algorithmes d'apprentissage complexes, les mémoires auto-associatives, les réseaux RBF (réseaux à fonctions de base radiale)…etc.

Rowley et al. (Rowley, Baluja, & Kanade, 1998) Ont développé un système de détection des visages basé sur un réseau de neurones comme montré dans Figure 5. Le réseau reçoit en entrée des images de 20 × 20 pixels, soit 400 unités d'entrée, et contient une seule couche

cachée de 26 unités où 4 unités contrôlent les 10 × 10 pixels, 16 unités pour les 5 × 5 pixel, et 6 unités pour vérifier pixels qui se chevauchent horizontalement.

Les auteurs, dans une autre approche (Rowley & S. Baluja, 1998), ont combiné leur système avec un autre réseau de neurones afin de détecter les visages à tous les angles dans le plan. Ils ont atteint un taux de détection de 79,6% dans les tests menés sur deux grandes bases de visages.

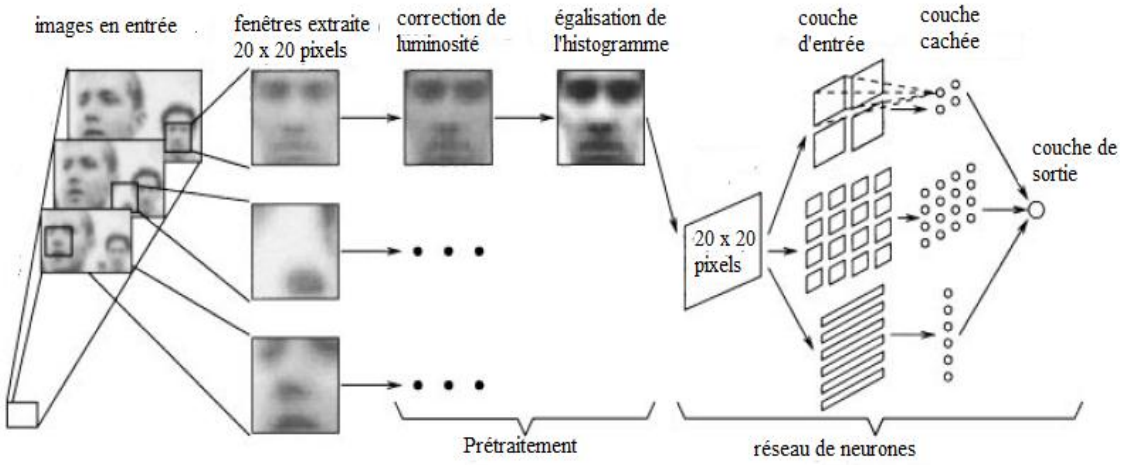

Figure 1.5. Architecture du système proposé par Rowley et al (Rowley & S. Baluja, 1998).

Paul et RON (Paul & Ron., 1996) ont développé une structure hiérarchique de réseaux de neurones pour la détection des visages. Leur système est composé de quatre réseaux, totalement connectés et rétro-propagés, dont trois sont entrainés pour reconnaitre respectivement, les yeux, la bouche, et le nez. Ces derniers sont appelés « réseaux fils ». Le réseau de neurones parent décide de la présence d'un visage si les trois réseaux fils détectent deux yeux, une bouche et un nez.

2.3 Les approches statistiques

les systèmes basés sur les règles de décision bayésienne, les machines à vecteurs de support, et la théorie de l'information sont appelés les approches statistiques.

Colmenarez et Huang (Colmenarez & Huang., 1997), ont développé une approche basée sur la divergence de Kullback; c'est une mesure positive de différence entre deux fonctions de densité de probabilité, P_{X^n} et M_{X^n} , pour un processus aléatoire X^n :

$$H_{P||M} = \sum_{X_n} P_{X_n} \ln \frac{P_{X^n}}{M_{X^n}} \quad (1.36)$$

Durant l'apprentissage, un histogramme est utilisé pour chaque paire de pixels afin de créer les fonctions de probabilité pour les classes « visage » et « non-visage ». La valeur d'un pixel dépend de son voisinage, X_n est considéré comme un processus du premier ordre de Markov, et le niveau de gris d'un pixel est quantifié à quatre niveaux. Les auteurs emploient, dans la phase d'apprentissage, des images 11 × 11 avec et sans visage pour en obtenir un tableau contenant des valeurs de vraisemblances.

Schneiderman and Kanade (Schneiderman & Kanade, 1998) (Schneiderman & Kanade, 2000) ont décrits deux détecteurs de visages basés sur la règle de décision de bayes suivante :

$$\frac{P(image \mid objet)}{P(image \mid non\ objet)} > \frac{P(non\ objet)}{P(\ objet)} \quad (1.37)$$

Si cette relation est vérifiée, alors l'objet localisé est un visage. L'avantage dans cette approche est la représentation exacte de $P(image \mid objet)$ et $P(image \mid non\ objet)$. Les auteurs dans (Duda & Hart, 1973) prouvent également l'optimalité de la règle de bayes.

La fonction de probabilité est basée sur les modifications suivantes :

1- La résolution de l'image est normalisée en 64 × 64 pixels
2- Chaque image est divisée en quatre 16 × 16 sous-régions non chevauchées.
3- Les sous-régions sont projetées sur un 12-dimensionnel sous-espace construit par le moyen d'une analyse en composante principale (ACP).
4- La région de visage est normalisée afin d'avoir la moyenne zéro et une variance unitaire.

Conclusion :

Nous avons brièvement présenté quelques techniques structurelles et globales pour la détection des visages. Ce processus est resté une problématique d'actualité, malgré l'avancement technologique dans ce domaine. Les chercheurs traitent toujours les problèmes de complexité d'environnement, dus aux changements de luminosités et d'échelles.
Les méthodes structurelles sont applicables pour les systèmes en temps réel où la couleur et le mouvement sont disponibles. Une analyse multi-résolution n'est pas toujours préférable, néanmoins, ces méthodes peuvent fournir des indices visuels qui méritent l'attention. Dans ces situations la technique la plus largement utilisée est la détection de la couleur de la peau. Pour le cas des images statiques, les approches analysant les niveaux de gris sont très promoteurs (Rowley & S. Baluja, 1998), montrant de bons résultats de détection en plus de rapidité des calculs.

Sung et Poggio (Sung & Poggio, 1998) considèrent les approches globales comme étant les techniques les plus robustes pour le traitement des images statiques en niveau de gris. Rowley et al. (Rowley & S. Baluja, 1998) ont défini les normes pour la recherche sur ce sujet, et les performances de leurs algorithmes sont encore comparables à la plus récente des approches globales.

Chapitre 2

Reconnaissance des visages

Introduction

La reconnaissance des visages est une phase primordiale inclue dans tout système d'analyse des visages. Ces derniers sont très nombreux et impliqués dans une panoplie de systèmes de sécurité, imagerie médicale, télé-enseignement…etc.

Les travaux de recherche concernés par ce créneau sont divisés en trois parties, à savoir, les approches globales, les approches locales, et enfin les approches hybrides. Chaque famille d'approches présente ses avantages et ses inconvénients vis-à-vis des problèmes liés aux conditions environnementales, le changement de l'échelle, les orientations des images, les positions de la tête…etc.

Les approches globales sont indépendantes des positions (haut ou bas) de la tête et même des orientations de l'image de visage. Ces méthodes sont très efficaces mais nécessitent, généralement, une lourde phase primaire d'apprentissage, et le résultat dépend quelques fois du nombre d'échantillons utilisé; (même la sensibilité aux bruits est à signaler). Autrement, les approches locales, qui sont basées sur la détection des objets du visage, sont très robustes aux changements de luminances. Néanmoins, l'imprédictibilité de la position de la tête ainsi que son orientation peuvent provoquer des lacunes dans le système.

Une alternative consiste à combiner les deux approches afin de profiter des avantages de l'une pour combler les inconvénients de l'autre, d'où vient la notion des approches hybrides.

1 Les approches holistiques :

1.1 Analyse en composante principale « Eigenface » :

Cette méthode expliquée dans le chapitre précédent, est utilisée également dans la reconnaissance des visages. Kirby et Sirovic (Kirby & Srovich., 1990) sont les premiers à mettre en œuvre cette conception, c'est un codage des informations pertinentes de l'image, suivi d'une comparaison avec une base d'images codées de la même façon. La théorie de « Eigenface » est inclue dans plusieurs approches de reconnaissance de visages (Moghaddam, Wahid, & Pentland, 1998) (Zhang, Yan, & Lades., 1997) et de l'émotion (Padgett & Cottrell, 1996). Les auteurs dans (Park, Oh, Ahn, & Lee., 2005), l'utilisent pour reconstruire un visage sans les lunettes qu'il porte.

Dans le chapitre précédent, on a montré les étapes à suivre pour réduire la taille de l'espace image. La nouvelle image est obtenue par projection du visage original dans l'espace de visages par la simple opération suivante : $w_k = U_k^T (\Gamma - \Psi)$ (38), avec $k = 1..M'$, où M', est le nombre de vecteurs propres correspondant aux valeurs propres significatives.

Selon Turk et Pentland (Turk & Pentland, 1991), le processus de classification d'un visage comprend les étapes suivantes :

1. Collecter un ensemble de caractéristiques des images pour un nombre connu d'individu. Les images de la même personne doivent être variées en termes de luminance et d'expressions faciales. Le nombre total des images est M.
2. Calculer la matrice de variances/covariances, puis choisir les M' vecteurs propres qui correspondent aux valeurs propres les plus élevées.
3. Combiner la matrice des images d'apprentissage avec les vecteurs propres pour obtenir les Eigenfaces.
4. Définir la classe de chaque individu de la population. Pour ce faire, on calcule pour chaque classe le vecteur de classe Ω_k défini par : $\Omega^T = [w_1, w_2, .., w_{M'}]$
5. Choisir le seuil θ_ϵ qui définit la distance maximale entre l'individu et la classe
6. Soit ϵ la distance entre l'image et l'espace des visages définie par :

   $\epsilon^2 = \left\|\Phi - \Phi_f\right\|^2$ (2.1) $\Phi = \Gamma - \Psi$ (40) et $\Psi_f = \sum_{i=1}^{M'} w_i u_i$ (2.2), soit aussi, $\epsilon_k^2 = \|\Omega - \Omega_k\|^2$, (2.3)

   Pour un nouveau visage, si $\epsilon_k < \theta_\epsilon$ et $\epsilon < \theta_\epsilon$ alors ce visage appartient à la classe k, sinon il est considéré comme un visage inconnu et utilisé en tant que nouvelle classe.
7. Si un nouveau visage est ajouté à une classe donnée, les Eigenfaces doivent être recalculés.

1.2 Eigenfaces probabilistes:

Les approches basées sur la distance Euclidienne ou la corrélation normalisée ne reflètent pas l'idée idéale de comparaison entre images. Parce qu'elles n'exploitent pas les caractéristiques dont les variations sont essentielles dans l'expression de similarité.

Moghaddam et al dans (Moghaddam, Wahid, & Pentland, 1998), définissent la mesure de similarité probabiliste, $\Delta = I_1 - I_2$ (2.4). L'idée consiste en une séparation de l'espace image en deux parties, à savoir, les variations intra-personnelles Ω_I , et les variations extra-personnelles Ω_E.

La mesure de similarité est donnée par : $S(I_1, I_2) = P(\Delta \in \Omega_I) = P(\Omega_I|\Delta)$ (2.5) où : $P(\Omega_I|\Delta)$ est la probabilité donnée par la règle de Bayes.

Les variations intra-personnelles représentent les variations des caractéristiques pour la même personne dans de différentes conditions environnementales, tandis que les variations extra-personnels sont les changements de caractéristiques entre deux personnes différents.

La similarité entre deux images est basée sur les fonctions de vraisemblance, $P(\Delta|\Omega_I)$, et $P(\Delta|\Omega_E)$. Elle est définie par $S(I_1, I_2) = P(\Omega_I|\Delta) = \frac{P(\Delta|\Omega_I)P(\Omega_I)}{P(\Delta|\Omega_I)P(\Omega_I) + (\Delta|\Omega_E)P(\Omega_E)}$ (2.6)

Le processus de comparaison, qui ressemble à un problème de classification binaire des formes, est très simple et résolu en utilisant la règle du maximum à postériori. Deux images représentent la même personne si $P(\Omega_I|\Delta) > P(\Omega_E|\Delta)$ ou si $S(I_1, I_2) > \frac{1}{2}$ (2.7)

Le problème des approches probabilistes est qu'elles consomment trop de mémoire, pour en remédier, Moghaddam et Pentland (Pentland, Moghaddam, & Starne., 1994) divisent l'espace R^N des caractéristiques en utilisant l'analyse en composante principale qui permet de réduire sa taille en $M \ll N$. Figure 2.1. L'espace R^N est divisé en deux sous-espaces : le sous-espace principal F qui contient les M composantes principales, et son complément orthogonal $\overline{F}$. Les composants de F représentent la distances entre l'espace des caractéristiques, et ceux de $\overline{F}$ représentent la distance dans l'espace de caractéristiques.

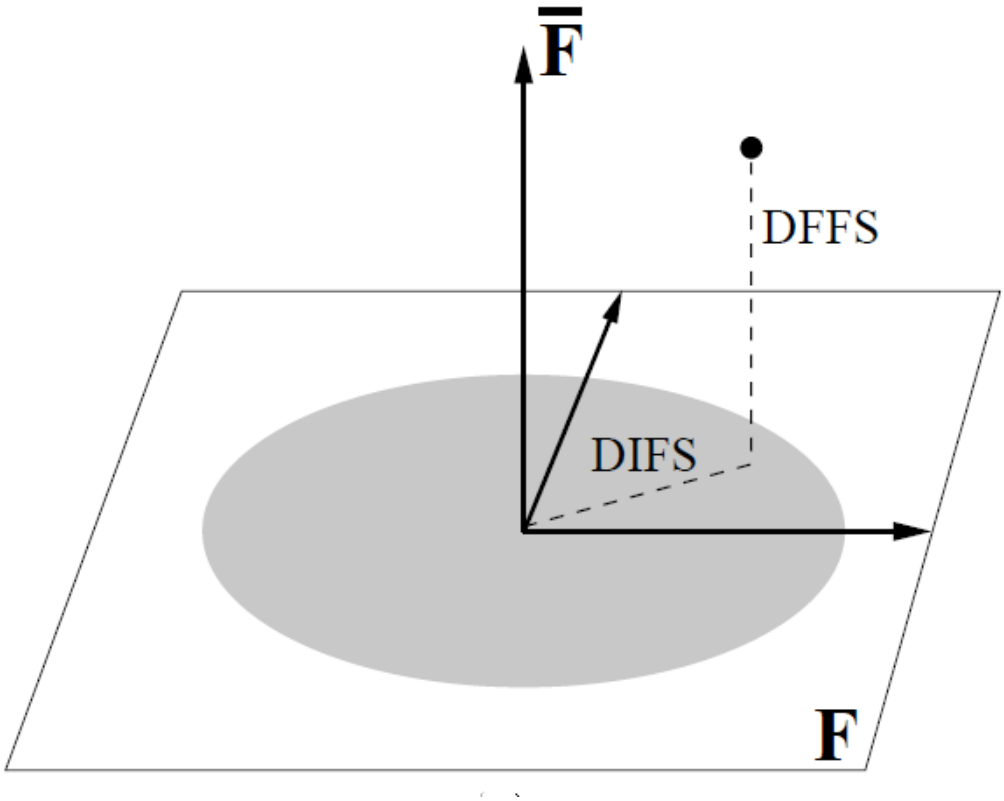

Figure 2.1. Décomposition de R^N en un sous-espace principal F et son complément orthogonal $\bar{F}$.

1.3 Analyse discriminante linéaire LDA

L'analyse discriminante linéaire est un très bon outil statistique, elle est appelée également, la méthode discriminante de Fisher (Sirovic & Kirby, 1987) (Belhumeur, Hespanha, & Kriegman., 1997). Cette méthode prend en compte les différentes variables d'un objet (visage), et cherche par la suite son groupe d'appartenance. Dans ce qui suit, nous détaillons l'algorithme d'analyse discriminante linéaire utilisé dans la reconnaissance des visages.

Soit l'ensemble d'apprentissage qui contient N images représentant c différents visages, les images sont représentées par des tableaux bidimensionnels et transformées sous forme de vecteurs de taille n. Chaque classe comprend plusieurs instances pour le même objet, et les objets distincts doivent appartenir à des classes différentes.

Les matrices de dispersion inter-classes S_B et intra-classe S_W sont définies ainsi :

$$S_B = \sum_{i=1}^{c} P_i(\mu_i - \mu)(\mu_i - \mu)^T \quad (2.8)$$

$$S_W = \sum_{i=1}^{c} P_i \sum_{j=1}^{N_i} (x_j^i - \mu_i)(x_j^i - \mu_i)^T \quad (2.9)$$

Où : x_j^i est le j^{ème} vecteur échantillon appartenant à la classe i, μ_i est le vecteur moyenne de la classe i, et μ est la moyenne globale de tous les échantillons.

$\mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_j^i$ (2.10) $\mu = \frac{1}{\sum_{i=1}^{c} N_i} \sum_{i=1}^{c} \sum_{j=1}^{N_i} x_j^i$ (2.11)

On choisit la matrice $W^{opt} \in R^{n \times k}$de colonnes orthogonales qui maximise le rapport du déterminant de la matrice de dispersion inter-classes sur le déterminant de la matrice de dispersion intra-classe.

$W^{opt} = argmax_w \frac{W^T S_B W}{W^T S_W W} = [w_1, w_2, \dots, w_k]$ (2.12), avec $w_i | i = 1,2,..k$ est l'ensemble des vecteurs propres de S_B et S_w correspondants aux valeurs propres $\lambda_i | i = 1,2,..k$ e

$S_B w_i = \lambda_i S_w w_i$ (2.13)

La classification dans l'analyse discriminante linéaire se déroule de la même façon que dans l'analyse en composante principale, l'image test est projetée dans le sous-espace Fisherface, puis la reconnaissance est faite par la distance Euclidienne. Les auteurs dans (Etemad & Chellappa., 1996) utilisent la moyenne absolue pondérée.

1.4 Les machines à support de vecteurs SVM

le SVM est un excellent outil qui permet de séparer entre deux classes, proposé par (Osuna, Freund, & girosi., 1997), ce classifieur est amplement utilisé dans la reconnaissance de formes (Osuna, Freund, & girosi., 1997) (Pontil & Verri., 1998) (Aouatif, Rziza, & Driss, 2008). La justification intuitive de cette méthode d'apprentissage est la suivante : si l'échantillon d'apprentissage est linéairement séparable, il semble naturel de séparer parfaitement les éléments des deux classes de telle sorte qu'ils soient le plus loin possible de la frontière choisie.

Pour résoudre un problème de classification à deux classes, il s'agit de trouver l'hyperplan le plus éloigné des points les plus proches de chaque classe. Ce qu'on appelle « l'hyperplan séparateur optimal » Figure 2.2.

Soit l'ensemble d'apprentissage : $\{(x_1, y_1), (x_2, y_2), \dots (x_n, y_n)\}$, où $x_i \in R^N$ et $y_i \in \{-1, +1\}$ et l'hyperplan : $wx + b = 0$ (2.14). On dit que l'ensemble des vecteurs est séparé d'une façon optimale par l'hyperplan, s'il est séparé sans erreur et avec une marge maximale.

L'hyperplan doit respecter les conditions suivantes :

$y_i[(w, x_i) + b] \geq 1, i = 1 \dots l$ (2.15)

La distance entre le point x et l'hyperplan est :

$d(w.b:x) = \frac{|w.x+b|}{\|w\|}$ (2.16)

La marge est $\frac{2}{\|w\|}$, et l'hyperplan optimal qui sépare les données est celui qui minimise $\Phi(w) = \frac{1}{2}\|w\|^2$ (2.17)

La solution est donnée par la fonction de Lagrange suivante :

$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i\{y_i[w.x_i + b] - 1\}$ (2.18), où : α_i sont les multiplicateurs de Lagrange.

La solution de Lagrange pour un problème dual est donnée par :

$\alpha = argmin_\alpha \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i\alpha_j Y_i Y_j X_i X_j$ (2.19)

Avec, $\alpha_i \geq 0, i = 1 \dots l$ et $\sum_{i=1}^{l} \alpha_i Y_i = 0$ (2.20).

Un réseau SVM ne résout qu'un problème de classification binaire. Lorsqu'on traite un problème multi-classes, telle que la reconnaissance des visages, une combinaison des SVM est accommodée. Il faut tenir aussi en compte que ces formules ne sont valables que pour les situations qui peuvent être séparées linéairement. Bien qu'on n'a pas présenté les cas non-linéaires, plus d'illustrations se trouvent dans (Aouatif, Rziza, & Driss, 2008).

1.5 Les lignes caractéristiques

La méthode de la ligne caractéristique la plus proche (Nearest Feature line) NFL, suppose qu'au moins deux points caractéristiques de prototype sont disponibles pour chaque classe, ce qui est généralement satisfait (Feng, Pan, & Yan., 2012) (Li & Lu., 1999). Elle tente de généraliser la capacité de représentation des prototypes disponibles pour faire face à divers changements en utilisant l'interpolation et l'extrapolation linéaire entre les points caractéristiques.

Soit une variation dans l'espace image depuis le point Z_1 jusqu'à Z_2, ce qui fait la variation dans l'espace caractéristiques depuis le point x_1 jusqu'à x_2. Le degré de changement est mesuré par : $\delta Z = \|Z_2 - Z_1\|$ (2.21) ou $\delta x = \|x_2 - x_1\|$ (2.22)

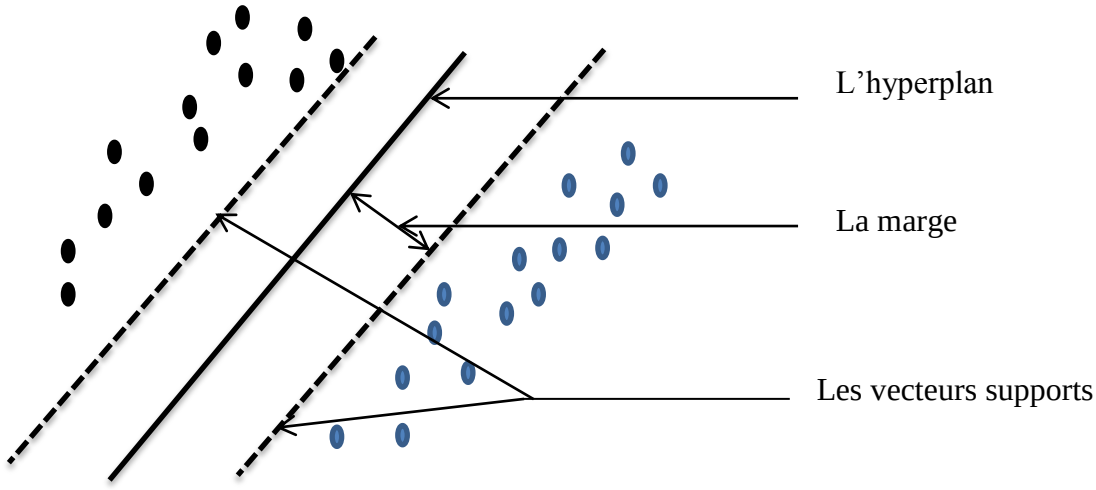

Figure 2.2. Séparation entre deux classes par un hyperplan

Quand $\delta Z \to 0$ alors $\delta x \to 0$

Suite aux changements, x est localisé approximativement, par une ligne droite, entre x_1 et x_2.

On définit par ligne caractéristique, la ligne droite qui traverse x_1 et x_2 , elle est notée par $\overline{x_1 x_2}$.

Si un point x est projeté dans l'espace « ligne caractéristique » en tant que point p, la distance entre x et $\overline{x_1 x_2}$ est définie par : $d(x, \overline{x_1 x_2}) = \|x - p\|$ (2.23), Figure 2.3.

Le point p est calculé ainsi : $p = x_1 + \mu(x_2 - x_1)$ (62), avec $\mu \in R$ est un paramètre de position calculé comme suit : comme $\overline{px}$ est perpendiculaire à $\overline{x_1 x_2}$ donc, $(p - x).(x_1 - x_2) = [x_2 + \mu(x_2 - x_1) - x].(x_2 - x_1) = 0$ (2.24) alors,

$$\mu = \frac{(x - x_1).(x_2 - x_1)}{(x_2 - x_1).(x_2 - x_1)} \quad (2.25)$$

μ, décrit la position de p relative à x_1 et x_2 :

1- Si $\mu = 0$ alors $p = x_1$.
2- Si $\mu = 1$ alors $p = x_2$.

3- Si $0 < \mu < 1$ alors p est un point d'interpolation entre x_1 et x_2.
4- Si $\mu > 1$, alors p est un point de l'extrapolation vers l'avant sur le côté de x_2.
5- Si $\mu < 0$, alors p est un point de l'extrapolation vers l'arrière sur le côté de x_1.

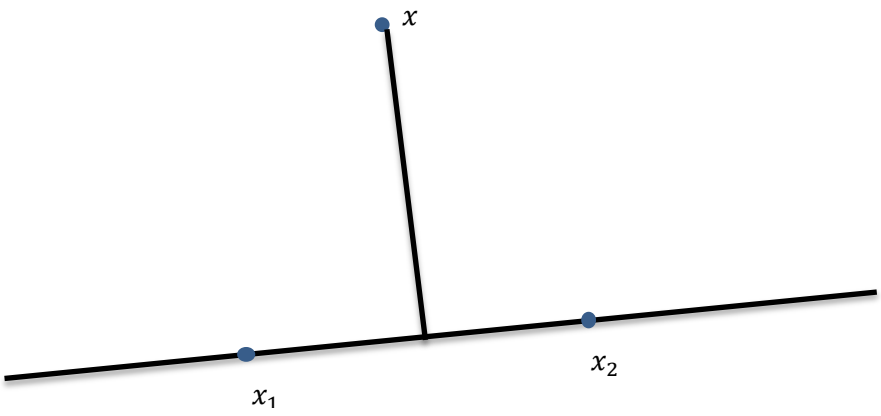

Figure 2.3. Représentation des points prototypes x_1 et x_2 par la ligne caractéristique $\overline{x_1 x_2}$

Le nombre de lignes caractéristiques dépend du nombre des caractéristiques, supposons que ceci est représenté par N_c et est supérieur à un, le nombre de ligne correspondant, est égale à $K_c = \frac{N_c(N_c-1)}{2}$ (2.26), le nombre de ligne pour Mclasses est : $N_{total} = \sum_{c=1}^{M} K_c$ (2.27)

1.6 Analyse en composantes indépendantes ICA

L'analyse en composantes indépendantes joue le même rôle que l'analyse en composantes principales, les deux font une projection linéaire depuis l'espace R^N vers R^M mais la première minimise les dépendances d'un ordre plus élevé avec les propriétés suivantes :

1. La reconstruction de l'image est approximative : $x \approx Ay$ (2.28)
2. les vecteurs propres sont non-orthogonaux : $A^T A \neq I$ (2.29)

L'analyse en composants indépendants consiste en une décomposition du signal en entrée (l'image) X en une combinaison linéaire de signaux de source indépendants (Fei, Jinsong, Xueyi, Zhenquan, & Bin, 2006) (Black & Yacoob, 1995). Par hypothèse, $X^T = As^T$ (2.30), où A est la matrice de mixage.

L'algorithme de L'ICA sert à trouver A ou la matrice de séparation W, tel que, $u^T = WX^T = WAs^T$ (2.31), avec les données en entrée qui sont organisées en M observations avec N variables, ce qui fait une matrice $N \times M$.

2. Les Approches locales (basée caractéristiques)

2.1 Les méthodes géométriques

Ces méthodes sont basées sur les caractéristiques locales du visage, L'analyse du visage humain est donnée par la description individuelle de ses parties et de leurs relations.
Le modèle géométrique imite la perception humaine, il se base sur les traits de visage ainsi que ses composants comme le nez, la bouche, les yeux, et les sourcils.

Les travaux de recherches réalisés (Brunnelli & Poggio., 1993) (Yuille, Hallinan, & Cohen, 1992) (Cootes, Taylor, Cooper, & Graham, 1995), font une extraction des caractéristiques faciales à partir d'une image de visage, et utilisent également un modèle de visage. Une modélisation de visage consiste en une définition des distances et des angles entre les points caractéristiques, cela peut se faire d'une façon automatique ou semi-automatique selon que la phase d'extraction des points est faite par l'ordinateur ou qu'elle est assistée par un utilisateur. La robustesse du système dépend de la façon d'extraction, de ces points caractéristiques ainsi que de leur distribution.

Au début des années 1990, Brunelli et Poggio (Brunnelli & Poggio., 1993) ont décrit un système de reconnaissance faciale qui extrait automatiquement 35 caractéristiques géométriques du visage. La similitude est calculée à l'aide de classifieurs de Bayes. Un taux d'identification de 90 % sur une base de données de 47 sujets a été rapporté par les auteurs. Le coût de stockage des techniques géométriques est très bas comparé à celui des autres techniques. Toutefois, les approches purement géométriques présentent quelques inconvénients, notamment :

1. les caractéristiques géométriques sont généralement difficiles à extraire, surtout dans des cas complexes : illumination variable, occultations, etc.
2. les caractéristiques géométriques seules ne suffisent pas pour représenter un visage, tandis que d'autres informations utiles comme les niveaux de gris de l'image ne sont pas du tout exploitées.

2.2 Elalstic Buch Graph Matching (EBGM)

La technique d'Elalstic Bunch Graph Matching (EBGM) introduite par (Wiskott, Fellous, Kruger, & Malsburg, 1997), utilise un seul graphe pour représenter les différentes variations d'apparence de visage. Chaque nœud contient un ensemble de 40 coefficients complexes d'Ondelette de Gabor, incluant la phase et l'amplitude. Ces coefficients sont connus sous le nom de jet. Ainsi, la géométrie d'un objet est codée par les arêtes du graphe, alors que les nœuds (jets) codent les variations des niveaux de gris.

La recherche des points de repère dans un nouveau visage nécessite une représentation générale au lieu d'un modèle spécifique de visage. Cette représentation doit tenir compte de différentes variations de formes des yeux, des nez et des lèvres, les différentes formes de barbes pour certains cas, la différence de sexe et de race…etc.

Deux inconvénients de la méthode EBGM ont été mis en évidence. Premièrement, elle exige un temps de calcul supérieur aux autres méthodes ce qui la rend plus difficile à mettre en œuvre en pratique. Et deuxièmement, seules les informations sur les positions clefs de l'image (par exemple : les yeux, le nez, la bouche) sont utilisées pour l'identification. Bien que ce soit un facteur crucial qui contribue à la robustesse de la méthode, la manière dont cette méthode gère une situation où les caractéristiques clefs sont occultées n'est pas claire.

Des améliorations ont été apportées à ce modèle par Kepenekci et al. (Kepenekci, Tek, & Akar, 2002), ils proposent une méthode basée sur les caractéristiques de Gabor. Au lieu de fixer le nombre de points caractéristiques du visage comme dans la méthode EBGM, ils ont utilisé un ensemble de matrices de filtres de Gabor pour parcourir les régions faciales locales. Les points caractéristiques obtenus avec la réponse fréquentielle la plus haute du filtre de Gabor sont automatiquement choisis pour être des candidats à la représentation de visage.

Puisque les points caractéristiques résultants sont différents d'un visage à un autre, la possibilité de trouver des classes spécifiques de caractéristiques s'en trouve donc augmentée. En plus des valeurs de la réponse de Gabor, la position de chaque point caractéristique est enregistrée, considérant ainsi implicitement la structure spatiale du visage. Des résultats expérimentaux sur l'ensemble de la base de données ORL montrent un taux d'identification de 95.25 % avec seulement une image d'apprentissage par personne. Un deuxième test sur la base FERET a démontré que cette méthode est moins coûteuse en de temps de calcul que la méthode EBGM. Cependant, sa flexibilité dans la détection des points d'intérêt augmente le risque des faux appariements en raison de la possibilité de la non-existence des caractéristiques dans la zone locale considérée.

Les méthodes basées sur les caractéristiques locales sont efficaces. Cependant leurs performances dépendent essentiellement de la précision de la localisation des points caractéristiques. Cette tâche reste très difficile en pratique, plus particulièrement dans des situations où la forme et l'apparence du visage peuvent fortement changer. Par exemple, la

sur-illumination peut provoquer une réflexion spéculaire sur le visage. Pour résoudre ce problème des méthodes basées sur l'apparence locale sont utilisées.

2.3 Les modèles de Markov (HMM) :

Les modèles cachés de Markov est un ensemble de modèles statistiques qui caractérisent les propriétés d'un signal (Nefian & Hayes, 1998). Un HMM consiste en un nombre fini d'état, une matrice des probabilités de transitions, un état initial de distribution des probabilités, et un ensemble de fonctions de densité de probabilité associé à chaque état.

Les éléments définissant un HMM sont les suivants :

1- Le nombre N d'état dans le modèle, S est l'ensemble des états alors $S = \{S_1, S_2, \dots S_n\}$, l'état du modèle à l'instant t est $q_t \in S$ avec $1 \leq t \leq T$, T est la durée de la séquence d'observation.

2- Distribution initiale des états, $\Pi = \{\pi_i\}$, avec $\pi_i = P[q_1 = S_i], 1 \leq i \leq N$ (2.32)

3- Matrice de probabilités de transition $A = \{a_{ij}\}$ et $a_{ij} = P\left[q_t = S_j \middle| q_{t-1} = S_i\right]$ (2.33), avec $1 \leq i, j \leq N$ et $0 \leq a_{ij} \leq 1$ et $\sum_{j=1}^{N} a_{ij} = 1$ (2.34)

La représentation d'un visage par un HMM est une tâche très délicate, avant d'en parler nous devons signaler que les HMM sont déjà pertinemment utilisés dans la reconnaissance de la parole ainsi que la définition des gestes. L'image d'un visage doit être divisée en régions faciales (les cheveux, le front, les yeux, le nez, les lèvres). Toutes les régions sont considérées par ordre même si l'image subissait des faibles, voire des fortes rotations. Chaque région faciale est assignée à un état HMM. Figure 2.4 montre l'ensemble des états reliés par probabilités de transition.

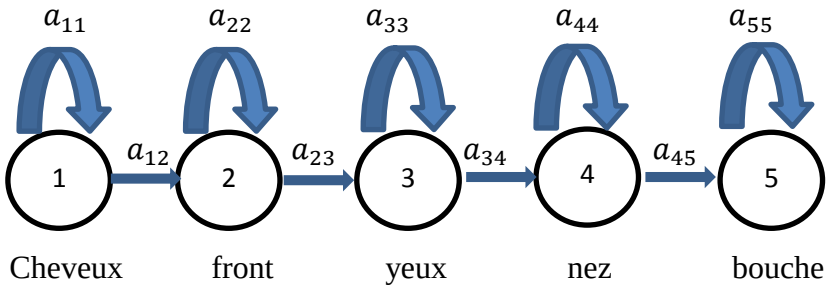

Figure 2.4 : HMM pour la reconnaissance des visages

2.4 Méthode de LBP (Local Binary Pattern)

La description LBP d'un visage est introduite par Ojala et al (Ojala, Pietikainen, & Harwood, 1996). Chaque pixel est labellisé en fonction de son voisinage 3×3 ou autres en considérant sa valeur binaire. Par la suite un histogramme des labels est conçu, et utilisé en tant que descripteur de texture.

Les approches actuelles ont appliqué l'opérateur LBP avec des voisinages de tailles différentes (Ojala, Pietikainen, & Maenpaa, 2002). En utilisant l'interpolation bilinéaire des valeurs des pixels et le voisinage circulaire, on peut se permettre d'utiliser n'importe quel rayon et nombre de pixels dans le voisinage.

Un voisinage est noté par (P, R) ce qui veut dire : P pixel dans le cercle de rayon R. Les Figures 10,11 montrent respectivement, le fonctionnement d'un LBP de base et un LBP circulaire (8,2).

Un code LBP est uniforme s'il contient au plus deux transition de 0 à 1 ou vice versa lorsque la chaîne binaire est considéré comme circulaire. Par exemple, 000000000 , 000001111000, 11111111000 sont des codes uniformes. Ojala et al estime que 90% des codes sont uniformes lorsqu'on utilise un voisinage (8,1), et 70% dans un voisinage (16,2).

L'opérateur LBP est noté par $LBP_{P,R}^{u2}$, $u2$ veut dire que tous les codes utilisés sont uniformes et le reste sont représentés par un seul code, et le voisinage utilisé est (P, R) .

Une fois tous les pixels sont codés, un histogramme est conçu afin de représenter l'image. Cet histogramme est défini par la relation suivante :

$$H_i = \sum_{x,y} I\{f(x,y) = i\}, i = 0 \dots n-1 \quad (2.35)$$

Où n est le nombre de labels produits par l'opérateur LBP, et

$$I\{A\} = \begin{cases} 1, & si\ A\ est\ vrai \\ 0, & sinon \end{cases} \quad (2.36)$$

Une autre idée consiste à diviser l'image en régions : $R_0, R_1, \dots, R_{m-1}$, afin de construire l'histogramme augmenté définit ainsi :

$$H_{i,j} = \sum_{x,y} I\{f(x,y) = i\}\ I\{(x,y) \in R_i\}, i = 0 \dots n-1, j = 0 \dots m-1 \quad (2.37)$$

La classification des patterns codés en LBP est fondée sur le principe du plus proche voisin, les trois mesures de dissimilarité les plus utilisés sont cité ci-dessous :

1- L'intersection d'histogramme : $D(S, M) = \sum_i \min(S_i, M_i)$ (2.38)
2- Log probabiliste : $L(S, M) = \sum_i S_i\, log M_i)$ (2.39)

3- Chi carrée statistiques : $\chi^2(S,M) = \sum_i \frac{(S_i - M_i)^2}{S_i + M_i}$ (2.40)

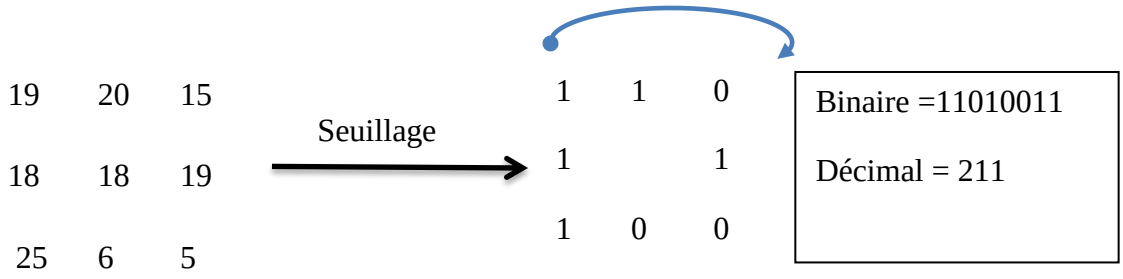

Figure 2.5. LBP de base

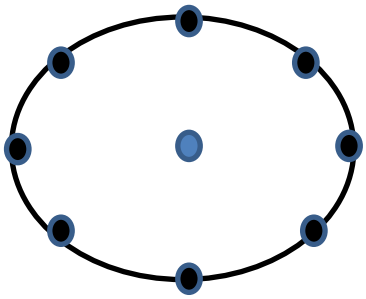

Figure 2.6. Voisinage circulaire(8,2)

3. Les méthodes hybrides

3.1 Eigenfaces modulaire

Cette approche a été introduite par Pentland et al. (Pentland, Moghaddam, & Starne., 1994) (Pentland & Choudhury., 2000) Le visage est divisé en trois régions faciales, à savoir, le visage en sa totalité, les yeux et le nez. Une analyse en composante principale est appliquée sur chacune de ces régions et les résultats de classifications obtenus sont agrégés. La bouche est exclue de la sélection car elle est trop sensible à des changements d'expression faciale, sa prise en compte peut engendrer une baisse du taux de reconnaissance.

Cette approche peut être qualifiée d'hybride, puisqu'elle utilise à la fois des caractéristiques globales et locales. Les auteurs ont montré qu'elle est plus efficace que les techniques globales ou strictement locales appliquées séparément.

Le problème des méthodes de discrimination linéaire réside dans le fait que le nombre d'échantillon doit être trop élevé. Chen et al. (Chen, Liu, & Zhou, 2004) ont proposé d'appliquer la méthode d'analyse de discrimination linéaire sur un exemple de petite taille. Ils ont partitionné chaque image de visage en un ensemble d'imagettes de même taille. Pour chaque classe, un ensemble d'imagettes est pris pour l'apprentissage.

Dans (Price, R.Jeffery, Gee, & Timothy, 2005), Price et Gee ont introduit une technique modulaire basée sur une variante de l'analyse discriminante linéaire. Les régions sélectionnées sont : la région faciale dans son ensemble, une bande faciale (de même largeur que la région faciale) s'étalant du front jusqu'au-dessous du nez, et une bande faciale contenant les yeux. Les résultats expérimentaux montrent que cette approche est plus performante que les techniques des Eigenfaces et des Fisherfaces, elle est notamment robuste aux changements dans les conditions d'illumination de visage, d'expression faciale et d'occlusion.

3.2 Les modèles d'apparences flexibles :

Ces méthodes utilisent des modèles flexibles pour la représentation des visages. Ces modèles sont contrôlés par un nombre réduit de paramètres qui sont utilisés pour coder l'apparence du visage dans l'image. Ces paramètres contrôlent également les variations inter-classes et intra-classes qui sont très utiles dans la phase de classification.

Les auteurs dans (Jingru & W.Y., 1999) ont défini un modèle représentant le visage humain en utilisant 152 point et 160 exemples d'apprentissage dont 8 échantillons pour 20 individus. Figure 12, montre les exemples apprentissages typiques, la forme moyenne, et les emplacements des points du modèle.

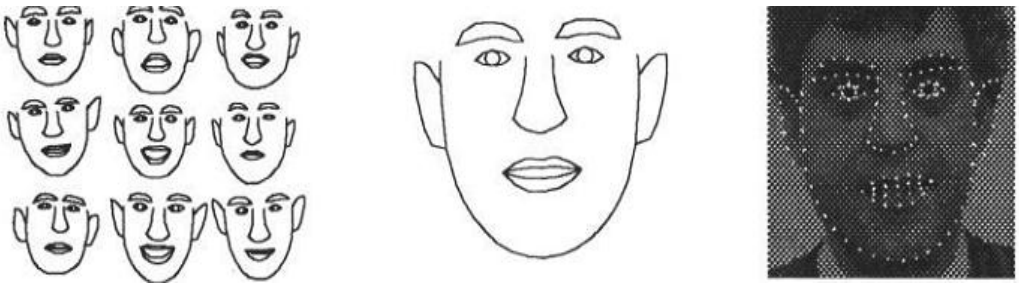

Figure 2.7. (a) (b) (c)

a : Formes typiques d'apprentissage, b : Visage moyen, c : Emplacement des points du model

3.3 Méthode linéaire hybride de Fisher

Des systèmes pratiques utilisent des méthodes hybrides de l'analyse en composantes principales et l'analyse linéaire de Fisher (Yang M. , 2002). Ce point de vue a été longtemps tenu à la communauté de la psychologie. Il semble mieux estimer eigenmodes / eigenfaces qui ont de grandes valeurs propres (et donc plus robustes contre le bruit), alors que pour une estimation de eigenmodes d'ordre supérieur, il est préférable d'utiliser l'analyse linéaire de Fisher. Pour soutenir ce point de vue, Penev et Atick (Penev & Atick, 1996) utilisent des eigenpictures principaux globales, intégrant des filtres de lissage qui sont efficaces dans la suppression du bruit.

L'analyse linéaire de Fisher est une méthode d'analyse des caractéristiques d'inspiration biologique (Yang M. , 2002). Sa motivation biologique vient du fait que, si une grande gamme de récepteurs (plus de six millions cônes) existe dans la rétine humaine, seulement une petite fraction d'entre eux est active, correspondant à des objets / signaux naturels qui sont statistiquement redondants. A partir de l'activité de ces récepteurs distribués, le cerveau doit découvrir où et comment les objets se trouvent dans le champ de vision et de récupérer leurs attributs. Par conséquent, on s'attend à représenter les objets naturels ou les signaux dans un sous-espace de dimension inférieure par le moyen d'un paramétrage approprié. Pour une catégorie limitée d'objets tels que des visages qui sont correctement alignés et mis à l'échelle, la dimensionnalité peut être encore plus faible. Un bon exemple est l'utilisation réussie de l'expansion de l'analyse en composante principale pour rapprocher des images de visages dans un sous-espace linéaire (Moghaddam & Pentland, 1997).

La région faciale est considérée comme un tableau à deux dimensions rempli par des récepteurs dont chacun correspond à un emplacement dans le visage, mais certains de ces récepteurs peuvent être inactifs. L'analyse linéaire de Fisher est utilisée pour extraire les caractéristiques topographiques locales contrairement à l'analyse en composantes principales.

La recherche de la meilleure topographie est basée sur l'erreur de reconstruction et appelé « sparsification » Penev et Atick (Penev & Atick, 1996). Deux points intéressants sont démontrés dans leur document:

1- En utilisant le même nombre d'axes principaux, la qualité de la reconstruction de la perception de l'analyse linéaire de Fisher est meilleure que celle de l'analyse en composantes principales.

2- Maintenir le deuxième eigenmodel de l'analyse en composantes principales dans la reconstruction à partir de l'analyse linéaire de Fisher réduit encore mieux l'erreur quadratique moyenne, ce qui suggère l'utilisation hybride de l'ACP et LFA.

3.4 Méthodes basées composants :

Cette méthode est fondée sur les progrès récents de reconnaissance à base de composants (Heisele, Serre, Pontil, & Poggio, 2001). L'idée de base de ce travail de recherche est de décomposer un visage dans un ensemble de composants tels que la bouche et les yeux qui sont interconnectés par un modèle géométrique flexible.

La motivation de l'utilisation des composants consiste en ce que les changements des positions de la tête conduisent principalement à des changements dans les positions des éléments du visage. Cependant, un inconvénient majeur de ce système est qu'il nécessite un grand nombre d'images d'apprentissage prises sous différents angles et dans de différentes conditions d'éclairage. Pour surmonter ce problème, le modèle de visage 3D (Roobaert, Nillius, & Eklundh, 2000) est appliqué pour générer des images de synthèse arbitraires sous différentes positions et variations de l'éclairage.

Seulement trois images de visage (frontale, semi-profil, profil) d'une personne sont nécessaires pour calculer le modèle de visage 3D. Une fois que le modèle 3D est construit, les images de synthèse sont générées pour l'apprentissage du classifieur. Quatorze composants du visage ont été utilisés pour la détection de visage, mais seulement neuf composantes qui n'ont pas été fortement engagés et contenaient des structures en niveau

de gris ont été utilisés pour la classification. En outre, la région contenant le visage a été ajoutée aux neuf composants pour former un vecteur caractéristique unique utilisé ensuite par un classifieur SVM (Pontil & Verri., 1998). L'apprentissage est fait sur trois images tandis que les tests sont élaborés sur 200 images par personne.

Conclusion

Dans ce chapitre nos avons brièvement présenté quelques méthodes de reconnaissance des visages. Les méthodes globales considèrent l'image dans sa globalité et procèdent généralement par une réduction de l'espace de données en définissant les axes principaux. Ces méthodes traitent mal les cas assujettis à des forts changements de luminosité ou de faibles contrastes… etc., mais très efficaces en cas d'occlusion partiale ou orientation des visages. Autrement, les méthodes locales traitent les images localement, certaines détectent les composants du visage (les yeux, le nez, et la bouche) et essayent d'en trouver des relations géométriques, d'autres distribuent des points à travers la surface faciale puis effectuent un codage correspondant aux emplacements désignés. Ces approches sont très robustes aux bruits, mais souffrent quand même des problèmes liés aux changements d'échelles et de fortes orientations et inclinaisons. Les méthodes hybrides sont des approches qui combinent les caractéristiques holistiques et locales afin d'améliorer les performances de la reconnaissance de visages. En effet, les caractéristiques locales et les caractéristiques globales ont des propriétés tout à fait différentes. On peut espérer pouvoir exploiter leur complémentarité pour améliorer la classification.

Chapitre 3

L’émotion

1 **Introduction et définition de l'émotion**

Dans ce chapitre nous nous focalisons sur la définition de l'émotion, les facteurs provoquant un changement d'émotion, ainsi que les différentes techniques employées.

Nous pouvons tirer une définition structurée en se basant sur la notion d'état affectif. Cette notion regroupe tout ce que nous pouvons ressentir et qui est divisée en cinq catégories: les émotions, les humeurs, les positions interpersonnelles, les préférences/attitudes, et les dispositions affectives (Pantic, Sebe, Cohn, & Huang, 2005).

Selon cette définition, une émotion est caractérisée par plusieurs points :

- une forte intensité ;
- une émotion est une réaction de courte durée ;
- une forte synchronisation : tout le corps réagit ;
- une émotion est directement liée à un évènement déclencheur ;
- une émotion est cependant soumise au traitement cognitif de l'évènement déclencheur ;
- une émotion peut changer très rapidement ;
- son impact sur le comportement est important.

Darwin, dans sa théorie sur l'évolution (Darwin, 1872), souligne que l'émotion est une réponse à l'environnement apparue de la même façon que bien d'autres phénomènes suite à la sélection naturelle. Une émotion induit des réactions physiologiques et psychologiques qui nous permettent de mieux répondre à l'environnement.

Ekman et al. dans (Ekman, 1999), fervent partisan de la théorie darwiniste, a extrait six émotions basiques en se basant sur six critères :

- L'émotion doit être déclenchée par des stimuli universels, soient communs à tous les membres de l'espèce.
- Elle doit apparaître spontanément.
- La réaction apparaît et disparaît rapidement.
- Le traitement cognitif du stimulus doit être automatique.
- Elle doit déclencher des pensées ou sensations spécifiques.
- Elle doit être présente chez d'autres êtres vivants que chez l'être humain.

Les six émotions extraites selon Ekman sont la colère, la tristesse, la peur, la joie, la surprise et le dégoût. Contrairement aux émotions de premier ordre, les émotions basiques, dont nous pouvons voir les manifestations chez le nourrisson, les émotions de second ordre sont celles qui sont passées par le filtre de l'expérience et de l'apprentissage social. Les six émotions

basiques d'Ekman, largement acceptées par la communauté des psychologues, fournissent un premier ensemble discret d'émotions.

William James (James, 1984) perçoit les émotions comme une réponse directe à la perception d'un événement contribuant à la survie de l'individu et insiste sur les changements induits sur son comportement corporel. Le corps répond d'abord d'une manière plus ou moins préprogrammée puis l'expérience de ce changement constitue ce que nous appelons les émotions. La rétroaction du corps et, notamment, les organes innervés par le système nerveux contribuent grandement à l'expérience des émotions.

Selon notre cadre de recherche, Il existe de nombreuses autres définitions (Wikipedia, 2008). Nous considérons qu'une émotion est une séquence de changement d'états intervenant dans cinq systèmes organiques (cognitif, neurophysiologique, moteur, motivationnel, moniteur), de manière interdépendante et synchronisée en réponse à l'évaluation de la pertinence d'un stimulus externe ou interne par rapport à un intérêt central pour l'organisme.

Une émotion correspond donc au départ à une réaction corporelle face à un événement, par exemple :

- la colère : fait affluer le sang vers les mains, ce qui permet à l'individu de s'emparer plus prestement d'une arme ou de frapper un ennemi et engendre une sécrétion massive d'hormones comme l'adrénaline qui libère l'énergie nécessaire à une action vigoureuse ;
- la peur : dirige le sang vers les muscles qui commandent le mouvement du corps comme les muscles des jambes ce qui prépare la fuite en faisant pâlir le visage ;
- le dégoût : entraîne une fermeture des narines avec retroussement de la lèvre supérieure face à une odeur désagréable ou pour recracher un aliment toxique.

Une émotion est une réaction à un événement qui apparaît soudainement et qui dure peu longtemps.

Selon Damasio (Damasio, 1994), les émotions se subdivisent en deux catégories : les émotions primaires et les émotions secondaires.

1.1 Les émotions primaires :

Au début de notre vie et selon Damasio nous sommes préprogrammés pour répondre par une réaction émotionnelle de façon instinctive et automatique à certains traits de stimulus survenant à la fois dans le monde externe (environnemental) et dans le monde interne (corporel). Ces réactions automatiques sont inscrites dans notre système nerveux, car elles nous aident à survivre. D'un point de vue neuronal, nous pouvons décomposer la réaction en trois étapes :

1. perception du stimulus ;
2. l'amygdale déclenche l'instauration d'un état du corps caractéristique de l'émotion;
3. nous percevons l'émotion en rapport avec le phénomène qui l'a déclenchée.

1.2 Les émotions secondaires :

Les émotions secondaires, telles qu'elles sont suscitées par les sensations : le plaisir, la joie, la tristesse, la colère, la peur et le dégoût.

Chaque émotion va d'un pôle positif à un pôle négatif. Ainsi, la joie, émotion positive, correspond à la tristesse, émotion négative, la sérénité correspond à la culpabilité, l'amour à l'hostilité, la fierté à la honte, le calme à l'angoisse, la perception constructive de soi à la dévalorisation de soi, etc. Toute émotion négative a sa contrepartie positive (Tableau 3.1).

1. Tristesse……	➢ …….……Joie
2. Culpabilité……	➢ ……..…….Sérénité
3. Hostilité…….	➢ ……..……Amour
4. Honte……	➢ ……..……..Fierté
5. Angoisse…….	➢ …….……..……Calme
6. Dévalorisation……	➢ ……..….Saine vision de Soi
7. Découragement…….	➢ ……..…..Encouragement
8. Désespoir……	➢ ……...……Espoir

Tableau 3.1 : émotions positives et négatives.

2. Les expressions faciales

L'expression faciale est définie par un ensemble des signes du visage qui traduisent un sentiment, une Émotion. L'expression faciale est un changement dans le visage, perceptible visuellement, dû à l'activation (volontaire ou non) de l'un ou de plusieurs des 44 muscles composant le visage. Ainsi, les expressions faciales représentent l'un des éléments les plus importants dans le processus de communication. Les expressions faciales peuvent permettre aussi de distinguer entre certaines émotions spécifiques et procurer de l'information à propos de l'intensité des émotions ressenties (Tremblay, Deschênes, Poulin, Roy, Kirouac, & Kappas, 1993).

Pourquoi les expressions faciales existent elles? On peut citer trois points de vue différents :

1. Les émotions sont au centre de l'explication des expressions faciales :

- sélection naturelle : 7 émotions de base ; 7 expressions faciales universelles (colère, dégoût, joie, mépris, peur, surprise, tristesse) ;

2. Les expressions faciales sont des signaux de conversation qui dépendent des intentions de l'émetteur, du comportement du récepteur et du contexte de l'interaction :
   - hausser les sourcils ⇔ hésitation, froncer les sourcils ⇔ incompréhension, hausser le menton ⇔ ignorance.
3. Les expressions faciales sont des activateurs et des régulateurs d'émotion dus à la plasticité du cerveau :
   - changent la température de l'hypothalamus pour faciliter ou inhiber la production de neurotransmetteurs reliés aux émotions.

Il est important de faire la distinction entre la reconnaissance d'expression faciale et la reconnaissance d'émotions.

Les émotions dépendent de plusieurs facteurs et peuvent être identifiées par la voix, la posture, les gestes, la direction du regard et les expressions faciales, etc.

Les émotions ne sont pas la seule origine des expressions faciales. En effet, celles-ci peuvent provenir de l'état d'esprit, par exemple la réflexion, de l'activité physiologique, comme la douleur ou la fatigue, et de la communication non verbale comme le clin d'œil, froncement des sourcils.

Sept émotions de base correspondent à une expression faciale unique, quelles que soient la culture de la personne en question. La reconnaissance des expressions faciales consiste à identifier les déformations des structures faciales et les mouvements faciaux uniquement à partir d'informations visuelles. La reconnaissance des émotions, quant à elle, est une tentative d'interprétation qui requiert une information contextuelle plus complète.

3. L'émotion

3.1 Les différentes émotions universelles

Ekman et al. (Ekman, W.Friesen, & Ellsworth, 1972) ont identifié six émotions universelles soit la joie, la surprise, le dégoût, la colère, la peur et la tristesse (Figure 3.1).

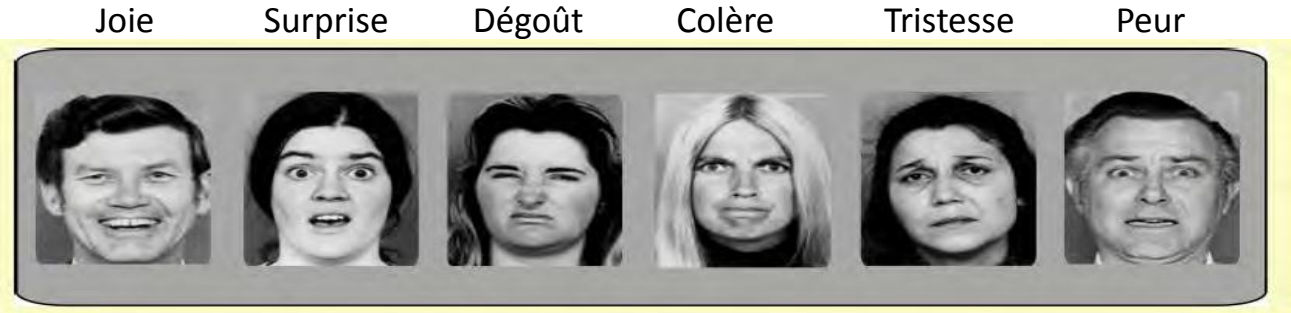

Figure 3.1 : Les six expressions faciales universelles (Ekman, W.Friesen, & Ellsworth, 1972)

3.2 Modèle théorique de l'émotion

Selon Scherer, les émotions sont les interfaces de l'organisme avec le monde extérieur et le processus émotionnel se décompose en trois principaux aspects (Scherer, Banse, Wallbot, t, & Goldbeck, 1991):

1. L'évaluation de la signification des stimuli par l'organisme (aspect cognitif);
2. La préparation aux niveaux physiologique et psychologique d'actions adaptées (aspect physiologique);
3. La communication par l'organisme des états et des intentions de l'individu à son environnement social (aspect expressif).

Ces trois aspects, cognitif, physiologique et expressif sont généralement acceptés comme constituants du phénomène émotionnel (Scherer, Banse, Wallbot, t, & Goldbeck, 1991).

Théorie physiologique :

William James[2] et Carl Lange[3] mettent l'accent sur le rôle essentiel des réactions émotionnelles dans le déclenchement de l'expérience émotionnelle. Cette théorie trouvera plus tard un prolongement avec l'hypothèse de la rétroaction faciale qui met en lumière la façon dont la prise de conscience des modifications corporelles et faciales intervient dans les modifications de l'état d'esprit de la personne.

D'autres chercheurs, jugent plutôt que les réactions émotionnelles résultent de l'activation de mécanismes sous-corticaux (Godefroid, 2008). Au même titre que le vécu psychologique.

[2] Psychologue américain 1884
3 Physiologiste danois 1885

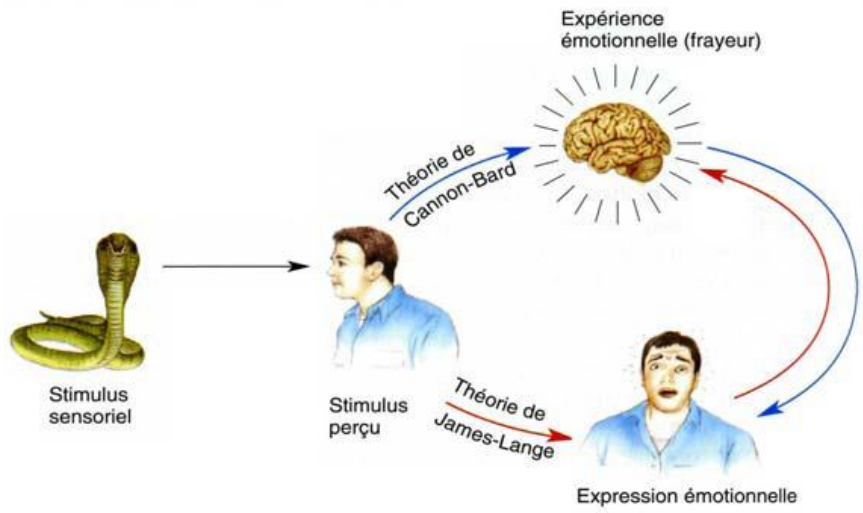

Figure 3.2 : Les théories physiologiques (Valat, 2008).

La figure ci-dessus (Figure 3.2) montre une comparaison schématique des théories des processus émotionnels de James-Lang et Cannon-Bard. Selon la théorie de James-Lang, la personne perçoit la présence de l'animal effrayant (Cobra), puis réagit. Ce comportement est déclenché en réponse à la perception de l'animal qui lui fait ressentir la frayeur.
Selon la théorie de Cannon-Bard, la peur résulte de la perception du stimulus et ensuite il s'en suit une réaction comportementale.

Théorie Néo-Darwinienne

La perspective évolutionniste tire son origine des travaux de Darwin. Elle étudie essentiellement la fonction communicative des émotions en donnant la prédominance aux expressions faciales.

Charles Darwin, en 1872, s'intéressait aux phénomènes émotionnels en publiant un ouvrage intitulé : l'expression des émotions chez l'homme et l'animal. Selon Darwin, les expressions émotionnelles de l'individu sont le reflet de la continuité de systèmes comportementaux complexes dérivés des autres espèces animales (Darwin, 1872). Il a eu recours a trois principes de base afin d'expliciter sa démarche :

1. Les habitudes associées : les expressions émotionnelles sont à l'origine des actes utilitaires qui rempliraient une fonction adaptative par rapport à l'environnement ;
2. L'antithèse : les états émotionnels sont souvent caractérisés par des manifestations motrices antagonistes ;
3. L'action directe sur le cerveau : effet de débordement et de dérivation de la force nerveuse engendrée par la stimulation.

Les théories néo-darwiniennes se sont essentiellement focalisées sur la détermination des émotions de base en étudiant les expressions faciales. Les diverses catégorisations des émotions de base proposées dans la littérature indiquent qu'il existe d'importantes divergences entre les auteurs. Ces diverses conceptions théoriques ont en commun de mettre l'accent sur la relation entre une configuration expressive faciale et une émotion spécifique.

3.3 Neurophysiologie des émotions

De nombreuses structures du cerveau participent à la physiologie des émotions. La figure 3.3 montre une conception de système comprenant un réseau de structures interconnectées qui contrôlent l'expression émotionnelle. Les principales structures comprennent le cortex cingulaire, l'hippocampe, l'amygdale et ses connexions étendues avec l'hypothalamus et le cortex, les corps mamillaires de l'hypothalamus et le cortex préfrontal. L'hypothalamus est le lieu où sont générés les comportements et le système limbique crée les émotions. Les régions préfrontales et sensorielles établissent des contacts avec le cortex cingulaire, l'hippocampe et l'amygdale. Les deux dernières structures établissent des connexions avec l'hypothalamus, qui, à son tour, établit des connexions avec le cortex cingulaire par le thalamus (Valat, 2008).

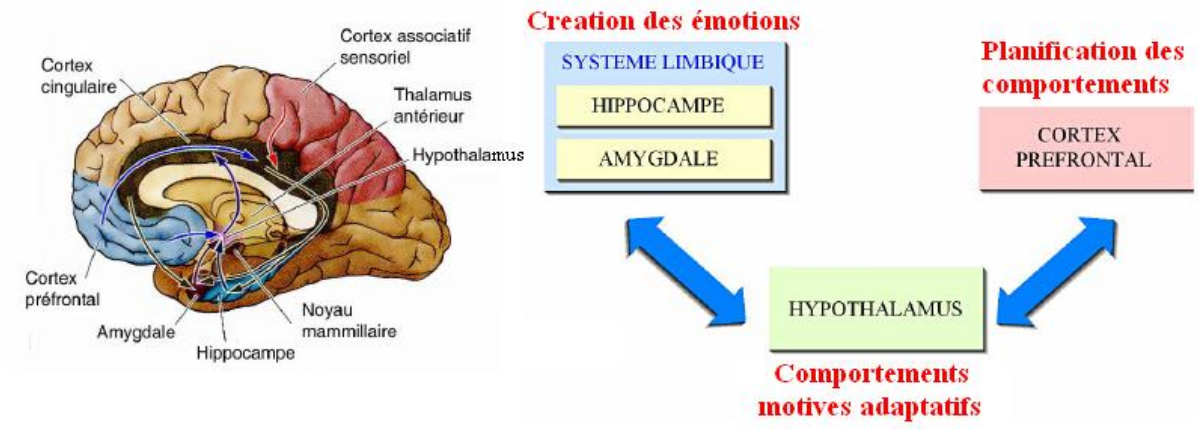

Figure 3.3. Représentation schématique des connexions principales du système limbique (Valat, 2008).

3.4 Représentation des émotions

La manipulation des émotions par ordinateur soulève de nombreuses problématiques. Dans un premier lieu, au niveau de leurs représentations, il s'agit de trouver un formalisme qui soit en

accord avec les résultats psychologiques existants, tout en permettant une manipulation simple. Dans un second lieu, pour un évènement donné, il faut pouvoir déterminer le potentiel émotionnel qui lui est associé. En se basant sur les travaux en psychologie, les états émotionnels sont considérés comme des catégories, ou un modèle multidimensionnel.

Approche catégorielle

C'est l'approche la plus répandue qui consiste à considérer les émotions comme des caractéristiques universelles (Ortony & Turner, 1990). Il suffit ensuite d'associer un mot du langage à ces caractéristiques. Le caractère universel des émotions entraine la définition d'un nombre d'émotions basiques (la peur, la colère, la joie, la tristesse, la surprise, etc.), qui ont pu être observées chez toutes les personnes, quel que soit leur culture (Tableau 3.2). Cette approche fait essentiellement la distinction entre ces émotions et propose de les classer sous forme de catégories discrètes. Ainsi les dénominations affectives qui ne trouvent pas leur place dans ces classifications sont considérées comme des mélanges d'émotions primaires. La justification principale de cette approche réside dans le fait que ces émotions basiques sont clairement identifiables chez la majorité des individus, notamment à travers la communication non verbale. Toutefois, leur nombre, le nom qu'il faut leur attribuer et leur caractérisation comme émotion basique, restent des questions ouvertes (Ortony & Turner, 1990). L'intérêt principal de l'approche catégorielle est qu'une fois que les émotions à traiter sont clairement identifiées, il devient simple de les manipuler, aussi bien pour les humains que pour les machines.

Approche	Emotions basiques
Ekman et al	Colère, dégoût, peur, joie, tristesse, surprise
Izard	Colère, mépris, dégoût, détresse, peur, culpabilité, intérêt, joie, honte, surprise
Plutchick	Acceptation, colère, anticipation, dégoût, peur, joie, tristesse, surprise
Tomkins	Colère, intérêt, mépris, dégoût, détresse, peur, joie, honte, surprise

Tableau 3.2 : Liste des émotions basiques selon différentes approches.

Approche dimensionnelle

L'approche dimensionnelle est une autre approche théorique très populaire en psychologie des émotions humaines (Russell, 1980) (Ekman, 1999) qui propose une représentation continue sur plusieurs axes ou dimensions par contraste aux catégories discrètes des émotions de base. (Tableau 3.3).

Trois facteurs ont été utilisés afin de mieux rendre compte des effets psychophysiologiques des différentes émotions : la valence, le degré d'activation physiologique (ou l'arousal) et la dominance (contrôle). En général, deux dimensions principales sont mises en avant (Figure 3.4).

D'une part, la valence émotionnelle, c'est-à-dire le caractère positif ou négatif de l'expérience émotionnelle et, d'autre part, la dimension de l'intensité ou le degré d'activation de l'expérience émotionnelle (l'arousal).

L'approche dimensionnelle permet de représenter facilement des émotions nuancées mais également des transitions entre différents états émotionnels.

Auteur	Axe choisi
Russel	Arousal\Valence
Cowie et al	Activation\Evaluation

Tableau 3.3: Axes choisis par quelques auteurs.

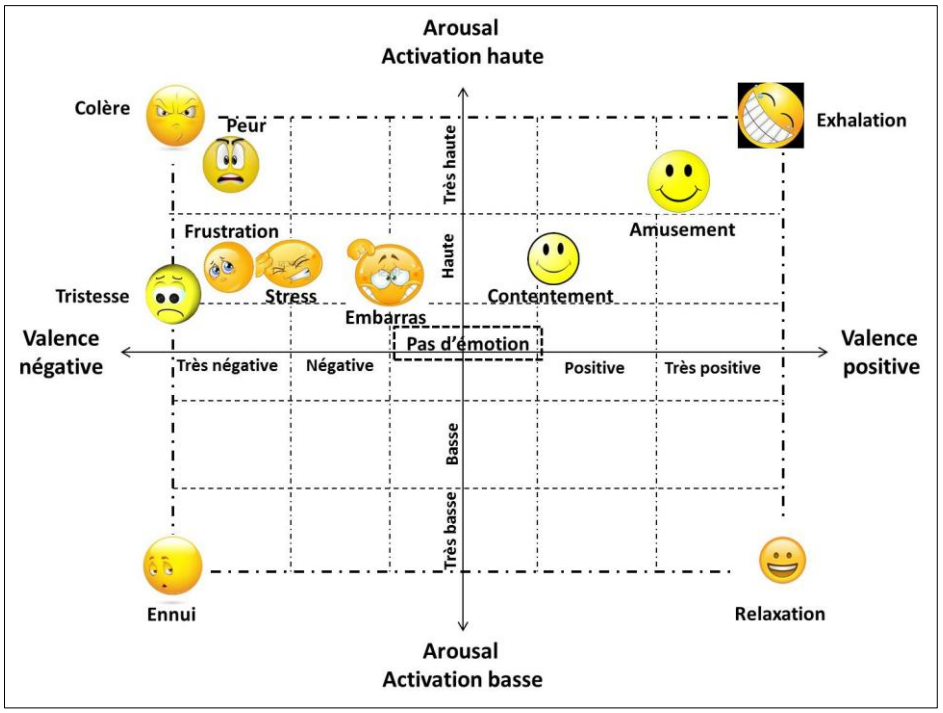

Figure 3.4: Représentation des émotions par rapport aux axes Arousal/Valence.

Plutchik en 1980 propose de placer les émotions primaires sur un cercle (Plutchik, 1980). Dans les encadrés de forme rectangulaire, on trouve les dyades primaires qui correspondent à des émotions secondaires. Elles résultent de la combinaison de deux émotions primaires représentées par des secteurs adjacents sur le cercle. Par exemple, la déception résulte de la tristesse et de la surprise (Figure 3.5).

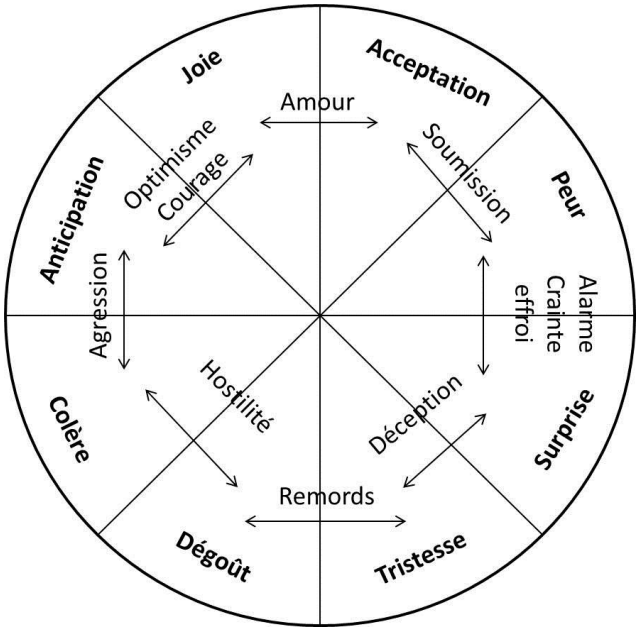

Figure 3.5 : les émotions mixtes.

4 Expression de l'émotion

4.1 Les canaux de l'expression

D'après Picard (Klein, Moon, & Picard, 2002), le corps humain est capable de transmettre de l'information par divers canaux; la communication émotionnelle est effectuée via ces différents canaux:

– les expressions faciales;

– les intonations de la voix;

– les réactions du Système Nerveux Autonome ;

– les mouvements du corps.

Les premières études effectuées ont montré que les expressions faciales constituent le moyen le plus évident pour reconnaître l'expression émotionnelle.
Lors de l'expérience d'une émotion par un individu, des muscles spécifiques sont activés permettant, par exemple, de sourire ou de froncer les sourcils.
Ces réactions ont pour but principal de communiquer l'émotion identifiée. Les signaux de la voix peuvent provenir à la fois d'une expression motrice (modulations, fréquence utilisée) et d'un processus neurophysiologique (par exemple, lorsqu'on a la gorge serrée par la tristesse). Les réactions du système nerveux autonome sont les réactions neurophysiologiques. Il s'agit, par exemple, d'une augmentation du rythme cardiaque, de la rétractation des vaisseaux pour mieux irriguer les muscles.
Enfin, les positions prises par la personne et les mouvements effectués sont également porteurs d'information émotionnelle. De l'extérieur, nous sommes capables de percevoir ces signaux chez un individu faisant l'expérience d'une émotion. Bien que la communication émotionnelle passe par chacun de ces canaux, ces derniers ne sont naturellement pas exclusivement dédiés à la communication émotionnelle : par exemple, parler et articuler déforme notre visage sans que ce soit à vocation de communication émotionnelle. Cela pose la difficulté de percevoir quelles caractéristiques expressives sont relatives à l'émotion. De plus, nous sommes capables de cacher nos émotions ou de contrefaire des émotions, c'est à dire d'exprimer des émotions que nous ne ressentons pas forcément. Morris a établi dans (Morris, Friedhoff, & Dubois, 1978) une hiérarchie de confiance dans ces différents canaux de communication émotionnelle, du plus crédible au moins crédible :
1. signaux d'expression physiologique;

2. positions et gestuelle;

3. expressions faciales;

4. expressions verbales.

Les signaux physiologiques sont très difficiles voire impossibles à contrefaire, bien que nous puissions être conscients de tels signaux. Par exemple, un individu rougissant de honte aura conscience de son rougissement mais ne pourra l'arrêter volontairement. De même, le rythme cardiaque est quasiment impossible à contrôler. Certaines intonations de la voix peuvent se situer dans cette catégorie, la réaction émotionnelle mettant en jeu les muscles utilisés pour parler ainsi que les cordes vocales. Ainsi, il peut être très difficile de cacher les effets d'une profonde tristesse dans notre voix.

Le canal de communication de la position et de la gestuelle peut encore être subdivisé selon le critère de crédibilité. Les mains sont les indicateurs les moins crédibles, ensuite le buste, et enfin les jambes et les pieds. En effet, ces derniers sont loin du centre d'attention lors d'une communication émotionnelle (le visage). Nous sommes donc moins habitués à les contrôler lorsqu'il s'agit d'exprimer une émotion non ressentie. Le buste reflète le tonus musculaire du corps ; il est difficile de le maintenir dans un état contraire à ce qui est ressenti (par exemple se tenir droit lorsqu'on est ennuyé). Enfin, les mains croisent sans arrêt notre regard lorsque nous communiquons : nous sommes donc plus à même de penser à les contrôler.

Enfin les expressions faciales et verbales sont des indicateurs peu dignes de confiance. Les expressions faciales étant le principal canal de communication émotionnelle, nous apprenons rapidement à les contrôler. L'expression verbale (ce que nous disons) est tout à fait contrôlable. Cette expression verbale est à différencier de l'expression non verbale de la voix : intonations, fréquences, etc., qui peuvent relever de processus neurophysiologiques.

4.2 Les variables captées pour reconnaître les émotions

La reconnaissance d'émotions consiste à se baser sur une perception des signaux émotionnels émis par un individu pour en inférer son expérience émotionnelle. Dans la communication humaine, nous le faisons chaque jour, de façon non forcément consciente (Ambady & Rosenthal, 1992). La recherche s'applique donc à identifier les caractéristiques pertinentes de l'émotion. Les variables non-verbales sont classées en trois catégories : les variables motrices, les variables distales et les variables proximales (Wallbott, 1998). Les variables motrices sont les mesures de l'activité corporelle d'un sujet. Par exemple, la tension de certains muscles, l'orientation des différents segments du corps humain, ou les battements du cœur. Les variables motrices peuvent être difficiles à capturer, et souvent par des dispositifs intrusifs qui

limitent leur utilisation aux expériences en laboratoire. Les variables distales sont la mesure des signaux produits par le corps et perçus par un observateur avant toute évaluation cognitive de la part de cet observateur. On peut, par exemple, enregistrer le son de la voix grâce à un microphone, ou enregistrer des mouvements grâce à une caméra vidéo. Enfin, les variables proximales sont liées à l'interprétation d'un observateur, et sont relevées, par exemple, grâce à des entretiens ou des questionnaires. Lors de la récolte de variables non verbales dans le cadre d'une expérimentation, il est possible de croiser des variables motrices, mesurant l'activité corporelle et physiologique, des variables distales, mesurant le sujet de l'extérieur mais de façon objective, et des variables proximales, passant par le filtre cognitif d'un évaluateur.

Il existe sur chacun des canaux d'expression de l'émotion une "bande passante émotionnelle", sur laquelle sont transmises les informations relatives à l'émotion que nous ressentons ou à l'émotion que nous souhaitons exprimer. Lorsque nous voulons cacher une émotion ressentie, la synchronisation émotionnelle lutte contre le contrôle que nous avons de nos expressions faciales, de nos mouvements, etc. Lorsque nous voulons simuler une émotion non ressentie, nous ne pouvons atteindre par le contrôle conscient de notre corps la même synchronisation que celle apportée par une émotion réellement ressentie. Ces deux cas de figure induisent une dissonance entre les différents signaux émotionnels des différents canaux.

5. Systèmes de reconnaissance de l'émotion

L'informatique affective (en anglais : Affective computing) est l'étude et le développement de systèmes et d'appareils ayant les capacités de reconnaître, d'exprimer, de synthétiser et modéliser les émotions humaines. C'est un domaine de recherche interdisciplinaire couvrant les domaines de l'informatique, de la psychologie et des sciences cognitives qui consiste à étudier l'interaction entre technologie et sentiments (Wikipédia).

Picard a divisé l'informatique affective en deux catégories : la synthèse et la reconnaissance d'émotions.

La synthèse de l'émotion consiste à générer des émotions virtuelles chez une machine à but de communication avec l'utilisateur. De nombreuses études ont été menées sur les avatars expressifs (Gratch & Marsella, 2001) (Karpouzis, Raouzaiou, & Kollias, 2003). Il s'agit d'un ensemble de personnages capables d'exprimer des émotions selon les canaux de communication affective et ceci dans le but d'améliorer l'interaction homme-machine.

La reconnaissance d'émotions permet de découvrir l'émotion exprimée ou ressentie chez un être humain afin de le prendre en compte et de l'analyser. Cette prise en compte peut être au cœur du système ou à des fins d'améliorations de l'interaction. Picard dans (Vyzas & Picard, 1999) a proposé un genre d'algorithme pour effectuer une reconnaissance affective :

1. Acquérir le signal en entrée et ceci par la mise en place des dispositifs d'acquisition ou de capture des données (microphone pour capturer la voix et une caméra pour le visage).
2. Reconnaître des formes dans le signal et ceci par l'extraction des caractéristiques pertinentes d'une émotion ou d'un état affectif.
3. Etre capable d'analyser les caractéristiques extraites de l'étape précédente afin d'identifier l'émotion plus probablement exprimée par l'utilisateur.
4. Entraîner la machine à reconnaître et classer l'émotion.
5. Délivrer l'émotion finalement interprétée.

La classification de l'émotion en catégories est une méthode d'interprétation trouvée fréquemment dans la littérature. Chaque émotion est considérée comme une catégorie, ainsi tout système de classification est capable d'identifier l'émotion correspondante à un vecteur de valeurs de caractéristiques. La plupart des systèmes de reconnaissance sont basés sur le modèle discret et en particulier sur les six émotions de base d'Ekman.

5.1 Les systèmes de reconnaissance émotionnels existants

La plupart des systèmes de reconnaissance émotionnels existants dans la littérature ont été classés selon deux paires : reconnaissance générique/personnalisée et reconnaissance active/passive.

Reconnaissance générique/personnalisée :

La reconnaissance de l'émotion peut être générique ou personnalisée. Pour s'adapter au plus grand nombre d'utilisateurs, un système doit être générique et pour obtenir de meilleurs résultats, il est préférable qu'il soit réglé en fonction de l'application et/ou adapté à un utilisateur en particulier. De nombreux travaux sont basés sur l'identification et la validation des caractéristiques émotionnelles dans la voix (Taskeed, Hasanul, & Oksam, 2010), le visage (Vyzas & Picard, 1999), les mouvements (Cowie, et al., 2001) (Picard R. W., 1999.) et les signaux physiologiques. Toutes ces études proposent des caractéristiques génériques.

Reconnaissance active / passive :

Une modalité est dite active si elle requiert une action explicite de l'utilisateur pour communiquer ou percevoir les données (taper une requête/commande, regarder un écran mobile) ; elle est dite passive si l'interaction ne requiert pas l'attention de l'utilisateur. Nous étendons donc ces notions à celles de reconnaissance active et passive d'émotions.

La reconnaissance active consiste à une action initiale par le système qui demande à l'utilisateur son état affectif courant. Ce dernier répond à l'interaction proposée par le système. Dans (Valat, 2008), les auteurs proposent un système où l'utilisateur indique, à tout instant et à l'aide d'un curseur, son degré de frustration. Ce mécanisme est réutilisé par (Klein, Moon, & Picard, 2002) pour permettre à l'utilisateur de se remettre d'états émotionnels négatifs.

Quant à la reconnaissance passive, elle n'implique pas l'utilisateur. Elle se base sur des systèmes d'observation de l'utilisateur à l'aide de divers dispositifs de capture (caméra, microphone, etc.). Les données sont traitées pour en extraire des caractéristiques utiles à l'identification des émotions. La majorité des systèmes existants se basent sur une reconnaissance passive.

5.2 Quelques capteurs utilisés pour la reconnaissance d'émotions

Généralement, la machine peut être équipée d'un ensemble de capteurs permettant de mesurer les comportements de l'utilisateur. Le dispositif de capture le plus utilisé pour la reconnaissance d'émotions est la caméra vidéo. Sachant qu'il existe un grand nombre d'outils de traitement d'image permettant le suivi d'un visage, la détection des points d'intérêt, et un suivi du corps dans l'espace. La caméra vidéo est ainsi parfaitement adaptée à la reconnaissance d'expressions faciales.

Il existe de nombreux autres capteurs tels que les capteurs du mouvement pour enregistrer les mouvements d'un individu, capteurs de pression sanguine, électrocardiogrammes, électro-encéphalogrammes, électromyogrammes, capteurs d'expansion de la cage thoracique pour mesurer la respiration... etc.

La « sentic mouse » (Kirsch, 1997) et la « pressure chair » (Clavel, 2007) sont deux exemples reconnus comme des dispositifs pour la reconnaissance d'émotions. La « sentic mouse » est une souris équipée d'un capteur de pression directionnelle sur le bouton gauche. La direction du clic donne une indication sur la valence de l'état émotionnel de l'utilisateur : un clic où le bouton est attiré à soi tend à indiquer une valence positive, tandis qu'un clic où le bouton est repoussé indique une valence négative. La pressure chair est un fauteuil de bureau

où sont placés des capteurs de pression dans l'assise et le dossier. Ces capteurs permettent de déterminer la position de l'utilisateur lorsqu'il est assis et, par conséquent, il est possible de détecter son état affectif.

Picard et al (Picard & Rosalind, 2000) (Picard R., 2001) décrivent plusieurs dispositifs portés pour la capture d'émotions. Les auteurs présentent une chaussure munie d'un capteur mesurant la conductivité de la peau, une boucle d'oreille munie d'un capteur de pression sanguine et une paire de lunettes munie de capteur de tension musculaire afin de détecter les froncements des sourcils. Ces systèmes présentent l'avantage d'accompagner l'individu et par conséquent ils permettent de relever des expressions propres de l'émotion.

5.3 Canaux de communication émotionnelle

Reconnaissance par expression faciale

Le visage constitue la première source de l'émotion. En effet, le visage est le canal communiquant le plus d'information émotionnelle (Ekman).

Paul Ekman a proposé un système d'encodage s'appuyant sur l'anatomie du visage FACS basé sur la définition d'unités d'action causant des mouvements sur le visage. Une unité d'action représente un point d'intérêt pour la dynamique du visage. Ces unités d'actions servent de base à de nombreux systèmes de reconnaissance d'émotions par le visage. La précision et la robustesse représentent les deux critères d'évaluation des systèmes qui effectuent une reconnaissance sur des modèles discrets d'émotions (entre trois et sept catégories d'émotions). Pantic et al (Pantic, Sebe, Cohn, & Huang, 2005) présentent des résultats ente 64% et 98% de reconnaissance selon les systèmes. Sachant qu'en absence d'un protocole expérimental, il est impossible d'approuver les pourcentages cités ci-dessus.

Reconnaissance par la voix :

Plusieurs travaux dans la littérature dont nous citons (Scherer, 2004) ont montré que la voix communique notre état émotionnel selon plusieurs caractéristiques. Elle permet de mieux communiquer l'activation d'une émotion que sa valence. Les caractéristiques de son pour la reconnaissance d'émotions telles que la vitesse du discours, la hauteur moyenne, l'intensité, la qualité de la voix, etc. sont présentées dans (L'Empire caché de nos Emotions, 2005).

Reconnaissance par le mouvement

Caractérisées par le mouvement des bras et des mains. Les trois caractéristiques les plus générales du mouvement sont : l'activité, l'expansion dans l'espace et l'énergie du mouvement.

Reconnaissance par les signaux physiologiques

Plusieurs types de signaux peuvent être utilisés pour reconnaître l'émotion. Le rythme cardiaque, la conductivité de la peau, l'activité musculaire, les variations de température de la peau, les variations de pression sanguine sont des signaux régulièrement utilisés pour la reconnaissance d'émotions.

Chanel et al. (Chanel, Kronegg, Grandjean, & Pun, 2006) se basent sur un électroencéphalogramme pour évaluer l'activation de quatre individus. Ils mesurent également la conductivité de la peau, la pression sanguine, les mouvements abdominaux et thoraciques dus à la respiration, et la température de la peau. L'électroencéphalogramme seul produit des performances de reconnaissance supérieures à 50%. L'intégration des autres signaux rend la reconnaissance plus robuste. Kim et al. (Kim, Bang, & Kim, 2006) utilisent un électrocardiogramme pour mesurer les variations de température de la peau et l'activité électrodermale pour reconnaître la tristesse, la colère, le stress et la surprise. Une étude sur cinquante individus donne des résultats de 61,8% dans la reconnaissance des quatre émotions citées ci-dessus. Vyzas et Picard (Vyzas & Picard, 1999) utilisent l'activité musculaire de la mâchoire, la pression sanguine et la conductivité de la peau des doigts, et mesurent l'expansion thoracique pour la respiration chez une actrice exprimant huit états affectifs : le neutre, la colère, la haine, le chagrin, l'amour platonique, l'amour romantique, la joie et le respect. Le système exposé offre un taux de reconnaissance de 81,5% pour ces huit émotions.

Conclusion :

Dans ce chapitre, nous avons donné une définition détaillée de l’émotion, nous avons montré également les liens entre l’expression faciale et l’émotion, et aussi l’impact de l’émotion sur le changement des caractéristiques du visage. Les chercheurs considèrent le visage comme étant le canal communiquant quiposséde le plus d’informations émotionnelles. Dans ce contexte nous pensons à développer un système de détection des émotions basé sur les caractéristiques faciales.

Chapitre 4

Présentation du système

Introduction

Dans ce chapitre nous présentons en détails notre système qui sert à détecter les visages et reconnaitre les personnes ainsi que leurs émotions. En fait, on a conçu deux sous-systèmes avec une méthode de détection commune et deux méthodes de reconnaissance différentes. Le premier sous-système combine les moments de Zernike avec les LBP étendus, et utilise une technique floue pour la sélection des caractéristiques pertinentes, la classification des visages et des émotions se fait par la méthode du plus proche voisin. Le deuxième sous-système combine, à son tour, les moments de Zernike avec les LBP étendus multi échelles. Ensuite une phase de réduction des paramètres est appliquée, celle-ci est basée sur la technique des mémoires auto-associatives dite aussi, l'analyse en composantes principales non-linéaire. La classification dans ce cas est assurée par un perceptron multicouches.

Les deux sous-systèmes comportent les étapes suivantes :

1. Une phase de prétraitement.
2. Une localisation des visages et élimination de l'arrière-plan.
3. Extraction des caractéristiques faciales (yeux, bouche).
4. Paramétrage global, appliqué sur tout le visage, et local appliqué sur la région des yeux et de celle la bouche.
5. Reconnaissance des visages suivie par une identification instantanée de l'émotion.

Les expérimentations sont élaborées sur deux bases de visages, notamment, la base JAFFE et la base FABO dont de plus amples définitions sont explorées dans le chapitre subséquent.

1. Présentation du premier système :

Ce système est appelé **FEAST** (Face and Emotion Analysis System for Smart Tablets), il sert à analyser les visages et les émotions pour la gestion des tablettes intelligentes. Figure 4.1 nous donne la structure globale du système.

Dans le contexte de tablettes intelligentes, le système FEAST doit prendre en compte les différents aspects de l'image, c'est à dire la position, les conditions d'éclairage,… etc. la caméra intégrée dans la tablette intelligente doit nécessairement capter le visage de l'utilisateur à n'importe quelle position ou condition d'éclairage.

Le but de notre système est de satisfaire les préférences de l'utilisateur en termes de programmes et applications installés dans la tablette. Cette dernière est dotée d'un module qui

permet d'enregistrer les favoris de chaque utilisateur dans son profil. Le profil se compose de deux parties: les données personnelles (nom, âge,…etc.) et les préférences (programmes et jeux préférés, fond, musique,… etc.). Chaque nouvel utilisateur, qui est généralement le propriétaire, un membre de la famille ou un ami, doit passer par cette phase d'authentification afin de pouvoir accéder au système. Autrement, la tablette est hors service. Dès que la tablette intelligente est activée, le système tente d'identifier la personne. Une fois reconnue, la tablette intelligente exécute le profil par défaut sélectionnée par l'utilisateur. Ensuite, le système de reconnaissance des émotions contrôle l'état émotionnel de l'utilisateur (une émotion toutes les 2 secondes). Par exemple, si l'utilisateur est dégoûté ou triste; le système lui propose une autre application (musique, jeux,…etc.) qui répond à ses besoins. Le nouveau choix sera basé sur les préférences de l'utilisateur stockées dans son profil, qui peut également être mis à jour à la demande des utilisateurs. Figure 4.2 explicite l'architecture du système.

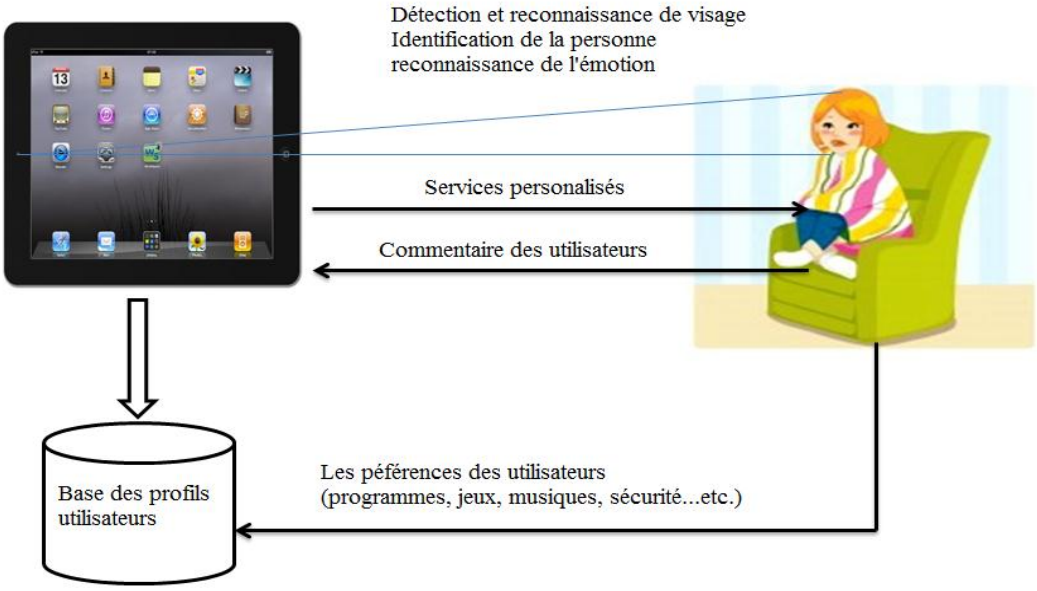

Figure 4.1. Structure générale du système.

2. Module de prétraitement:

Tout système de reconnaissance des visages ou des caractéristiques faciales doit passer par une étape de filtrage. L'étude des cas réels doit tenir compte des conditions environnementales et des circonstances d'acquisition des vidéos ou des images, notamment, la

luminosité, la position de la tête, la couleur de l'arrière-plan, le port de lunettes, la barbe, les moustaches...etc.

L'étape de prétraitement contribue essentiellement par l'élimination des effets secondaires, et facilite aussi la tâche au module de détection des visages et des caractéristiques faciales. En fait dans notre cas, le module de filtrage est consulté à un stade précoce pour améliorer la qualité de l'image et l'adapter à un traitement spécifique. Toutes les autres phases de traitement sont connectées en permanence avec ce module. Voir Figure 19. Les quatre opérations de base appliquées dans la première étape sont les suivantes :

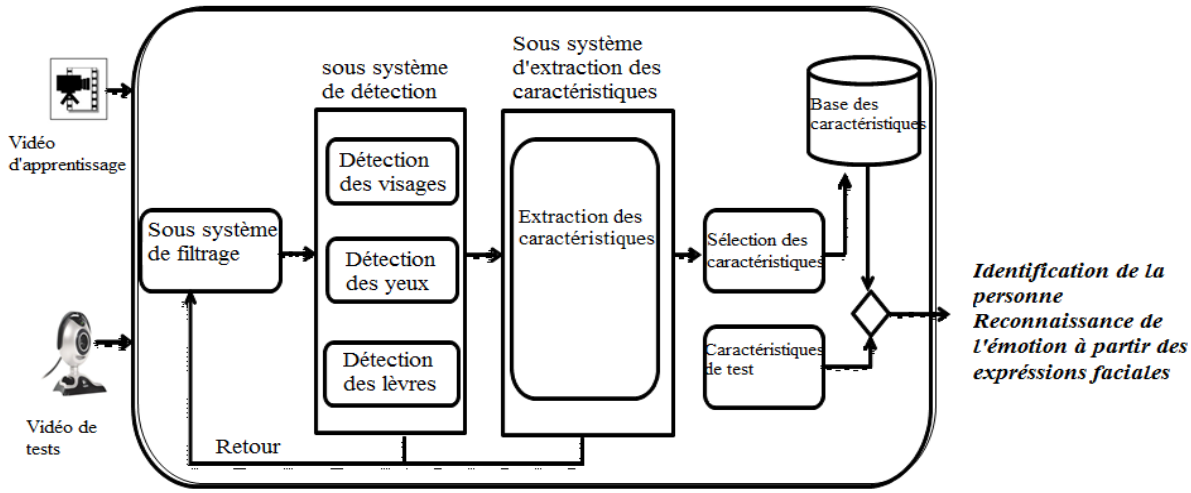

Figure 4.2. Architecture du système.

2.1 Redimensionnement de l'image

Les images traitées peuvent être représentées de n'importe quelle taille, tout dépend de la résolution choisie durant la phase d'acquisition. Le problème se pose surtout dans le cas où nous traitons des images à hautes résolutions, par exemple, avec une image de résolution de un méga octet, il faut parcourir un million de pixel pour en extraire les caractéristiques ! Ce qui est trop lent même pour une machine très puissante. Pour y remédier, nous procédons par un redimensionnement de l'image, sa taille est réduite au préalable afin de permettre un traitement postérieur accéléré tout en évitant la perte de ses qualités. Dans notre cas, chaque image est réduite à 300 × 300 pixels.

2.2 Correction de l'angle d'inclinaison

Cette étape de filtrage vient juste avant la localisation de visage. La détection des yeux nécessite qu'ils soient orientés horizontalement, pour ce faire une correction de l'angle d'inclinaison est promulguée, Figure 4.3.

Dans notre cas la détection des visages se fait par le biais des moments géométriques, nous les verrons en détails dans le module correspondant. Parmi les informations fournies par ces moments, on a l'angle d'inclinaison, θ, de l'ensemble des composants connexes trouvé. Selon l'angle θ, l'ajustement de l'inclinaison se fait au niveau de chaque pixel comme suit :

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix} \times \begin{pmatrix} x \\ y \end{pmatrix} \qquad (80)$$

Où : x, y : sont les coordonnées originales des pixels, x', y' : sont les nouvelles coordonnées obtenues.

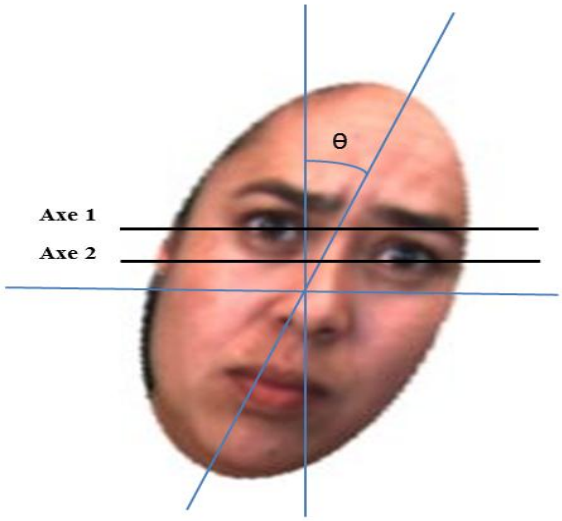

Figure 4.3. Angle d'orientation et alignement des yeux

2.3 Elimination de la luminance

Le système de détection des visages peut échouer dans certains cas, surtout lorsque les images présentent une forte luminosité. De nombreuses approches utilisent les techniques d'égalisation d'histogramme pour remédier au problème. Pour pallier cet inconvénient, nous appliquons la méthode proposée par Yu-Tzu et al. (Yu-Tzu, Ruei-Yan, L.Yu-Chih, 2012) avec une légère amélioration.

Il est vrai que les images en couleurs fournissent des informations plus riches que celles qui sont en niveau de gris, néanmoins, dans notre cas nous devons opter pour les méthodes qui sont à la fois efficaces et rapides. Pour cela nous traitons les images avec la taille la plus réduite que possible. Donc toutes les images sont transformées en niveau de gris par la formule suivante :

$$I(x,y) = 0{,}2126 \times R(x,y) + 0{,}7152 \times G(x,y) + 0{,}0722 \times B(x,y) \quad (4.1)$$

Où : $I(x,y)$, est l'intensité du pixel (x,y) de l'image en niveau de gris, $R(x,y), G(x,y), B(x,y)$, représentent respectivement, les composant R, G, B du pixel (x,y). Une fois l'image transformée en niveau de gris, une élimination de l'effet de luminance est entamée, cela se fait par la transformation de l'intensité $I(x,y)$ en $I'(x,y)$ par la formule suivante : $I'(x,y) = \begin{cases} I(x,y), & si\ I(x,y) \leq M \\ M, & si\ I(x,y) > M \end{cases}$ (4.2)

Où : $I'(x,y)$, est l'image résultante, M, est la valeur médiane calculée en niveau de chaque pixel appartenant à l'ensemble des pixels contenant la couleur de la peau, on les appelle aussi les skins pixels.

2.4 Renforcement de contraste

Pour assurer une détection précise des yeux, nous appliquons des opérations morphologiques telles que l'ouverture et l'érosion. La morphologie mathématique est basée sur la théorie des ensembles. Les opérations morphologiques appliquées sur une image binaire sont dites morphologie binaire dont l'image est représentée comme un ensemble $X \subseteq R^2$, où R est l'ensemble des nombres réels. Les pixels de l'image du premier-plan appartiennent à X, et ceux de l'arrière-plan appartiennent au complément de X, noté par X^c. La transformation de l'image se fait par un autre ensemble appelé l'élément structurant, dont sa forme et sa taille déterminent l'image résultante (Vincent, 1993).

La dilatation, l'érosion et l'ouverture pour un ensemble binaire se font respectivement par les opérations suivantes :

$$X \oplus H = \left\{(x,y): H_{(x,y)} \bigcap X \neq \Phi \right\} \quad (4.3)$$

$$X \ominus H = \left\{(x,y): H_{(x,y)} \subseteq X\right\} \quad (4.4)$$

$$X \circ H = (X \ominus H) \oplus H\} \quad (4.5)$$

Où : X, est l'image originale, $H \subseteq R^2$, est l'élément structurant, $H_{(x,y)}$, est la translation de l'ensemble H, par le vecteur $(x, y) \in R^2$.

Les opérations morphologiques appliquées sur une image au niveau de gris, qu'on a utilisé dans notre système, sont une extension des opérations morphologiques binaires, dans ce cas l'image est représentée par la fonction $f(x, y)$, avec $(x, y) \in R^2$, l'élément structurant est défini par la fonction $h(x, y)$ ou simplement par h. Les opérations de dilatation, d'érosion et d'ouverture correspondantes sont définies respectivement par :

$$(f \oplus h)(x, y) = sup_{(r,s)\in H} \{f(x - r, y - s) + h(r, s)\} \quad (4.6)$$

$$(f \ominus h)(x, y) = inf_{(r,s)\in H} \{f(x + r, y + s) + h(r, s)\} \quad (4.7)$$

$$f \circ H = (f \ominus H) \oplus H \quad (4.8)$$

Où : sup{} et inf{} sont les opérateurs supremum et infimum, et $H \subseteq R^2$.

3. Module détection de visage :

La détection des visages est basée sur les informations de la couleur et de la forme. Premièrement, toutes les images sont transformées depuis l'espace de couleurs RGB vers l'espace HSV (Hue, Saturation, Value), ensuite les pixels qui ressemblent à la couleur de la peau sont sélectionnés selon la formule de Pitas et al (Sobottka & Pitas, 1996) ci-dessous :

$$\begin{cases} 0^\circ \leq H \leq 25^\circ \ et \ 335^\circ \leq H \leq 360^\circ \\ 0{,}2 \leq S \leq 0{,}6 \ et \ V \geq 0{,}4 \end{cases} \quad (4.9)$$

Généralement, le visage humain est d'une forme elliptique. Une fois labélisés, les pixels qui ressemblent à la couleur de la peau sont regroupés en des ensembles de composants connexes, ensuite, sur chaque ensemble nous appliquons la méthode basée sur les moments géométriques pour chercher l'ellipse de meilleur ajustement. Cette méthode est utilisée dans (Haddadnia, Ahmadi, & Faez, 2003).

Pour trouver la région faciale, nous utilisons un modèle d'ellipse avec cinq paramètres, à savoir, x_0, y_0, qui représentent les coordonnées du centre de l'ellipse, l'angle d'orientation θ, α et β, qui indiquent respectivement, l'axe majeur et l'axe mineur. Voir Figure 4.4.

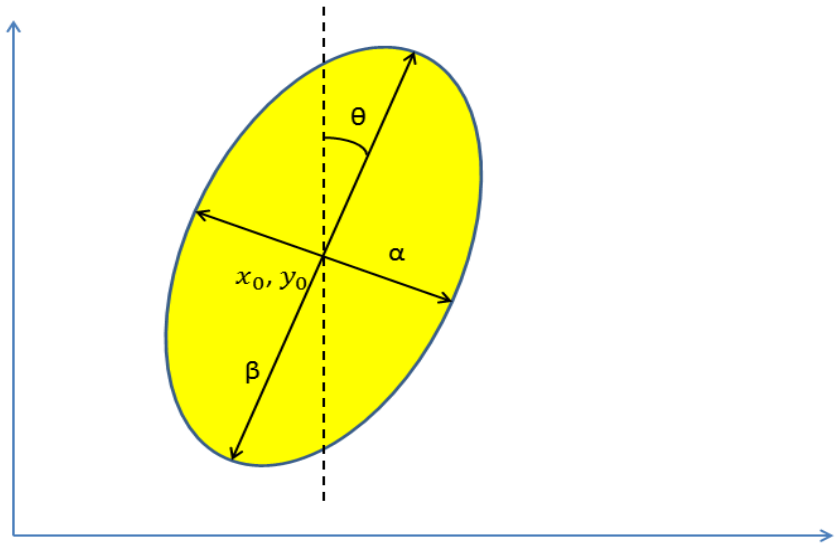

Figure 4.4: modèle elliptique de visage

Pour calculer les cinq paramètres de l'ellipse, les moments géométriques, d'ordre $p+q$ de l'image $f(x,y)$, sont définis comme suit :

$$M_{pq} = \sum_x \sum_y f(x,y)\, x^p y^q \qquad (4.10)$$

Où : $p, q = 0,1,2,\ldots$ et $f(x,y)$ est le niveau de gris de l'emplacement (x,y) dans l'image.

Les moments centraux invariants à la translation sont calculés en mettant les origines au centre de l'image, c'est-à-dire :

$$\mu_{p,q} = \sum_x \sum_y f(x-x_0, y-y_0)\,(x-x_0)^p (y-y_0)^q \qquad (4.11)$$

Avec : $x_0 = \frac{M_{10}}{M_{00}}$ (4.12) et $y_0 = \frac{M_{01}}{M_{00}}$ (4.13) sont les centres des composants connexes.

L'angle d'orientation θ est donné par :

$$\theta = \frac{1}{2} \times \arctan(2\mu_{11}/(\mu_{20} - \mu_{02})) \qquad (4.14)$$

Où : μ_{pq}, est le moment central défini dans la formule (4.11)

Les longueurs respectives de l'axe mineur et majeur sont calculées en fonction des moments d'inertie, ces derniers sont définis par les deux formules correspondantes :

$$I_{Min} = \sum_x \sum_y ((x - x_0)\, cos\theta - (y - y_0) \sin\theta)^2 \qquad (4.15)$$

$$I_{Max} = \sum_x \sum_y ((x - x_0)\, sin\theta - (y - y_0) \cos\theta)^2 \qquad (4.16)$$

Et finalement les longueurs des axes, mineur et majeur, sont calculées comme suit :

$$\alpha = \frac{1}{\pi}\, (I_{Max}^3 / I_{Min})^{1/8} \qquad (4.17)$$

$$\beta = \frac{1}{\pi}\, (I_{Min}^3 / I_{Max})^{1/8} \qquad (4.18)$$

Une fois les cinq paramètres sont déterminés pour chaque ensemble des composants connexes, une opération de recherche des yeux et de la bouche est lancée. Ce processus permet, d'une part, de définir les caractéristiques faciales du visage, et d'autre part de confirmer si l'ensemble de composants connexes définit bien un visage humain.

4. Détection des yeux et des lèvres

Après la détection du visage et la normalisation de l'angle d'inclinaison, le processus de recherche des caractéristiques faciales est déclenché. Ce dernier est basé sur la méthode d'analyse min-max (Sobotka & Pitas, 1996). Dans ce stade, la normalisation de l'angle d'inclinaison est nécessaire, car les yeux et la bouche sont orientés horizontalement.

L'analyse min-max commence par une détermination des y-projections, à cet effet, la moyenne des niveaux de gris est calculée au niveau de chaque ligne, et le y-relief résultant subit un amincissement par le moyen d'un filtre moyenneur de largeur 3. Ensuite, les minimas et les maximas sont extraits. Puis pour chaque minima significatif, un x-relief est calculé ; c'est une moyenne des niveaux de gris calculée sur trois lignes au voisinage de chaque colonne. Les minimas et maximas de chaque x-relief sont calculés.

La position des yeux est définie par la recherche du deuxième plus haut minima dans le y-relief, ensuite une recherche des minimas dans le x-relief est lancée afin de trouver les deux minima qui répondent aux exigences relatives à la position des yeux notés ci-dessous :

1. Les yeux sont localisés dans la partie supérieure du visage.

2. Deux minimas significatifs qui correspondent à l'œil gauche et droit doivent être trouvés.
3. Les minimas doivent avoir le même niveau de gris.
4. Entre les deux minimas des yeux, il doit y avoir un maxima significatif.
5. Le rapport de la distance entre les deux yeux et la largeur du visage doit respecter une certaine fourchette.

Enfin, la détection des lèvres est effectuée selon un modèle géométrique de visage comme dans (Frank & Chao-Fa, 2004). En fait, nous pourrions réaliser cela en utilisant les profils de projections. Mais pour éviter de nombreux inconvénients; causé, notamment, par la présence de la moustache, barbe et parfois par l'ombre, nous préférons la méthode de modèle géométrique qui est rapide et efficace. En se référant à la figure 22, le modèle de visage géométrique est décrit comme suit:

1. La distance horizontale entre les deux yeux est D.
2. La distance verticale entre la ligne passant par les yeux et la ligne passant par la bouche est D.
3. La hauteur de la bouche est de 0,4 D.
4. La longueur de l'axe mineur est 2D.
5. La longueur de l'axe majeur est de 1,5 × axe mineur.

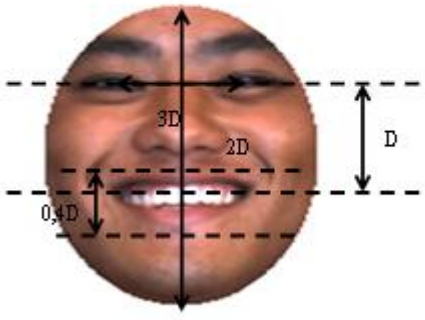

Figure 4.5. Modèle géométrique de visage.

5 Réglage de la détection des visages

Selon le modèle de visage, l'ellipse détectée est ajustée en effectuant des corrections dans la taille des axes majeur et mineur afin de lui rendre le plus proche que possible à la forme réelle du visage. Pendant la segmentation, les problèmes les plus fréquemment rencontrés sont dus à une luminosité élevée ou à un faible contraste. Figure 23 nous montre les différents cas rencontrés.

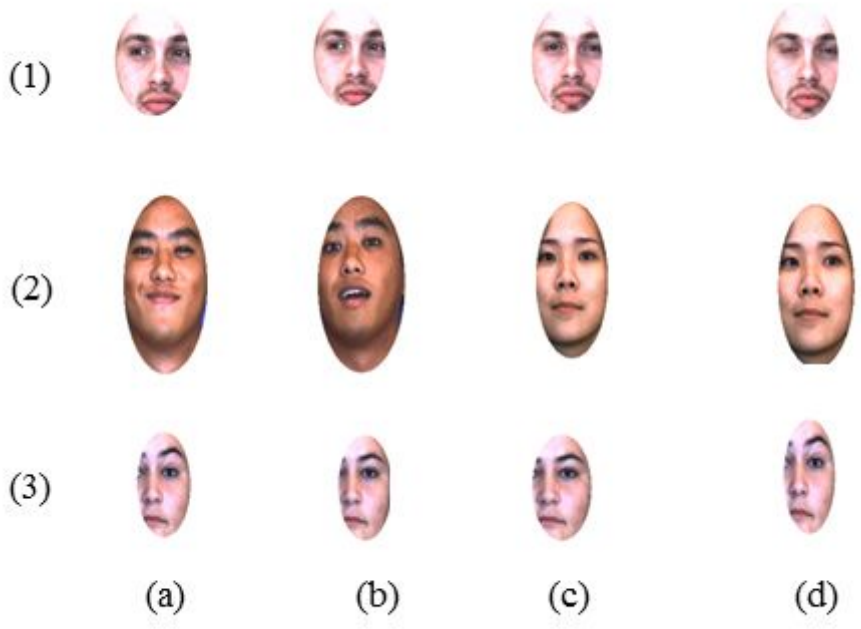

Figure 4.6. Différent cas présentant un défaut de détection.

A partir des images exhibées dans la figure 4.6, on distingue les cas suivants:

1. Les images présentent une mauvaise détection précisément dans le côté gauche et en bas. Après la détection des yeux, on ajuste respectivement la taille du petit et du grand axe en fonction de la distance horizontale entre les yeux et le modèle de visage.
2. Les images présentent une mauvaise détection en bas. Le col est inclus dans l'ellipse entourant le visage, ce qui peut affecter la précision de la reconnaissance. Dans ce cas, la dimension du grand axe est corrigée en fonction de la distance horizontale entre les yeux et la taille de l'axe mineur.
3. Il s'agit d'un cas délicat. Ici, le faux négatif est remarqué. Dans cette situation, nous nous retournons d'abord au module de filtrage afin d'éliminer l'effet d'éclairage, puis nous réessayons encore de trouver le visage et les yeux.

Les résultats de traitement des cas 1,2, 3 dans Figure 4.6 sont exposés dans la figure 4.7.

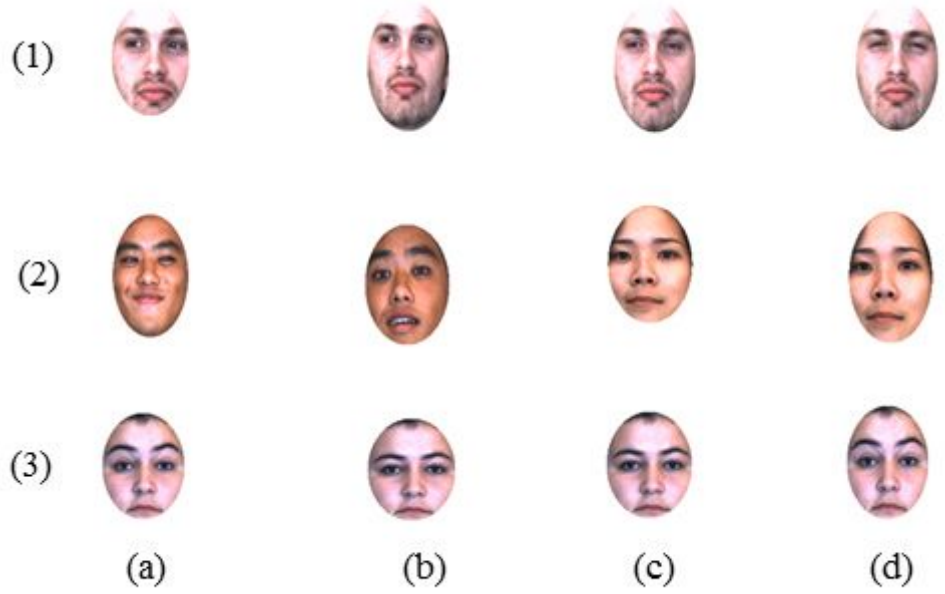

Figure 4.7. Réparation des différentes situations présentées dans la figure 4.6.

6 Extraction des paramètres

Dans la phase de reconnaissance, nous utilisons deux espaces de paramètres, notamment, les moments de Zernike et EAR-LBP (Extended Asymmetric Region Local Binary Pattern) opérateur. Les moments de Zernike sont calculés sur la totalité du visage, et EAR-LBP sont extraites seulement de deux parties du visage. C'est la partie des yeux et de la bouche, qui sont absolument les parties les plus expressives du visage. La Figure 4.8 en donne l'exemple, la combinaison de ces deux caractéristiques rend le système plus robuste, néanmoins, nous devons prendre en compte les performances de classification et le temps de réponse. Pour cela, on procède par une réduction de données. Ensuite, les deux ensembles de caractéristiques sont regroupées dans un seul vecteur, qui à son tour, passe par une étape de sélection de caractéristiques pertinentes basée sur la technique d'entropie floue proposé dans (Jen-Da & Shyi-Ming, 2007). Cette étape permet de réduire la taille du vecteur de paramètres en choisissant uniquement les paramètres pertinents et sans pour autant perdre de précision de la classification. Dans ce qui suit, nous présentons à la fois l'opérateur EAR-LBP et les concepts de moments de Zernike. Par la suite, nous détaillons l'algorithme d'entropie floue.

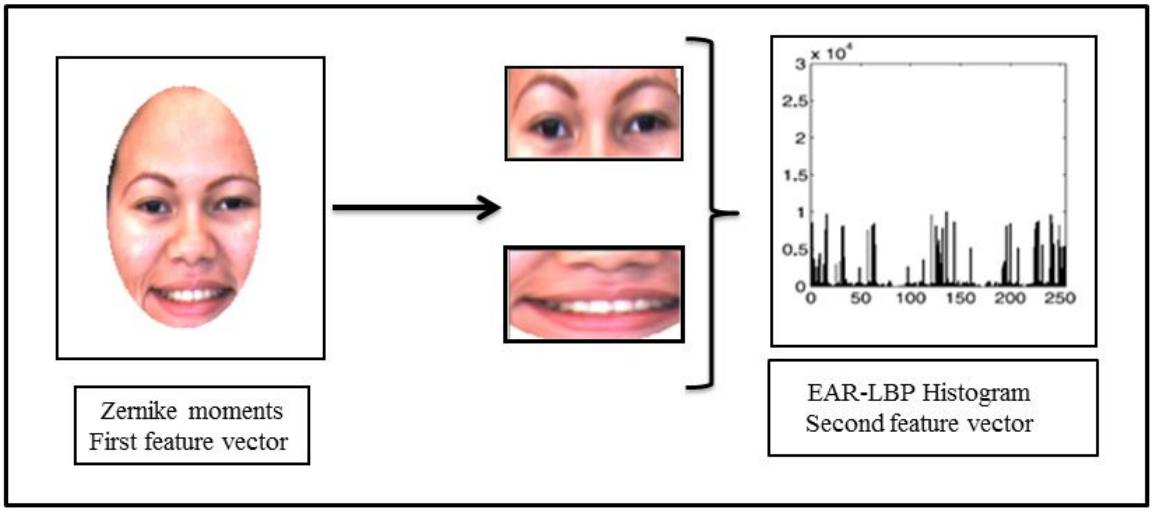

Figure 4.8. Les caractéristiques extraites du visage

6.1 EAR-LBP :

LBP (Local Binary Pattern) est un excellent descripteur, introduit par Ojala et al. (Ojala, Pietikainen, & Harwood, 1996), cet opérateur est largement utilisé dans la classification de texture. Il est incorporé en abondance dans les travaux de recherche récents (Ojala, Pietikainen, & Maenpaa, 2002) (Pei-zhi & Shui-li, 2010) (Hong, Siyu, Lizuo, & Liangzheng, 2011) (Hua, Yihua, Hang, & Yawen, 2012). Dans ce travail, nous utilisons l'opérateur EAR-LBP, proposée dans (Venugopal & Patnaik, 2012), pour décrire les parties des yeux et des lèvres extraites dans l'étape de détection de visage. Chaque image est divisée en sous-régions de 5 × 5 pixels, nous choisissons cette petite taille pour capturer au maximum la variation de l'information dans l'image et par conséquent, augmenter la capacité de discrimination. Le code EAR-LBP est donné par:

$$EAR - LBP(x_c, y_c) = \sum_{i=1}^{7} S(a_i - a_c) \times 2^i \quad (4.19)$$

$$S(x) = \begin{cases} 1, si\ x \geq 0 \\ 0, autrement. \end{cases} \quad (4.20)$$

Où : a_i est la valeur moyenne des pixels appartenant à la région qui entoure les pixels de bordure. a_c, est le pixel central.

La figure 4.9 nous donne plus de détails.

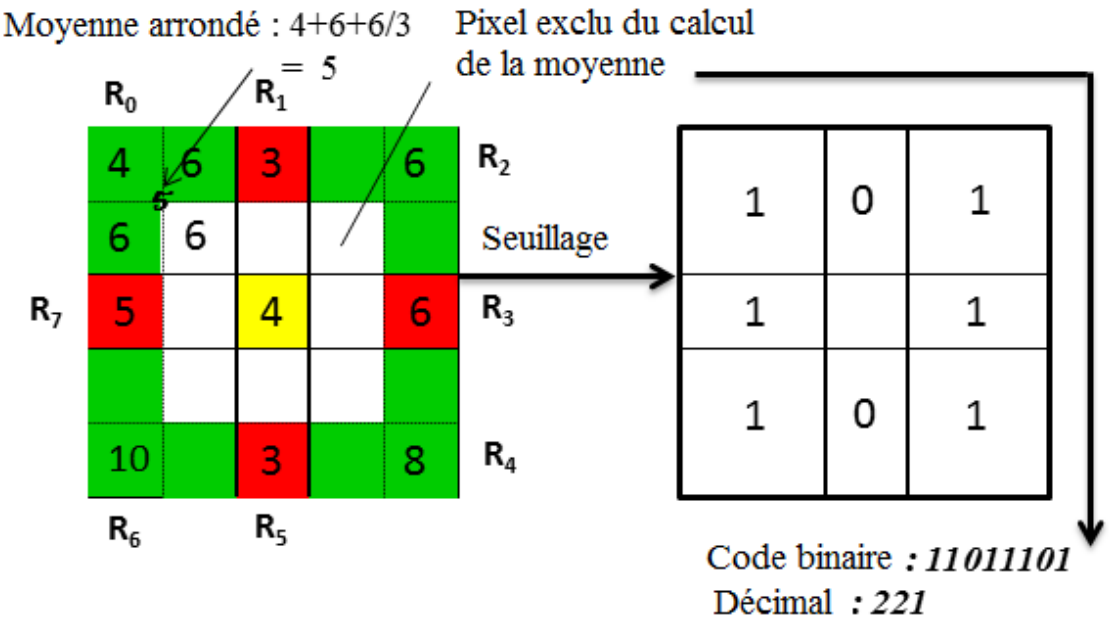

Figure 4.9 : Exemple de 5×5 EAR-LBP

6.2 Les moments de Zernike

Les Moments de Zernike sont qualifiés d'excellentes primitives globales. Ils font partie des moments géométriques, et caractérisés par leur invariance à la translation, au changement de l'échelle et d'orientation. Les polynômes de Zernike sont définis comme suit:

$$V_{n,m} = R_{n,m}(x,y)\exp\left(jm\,tang^{-1}\left(\frac{y}{x}\right)\right) \qquad (4.21)$$

Avec : $x^2 + y^2 \leq 1, n \geq 0, |m| < n, et\ n - |m|$ est pair.

Les polynômes radiaux $R_{n,m}$ sont définis comme suit :

$$R_{n,m}(x,y) = \sum_{s=0}^{(n-|m|)/2} S_{n,|m|,s}(x^2+y^2)^{\frac{n-2s}{2}} \qquad (4.22)$$

$$S_{n,|m|,s} = (-1)^s \frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!} \qquad (4.23)$$

7 Sélection des paramètres

La sélection des caractéristiques peut être considérée comme un problème de réduction de la dimension (Hoai & Nguyen). Elle est sollicitée par de nombreuses approches récentes de

reconnaissance des visages (Aouatif, Rziza, & Driss, 2008) (Satyanadh & Vijayan, 2007) (Yazhou, Hongxun, Wen, & Debin, 2005) (Zhou & Lipo, 2009) (Zhiping, L. Xi, & Qing, 2012). Dans ce premier système, nous utilisons L'algorithme de sélection de sous-ensemble basé sur le travail élaboré dans (Jen-Da & Shyi-Ming, 2007), qui utilise des échantillons extrêmes. Ce sont les modèles mal classés. Également dans cette méthode nous impliquons l'algorithme k-means et nous connaissons d'avance le nombre de classes aussi bien pour la reconnaissance du visage que pour la reconnaissance des émotions. Avant de détailler l'algorithme, nous donnons d'abord, quelques définitions :

Définition 1 : La matrice d'extension EM_f, qui contient les degrés d'appartenance de la caractéristique f appartenant à des ensembles flous, est définie comme suit:

$$EM_f = \begin{bmatrix} U_{v1}(r_{1f}) & \cdots & U_{vm}(r_{1f}) \\ \vdots & \ddots & \vdots \\ U_{v1}(r_{nf}) & \cdots & U_{vm}(r_{nf}) \end{bmatrix} \quad (4.24)$$

Où : n est le nombre des échantillons, m est le nombre des ensembles flous de l'ensembles des caractéristiques f, et $U_{vz}(r_{pf})$ dénote le grade d'appartenance de la valeur r_{pf} de l'ensemble f, de l'échantillon r_p appartenant à l'ensemble flou v_z . avec $1 \leq p \leq n$ et $1 \leq z \leq m$

Définition 2 : La classe des degrés $CD_c(v)$ des échantillons de la classe c appartenant à l'ensemble flou v est définie comme suit :

$$CD_c(v) = \frac{\sum_{r \in R_c} EM_f(|r|, |v|)}{\sum_{r \in R} EM_f(|r|, |v|)} \quad (4.25)$$

Où : R est l'ensemble des échantillons, R_c est l'ensemble des échantillons appartenant à la classe c, $1 \leq r \leq n$ et $1 \leq v \leq m$

Définition 3 : l'entropie floue $FFE(f)$ de l'ensemble de caractéristiques f est donnée par :

$$FFE(f) = \sum_{v \in V} \left[\sum_{v \in V} \frac{S_v}{S} \times \sum_{c \in C} \left(-CD_c(v)\, log_2 CD_c(v) \right) \right] \quad (4.26)$$

Définition 4 : La matrice d'extension contient les grades d'appartenance des valeurs des caractéristiques du sous-ensemble $\{f_1, f_2\}$, elle est notée $CEM(f_1, f_2, T_r)$, et définie comme suit :

$$CEM(f_1, f_2, T_r)$$
$$= \begin{pmatrix} U_{v11}(r_{1f_1}) \wedge U_{v21}(r_{1f_2}) & \cdots & U_{v11}(r_{1f_1}) \wedge U_{v2j}(r_{1f_2}) & \ldots & U_{v11}(r_{1f_1}) \wedge U_{v21}(r_{1f_2}) & \ldots & U_{v11}(r_{1f_1}) \wedge U_{v21}(r_{1f_2}) \\ \vdots & \ldots & \vdots & \ldots & \vdots & \ldots & \vdots \\ U_{v11}(r_{1f_1}) \wedge U_{v21}(r_{1f_2}) & \ldots & U_{v11}(r_{1f_1}) \wedge U_{v21}(r_{1f_2}) & \ldots & U_{v11}(r_{1f_1}) \wedge U_{v21}(r_{1f_2}) & \ldots & U_{v11}(r_{1f_1}) \wedge U_{v21}(r_{1f_2}) \end{pmatrix} \quad (4.27)$$

$T_r \in [0,1]$ est un seuil défini par l'utilisateur, dans notre cas il est fixé à 0,5. i, j dénotent, respectivement, le nombre d'ensemble flou de la caractéristique f_1, f_2 dont la classe des degrés est inférieure à T_r .

$U_{v1x}(r_{pf_1})$ est le grade d'appartenance de la valeur r_{pf_1} de la caractéristique f_1 de l'échantillon r_p, et $U_{v2y}(r_{pf_2})$ est le grade d'appartenance de la valeur r_{pf_2} de la caractéristique f_2 de l'échantillon r_p. Avec, $1 \leq x \leq i$ et $1 \leq y \leq j$, $\wedge$ est l'opérateur minimum.

Définition 5 : l'entropie floue du sous-ensemble de caractéristiques $\{f_1, f_2\}$ basée sur les échantillons extrêmes est définie ainsi :

$$BSFFE(f_1, f_2) = \begin{cases} \frac{S_{1B}}{S_1} \times \sum_{w \in v_{FS}} \frac{S_w}{S_{FS}} FE(w) + \sum_{v_1 \in v_{1 \cup B}} \frac{S_{v1}}{S_1} FE(v_1) \\ \frac{S_{2B}}{S_2} \times \sum_{w \in v_{FS}} \frac{S_w}{S_{FS}} FE(w) + \sum_{v_2 \in v_{2 \cup B}} \frac{S_{v2}}{S_2} FE(v_2) \end{cases} \quad (4.28)$$

L'algorithme de sélection des caractéristiques est montré dans la figure 4.10.

1. Use the k-means to generate the k cluster centers (k is the number of class)
2. Construct the membership of the fuzzy sets based on these k cluster centers
3. Construct the extension matrix EM_f
4. Calculate the fuzzy entropy $FE(v)$ of each fuzzy set v of the feature f
5. Calculate the fuzzy entropy $FFE(f)$ of the feature f
6. Let $\hat{f} = argmin_{f \in F} FFE(f)$
7. Let $E_{FS} = FFE(\hat{f})$
8. Let $FS = \{\hat{f}\}$
9. Let $F = F - \{\hat{f}\}$
10. For each $f \in F$ do
11. {
12. Construct the extension matrix of $FS \cup \{f\} = CEM(FS, f, T_r)$
13. Calculate $CD_c(v)$ of the samples of each combined fuzzy set of the feature subset $FS \cup \{f\}$
14. Calculate the fuzzy entropy $FE(v)$ of each combined fuzzy set of the feature subset $FS \cup \{f\}$
15. Calculate the fuzzy entropy $BSFFE(FS, f)$ of the feature subset $FS \cup \{f\}$
16. Let $\hat{f} = argmin_{f \in F} BSFFE(f)$
17. Let $D = E_{FS} - BSFFE(FS, \hat{f})$
18. Let $E_{FS} = BSFFE(FS, \hat{f})$
19. Let $FS = FS \cup \{\hat{f}\}$
20. Let $F = F - \{\hat{f}\}$
21. }until ($E_{FS} = 0$ **or** $D \leq 0$ **or** $F = \emptyset$

Figure 4.10. Algorithme de sélection des paramètres.

8 Présentation du deuxième système

Ce système emploie les mêmes modules de prétraitement des images et de sélection des visages ainsi que les caractéristiques faciales. Même les paramètres utilisés sont les moments de Zernike et les EAR-LBP. La seule différence c'est que ces derniers sont appliqués en multi-échelles. Même la sélection des caractéristiques est substituée par une méthode de réduction des paramètres qui utilise les mémoires auto-associatives.

Le but de ce système est de satisfaire les préférences de l'utilisateur en termes de programmes de télévision. Pour cela, la Smart TV (téléviseur intelligent) contient un module qui permet d'enregistrer les préférences de chaque utilisateur comme étant son profil. Le profil se compose de deux parties: les données personnelles (nom, âge, etc.) et les préférences de l'utilisateur (programmes favoris et les canaux, le temps de diffusion, langue, etc.). Chaque nouvel utilisateur, qui est généralement un membre de la famille, doit passer par cette étape d'authentification afin de profiter de ce système. Dans le cas contraire, la Smart TV sera utilisée d'une manière classique. Dès que la télévision est mise en service, le système cherche

à identifier la personne. Une fois reconnue, la Smart TV diffuse le canal par défaut sélectionné par l'utilisateur. Ensuite, le système de reconnaissance des émotions se déclenche et contrôle l'état émotionnel de l'utilisateur. Par exemple, s'il détecte que l'utilisateur est dégoûté ou triste; il cherche une autre chaine qui satisfait ses besoins, en se référant aux préférences de l'utilisateur stockées dans son profil. Ce dernier peut aussi être mis à jour à la demande de l'utilisateur. L'architecture du système est montrée dans Figure 4.11.

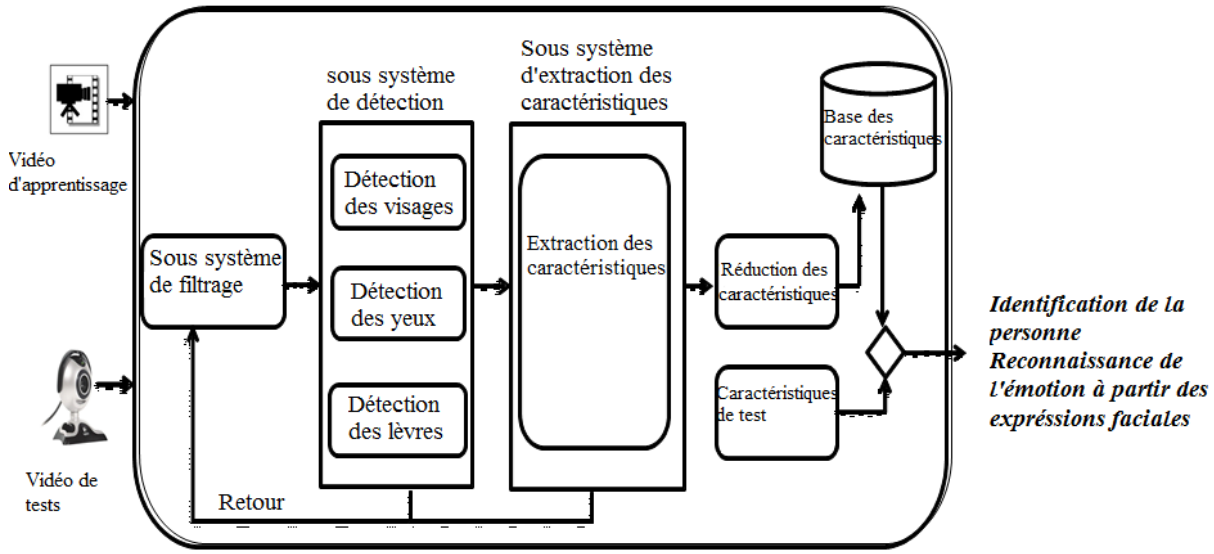

Figure 4.11. Architecture du système.

Les modules de prétraitement, de détection, et de paramétrage sont les mêmes utilisés dans le premier système, il est donc inutile de les réexpliquer. Dans ce qui suit, nous allons détailler la partie de réduction des caractéristiques et de classification.

Dans cette approche, nous utilisons deux espaces de paramètres pour représenter les visages, ce qui donne un vecteur caractéristique de taille très importante qui peut affecter la phase de classification. Pour remédier à cet inconvénient, nous utilisons une méthode de réduction des paramètres basée sur les mémoires auto-associative. Certains travaux la considèrent comme une ACP (analyse en composantes principales) non-linéaire (Reyes, Vellasco, & Tanscheit, 2013) (Parviainen & Bottleneck, 2010) (Scholkopf, Smola, & Muller, 1998), car elle permet de ne retenir que les primitives non-linéairement corrélés. Les mémoires auto-associatives sont des réseaux de neurones qui reproduisent en sortie les données injectées en entrée. Ces réseaux sont composés de cinq couches dont troisième contient le plus petit nombre d'unités,

elle sert d'une couche de compression. Dans la figure 4.12, nous retrouvons plus d'explications.

Les mémoires auto- associatives servent à reproduire les données d'entrée $(x_1, x_2, .., x_n)$ dans la couche de sortie $(\hat{x}_1, \hat{x}_2, .. \hat{x}_n)$ qui est appelée mappage d'identité (Pentland & Choudhury., 2000). En fait, ce processus comporte deux étapes. Premièrement, les données en entrée sont réduites dans la couche d'étranglement, c'est une phase de codage. Deuxièmement, les données de la couche d'étranglement sont décompressées à la couche de sortie, il s'agit d'une phase de décodage. Le processus d'apprentissage est arrêté lorsque les données en sortie sont plus proches que possible à ceux de l'entrée. Processus d'encodage est interprété comme suit:

$$v_k = \sum_{j=1}^{N_2} w_{2jk}\, \sigma\left(\sum_{i=1}^{N_1} w_{1ij}\, x_i + \theta_i\right), k = 1..M \quad (4.29)$$

Avec : N_1 est le nombre de neurones dans la couche d'entrée, N_2 est le nombre de neurones dans les couches cachées, M étant la taille de la couche d'étranglement, w représente le poids et θ correspond au biais. σ est une fonction sigmoïde définie comme ainsi :

$$\sigma = \frac{1}{1 + e^x} \qquad (4.30)$$

Le processus de décodage à son tour peut être reformulé comme suis :

$$\hat{x}_k = \sum_{j=1}^{N_2} w_{4jk}\, \sigma\left(\sum_{i=1}^{M} w_{3ij}\, v_i + \theta_i\right), k = 1..N_1 \quad (4.31)$$

Le processus d'apprentissage s'arrête dès que l'erreur E soit minimale, elle est calculée par la formule suivante:

$$E = \sum_{n=1}^{N} \sum_{i=1}^{N_1} (x_{in} - \hat{x}_{in})^2 \qquad (4.32)$$

Où : N est le nombre d'échantillon.

Dans notre étude, les vecteurs de caractéristiques qui correspondent respectivement au Pseudo moments de ZERNIKE et MS-EAR-LBP, les deux sont regroupés en un vecteur unique qui sert d'entrée pour le réseau de mémoires auto-associatives. La couche d'entrée contient 303

unités (neurone) ce qui équivaut à la taille du vecteur caractéristique. La deuxième couche contient à son tour, l'unité 909, et la couche de compression ou d'étranglement contient 60 unités. Ensuite, les données dans la ($3^{ème}$ couche) sont utilisées comme entrée pour un réseau MLP (perceptron multicouches) qui fournit les résultats de la classification.

Le réseau MLP contient 60 neurones dans la couche d'entrée, 120 neurones dans la couche cachée et 7 neurones dans la couche de sortie pour la reconnaissance des émotions (type d'émotion) ou 10 neurones (correspondants au nombre de personnes dans le base d'apprentissage) pour la reconnaissance du visage (voir la figures 4.12, 4.13). Ce classifieur a été mis en œuvre selon l'architecture du perceptron multicouche, à l'aide du logiciel NeurophStudio (http://neuroph.sourceforge.net). Après les expérimentations de différentes topologies de réseau, la meilleure précision a été trouvée avec 120 neurones dans la couche caché. Le critère d'arrêt précoce a été déterminé après « l » validation sur l'ensemble d'apprentissage, et le nombre d'époques a été sélectionné à 1500. Cela garantit que le processus d'apprentissage s'arrête quand l'erreur moyenne carrée MSE (Mean Squares Error) commence à augmenter. Le taux de l'apprentissage a été laissé à 0,3. L'algorithme de rétro-propagation d'erreur a été utilisé pour l'apprentissage. En outre, tous les neurones de notre architecture suivent la fonction d'activation sigmoïde, tandis que tous les attributs ont été normalisés pour l'amélioration des performances du réseau.

Conclusion

A la fin de chapitre, nous rappelons que la validation de ces concepts a été menée sur deux bases de données, une base d'image (JAFFE) et l'autre des vidéos (FABO). Dans le chapitre subséquent, nous donnons les résultats des expérimentations et nous dénudons les points forts et les faiblesses de tous les modules inclus dans les deux systèmes.

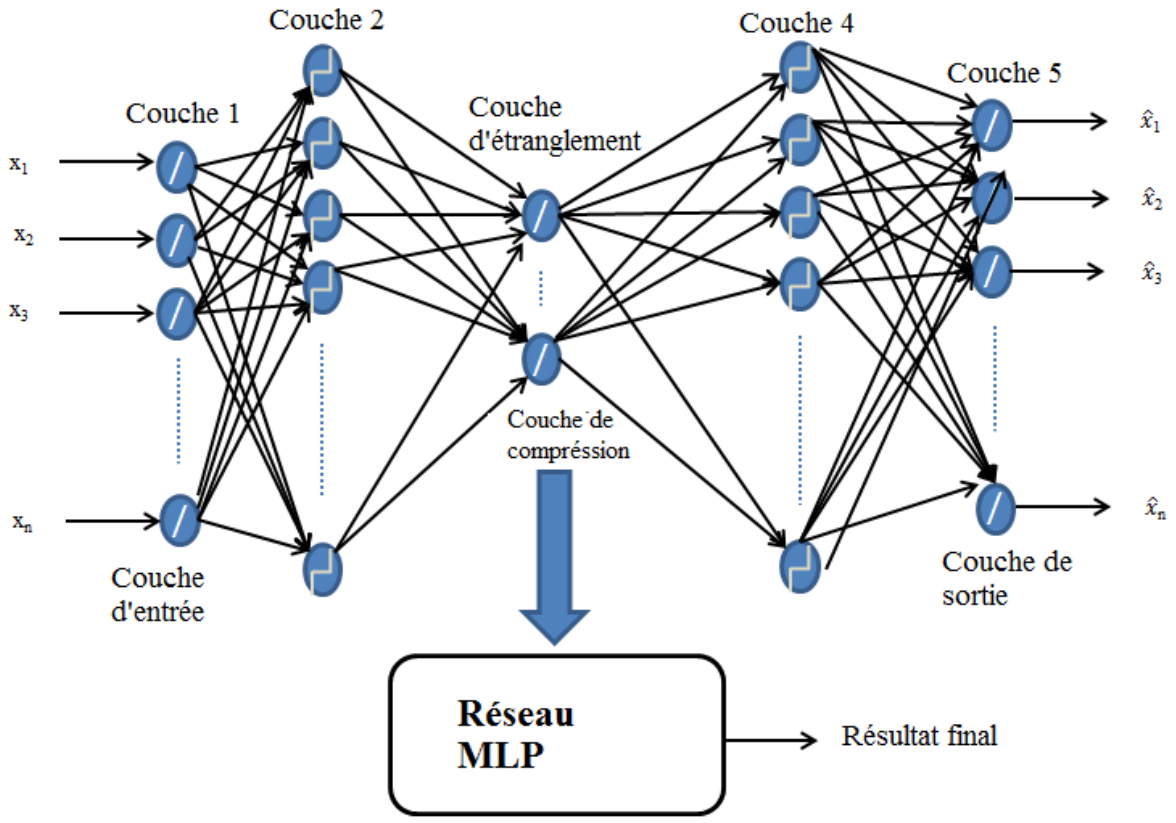

Figure 4.12. Architecture du système de classification.

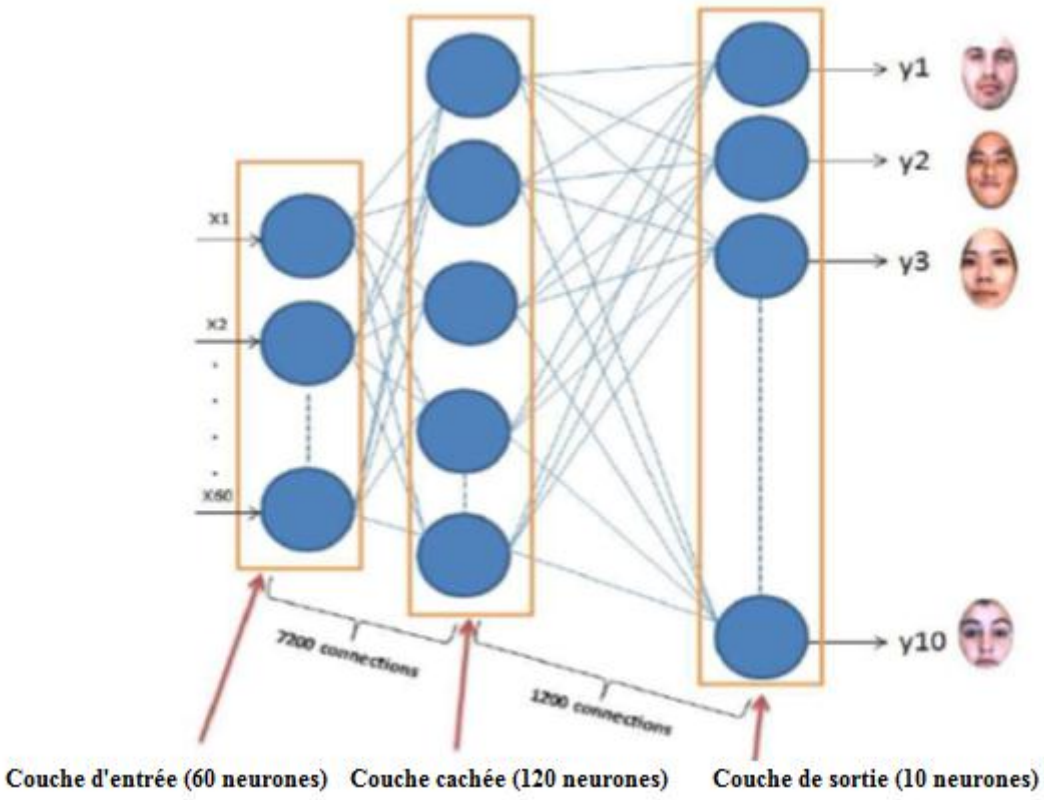

Figure 4.13. Architecture du système de reconnaissance des visages.

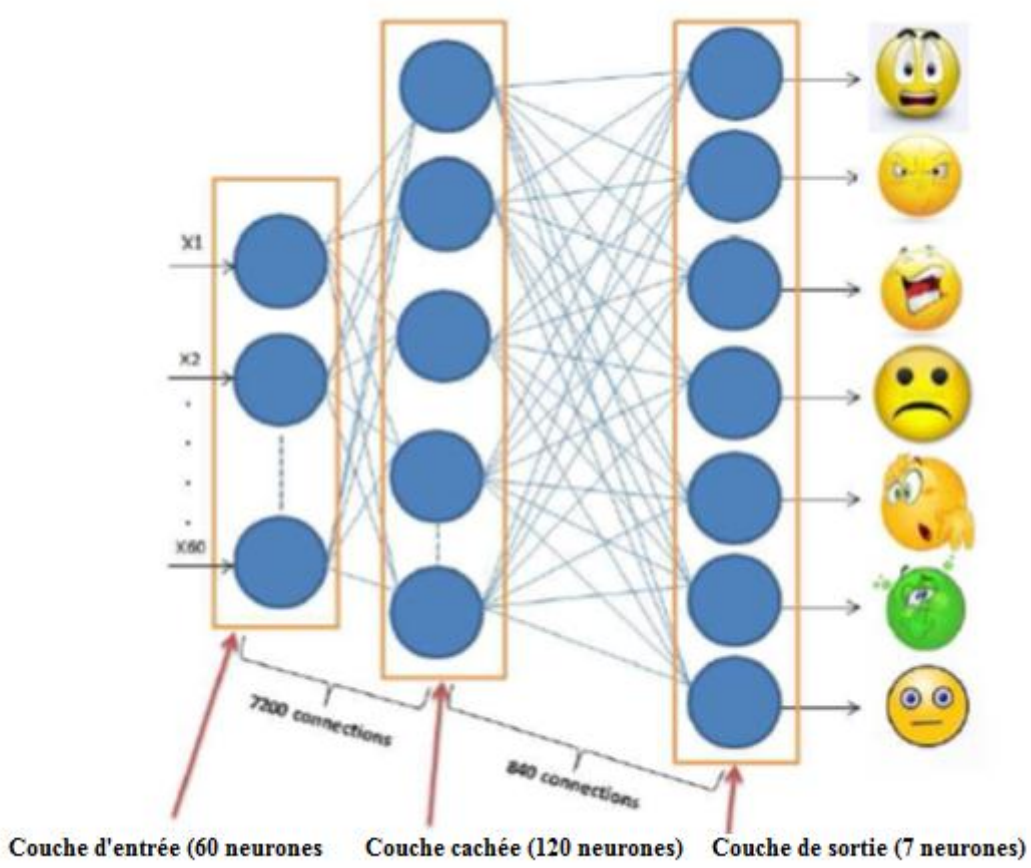

Figure 4.14. Architecture du système de reconnaissance des émotions.

Chapitre 5

Expérimentations

Introduction

Dans cette partie nous déployons la démarche suivie pour expérimenter les deux approches montrées dans le chapitre précédent. Pour ce faire, nous avons effectué des tests pour tous les niveaux de traitement, commençant par la détection des visages jusqu'à la reconnaissance de l'état émotionnel. Nous avons aussi montré l'effet des modules de prétraitement dans l'amélioration, à tous les niveaux, de la qualité du traitement.

Nous commençons d'abord à définir les bases de données utilisées dans les tests, puis nous donnons les résultats des expérimentations établies par les deux systèmes. Nous dressons également pour chaque approche un état comparatif qui permet de mesurer chaque système par rapport à ceux existants.

1. Les bases des visages

1.1 La base Jaffe

Cette base de données contient 213 images contenant 7 expressions faciales (6 expressions de base + 1 neutre) représentées par les dix modèles féminins japonais. Chaque image a été évaluée sur six émotions différentes. La base Jaffe a été conçue et assemblée par Michael Lyons, Miyuki Kamachi, et Jiro Gyoba. Les photos ont été prises au département de psychologie de l'université de Kyushu.

1.2 La base FABO

La base bimodale des visages et des gestes corporels FABO est destinée pour l'analyse automatique du comportement émotionnel non verbal de l'humain. Elle a été créée à la Faculté de Technologie de l'information à l'Université de Technologie de Sydney (UTS) par Hatice Gunes et Massimo Piccardi en 2005. Les données visuelles prises ont été recueillis par des bénévoles dans un environnement de laboratoire en demandant et dirigeant les participants sur les actions et les mouvements nécessaires.

La base FABO contient des vidéos reflétant les expressions du visage et du corps humain enregistrées simultanément par des caméras du visage et du corps. Cette base de données est la première à combiner les expressions faciales et les gestes corporels d'une manière vraiment bimodale, donc permettant une future progression significative dans la recherche affective en informatique.

23 est le nombre total de sujets dont 10 d'Europe, 2 de Moyen-Orient, 3 d'Amérique latine, 7 d'Asie et 1 d'Australie. L'âge est entre 18 et 50 ans. La base se compose de 12 femmes et 11 hommes.

2. Expérimentation du premier système

2.1 Reconnaissance de l'émotion à partir des expressions faciales

La détection des visages

Tout d'abord, avec la base de données FABO nous testons la reconnaissance du visage; nous considérons tous les cas qui peuvent survenir, en particulier, les différentes positions de la tête avec également l'ensemble des cas exposés aux différents changements de luminance. Nous montrons aussi l'impact de la phase de filtrage dans l'amélioration de la qualité de la détection et ainsi l'augmentation de la précision.

Le tableau 5.1 nous montre les différents taux de détection obtenus lors des tests effectués sur des vidéos de neuf personnes dont chaque vidéo contient plus d'une centaine de séquences.

Nous utilisons quatre vidéos de chaque personne. Pour ceux ayant une luminance élevée, nous avons appliqué un filtrage afin d'améliorer la qualité de détection. Le tableau 5.2 nous montre les résultats obtenus. Deuxièmement, nous testons avec base de données JAFFE où les images sont petites, déjà traitées et en niveau de gris. Dans cette deuxième partie, il est donc faisable de détecter des visages sans passer par le prétraitement. Nous n'accordons pas d'importance au temps d'exécution, qui est très court, et à l'échelle de l'image. Les résultats sont présentés dans le tableau 5.7.

Vidéo	Nombre de séquences	Avant correction	Après correction
Personne 1	498	477 (95,78 %)	495 (99,39 %)
Personne 2	776	752 (96,90 %)	773 (99,61 %)
Personne 3	958	958 (100 %)	958 (100%)
Personne 4	1081	1073 (99,26 %)	1080 (99,90 %)
Personne 5	406	406 (100 %)	406 (100 %)
Personne 6	593	566 (95,44 %)	593 (100 %)
Personne 7	404	404 (100 %)	404 (100 %)
Personne 8	632	632 (100 %)	632 (100 %)
Personne 9	704	704 (100 %)	704 (100 %)
Total	6052	5972 (98.67 %)	6045 (99.88 %)

Tableau 5.1. taux de détection des visages obtenus pendant les tests avec la base FABO

Vidéo	Nombre de séquences	Avant correction	Après correction
Personne 1	161	152 (94,40%)	159 (98,75 %)
Personne 4	112	80 (71,42 %)	112 (100 %)
Personne 6	99	21 (21,21 %)	96 (96,96 %)
Personne 8	240	145 (60,41 %)	240 (100 %)
Personne 10	425	0 (0 %)	420 (98,82 %)
Total	1037	398 (38,37%)	1027 (99.03 %)

Tableau 5.2. taux de détection avant et après filtrage pour les vidéos à haute luminance

Images	Nombre d'images	Bonne détection de visage	Bonne détection des yeux et lèvres	Taux de reconnaissance
Personne 1	21	21 (100 %)	21 (100%)	21 (100%)
Personne 2	22	22 (100%)	21 (95,45%)	21 (100%)
Personne 3	22	21 (95,45%)	20 (95,23%)	20 (100%)
Personne 4	20	20 (100 %)	20 (100 %)	20 (100 %)
Personne 5	21	20 (95,23 %)	20 (100%)	19 (95%)
Personne 6	21	20 (95,23 %)	20 (100%)	19 (95%)
Personne 7	20	20 (100 %)	19 (95%)	19 (100%)
Personne 8	21	19 (90,47%)	19 (100%)	19 (100%)
Personne 9	21	21 (100 %)	21 (100 %)	21 (100 %)
Personne 10	22	22 (100 %)	20 (90,91%)	19 (95%)
Total	211	206 (97.76 %)	201 (97,57%)	98,50%

Tableau 5.3. taux de détection des visages, des yeux et des lèvres, et de reconnaissance des visages avec la base Jaffe.

Détection des yeux et des lèvres

Après l'étape de détection, pour chaque visage localisé, l'extraction des yeux et des lèvres est effectuée. Pour la base de données FABO, l'opération de filtrage est effectuée pour améliorer le contraste. Dans cette étape, nous testons les séquences d'images en différents échelles, et nous calculons le temps de traitement dans chaque étape. Les résultats des tests élaborés sur les bases de données FABO et JAFFE sont présentés respectivement dans les tableaux 5.3 et 5.4.

Vidéos	Nb. de séquences	Echelle 1		Echelle 2		Echelle 3		Echelle 4	
		Temps (s)	Taux %	Temps (s)	Taux %	Temps (s)	Taux %	Temps (s)	Taux %
Personne 1	659	3702	28,52	2334	42,33	923	90,74	453	80,42
Personne 2	776	4464	14,69	2927	16,11	1045	64,56	467	93,68
Personne 3	958	5826	0,005	3510	41,44	1284	85,07	582	99,68
Personne 4	1193	6308	0,004	4794	0,08	1597	87,17	699	88,76
Personne 5	406	2641	32,26	1428	94,58	596	100	272	100
Personne 6	692	3979	55,06	2300	97,25	916	99,85	476	98,55
Personne 7	404	2862	0	1469	41,83	539	42,27	254	64,85
Personne 8	872	4765	19,26	3320	37,84	1152	88,42	529	97,82
Personne 9	704	3603	22,30	2293	45,03	1053	87,50	473	92,33
Personne 10	425	2655	13,88	1495	41,88	623	90,35	287	98,82

Tableau 5.4. Taux de détection des yeux et des lèvres avec 4 échelles : échelle 1= 1026×770, échelle 2= 800×600, échelle 3= 500×400, et l'échelle 4= 300×300

La reconnaissance des visages

Après l'extraction des caractéristiques qui sont, respectivement, les moments de Zernike extraites sur l'ensemble du visage détecté dans une ellipse et les EAR- LBP extraites sur les parties des yeux et des lèvres, nous commençons l'apprentissage pour sélectionner des caractéristiques pertinentes. Dans notre cas, nous testons avec des images qui représentent respectivement dix personnes de la base de données FABO et dix autres de base de données JAFFE. Avec FABO, nous prenons cinq vidéos pour chaque personne où cinq cents séquences sont utilisées dans l'apprentissage. Avec JAFFE, nous utilisons seulement cinq images de chaque personne pour l'apprentissage.

On définit le vecteur V_1 qui contient les moments de Zernike et le vecteur V_2 qui contient les EAR-LBP (voir la figure 4.9).

La dimension de V_1 est 47, $V_1 = MZ_1, MZ_2, \ldots, MZ_{47}$, et la taille du vecteur V_2 est de 256, V_1 et V_2 sont regroupés dans un troisième vecteur noté $V_3 = V_1 + V_2$. Une fois la sélection des caractéristiques est établie, la taille du vecteur V_3 est réduite à 122 au lieu de 303 pour FABO et à 112 pour JAFFE. Tableau 5.5 et le Tableau 5.3 contiennent respectivement les résultats de l'étape de reconnaissance de visage pour les bases de données FABO et Jaffe.

D'après le tableau 8, on voit que la dimension 300 x 300 est meilleure, d'une part parce que le temps de réponse est trop court (2 sec. Par séquence) et, d'autre part, le taux de détection est quasiment élevé. Dans certains cas, l'échelle de 500 × 400 donne de meilleurs résultats, mais le temps de traitement est trop long, ce qui retarde le temps de réponse du système.

D'après les résultats présentés dans le tableau 5.3, nous calculons un taux de reconnaissance moyen de 98,50%. Ce dernier présente un taux très satisfaisant sauf dans les cas de personne 5, personne 6 et personne 10. Par exemple, pour personne 5 et personne 10, les traits du visage changent complètement entre la peur et l'émotion surprise. D'après les résultats présentés dans le tableau 5.5, nous calculons le taux de reconnaissance moyen qui est de 96,80%. Pour les cas de personne 5, personne 7, personne 9, et personne 10, le taux de reconnaissance est de 100% parce que ces vidéos ont des séquences d'images de faible contraste, et ne sont pas bruitées. Pour le cas de personne 2, personne 3 et personne 6, les taux de reconnaissance sont en dessous de la moyenne, car certaines séquences de ces vidéos ont subi plusieurs opérations de filtrage pour pallier aux problèmes liés à la présence de la barbe pour la personne 2, et le problème de luminance pour personne 3 et personne 6.

	Vidéo 1	**Vidéo 2**	**Vidéo 3**	**Vidéo 4**	**Vidéo 5**	**Nb. de séquences.**	**Taux moyen**
Personne 1	97.32 %	100 %	94.68 %	94.40 %	93.78 %	655	96.03 %
Personne 2	92.06 %	92.59 %	94.77 %	94.11 %	—	776	93.81 %
Personne 3	89.18 %	94.69 %	96.22 %	95.20 %	—	958	92.90 %
Personne 4	99.09 %	100 %	100 %	95.84 %	91.07 %	1193	96.89 %
Personne 5	100 %	100 %	100 %	100 %	—	406	100 %
Personne 6	87.64 %	100 %	93.93 %	90.15 %	94.92 %	692	93.35 %
Personne 7	100 %	100 %	% 100	100 %	—	393	100 %
Personne 8	100 %	93.82 %	93.89 %	97.14 %	91.66 %	872	95.06 %
Personne 9	100 %	100 %	100 %	100 %	—	701	100 %
Personne 10	100 %	100 %	100 %	—	—	425	100 %

Tableau 5.5. taux de reconnaissance des visages pour la base FABO

Reconnaissance de l'émotion :

Pour la reconnaissance des émotions, nous employons les mêmes caractéristiques. Nous prenons en compte les sept émotions connues: la joie, la tristesse, la peur, la surprise, la colère, le dégoût et l'émotion neutre. Pour l'évaluation de la reconnaissance des émotion, les mesures de précision et de rappel sont utilisées. Les paramètres sont définis comme suit:

$$Rappel = \frac{nombre\ de\ bonne\ détection\ trouvées}{nombre\ d'images\ existantes} \qquad (5.1)$$

$$Précision = \frac{nombre\ de\ bonne\ détection\ trouvées}{nombre\ d'image\ associées\ à\ l'émotion} \qquad (5.2)$$

Les résultats de la reconnaissance de l'émotion sont présentés dans les tableaux 10 et 11. Pour la base de données FABO, nous ne présentons que 2 vidéos pour 8 personnes. Comme on peut le distinguer, la métrique de rappel est assez élevée, elle atteint 85%.

Le tableau 5.6 récapitule les résultats de la classification de la reconnaissance des émotions pour chaque personne dans la base FABO (première colonne). Nous notons l'émotion figurée

dans la Base des visages par (EDB), elle représente le nombre d'occurrences de chaque émotion dans la vidéo, une bonne détection (GD), qui représente le nombre des émotions détectées correctement, fausse détection (FD), qui représente le nombre des émotions incorrectement détectées, les mesures de rappel et de précision. Nous utilisons cinq vidéos par personne (dix personnes), mais en raison de contraintes d'espace, nous ne transmettons ici que deux vidéos par personne et pour 8 individus uniquement.

Le tableau 5.7 résume les résultats de la reconnaissance des émotions pour chaque personne dans la base Jaffe (première colonne); nous définissons la base des émotions par (EDB), qui représente le nombre de chaque émotion, une bonne détection par (GD), qui représente le nombre d'émotion détecté correctement, et fausse détection (FD) qui représente le nombre d'émotion incorrectement détecté. Nous donnons aussi les métriques (rappel et précision).

		Vidéo 1							**Vidéo 2**						
		F	**A**	**H**	**N**	**S**	**D**	**SU**	**F**	**A**	**H**	**N**	**S**	**D**	**SU**
Personne 1	EDB	50	0	0	50	12	0	0	0	0	60	60	0	0	25
	GD	40	0	0	43	10	0	0	0	0	57	53	0	0	19
	FD	1	6	0	3	0	7	2	0	0	5	4	3	2	2
	Rappel	0,8	–	–	0,86	0,83	–	–	–	–	0,95	0,88	–	–	0,76
	Précision	0,97	–	–	0,93	0,83	–	–	–	–	0,91	0,92	–	–	0,90
Personne 2	EDB	0	40	0	50	36	0	0	40	0	0	40	0	28	0
	GD	0	36	0	41	21	0	0	31	0	0	29	0	22	0
	FD	7	7	3	5	3	3	0	10	4	0	3	5	4	0
	Rappel	–	0,9	–	0,82	0,58	–	–	0,77	–	–	0,72	–	0,78	–
	Précision	–	0,83	–	0,89	0,87	–	–	0,75	–	–	0,90	–	0,84	–
Personne 3	EDB	85	0	0	100	82	0	103	142	0	0	103	0	38	0
	GD	74	0	0	83	77	0	89	125	0	0	89	0	29	0
	FD	17	1	5	12	8	0	4	15	2	2	12	1	8	0
	Rappel	0,87	–	–	0,83	0,93	–	0,86	0,88	–	–	0,86	–	0,76	–

	Précision	0,81	–	–	0,87	0,90	–	0,95	0,89	–	–	0,88	–	0,78	–
Personne 4	EDB	51	0	0	50	10	0	0	0	40	0	29	0	0	30
	GD	45	0	0	43	8	0	0	0	33	0	24	0	0	26
	FD	2	2	2	4	5	0	0	0	6	2	5	0	0	3
	Rappel	0,88	–	–	0,86	0,8	–	–	–	0,82	–	0,82	–	–	0,86
	Précision	0,95	–	–	0,91	0,61	–	–	–	0,84	–	0,82	–	–	0,89
Personne 5	EDB	0	0	68	26	0	0	37	70	0	0	3	28	0	0
	GD	0	0	68	21	0	0	31	62	0	0	2	22	0	0
	FD	0	0	4	2	0	0	5	3	0	1	3	7	0	1
	Rappel	–	–	1	0,80	–	–	0,83	0,88	0	–	0,66	0,78	–	–
	Précision	–	–	0.94	0,91	–	–	0,86	0,95	–	–	0,40	0,75	–	–
Personne 6	EDB	0	0	29	29	0	0	31	47	0	0	34	33	0	0
	GD	0	0	22	24	0	0	29	42	0	0	30	29	0	0
	FD	0	1	3	1	2	1	6	2	2	0	6	3	0	0
	Rappel	–	–	0,75	0,82	–	–	0,93	0,89	–	–	0,88	0,87	–	–
	Précision	–	–	0,88	0,96	–	–	0,82	0,95	–	–	0,83	0,90	–	–
Personne 7	EDB	0	0	0	16	0	45	17	0	31	0	19	0	48	0
	GD	0	0	0	12	0	43	14	0	28	0	15	0	42	0
	FD	2	0	0	4	0	2	1	1	2	0	4	2	3	1
	Rappel	–	–	–	0,75	–	0,95	0,82	–	0,90	–	0,78	–	0,87	–
	Précision	–	–	–	0,75	–	0,95	0,93	–	0,93	–	0,78	–	0,93	–
Personne 8	EDB	0	0	74	29	0	0	61	75	0	0	28	59	0	0
	GD	0	0	70	25	0	0	58	69	0	0	26	54	0	0
	FD	2	1	3	3	0	0	2	4	1	0	4	4	0	0
	Rappel	–	–	0,94	0,86	–	–	0,95	0,92	–	–	0,92	0,91	–	–

	Précision	–	–	0,95	0,89	–	–	0,96	0,94	–	–	0,86	0,93	–	–

Tableau 5.6. taux de reconnaissance des émotions pour la base FABO

		Peur	**Colère**	**Joie**	**Neutre**	**Triste**	**Dégout**	**Surpris**
Personne 1	EDB	4	3	4	3	3	3	3
	GD	3	3	4	2	3	3	3
	FD	0	2	0	0	0	0	0
	Rappel	0,75	1	1	0,66	1	1	1
	Précision	1	0,6	1	1	1	1	1
Personne 2	EDB	3	3	3	3	3	4	3
	GD	3	3	3	2	3	3	2
	FD	1	2	0	0	0	0	0
	Rappel	1	1	1	0,66	1	0,75	0,66
	Précision	0,75	0,6	1	1	1	1	1
Personne 3	EDB	3	3	4	4	3	2	3
	GD	2	1	4	3	3	2	1
	FD	0	0	1	6	0	0	1
	Rappel	0,66	0,33	1	0,75	1	1	0,33
	Précision	1	1	0,8	0,33	1	1	0,50
Personne 4	EDB	3	3	2	3	3	3	3
	GD	2	3	2	2	2	3	3
	FD	0	0	0	1	1	1	0
	Rappel	0,66	1	1	0,66	0,66	1	1
	Précision	1	1	1	0,66	0,66	0,75	1
Personne 5	EDB	3	3	3	3	3	3	3
	GD	2	3	2	3	3	3	2

	FD	0	2	0	1	0	0	0
	Rappel	0,66	1	0,66	1	1	1	0,66
	Précision	1	0,60	1	0,75	1	1	1
Personne 6	EDB	3	3	3	3	3	3	3
	GD	3	2	3	3	3	3	3
	FD	0	0	0	0	1	0	0
	Rappel	1	0,66	1	1	1	1	1
	Précision	1	1	1	1	0,75	1	1
Personne 7	EDB	3	3	3	3	3	2	3
	GD	3	3	3	3	2	1	3
	FD	0	1	0	1	0	0	0
	Rappel	1	1	1	1	0,66	0,50	1
	Précision	1	0,75	1	0,75	1	1	1
Personne 8	EDB	3	3	3	3	3	3	3
	GD	3	3	3	3	3	2	2
	FD	0	0	0	1	1	0	0
	Rappel	1	1	1	1	1	0,66	0,66
	Précision	1	1	1	0,75	0,75	1	1
Personne 9	EDB	3	3	3	3	3	3	3
	GD	3	3	3	3	2	2	3
	FD	0	1	0	1	0	0	0
	Rappel	1	1	1	1	0,66	0,66	1
	Précision	1	0,75	1	0,75	1	1	1
Personne 10	EDB	4	3	3	3	3	3	3
	GD	4	3	3	2	2	3	3
	FD	2	0	0	0	0	0	0

	Rappel	1	1	1	0,66	0,66	1	1
	Précision	0,66	1	1	1	1	1	1

Tableau 5.7. Taux de reconnaissance des émotions pour la base Jaffe

3. Expérimentation du deuxième système :

	Nombre de séquences	**Pseudo moments de Zernike**		**EAR-LBP multi-échelle**		**Fusion des deux méthodes**	
		Nombre rec.	Taux	Nombre rec.	Taux	Nombre rec.	Taux
Personne 1	655	655	100%	655	100%	655	100%
Personne 2	776	776	100%	776	100%	776	100%
Personne 3	958	768	80,16%	796	83,08%	902	94,15 %
Personne 4	1193	1193	100 %	1193	100 %	1193	100 %
Personne 5	406	330	81,28 %	406	100 %	406	100 %
Personne 6	692	560	80,92%	621	89,73%	686	99,13 %
Personne 7	393	393	100 %	393	100 %	393	100 %
Personne 8	872	783	89.79%	785	90.02%	853	97.82%
Personne 9	701	701	100 %	701	100 %	701	100 %
Personne 10	425	425	100 %	425	100 %	425	100 %

Tableau 5.8. Taux de reconnaissance des visages en utilisant, les moments de Zernike, MS-EAR-LBP, et la méthode de fusion.

Les résultats promulgués dans Tableau 5.8 sont montrés dans le schéma de la figure 5.1 ci-dessous :

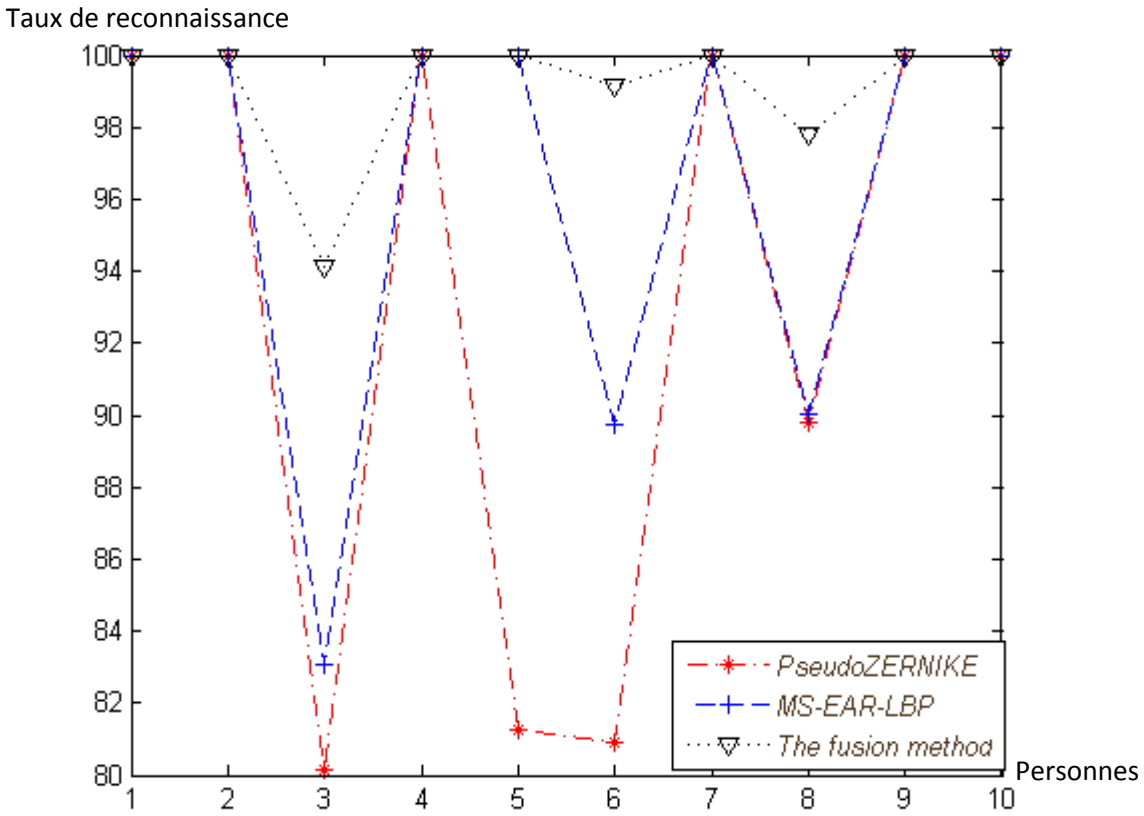

Figure 5.1. Résultat de la reconnaissance des visages

Le Tableau 5.9 nous donne les résultats des expérimentations pour la reconnaissance de l'émotion. Il s'agit de séquences vidéo de cinq personnes prises de la base FABO.

		Peur	Colère	Joie	Triste	Surpris	Dégout	Neutre
Personne 1	EDB	0	0	30	20	20	30	0
	GD	0	0	30	20	20	20	0
	FD	0	0	10	0	0	0	0
	Rappel	–	–	1	1	1	0,67	–
	Précision	–	–	0,75	1	1	1	–
Personne 2	EDB	0	30	30	0	0	30	10
	GD	0	30	12	0	0	24	10
	FD	0	24	0	0	0	0	0
	Rappel	–	1	0,4	–	–	0,8	1
	Précision	–	0,55	1	–	–	1	1
Personne 3	EDB	0	20	20	20	0	20	20
	GD	0	20	13	20	0	11	20
	FD	0	10	0	0	0	6	0
	Rappel	–	1	0,65	1	–	0,55	1
	Précision	–	0,66	1	1	–	0,64	1
Personne 4	EDB	0	0	30	0	0	20	50
	GD	0	0	30	0	0	15	40
	FD	0	0	15	0	0	0	0
	Rappel	–	–	1	–	–	0,75	0,80
	Précision	–	–	0,66	–	–	1	1
Personne 5	EDB	0	20	20	10	10	0	40
	GD	0	10	20	10	9	0	30
	FD	0	0	20	1	0	0	0
	Rappel	–	0,5	1	1	0,9	–	0,75

	Précision	–	1	0,5	0,5	0.9	–	1

Tableau 5.9. Résultats de reconnaissance de l'émotion sur la Base FABO.

4. Comparaison avec les travaux existants

A titre de comparaison avec des travaux récents, nous avons choisi celui de (Hoai & Nguyen) et (Taskeed, Hasanul, & Oksam, 2010) pour la base de données JAFFE et l'autre de (Shizhi, YingLi, Qingshan, & Dimitris, 2011) pour la base de données FABO. Les résumés des résultats obtenus sont présentés respectivement dans le tableau 5.10, la figure 5.2, tableau 5.11, et la figure 5.3. Les tests sur la base de données FABO dans (Shizhi, YingLi, Qingshan, & Dimitris, 2011) ne considèrent pas l'état neutre.

Emotion	**Temporal normalization %**	**Bag of words**	**Notre méthode %**
Colère	92	92	84,7
Dégout	82	36	84
Peur	43	43	86,1
Joie	94	61	89
Tristesse	17	58	81,4
Surprise	25	38	85,8

Tableau 5.10. Comparaison avec des travaux utilisant la base FABO

Emotion	**Canny, PCA, et ANN %**	**Local Directional Pattern %**	**Notre méthode %**
Colère	90	94,3	89,9
Dégout	90	80,1	85,7
Peur	80	86,3	87,3
Joie	80	95,2	96,6
Tristesse	90	77,5	80,6
Surprise	90	89,6	86,4
Neutre	80	84,5	83,1

Tableau 5.11. Comparaison avec des travaux utilisant la base JAFFE

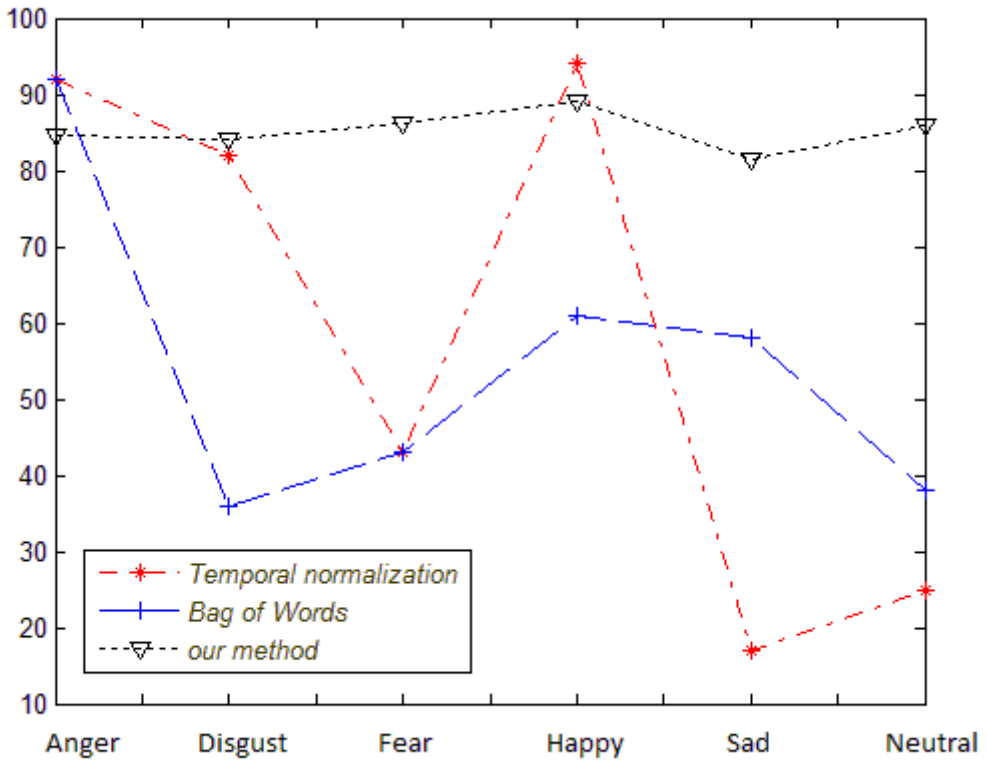

Figure 5.2. Comparaison avec des travaux utilisant la base FABO

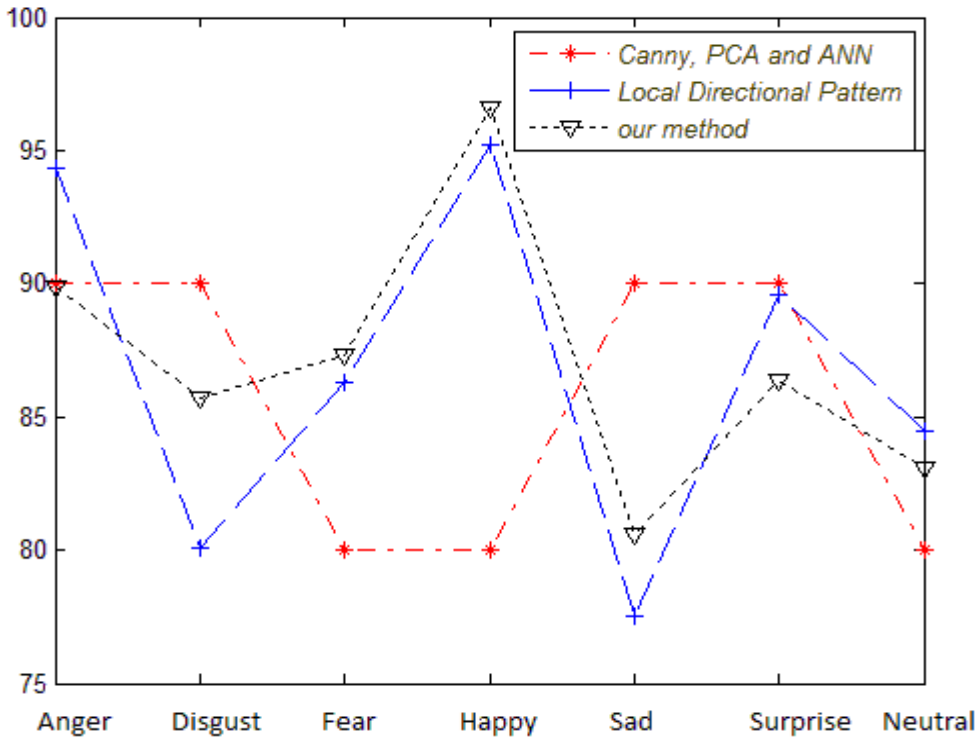

Figure 5.3 : comparaison avec des travaux utilisant la base JAFFE

Conclusion:

Nous avons montré dans ce chapitre les étapes de validation de toutes les phases du système, l'efficacité est mesurée ainsi en termes de rappel et précision. Par la suite nous avons mené une comparaison avec des travaux récents, et les résultats obtenus montrent que notre approche se classe au même rang et parfois mieux que des travaux existants.

Conclusion générale et perspectives:

Les nouvelles technologies sont présentes dans notre vie quotidienne, elles ne fournissent pas une interface adéquate qui les rend plus abordables pour les utilisateurs. Par conséquent, l'informatique affective qui améliore l'interaction homme-ordinateur, permet aux ordinateurs d'être plus adaptés à l'homme et non pas l'inverse. Dans ce contexte notre recherche se concentre sur la reconnaissance des émotions à partir d'expressions faciales afin de fournir un système cognitif d'analyse de l'émotion.

Ce travail présente deux systèmes de contrôle appliqués respectivement à la tablette intelligente et le téléviseur intelligent. Ils reposent essentiellement sur la détection et la reconnaissance des visages ainsi que leurs émotions. Pour ce faire, deux espaces de caractéristiques sont fusionnées, à savoir, des moments de Zernike et EAR-LBP. Ensuite, le premier système utilise une technique de sélection de paramètres pertinents pour réduire la taille de l'espace des primitives tout en conservant seulement ceux les plus pertinents. Tandis que le deuxième système procède par une réduction de données en se basant sur les mémoires auto-associatives. Les tests sur chaque système sont développés sur les bases de données Fabo et Jaffe, et les résultats obtenus sont significatifs. Entre autres, le module de prétraitement améliore considérablement la robustesse de chaque système vis-à-vis des différents changements d'illumination, mais toujours reste-il impertinent devant les fortes orientations de forte tête.

Le travail d'avenir consiste en l'amélioration du module de prétraitement et de paramétrage afin de remédier aux problèmes qui peuvent aussi survenir quand une personne porte des lunettes, barbue ou avec des moustaches. Pour ce faire, nous proposons l'intégration de filtres qui sont destinés à combler ces lacunes, si nécessaire, nous utilisons des systèmes permettant de raser la barbe et / ou enlever les lunettes.

Bibliographie

L'Empire caché de nos Emotions. (2005, septembre). *Science & Vie Hors-Série n° 232*. Paris , France: Excelsior Publications.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences : A meta-analysis. *Psychological Bulletin*, 256-274.

Aouatif, A., Rziza, M., & Driss, A. (2008). SVM-Based Face Recognition Using Genetic Search for Frequency-Feature Subset Selection. *Image and Signal Processing Lecture Notes in Computer Science*, 321-328.

Beek, P. J., Reinders, M. J., Sankur, B., & Lubbe., J. C. (1992). Semantic segmentation of videophone image sequences., (pp. 1182-1193.).

Belhumeur, P., Hespanha, J., & Kriegman., D. (1997). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* .

Black, & Yacoob, Y. (1995). Tracking and recognizing rigid and non-rigid facial motions using local parametric model of image motion. *Proceedings of the International Conference on Computer Vision* (pp. 374-381). IEEEComputer Society.

Brunnelli, R., & Poggio., T. (1993). Faces Recognition: Features versus Templates. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1042-1052.

Burl, M. C., & Perona, P. (1996). Recognition of planar object classes. *Int. Conf. on Computer Vision and Pattern Recognition.* IEEE.

Burl, M. C., Leung, T. K., & Perona, P. (1995). Face localization via shape statistics. *Int. Workshop on Automatic Face and Gesture Recognition.* Zurich.

Canny, J. (s.d.). A Computational Approach to Edge Detection. *Pattern Analysis.* IEEE Trans.

Chanel, G., Kronegg, J., Grandjean, D., & Pun, T. (2006). Emotion assessment. Arousal evaluation using EEG's and peripheral physiological signals. *Lecture Notes in Computer Science*.

Chen, S., Liu, J., & Zhou, Z.-H. (2004). Making FLDA applicable to face recognition with one sample per person. *Pattern Recognition*, 1553-1555.

Choi, J., Kim, S., & Rhee, P. (1999.). Facial components segmentation for extracting facial feature. *Second International Conference on Audio- and Video-based Biometric Person Authentication* .

Clavel, C. (2007). *Analyse et reconnaissance des manifestations acoustiques des émotions de type peur en situations anormales.* These en signal et images, l'Ecole Nationale Superieure des Telecommunications.

Colmenarez, A. J., & Huang., T. S. (1997). Face detection with information-based maximum discrimination. *Int. Conf. on Computer Vision and Pattern Recognition.* IEEE.

Cootes, T. F., & Taylor., C. J. (1992). Active shape models - ‘smart snakes'. *British Machine Vision Conference*, (pp. 266–275.).

Cootes, T., Taylor, C., Cooper, D., & Graham, J. (1995). Active shape models—their training and application. *Comput. Vis. Image Understand*, 18-23.

Cowie,]. R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., et al. (2001). Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 32-80.

Crowley, J. L., & Berard, F. (1997). Multi-model tracking of faces for video communications. *Int,Conf. on Computer Vision and Pattern Recognition.* Puerto Rico: IEEE.

Damasio, A. (1994). L'erreur de Descartes: la raison des émotions. *Odile Jacob.* Paris.

Darwin, C. (1872). *The expression of the emotions in man and animals.*

Duda, R. O., & Hart, P. E. (1973). Pattern Classification and Scene Analysis. Wiley, New York.

Ekman, P. (1999). *Basic emotions. Handbook of cognition and emotion.*

Ekman, P., Friesen, W., & Hager, J. (s.d.). *Facial Action Coding System : the manual.*

Ekman, P., W.Friesen, & Ellsworth, P. (1972). Emotion in the human face. *Pergamon Press.* New York.

Etemad, K., & Chellappa., R. (1996). Face recognition using discriminant eigenvectors. *Acoustics, Speech, and Signal Processing, ICASSP-96.*

Fei, L., Jinsong, H., Xueyi, Y., Zhenquan, Z., & Bin, L. (2006). Discriminant Independent Component Analysis as a subspace representation. *Journal of Electronics*.

Feng, Q., Pan, J., & Yan., L. (2012). Restricted Nearest Feature Line with Ellipse for Face Recognition. *Journal of Information Hiding and Multimedia Signal Processing*.

Frank, Y., & Chao-Fa, C. (2004). Automatic extraction of head and face boundaries and facial features. *Information Sciences*, 117-130.

Godefroid, J. (2008). *Psychologie : Science humaine et science cognitive.* de boeck.

Graf, H. P., Cosatto, E., Gibson, D., Petajan, E., & Kocheisen, M. (1996). Multi-modal system for locating heads and faces. *2nd Int. Conf. on Automatic Face and Gesture Recognition* (pp. 277-282). Vermont: IEEE Proc.

Gratch, J., & Marsella, S. (2001). Tears and fears : modeling emotions and emotional behaviors in synthetic agents. *fifth international conference on Autonomous agents*, (pp. 278–285).

H, K. S., & Wallbott, G. (1985). Analysis of nonverbal behavior . *Handbook of discourse analysis,*, pp. 199-230.

Haddadnia, J., Ahmadi, M., & Faez, K. (2003). An efficient feature extraction method with pseudo-Zernike moment in RBF neural network-based human face recognition system. *EURASIP journal on applied signal processing*, 890-901.

Heisele, B., Serre, T., Pontil, M., & Poggio, T. (2001). Component-based face detection. . *IEEE Conference on Computer Vision and Pattern Recognition.*

Herpers, R., Kattner, H., Rodax, H., & Sommer, G. (1995). Gaze: An attentive processing strategy to detect and analyze the prominent facial regions. *Int. Workshop on Automatic Face- and Gesture-Recognition* (pp. 214-220). Zurich: IEEE.

Hjelmas., E. (2001). Face Detection: A Survey. *Computer Vision and Image Understanding*, 236-274.

Hoai, B. L., & Nguyen, A. T. (s.d.). Using Rough Set in Feature Selection and Reduction in Face Recognition Problem Advances in Knowledge Discovery and Data Mining . *Lecture Notes in Computer Science*, 226-233.

Hong, P., Siyu, X., Lizuo, J., & Liangzheng, X. (2011). Efficient Face Recognition Fusing Dy- namic Morphological Quotient Image with Local Binary Pattern. *IWANN, Part II* , 228-235.

Hsu, R. L., Abdel-Mottaleb, M., & Jain, A. K. (2002). Face Detection in Colour Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 696–706.

http://neuroph.sourceforge.net. (s.d.).

Hua, Y., Yihua, C., Hang, S., & Yawen, F. (2012). The large-scale crowd analysis based on sparse spatial-temporal local binary pattern. *Mutimed Tools Appl*.

James, W. (1984). What is emotion? *Mind*, 188-205.

Jebara, T. S., & Pentland., A. (1997). Parametrized structure from motion for 3D adaptive feedback tracking of faces. *Int. Conf. on Computer Vision and Pattern Recognition.* Puerto Rico: IEEE.

Jen-Da, S., & Shyi-Ming, C. (2007). Feature subset selection based on fuzzy entropy measures for handling classification problems. *Appl. Intell.*, 69-82.

Jeng, S. H., Liao, H. Y., Han, C. C., Chern, M. Y., & Liu, Y. T. (1998). Facial feature detection using geometrical face model: An efficient approach,. *Pattern Recog.*

Jingru, W. M., & W.Y., G. (1999). Knowledge-Based Edge Detection and Feature Extraction of Human-Face Organs. *Pattern Recognition and Artificial Intelligence*, 340-346.

Karpouzis, K., Raouzaiou, A., & Kollias, S. (2003). Moving'avatars : emotion synthesis in virtual worlds. Human-Computer Interaction . *Theory and Practice*, 503-507.

Kass, M., Witkin, A., & Terzopoulos, D. (1987). Snakes: active contour models. *1st Int Conf. On Computer Vision.* London.

Kendall., D. G. (1989). A survey of the statistical theory of shape. *Statistical Science*, 87-120.

Kepenekci, B., Tek, F., & Akar, G. B. (2002). Occluded face recognition based on Gabor wavelets. *ICIP.*

Kim, K., Bang, S., & Kim, S. (2006). Emotion recognition system using short-term monitoring of physiological signals. *Medical and biological engineering and computing*, 419-427.

Kirby,]. M., & Srovich., L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. . *IEEE Trans. Patt. Anal.Mach. Intell.*

Kirsch, D. (1997). *The Sentic Mouse : Developing a tool for measuring emotional valence.* MIT Media Laboratory Perceptual Computing Section.

Klein, J., Moon, Y., & Picard, R. (2002). This computer responds to user frustration : Theory, design, and results. *Interacting with computers*, 119-140.

Lanitis, A., Taylor, C. J., & Cootes, T. F. (1994). . Automatic tracking, coding and reconstruction of human faces, using flexible appearance models. *IEEE Electron. Lett.*, 1578-1579.

Lanitis, A., Taylor, C., & Cootes, T. (1995). An Automatic Face Identification System Using Flexible Appearance Models. *Image and Vision Computing*, 393-401.

Le, H. T., Nguyen, D. T., & Tran, S. H. (2011). A Facial Expression Classification System Integrating Canny Principal Component Analysis and Artificial Neural Network. *International Journal of Machine Learning and Computing*.

Leonardis, A., & Bischof, H. (2000). Robust Recognition Using Eigen images and Image Understanding. *Computer Vision*, 99-118.

Leymarie, F., & Levine., M. D. (1993,). Tracking deformable objects in the plane using an active countour model. *IEEE transaction of pattern analysis and machine intelligence*, 617-634.

Li, S. Z., & Lu., J. (1999). Face Recognition Using the Nearest Feature Line Method. *IEEE TRANSACTIONS ON Neural Networks*.

Lin, C. C., & Lin, W. C. (1996). Extracting facial features by an inhibitory mechanism based on gradient distributions. *Pattern Recog.* , 2079–2101. .

Low B. K. Computer Extraction of Human Faces, P. t. (1998). *Computer Extraction of Human Faces, PhD thesis, Dept. of Electronic and Electrical Engineering, De Montfort University, 1998.* PhD thesis, Dept. of Electronic and Electrical Engineering, De Montfort University.

Low, B. K., & Ibrahim, M. K. (1997). A fast and accurate algorithm for facial feature segmentation,. *International Conference on Image Processing.*

Luthon, F., & Lievin, M. (1997). Lip motion automatic detection. *Scandinavian Conference on Image Analysis.*

Maio, D., & Maltoni, D. (2000). Real-time face location on gray-scale static images . *Pattern Recognition*, 1525-1539.

McKenna, S., Gong, S., & Liddell, H. (1995). Real-time tracking for an integrated face recognition system . *2nd Workshop on Parallel Modelling of Neural Operators.* Faro, Portugal.

Minsky, M. (1985). *The Society of Mind. .* Simon and Schuster editors.

Moghaddam, B., & Pentland, A. (1997). Probabilistic visual learning for object representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. *PAMI*, 696-710.

Moghaddam, B., Wahid, W., & Pentland, A. (1998). Beyond Eigenfaces: Probabilistic Matching for Face Recognition. *The 3rd IEEE Int'l Conference on Automatic Face & Gesture Recognition.* IEEE.

Morris, D., Friedhoff, H., & Dubois, Y. (1978). *La clé des gestes.*

Nefian, A., & Hayes, M. H. (1998). Face detection and recognition using hidden Markov models. *Image Processing, ICIP.*

Neji, M., Benammar, M., Wali, A., & Alimi, A. M. (2013). Towards an intelligent information research system based on the human behavior: Recognition of user emotional state. *International Conference on Computer and Information Science*, (pp. 371-376).

Ojala, T., Pietikainen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition* , 51-59.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 971–987.

Ortony, A., & Turner, T. (1990). What's basic about basic emotions. *Psycological review*.

Osuna, E., Freund, R., & girosi., F. (1997). Training support vector machines: an application to face detection. *CVPR.*

Padgett, C., & Cottrell, G. (1996). Representing face images for emotion recognition. *Advances in Neural Information Processing Systems, NIPS.*

Pantic, M., Sebe, N., Cohn, J. F., & Huang, T. (2005). Affective multimodal human-computer interaction. . *In MULTIMEDIA '05 : Proceedings of the 13th annual ACM international conference on Multimedia*, (pp. 669–676).

Park, J., Oh, Y., Ahn, S., & Lee., S. (2005). Glasses Removal from Facial Image Using Recursive Error Compensation. *IEEE Transactions On Pattern Analysis and Machine Intelligence*.

Parviainen, E., & Bottleneck, D. (2010). Classifiers in Supervised Dimension Reduction, Artificial Neural Networks, ICANN, 6354, 2010, 1-10. . *Artificial Neural Networks, ICANN*, 1-10.

Paul, J., & Ron., M. (1996). A hierarchical neural network for human face detection,. *Pattern recognition*, 781-787.

Peer, P., & Solina, F. (1999). An Automatic Human Face Detection Method. *the 4th Computer Vision Winter Workshop (CVWW'99)*, (pp. 122-130).

Pei-zhi, C., & Shui-li, C. (2010). A New Face Recognition Algorithm Based on DCT and LBP. *Advances in Intelligent and Soft Computing*, 811-818.

Penev, P., & Atick, J. (1996). Local feature analysis: A general statistical theory for objecct representation. *Netw.: Computat. Neural Syst.*, 477-500.

Pentland, A., & Choudhury., T. (2000). Face recognition for smart environments. *IEEE Computer*, 50-55.

Pentland, A., Moghaddam, B., & Starne., T. (1994). View-Based and Modular Eigenspaces for face Recognition. *IEEE Conference on Computer Vision & Pattern Recognition.* IEEE.

Pentland, M. T. (1991). Eigenfaces for recognition. *J. Cog. Neurosci.*, 71-86.

Picard,]. R., & Rosalind, W. (2000). Toward agents that recognize emotion. *VIVEKBOMBAY*, 3-13.

Picard, R. (1997). *Affective computing.* MIT press.

Picard, R. (2001). Building HAL: Computers that sense, recognize, and respond to human emotion. *Conference on Human Vision and Electronic Imaging*, (pp. 518-523).

Picard, R. W. (1999.). *Affective Computing for HCI.* HCI.

Plutchik, R. (1980). *A general psychoevolutionary theory of emotion.*

Pontil, M., & Verri., A. (1998). Support vector machines for 3-d object recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 637-646.

Price, J., R.Jeffery, Gee, T., & Timothy, F. (2005). Face recognition using direct, weighted linear discriminant analysis and modular subspaces. *Pattern recognition*, (pp. 209-219).

Reisfeld, D., & Yeshurun, Y. (1998). Preprocessing of face images: Detection of features and pose normalization. *Comput. Vision Image Understanding*.

Reyes, J., Vellasco, M., & Tanscheit, R. (2013). Fault detection and measurements correction for multiple sensors using a modified auto-associative neural network. *Neural Comput & Applic*.

Roobaert, D., Nillius, P., & Eklundh, J. (2000). Comparison of learning approaches to appearance-based 3d object recognition with and without cluttered background. *ACCV.*

Rowley, H. A., & S. Baluja, T. K. (1998). Rotation invariant neural network-based face detection. *Intl. Conf. on Computer Vision and Pattern Recognition* (pp. 38-44). IEEE.

Rowley, H. A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. *Pattern Anal.Mach Intell.*, 23-38.

Russell, J. (1980). A circumplex model of affect. *Journal of personality and social Psychology*.

Satyanadh, G., & Vijayan, A. (2007). Selection for Improved Face Recognition in Multisensor Images. *Signals and Communication Technology*, 109-120.

Scherer, K. (2004). Feelings integrate the central representation of appraisal-driven response organization in emotion. *In Feelings and emotions : The Amsterdam symposium*, (pp. 136-157). Amsterdam.

Scherer, K., Banse, R., Wallbot, H., t, & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion* , 123-148.

Schneiderman, H., & Kanade, T. (1998). Probabilistic modeling of local appearance and spatial relationships for object recognition. *IEEE Conference on Computer Vision and Pattern Recognition.* IEEE.

Schneiderman, H., & Kanade, T. (2000). A statistical model for 3D object detection applied to faces and cars. *IEEE Conference on Computer Vision and Pattern Recognition.* IEEE.

Scholkopf, B., Smola, A. J., & Muller, K. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 1299-1319.

Schubert, A. (2000). Detection and tracking of facial features in real time using a synergistic approach of spatiotemporal models and generalized hough-transform techniques. *Fourth IEEE InternationalConference on Automatic Face and Gesture Recognition.*

Shizhi, C., YingLi, T., Qingshan, L., & Dimitris, N. M. (2011). Recognizing Expressions from Face and Body Gesture by Temporal Normalized Motion and Appearance Features. *Computer Vision and Pattern Recognition Workshops (CVPRW).*

Sirovic, L., & Kirby, M. (1987). Low-dimensional procedure for the caracterization of human faces. *Journal of optical society of America*.

Sobotka, K., & Pitas, L. (1996). Segmentation and tracking of faces in color images. *int. Conférence on Automatic face & gesture recognition.*

Sobottka, K., & Pitas, L. (1996). Extraction of facial regions and features using color and shape information. *int. Conf. On pattern recognition (ICPR)* , (pp. 421-425). Vienna.

Solina, F., Peer, P., Batagelj, B., & Juvan, S. (2002). 15 seconds of fame - an interactive, computer-vision based art installation. *ICARCV*, (pp. 198-204).

Sung, K. K. (1996). *Learning and Example Selection for Object and Pattern Detection.* Massachusetts: Massachusetts Institute of Technology.

Sung, K. K., & Poggio, T. (1998). Example-based learning for view-based human face detection. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 39-51.

Taskeed, J., Hasanul, K. M., & Oksam, C. (2010). Robust Facial Expression Recognition Based on Local Directional Pattern. *ETRI Journal*.

Tremblay, A., Deschênes, J., Poulin, B., Roy, J., Kirouac, G., & Kappas, A. (1993). Identification de régions faciales critiques dans le jugement d'authenticité d'un sourire d'expressions de joie véritablement ressentie. *XVIe Congrès annuel de la SQRP.*

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition, J. Cog. Neurosci. 3, 1991, 71-86. *J. Cog. Neurosci*, 71-86.

Valat, J. (2008). *Le comportement emotionnel.* Montpellier: Licence de Psychologie, Université Montpellier II.

Venugopal, K. R., & Patnaik, L. M. (2012). Automatic Facial Expression Recognition Using Extended AR-LBP. *ICIP CCIS*, 244-252.

Vincent, L. (1993). Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE trans. Image processing*, 176-201.

Vyzas, E., & Picard, R. W. (1999). Offline and Online Recognition of Emotion Expression from Physiological Data. In Emotion-Based Agent Architectures Workshop Notes. *International Conference on Autonomous Agents*, (pp. 135-142).

Wallbott, H. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, 879–896.

Wiskott, L., Fellous, R., Kruger, N., & Malsburg, C. V. (1997). Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell*, 775–779.

Yang, G., & Huang, T. S. (1994). Human face detection in a complex background. *Pattern Recog*, 53–63.

Yang, M. (2002). Kernel Eigenfaces Vs Kernel Fisherfaces : Face recognition Using Kernel methods. *fifth international conference on automatic face and gesture recognition*, (pp. 215-220).

Yazhou, L., Hongxun, Y., Wen, G., & Debin, Z. (2005). Illumination Invariant Feature Selection for Face Recognition, Advances in Multimedia Information Processing - PCM. *Lecture Notes in Computer Science*, 946-957.

Yuille, A. L., Hallinan, P. W., & Cohen, D. S. (1992). Feature extraction from faces using deformable templates. *Int. J. Comput. Vision*, 99-111.

Yu-Tzu, L., Ruei-Yan, L., L.Yu-Chih, & C.Greg. (2012). Real-time eye-gaze estimation using a low-resolution webcam. *Multimedia Tools Appl*, 543-568.

Yu-Tzu, L., Ruei-Yan, L., L.Yu-Chih, & C.Greg. (2012). Real-time eye-gaze estimation using a low-resolution webcam. *Multimedia Tools Appl.*, 543-568.

Zhang, J., Yan, Y., & Lades., M. (1997). Face recognition: eigenface, elastic matching and neural nets. *Proc. IEEE*, 1423-1435.

Zhiping, S., L. Xi, L. Q., & Qing, H. (2012). Extracting discriminative features for CBIR. *Multimedia Tools Appl*, 263-279.

Zhou, N., & Lipo, W. (2009). Class-Dependant Feature Selection for Face Recognition, Ad-vances in Neuro-Information Processing. *Lecture Notes in Computer Science*, 551-558.