

# Analyse dynamique des transformations des traits du visage lors de la production d'une émotion

## Facial features boundaries segmentation and facial expressions analysis

Zakia Hammal, Alice Caplier

Laboratoire des Images et des Signaux,  
46 avenue Félix Viallet, 38031 Grenoble, France  
alice.caplier@inpg.fr

### Résumé

L'objectif de notre travail est l'étude dynamique des émotions faciales. Puisque certaines expériences en psychologie ont mis en évidence le fait que l'être humain est capable de reconnaître les émotions sur un visage à partir de la visualisation de l'évolution temporelle de ses certains points caractéristiques, nous travaillons à l'extraction automatique des contours des traits du visage.

Dans un travail précédent, nous avons développé un algorithme de segmentation automatique du contour des lèvres. Dans cet article, on s'intéresse au problème de la segmentation des yeux et des sourcils. L'ensemble des deux algorithmes conduit à l'extraction de ce que nous appelons des "squelettes d'émotions".

L'algorithme proposé pour l'extraction des contours des yeux et des sourcils met en jeu une segmentation statique lors de la phase d'initialisation et une segmentation dynamique pour les images suivantes. Dans les deux cas, la segmentation est constituée de trois étapes : d'abord, on définit un modèle paramétrique pour modéliser au mieux le contour de chaque trait considéré; ensuite, les modèles choisis sont initialisés sur les images à segmenter grâce à l'extraction d'un ensemble de points caractéristiques ; enfin, les modèles initiaux sont ajustés finement en tenant compte d'information de gradient de luminance. Des séquences de squelettes constitués des contours extraits sont construites pour chaque type d'émotion. Plusieurs distances caractéristiques compatibles avec la description MPEG-4 des six émotions universelles sont définies sur chaque squelette. Nous montrons que l'analyse temporelle de l'évolution de ces distances peut conduire à la mise au point d'un système automatique de reconnaissance émotionnelle

### Abstract

The aim of our work is the dynamic analysis of facial emotions. Since some experiences in psychology have

shown that humans are able to recognize facial expressions by the analysis of the temporal evolution of keypoints of facial features, we are working on facial features extraction.

In a previous work, we have developed a robust and fast algorithm for accurate lip contour segmentation. In this paper, we are focusing on the problem of eyes and eyebrows boundaries extraction. Both algorithms yields to the extraction of what we call "emotion skeleton".

The proposed algorithm for eyes and eyebrows segmentation involves a static segmentation for the initialization and a dynamic segmentation for the following frames. In both cases, the segmentation requires three steps: first, an accurate model based on flexible curves is defined for each feature; second, models are initialized on the image to be processed after the detection of characteristic points; third, models are accurately fitted to the facial features according to some information of luminance gradient. Temporal sequences with the extracted boundaries (or skeletons) of facial features are build for each type of emotion. Several distances compliant with the MPEG-4 description of the six basic emotions are computed on each skeleton. We show that the temporal analysis of these distances will yield to the automatic classification of the six universal emotions.

### 1. Introduction

Récemment, de nombreux travaux ont porté sur la localisation automatique du visage et de ses traits caractéristiques. En effet, l'objectif est de développer des systèmes interactifs capable d'analyser et d'interpréter le comportement de l'utilisateur. Dans notre laboratoire, nous nous intéressons à l'extraction automatique des traits du visage que sont la bouche, les yeux et les sourcils. Nous avons déjà développé un algorithme robuste et rapide d'extraction du contour des lèvres [1] afin d'améliorer la perception d'un message en environnement bruité.

Dans cet article, nous nous intéressons à l'extraction automatique des contours des yeux et des sourcils. Les travaux existants souffrent de deux limitations principales : certains proposent une localisation globale grossière de ces traits par l'extraction d'une boîte englobante [2,3] ; d'autres essaient d'extraire plus précisément les contours mais les modèles choisis sont trop simples et peu réalistes [4] et les algorithmes nécessitent une phase de sélection manuelle de points dans la première image. La méthode proposée s'efforce de pallier ces problèmes afin d'obtenir des contours de suffisamment bonne qualité pour pouvoir être utilisés dans le cadre d'une reconnaissance d'expressions faciales.

Les systèmes de reconnaissance d'émotions travaillent pour la plupart à partir de l'analyse d'images statiques ce qui suppose d'être capables de localiser la « bonne » image à traiter ce qui est peu réaliste dans le cadre d'une application réelle. Des méthodes telles que l'ACP sont utilisées pour la classification [5, 6]. Une méthode proche de ce que nous proposons est décrite dans [7]. Les différences principales sont les suivantes :

- Nous proposons une méthode d'extraction automatique des contours et des points caractéristiques;
- Seule une analyse dynamique sur des séquences d'images est faite sachant qu'un système réaliste d'analyse d'expressions faciales devra faire une interprétation dynamique du comportement de l'utilisateur.
- Les critères d'évolution définis pour l'analyse des expressions faciales sont plus simples bien qu'ils soient compatibles avec la description des émotions donnée dans la norme MPEG-4.

Dans la partie 2, nous décrivons la phase de segmentation statique. La partie 3 présente la segmentation dynamique qui permet d'introduire une cohérence et une stabilité temporelles en tenant compte des informations issues de la segmentation de l'image précédente. La partie 4 définit la notion de squelettes d'expressions faciales ainsi que des distances particulières calculées sur ces squelettes. On montre que les distances choisies ont un rapport direct avec la description des six émotions universelles (joie, colère, dégoût, tristesse, peur et surprise) donnée dans la norme MPEG-4 et que l'analyse de l'évolution temporelle de ces distances peut conduire à un système de reconnaissance des expressions faciales.

## 2. Segmentation statique : initialisation

Le problème de l'extraction automatique du visage dans une image a été largement traité ces dernières

années [8,9]. Ceci n'est donc pas l'objet de ce papier. Ici nous supposons qu'il est possible d'extraire une boîte rectangulaire entourant le visage dans la première image de la séquence. Pour les images suivantes, la boîte englobant le visage est suivie par mise en correspondance.

### 2.1 Détection de l'iris

Le contour de l'iris représente la frontière entre le blanc de l'oeil et la zone sombre associée à l'iris. De ce fait, on recherche le contour de l'iris comme un cercle constitué de points de gradient de luminance maximum. Chaque cercle contour de l'iris maximise la quantité :  $E = \sum_{p \in C} \vec{\nabla} I(p) \cdot \vec{n}(p)$  où  $I$  est la luminance

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

La zone de recherche de chaque iris est limitée aux parties hautes gauche et droite de la boîte englobant le visage. Les dimensions de chaque boîte de recherche de l'iris ont été déterminées suite à une phase d'apprentissage. Sur chaque image de la base ORL [10], une boîte englobant chaque œil a été sélectionnée à la main et les dimensions respectives de ces boîtes ont été étudiées. On en a déduit les relations suivantes entre les dimensions de la boîte englobant le visage et celles de la boîte englobant chaque œil :

$$Hauteur_{visage} = 4 * Hauteur_{oeil} ; L arg eur_{visage} = 2.5 * L arg eur_{oeil}$$

### 2.2 Modèles paramétriques pour les yeux et les sourcils

Dans beaucoup de travaux, les yeux sont modélisés par deux paraboles. Une étude précise d'un certain nombre de cas montre qu'une parabole n'est pas la courbe la plus adaptée car elle induit une contrainte de symétrie verticale qui n'est pas toujours satisfaite ni pour les yeux ni pour les sourcils (voir figures 3 et 4). Nous proposons d'utiliser une courbe de Bézier à trois points de contrôle. La Figure 1 montre le modèle choisi pour chaque œil : pour le contour inférieur, une parabole passant par les trois points  $P_1, P_2, P_4$  est considérée et pour le contour supérieur, une courbe de Bézier à trois points de contrôle  $P_1, P_2, P_3$  est définie.

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

proposons de définir, comme modèle adapté et réaliste, une courbe de Bézier définie par les trois points de contrôle  $P_5, P_6, P_7$  (voir Figure 1).

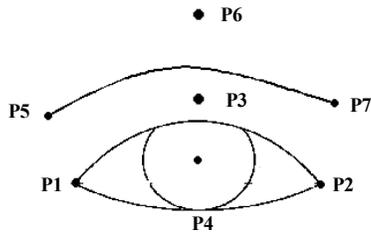


Figure 1 : modèle pour l'oeil et le sourcil et points caractéristiques  $P_i$

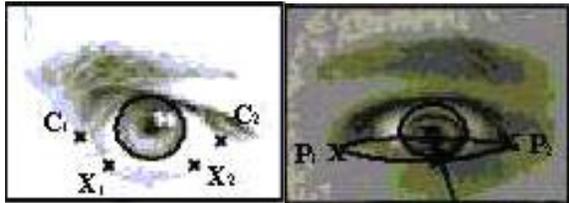


Figure 2: à gauche : détection des coins  $C_1$  et  $C_2$  des yeux; à droite : initialisation du modèle de l'œil.

### 2.3 Segmentation des yeux

Le modèle étant défini, ce dernier est placé sur l'image à traiter grâce à l'extraction automatique des points caractéristiques ( $P_1, P_2, P_3$  et  $P_4$ ) et par la déformation de ce modèle conformément à des informations de maximum de gradient de luminance. En effet, les contours des yeux sont à la limite entre le blanc des yeux (zone claire) et la peau (zone plus sombre).

Un processus de suivi de points de gradient de luminance maximum est utilisé pour détecter les coins des yeux. La partie gauche de la Figure 2 donne une illustration de la méthode : partant des points  $X_1$  et  $X_2$ , deux points de gradient de luminance maximum situés près de la verticale des limites du cercle détecté pour l'iris, un processus de suivi, vers la gauche, des points de gradient de luminance maximum conduit à la détection du coin  $C_1$ . Le même processus de suivi vers la droite conduit à la détection du coin  $C_2$ . Le chemin suivi entre  $X_1$  et  $C_1$  (resp.  $X_2$  et  $C_2$ ) ne contient que des points de gradient de luminance maximum.

Les points  $P_1$  et  $P_2$  du modèle sont placés sur les deux coins détectés  $C_1$  et  $C_2$ . Le point  $P_4$  de la parabole est aligné avec le point le plus bas du cercle détecté pour l'iris et le point  $P_3$  coïncides avec le centre de ce même cercle. La partie droite de la Figure 2 présente un exemple d'initialisation du modèle associé à l'œil.

L'idée pour la déformation du modèle initial reste la même : chaque contour est un ensemble de points de gradient de luminance maximum. La courbe

sélectionnée est celle qui maximise le flux du gradient de luminance à travers le contour.

Les points ( $P_1, P_2$  et  $P_4$ ) sont détectés avec suffisamment de précision pour conduire à une parabole qui n'a plus besoin d'être ajustée. A l'opposé, puisque le point de contrôle  $P_3$  de la courbe de Bézier est initialisée au niveau du centre de l'iris, on sait que cette courbe doit être déformée. En particulier, il faut que le point  $P_3$  se déplace vers le haut (les points  $P_1$  et  $P_2$  restant fixes) jusqu'à ce que le flux du gradient de luminance à travers le contour soit maximum.

### 2.4 Segmentation des sourcils

$P_5$  et  $P_7$  sont les coins de chaque sourcil. Les abscisses  $x_5$  et  $x_7$  de ces deux points correspondent aux maximums droit et gauche de la quantité:

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

et les ordonnées  $y_5$  et  $y_7$  correspondent aux maximums droit et gauche de la quantité :  $V(y) = \sum_{x=1}^{N_x} [255 - I(x, y)]$  où  $I(x, y)$  est la

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

Le troisième point de contrôle  $P_6$  est déduit des positions de  $P_5$  et  $P_7$  de la manière suivante :  $\{x_6 = (x_5 + x_7)/2; y_6 = y_7\}$ . De nombreux tests sur différentes images ont montré que la détection de ces points pouvaient conduire à des positionnements grossiers si bien que  $P_5$  et  $P_7$  doivent être ajustés lors de la phase d'ajustement du modèle aux données. Cet ajustement utilise la maximisation du flux de gradient de luminance, chaque point de contrôle étant déplacé tour à tour.

### 3. Segmentation dynamique

Afin d'obtenir une segmentation stable au cours du temps, l'extraction des contours dans l'image suivante tient compte des résultats obtenus lors du traitement de l'image précédente. Ceci permet également un traitement plus rapide.

Chaque point  $P_1, P_2, P_5$  et  $P_7$  est suivi d'image en image en utilisant l'algorithme proposé par Kanade dans [11]. Cette procédure de suivi fournit de très bons résultats pour les coins des yeux (qui ont des caractéristiques de luminance marquées) mais les points obtenus pour les coins des sourcils sont plus approximatifs. De ce fait, nous réalisons également une estimation de ces points par la méthode décrite au paragraphe 2.4. Les positions initiales des points  $P_5$  et

$P_7$  sont alors les positions moyennes entre les positions suivies et les positions estimées.

Le même processus de maximisation du flux de gradient est utilisé pour ajuster le modèle initial aux yeux et aux sourcils présents sur l'image à segmenter. Les Figures 3 et 4 présentent des résultats de segmentation. Dans le cas de la figure 3, les sourcils sont asymétriques, cette forme est bien rendue grâce au choix d'une modélisation par courbe de Bézier. De même, l'oeil droit de la figure 4 présente un contour supérieur asymétrique qui ne pourrait pas être correctement détecté par l'utilisation d'un modèle à base de paraboles seulement.

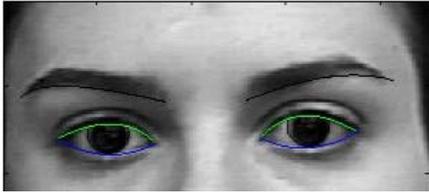


Figure 3 : extraction des contours (sourcils asymétriques)

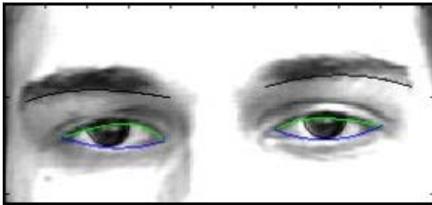


Figure 4 : extraction des contours (yeux asymétriques)

## 4. Analyse des squelettes d'émotion

### 4.1. Définition de la notion de squelette

Associé à notre algorithme de segmentation des contours de lèvres décrit dans [1], la méthode proposée permet de construire des squelettes pour chaque expression du visage. Ces squelettes sont constitués de l'ensemble des contours extraits pour les traits du visage. La Figure 5 montre des exemples de séquences de squelettes dans le cas des émotions joie et surprise. On constate que ces séquences sont typiques d'une émotion donnée : il est possible de reconnaître l'émotion lors de la visualisation de ces seuls squelettes.

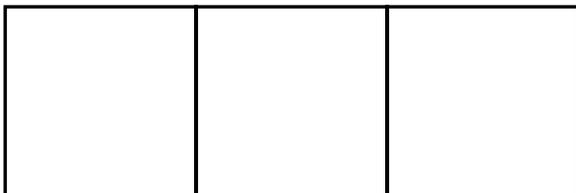
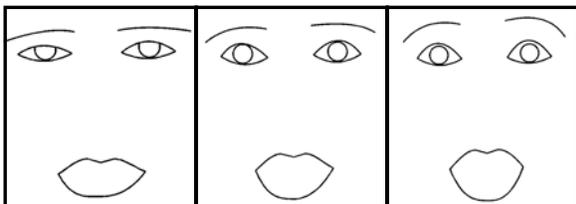


Figure 5 : séquences de squelettes : cas de la surprise (en haut) et de la joie (en bas).

### 4.2 Analyse compatible MPEG-4 des squelettes émotionnels

La norme MPEG-4 fournit une description des transformations subies par chacun des traits du visage lors de la production de chacune des six émotions universelles. Cette description est la suivante [12] :

- Joie : {la bouche s'ouvre}, {les commissures se retirent en arrière en direction des oreilles}, {les sourcils sont décontractés}
- Tristesse : {les coins intérieurs des sourcils se courbent vers le haut}, {les yeux se ferment légèrement}, {la bouche est décontractée}
- Colère : {les coins intérieurs des sourcils s'abaissent ensemble}, {les yeux s'ouvrent largement}, {les lèvres se serrent l'une contre l'autre ou bien elles s'ouvrent pour laisser apparaître les dents}
- Peur : {les sourcils se lèvent ensemble et leur partie intérieure est courbée vers le haut}, {les yeux sont contractés et en état d'alerte}
- Dégoût : {la lèvre supérieure se lève et se courbe souvent de manière asymétrique}, {les sourcils et les paupières sont décontractés}
- Surprise: {les sourcils se lèvent}, {les paupières supérieures s'ouvrent}, {la bouche s'ouvre}, {les paupières inférieures sont relâchées}

Afin de traduire numériquement toutes ces descriptions, nous avons défini un ensemble de distances particulières sur chaque squelette : la Figure 6 décrit toutes les distances  $D_i$  considérées. Chaque distance  $D_i$  a été normalisée par rapport à sa valeur sur le visage lorsqu' aucune émotion n'est présente.

$D_2$  et  $D_7$  donnent une mesure de la distance entre les yeux et les sourcils.  $D_6$  mesure la distance entre les yeux et la bouche.  $D_3$  et  $D_4$  mesurent le degré d'ouverture de la bouche,  $D_1$  mesure le degré d'ouverture des yeux.

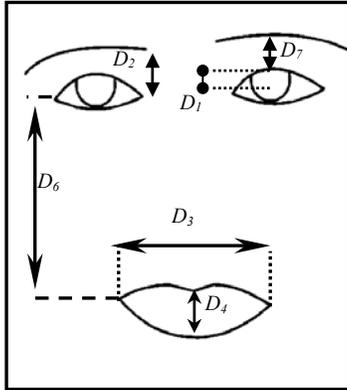


Figure 4 : définition des distances  $D_i$

Nous associons les critères suivants à chaque description d'émotions :

- Joie :  $\{D_4$  augmente $\}$ ,  $\{D_3$  diminue ET  $D_6$  diminue $\}$ ,  $\{\text{les autres distances restent constantes}\}$
- Tristesse :  $\{D_2$  augmente ET  $D_7$  diminue $\}$ ,  $\{D_1$  diminue $\}$ ,  $\{\text{les autres distances restent constantes}\}$
- Colère :  $\{D_2$  diminue $\}$ ,  $\{D_1$  augmente $\}$ ,  $\{D_4$  diminue ou  $D_4$  augmente $\}$
- Peur :  $\{D_2$  augmente ET  $D_7$  augmente mais plus que  $D_2$  $\}$ ,  $\{?\}$
- Dégoût :  $\{D_3$  augmente ET  $D_4$  augmente $\}$ ,  $\{\text{les autres distances restent constantes}\}$
- Surprise :  $\{D_2$  augmente $\}$ ,  $\{D_1$  augmente $\}$ ,  $\{D_4$  augmente $\}$ ,  $\{\text{les autres distances restent constantes}\}$

Pour chaque émotion, chaque expression entre accolade de la description précédente doit être mise en relation avec l'évolution d'une distance  $D_i$  particulière. Pour la peur, la condition  $\{?\}$  traduit le fait que nous ne savons pas comment traduire la caractéristique  $\{\text{les yeux sont contractés et en état d'alerte}\}$ .

Pour chacune des six émotions de base, nous avons étudié l'évolution des distances  $D_i$  pour 20 personnes différentes afin de valider les scénarios proposés. Nous avons utilisé une base de séquences d'images acquises au laboratoire. Pour chaque séquence, nous avons d'abord une expression neutre, puis une émotion particulière puis un retour vers le neutre. Chaque séquence dure environ 5 secondes (soit 125 images pour une cadence d'acquisition de 25 images par seconde).

La Figure 5 montre l'évolution des distances  $D_i$  moyenne dans le cas de la surprise. On constate qu'effectivement, les distances  $D_1$ ,  $D_2$  et  $D_4$  subissent une augmentation significative lorsque la surprise apparaît sur le visage. A l'opposé, les distances  $D_3$ ,  $D_6$  et  $D_7$  restent à peu près constantes.

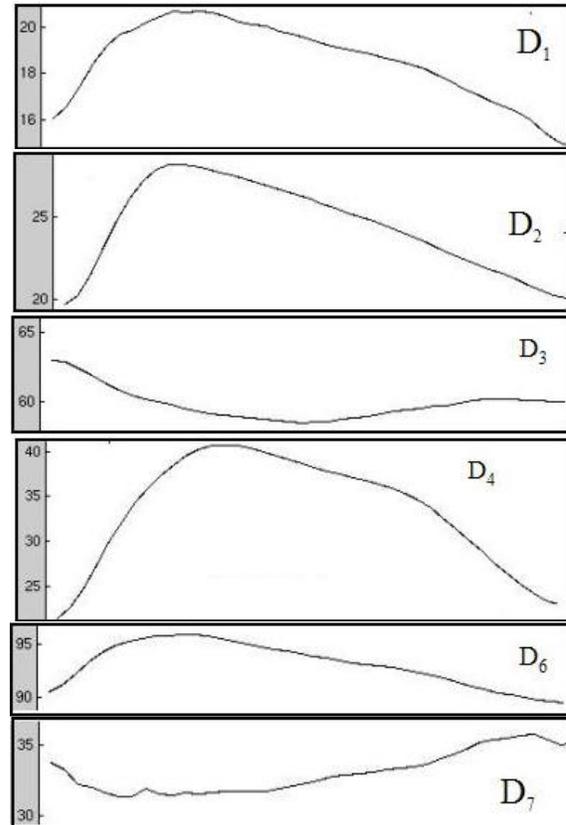
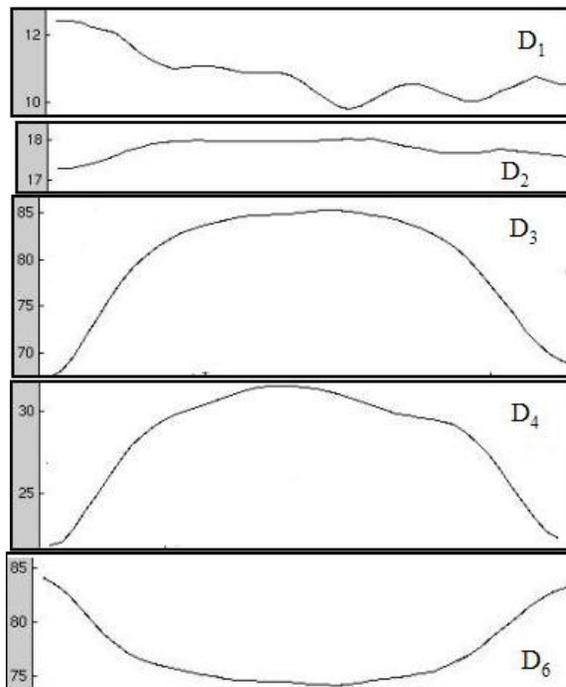


Figure 5 : courbes des  $D_i$  moyennes pour la surprise



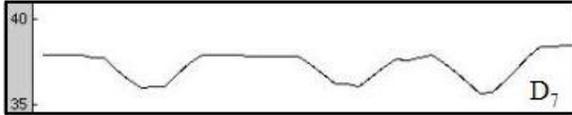


Figure 6 : courbes des  $D_i$  moyennes pour la joie.

La Figure 6 montre l'évolution des distances moyennes  $D_i$  dans le cas de la joie. Une évolution significative est détectée pour les distances  $D_3$ ,  $D_4$ , et  $D_6$ ; les autres distances restent à peu près constantes.

Si on compare les Figures 5 et 6, on constate que les émotions joie et surprise diffèrent par l'évolution des distances  $D_1$ ,  $D_2$ ,  $D_3$  et  $D_6$ . Ceci permet une distinction entre ces deux émotions.

### 4.3 Reconnaissance d'émotions

Le tableau 1 donne un résumé des scénarios d'évolution des distances  $D_i$  pour chacune des six émotions ("↗" signifie augmentation, "↘" signifie diminution et "=" traduit l'absence d'évolution). Remarquons que pour la peur, on ne fait aucune hypothèse sur l'évolution de  $D_1$  car nous ne savons pas comment traduire la condition {les yeux sont contractés et en état d'alerte}.

	$D_1$	$D_2$	$D_3$	$D_4$	$D_6$	$D_7$
joie	=	=	↗	↗	↘	=
tristesse	↘	↗	=	=	=	↘
colère	↗	↘	=	↗or↘	=	=
peur	?	↗	=	=	=	↗
dégoût	=	=	↗	↗	=	=
surprise	↗	↗	=	↗	=	=

Table 1 : évolution des  $D_i$  pour chaque émotion

L'analyse de ce tableau montre qu'il sera possible de différencier entre elles toutes ces émotions en s'intéressant en priorité aux distances  $D_i$  qui subissent des modifications significatives. En effet, il y a toujours au moins une évolution différente dans chaque scénario. Par exemple, la joie et le dégoût diffèrent par l'évolution de la distance  $D_6$ .

Un système de classification dynamique travaillant à partir de ces règles et utilisant la théorie de l'évidence est à l'heure actuelle à l'étude.

## 5. Conclusion

Nous avons présenté une méthode de segmentation des contours des yeux et des sourcils qui, associée à un algorithme d'extraction des contours des lèvres conduit à la construction de séquences de squelettes d'émotions. L'analyse de l'évolution temporelle de distances  $D_i$  calculées sur chaque squelette permet la définition pour chaque émotion d'un ensemble de caractéristiques compatibles MPEG-4. Bien que nous

soyons capables d'extraire les contours des traits du visage avec une bonne précision, aucune information de forme n'est prise en compte jusqu'ici dans la phase d'analyse et de reconnaissance des émotions. L'influence de la forme proprement dite est à l'étude à l'heure actuelle.

## 6. Bibliographie

- [1] N. Eveno, A. Caplier, and P.Y. Coulon "Automatic and Accurate Lip Tracking". To appear in *IEEE Trans. On Circuits and Systems for Video Technology*, May 2004.
- [2] M. Pantic, and L. Rothkrantz. "Automatic Analysis of Facial Expressions : the State of the Art". *IEEE PAMI*, Vol.22, N°12, pp. 1424-1445, December 2000.
- [3] A. Haro, M. Flickner, and I. Essa "Detecting and tracking eyes by using their physiological properties, dynamics and appearance". *Proc. IEEE CVPR*, South Carolina, June 2000.
- [4] Y. Tian, T. Kanade., and J. Cohn "Dual state Parametric Eye Tracking". *Proc. of the 4th IEEE Inter. Conf. on Automatic Face and Gesture Recognition*, March 2000, pp. 110 – 115.
- [5] I. Buciu, C. Kotropoulos, I. Pitas « ICA and Gabor representation for facial expression recognition » *IEEE Int. Conf. On Image Processing*, Spain, September 2003.
- [6] B. Abboud, F. Davoine, M. Dang « Statistical modeling for facial expression analysis and synthesis » *IEEE Int. Conf. On Image Processing*, Spain, September 2003.
- [7] A. Raouzaoui, N. Tsapatsoulis, K. Karpouzis, and S. Kollias "Parameterized Facial Expression Synthesis Based on MPEG-4". *Eurasip Journal on Applied Signal Processing*, 10, pp.1021-1038, 2002.
- [8] M.H. Yang, D. Kriegman, and N. Ahuja "Detecting face in images : a survey". *IEEE Trans on PAMI*, vol.24, n°1, pp. 34-58, January, 2002.
- [9] H. Hjelmas, B. Low "Face detection: a survey". *Computer Vision and Image Understanding*, 83, 2001, pp. 236-274.
- [10] ORL face database [http://www.uk.research.att.com/pub/data/att\\_faces.zip](http://www.uk.research.att.com/pub/data/att_faces.zip)
- [11] B. Lucas, and T. Kanade "An iterative image registration technique with an application to stereovision". *Proc. IJCAI'81*, Vancouver, 1981, pp.674-679.
- [12] M. Malciu « Approches orientées modèle pour la capture des mouvements du visage en vision par ordinateur ». Thèse de l'université René Descartes, Paris V, décembre 2001.