

Extraction des traits caractéristiques du visage à l'aide de modèles paramétriques adaptés

Parametric models for facial features segmentation

Zakia Hammal, Nicolas Eveno, Alice Caplier, Pierre-Yves Coulon

LIS, INPG, 46 avenue Félix Viallet, 38031 Grenoble Cedex
nom@lis.inpg.fr

Manuscrit reçu le 18 décembre 2003

Résumé et mots clés

Dans cet article, nous nous intéressons à l'extraction automatique des contours des traits permanents du visage à savoir : les yeux, les sourcils et les lèvres. Pour chacun des traits considérés, un modèle paramétrique spécifique capable de rendre compte de toutes les déformations possibles est défini. Lors de la phase d'initialisation, des points caractéristiques du visage sont extraits (coins des yeux et de la bouche par exemple) et servent de points d'ancrage initiaux pour chacun des modèles. Dans la phase d'évolution, chaque modèle est déformé afin de coïncider au mieux avec les contours des traits présents sur le visage analysé. Cette déformation se fait par maximisation d'un flux de gradient (de luminance et/ou de chrominance) le long des contours définis par chaque courbe du modèle. La définition de modèles permet d'introduire naturellement une contrainte de régularisation sur les contours recherchés. Néanmoins, les modèles choisis restent suffisamment flexibles pour permettre une extraction réaliste des contours des yeux, des sourcils et de la bouche. L'extraction précise des contours des principaux traits du visage constitue la première étape d'un ensemble d'applications multimédia.

Modèles paramétriques, traits du visage, segmentation.

Abstract and key words

In this paper, we are dealing with the problem of facial features segmentation (mouth, eyes and eyebrows). A specific parametric model is defined for each feature, each model being able to take into account all the possible deformations. In order to initialize each model, some characteristic points are extracted on each image to be processed (for example, the corners of the eyes, mouth and eyebrows). In order to fit the model with the contours to be extracted, a gradient flow (of luminance or chrominance) through the estimated contour is maximized because at each point of the searched contour, the gradient (of luminance or chrominance) is normal. The advantage of the definition of a model associated to each feature is to be able to introduce a regularisation constraint. However, the chosen models are flexible enough in order to produce realistic contours for the mouth, the eyes and eyebrows. This facial features segmentation is the first step of a set of multi-media applications.

Parametric models, facial features, segmentation.

1. Introduction

Les traits du visage participent à l'acte de communication entre êtres humains : par exemple, la visualisation des lèvres et de leur mouvement permet d'améliorer la compréhension d'un message en environnement bruité et elle permet même la communication avec des personnes malentendantes.

Néanmoins, il faut remarquer que la visualisation des traits du visage n'est un support à la communication que si ces traits sont extraits avec une précision suffisante. Dans le cas contraire, l'information issue d'une analyse trop sommaire constitue plus une gêne qu'une aide.

L'objet de notre travail est d'obtenir une extraction automatique et précise des contours des traits du visage en vue d'améliorer la communication homme/homme via une machine. Plusieurs applications sont envisagées :

- Le contour extérieur des lèvres est destiné à servir dans une application de type téléphonie mobile dans le but d'améliorer la compréhension d'un message lorsque l'environnement est bruité ;

- Ce même contour doit aussi être utilisé dans un projet de téléphonie à l'usage des malentendants qui nécessite la synthèse du signal de parole à partir de l'analyse de la forme et des mouvements des lèvres ;

- La détection du contour des iris est destinée d'une part à être utilisée dans une application de surveillance de niveau de vigilance ou niveau d'intérêt d'un utilisateur par analyse de la fréquence des clignements de ses yeux et d'autre part à être utilisée dans une application d'estimation de la direction du regard d'un utilisateur devant un écran de PC ;

- L'ensemble des traits détectés (yeux + sourcils + lèvres) doit servir d'information de base d'un système de reconnaissance automatique des émotions sur le visage d'une personne devant un PC et en situation de travail (seule cette application sera un peu plus précisément développée à la fin de cet article).

Les contraintes de précision des contours obtenus et de traitement temps réel sont communes à l'ensemble des applications envisagées.

Dans cet article, nous proposons de nouveaux modèles paramétriques spécifiques et adaptés à l'extraction des contours des yeux, des sourcils et des lèvres. Chaque modèle est initialisé suite à l'extraction de points caractéristiques sur le visage (coins des yeux et de la bouche par exemple) puis est déformé afin de maximiser un flux de gradient (de luminance et/ou de chrominance). La particularité de notre approche réside dans la définition de modèles adaptés aux déformations possibles de chaque trait ce qui conduit à l'extraction de contours réalistes.

Ces dernières années, de nombreuses méthodes ont été proposées pour résoudre le problème de la segmentation des lèvres. Certaines d'entre elles n'utilisent que des informations spatiales de bas niveau telles que la couleur et les contours [27]. Aucune contrainte de lissage, ni de forme n'est introduite si bien que la segmentation obtenue est souvent grossière. Dans [17], une *analyse linéaire discriminante* (LDA) est utilisée pour séparer

les pixels lèvres des pixels peau ce qui permet d'aboutir aux contours des lèvres. Bien que l'analyse discriminante soit suivie d'une opération de lissage, la segmentation obtenue est trop souvent bruitée.

Les *snakes* [15] ont été très utilisés en segmentation des lèvres ces dernières années ([1][19]) du fait de leur aptitude à prendre en compte dans un même formalisme des contraintes de lissage et d'élasticité. Les méthodes à base de *snakes* conduisent à des résultats intéressants mais malheureusement le réglage des nombreux paramètres associés est très difficile et le résultat de segmentation obtenu dépend fortement de la qualité de l'initialisation. Afin d'obtenir une segmentation plus fine et réaliste, il est possible d'utiliser un modèle de forme *a priori* ce qui conduit à un contour plus régulier. Une telle contrainte garantit que le contour extrait appartient à l'espace des formes considérées comme admissibles pour les lèvres. Par exemple, les modèles de forme actifs (ASM) ont été utilisés [16], mais dans ce cas, l'ensemble des images servant à la phase d'apprentissage doit être très grand afin de rendre compte de l'ensemble des déformations possibles pour la bouche. De plus, chaque image de la base d'apprentissage doit être soigneusement calibrée ; l'orientation du visage ainsi que les conditions d'illumination doivent être constantes sous peine de résultats erronés.

Une solution pour s'affranchir de toute phase d'apprentissage est de se tourner vers des modèles paramétriques comme l'a proposé Yuille dans [26].

En ce qui concerne l'extraction des yeux et des sourcils, le problème est un peu plus simple car les contours de ces traits sont moins déformables que ceux de la bouche. Dans la littérature, deux approches existent quant à la localisation de ces traits. Tout d'abord, on recherche une localisation grossière sans soucis de précision au niveau des contours en utilisant des informations issues de la luminance (image vallée par exemple) ([18], [23]) mais ces approches ne conviennent pas car elles ne conduisent pas à une détection réaliste des contours des yeux et des sourcils. La seconde approche consiste à introduire des modèles à associer aux contours recherchés [21].

Dans notre approche, nous avons opté pour le choix de modèles paramétriques. Dans la partie 2, nous décrivons et justifions les modèles choisis. Dans une phase de pré-traitement décrite dans la partie 3, on montre comment d'une part, on s'affranchit des variations d'illumination sur le visage et d'autre part comment on extrait le contour circulaire associé à l'iris. La partie 4 décrit les algorithmes de détection automatique de points caractéristiques sur le visage afin d'initialiser les modèles associés à chacun des traits du visage. La partie 5 montre comment ces modèles initiaux sont déformés en tenant compte d'informations de gradient afin qu'ils coïncident avec les contours présents dans l'image traitée. Enfin, la partie 6 présente un ensemble de résultats qualitatifs et quantitatifs permettant de juger d'une part de la pertinence des modèles choisis et d'autre part de la qualité de la segmentation. Les résultats obtenus nous permettent d'envisager l'utilisation de ces contours dans une application de reconnaissance des émotions présentes sur un visage.

2. Choix des modèles

L'analyse d'un ensemble d'images extraites de bases d'images de visage a permis de mettre en évidence que les modèles proposés jusqu'ici pour les yeux, les sourcils et les lèvres sont trop rigides pour permettre d'obtenir des contours réalistes.

2.1. Modèle de bouche

Plusieurs modèles paramétriques ont déjà été proposés pour modéliser le contour des lèvres. Tian [20] utilise un modèle constitué de deux paraboles. C'est simple à calculer mais la précision obtenue est très limitée (voir Figure 1-a). D'autres auteurs ont proposé de modéliser le contour supérieur des lèvres à l'aide de deux paraboles au lieu d'une [5] ou encore d'utiliser des quartiques [12]. Un gain en précision a été obtenu, néanmoins tous ces modèles sont encore limités par leur trop grande rigidité, en particulier dans le cas d'une bouche non symétrique (voir Figure 1-a, Figure 1-b, Figure 1-c).

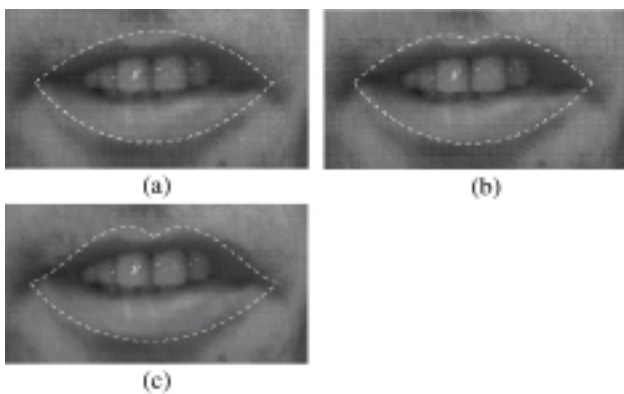


Figure 1. a) modèle à 2 paraboles ; b) modèle à 3 paraboles ; c) modèle à base de quartiques.

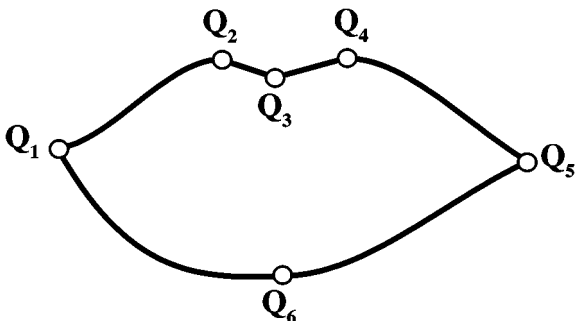


Figure 2. Modèle choisi pour la bouche.

Le choix d'un modèle adapté pour modéliser les lèvres est très délicat car la forme des lèvres est très variable. L'utilisation d'un modèle *a priori* lors de la phase de segmentation permet une régularisation des contours recherchés. Mais si le modèle choisi

n'est pas bien adapté, le résultat de la segmentation ne sera pas de bonne qualité. Le modèle que nous proposons est composé de 5 courbes indépendantes, chacune d'entre elles décrivant une partie du contour labial. Entre Q_2 et Q_4 , l'arc de Cupidon est décrit par une ligne brisée tandis que les autres portions du contour sont décrites par des courbes polynomiales cubiques γ_i (voir Figure 2). De plus, on impose à chaque cubique d'avoir une dérivée nulle au point Q_2 , Q_4 ou Q_6 . Par exemple, γ_1 (cubique entre Q_1 et Q_2) doit avoir une dérivée nulle en Q_2 .

2.2. Modèle pour les yeux et les sourcils

Le modèle le plus courant proposé jusqu'ici pour les yeux est un modèle à base de paraboles pour la modélisation de la frontière entre les yeux et les paupières et un modèle circulaire pour la modélisation de l'iris ([5], [21]). L'étude plus précise des contours des yeux sur la base d'images ORL [10] a montré que le contour associé à la paupière supérieure ne présentait pas nécessairement une symétrie verticale. Sur la Figure 3, nous présentons le cas d'un œil à paupière supérieure non symétrique qui est mal approché par une parabole (Fig.3-a) mais qui est bien délimité par une courbe de Bézier plus flexible (Fig.3-b). En revanche, le contour de la paupière inférieure présente une telle symétrie d'où le choix d'une parabole, l'idée étant également de choisir pour chaque contour la courbe la plus adaptée mais aussi la plus simple possible.

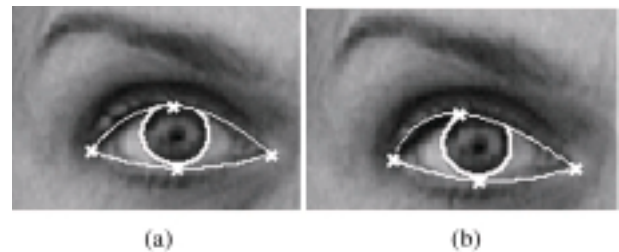


Figure 3. a) modèle à deux paraboles ; b) modèle à une courbe de Bézier pour le contour supérieur et une parabole pour le contour inférieur.

En ce qui concerne les sourcils, ceux-ci sont modélisés la plupart du temps par deux lignes brisées passant par les deux coins et un point milieu ce qui est très rudimentaire.

Ici, nous proposons un modèle paramétrique qui est plus adapté à la diversité des formes des yeux et des sourcils (voir Figure 4):

- Pour chaque œil : un cercle pour l'iris (éventuellement incomplet si l'œil est semi-ouvert) ; pour le contour inférieur, une parabole définie par trois points $\{P_1, P_2, P_4\}$; pour le contour supérieur, une courbe de Bézier à trois points de contrôle $\{P_1, P_2, P_3\}$; on se limite à une droite passant par P_1 et P_2 dans le cas d'un œil fermé.

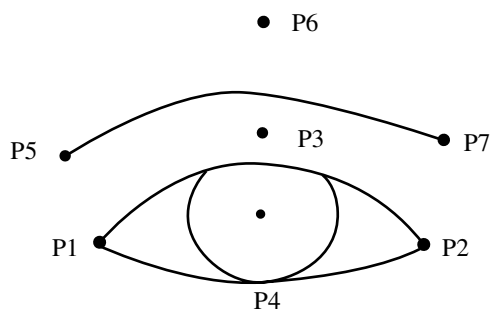


Figure 4. Modèles choisis pour l'œil et le sourcil.

- Pour les sourcils: une courbe de Bézier à trois points de contrôle $\{P_5, P_6, P_7\}$ pour le contour inférieur (on se limite à ce contour).

De manière générale, si on note $A(x_a, y_a)$, $B(x_b, y_b)$, $C(x_c, y_c)$ les coordonnées de trois points de contrôle, les coordonnées (x, y) de la courbe de Bézier associée sont définies par :

$$x = (1 - t)^2 x_a + 2t(1 - t)(x_c - x_a) / t^2 x_b$$

$$y = (1 - t)^2 y_a + 2t(1 - t)(y_c - y_a) / t^2 y_b$$

où t est un paramètre appartenant à $[0, 1]$.

Le modèle proposé est moins flexible que celui de la bouche car les yeux et les sourcils sont beaucoup moins déformables. Dans chaque cas, l'objectif est de proposer le modèle adapté au contour à extraire le plus simple possible afin de réduire la complexité. La contrainte de simplicité nous amène également à ne pas introduire dans le modèle un lien flexible explicite entre les positions relatives des iris et des contours des yeux. Ce type de lien est néanmoins envisagé sous forme de connaissance *a priori* dans la phase (en cours de développement) de suivi temporel des segmentations.

3. Pré-traitements

3.1. Conditions d'acquisition vidéo et de traitement

Les séquences analysées ont été acquises avec une caméra couleur standard de type Webcam. L'acquisition se fait en lumière naturelle donc de type lumière blanche (notre objectif est de nous placer dans des conditions « naturelles » d'acquisition compatibles avec les applications envisagées). La caméra est réglée de manière à faire un gros plan sur le visage de la personne présente dans la scène (la taille minimale du visage dans l'image doit être de l'ordre de 90×90 pixels). Dans la première image de la séquence, il est nécessaire que le visage de la personne soit vertical et face à la caméra et que les yeux soient ouverts. Par la suite, il est nécessaire que le visage de la personne reste proche de la verticale.

Il existe de nombreux travaux sur l'extraction automatique du visage dans une image (voir les articles de synthèse [13], [25]) et l'extraction automatique de la zone contenant le visage n'est pas le propos de ce travail. Dans cet article, la méthode proposée impose, sur la première image de la séquence, l'extraction à la main d'un rectangle englobant le visage ainsi que l'estimation du rayon de l'iris (cf. Figure 7). En première approximation, ce rectangle passe au ras des oreilles pour ses limites droite et gauche, il est limité en bas par le menton et en haut, il est défini de manière à inclure dans la zone englobante le moins de cheveux possible. Cette phase d'initialisation sur la première image de la séquence ne nous semble pas rédhibitoire pour les applications envisagées. À terme, il serait souhaitable de pouvoir extraire automatiquement la zone du visage. Les méthodes de détection de visage utilisent fréquemment l'information de chrominance typique associée à la peau. Voilà pourquoi nous avons défini un rectangle autour le visage contenant majoritairement des points de peau. Le cas d'un visage pour lequel les cheveux sont très longs et recouvrent totalement le front et éventuellement les sourcils et les yeux n'a pas été explicitement abordé: les algorithmes d'extraction des contours des traits du visage supposent que ces contours sont présents et visibles dans les images traitées.

Les dimensions du rectangle englobant le visage sont supposées constantes tout au long de la séquence et la zone englobant le visage est suivie d'image en image par *block matching*. Cette hypothèse limite l'amplitude des déplacements de l'utilisateur perpendiculairement à l'axe de la caméra (effet de zoom) ce qui n'est pas trop gênant dans le cas d'une personne travaillant devant un PC. Des mouvements de rotation (droite/gauche) et (haut/bas) de la tête sont possibles dans la mesure où tous les contours à extraire restent visibles.

3.2. Prise en compte des variations d'illumination

Dans la phase de pré-traitement, nous nous affranchissons des variations d'illumination en utilisant un filtrage adapté inspiré du comportement de la rétine [2]. Ce filtre permet de réaliser un lissage local des variations d'éclairement. Nous donnerons ici les étapes successives des traitements réalisés, les justifications sont détaillées dans [2]. Le filtrage rétinien réalise une succession de filtrages et de compressions adaptatives. Soit G un filtre gaussien de taille 15×15 et d'écart type $\sigma = 2$. Soit I_{in} l'image initiale et I_1 le résultat de son filtrage par G . À partir de l'image I_1 , on définit l'image X_0 par la relation :

$$X_0 = \frac{0,1 + 410 I_1}{105,5 + I_1}$$

L'image X_0 permet de définir la fonction de compression (notée C) par la relation :

$$C : I \longrightarrow \frac{(255 + X_0)I}{X_0 + I}$$

où I est une image quelconque.

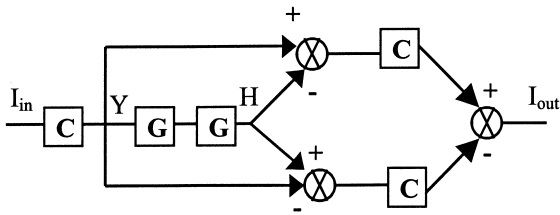


Figure 5. Filtre rétinien.



Figure 6. À gauche, image à forte différence d'éclairage ; à droite, image en sortie du filtre rétinien.

La figure 5 donne le schéma bloc du filtrage rétinien, la sortie de ce filtrage est notée I_{out} .

La Figure 6 présente l'effet de ce filtre sur un visage éclairé latéralement qui présente de ce fait une variation importante de luminosité entre les parties gauche et droite du visage. À l'issue du filtrage, les variations de luminosité ont été fortement atténuées.

3.3. Localisation des deux iris

Le contour de l'iris étant la frontière entre une zone sombre (l'iris) et une zone claire (le blanc de l'œil), ce contour est recherché sous la forme d'un cercle constitué de points de gradient de luminosité maximal. De plus, comme les yeux peuvent éventuellement être semi-ouverts, il arrive que le contour supérieur de l'iris soit occulté. Ainsi, dans chaque quart supérieur droit et gauche du rectangle englobant le visage, on recherche le

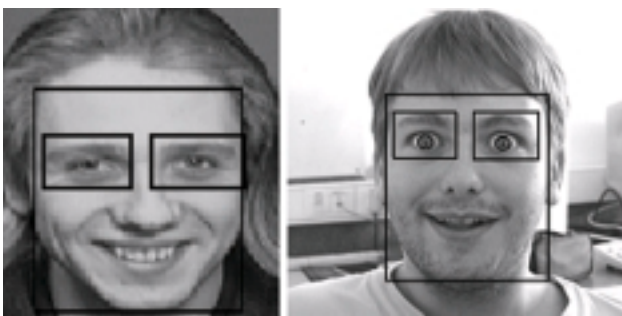


Figure 7. Rectangle englobant sur le visage ; détection des iris et zones rectangulaires de recherche des yeux et des contours (l'image de gauche provient de la base ORL [10]).

demi-cercle qui maximise le Flux de Gradient de luminosité Normalisé :

$$FGN = \frac{1}{length\ SC} \sum_{p \in SC} \vec{\nabla} I(p) \cdot \vec{n}$$

Avec $I(p)$ luminosité au point p , SC le demi-cercle cherché et \vec{n} la normale au contour au point p . En effet, en chacun des points du cercle cherché, le gradient de luminosité est normal au contour.

À l'heure actuelle, on suppose que le rayon de l'iris est connu (cf. 3.1) et seule la position du centre dans la zone de recherche est modifiée d'un demi-cercle à l'autre. Tous les points de la zone de recherche de l'iris sont testés comme centre potentiel. La méthode de maximisation du FGN présente l'avantage d'être très rapide (par rapport à une méthode de type descente de gradient par exemple), sans réglage de paramètre et elle conduit sans ambiguïté à la sélection du bon demi-cercle puisque le FGN présente toujours un pic très marqué (cf. Figure 8). Ce pic correspond à la position correcte pour le demi-cercle recherché. Il serait possible d'envisager de tester plusieurs valeurs pour le rayon, la gamme des rayons possibles étant elle-aussi corrélée aux dimensions du visage.

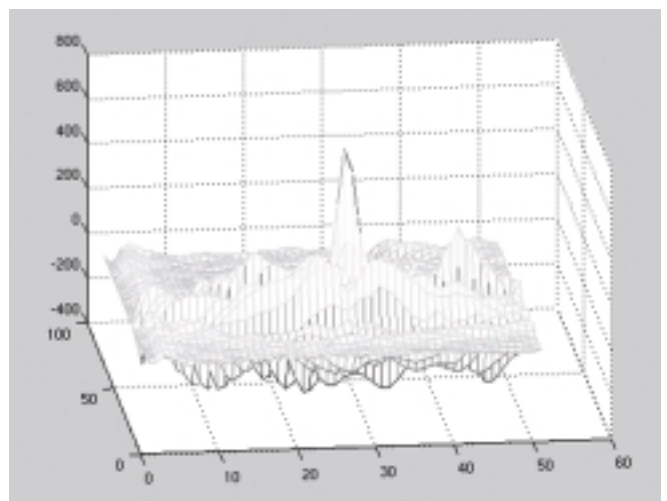


Figure 8. Evolution du FGN lors du balayage de la zone de recherche de l'iris (quart supérieur droit ou gauche du rectangle englobant le visage).

4. Extraction de points caractéristiques

Le positionnement initial de chaque modèle sur l'image à traiter a lieu après extraction automatique de points caractéristiques sur le visage.

4.1. Cas des yeux

Un processus de suivi de points de gradient de luminance maximum est utilisé pour détecter les coins des yeux.

À partir de la boîte englobant le visage et de ses dimensions (H_{visage}, L_{visage}), nous avons défini les dimensions (H_{oeil}, L_{oeil}) des rectangles englobants les yeux et les sourcils à partir des relations suivantes (voir Figure 7) :

$$H_{visage} = 4 * H_{oeil} \text{ et } L_{visage} = 2,5 * L_{oeil}$$

Ces relations ont été apprises suite à l'extraction à la main de ces rectangles englobants sur les 400 images de la base ORL [10]. On constate sur la Figure 7 que les zones englobants les sourcils et les yeux sont plus petites que les zones utilisées au § 3.3 (quarts supérieurs gauche et droit du visage) pour localiser le centre de chaque iris.

Le centre des rectangles englobant les yeux et les sourcils est aligné sur le centre de chaque iris détecté. Ceci nous permet de limiter et de positionner efficacement la zone de recherche des yeux et des sourcils.

La Figure 9 donne une illustration de la méthode de détection des coins des yeux : en partant du point X_1 déterminé par translation de 3 pixels à gauche à partir de la position du point le plus bas du cercle détecté pour l'iris, un algorithme de suivi, vers la gauche, de points de gradient maximum conduit à la détection du coin C_1 . Un écart de 3 pixels du cercle de l'iris est nécessaire afin d'éviter de suivre les points de gradient maximum situés sur ce même cercle. Puisque le point initial X_1 est au-dessous

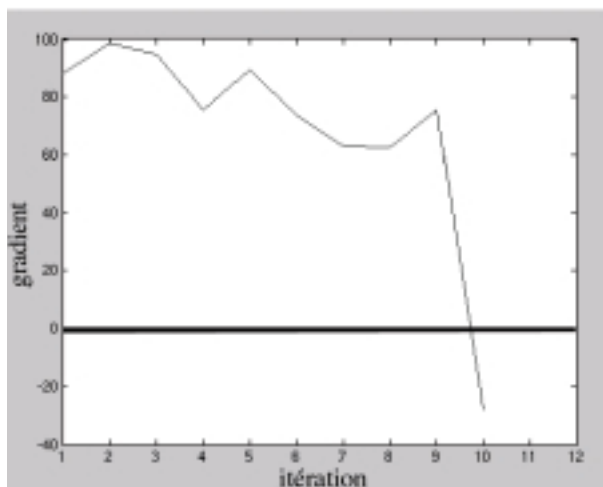
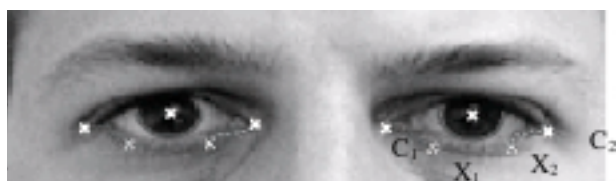


Figure 9. Détection des coins des yeux: en haut, processus de suivi; en bas, évolution du gradient le long de la courbe X_1C_1 .

du coin cherché C_1 , seuls les trois points noirs voisins suivants

$\begin{pmatrix} \bullet & \bullet \\ \bullet & X_1 \end{pmatrix}$ situés à gauche et au-dessus de la position courante

de X_1 sont testés. La courbe joignant les points X_1 et C_1 est constituée d'un ensemble de points de gradient de luminance localement maximum. Le suivi de point s'arrête lorsque le gradient de luminance devient négatif (à l'itération 10 pour l'exemple de la Figure 9) puisqu'un point de peau est plus clair que le pixel associé au coin de l'œil. Un processus de suivi similaire permet de détecter le second coin C_2 en partant de X_2 .

La parabole, modèle choisi pour le contour inférieur de l'œil, est initialisée par les deux coins C_1 et C_2 et le point le plus bas du cercle détecté pour l'iris; la courbe de Bézier, modèle choisi pour le contour supérieur de l'œil, est initialisée par les deux coins des yeux et le centre du cercle détecté pour l'iris. La Figure 10 montre le résultat de l'initialisation du modèle pour les yeux.



Figure 10. Positionnement du modèle initial pour les yeux.

4.2. Cas des sourcils

Pour l'initialisation de la courbe de Bézier associée aux sourcils, on extrait les deux coins intérieur et extérieur de chaque sourcil notés S_1 et S_2 .

Pour chaque sourcil, on réduit la zone de recherche de ces points à la zone de l'image située au-dessus de l'iris détecté lors de la phase de pré-traitement. Sur cette zone de recherche, on calcule l'image des vallées définie par: $Val(x,y) = 255 - I(x,y)$ où $I(x,y)$ est la luminance au point de coordonnées (x,y) .

Pour la détection des abscisses x_1 du coin intérieur S_1 et x_2 du coin extérieur S_2 , on recherche les abscisses des deux points pour lesquels il y a changement de signe ou annulation de la dérivée de la projection horizontale de l'image vallée définie par: $H(x) = \sum_{y=1}^{N_y} Val(x,y)$ où N_y représente le nombre de lignes de l'image.

Pour la détection des ordonnées y_1 du coin intérieur S_1 et y_2 du coin extérieur S_2 , on recherche l'abscisse du maximum de la projection verticale de l'image vallée définie par: $V(y) = \sum_{x=1}^{N_x} Val(x,y)$ où N_x représente le nombre de colonnes de l'image.

S_1, S_2 et le point S_3 défini par les coordonnées: $\{(x_1 + x_2)/2; y_2\}$ servent de points de contrôle initiaux pour la courbe de Bézier associée à chaque sourcil.

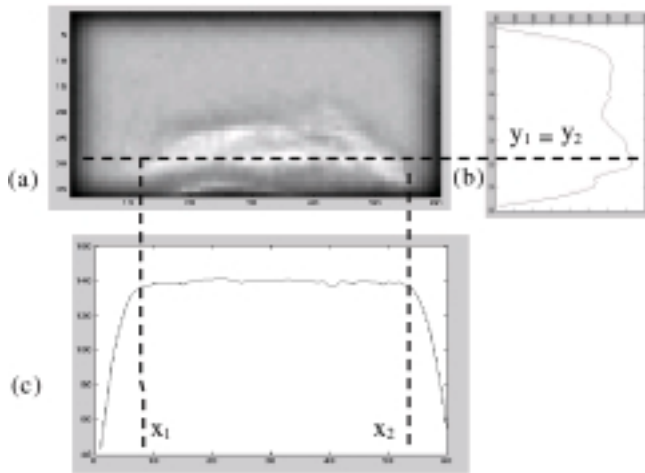


Figure 11. a) image vallée; b) projection verticale; c) projection horizontale.

La détection des coins des sourcils est une tâche difficile (tout particulièrement pour le coin extérieur qui n'est pas toujours très bien défini sur l'image) et elle suppose que les sourcils sont visibles et sont plus sombres que la peau (ce qui est très majoritairement le cas). La méthode proposée ne permet pas une détection très précise de ces coins dans tous les cas car elle peut être sujette au bruit. Voilà pourquoi ces points seront réajustés lors de la phase de déformation du modèle associé aux sourcils.

4.3. Cas de la bouche

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

4.3.1. Analyse comparative de la couleur des lèvres et de la peau

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

Dans [27], l'espace HSV est utilisé pour la détection des lèvres car la teinte est discriminante. Même pour des personnes différentes, la teinte des lèvres et la teinte de la peau présentent des valeurs différentes. Cependant, la teinte est une grandeur qui est souvent bruitée: du fait de sa définition angulaire, les faibles valeurs de teinte sont proches des valeurs élevées. Par ailleurs, la teinte n'est pas du tout fiable pour les pixels à faible saturation. Les espaces CIELUV et YCrCb sont utilisés également en analyse de visages (une définition de ces espaces ainsi qu'une

caractérisation de leurs performances à discriminer les pixels de lèvres des pixels de peau est donnée dans [9]). Il a été mis en évidence que les pixels de peau étaient concentrés sur une faible portion des espaces (Cr, Cb) ou (u, v) ([21][23]). Néanmoins, même dans ces deux espaces, il y a recouvrement entre les pixels de lèvres et les pixels de peau ce qui les rend peu adaptés à une segmentation précise.

Dans l'espace RVB, les pixels peau et lèvre ont des distributions différentes. Certes pour les deux types de pixels, le rouge prédomine. Cependant, il y a plus de vert que de bleu dans le mélange de couleurs associé à la peau alors que ces deux composantes sont équivalentes pour les lèvres [6]. La peau semble plus jaune que les lèvres du fait de la différence entre le rouge et le vert qui est plus marquée pour les lèvres que pour la peau. Hulbert et Poggio [14] proposent de définir une pseudo-teinte h qui met en évidence cette différence :

$$h(x, y) = \frac{R(x, y)}{G(x, y) + R(x, y)}$$

où $R(x, y)$ et $G(x, y)$ sont respectivement les composantes rouge et verte au pixel (x, y) . Contrairement à la teinte usuelle, la pseudo-teinte est une fonction bijective. Elle est plus élevée pour les lèvres que pour la peau [6].

La luminance est également une information intéressante à considérer. En général, la lumière provient du dessus du personnage ce qui fait que le contour supérieur des lèvres est très bien éclairé alors que le contour inférieur se retrouve dans l'ombre. Afin de tenir compte de la couleur et de la luminance, on utilise l'information hybride $\vec{R}_{top}(x, y)$, définie dans [7]. Elle est calculée de la manière suivante :

$$\vec{R}_{top}(x, y) = \vec{\nabla} [h_N(x, y) - I_N(x, y)]$$

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

4.3.2. L'algorithme du Jumping Snake

Les snakes [15] ont été avantageusement utilisés en segmentation d'images. Cependant, ni le problème difficile du réglage des paramètres du snake, ni celui de sa haute dépendance à la position initiale n'ont pu être résolus. La méthode que nous proposons permet de s'affranchir de ces deux difficultés.

Pour détecter des points caractéristiques du contour supérieur de la lèvre, nous définissons un nouveau type de contour actif que nous avons désigné sous le nom de *jumping snake* car sa convergence fait intervenir successivement des phases de croissance et des phases de sauts [8]. Il est initialisé avec un germe noté S^0 positionné automatiquement dans la zone comprise entre le nez et la bouche près de l'axe de symétrie verticale grâce à la connaissance de la position des yeux et une étude des

positions respectives moyennes des yeux et de la bouche effectuée sur la base ORL (voir Figure 12). Il est important que le germe initial soit positionné au-dessus de la bouche et en dessous du nez mais ce positionnement ne requiert pas une très grande précision (par exemple, toutes les positions initiales à l'intérieur du rectangle noir défini de manière expérimentale sur la Figure 12 sont possibles) si bien que la position précise du nez n'a pas à être connue. Le *jumping snake* grandit depuis le germe jusqu'à atteindre un nombre prédéfini de points. Cette phase de croissance est semblable à ce qui est proposé par Berger et Mohr [3], en ce sens que le *jumping snake* est initialisé avec un seul point et qu'il s'accroît progressivement jusqu'à son point final. Dans une seconde phase, le germe « saute » vers une nouvelle position plus proche du contour recherché (voir Figure 12). Le processus s'arrête lorsque la taille autorisée pour le saut devient plus petite qu'un pixel (ce qui requiert en moyenne 4 à 5 itérations).

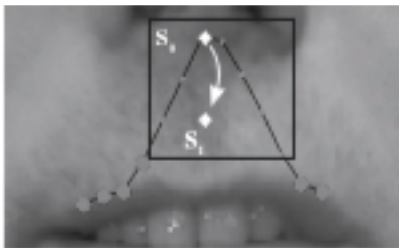


Figure 12. Initialisation du *jumping snake*.

La position du germe S^1 est calculée à partir de S^0 (initialisé à l'intérieur du cadre noir) et des points (gros ronds) associés aux flux moyens les plus élevés.

Durant la phase de croissance du *jumping snake*, les meilleurs points à ajouter à droite et à gauche, notés respectivement $M_{-(i+1)}$ et M_{i+1} , sont les points qui maximisent le flux moyen de \vec{R}_{top} à travers les segments $M_{-(i+1)}M_{-i}$ et M_iM_{i+1} (voir Figure 13). Les deux flux moyens sont définis par :

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

où $d\vec{n}$ est la normale au segment. Les maximisations de ϕ_{-i-1} et ϕ_{i+1} s'obtiennent en comparant les valeurs obtenues pour un petit ensemble de points candidats. Une méthode de maximisation de type descente de gradient n'a pas été retenue car, le nombre de points candidat à tester étant faible, (8 points environ pour $(\theta_{inf}, \theta_{sup}) = (-\frac{\pi}{3}, \frac{\pi}{5})$), la méthode d'optimisation exhaustive proposée permet d'obtenir une solution beaucoup plus rapidement.

Lorsque le *jumping snake* a atteint un nombre prédéfini de points noté $2N + 1$, la phase de croissance s'arrête.

Commence alors la **phase de saut** de l'algorithme. Soient $\{M_{-N}, \dots, M_{-1}, S^0, M_1, \dots, M_N\}$ l'ensemble des points du snake et soient $\{\phi_{-N}, \dots, \phi_{-1}, \phi_1, \dots, \phi_N\}$ l'ensemble des flux moyens à travers les $2N$ segments. Le nouveau germe S^1 doit être proche d'une zone de fort gradient, à savoir d'une zone de flux moyen élevé. On considère que S^1 est le barycentre de S^0 et des points situés dans les zones de gradients les plus élevés (cf. les gros points sur la Figure 12). Soit $\{i_1, \dots, i_N\}$ les indices des N plus grands flux moyens, l'ordonnée du nouveau germe S^1 est calculée par la relation :

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

où $y(i_k)$ est l'ordonnée du point M_{i_k} . L'abscisse x_{S^1} du germe est maintenue constante.



Figure 13. Depuis le germe S^0 , le snake s'accroît par ajout de points à droite et à gauche suite à la maximisation de \vec{R}_{top} à travers chaque nouveau segment.

Le processus complet est alors itéré : un nouveau *jumping snake* s'accroît à partir de ce nouveau germe jusqu'à ce qu'il atteigne la longueur pré-définie et une nouvelle phase de saut se produit. La convergence du *jumping snake* est atteinte lorsque l'amplitude du saut associé au germe devient inférieure au pixel.

L'algorithme du *jumping snake* fait intervenir 4 paramètres : $(\theta_{inf}, \theta_{sup})$ définissant le secteur angulaire d'évolution possible pour chaque branche du *jumping snake*; Δ définissant la distance horizontale entre deux points consécutifs du snake et N réglant le nombre de points total de chaque branche du snake. Pour tous les résultats présentés dans cet article, ces paramètres ont été fixés à $(\theta_{inf}, \theta_{sup}, N, \Delta) = (-\frac{\pi}{3}, \frac{\pi}{5}, 6, 5)$. Pour le choix du secteur angulaire, il est impératif que $|\theta_{inf}| > |\theta_{sup}|$ afin de faire évoluer le *jumping snake* vers le bas en direction de la bouche. Le choix de N et de Δ résulte d'un compromis entre précision du contour et rapidité de convergence. Une étude complète de l'influence de chacun de ces 4 paramètres est donnée dans [9].

4.3.3. Détection des points caractéristiques

On considère 6 points principaux (voir Figure 2) : les coins droit et gauche de la bouche (Q_1 et Q_5), le point central le plus bas de la lèvre inférieure (Q_6) et les trois points de l'arc de Cupidon (Q_2 , Q_3 et Q_4). Deux points secondaires situés à l'intérieur de la bouche sont également considérés : Q_7 et Q_8 (voir Figure 14). Ils sont utilisés pour détecter automatiquement la position du point central bas Q_6 .

Les trois points supérieurs sont situés sur le contour résultant de la convergence du *jumping snake* : Q_2 et Q_4 sont les points les plus hauts de part et d'autre du germe final. Q_3 est le point le plus bas du contour situé entre Q_2 et Q_4 (voir Figure 14).

Les points Q_6 , Q_7 et Q_8 sont détectés par analyse de $\nabla_y(h)$, gradient 1D de la pseudo-teinte le long de l'axe vertical passant par Q_3 (voir Figure 14). La pseudo-teinte est plus forte pour les lèvres que pour la peau, la langue ou les dents. Le maximum de $\nabla_y(h)$ au-dessous du contour supérieur donne la position de Q_7 . Q_6 et Q_8 sont les minima de $\nabla_y(h)$ en-dessous et au-dessus de Q_7 respectivement. Ceci suppose que le visage est aligné sur la verticale et donc que les lèvres sont horizontales.

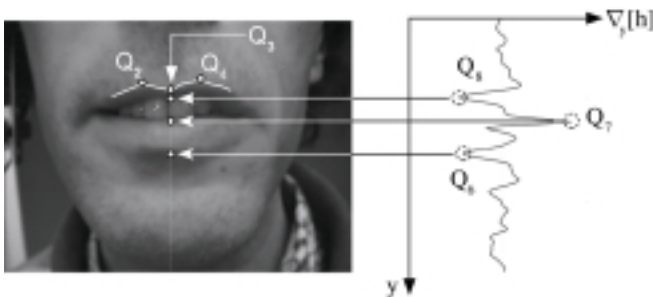


Figure 14. Les trois points du haut proviennent du *jumping snake* (ligne blanche). Q_6 , Q_7 et Q_8 sont en dessous de Q_3 sur les extrêmes de $\nabla_y(h)$.

5. Evolution des modèles

5.1. Cas des yeux et des sourcils

La Figure 15 présente un exemple d'initialisation des modèles associés à l'œil et au sourcil.

L'idée pour la déformation du modèle initial reste la même : chaque contour recherché est un ensemble de points de gradient de luminosité maximum. La courbe sélectionnée est celle qui maximise le flux du gradient de luminosité normalisé 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. À l'opposé, puisque le point de contrôle P_3 de la



Figure 15. Initialisation des contours de l'œil et du sourcil.

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 sur un axe vertical (les points P_1 et P_2 restant fixes) jusqu'à ce que le flux du gradient de luminosité à travers le contour soit maximum.

Pour la déformation de la courbe associée à chaque sourcil, le principe est le même sachant que le contour du sourcil représente un ensemble de points de gradient maximum localement. La différence réside dans la stratégie de déplacement des points de contrôle. Comme les points initiaux P_5 , P_6 et P_7 sont initialisés avec moins de précision que dans les cas des coins des yeux, on envisage un déplacement de chacun des trois points de contrôle : le déplacement est autorisé sur un axe vertical seulement pour les points P_5 , P_7 et il est autorisé dans toutes les directions mais vers le haut pour le point P_6 .

5.2. Cas de la bouche

Pour la déformation des cubiques associées aux lèvres, la stratégie est différente. On va rechercher en même temps les meilleures courbes et les positions des coins de la bouche car l'ajustement du modèle et la détection précise des commissures sont deux processus liés. Lorsqu'un être humain recherche les commissures, il utilise implicitement la connaissance qu'il a de la forme globale de la bouche : il suit les contours supérieur et inférieur des lèvres et même en cas de contours devenant peu marqués, il les prolonge et place les commissures à l'intersection (éventuellement interpolée) de ces deux contours.

A priori, une cubique est définie de manière unique par la connaissance de 4 paramètres. Dans notre cas, chaque cubique passe par l'un des points caractéristiques Q_2 , Q_4 ou Q_6 et doit y avoir une dérivée nulle. Ces considérations permettent de réduire le nombre de paramètres à estimer pour chaque cubique de 4 à 2. Il manque donc deux points supplémentaires afin d'obtenir une définition unique de chaque cubique. Les points manquants sont sélectionnés parmi les points les plus fiables des contours à savoir près de Q_2 , Q_4 ou Q_6 . Pour les deux cubiques associées à la lèvre supérieure, les points manquants sont choisis sur le *jumping snake*. En ce qui concerne les contours de la lèvre inférieure, seul le point Q_6 est connu. Pour obtenir des points supplémentaires, on fait croître un *snake* à partir du germe Q_6 . La croissance s'arrête après l'ajout de quelques points (voir les points blancs de la Figure 16).

Maintenant que nous disposons de suffisamment de points sur chaque partie du contour des lèvres, il est possible de calculer les cubiques γ_i passant par ces points et de trouver les commissures en tant que points d'intersection de deux courbes. Afin de rendre les résultats plus robustes, on suppose de plus que les positions des coins (Q_1, Q_5) sont connues. Ceci permet d'ajouter une contrainte supplémentaire qui réduit encore d'une unité le nombre de paramètres à estimer pour chaque cubique. De là, l'estimation du paramètre manquant est faite très rapidement par minimisation au sens des moindres carrés et les résultats obtenus sont moins sensibles à la position des points supplémentaires choisis. En d'autres termes, pour une position donnée des commissures, il correspond un unique couple de courbes rapide à calculer. Donc finalement, l'ajustement des cubiques se fait en recherchant les coins qui conduisent aux contours les plus précis. Bien entendu, un test exhaustif de l'ensemble des points de l'image en tant que coins potentiels est hors de question. En remarquant que les coins de la bouche sont situés dans des zones sombres de l'image (ce qui est très majoritairement le cas car les commissures se trouvent dans une zone de replis en retrait des deux lèvres), nous pouvons réduire considérablement la zone de recherche. Ainsi, nous recherchons le pixel de luminance minimale sur chaque colonne située entre les frontières supérieure et inférieure ce qui conduit à la construction d'une ligne de minima de luminance notée L_{mini} (voir Figure 16). On suppose alors que les coins sont sur cette ligne. À chaque coin (droit ou gauche) correspond un unique couple de cubiques, l'une pour la moitié de la lèvre supérieure et l'autre pour la moitié de la lèvre inférieure (cf. les courbes en pointillés sur la Figure 16). L'ajustement du modèle consiste à trouver les coins qui conduisent aux cubiques les plus proches des contours réels qui sont par hypothèse ceux qui satisfont au critère d'optimalité du maximum de flux de \vec{R}_{top} . Une analyse qualitative visuelle des contours obtenus en maximisant un flux de gradient de luminance seule ou un flux de gradient de chrominance seule a été faite dans [9] et a mis en évidence la supériorité de l'information hybride \vec{R}_{top} pour obtenir une détection robuste des contours de lèvres.

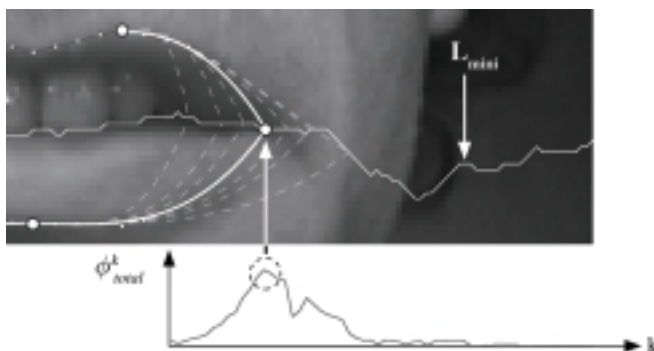


Figure 16. Le maximum de ϕ_{total}^k donne la position du coin sur L_{mini} . Les courbes en pointillés sont les cubiques associées aux différents points k testés le long de L_{mini} .

Si les cubiques estimées pour le contour supérieur γ_1 et γ_2 coïncident parfaitement avec les contours réels, elles sont orthogonales au champ de gradient \vec{R}_{top} . De même, les courbes γ_3 et γ_4 pour le contour inférieur doivent être normales en tout point au champ de gradient $\vec{\nabla}[h_N]$. On calcule $\phi_{\text{top},i}$ et $\phi_{\text{low},i}$, flux moyens à travers les courbes du haut et du bas respectivement par les relations :

$$\phi_{\text{top},i} = \frac{\int_{\gamma_i} \vec{R}_{\text{top}} \cdot d\vec{n}}{\int_{\gamma_i} ds} \quad i \in \{1,2\}$$

$$\phi_{\text{low},i} = \frac{\int_{\gamma_i} \vec{\nabla}[h_N] \cdot d\vec{n}}{\int_{\gamma_i} ds} \quad i \in \{3,4\}$$

où $d\vec{n}$ et ds représentent la normale en chaque point de contour et l'abscisse curviligne. On considère n positions possibles le long de L_{mini} pour chaque point Q_1 et Q_5 . La meilleure position donne une valeur élevée pour $\phi_{\text{top},i}$ et une valeur fortement négative pour $\phi_{\text{low},i}$. De chaque côté, nous devons donc maximiser la quantité suivante :

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

avec :

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

ϕ_{top}^k et ϕ_{low}^k étant associés au k ième point testé. $\phi_{\text{top},\text{normalisé}}^k$ et $\phi_{\text{low},\text{normalisé}}^k$ sont des valeurs normalisées sur l'ensemble des points testés. Dès lors que ϕ_{total}^k est grand, la position du coin est fiable (une estimation quantitative de la précision obtenue lors de la détection des coins de la bouche est donnée dans la partie 6) car les courbes estimées coïncident bien (visuellement) avec les contours réels ce qui montre qu'ajustement du modèle et détection des commissures sont réalisés en une seule opération. La Figure 16 montre l'évolution de ϕ_{total}^k pour différentes valeurs de k . Le maximum de ϕ_{total}^k donne la position du coin le long de L_{mini} .

6. Résultats

La méthode de segmentation décrite a été testée sur une base de séquences vidéos acquises au laboratoire constituées de visages plus ou moins éloignés de la caméra. L'éclairage est *a priori* quelconque. De plus, nous avons également testé nos méthodes de segmentation sur la base de séquences de visages avec émotions de Kanade et Cohn [4]. Seule l'extraction des contours des yeux et des sourcils a pu être testée car les séquences vidéos de

cette base sont en niveau de gris. Or l'extraction du contour de la bouche utilise également l'information de chrominance.

6.1. Pertinence des modèles et qualité de segmentation

La Figure 17 permet de juger visuellement de la pertinence du modèle choisi pour le contour extérieur des lèvres ainsi que de sa flexibilité. Le choix d'un modèle pour ce contour est un véritable challenge car c'est lui qui présente le plus de déformations potentielles. On constate qu'avec le modèle proposé, il est possible de segmenter ce contour même dans le cas de bouches très déformées (cas d'une grimace). Ceci est très important puisqu'une des applications visées est l'utilisation des informations dynamiques relatives à la forme de la bouche afin d'améliorer la compréhension d'un message vocal lors d'une transmission bruitée. L'utilisation d'une information visuelle pour augmenter l'intelligibilité d'un message n'a de sens que si cette information est de bonne qualité (en particulier, contours extraits fidèles).

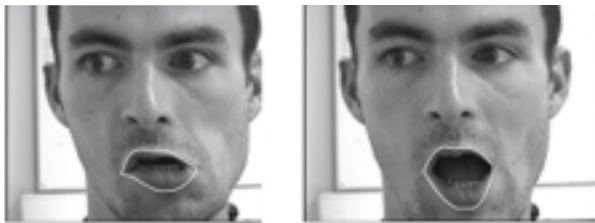


Figure 17. Contours extraits pour la bouche dans le cas d'une grimace.

La Figure 18 présente des résultats d'extraction de l'ensemble des traits (bouche, yeux, sourcils) du visage. Les contours obtenus sont précis et réalistes même dans le cas où le sujet porte des lunettes. Les lunettes ne présentent pas un problème particulier puisque, pour l'extraction des contours des yeux, l'étape préalable est la détection de l'iris et qu'en général, les lunettes sont plus grosses que les yeux. Pour l'extraction du contour associé aux sourcils, il est possible, selon la taille des lunettes, que ce contour coïncide avec celui des lunettes. Mais alors même visuellement, il n'est pas possible de faire la différence. Afin de faire une évaluation quantitative de la qualité des résultats obtenus à l'issue de la segmentation, nous avons comparé les résultats obtenus par notre algorithme avec ceux issus d'une segmentation manuelle. La précision est évaluée au niveau des points caractéristiques Q_1 à Q_6 pour la bouche, aux points P_1 , P_2 pour chaque œil et aux points P_5 et P_7 pour chaque sourcil. Pour cette comparaison, 300 images issues de 8 séquences différentes ont été d'une part étiquetées à la main par 5 personnes différentes et d'autre part, traitées par nos algorithmes de segmentation. Les positions respectives des points obtenus ont été comparées. La Figure 19 donne une représentation de l'erreur moyenne obtenue pour chaque point : associé à chaque point caractéristique est représenté un cercle dont le rayon est égal à



Figure 18. Contours extraits pour différentes personnes.

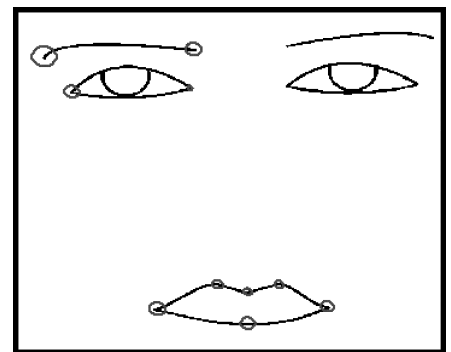


Figure 19. Étude quantitative de la qualité de la segmentation.

cette erreur moyenne. On constate que, pour la bouche, les points caractéristiques extraits sont moins précis au niveau des commissures (remarquons que ces points sont parfois difficiles à déterminer précisément même à l'œil nu). Pour les yeux et les sourcils, la précision obtenue est moins bonne pour les coins extérieurs, plus difficiles à caractériser sur l'image que les coins intérieurs.

La précision obtenue pour chaque point est donnée dans le tableau 1. Les erreurs sont exprimées en pourcentage des dimensions de la bouche, des yeux et des sourcils respectivement. À titre de comparaison, nous avons également donné la dispersion obtenue lors de l'étiquetage manuel de ces mêmes séquences par 5 personnes différentes. Même lors d'un étiquetage manuel, nous obtenons une erreur moyenne non nulle (cf. tableau 1) et celle-ci est plus importante au niveau des points caractéristiques extérieurs des yeux et des sourcils (car plus difficiles à définir précisément même à l'œil nu).

Pour une étude quantitative plus complète de la qualité des contours obtenus lors de la segmentation des lèvres, nous envisageons deux approches : synthèse des mouvements des lèvres

Tableau 1. Erreurs exprimées en % pour l'extraction des points caractéristiques automatiquement (P_{algo}) et manuellement (P_{main}).

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	P_1	P_2	P_5	P_7
P_{algo}	4.5	2.9	2.3	2.4	3.8	4.1	6.1	2.9	7.8	4.9
P_{main}	1.5	1.1	1.1	1.2	1.6	1.6	2.5	1.7	3.2	1.9

d'un clone parlant à partir des paramètres des contours extraits et reconnaissance de sons (cette validation est en cours de développement).

6.2. Application à la reconnaissance d'émotions

Les contours extraits étant suffisamment réalistes, nous les utilisons dans un système d'analyse et de reconnaissance des six émotions universelles sur le visage (joie, surprise, peur, dégoût, colère, tristesse). Ceci a été validé par les résultats d'une expérience pour laquelle nous avons présenté les contours extraits seuls (sans visage en arrière plan, cf. Figure 20) à des personnes juges (60 sujets masculins et 60 sujets féminins) et nous leur avons demandé de reconnaître l'émotion présente. Le taux de reconnaissance obtenu est de l'ordre de 65% lorsque les personnes juges ont à choisir la réponse parmi un nombre d'émotions limitées et suggérées (joie, surprise, dégoût, neutre qui représentent les émotions susceptibles d'être présentes sur les séquences de contours proposées +3 émotions servant de leurres: colère, tristesse, peur).

La classification ou reconnaissance d'une émotion se base l'évolution temporelle des informations contenues dans le « squelette » issu de cette étape de segmentation (évolution temporelle de cinq distances caractéristiques). La Figure 20 donne des exemples de squelettes émotionnels dans le cas de la joie et de la surprise ainsi que la définition des distances considérées (la distance D_5 non représentée sert à mesurer le degré d'ou-

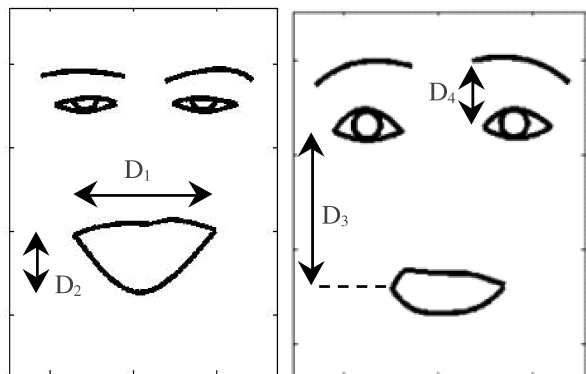


Figure 20. Squelettes d'émotion pour la joie (à gauche) et pour la surprise (à droite); distances caractéristiques.

verture de l'œil). On constate que la joie se caractérise essentiellement par une extension des dimensions du contour associé à la bouche et un plissement des yeux. La surprise quant à elle se caractérise par un écarquillement des yeux d'où un mouvement des sourcils vers le haut et une grande ouverture des yeux. La description complète d'un système de reconnaissance d'émotions basée sur la théorie de l'évidence et utilisant les distances D_i comme mesures est proposée dans [11].

6.3. Bilan des algorithmes proposés

Les algorithmes proposés font intervenir les paramètres suivants :

- Pour les lèvres, 4 paramètres ($\theta_{inf}, \theta_{sup}, N, \Delta$) sont associés à la création du *jumping snake*. L'ordre de grandeur de ces paramètres est relié à la taille de la bouche ainsi qu'à un compromis entre vitesse et précision [9]. Les valeurs utilisées pour l'ensemble des exemples présentés ont été données à la section 4.3.2
- Pour les yeux et les sourcils, le paramètre à estimer manuellement au préalable est le rayon de l'iris. En respectant les contraintes d'acquisition décrites au paragraphe 3.1, les résultats sont robustes vis à vis des rectangles englobants définis pour le visage et les yeux.

En ce qui concerne les temps de calcul, l'extraction des contours des lèvres est faite à la cadence de 15 images par seconde pour un code en langage C et sur un PC standard. L'extraction des contours de l'iris et des yeux conduit également à une cadence de l'ordre de 15 images par seconde. L'extraction des contours des sourcils est plus lente car l'ensemble des points de contrôle du modèle du sourcil doit être ajusté.

Les algorithmes proposés sont adaptés aux conditions d'acquisition des images pour chacune des applications envisagées. Néanmoins, nous travaillons sur l'estimation des mouvements rigides de la tête ce qui nous permettra de lever les contraintes d'alignement du visage sur la verticale et de mouvements limités du visage perpendiculairement à l'axe de la caméra (effet de zoom).

En revanche, nous avons mis en évidence la pertinence des modèles proposés : ils sont suffisamment flexibles pour décrire les contours considérés.

7. Conclusion

Dans cet article, nous avons proposé des modèles paramétriques associés à une méthode de segmentation permettant d'extraire automatiquement les contours des principaux traits du visage. Les contours ainsi extraits sont suffisamment réalistes pour envisager de les utiliser dans des applications conduisant à une interprétation de plus haut niveau telle par exemple la reconnaissance des émotions faciales.

Références

- [1] P.S. ALEKSIC, J.J. WILLIAMS, Z. WU et K. KATSAGGELOS, « Audio-Visual Speech Recognition Using MPEG-4 Compliant Visual Features », *EURASIP Journal on Applied Signal Processing, Special Issue on Joint Audio-Visual Speech Processing*, September 2002, pp.1213-1227.
- [2] W. BEAUDOT, The neural information processing in the vertebrate retina: a melting pot of ideas for artificial vision. *Phd thesis, tirflaboratory*, Grenoble, France, 1994.
- [3] M.O. BERGER, R. MOHR, «Towards Autonomy in Active Contour Models ». In *Proc. ICPR'90*, June 1990, pp.847-851.
- [4] Base de séquences vidéos émotions de Kanade et Cohn URL: <http://www.cs.cmu.edu/~face>
- [5] T. COIANIZ, L. TORRESANI, B. CAPRILE, «2D Deformable Models for Visual Speech Analysis». In *NATO Advanced Study Institute: Speech reading by Man and Machine*, 1995, pp.391-398.
- [6] N. EVENO, A. CAPLIER, P.Y. COULON, «A new color transformation for lip segmentation». In *Proc. IEEE MSSP'01*, Cannes, France, September 2001.
- [7] N. EVENO, A. CAPLIER, P.Y. COULON, A parametric model for realistic lip segmentation. *International Conference on Control, Automation, Robotics and Vision (ICARV'02)*, Singapore, December 2002.
- [8] N. EVENO, A. CAPLIER, P.Y. COULON, «Jumping snakes and parametric model for lip segmentation». *International Conference on Image Processing*, Barcelone, Espagne, Septembre 2003.
- [9] N. EVENO, « Segmentation des lèvres par un modèle déformable analytique ». *Thèse de l'INPG*, novembre 2003.
- [10] The Database of Faces, Cambridge University Department, <http://www.uk.research.att.com/facedatabase.html>
- [11] Z. HAMMAL, A. CAPLIER, M. ROMBAUT, « Classification des émotions faciales par la théorie de l'évidence ». *12^{ème} Colloque sur la Logique Floue et ses Applications*, Nantes, France, novembre 2004.
- [12] M.E. HENNECKE, K.V. PRASAD, D.G. STORCK, «Using Deformable Templates to Infer Visual Speech Dynamics ». In *Proc. 28th Annual Asilomar Conference on Signals, Systems and computers*, 1994, pp. 578-582.
- [13] H. HJELMÅS, B. LOW, « Face detection: a survey ». *Computer Vision and Image Understanding*, 83, 2001, pp. 236-274.
- [14] A. HULBERT, T. POGGIO, « Synthesizing a Color Algorithm from Examples ». *Science*, Vol.239, 1998, pp.482-485.
- [15] M. KASS, A. WITKINS, D. TERZOPOULOS, « Snakes: Actives Contours Models », *International Journal of computer vision*, 1(4), January 1988, pp.321-331.
- [16] J. LUETTIN, N.A. TRACKER et S.W. BEET, « Active Shape Model for Visual Speech Feature Extraction » University of Sheffield, U.K., Electronic System Group Report N°95/44, 1995.
- [17] A. NEFIAN, L. LIANG, X. PI, L. XIAOXIANG, C. MAO et K. MURPHY, « A couple HMM for Audio-visual Speech Recognition ». In *Proc. ICASSP'02*, 2002, pp.2013-2016.
- [18] K. SOBOTTKA and I. PITAS, « Looking for Faces and Facial Features in Color Images », *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications, Russian Academy of Sciences*, Vol. 7, No. 1, 1997.
- [19] D. TERZOPOULOS et K. WATERS, « Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models ». *IEEE Trans. On PAMI*, 15(6), June 1993, pp.569-579.
- [20] Y. TIAN, T. KANADE, J. COHN, « Robust Lip Tracking by Combining Shape, Color and Motion ». *Proc ACCV'00*, 2000.
- [21] Y. TIAN, T. KANADE, and J. COHN, « Dual state Parametric Eye Tracking ». *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, pp. 110-115, March 2000.
- [22] N. TSAPATSOUKLIS, Y. AVRITIS, S. KOLLIAS, « Efficient Face Detection for Multimedia Applications ». In *Proc. ICIP'00*, Vancouver, Canada, September 2000.
- [23] R. WANG and Y. WANG, « Facial Feature Extraction and Tracking in Video Sequences ». *IEEE Signal Processing Society 1997 Workshop on Multimedia Signal Processing* June 23 – 25, 1997, Princeton, New Jersey, USA Electronic Proceedings. pp. 233-238.
- [24] M.H. YANG, N. AHUJA, « Gaussian Mixture Model for Human Skin Color and Its Application in Image and Video Database ». In *Proc. Of the SPIE: Conf. On Storage and Retrieval for Images and video Databases*, vol.3656, 1999, pp.458-466.
- [25] 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.
- [26] A. YUILLE, P. HALLINAN et D. COHEN, « Feature Extraction from faces using deformable templates ». *Int. Journal of computer Vision*, 8(2), 1992, pp.99-111.
- [27] X. ZHANG, R.M. MERSERAU, M.A. CLEMENTS, C.C BROWN, « Visual Speech Feature Extraction for Improved Speech Recognition ». In *Proc. ICASSP'02*, 2002, pp. 1993-1996.



Zakia **Hammal**

Zakia Hammal est diplômée de l'institut national d'informatique d'Alger. Elle a obtenu un DEA en Algorithmique et intelligence artificielle à Caen. A l'heure actuelle, elle est en 3ème année de thèse au laboratoire des images et des signaux de Grenoble. Son travail porte sur l'analyse et la reconnaissance des gestes faciaux.



Nicolas **Eveno**

Nicolas Eveno est diplômé en Physique Appliquée de l'ENS de Cachan. Il a obtenu un DEA puis un doctorat de l'INP-Grenoble en traitement d'images. Ses centres d'intérêts se situent dans l'analyse du geste humain et couvrent notamment les problèmes de détection, de suivi et d'identification de personnes dans une scène. Il est actuellement post-doctorant au laboratoire CLIPS de Grenoble et travaille sur la biométrie faciale bimodale.



Alice **Caplier**

Alice Caplier est maître de Conférences à l'INPG depuis 1997 et effectue sa recherche au LIS. Ses centres d'intérêt portent sur l'analyse et l'interprétation haut niveau du mouvement humain avec comme cadre d'application le développement d'interfaces multi-modales. Ses recherches portent plus précisément sur l'analyse et la reconnaissance des gestes faciaux (expressions faciales, mouvement rigide de la tête) ainsi que sur l'analyse de comportements humains.



Pierre-Yves **Coulon**

Pierre-Yves Coulon a obtenu un doctorat en Automatique en 1982 et il est depuis 1996 Professeur à l'Institut National Polytechnique de Grenoble (INPG). Ses travaux de recherche ont porté successivement sur la poursuite de cible, les architectures matérielles et la segmentation d'images. Actuellement chercheur au Laboratoire des Images et des Signaux (LIS), son activité est orientée vers la segmentation du visage, notamment les lèvres, par contours actifs et modèles d'apparence.