

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

Z. HAMMAL, N.EVENO, A.CAPLIER, PY.COULON

LIS, INPG, 46 avenue Félix Viallet, 38031 Grenoble Cedex

nom@lis.inpg.fr

Résumé – 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 plaquer au mieux sur les contours des traits présents dans 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. Mais, 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 système de reconnaissance des dynamiques émotionnelles.

Abstract – 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 process (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 (luminance or chrominance) through the estimated contour is maximized because at each point of the searched contour, the gradient (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 procure realistic contours for the mouth, the eyes and eyebrows. This facial features segmentation is the first step of a global system of facial emotions recognition.

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 mal-entendantes. 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 en se focalisant plus particulièrement sur le problème de la reconnaissance automatique des émotions sur un visage.

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.

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.

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 (souvent à base de paraboles) [1,2] ne sont pas suffisamment précis pour permettre d'obtenir des contours réalistes. Nous proposons alors les modèles suivants (cf. figures 1 et 2) :

- ✓ Pour les yeux : 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.
- ✓ 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) ;
- ✓ Pour les lèvres : un ensemble de 4 cubiques (γ_i) passant chacune par deux points Q_i caractéristiques et auxquelles on impose une contrainte de dérivée nulle en Q_2, Q_4 ou Q_6 ([3]). Par exemple, on impose à la cubique passant par Q_1 et Q_2 d'avoir une dérivée nulle en Q_2 .

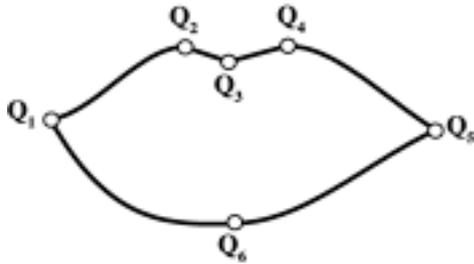


Figure 1 : modèle choisi pour la bouche.

L'utilisation de courbes de Bézier pour les yeux et les sourcils permet de s'affranchir de la contrainte de symétrie verticale trop restrictive imposée par une parabole. L'utilisation de cubiques pour les lèvres permet de rendre compte d'une grande variabilité de formes différentes pour les lèvres (cf. figure 5).

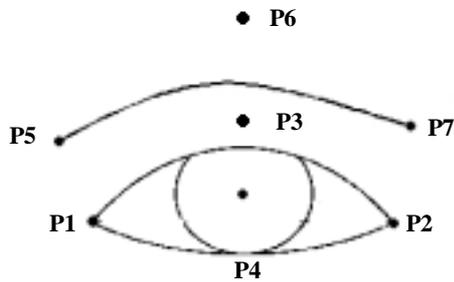


Figure 2 : modèle choisi pour l'œil et le sourcil.

3. Pré-traitement

Dans la phase de pré-traitement, nous nous affranchissons des variations d'illumination par l'utilisation d'un filtrage adapté inspiré du comportement de la rétine [4]. La figure 3 présente un schéma bloc du filtrage réalisé où G est un filtre de type gaussien.

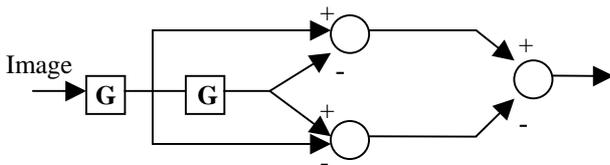


Figure 3 : filtre de variations d'illumination

Grâce à un double seuillage sur des informations de chrominance (spécificité des points de peau dans l'espace C_b, C_r [6]), nous extrayons un rectangle englobant le visage. Sur cette boîte englobante, le contour de l'iris droit (resp. gauche) est recherché dans la partie supérieure droite (resp. gauche) sous la forme d'un cercle constitué de points de gradient de luminance maximal (le contour de l'iris représente la frontière entre l'iris et le blanc de l'œil). Dans toute la zone de localisation de chaque iris, on recherche le contour de type circulaire qui maximise la quantité :

$$E = \sum_{p \in C} \bar{\nabla} I(p) \cdot \bar{n}$$

Avec $I(p)$ luminance au point p , C le cercle cherché et n la normale au contour au point p . En effet, en chacun des points du cercle cherché, le gradient de luminance est normal au contour.

La localisation initiale du contour de la lèvre supérieure est effectuée grâce à un nouveau type de contour actif ouvert : le *jumping snake* (voir [7] pour les détails de la définition et de la mise en œuvre). Contrairement aux *snakes* classiques, il peut être initialisé assez loin du contour final et le réglage de ses paramètres est simple et intuitif. Sa convergence est une alternance de phases de croissance à partir d'un germe et de phases de saut. Le germe initial est positionné au-dessus des lèvres en se servant de la position des yeux. Puis, le *jumping snake* croît et se déplace progressivement vers les lèvres. A chaque étape de la progression, on s'assure que le *jumping snake* trouvé maximise le flux à travers le contour d'une information de gradient hybride tenant compte de I_N (luminance normalisée dans $[0,1]$) et h_N (pseudo-teinte normalisée dans $[0,1]$) définie par :

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

A la fin du processus, il coïncide avec le contour supérieur de la bouche.

4. Initialisation

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

On extrait le coin extérieur C_2 (resp. intérieur C_1) de l'œil grâce à un suivi vers l'extérieur (resp. vers l'intérieur) des points de gradient de luminance maximal en partant du point X_2 (resp. X_1) déterminé par translation de quelques pixels à gauche (resp. à droite) à partir de la position du point le plus bas du cercle détecté pour l'iris (cf. figure 4). Le point initial pour la recherche de chaque coin ne peut pas appartenir au cercle associé à l'iris car ce cercle est lui-même constitué de points de gradient maximal.

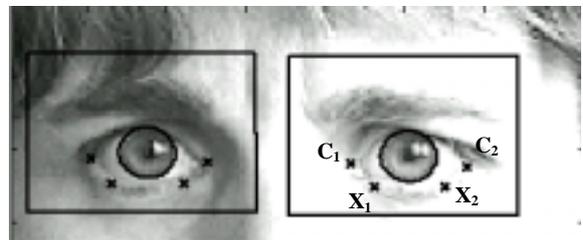


Figure 4 : détection des coins des yeux.

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.

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 définies 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 maxima 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 les abscisses des deux maxima 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 :

$$\begin{cases} (x_1 + x_2) / 2 \\ y_2 \end{cases}$$

servent de points de contrôle initiaux pour la courbe de Bézier associée à chaque sourcil.

4.3 Cas de la bouche

On extrait automatiquement les deux coins et les deux sommets de l'arc de Cupidon en utilisant des gradients hybrides ([7]) et en plaquant le modèle sur le *jumping snake* final obtenu à l'issue du pré-traitement (cf. section 3).

5. Evolution

La parabole associée au contour inférieur de l'œil étant bien initialisée, elle n'est plus modifiée par la suite. La courbe de Bézier associée au contour supérieur de l'œil est déformée vers le haut par déplacement successif de chacun des 3 points de contrôle. De même que pour le cercle associé à l'iris, la courbe définitive retenue est celle qui maximise le flux du gradient de luminance à travers le contour sachant que le contour supérieur correspond à la frontière entre le blanc de l'œil et la peau plus sombre. Pour la déformation de la courbe associée à chaque sourcil, on utilise la même méthode sachant que le contour du sourcil représente la frontière entre la peau et le sourcil plus sombre.

On déforme ensuite les cubiques associées aux lèvres de manière à maximiser le flux associé au gradient hybride défini dans la section 3 [7].

6. Résultats

6.1 Efficacité des modèles et de la segmentation

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 séquence de visages avec émotions de Kanade et Cohn [8]. 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.

La figure 5 permet de juger de la pertinence du modèle choisi pour le contour extérieur des lèvres. 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 de manière réaliste ce contour même dans le cas de bouches très déformées (cas d'une grimace sur la figure 5). 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 visuelle est de bonne qualité (en particulier, contours extraits fidèles).



Figure 5 : contours extraits pour la bouche dans le cas d'une grimace.

La figure 7 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.

6.2 Application à la reconnaissance d'émotions

Les contours extraits étant suffisamment réalistes, nous les utilisons dans un système (en cours de développement) d'analyse des émotions sur le visage. La classification ou reconnaissance d'une émotion se base l'évolution temporelle des informations contenues dans le « squelette » issu de cette étape de segmentation. La figure 6 donne un exemple de squelette émotionnel dans le cas de la joie. Les squelettes issus de la phase de segmentation pour deux images non consécutives sont présentés. On constate que la joie se caractérise essentiellement par une extension des dimensions du contour associé à la bouche et un plissement des yeux.

[8] Base de séquences vidéos émotions de Kanade et Cohn
URL : <http://www.cs.cmu.edu/~face>

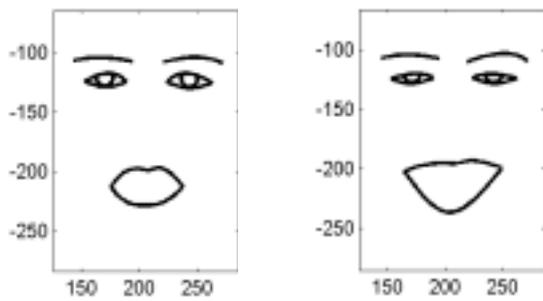


Figure 6 : squelette d'émotion

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 analyser afin d'en déduire une interprétation de plus haut niveau telle par exemple une reconnaissance des émotions faciales.

Références

- [1] 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.
- [2] 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.
- [3] 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.
- [4] W. Beaudot. *The neural information processing in the vertebrate retina : a melting pot of ideas for artificial vision*. Phd thesis, tirt laboratory, Grenoble, France, 1994.
- [5] N. Eveno, A. Caplier, P.Y. Coulon. *Key Points Based Segmentation of Lip*. International Conference on Multimedia and Expo (ICME'02), Switzerland, August 2002.
- [6] V. Girondel, L. Bonnaud, A.Caplier. *Hands Detection and Tracking For Interactive Multimedia Applications*, International Conference on Computer Vision and Graphics, 25-29 September, Zakopane, Poland, 2002.
- [7] N. Eveno, A. Caplier, P.Y. Coulon. *Jumping snakes and parametric model for lip segmentation*. International Conference on Image Processing, Barcelone, Espagne, Septembre 2003.

Figure 7 : segmentation des traits du visage.

