Contours Actifs Adaptés pour la Labiométrie

3 avril 2000

P. Delmas P.Y Coulon et V. Fristot
Laboratoire des Images et des Signaux, Institut National Polytechnique de Grenoble, LIS, INPG, 46 av. Félix-Viallet, 38031 Grenoble Cedex, France
email : delmas,coulon,fristot@tirf.inpg.fr fax : +33 (0)4 76 57 47 90

Résumé

Les contours actifs sont largement utilisés, depuis une dizaine d'années, en segmentation d'images pour leur capacité à intégrer les processus de détection et de chaînage des contours en un seul processus de minimisation d'énergie. Toutefois l'estimation des paramètres et les problèmes d'initialisation font des contours actifs une méthode difficile à calibrer. Les performances des contours actifs sont généralement améliorées par une initialisation proche des solutions désirées. Nous développons ici un algorithme d'extraction de zone d'intérêt, centrée sur la bouche, exploitant des informations locales de type niveaux de gris et gradient. Nous présentons ensuite un algorithme original de contours actifs. Notre méthode utilise des paramètres de rigidité et tension variables spatialement ce qui permet au contour actif de conserver globalement une forme de bouche au cours de son évolution temporelle. Nos expérimentations sur une banque d'images étendues démontrent la robustesse de nos algorithmes de détections de coins et de contours actifs "adaptés". Les applications principales de nos travaux sont la visiophonie haute qualité bas-débit, les services multimédias et les réalités virtuelles (restitution de parole audiovisuelle par avatars, reconnaissance robuste de la parole ...).

1 Introduction

Nos travaux font partie du projet LABIOPHONE, plate-forme de communication audiovisuelle visant l'acquisition automatique de contours labiaux d'un locuteur et l'exploitation de paramètres caractéristiques de ces contours par des algorithmes d'animation d'acteurs de synthèse. La caméra est solidaire d'un casque fixé sur la tête du locuteur ce qui permet de maintenir le visage centré dans l'image. Ce projet, impliquant plusieurs laboratoires (Institut de la Communication Parlée-(ICP), Laboratoire des Images et des Signaux-(LIS)), est développé par la fédération ELESA n°8 (CNRS/INPG). Notre objectif est d'arriver à extraire dans un temps relativement court, à terme en temps réel, les caractéristiques labiales d'un locuteur. Les informations obtenues devront être

suffisamment fines pour permettre une reconstitution réaliste des mouvements et de la physionomie de la zone labiale. Pour leur capacité à segmenter de façon précise des objets tout en gardant une certaine cohérence géométrique, les contours actifs ont été choisis pour mener à bien nos objectifs. Introduits par Kass et al. [6], les contours actifs ou "snakes" ont été conçus en tant que processus semi-automatique de segmentation visant, à l'aide d'une interface graphique, une reconnaissance guidée des différentes formes présentes dans une image.

Nous proposons ici un algorithme d'extraction de contour des lèvres robuste vis à vis du changement de locuteur. Une phase de pré-segmentation, permettant de localiser précisément les commissures ainsi que les limites verticales et horizontales de la bouche, est effectuée à partir d'informations locales sur les niveaux de gris de l'image. Les contours actifs sont ensuite appliqués à partir de la position initiale précédemment définie et optimisés afin de conduire à une détection fine des contours des lèvres avec une complexité calculatoire la plus faible possible.

L'article se décompose en trois parties : les contours actifs sont présentés ainsi que les différents compléments théoriques et techniques apportés au modèle initial dans la première partie. La seconde traitera de la phase de détection des extréma (coins et limites horizontales) de la bouche, nécessaire à une bonne initialisation des contours actifs. Nous montrerons ensuite dans la dernière partie les résultats comparés de deux types de contours actifs ainsi que quelques améliorations introduites.

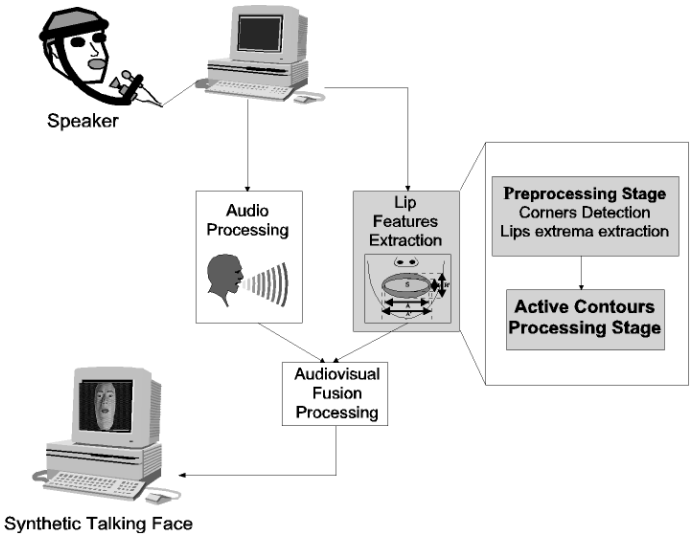

FIG. 1: *Le labiophone : A partir de l'image d'un locuteur, le suivi des lèvres fournit des paramètres pertinents pour la synthèse d'un visage parlant.*

2 Les contours actifs pour la détection des lèvres

Les contours actifs (ou "snakes") évoluent au gré de la minimisation de la fonctionelle d'énergie, $\phi(v)$, qui leur est associée. Ils se déplacent au sein de l'image d'une position initiale vers une configuration finale qui dépendra de l'influence respective des divers termes d'énergie en présence. L'énergie des "snakes" comprend un terme d'énergie interne appelé énergie de régularisation et d'un terme d'énergie externe ou d'adéquation aux données.

Soit s l'abscisse curviligne et v(s) la position d'un point sur la courbe décrivant un contour actif.

$$v(s) = [x(s), y(s)], \quad s\epsilon[0,1]. \tag{1}$$

$$\phi : v(s) \longrightarrow \int_0^1 \left(E_{int}(v(s)) + E_{ext}(v(s))\right) ds \tag{2}$$

L'énergie interne dérive d'un stabilisateur de Tikhonov d'ordre 2 qui prend en compte les déformations de courbure et de tension de la courbe au travers des coefficients α et β.

$$E_{int}(s) = \alpha(s)|v'(s)|^2 + \beta(s)|v''(s)|^2 \tag{3}$$

$$E_{ext}(s) = -\left|\nabla \left(G_\sigma \otimes H\right)(v(s))\right|^2 \tag{4}$$

G_σ représente un filtre gaussien 2D, H le plan image teinte et ∇ l'opérateur gradient. Le terme d'énergie externe prend en compte les informations liées à l'image. Ici, il s'agira d'informations de type contours i.e. de type gradient. La minimisation de cette énergie par la méthode d'Euler-Lagrange et la discrétisation par différences finies de l'équation différentielle obtenue conduit à la modélisation de l'évolution des contours actifs sous la forme matricielle classique :

$$V(t) = (A + \gamma I_d)^{-1} \left(\gamma V(t-1) - F(V(t-1))\right) \tag{5}$$

I_d est la matrice identité, A matrice de toeplitz, V le vecteur contenant les points du snake, F l'ensemble des forces externes qui dérivent de l'énergie externe, et $\frac{1}{\gamma}$ le pas d'évolution temporelle.

2.1 Principales améliorations

Plusieurs améliorations ont été apportées depuis, à la formulation traditionnelle des contours actifs. Berger [1] a étudié la stabilité des "snakes" par l'étude de son conditionnement. Son calcul des valeurs propres de la matrice d'évolution des contours actifs a permis de donner des intervalles de confiance pour les coefficients d'élasticité (α) et de rigidité (β). L.D Cohen [2] a introduit une force externe supplémentaire dite "force de pression" qui permet d'accélérer la convergence du "snake" initialisé loin des zones d'intérêt, et ainsi de s'affranchir des zones de gradients parasites. Cette force, suivant son orientation, permet de plus, aux contours actifs, d'acquérir des propriétés de dilatation. Toutefois le réglage du paramètre de cette force est délicat et alourdit d'autant le processus

d'optimisation du snake. Chen et al. [9] ont proposé une nouvelle force externe dite "gradient vector flow" visant une plus grande indépendance du positionnement initial du "snake" et de lui conférer des propriétés de dilatation. Ceci est obtenu par diffusion du gradient de l'image. Toutefois ce processus est très lent et peut poser des problèmes d'interférences dans le cas de gradients parasites de forte intensité. Gun et al. [5] font converger l'un vers l'autre deux contours actifs. Le premier est initialisé à l'intérieur de l'objet, l'autre à l'extérieur. Une énergie mutuelle assure la convergence des "snakes" l'un vers l'autre. Cette méthode fournit des résultats intéressants mais nécessite un objet sans contours multiples (ce qui n'est pas le cas de la bouche) ainsi qu'une connaissance a priori de l'intérieur et de l'extérieur de l'objet.

2.2 Adaptation du "snake" à un modèle morphologique

En théorie les coefficients α et β sont variables le long de la courbe caractérisant les contours actifs. Pour un point donné sur la courbe, ils sont liés aux informations issues de l'image (niveaux de gris, gradient) et à la courbure (pour β) ou tension (α) de la courbe autour de ce point [3]. Ainsi Gao et al. [4] ont déterminé empiriquement les relations liants les coefficients α et β à la distance entre les points du snake et à la courbure en ces points. En pratique, la plupart des études sur les contours actifs ont été effectuées avec des coefficients constants. En effet, les objets recherchés ne sont pas forcément géométriquement connus. Ici la structure des lèvres, bien que variable selon les individus, garde la même forme générale. Nous utiliserons donc des "snakes" à coefficients variables spatialement afin de s'adapter au mieux aux formes recherchées. Ainsi β sera pris plus grand au centre de la lèvre inférieure que vers les bords. Les points extrémités seront choisis fixes et à β nul afin de favoriser les discontinuités de courbure. Pour la lèvre supérieure, β sera choisi symétrique par rapport à l'arc de cupidon, à valeur forte dans les zones de forte courbure (arc de cupidon) et faible proche des commissures.

2.3 Mise en œuvre de l'algorithme

Les contours actifs nécessitent quelques améliorations afin de converger rapidement vers les solutions désirées. Le caractère discret des images n'est pas favorable à la stabilité de cette méthode. Nous avons donc choisi d'effectuer une interpolation linéaire pour gérer des valeurs non entières des points de contrôle du snake. La condition d'équidistance des points du "snake" est implicitement imposée lors de la caractérisation théorique des contours actifs. En pratique, le "snake" doit être ré- échantillonné périodiquement pour éviter une accumulation de ses points sur les zones de fort gradient. Un rééchantillonnage par fonction spline est donc effectué toutes les N itérations. Une autre faiblesse des contours actifs est leur instabilité en position d'équilibre final. Les "snakes" ont tendance à osciller autour de leur position d'équilibre, rendant ainsi difficile la caractérisation de leur convergence. Nous avons choisi un critère de distance quadratique entre deux ré-échantillonnages successifs. Soit ϵ le critère d'arrêt :

$$\epsilon = \sum_{i \in [0..N-1]} |V_i(t) - V_i(t - N_{ech})| \quad (6)$$

N_{ech} est le pas temporel de ré-échantillonnage du "snake" et N le nombre de points de contrôle du snake.

3 Phase d'initialisation

La phase d'initialisation est la première étape nécessaire à l'utilisation des "snakes" . En effet, pour obtenir une convergence rapide vers les formes désirées, les contours actifs nécessitent la meilleure initialisation possible. Généralement, elle prend en compte des informations provenant de l'image et/ou des objets à détecter. Par exemple, Radeva [8] propose une localisation précise des différents éléments constitutifs du visage à l'aide de projections horizontales et verticales des lignes et colonnes de l'image. Toutefois, cette méthode n'est pas robuste vis à vis des changements de locuteurs ou des conditions d'éclairage. Nous allons développer ici une méthode de détection des commissures et des extrema horizontaux de la bouche, fondée sur la recherche des minima du plan Luminance et du gradient associé, qui nous aidera à positionner le "snake" initial au plus près des lèvres.

3.1 Prétraitement

L'acquisition des images a été obtenue à cadence vidéo (25 images couleurs tramées par seconde) à partir d'une caméra couleur monoCCD fixe par rapport au visage. Les acquisitions ont été faites sous condition d'éclairage naturel, sans maquillage ni marqueur.

Notre but est d'extraire les frontières des lèvres. Les informations du type frontière étant plus pertinentes dans le plan teinte que dans le plan luminance, nous avons décidé d'utiliser l'information gradient issue du plan teinte pour extraire les contours de l'image. Nous utiliserons pour cela un filtre gradient classique (de type Sobel ou Canny-Deriche). Le plan Teinte est obtenu à partir de la transformée logarithmique de l'espace des couleurs (RGB vers HI)[7].

3.2 Détection des extrema des lèvres

Les zones les plus sombres du plan luminance apparaissent au niveau des commissures de la bouche ainsi qu'à l'intérieur de la bouche, qu'elle soit ouverte ou fermée. Il apparait donc intéressant de déterminer les minima de niveaux de gris sur le plan Luminance le long des verticales de l'image. Afin de tenir compte de la position verticale centrée de la bouche et de sa symétrie, nous avons introduit une fonction de pondération favorisant les minima verticaux proches du centre de l'image plutôt que ceux situés sur les bords (fig. 2).

$$\zeta_j = e^{-4(1-j/N_{col})^2} \quad (7)$$

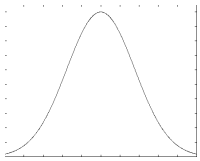

FIG. 2: *fonction de pondération*

On construit un vecteur d'accumulation appelé V_{RI}, projections pondérées des minima par colonne. Considérons que pour une colonne j le minimum de luminance se trouve à la ligne i. La composante i du vecteur V_{RI} est alors incrémentée d'une valeur ζ_j correspondant à la pondération de la colonne j. La composante de plus forte amplitude de V_{RI} nous donne alors la position horizontale de la bouche. Pour obtenir les commissures, on utilise les minimums locaux par colonnes. On effectue alors un chaînage horizontal de ces minimums locaux en partant du centre de la bouche vers les extrémitées (commissures des lèvres) (fig. 3). Une région d'intérêt est alors déterminée (largeur : distance entre les commissures, rapport hauteur/largeur = 2/3). Bien qu'incomplètes, les lignes de gradient issues des contours extérieurs des lèvres peuvent nous aider à déterminer les limites verticales de la bouche. Pour cela, le gradient est seuillé par la valeur moyenne calculée dans la région d'intérêt (Fig. 4). La projection de quelques colonnes centrales du gradient nous fournira les extrema de la lèvre supérieure (premier pic vers le haut à partir du pic central), respectivement inférieure (premier pic vers le bas à partir du pic central). Une parabole (pour la lèvre inférieure) et deux quartiques (pour la lèvre supérieure) fournissent une première approximation du contour extérieur de la bouche à partir des positions extrêmes de la bouche précédemment calculées (fig. 5). L'initialisation du "snake" se fera alors par échantillonnage le long de ces courbes.

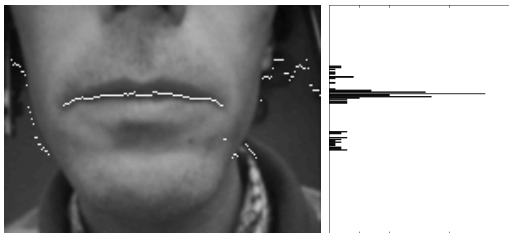

FIG. 3: *A gauche : Minima (en blanc) issus des verticales sur une image a niveaux de gris. A droite : Le vecteur RI (en noir) qui indique le positionnement horizontal de la bouche.*

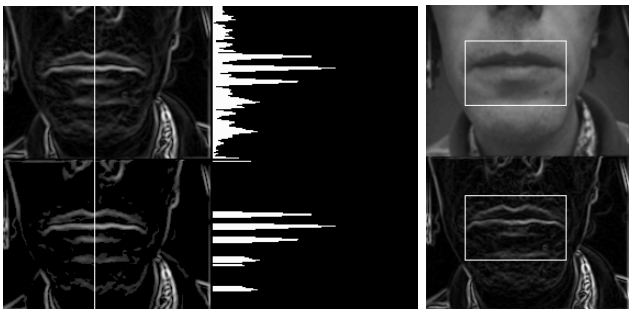

FIG. 4: *A gauche : Haut : gradient de l'image et projection des colonnes centrales. Bas : gradient seuillé et projection correspondante. A droite : région d'intérêt positionnée sur l'image à niveaux de gris (haut) et sur l'image gradient (bas).*

FIG. 5: *Initialisation du "snake" à partir des résultats de la pré-segmentation.*

4 Résultats Expérimentaux

Notre recherche de coins et limites verticales des lèvres est associée à l'algorithme de contours actifs. Les tests sont effectués, à coefficients α et β constants ou variables spatialement, sur plusieurs locuteurs, bouches ouvertes ou fermées. Le "snake" est initialisé à proximité des lèvres grâce à notre segmentation préalable.

4.1 Sensibilité à l'initialisation

Un problème important subsiste une fois les paramètres de réglage des contours actifs "optimisés" : le positionnement initial. Quand un contour actif est positionné trop loin des formes à segmenter, il peut rencontrer des zones de gradient parasite qui vont l'empêcher de converger. De même, initialisé trop à l'intérieur des objets, le contour actif sera incapable de se dilater suffisamment pour atteindre les contours désirés. Les figures suivantes montrent les résultats de mauvaises initialisations ainsi que les problèmes occasionnés par les zones d'ombres. Une zone de gradient parasite est présente sous la lèvre inférieure : elle empêchera la convergence du contour actif vers les bonnes frontières. Les résultats sur Benny (fig. 6 à droite) sont bons pour la lèvre supérieure et mauvais pour la lèvre inférieure. Pour Nico (fig. 6 a gauche), une initialisation trop à l'intérieur de la lèvre inférieure produit une convergence vers les dents (zone

de fort gradient) (fig. 6). Ces deux types de problèmes pourront être réglés par l'utilisation de la teinte en lieu et place de la luminance et par une recherche d'une initialisation toujours plus proche des lèvres.

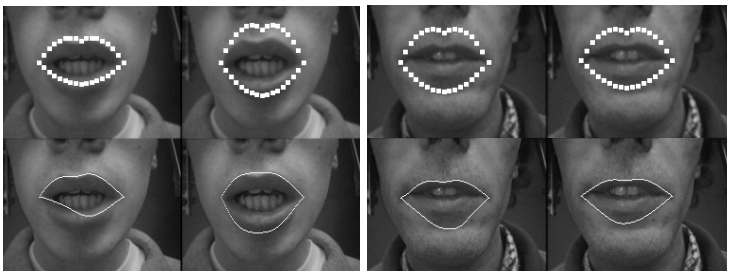

FIG. 6: Mauvaise intialisation des "snakes" et résultat sur l'utilisation du plan Luminance pour Nico (à gauche) et Benny (à droite).

4.2 Snake adapté vs Snake traditionnel

Nous comparons ici un contour actif à coefficients variables spatialement avec un autre à coefficients constants. Le premier possède des coefficients (spatialement) adaptés à la géométrie des lèvres qui ne seront pas modifiés selon les images traitées. Les coefficient du "snake" traditionnel sont constants spatialement mais seront ajustés manuellement pour chaque série d'images afin d'être le plus performant possible. Le "snake" traditionnel converge dans quasiment tous les cas vers les bons contours des lèvres. Il n'arrive pas à détecter l'arc de cupi-

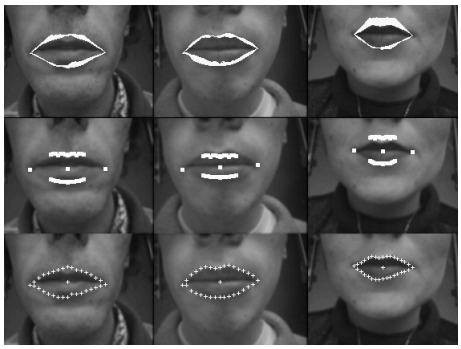

FIG. 7: *Convergence du "snake" avec des paramètres fixés manuellement. Haut : Minima des colonnes de l'image. Milieu : Initialisation du contour actif. Bas : Résultat après convergence.*

don sur l'image Benny (Fig. 7 gauche) mais il détecte correctement les contours extérieurs de la bouche dans les autres images. Le "snake" adapté converge correctement et plus rapidement sur les images testées (Fig. 8), la différence

de rapidité pouvant aller de quelques pourcents à plus de 50%. Il semble donc qu'un contour actif à coefficients variables soit plus rapide et mieux adapté aux formes que nous recherchons. De plus nous avons testé nos algorithmes sur des bouches ouvertes en conservant les mêmes coefficients de réglage du snake. La détection des extrema de la bouche et l'application des contours actifs ont permis de détecter correctement les contours extérieurs des lèvres (Fig. 8).

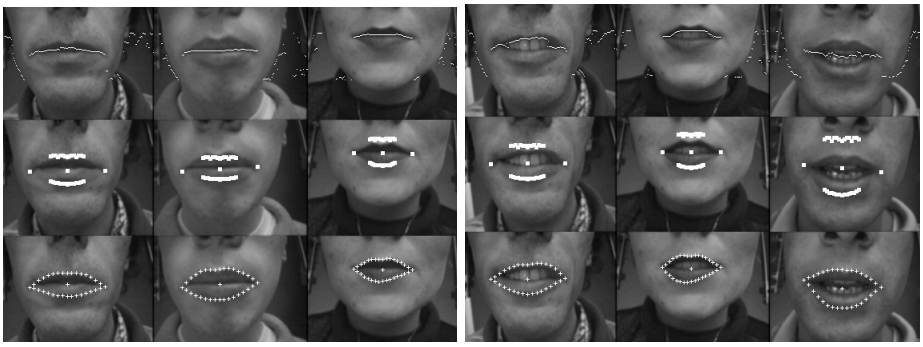

FIG. 8: *A gauche : Convergence des contours actifs adaptés sur des bouches fermées. A droite : Convergence des contours actifs adaptés sur des bouches ouvertes. Haut : Minima des colonnes de l'image. Milieu : Initialisation du contour actif. Bas : Résultat après convergence.*

4.3 Résultats statistiques

Nous avons testé les algorithmes de détection des commissures et de contour actif sur les 130 images de notre base (6 locuteurs différents, 84 images de bouches ouvertes et 46 de bouches fermées). Nous avons obtenu 98% de bons résultats en détection des coins de la bouche sur notre base et 90% de réussite dans l'estimation de la Région d'Intérêt de la bouche ; à savoir : une détection des coins des lèvres, une bonne estimation de l'orientation de la bouche et un positionnement initial du "snake" correct. Il faut noter que notre base contient des séquences sous-éclairées ou avec des bouches fortement distordues afin de valider nos algorithmes dans des conditions "difficiles". En faisant évoluer notre algorithme de contours actifs à partir de l'initialisation précédement établie, nous obtenons 82% de convergence sur les frontières des lèvres. La plupart des échecs sont dus à des "snakes" initialisés trop loin de la bouche et attirés vers d'autres zones de gradient du visage (nez, fossette du menton). La plupart des problèmes rencontrés ont été résolus à l'aide d'une initialisation plus fine des "snakes" autour des lèvres obtenus à partir de masques spatio-temporels (Fig. 10).

4.4 Initialisation par segmentation spatio-temporelle

Nos algorithmes de contours actifs et de détection de coins ont été associés aux travaux d'un autre doctorant du laboratoire portant sur la segmentation spatio-temporelle des lèvres par approche région [7]. L'approche statistique globale utilisée permet l'extraction de masques des zones des lèvres en mouvement (Fig. 9). Les masques obtenus sont robustes vis à vis du changement de locuteur et de l'éclairage. Toutefois, ceux ci ne sont pas précis dans la détection des commissures. Notre algorithme de détection des commissures des lèvres est associé aux masques spatio-temporels. Un détecteur de contours est appliqué sur les masques. Les contours déterminés sont alors échantillonnés et reliés aux coins des lèvres précédemment détectés. Ce processus fournit aux contours actifs une initialisation très précise sur les limites intérieures et extérieures des lèvres (Fig. 10 haut et milieu). Cette fusion des deux méthodes nous a permis d'améliorer significativement la robustesse de notre algorithme de détection des lèvres. Nous obtenons alors 98% de réussite dans le positionnement initial du snake. La convergence sur les contours intérieurs et extérieurs des lévres se fait en moins de 100 itérations pour toutes les images. Aucun échec n'est constaté pour la détection du contour extŕieur des lèvres. La détection du contour intérieur est plus délicate en raison de la présence de la langue et des gencives à proximité des zones à segmenter. On peut observer la détection fine des contours intérieurs et extérieurs de la bouche sur 7 images successives d'un locuteur fermant la bouche (Fig. 10).

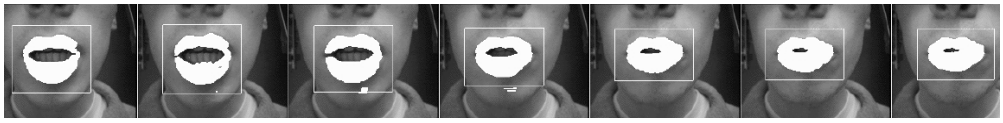

FIG. 9: *et région d'intérêt correspondante.*

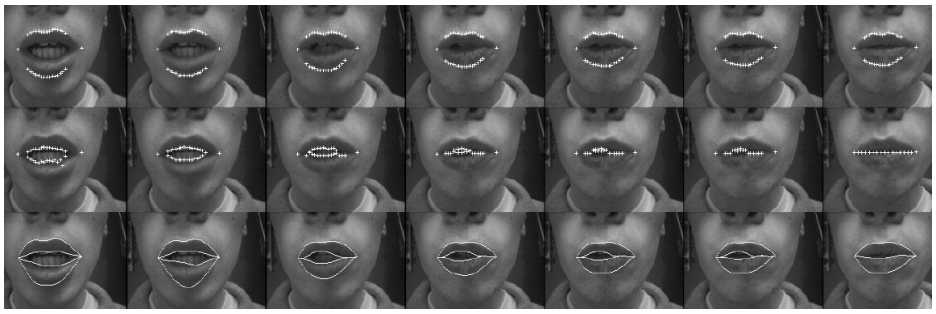

FIG. 10: *Détection des lèvres par approche conjointe. Haut : Initialisation pour le "snake" extérieur. Milieu : Initialisation pour le "snake" intérieur. Bas : Résultats après convergence.*

5 Conclusion et perspectives

Nous avons présenté un algorithme de détection des contours des lèvres intégrant une détection préalable des coins et des extréma horizontaux de la bouche. Cette localisation initiale des lèvres précède l'application d'un algorithme de contours actifs à coefficients variables spatialement, particulièrement adapté aux formes géométriques recherchées. Les résultats obtenus en gardant le même réglage des paramètres pour chaque expérimentation sont satisfaisants. La phase d'initialisation se révélant être une étape critique, nous avons combiné une approche statistique globale à notre approche locale de détection des coins de la bouche et obtenu ainsi des résultats intéressants sur la détection des contours intérieurs et extérieurs des lèvres. Quelques améliorations sont encore possibles : nous étudions notamment l'adaptation automatique des coefficients de réglage (α et β) du "snake" aux informations liées à l'image (niveau de gradient,). La détection des contours intérieurs des lèvres est rendu difficile par la présence éventuelle de la langue et/ou des dents. Une segmentation plus détaillée de la zone labiale, incluant une séparation des zones dents, langue et intérieur de la bouche est alors nécessaire.

Références

[1] Berger M.O., "Les contours actifs : modélisation, comportement et convergence". *Thèse*, Institut National Polytechnique de Lorraine, 1991.

[2] Cohen L.D., "On Active Contour Models and Balloons". In *Comp. Vis. Graph. and Image Processing*, 53(2) : 211-218, 1991.

[3] Cohen I., "Modèles Déformables 2-D et 3-D : Application à la segmentation d'images médicales". *Thèse*, Université Paris IX-Dauphine, 1992.

[4] Gao J., Kosaka A., Kak A., "A deformable model for human organ extraction". In *Proceedings of the IEEE Int. Conf. on Image Processing*, Chicago, 1998.

[5] Gunn S.R. and Nixon M.S., "Improving "snake" performance via a dual active contour". In 6^{th} *Int. Conf. CAIP'95*, Prague, 1995.

[6] Kass M., Witkins A. and Terzopoulos D., "Snakes : Active Contours Models". *International Journal of Computer Vision*, 1(4) : 321-331, 1987.

[7] Liévin M., Delmas P. et al., "Automatic lip tracking : Bayesian segmentation and active contours in a cooperative scheme". *Proc. of the* 6^th *IEEE Int. Conf. on Multimedia Computing and Systems*, Florence, Italie, 1999.

[8] Radeva P. and Marti E., "Facial Features Segmentation by Model-Based Snakes". *Int. Conf. on Comp. Anal. and Image Processing*, Prague, 1995.

[9] Xu C. and Prince J.L., "Gradient Vector Flow : A New External Force for Snakes". *IEEE Proc. Conf. on Comp. Vis. Patt. Recog.*, pp 66-71, Puerto Rico, 1997.