AUTOMATIC SNAKES FOR ROBUST LIP BOUNDARIES EXTRACTION

P. Delmas P.Y Coulon and V. Fristot

Signal and Image Laboratory, Grenoble National Polytechnical Institute,
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

ABSTRACT

Active contours or snakes are widely used in object segmentation for their ability to integrate features extraction and pixel candidate linking in a single energy minimizing process. But the sensitivity to parameters values and initialization is also a widely known problem. Performance of snakes can be enhanced by better initialization close to the desired solution. We present here a fine mouth region of interest (ROI) extraction using gray level image and corresponding gradient informations. We link this technique with an original snake method. The Automatic Snakes use spatially varying coefficients to remain along its evolution in a mouth-like shape. Our experimentations on a large image base proove its robustness regarding speakers change of the ROI mouth extraction and automatic snakes algorithms. The main application of our algorithms is video-conferencing.

1. INTRODUCTION

A wide range of papers describe the applications of active contours for object boundaries detection. Snakes, also known as active contours, are energy minimizing curves. From a given initial position, it evolves through the minimization of its energy towards image features to the desired final position. Three major problems were encountered while using snakes: initialization, parameters estimation and convergence of the algorithm. Initialization is commonly hand done, close to the object to provide good convergence. Snake evolution is sensitive to parameters values which are also evaluated manually after several tests. Snake convergence needs a good adequation between its energy and the desired image features.

Chen [6] solve most of the initialization and convergence problems for concave areas by creating a new external force called gradient vector flow. But the diffusion process required too much time and therefore has no quasi real time applications. Gun [2] propose a dual active contour: a combinaison of two snakes converging together until they coincide exactly. But this requires no inner or outer frontiers between them and the final contour. L.D Cohen [1] normalised external forces and introduced balloon forces to provide inflation of snakes or push it far away from its initial guess. This approach helps to ensure the snake convergence. Radeva [5] uses horizontal and vertical projections of a gray level image of face to find mouth location but it is not robust in regards to mouth shape changing.

We present in this paper an original algorithm for lip boundaries extraction robust to change of speakers. Our algorithm is based on unsupervised initialization of the snake called ROI mouth finding close to the desired object and an automatic snake. ROI mouth finding is achieved by the use of pixel local intensity information in the Intensity channel (extracted from RGB images) and the corresponding gradient. The automatic snake is a snake with spatially varying coefficients. Advantages of automatic snakes over traditionnal ones is their abilities to fit mouth contours through different speakers.

Our research is part of the labiophone project, an advanced audio-visual communication tool which integrates both audio and visual features. The camera is interdependant with a helmet in order to center the mouth within the image. This project, which involves several laboratories, aims to provide a very low rate coding communication system. It is supported by ELESA Federation n. 8 (CNRS-INPG).

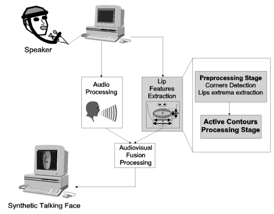

Figure 1: *Context of Lipreading.*

2. INITIALIZATION STEP

2.1. Early processing

Frame capture is achieved by a mono CCD color camera fixed w.r.t to the head. Image acquisition is made under natural light conditions and without any make-up or markers. We are focused on lip boundaries. We also observed on our image base that Hue plan provide more boundaries information than Luminance one. We then decided to use the gradient of Hue image to extract edge points. To do so we use a classical gradient filter (such as Sobel or Canny-Deriche). Hue image is obtained by logarithmic color-space transform, RGB to HI (Hue, Intensity) [4].

2.2. ROI extraction

Mouth position has been founded but corners detection and lips separation for closed mouth are still unknown. We observe that areas of darkness occured at the inner border of lips of Luminance image. It seems natural to search the vertical luminance minima to locate the frontier between lips. First we search the gray level minima pixel over image columns. We construct line size vector called ROI vector representing line repartition of column minimum pixels. We also want to encourage the vertical centered minima rather than those located at the image extremities. For each column j, if the vertical gray level minimum is found at line i then we add ζ_j (Eq. 1) to the ith coordinate of the ROI vector. The weighting coefficient ζ_j varies from 0 to 1, from the border to the center of the image.

$$\zeta_j = 4 \times \frac{j * (N_{col} - j)}{(N_{col})^2} \quad (1)$$

N_{col} is the number of column of the image. The highest peak of V_{ROI} gives the horizontal symetry axis of the mouth. The interval where lips minima are supposed to belong is deduced from the width of this peak.

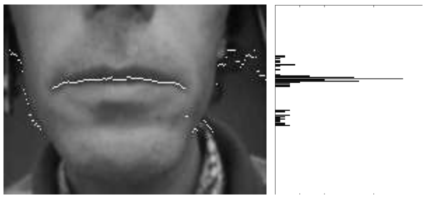

Figure 2: *Left: vertical minima (in white) on a gray level image. Right: The ROI vector (in black)*

We only have to follow, from the center of the image to the left (respectively to the right), the line of minima to find lips commisures. We finally obtain a rather good estimation of width and orientation angle of the mouth.

We treshold the gradient modulus by the mean value calculated over the ROI rectangle (height: half mouth size, width: from left to right mouth corner). We do a projection of a few mouth centered gradient column and search the first peak to the top (respectively to the bottom) of this line. This lead us to upper and lower extrema of the lips. Construction of the snake initialization is done by sampling points around lips.

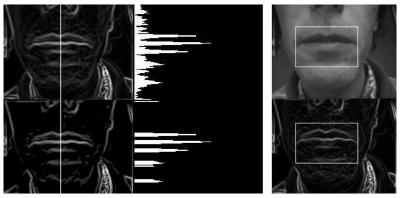

Figure 3: *Top: gradient and the projection of central verticals. Bottom: treshold gradient and the correspondant projection*

Figure 4: *Snake initialization obtained from previous figure*

3. LIP BOUNDARIES DETECTION BY ACTIVE CONTOURS

3.1. Active contours

Introduced by Kass and al [3] active contours were designed for interactive interpretation in which the user guides (by external forces modification) the snake near the desired solution. A snake is a curve v parametrically defined (Eq. 2) by its cartesian coordinates x and y along the curvilinear abscissa s which evolves through the minimization of its functionnal (Eq. 3).

$$v(s) \quad = \quad [x(s), y(s)]\,, \quad s\epsilon[0,1]. \quad (2)$$

$$F : v(s) \quad \longrightarrow \quad \int_0^1 \left(E_{int}(s) + E_{ext}(s)\right) ds \quad (3)$$

The internal energy (Eq. 4) is the regularisation term derived from ill-posed problems theory. It controls the curve smoothness via weighting parameters α and β. α controls the snake tension and β its curvature. External energy (Eq. 5) represents the adequation of image data to current vector.

$$E_{internal}(s) \quad = \quad \alpha|v'(s)|^2 + \beta|v''(s)|^2 \quad (4)$$

$$E_{external}(s) \quad = \quad -\left|\nabla\left(G_\sigma \otimes H\right)(v(s))\right|^2 \quad (5)$$

∇ represents the gradient operator, G_σ the 2D gaussian kernel and H the Hue image. This lead us to the classical dynamic scheme (Eq. 6) where I_d is the identity matrix, A the Toeplitz matrix, V the column vector of the snake points and $\frac{1}{\gamma}$ the time step coefficient.

$$V(t) = (A + \gamma I_d)^{-1} (\gamma V(t-1) - F(V(t-1))) \quad (6)$$

F represents forces derived from external energy. Higher level information forces such as Distance map or Balloon forces [1] can be added there.

3.2. Automatic snakes

Our automatic snakes integrate α et β non constants spatially. The Toeplitz matrix obtained is not detailled here. We do an LU inversion which fits to triangular inverse matrix computing. Forces calculation is done by bilinear interpolation to solve numerical instabilities which occurs through snake energy minimization. We also do a resampling by spline curve every N iterations to enforce constant distance between snake points and solve convergence problems. Finally we impose all parameters constant through different images for a given number of snake points. As mouth corners are correctly found by ROI mouth extraction we choose a snake with fixed extremities. That kind of active contours is less unstable than the traditionnal one. Test of convergence is done after each resampling. We autorize a maximum quadratic error (Eq. 7) $\epsilon = 10^{-3}$ between two consecutive snake vectors.

$$\epsilon = \sum_{i \in [0..N-1]} |V_i(t) - V_i(t-1)| \quad (7)$$

3.3. Shape constraint

Without external forces action, snake naturally shrinks to a point after several iterations conserving the minimal energy shape (a circle). Our aim is to maintain a mouth shaped snake even without external constraints. Thus we test non spacially constant snake parameters derived from physical considerations based on mouth geometry. For example the lower lip contour usually has a curvature that is minimal at the middle and maximal at the corners. We choose a higher β coefficient at the middle and a lower one around corners. We do the same kind of adjustment for the upper lip.

4. RESULTS

4.1. Comparison with usual snakes techniques

We linked our mouth ROI extraction with active contours. We test 3 different snakes on 3 different speakers. The first algorithm is the traditionnal snake with constant hand fitted parameters. For the second method we add balloon forces to help the snake deformation to fit the mouth shape. The third test is done on our automatic snake with spatially varying α and β. Snake initialization is done close to the lips thanks to good mouth ROI extraction.

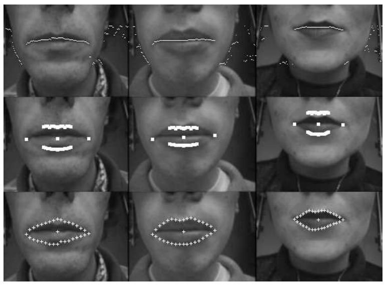

Figure 5: *Snake convergence with hand fitted parameters without ballon forces. Top: minima along image. Middle: snake initialization. Bottom: snake results.*

Results on the first image (Fig. 4) are not sufficient to extract good lips features. The traditionnal snake has not succeeded in giving external lip boundaries for that image.

Fig. 5 shows that balloon forces can help snake convergence but it also provides deformation of shapes and gives irregular lip contours results. Balloon forces have difficulties to follow varying shapes such as mouth without changing its weighting coefficient. These results demonstrate that with or without balloon forces a traditionnal snake needs different hand fitted bending and curvature parameters to extract lip boundaries from different speakers.

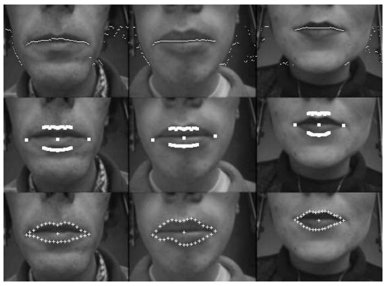

Figure 6: *Hand fitted snake plus balloon forces. Top: minima along image. Middle: snake initialization. Bottom: snake results.*

Results (Fig. 6) are mouth shape speaking correct. Our automatic snake has succeed in extracting outer lips boundaries without changing any parameters. We also shows (Fig. 7) snake evolutions on open mouths. Even with a bad initialization (third image) our snake has again reached good mouth boundaries.

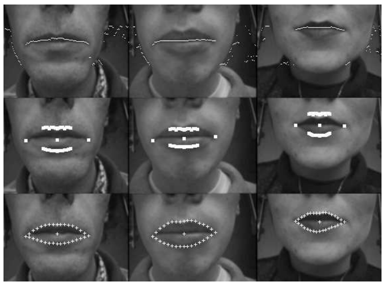

Figure 7: *Automatic snake process. Top: minima along image. Middle: snake initialization. Bottom: snake results.*

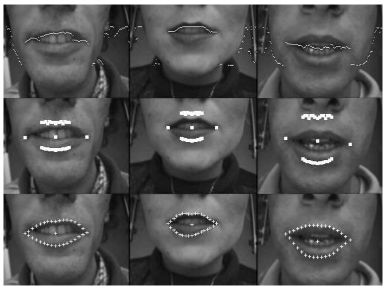

Figure 8: *Automatic snake process on open mouth. Top: minima along image. Middle: snake initialization. Bottom: snake results.*

4.2. Computationnal results

The next table present indications about the computation time for previous tests (Fig.4, Fig.5 and Fig.6) by giving the number of iteration. 99n mean that our algorithm stopped after 99 iterations but without good mouth boundaries extraction.

Methods	AS	TS	TSBF
Image1	115	184	99n
Image2	46	116	57
Image3	64	76	184

Table 1: *Computationnal complexity. AS: Automatic snake (Fig. 6) TS: Traditionnal snake (Fig. 4) TSBF: Traditionnal snake with balloon forces (Fig. 5)*

The worse execution time takes no longer than a few secondes without any optimisation on C routines. We insist on saying that the manual optimisation of snakes coefficient is supposed to be the best known possibility. Our automatic snake is therefore the faster one.

4.3. Statistical results

We hold all the coefficient constant through our test. We used 120 images from our image base including openned and closed mouths from different faces. We obtain 90 percent of good mouth ROI (mouth corners, orientation correctly founded and rather good snake initialization). With good snake initialization (correct ROI mouth extraction) we had 82 percent of convergence on lips boundaries. The 18 percent missing are converging snakes attracted by other image features. This is due to vertical snake initialization which is too far from lips edge points. We are dealing with this problem. It is evident that better snake initialization will solve most of convergence problems.

5. CONCLUSION

We presented an automatic snake which enables us to be less dependent on parameters determination and convergence to obtain the desired solution. Our algorithm is robust considering mouth shapes. The non varying coefficient allowed the computation of images in a short time. But we still think that more accurate ROI mouth finding can be done by linking minima information with vertical global minima. This will help us to reduce the snake computationnal time. Future work will include inner mouth fine detection provided by mouth masks [4].

We will also focus on image sequences processing to enhance lip boundaries extraction robustness. Our main goal is to obtain an unsupervised snake (both initialization and parameters estimation) which lead us to a fine lip boundary extraction in realtime.

6. REFERENCES

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

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

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

[4] Liévin M. and Luthon F. "Lip features automatic extraction". *Proc. of the* $5^t h$ *IEEE Int. Conf. on Image Processing*,Chicago, Illinois, 1998.

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

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