Lip Contour Extraction based on Active Shape Model and Snakes

Kyung Shik Jang

Department of Multimedia Engineering , Dong-Eui University, Busan, Korea

Summary
In this paper, we propose an efficient method for extracting lip contour. Lip shape deformation is modeled by a statistically deformable model based on Active Shape Model(ASM). In the traditional ASM, each landmark point is moved independently to the best matching point with lip profile model, so it deforms the lip shape to implausible one in many cases and may cause many errors for extracting a correct lip contour. In this paper, we have defined an energy function which is consisted of an internal energy and an external energy. The landmark points are moved to positions at which the total energy is minimized. The internal energy is used to minimize the difference of the displacements of adjacent landmark points and keeps the contour from bending abruptly. The external energy is what derives the contour towards matching the profile model. The experiments have been performed for many lip images, and showed very encouraging result.

Key words:
Lip contour, Shape deformation, Active shape model, Snake

1. Introduction

Recently, there is an increasing requirement for a system to track and locate human lip[1, 2]. Human lip has much more information than any other face features, so the lip information could be used in image coding[2]. To improve the performance of speech recognition, the lip information is used together with the acoustic signal[3, 4]. The information is also be applied to the graphic animation systems, which need it for generating the lip shape of the speaker[2, 4].

Accurately and robustly tracking lip motion in image sequences is especially difficult because lips are highly deformable and they vary in shape and color. Gradient based techniques[5, 6] for edge detection of lip often fail due to the poor contrast between lip and surrounding skin region. For methods using color information to build a parametric deformable model for the lip contour, these require optimization technique to refine estimates of contour model to the human lip[7, 8]. Many papers have described the applications of active contour model(snake) for lip boundary detection[9, 10]. The snake methods are able to resolve fine contour details but shape constraints are difficult to incorporate. Furthermore, the snake methods often converge to the wrong result when the lip edges are not distinct or when the lip color is very close to the face skin. Many methods have localized only the outer lip contour when the mouth is closed because the presence of tongue and teeth could obscure the inner contour when the mouth is open. Delmas[9] and Lievin[10] have proposed a statistical approach based on markov random field to segment mouth using the color information and the motion in a spatiotemporal neighborhood. This method has shown good result at the outer lip contour, but shown bad result at the inner lip contour.

Methods using Active Shape Model(ASM) have shown good results in extracting not only the outer lip contour but also the inner one[4, 11]. However they have shown poor results in some cases. In the traditional ASM[12, 13], each landmark point is moved to the best matching point with its local profile model independently, so it deforms the lip shape to much implausible one. That is, the resultant lip shape has a deformed contour with abruptly varying displacements of adjacent points or a deformed contour with bending abruptly. As a result, it may cause many errors for extracting a correct lip contour. Though the updated shape is resolved by projecting the shape into the shape parameter space, most of errors occurred because each landmark points are moved independently.

In this paper, we propose an effective method for extracting lip contour. The lip shape is represented as a set of landmark points and the lip deformation is modeled by a statistically deformable model based ASM. In the traditional ASM, each landmark point is moved independently to the best matching point with its local profile model, so it deforms the lip shape to implausible one and may cause many errors for locating a correct lip contour. In this paper, we have defined an energy function which is consisted of an internal energy and an external energy. The internal energy is used to minimize the difference of the displacements of adjacent landmark points and keeps the contour from bending abruptly. The external energy is what derives the contour towards matching the profile model. The landmark points are moved to positions at which the total energy is minimized. The experiments have been performed for many lip images of Tulip 1 database, and showed that our method extract lip contour than a traditional ASM more exactly.

Manuscript received October 5, 2007
Manuscript revised October 20, 2007

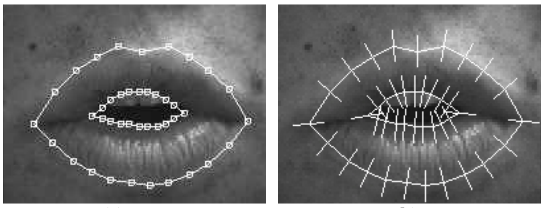
2. Lip Model

2.1 Lip Shape Model

Lip shape is described by a set of 41 landmark points as shown in Fig. 1(a). The lip shape of *i*th training image is described by a shape vector containing the coordinates of the landmark points as shown in equation (1). The lip shape model is represented using equation (2), where $\bar{h}$ is a mean shape, P is a matrix of the first t column eigenvectors corresponding to the largest eigenvalues and shape parameter b is a vector containing the weights for each eigenvector[4].

$$h_i = (x_{1i}, y_{1i}, x_{2i}, y_{2i}, \cdots, x_{41i}, y_{41i})^T \tag{1}$$

$$\begin{aligned} & h = \bar{h} + Pb \\ & where,\ P = [P_1\ P_2\ P_3 \cdots P_t], \\ & \lambda_i \geq \lambda_{i+1}\ ,\quad b = (b_1, b_2, \cdots, b_t)^T \end{aligned} \tag{2}$$



(a) (b)

Fig. 1 Lip shape model and Intensity profile

2.2 Lip Profile Model

Lip profile model is constructed by statistical analysis of the appearance of the image intensity in the neighborhood of each landmark point. As shown in Fig. 1 (b), for every landmark point *j* in the image *i* of the training set, we choose to sample one dimensional profile g_{ij} of length N_p perpendicular to the contour and centered at the point *j* as shown in equation (3). It is sensitive to the light condition to use the absolute gray level value, so we normalize the profile g_{ij} to obtain g_{ij}^n using equation (4). A mean profile $\bar{g}_{ij}^n$ and a covariance matrix S_j are derived. This is repeated for all landmark points at outer and inner contour, giving a profile model for each landmark point.

$$\begin{aligned} & g_{ij} = (g_{ij1}, g_{ij2}, g_{ij3}, \cdots, g_{ijNp})^T \\ & where,\quad g_{ijk}\ is\ a\ gray\ value. \end{aligned} \tag{3}$$

During searching lip contour, we sample a profile F_j of length M_p(M_p>N_p) either side of a point *j* of the lip contour produced by the shape model. The normalized profile F_j^n is derived by using equation (4). We make f_j^n by selecting a sample of length N_p at each of the M_p-N_p+1 possible positions along F_j^n and compare it with $\bar{g}_j^n$ which is the profile model at *j*th landmark point. The updated location for current point is selected by choosing a point that minimizes the function in equation (5).

$$\begin{aligned} & g_{ij}^n = \frac{g_{ij}^{'}}{\sum |g_{ijk}^{'}|} \qquad where, \\ & g_{ij}^{'} = \{ g_{ijk} \mid g_{ij(k+2)} - g_{ijk},\ k = 1, \cdots, N_p - 2 \} \end{aligned} \tag{4}$$

$$E_{profile} = (f_j^n - \overline{g_j^n})^T \mathbf{S}_j^{-1} (f_j^n - \overline{g_j^n}) \tag{5}$$

3. Snake

A snake is an energy-minimizing spline guided by external constraint forces and influenced by image forces that pull it toward features such as edges[14, 15, 16]. The snake is represented by a vector, $v(s) = (x(s), y(s))$ having arc length, s, as parameter. The total energy for the snake is the integral of the energy at each point as equation (6).

$$E = \int_0^1 E_{snake}(v(s))ds = \int_0^1 (E_{\text{int}}(v(s)) + E_{image}(v(s)))ds \tag{6}$$

The internal energy is represented as equation (7). The internal energy term tries to keep the snake smooth. In the original snakes implementation, the first derivative term makes the snake act like a membrane(elasticity) and is used to keep the snake from stretching or contracting along its length. The second derivative term makes it act like a thin plate (stiffness) and is used to keep the snake from bending. Parameters $\alpha(s)$ and $\beta(s)$ are weighting parameters, and control the sensitivity with respect to the first and second derivative respectively. The derivatives may be approximated by finite differences. If $v_i = (x_i, y_i)$ is a point on the contour, the approximations in equation (8) is used.

$$E_{\text{int}}(v(s)) = \alpha(s) | v_s(s) |^2 + \beta(s) | v_{ss}(s) |^2 \tag{7}$$

$$v_s(s) \cong | v_i - v_{i-1} | \qquad v_{ss}(s) \cong | v_{i-1} - 2v_i + v_{i+1} | \tag{8}$$

The image energy term is what derives the snake towards matching the image. It is usually inversely based on image intensity, gradient magnitude(edges), or similar image features. In equation (9), $\gamma(s)$ is an approximate weighting function. In equation (10), the image energy term is represented as gradient magnitude, where $I(x, y)$ is a pixel value and ∇ is a gradient operator.

$$E_{image}(v(s)) = \gamma(s) E_{edge}(v(s)) \tag{9}$$

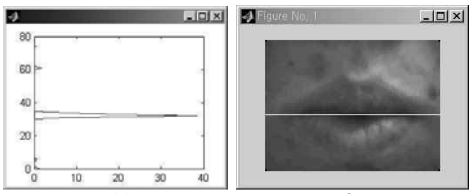
$$E_{edge}(v(s)) = -|\nabla I(x, y)|^2 \tag{10}$$

4. Lip Contour Extraction

To reduce the dependency of initial position, we find out a center line of lip which connects two corner points of lip. The *y*-coordinate of the line is used as *y*-coordinate for the initial search. For each column of the image, *y*-coordinate of the smallest pixel value is found by using equation (11), where *I(x, y)* represents a pixel value at coordinate *(x, y)*. *H* and *W* represent image height and image width, respectively. K_1 and K_2 are constants. The highest peak of *L(y)* in equation (12) gives *y*-coordinate of the center line. *L(y)* for a sample image is shown in Fig. 2(a) and the result is shown in Fig. 2(b). In Fig. 2(a), *y*-axis represents an image height, and *x*-axis does *L(y)*.

$$M(x) = \arg \min_{y \in H} I(x, y) \tag{11}$$

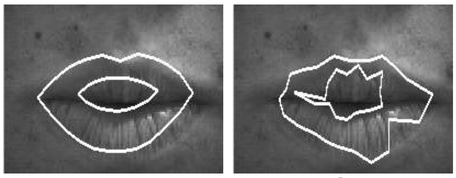
$$L(y) = \sum_{x \in W} \frac{K_1 - \cosh\left(\dfrac{M(x) - \dfrac{H}{2}}{\dfrac{H}{K_2}}\right)}{K_1} \tag{12}$$



(a) (b)

Fig. 2 Lip Center Line

ASM is a method that finds contour based on the standard model and each landmark point is moved independently. On the other hand, snake has no standard model and each point is moved based on the energy of the contour. In ASM, image region in a neighborhood of each landmark is examined and each landmark point is moved to the best matching point with its profile model. Because each point is moved independently without any constraint for the shape, the search result is an implausible shape as shown in Fig. 3. In Fig 3, (b) is a search result from (a), in which multiple adjacent points are moved towards an incorrect boundary so that the contour locally fail to converge towards the correct boundary. In this paper, we have defined an energy function similar to an energy function used in snake. Each landmark point is moved to position at which the total energy is minimized using the energy function and lip profile model.



(a) (b)

Fig. 3 Lip Shape after Search

A lip contour is generated using lip shape model in equation (2) and an initial search point is examined. A region of an image in a neighborhood of each landmark point is examined, and image feature profile for the landmark point is generated. Each landmark point is moved to a point that minimizes an energy function using its profile model and the image feature profile. The resultant shape is resolved by projecting the shape into the shape parameter space and by rescaling the shape parameter. The iterative search process is described below, where θ, *t* and *s* are rotation, translation (t_x, t_y) and scale factor respectively.

Step 1. Generate a lip contour *h* from the lip shape model in equation (2) using shape parameter *b*, and perform linear transform using equation (13).

Step 2. Examine the image region in a neighborhood of each landmark to find points that minimize an energy function as equation (6), and find out new contour H^{new}.

Step 3. Obtain θ^{new}, t^{new}, s^{new} and h^{new} using equation (14) and (15).

Step 4. Find out new shape parameter b^{new} using equation (16).

Step 5. Check the condition for b^{new} using equation (17). If not satisfied, it is rescaled using equation (18).

Step 6. Iterate from step 1 to step 5 until *b* and (θ, t, s) converge.

$$T(s, \theta, t)\begin{pmatrix} x \\ y \end{pmatrix} = s\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \tag{13}$$

$$\begin{aligned}(s^{new}, \theta^{new}, t^{new}) &= \arg\min \\ &[(T(s,\theta,t)(\bar{h}) - H^{new})^T (T(s,\theta,t)(\bar{h}) - H^{new})]\end{aligned} \tag{14}$$

$$h^{new} = T^{-1}(s^{new}, \theta^{new}, t^{new})(H^{new}) \quad (15)$$

$$\mathbf{b}^{new} = \mathbf{P}^T(h^{new} - \bar{h}) \quad (16)$$

$$D^2 = \sum_{i=k}^{t} \left(\frac{b_i^{new\,2}}{\lambda_i}\right) \le D^2_{\max} \quad (17)$$

$$\mathbf{b}^{new} = \mathbf{b}^{new} \frac{D_{\max}}{D} \quad (18)$$

The internal energy in equation (6) is represented as equation (19). The first term is approximated as equation (20), where ∇v_i is the displacement of *i*th landmark point. The term is used to minimize the difference of the displacements of adjacent points. The term penalizes a contour for which the displacements of adjacent points vary abruptly along the contour and favors a contour with smoothly varying displacement of adjacent points. The second term is used to keep the contour from bending abruptly. As a result, the internal energy term tries to keep the snake smooth. The image energy term is usually inversely based on gradient magnitude in snake. However, in this paper, it is what derives the landmark points towards matching the profile model by using equation (5) and (21). In equation (19) and (21), $\alpha(s), \beta(s)$ and $\gamma(s)$ are approximate weighting functions.

$$E_{\text{int}}(v(s)) = \alpha(s)\,|\,\Delta v(s)\,|^2 + \beta(s)\,|\,v_{ss}(s)\,|^2 \quad (19)$$

$$\Delta v(s) \cong \Delta v_i - \Delta v_{i-1} \quad (20)$$

$$E_{image}(v(s)) = \gamma(s)\, E_{profile}(v(s)) \quad (21)$$

5. Experimental Result

The experiments have been performed for the images of Tulip 1 database of isolated digits[17]. It consists of 96 gray image sequence of 12 speakers each saying the first four digits in English twice. We referred to the set of words spoken the first times as Set 1 and the set of words spoken the second times as Set 2. In this paper, the lip shape model was built using 468 images from Set 1. We used 12 shape modes t for the lip shape model. For experiment, 411 images in Set 2 were used. An initial value for parameter (θ, t, s) is (0, (Y_{Lip}, *W/2*), 70). Y_{Lip} represents *y*-coordinate of a line connecting two corner points of lip, and *W* is an image width.

To evaluate the proposed method, we implemented our method and a traditional ASM using Matlab. Fig. 4 shows a result by our method, and Fig. 5 shows one by traditional ASM. The result using our method is more similar to real lip shape than a result using the traditional ASM. Examples of lip contour extraction for all subjects are shown in Fig. 6. Table 1 shows the lip contour extraction results by our method along with ones by the traditional ASM for all test images. All the test images were labeled by hand and were used as the ground truth. True-positive means that lip pixels are classified to lip pixels, and false-negative means that lip pixels are classified to non-lip pixels. True-negative means that non-lip pixels are classified to non-lip pixels, and false-positive means that non-lip pixels are classified to lip pixels.

Two methods are compared in Fig. 7. Fig. 7(a) is a test image and Fig. 7(b) is an initial state. In the traditional ASM, each point is moved independently without constraint. Therefore it would generate a contour along which the displacements of adjacent points vary abruptly. As a result, it would generate an implausible shape as shown Fig. 7(c), and the contour may fail to converge towards the correct boundary. However, in our method, the internal energy is used to minimize the difference of the displacements of adjacent points and keeps the contour from bending abruptly. As a result, we can get a plausible shape as shown Fig 7(d), and the contour may converge towards the correct boundary. The result of our method and the traditional ASM are shown in Fig. 7(e) and Fig. 7(f) respectively.

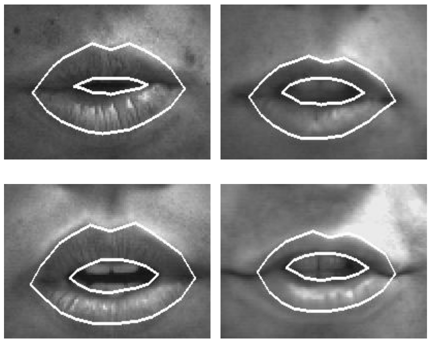


Fig. 4 Result by our method

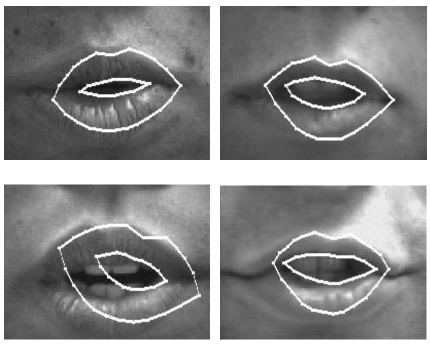


Fig. 5 Result by the traditional ASM

The errors in the inner lip contour occurred mainly due to the gradient originating from the teeth and tongue. For the outer lower lip, some errors occurred because there was no obvious gray level difference between the lip and their neighborhood skin.

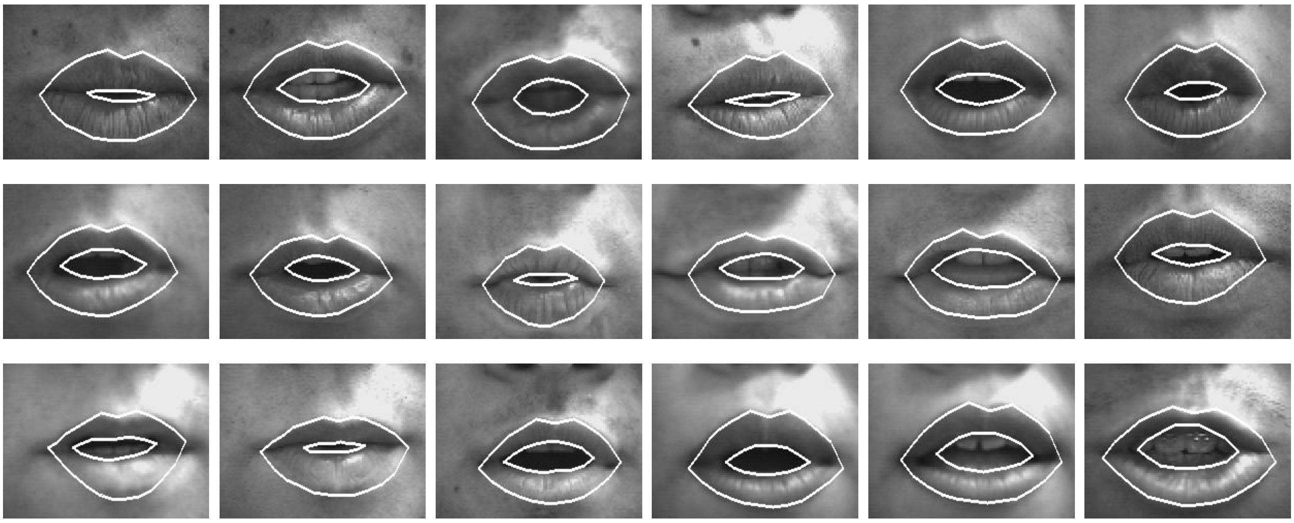


Fig. 6 Lip localization examples

Table 1: Lip contour extraction performance

	Proposed method	ASM
Truth Positive %	74.23	63.34
False Negative %	25.78	36.66
True Negative %	96.00	96.25
False Positive %	4.00	3.75
Error %	9.66	12.25

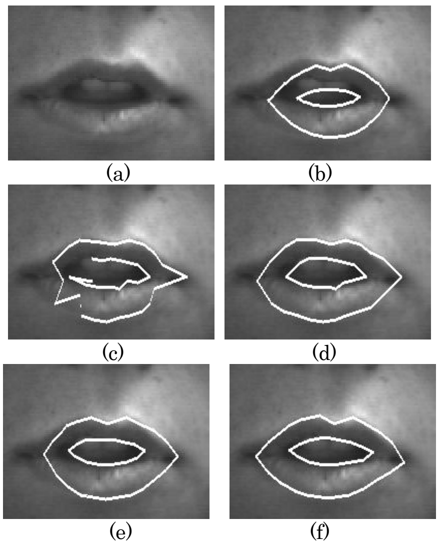


Fig. 7. Result comparison

6. Conclusion

In this paper, we propose an efficient method for extracting lip contour. We define an energy function which is consisted of an internal energy and an external energy, and the landmark points are moved to positions at which the total energy is minimized. The proposed method was tested to many samples of various shapes and the result showed that it extracted correctly lip shapes that were not extracted by a traditional ASM. The better performance may be obtained by defining the internal energy with more global information on lip shape and this is left to the next topic.

References

[1] Mlrhosseinl A. R., H. Yan and K. M. Lam, "Adaptive Deformable Model for Mouth Boundary Detection", Optical Engineering, Vol. 37 No. 3(1998), pp. 869-875.

[2] Oliver N., A. Pentland, "LAFTER: Lips and Face Real Time Tracker", Proceedings of the 1997 Conf. on Computer Vision and Pattern Recognition, (1997), pp. 123-129.

[3] Kaucic R.. A. Blake, "Accurate, Real-Time, Unadorned Lip Tracking", Proceedings of the 6th International Conf. on Computer Vision, pp. 370-375, 1998.

[4] Iain Matthnews, Timothy F. Cootes, J. Andrew Banghan, Stephen Cox and Richard Marvey, "Extractoin of Visual Features for Lipreading", IEEE Tans. on Pattern Recognition and Machine Analysis, Vol 24, No. 2, , pp. 198-213, Feb. 2002.

[5] L. Zhang, "Estimation of the mouth features using deformable templates", IEEE International Conference on Image Processing, Vol. III, pp. 328-331, 1997.

[6] M. Lievin, F. Luthon, "A Hierarchical Segmentation Algorithm for Face Analysis : Application to Lipreading", IEEE Conf. on Multimedia & Exposition'2000, August, 2000.
[7] Wark T., Sridharan and V. Chandran, "An Approach to Statistical Lip Modelling for Speaker Identification via Chromatic Feature Extraction", Proceedings of the 14th International Conf. on Pattern Recognition, Vol. 1, pp. 123-125, 1998.
[8] T. Wark, S. Sridharan, "A Syntatic Approach to Automatic Lip Feature Extraction for Speaker Identification", IEEE International Conference on Acoustics, Speech and Signal Processing, p 1227 - , 1998
[9] Delmas P., Y. Coulon and V. Fristot, "Automatic Snakes for Robust Lip Boundaries Extraction", IEEE International Conf. on Acoustics, Speech and Signal Processing, Vol. 6, pp. 3069-3072, 1999.
[10] Lievin M., F. Luthon, "Unsupervised Lip Segmentation under Natural Conditions", IEEE International Conf. on Acoustics, Speech and Signal Processing, Vol. 6, pp. 3065-3068, 1999.
[11] Luettin, J, and Thacker, NA, "Speechreading using probabilistic models," Computer Vision and Image Understanding, vol. 65, pp. 163-178, 1997.
[12] M. B. Stegmann, R. Fisker, "On Properties of Active Shape Models", Informatics and Mathematical Modelling, Technical University of Denmark, 2000.
[13] Cootes, Taylor, Cooper and Graham, "Active Shape Models-Their Training and Application," Computer Vision and Image Understanding, Vol. 61, No. 1, pp. 38-59, 1995.
[14] Michael Kass, Andrew Witkin and Demetri Terzopoulos, “Snake : Active Contour Model,” International Journal of Computer Vision. v. 1, n. 4, pp. 321-331, 1987.

Kyung Shik Jang received the M.S. and Ph.D. degrees in Electronics Engineering from Yonsei University, Seoul, Korea, in 1991 and 1996, respectively. During 1996-1998, he stayed in Daewoo Electronics Co. Ltd., where he developed Video on Demand system. Since 1998, he joined the Multimedia Engineering, Dongeui University, Busan, Korea. Currently he is an associate professor. His main research interests are face detection, face recognition, image processing and multimedia system.