# A NEW REAL-TIME LIP CONTOUR EXTRACTION ALGORITHM

S.L.Wang[+], W.H.Lau[+] and S.H.Leung[*]
*Department of Electronic Engineering
+Department of Computer Engineering and Information Technology,
City University of Hong Kong, 83 Tat Chee Avenue, Hong Kong

## ABSTRACT

A new lip contour extraction algorithm that combines the merits of the point-based lip model and the parametric lip model is presented in this paper. A 16-point lip model is used to describe the lip contour. With the aid of the FCMS (fuzzy clustering method incorporating shape function), a robust probability map is generated and a region-based cost function can be established. An iterative optimization procedure has been developed to fit the lip model to the probability map. In each iteration, the adjustment of the 16 lip points is governed by three pieces of quadratic curves which constrain the points to form a physical lip shape. Experimental results show that the proposed approach provides satisfactory results for 5,000 lip images of over 20 individuals. A real-time lip contour extraction system has also been implemented.

## 1. INTRODUCTION

In recent years, there is a growing interest in incorporating visual information into automatic speech recognition system. It has been shown that visual cue such as lip movements can greatly enhance the accuracy of the speech recognition system, especially in noisy environments [1,2]. Hence, the development of an accurate and robust algorithm for lip contour extraction becomes vital. However, the presence of variations in different images caused by speakers, utterances, illumination and poor color contrast makes this task difficult.

Many lip models have been proposed for lip contour extraction. Among these, deformable template [3] and active shape model (ASM) [4,5,6] are widely used. Deformable template uses a parametric model to describe the physical shape of the lip. Generally this model can describe the lip shape using only a small number of parameters. However, it is a time consuming process in determining the model parameters mainly due to the complexity of computing the region-based cost function (2D integration over a curved shape) when optimizing the deformable model. ASM is basically a shape-constrained iterative fitting approach. Most of the ASMs use point distribute model (PDM) to describe the lip contour and the valid lip shapes are derived from the training data set via principal component analysis (PCA). One major advantage of using ASM is that no heuristic assumptions are made to the legal shape. In addition, the convergence time can be greatly reduced with the use of the optimization method described in [8]. However, building such bases requires a large training data set and the training is quite time-consuming and tedious. Moreover the PDM may not be able to provide a good fit to those lip shapes that are quite different from the training data set.

In this paper, we present a new lip contour extraction method that combines the advantages of the two aforementioned methods. We will use a PDM to describe the lip contour. However, our algorithm uses a geometric lip model to constrain the points' deformations and thus training data set is no longer required. Since the lip region is defined by a piece-wise linear model, the computation of the region-based cost function is made easier than that for a general parametric model.

## 2. LIP MODEL

In our approach, a 16-point model is used to describe the lip contour. The lip points $p_0$ to $p_{15}$ are labeled in anti-clockwise direction as shown in Fig.1. The point-based lip model parameter set is described by $\lambda_p = \{x_{pi}, y_{pi}\}$ $i = 0,..,15$ . These points are divided into three groups: $p_0$ to $p_7$ and $p_{15}$ describe the lower lip, $p_7$ to $p_{11}$ and $p_{11}$ to $p_{15}$ describe the upper-left lip and upper-right lip, respectively. Three quadratic curves are used to fit these points in least square sense.

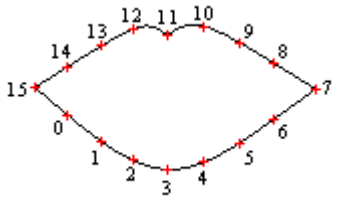



Fig.1: The 16-point lip model and the quadratic curves

### *2.1 Normalization*

In order to reduce the number of parameters of the quadratic models, a normalization process is used to translate the lip corner points $p_7$ and $p_{15}$ to lie on the $x$-axis and the dip point $p_{11}$ on the $y$-axis. The center origin of the lip is set to be the midpoint between the two lip corner points, $p_7$ and $p_{15}$, i.e.,

The work described in this paper is fully supported by a research grant (CityU 1215/01E) from the RGC of the HKSAR, China.

$$x_c = \frac{x_{p7} + x_{p15}}{2}, \quad y_c = \frac{y_{p7} + y_{p15}}{2}. \qquad (1)$$

The normalized radius and rotation angle are given by

$$r = \sqrt{(x_{p7} - x_c)^2 + (y_{p7} - y_c)^2} \qquad (2)$$

$$\theta = \tan^{-1}\left((y_{p15} - y_{p7})/(x_{p15} - x_{p7})\right) \qquad (3)$$

The positions of all the points can be normalized as follows:

$$\begin{aligned} x_{pi,normalized} &= \left[(x_{pi} - x_c)\cos\theta + (y_{pi} - y_c)\sin\theta\right]/r \\ y_{pi,normalized} &= \left[(y_{pi} - y_c)\cos\theta - (x_{pi} - x_c)\sin\theta\right]/r \end{aligned} \qquad (4)$$

$$i = 0,...,15.$$

The upper lip points are then translated to obtain a symmetric lip shape with dip point $p_{11}$ aligned to the center $y$-axis, i.e., $x_{dip} = x_{p11,normalized}$, $y_{dip} = y_{p11,normalized}$, the $x$-coordinates of the upper lip points are modified as follows

$$\begin{aligned} x'_{pi,normalized} &= (x_{pi,normalized} - x_{dip})/(1 - x_{dip}) \quad i = 7,...,10 \\ x'_{pi,normalized} &= (x_{pi,normalized} - x_{dip})/(1 + x_{dip}) \quad i = 12,...,15 \end{aligned} \qquad (5)$$

### *2.2 Geometric Lip Model*

After normalization, three quadratic curves are used for the geometric lip model to fit the lip points as shown in Fig.1. The equations of these curves are given as follows:

(i) for lower lip points from $p_0$ to $p_7$, and point $p_{15}$:

$$x^2 + b_1 xy + c_1 y^2 + a_1 y - 1 = 0 \qquad (6)$$

(ii) for upper-left lip points from $p_{11}$ to $p_{15}$:

$$(x - a_2)^2 + b_2(x - a_2)y + c_2 y^2 + d_2 y - (-1 - a_2)^2 = 0 \qquad (7)$$

where $d_2 = a_2 b_2 - c_2 y_{dip} + (1 + 2a_2)/y_{dip}$.

(iii) for upper-right lip points from $p_7$ to $p_{11}$:

$$(x - a_3)^2 + b_3(x - a_3)y + c_3 y^2 + d_3 y - (1 - a_3)^2 = 0 \qquad (8)$$

where $d_3 = a_3 b_3 - c_3 y_{dip} + (1 + 2a_3)/y_{dip}$.

The parameter set of this geometric model is then given by $\lambda_g = \{x_c, y_c, r, \theta, x_{dip}, y_{dip}, a_1, b_1, c_1, a_2, b_2, c_2, a_3, b_3, c_3\}$.

## 3. IMPLEMENTATION

### *3.1 Probability map generation*

A probability map of a lip image, as shown in Fig.2, can be generated using the FCMS (fuzzy clustering method incorporating shape function) [7], in which the value at each pixel represents the probability of that pixel being a lip pixel. This clustering method takes both the color and shape information into account while most of the current clustering methods only deal with the former. An elliptic cost function is incorporated in the fuzzy measure for enhancing the fuzzy membership.

### *3.2 Cost Function Definition*

In order to obtain an accurate model, a cost function must be defined for determining the optimum model parameters such that the pixels inside the lip region have high probability being lip pixels while those outside pixels have low probability. Assuming that the lip and the non-lip regions are not overlapped, the optimum partition is found when the cost function in (9) is maximized.

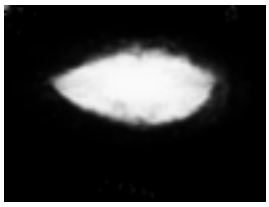

(a)

(b)

Fig.2 (a) The original lip image, (b) The probability map of (a)

$$\max C(\lambda_p) = \prod_{(x,y)\in R_l(\lambda_p)} prob_l(x,y) \prod_{(x,y)\in R_{nl}(\lambda_p)} prob_{nl}(x,y) \qquad (9)$$

where $\lambda_p$ is the 16-point model parameters and $R_l$ and $R_{nl}$ are the regions enclosed by the point model and the region outside, respectively; $prob_l(x,y)$ is the probability of pixel at location $(x,y)$ being a lip pixel and $prob_{nl}(x,y)$ is the probability being a non-lip pixel.

By taking logarithm and extending it to a continuous domain, (9) can be reformulated to minimize $F_B$ as follows:

$$F_B = -\sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} \int_0^{g_i(u)} f(u,v)\,dv\,du \qquad (10)$$

where $f(u,v) = \log(prob_l(u,v)) - \log(prob_{nl}(u,v))$ is the difference of log probability between lip and non-lip regions, $g_i(u)$ is the lip boundary between points $i$ and $i$+1 using the first order interpolation, i.e.,

$$g_i(u) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}(u - x_i) + y_i \qquad (11)$$

where $(x_i, y_i)$ is the coordinate of the $i^{th}$ point, with $x_N = x_0$ and $y_N = y_0$ ($N$=16).

### *3.3 Model Initialization*

With the probability map generated by the FCMS, the initial 16-point model can easily be derived. The left and right lip corners can be found by scanning the columns. The top and bottom boundaries can then be derived from the center axis between the left and right corner points. Finally, standard quadratic equations are used to fit the upper and lower lip points separately and the initial geometric parameters can be derived by setting $a_i = b_i = 0$ for $i$ = 1,2,3.

### *3.4 Lip Contour Points Calculation*

The 16-point lip points can be derived from the geometric parameters by evenly assigning the points along the quadratic curves. The normalized $x$ coordinates of these points are evenly assigned in the $x$-axis, i.e., $x_{p0}$ to $x_{p6}$ equal to {-3/4, -1/2, -1/4, 0, 1/4, 1/2, 3/4} for the lower lip contour, and $x_{p7}$ to $x_{p15}$ equal to {1, 3/4, 1/2, 1/4, 0, -1/4, -1/2, -3/4, -1} for the upper lip contour, respectively. The normalized $y$ coordinates of the lip points can then be calculated easily according to the curve parameters $a_i$, $b_i$, $c_i$ $i$ = 1,2,3. Finally, the normalized coordinates can be translated back to their original coordinates by reversing the process of normalization.

### *3.5 Optimization Procedure*

The optimization procedure is an iterative process and the lip points are adjusted in order to reduce the cost function $F_B$ at each iteration. This step can be realized using the optimization method described in [8]. The displacement vector $\Delta\lambda_p$ and the new lip model can be updated as follows:

$$\Delta\lambda_p = \{dx_i, dy_i\} \quad i = 0,1,\cdots,15 \tag{12}$$

$$\lambda_{p,new} = \lambda_{p,old} + w\Delta\lambda_p \tag{13}$$

where $w$ is the step size and is set to 0.05 empirically.

After normalization, three quadratic curves are then fitted to the 16 lip points accordingly and the geometric parameters $\lambda_g$ can be derived. The constrained 16-point model $\lambda_{p,new,contrainted}$ can then be derived from these parameters.

The entire lip contour extraction method runs as follow:

i. Generate the probability map using the FCMS algorithm and obtain the initial geometric parameters $\lambda_g$ and the 16-point model parameters $\lambda_p$.
ii. Calculate the displacement vector $\Delta\lambda_p$ and the new lip contour $\lambda_{p,\,new}$ using the optimization method described in [8].
iii. Normalize the 16 lip points and fit them to the three quadratic curves, and obtain the geometric parameters $\lambda_g$ using curve-fitting method (see Appendix).
iv. Calculate the new lip points, $\lambda_{p,\,new,\,constrained}$, from the geometric model.
v. Compute the cost function $F_B$ and repeat step (ii) until the change of cost function is less than a pre-defined small threshold ε.

## 4. EXPERIMENTAL RESULTS

### *4.1 Lip Extraction Results*

5,000 lip images of size 110×90 for over 20 different individuals have been analyzed using the proposed lip extraction algorithm. Over 95% of these images have been successfully analyzed and all the unsatisfactory cases are associated with the poor probability maps caused by the poor contrast between the lip and background. Fig. 3 shows some of the lip extraction results.

Fig.3 Lip extraction results for different lip shapes

The algorithm has been implemented on a PIV 1.9GHz PC and applied to track the lip movement in video sequence. By using the model parameters obtained for the current field as the initial parameters for the next field, the number of iterations has been shown to decrease tremendously and thus increase the extraction rate. Experimental results show that our real-time system can perform lip contour extraction at a rate of 50 fields/sec for PAL system. Figure 4 shows a fragment of lip tracking results.

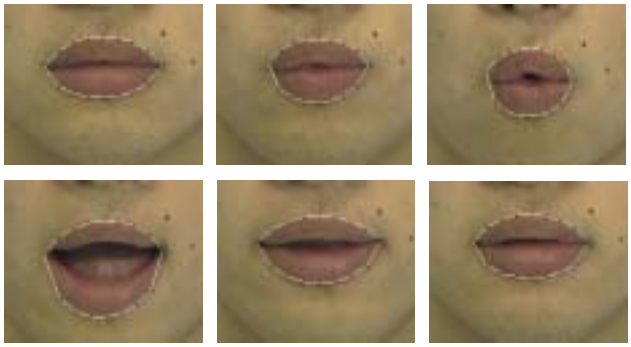


Fig.4. Lip contour tracking results of several frames of a lip image sequence (an utterance of "one" in English)

### *4.2 Selection of The Number of Model Points*

The quality of the lip fitting for the proposed 16-point model is clearly related to the number of points defined in the model. The performance of the algorithm has been examined for different number of lip points. It has been observed that longer computation time is required for a model with higher number of points and it is also more difficult to obtain a good fit. However, the lip shape cannot be described accurately by using small number of points. Furthermore, the number of iterations will also increase since the cost function changes drastically in each update.

The evaluation of the performance in lip contour extraction is a subjective measure to some extent. Here we use the converged value of the cost function as an objective measure to denote the accuracy of the extraction result. After analyzing 5,000 lip images, the normalized cost function and averaging processing time are shown in Fig.5. It should be noted that the cost function is normalized by the 16-point model. In the *x*-axis, *m*-2-*n* represents that *m*, 2 and *n* points are used to describe the lower lip, lip corners and upper lip, respectively. From this figure, it is obvious that the 16-point model (7-2-7) is the best choice which not only gives more accurate results but also more efficient.

## 5. CONCLUSIONS

In this paper, a new lip contour extraction algorithm that combines the advantages of parametric model and point distribute model is presented. Experimental results show that the new method can provide an accurate lip contour in a very efficient manner. The algorithm has also been implemented in a real-time lip contour extraction system.

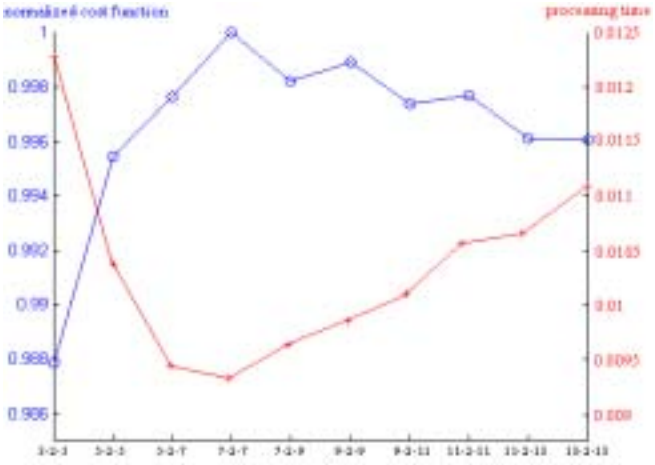


Fig.5. Performance comparison versus number of lip points.

## 6. APPENDIX: CURVE-FITTING METHOD

The least square method and iterative method are used to find the best fitting curves for the lower lip and upper lips, respectively. Since the change of the geometric parameters is relatively small, the number of iterations is also small and the whole process is fairly efficient comparing with the computation of the cost function $F_B$.

(i) Lower lip: Since the model is linear, we can use least square method to solve for the parameters.

$$Set \quad A=\begin{bmatrix} y_i & x_i y_i & y_i^2 \\ y_{i+1} & x_{i+1} y_{i+1} & y_{i+1}^2 \\ \cdots & \cdots & \cdots \\ y_{i+n} & x_{i+n} y_{i+n} & y_{i+n}^2 \end{bmatrix}, u=\begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix}, v=\begin{bmatrix} 1-x_i^2 \\ 1-x_{i+1}^2 \\ ... \\ 1-x_{i+n}^2 \end{bmatrix}$$

$$\because A\,u = v$$

$$\Rightarrow A^T A\,u = A^T v$$

$$\Rightarrow u = \left(A^T A\right)^{-1} A^T v$$

(ii) Upper lip: Let

$\Delta_2 = \sum_i \left((x-a_2)^2 + b_2(x-a_2)y + c_2 y^2 + d_2 y - (-1-a_2)^2\right)^2$ and

$\Delta_3 = \sum_i \left((x-a_3)^2 + b_3(x-a_3)y + c_3 y^2 + d_3 y - (1-a_3)^2\right)^2$ be the fitting error for the upper-left lip and upper-right lip, respectively. Using the steepest descent method, we have the parameters of the quadratic curves given by the following equations:

$$a_2 = -\frac{\sum_i \left[x_i^2 + b_2 x_i y_i + c_2\left(y_i^2 - y_{dip} y_i\right) + \frac{y_i}{y_{dip}} - 1\right]\left(-2x_i + \frac{2y_i}{y_{dip}} - 2\right)}{\sum_i \left(-2x_i + \frac{2y_i}{y_{dip}} - 2\right)^2}$$

$$b_2 = -\frac{\sum_i \left[x_i^2 + c_2\left(y_i^2 - y_{dip} y_i\right) + a_2\left(-2x_i + \frac{2y_i}{y_{dip}} - 2\right) + \frac{y_i}{y_{dip}} - 1\right](x_i y_i)}{\sum_i (x_i y_i)^2}$$

$$c_2 = -\frac{\sum_i \left[x_i^2 + b_2 x_i y_i + a_2\left(-2x_i + \frac{2y_i}{y_{dip}} - 2\right) + \frac{y_i}{y_{dip}} - 1\right]\left(y_i^2 - y_{dip} y_i\right)}{\sum_i \left(y_i^2 - y_{dip} y_i\right)^2}$$

$$a_3 = -\frac{\sum_i \left[x_i^2 + b_3 x_i y_i + c_3\left(y_i^2 - y_{dip} y_i\right) + \frac{y_i}{y_{dip}} - 1\right]\left(-2x_i - \frac{2y_i}{y_{dip}} + 2\right)}{\sum_i \left(-2x_i - \frac{2y_i}{y_{dip}} + 2\right)^2}$$

$$b_3 = -\frac{\sum_i \left[x_i^2 + c_3\left(y_i^2 - y_{dip} y_i\right) + a_3\left(-2x_i - \frac{2y_i}{y_{dip}} + 2\right) + \frac{y_i}{y_{dip}} - 1\right](x_i y_i)}{\sum_i (x_i y_i)^2}$$

$$c_3 = -\frac{\sum_i \left[x_i^2 + b_3 x_i y_i + a_3\left(-2x_i - \frac{2y_i}{y_{dip}} + 2\right) + \frac{y_i}{y_{dip}} - 1\right]\left(y_i^2 - y_{dip} y_i\right)}{\sum_i \left(y_i^2 - y_{dip} y_i\right)^2}$$

## 7. REFERENCES


[1] E. D. Petajan, "Automatic Lipreading to enhance Speech Recognition," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 40-47, 1985.

[2] Y. Zhang, S. Levinson, T. Huang, "Speaker independent audio-visual speech recognition," *Proc. of IEEE Int. Conf. on Multimedia and Expo*, pp.1073-1076, vol.2, New York, 2000.

[3] T. Coianiz, L. Torresani and B. Caprile, "2D Deformable Models for Visual Speech Analysis," *Speechreading by Humans and Machines*, Springer, 1996.

[4] T.A. Faruquie, A. Majumdar, N. Rajput, L.V. Subramaniam, "Large vocabulary audio-visual speech recognition using active shape models," *Proc. of IEEE Int. Conf. on Pattern Recognition*, pp.106-109, vol.3, Barcelona, 2000.

[5] I. Mattews, T.F Cootes, J.A Bangham, S. Cox, R. Harvey, "Extraction of visual features for lipreading", *IEEE Tran. on PAMI*, pp.198-213, vol.24, Feb. 2002.

[6] J. Luettin, Neil A. Thacker and Steve W. Beet, "Active Shape Models for Visual Speech Analysis," *Speechreading by Humans and Machines*, Springer, 1996.

[7] S.L. Wang, S.H. Leung and W.H. Lau, "Lip segmentation by fuzzy clustering incorporating with shape function," *Proc. of IEEE ICASSP'2002*, pp.1077-1080, vol.1, Orlando, 2002.

[8] K.L. Sum, W.H. Lau, S.H. Leung, A.W.C Liew, K.W. Tse, "A new optimization procedure for extracting the point-based lip contour using active shape model," *Proc. of IEEE ICASSP'2001*, pp.1485-1488, vol.3, Salt Lake City, 2001.