

A NEW OPTIMIZATION PROCEDURE FOR EXTRACTING THE POINT-BASED LIP CONTOUR USING ACTIVE SHAPE MODEL

K.L. Sum, W.H. Lau, S.H. Leung, Alan W.C. Liew and K.W. Tse

Department of Electronic Engineering
City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong

Abstract

This paper presents a new optimization procedure for extracting the point-based lip contour using Active Shape Model (ASM). A 14-point ASM lip model is used to describe the lip contour. With the aid of fuzzy clustering analysis, a probability map of the color lip image is obtained and a region-based cost function is established. The new optimization procedure operates on the spatial domain (actual contour points) and all the points are pulled towards their desirable locations in each iteration. Hence, the lip contour evolution becomes better controlled and consequently fast convergence is achieved. The new procedure can also achieve real-time performance on lip contour extraction and tracking from lip image sequence.¹

1. INTRODUCTION

It has been demonstrated that incorporating the lip features into an automatic speech recognition system can significantly improve the performance especially in noisy environment [1]. In order to obtain the features, one has to extract the lip contour from an image. Many lip models, such as the Active Contour Models (or Snakes) [2], Deformable Templates [3,4] and Active Shape Model (ASM) [5], have been proposed for lip contour extraction. All these models have their own advantages. Snakes have enough variability for most shapes description. Deformable templates can describe the shape with small set of parameters. ASM can describe shapes within the seen set without any heuristic legal shape assumptions.

In order to perform lip contour extraction automatically, we have to define the lip contour model first. Using the lip model parameters, a cost function will then be established. Finally, an optimization procedure is used to find the model parameters.

Many optimization algorithms have been proposed for lip contour extraction. Since the lip contour model is generally described by a lip parameter vector, it is natural to use a multidimensional optimization algorithm, e.g. Downhill Simplex Method [5], to find the solution. This kind of optimization algorithms usually operates on the *parameter*

domain directly. Fig. 1 shows the lip contours obtained from 3 different iterations using the Downhill Simplex Method. It can be observed that only some of the contour points are *pulled* towards their correct locations in each iteration. Hence more iterations are required to achieve a good fit.

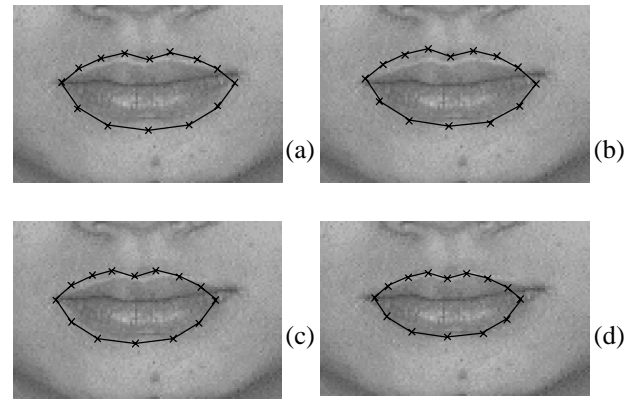


Fig. 1. The lip contours obtained using the Downhill Simplex Method: (a) the initial contour, (b) the 5th iteration, (c) the 8th iteration, and (d) the 16th iteration.

Instead of operating on the parameter domain, other optimization procedures operated on the *spatial domain* (actual contour points) have also been proposed. Greedy Search [6] is an exhaustive approach in which 8 neighborhood pixels of each snake point are being examined during each iteration. The search is repeated until a predefined number of points stop moving. This method will converge slowly if its initial contour is far away from the desired contour. In addition, it is computationally inefficient due to its exhaustive search nature.

Another spatial domain based optimization method for ASM is described in [7]. The model points are evolved to the desired contour position by finding the image evidences, such as the edges and intensity, along the perpendicular direction to the contour of each point. Similar to the Greedy Search, the edges energy functional is used for optimization. However, this approach is also not suitable for locating the desired contour that is far away from the initial position.

In this paper, a new optimization procedure for extracting the lip contour from a point-based model using ASM will be described. The new approach is operated on the spatial domain and a region-based cost function is used for the

¹ The authors wish to acknowledge that this work is supported by Hong Kong RGC grant No. CityU 1036/97E.

optimization. It will be shown that the search direction in each iteration is well controlled and fast convergence is hence achieved. This new optimization procedure has been applied to both lip contour extraction and tracking and the results are presented in Section 3.

2. LIP CONTOUR EXTRACTION

By considering the advantages and disadvantages of different approaches, Active Shape Model and a region-based cost function are employed in our algorithm.

2.1. Lip model

Active Shape Model is a statistical model and the shape is described by a set of points. These points are allowed to deform within the main deformation modes which are obtained from a training set of images via Principal Component Analysis. One major advantage of using ASM is that no heuristic assumptions are made to the legal shape deformation. In addition, the ASM is flexible enough to capture the shape details with the use of linear combination of a small set of deformation modes. In general, the coordinates of the contour points of an arbitrary shape \mathbf{x} represented by ASM can be approximated by (1).

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b} \quad (1)$$

where $\bar{\mathbf{x}}$ is the mean shape, \mathbf{P} is the matrix of eigenvectors of the covariance matrix and \mathbf{b} is the weight vector for each eigenvector. It should be noted that only the first few eigenvectors corresponding to the largest eigenvalues are sufficient for modeling the shape variation.

In our approach, a 14-point ASM lip model is used for lip contour description. The contour points p_1 to p_{14} are labeled in anti-clockwise direction as shown in Fig 2. A set of lip parameters $\lambda_{\mathbf{B}} = \{s, \theta, x_c, y_c, b_1, b_2, \dots, b_6\}$ representing the scale, rotating angle, model center offsets, and weights of six dominant bases, are used to describe the lip contour.

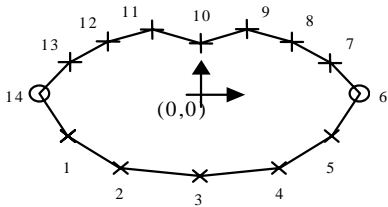


Fig. 2. The 14-point ASM lip model

2.2. Cost function

For a color lip image as shown in Fig. 3(a), its RGB representation is transformed into the uniform CIELAB and CIELUV color spaces [8]. Color feature vectors consisted of color features $\{a, b, u, v, \text{hue}_{ab}, \text{hue}_{uv}, \text{chroma}_{uv}\}$ is then generated for each pixel. Applying our fuzzy clustering analysis [9] to the color feature vectors, a probability map is obtained as shown in Fig. 3(b). A pixel with higher probability value is more likely to be within the lip region.

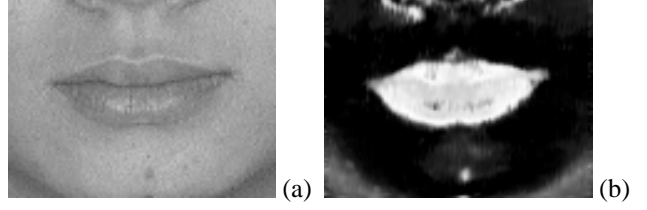


Fig. 3. (a) An RGB lip image. (b) The probability map of the lip image in (a).

To apply the ASM lip model, one has to find the lip region from an image. The lip image is assumed to have two non-overlapped regions, i.e. a lip and a non-lip region. The lip parameter $\lambda_{\mathbf{B}}$ can be used to define the boundary between the lip region (R_l) enclosed by the lip model \mathbf{x} and the region outside (R_{nl}). An optimal parameter set is found when the cost function $C(\lambda_{\mathbf{B}})$ is maximized, i.e.

$$\max_{\lambda_{\mathbf{B}}} C(\lambda_{\mathbf{B}}) = \left\{ \prod_{e \in R_l(\lambda_{\mathbf{B}})} \text{Pr}_l(e) \prod_{e \in R_{nl}(\lambda_{\mathbf{B}})} \text{Pr}_{nl}(e) \right\} \quad (2)$$

where $\text{Pr}_l(e)$ and $\text{Pr}_{nl}(e)$ are the probabilities of a pixel e that belong to the lip and non-lip region, respectively.

By taking logarithm and extending it to a continuous domain, (2) can be reformulated to minimize F_B given in (3).

$$F_B = \sum_{i=1}^N \hat{f}_i = - \sum_{i=1}^N \int_{x_i}^{x_{i+1}} \int_0^{g_i(t_1)} f(t_1, t_2) dt_2 dt_1 \quad (3)$$

where $f(t_1, t_2) = \log(\text{Pr}_l(t_1, t_2)) - \log(\text{Pr}_{nl}(t_1, t_2))$ is the difference of the log probability of point (t_1, t_2) being in the lip and non-lip region; $g_i(t_1)$ is the boundary of the lip contour between the i^{th} and $i+1^{\text{th}}$ points using first order interpolation, i.e.,

$$g_i(t_1) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} (t_1 - x_i) + y_i \quad (4)$$

where (x_i, y_i) is the coordinate of the i^{th} point; $x_{N+1} = x_1$ and $y_{N+1} = y_1$.

2.3. Optimization procedure for lip contour extraction

With the lip model and probability map, we would like to find the lip parameters using the region-based contour fitting approach. In this paper, we proposed a new optimization procedure, which is a modification to that proposed in [7], to find $\lambda_{\mathbf{B}}$. Assuming that the lip contour is given by

$$\mathbf{L} = \mathbf{M}(s, \theta)[\mathbf{x}] + \mathbf{X}_c \quad (5)$$

where \mathbf{L} is a vector containing the coordinates of the lip contour points, i.e.,

$$\mathbf{L} = [x_1 \ y_1 \ x_2 \ y_2 \ \dots \ x_N \ y_N]^T \quad (6)$$

$\mathbf{M}(s, \theta)[\mathbf{x}]$ is the affine transformation of \mathbf{x} [7] and $\mathbf{X}_c = [x_c \ y_c \ x_c \ y_c \ \cdots \ x_c \ y_c]^T$ is the center position of the lip model.

The optimization procedure is an iterative process. A local displacement made to each lip contour point is determined in each iteration. The displacement of the i^{th} point is determined by differentiating F_B with respect to its x and y coordinates (see appendix), i.e.,

$$dp_i = \begin{bmatrix} dp_{x_i} \\ dp_{y_i} \end{bmatrix} = \begin{bmatrix} -\frac{\partial F_B}{\partial x_i} \\ -\frac{\partial F_B}{\partial y_i} \end{bmatrix} = - \begin{bmatrix} \frac{\partial \hat{f}_{i+1}}{\partial x_i} + \frac{\partial \hat{f}_i}{\partial x_i} \\ \frac{\partial \hat{f}_{i+1}}{\partial y_i} + \frac{\partial \hat{f}_i}{\partial y_i} \end{bmatrix} \quad (7)$$

The displacement vector $\Delta \mathbf{L}$ of the lip contour points can be written as:

$$\Delta \mathbf{L} = [dp_{x_1} \ dp_{y_1} \ \cdots \ dp_{x_N} \ dp_{y_N}]^T \quad (8)$$

The lip contour is then given by

$$\mathbf{L}_{new} = \mathbf{L}_{old} + w \Delta \mathbf{L} \quad (9)$$

where w is step size of the displacement vector.

Since \mathbf{L}_{new} will be described by $\mathbf{x}_{new} = \bar{\mathbf{x}} + \mathbf{P} \mathbf{b}_{new}$ using the ASM lip model in (1), the lip model \mathbf{x}_{new} can be obtained from (5) as follows:

$$\mathbf{x}_{new} = \mathbf{M}(s_{new}^{-1}, \theta_{new}^{-1})[\mathbf{L}_{new} - \mathbf{X}_{cnew}] \quad (10)$$

where s_{new}, θ_{new} and \mathbf{X}_{cnew} can be obtained using the minimization method described in [7].

The weight vector is then given as $\mathbf{b}_{new} = \mathbf{P}^{-1}(\mathbf{x}_{new} - \bar{\mathbf{x}})$ according to (1). It should be noted that \mathbf{P}^{-1} equals \mathbf{P}^T since \mathbf{P} is an orthogonal matrix, i.e.

$$\mathbf{b}_{new} = \mathbf{P}^T (\mathbf{x}_{new} - \bar{\mathbf{x}}) \quad (11)$$

In order to guarantee that the updated lip contour described by \mathbf{x}_{new} is within the allowable shape variation, each of the weights in \mathbf{b}_{new} is constrained within ± 3 s.d. of that obtained from the training set. In addition, this also ensures that the lip contour is varied in a controlled manner during each iteration. Finally, the updated lip parameter is given in (12) and the updated lip contour is obtained using (5).

$$\lambda_{\mathbf{B}new} = \{s_{new}, \theta_{new}, x_{cnew}, y_{cnew}, \mathbf{b}_{new}\} \quad (12)$$

In summary, the optimization procedure is as follow:

Step 1: Overlay an initial lip contour (e.g. $\bar{\mathbf{x}}$) to the lip region and compute the cost function F_B using (3). *Note that an initial lip contour can be estimated using the probability map as shown in Fig. 3(b).*

Step 2: Compute the displacement vector $\Delta \mathbf{L}$ and the lip contour \mathbf{L}_{new} using (7), (8) and (9).

Step 3: Compute s_{new}, θ_{new} and \mathbf{X}_{cnew} using the minimization method in [7].

Step 4: Compute the lip model \mathbf{x}_{new} using (10).

Step 5: Construct $\lambda_{\mathbf{B}new}$ using (11) and the results obtained from *Step 3*, and compute the updated lip contour using $\lambda_{\mathbf{B}new}$, (1) and (5).

Step 6: Compute the cost function F_B and repeat *Step 2* if the change of cost function exceeds a pre-defined threshold, ϵ .

3. RESULTS

A 14-point ASM lip model was built using 200 lip images of 2 males and 2 females. The image size is 108×81 and in 24-bit true color (RGB). The ASM lip model has 26 bases, but only the first 6 significant bases are used in model (1). The threshold ϵ is set to 0.001 and lip contour step size w in the update of lip contour is set to 0.6.

3.1. Lip contour extraction

A total of 1320 images have been processed and a successful rate of 95% is recorded in locating the correct lip contour. It is found that all the failed cases are associated with the poor probability map of an image where the color difference between the lip and the face is not very distinct.

Instead of using the edge-based information to define the cost function [7], we use a region-based approach to carry out the optimization. This has an advantage that the optimization procedure becomes insensitive to the position and size of the initial lip contour. Fig. 4 shows two examples of the lip contour extraction, one with an off-centered initial lip contour and the other with an off-scaled initial lip contour.

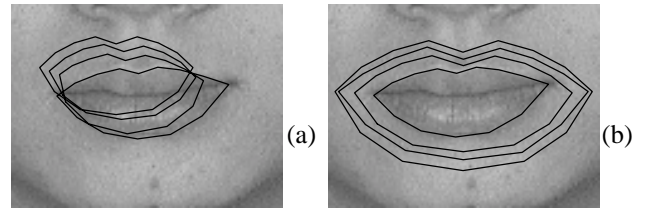


Fig. 4. The evolution of the lip contours during optimization: (a) with off-center initialization; and (b) with off-scale initialization. *Noted that only 4 steps are shown.*

It is observed that all the contour points are pulled towards their own desirable locations in each iteration. The lip contour hence gradually converges towards the correct location for both cases. By comparing with the results given in Fig. 1, it is clearly shown that our optimization procedure has a better control of the lip contour evolution during each iteration and consequently converges faster.

3.2. Lip contour tracking

For a lip image sequence, the procedure for finding the lip contour for each individual image is identical. The lip contour obtained from the previous image will serve as an initial shape for the present image. Fig. 5 shows the lip contour tracking results of a lip image sequence. Experimental results show that our new lip contour optimization procedure achieves real-time operation, i.e., more than 50 fields/sec, using a 933 MHz PC. The computation includes the generation of probability map from each field of the image sequence.

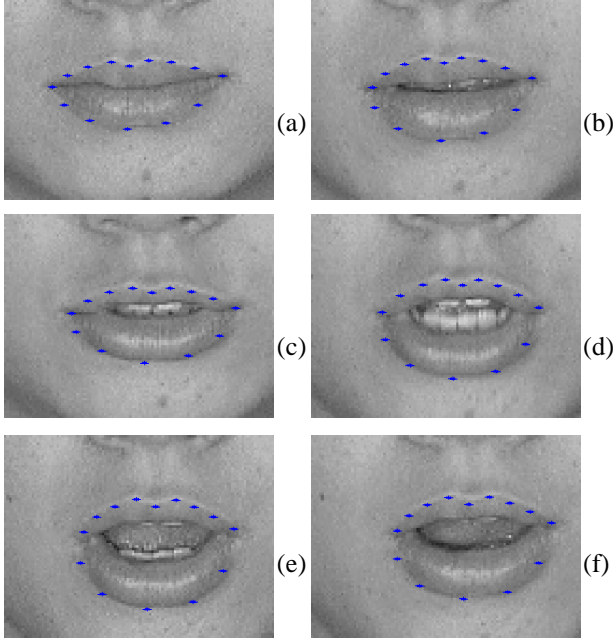


Fig. 5. The lip contour tracking results of a lip image sequence for field number: (a)1, (b) 26, (c) 36, (d) 42, (e) 45, (f) 50.

4. CONCLUSIONS

In this paper, a new optimization procedure for lip contour extraction is proposed. Instead of using the edge information to define the cost function, we choose to use a region-based approach. Since the optimization is operated directly on the spatial domain not the parameter domain, the lip contour evolution in each iteration becomes better controlled and fast convergence is hence achieved. In addition, our approach is also insensitive to the position and size of the initial lip contour. Real-time performance on lip contour extraction from lip image sequence is also achieved.

5. APPENDIX

Given $\hat{f}_i = \int_{x_i}^{x_{i+1}} \int_0^{g_i(t_1)} f(t_1, t_2) dt_1 dt_2$

and $g_i(t_1) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} (t_1 - x_i) + y_i$

We have,

$$\frac{\partial \hat{f}_i}{\partial x_i} = - \int_{x_i}^{x_{i+1}} f(t_1, g_i(t_1)) \frac{\partial g_i}{\partial x_i} dt_1 + \int_0^{g_i(x_i)} f(x_i, t_2) dt_2$$

$$\frac{\partial \hat{f}_i}{\partial y_i} = - \int_{x_i}^{x_{i+1}} f(t_1, g_i(t_1)) \frac{\partial g_i}{\partial y_i} dt_1$$

$$\frac{\partial \hat{f}_i}{\partial x_{i+1}} = - \int_{x_i}^{x_{i+1}} f(t_1, g_i(t_1)) \frac{\partial g_i}{\partial x_{i+1}} dt_1 - \int_0^{g_i(x_{i+1})} f(x_{i+1}, t_2) dt_2$$

$$\frac{\partial \hat{f}_i}{\partial y_{i+1}} = - \int_{x_i}^{x_{i+1}} f(t_1, g_i(t_1)) \frac{\partial g_i}{\partial y_{i+1}} dt_1$$

6. REFERENCES

- [1] E. D. Petajan, "Automatic Lipreading to Enhance Speech Recognition," *Proc. of IEEE Global Telecomm. Conf.*, Atlanta, Georgia, 1984, pp.265-272.
- [2] M. Kass, A. Witkin and D. Terzopoulos, "SNAKES: Active Contour Models," *Int. J. of Computer Vision*, 1(4):321-331, 1987.
- [3] Ram R. Rao and Russell M. Mersereau, "Lip Modelling for Visual Speech recognition," *Proc. of the 28th Asilomar Conf. on Signals, Systems & Computers*, Vol. 1, pp. 587-590, 1994.
- [4] T. Coianiz, L. Torresani, and B. Caprile, "2D Deformable Models for Visual Speech Analysis", *Speechreading by Humans and Machines*, Springer, NY, 1996.
- [5] J. Luetttin, Neil A. Thacker and Steve W. Beet, "Active Shape Models for Visuals Speech Feature Extraction," *Speechreading by Humans and Machines*, Springer, NY, 1996.
- [6] Donna J. Williams and Mubarak Shah, "A Fast Algorithm for Active Contours and Curvature Estimation," *CVGIP: Image Understanding*, Vol. 55, No. 1, pp. 14-26, 1992.
- [7] T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham, "Active Shape Models – Their Training and Application," *Computer Vision and Image Understanding*, Vol. 61, No. 1, pp. 38-59, 1995.
- [8] R.W.G. Hunt, *Measuring Colour*, 2nd Ed., Ellis Horwood Series in Applied Science and Industrial Technology, Ellis Horwood Ltd., 1991.
- [9] Alan W.C. Liew, K.L. Sum, S.H. Leung and W.H. Lau, "Fuzzy Segmentation of Lip Image Using Cluster Analysis," *Proc. of Eurospeech'99*, Vol. 1, pp. 335-338, Sept, 1999.