A Local Region-Based Approach For Lip Contour Extraction Using Localized Active Contour Model

Xin Liu

Abstract

Lip contour extraction is crucial to the success of a lipreading system. This paper presents a local region-based approach for lip contour extraction using the localized active contour model. The proposed approach utilizes a combined semi-ellipse as the initial evolving curve to split the local neighborhoods into local-interior and local-exterior respectively, and then computes the localized energy for evolving and extracting. This method is robust against the noise, rotation, deformation and teeth appearance, and not limited to a fixed lip model. Experiments show its promising result in comparison with the existing methods.

1 Introduction

Lip contour extraction has been extensively studied in recent years [2, 13, 7]. It is one of the most important techniques for human-machine interface applications such as lip reading [10], audio-visual speech recognition [8] and facial expression analysis [6]. Nevertheless, it is a non-trivial task to find a robust and accurate method for lip contour extraction due to large variations caused by different speakers, noise, illumination conditions, low contrast between lip and skin, high deformable level of lips, and so forth.

In the past decade, a number of techniques have been proposed to achieve lip contour extraction, which can be categorized into two major classes: the edge-based approach and the model-based approach. The edge-based method mainly utilizes the low level spatial cues such as color and edges to achieve lip localization and extraction [13]. Often, the performance of such a method will deteriorate when there is a poor contrast between lip and surrounding skin regions. In contrast, the model-based approach, which builds a lip model with a small set of parameters, generally outperforms the former one. Examples include deformable template (DT) [3] [7] and active shape model (ASM) [9]. The DT algorithm utilizes a parametric model to describe the lip contour, which is, however, sensitive to the deformation and irregularity of the lips. The ASM utilizes a series of landmark points, which are controlled within a few modes derived from a training data, to describe the lip shape. In general, the training process of the ASM is quite time-consuming. Besides the DT and ASM methods, the active contour model (ACM) or snakes [2] is another typical example along this line. The conventional ACM allows an initial contour to deform by minimizing a specific global energy function to produce the desired segmentation. Paper [2] has shown the success of this method in its application domain, but this method is somewhat sensitive to the parameter initialization and the image noise.

Furthermore, when objects have heterogeneous statistics, it is found that the localized active contour model(LACM) [4] can generally achieve a satisfactory segmentation results while the conventional ACM fails. In the LACM, the evolving curve splits the local neighborhoods into local-interior and local-exterior respectively. Subsequently, the localized energy for evolving and extracting can be computed. However, improper parameters such as large radius or far away evolving curve in LACM can lead to the wrong extracting results. See the Fig. 1 and Fig. 2 for the examples.

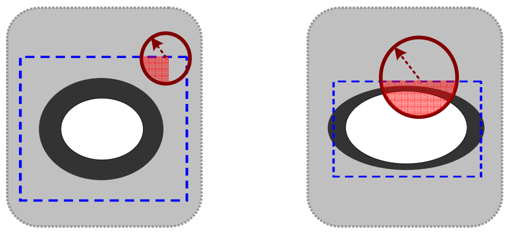


Figure 1. Improper parameters expression.(a)ulterior evolving curve with small local radius, (b)proper evolving curve with large local radius.

In this paper, we propose a local region-based approach for lip contour extraction method using the localized active contour model. We find a combined semi-ellipse as the initial evolving curve which can be fitted well in the LACM, meanwhile, the proper parameters in LACM can be automatic selected. Experiments have shown the promising results of the proposed algorithm in comparison with the existing methods.

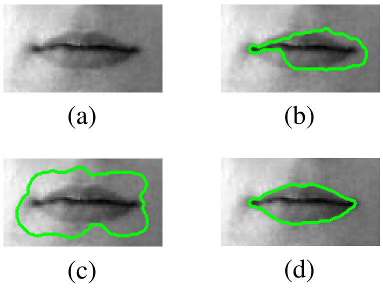


Figure 2. (a) Lip image with noise affects, (b) Conventional ACM based extracting result, (c) LACM based extracting result with improper parameters, (d) LACM based extracting result with proper parameters.

2. Overview of LACM

This section will overview the framework of LACM [4], in which the only assumption is that the foreground and background regions will be locally different.

The statistical analysis of local regions leads to the construction of a group of local energies about each point on the evolving curve, in order to optimize these local energies in its own local region, each point is considered individually, consequently, the point's component of the local energy is computed by splitting the local neighborhoods into local-interior and local-exterior using the evolving curve.

In this paper, I denotes a pre-specified image defined on the domain Ω, C denotes a closed contour represented as the zero level set of a signed distance function ϕ, i.e., $C = \{u\,|\phi(u) = 0\}$ [4]. The interior of C is specified by the following approximation of the smoothed Heaviside function:

$$\mathcal{H}\phi(u) = \begin{cases} 1, & \phi(u) < -\varepsilon \\ 0, & \phi(u) > \varepsilon \\ \frac{1}{2}\left\{1 + \frac{\phi}{\varepsilon} + \frac{1}{\pi}\sin\left(\frac{\pi\phi(u)}{\varepsilon}\right)\right\}, & otherwise. \end{cases} \quad (1)$$

Similarly, the exterior C can be defined as $(1 - \mathcal{H}\phi(u))$.

The derivative of $\mathcal{H}\phi(u)$, a smoothed version of the Dirac delta is used to specify the area adjacent to the curve.

$$\delta\phi(u) = \begin{cases} 1, & \phi(u) = 0 \\ 0, & |\phi(u)| < \varepsilon \\ \frac{1}{2\varepsilon}\left\{1 + \cos\left(\frac{\pi\phi(u)}{\varepsilon}\right)\right\}, & otherwise. \end{cases} \quad (2)$$

Parameters u and v are expressed as independent spatial variables to represent a single point in Ω, respectively. Using this notation, the characteristic function $\mathcal{B}(u, v)$ marked the local regions in terms of a radius parameter r can be described as follows:

$$\mathcal{B}(u, v) = \begin{cases} 1, & \|u - v\| < r \\ 0, & otherwise. \end{cases} \quad (3)$$

Using $\mathcal{B}(u, v)$, we define an energy functional in terms of a generic internal energy functional F, the resulting energy functional $\mathrm{E}(\phi)$:

$$\mathrm{E}(\phi) = \int_{\Omega_u} \delta\phi(u) \int_{\Omega_v} \mathcal{B}(u, v) \cdot F(I(v), \phi(v))dudv, \quad (4)$$

where the functional F is a generic internal energy measure used to represent local adherence to a given model at each point along the contour. This energy relies on the assumption that foreground and background regions should have maximally separate mean intensities which can cause the curve to move.

Therefore, a localized region-based energy formed from the global energy by substituting local means for global ones is shown here [12]:

$$F = -(\mu_{in}(u) - \mu_{out}(u))^2, \quad (5)$$

$$\mu_{in}(u) = \frac{\int_{\Omega_v} \mathcal{B}(u, v) \cdot \mathcal{H}\phi(v) \cdot I(v)dv}{\int_{\Omega_v} \mathcal{B}(u, v) \cdot \mathcal{H}\phi(v)dv}, \quad (6)$$

$$\mu_{out}(u) = \frac{\int_{\Omega_v} \mathcal{B}(u, v) \cdot (1 - \mathcal{H}\phi(v)) \cdot I(v)dv}{\int_{\Omega_v} \mathcal{B}(u, v) \cdot (1 - \mathcal{H}\phi(v))dv}, \quad (7)$$

where the localized versions of the means $\mu_{in}(u)$ and $\mu_{out}(u)$ represent the intensity mean in local interior and exterior regions around a point u, respectively.

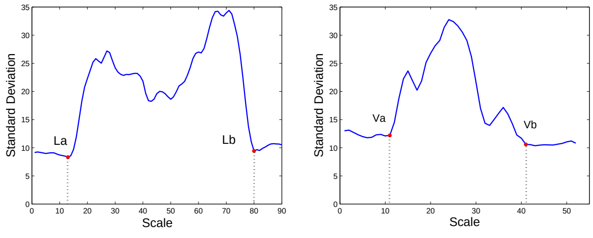
By ignoring the image complexity that may arise outside local region, only contributions from the points within the radius r of the contour are considered. Finally, for the purpose of keeping the curve smooth, a regularization term is added as is commonly done in active contour segmentation energies. Meanwhile, the arclength of the curve is penalized and weighted by a parameter λ, and the final energy $\mathrm{E}(\phi)$ is given in the following:

$$\mathrm{E}(\phi) = \int_{\Omega_u} \delta\phi(u) \int_{\Omega_v} \mathcal{B}(u, v) \cdot F(I(v), \phi(v))dvdu + \lambda \int_{\Omega_u} \delta\phi(u)\,\|\nabla(u)\|du. \quad (8)$$

By taking the first variation of this energy with respect to ϕ, the following evolution equation is obtained:

$$\frac{\partial\phi}{\partial t}(u) = \delta\phi(u) \int_{\Omega_v} \mathcal{B}(u, v) \cdot \nabla_{\phi(v)} F(I(v), \phi(v))dv + \lambda\delta\phi(u)div(\frac{\nabla\phi(u)}{|\nabla\phi(u)|})\,\|\nabla\phi(u)\|. \quad (9)$$

It is certainly noteworthy that, any region-based segmentation energy can be put into this framework.



(a) (b)

Figure 3. Standard deviation of columns and rows.

3. The Proposed Algorithm

Our proposed automatic lip contour extraction system includes an initialization step and a lip contour extraction process.

3.1 Initialization

Empirical studies have found that a lip shape is usually close to an rectangular region [5] [7]. Furthermore, it can also be approximatively surrounded by various of semi ellipses. How to find a combined semi-ellipse as the initial evolving curve is of crucial importance to extract the lip contours in our method.

For the purpose of finding the a combined semi-ellipse of lip region, the detection of lip corner dots is needed. Specifically, $I(x, y)$ represents a pixel value at coordinate (x, y), m, n express the maximum values of rows and columns. The left corner, right corner, up corner, down corner are denoted as La, Lb, Va, Vb.

We project the RGB-based lip image into the gray-level one, from the practical viewpoint, it is inevitable that there exists the noises or uneven illumination affects. Hence, each lip image is performed with a 3×3 mean filter and a contrast stretching adjustment.

According to the statistical methods, the really lip regions are usually of the big standard deviation value. Therefore, the lip corner dots of horizontal or vertical can be detected by computing the first and last value of standard deviation , which is changing obviously compared with the adjacent ones. For example, in order to find the horizontal lip corner columns quickly, we can use the following equations:

$$mean_j = \frac{1}{m}\sum_{i=1}^{m}(I(i,j)), j = 1, 2, \cdots n, \tag{10}$$

$$Std_j = \left(\frac{1}{m}\sum_{i=1}^{m}(I(i,j) - mean_j)^2\right)^{\frac{1}{2}}, \tag{11}$$

$$|Std_j - Std_{j+2}| < \Delta S, |Std_j - Std_{j-2}| < \Delta S, \tag{12}$$

where ΔS is a changing threshold. Fig. 3 and Fig. 4gives an example. We can easily obtain the coordinate value of La_x, Lb_x. Afterwards, the value of La_y and Lb_y can be computed through the mean coordinate with the minimum gray value of $I(La_x, j), i = 1, \cdots, n$. The other approximative corner dots can also be computed through the above method.

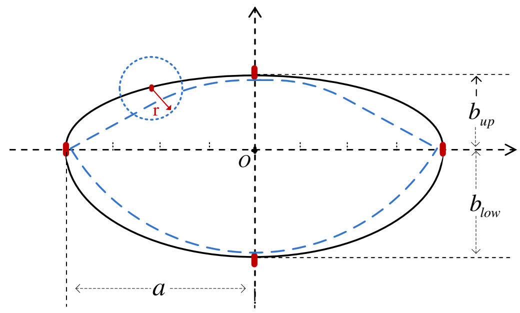


Figure 4. A combined semi-ellipse around the lip.

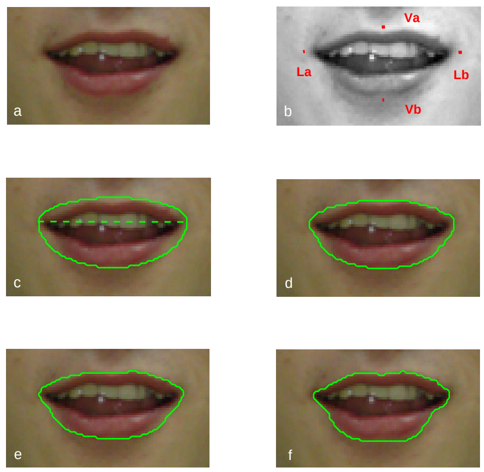


Figure 5. The extracting procedure. (a) lip image, (b) lip corner dots, (c) the combined semi-ellipse, (d) after 10 iterations, (d) after 20 iterations, (d) after 30 iterations.

Let the (x_c, y_c) be the origin center of the combined semi-ellipse, through which the mathematical equations are as follows:

$$x_c = \frac{1}{2}(La_x + Lb_x), y_c = \frac{1}{2}(La_y + Lb_y),$$

$$\theta = \arctan\left(\frac{Lb_y - La_y}{Lb_x - La_x}\right),$$

$$a = \frac{1}{2}((Lb_x - La_x)^2 + (Lb_y - La_y)^2)^{\frac{1}{2}},$$

$$b_{up} = ((Va_x - x_c)^2 + (Va_y - y_c)^2)^{\frac{1}{2}},$$

$$b_{low} = ((Vb_x - x_c)^2 + (Vb_y - y_c)^2)^{\frac{1}{2}},$$

$$X = (x - x_c) \cdot \cos\theta + (y - y_c) \cdot \sin\theta,$$
$$Y = (y - y_c) \cdot \cos\theta - (x - x_c) \cdot \sin\theta,$$

$$\frac{X_{up}^2}{a^2} + \frac{Y_{up}^2}{b_{up}^2} = 1, \frac{X_{low}^2}{a^2} + \frac{Y_{low}^2}{b_{low}^2} = 1, \tag{13}$$

where a is the semi-major axes, b_{up} and b_{low} are the up and low semi-minor axes, respectively. θ is the inclined angle, which is positively defined in the counter-clockwise direction. θ is the inclined angle , and it is defined to be positive in the counter-clockwise direction.

3.2 Contour Extraction

After successfully finding the combined semi-ellipse of the lip region, we can let it be the evolving curve represented the zero level set C just as is stated in part 2, which can be fitted well in the LACM. Subsequently, local neighborhoods of the points can be split into local interior and local exterior by the evolving curve.

By computing the local energies at each point along the curve, the evolving curve will deform so as to minimizing the local energies in order to produce the desired lip segmentation. See Fig. 5 for the details.

The concise extraction steps are as follows:

- Locate the lip region, preprocess;
- Obtain the combined semi-ellipse;
- Evolve with iteration;
- Extract the lip contours.

The radius r selected by the function $\mathcal{B}(u, v)$ is an important parameter in LACM. By rule of thumb, $r = \frac{r_b}{2}$ is appropriate in most cases.

4. Experimental Result

We have applied our approach to the 500 frontal face images with the different mouth shapes. The database consists of 200 face images from the CVL face database [11], 200 face images from the GTAV face database [1] and 100 lip images from our laboratorial database. In our experiments, we set the parameter λ is equal to 0.3.

Examples of lip contour extraction are shown in Fig. 6 and Fig. 7. It can be clearly seen that, the accurate lip contours can be extracted using our algorithm. Table 1 presents the details in comparison with other two existing methods, in which, we define the extracted performance as $\frac{n_c}{n}$, where the n_c is the correct extracted numbers, n is the total numbers of the test database.

Figure 6. The extracted results of lip images form the CVL database by our proposed approach.

Figure 7. The extracted results of lip images from our laboratory database by our proposed approach.

Table 1. Compared with the existing methods.

Method	Lip model	Automatic	Performance
ACM	No	No	78.6%
DT	Yes	No	92.6%
Our approach	No	Yes	96.4%

As is shown above, the deformable or irregular lip contours can be extracted successfully using the proposed method, which also has a better extracting performance. Furthermore, the proposed approach is not limited to a fixed lip model, meanwhile , it is usually more tolerant to the affects of noise, rotation and the deformation. More important, it is also worth noting that our proposed algorithm can reach completely automation, though which the combined semi-ellipse can be found as the initial evolving curve in LACM.

We have also examined the unsatisfactory results (less than 4%), and found that they all have the very poor contrast between the lip and skin region, or have obvious beard effects around the lips.

5. Conclusion

This paper has proposed an automatic and robust lip contour extraction algorithm using localized active contour model. We obtain a combined semi-ellipse as the initial evolving curve through which an optimum extraction of the lip image into lip and non-lip regions can be found. This algorithm is robust against the the noise, rotation , deformation and the appearance of teeth.

References

[1] http://gps-tsc.upc.es/gtav/researchareas/upcfacedatabase /gtavfacedatabase.htm.

[2] G. Chiou and H. Jenq Neng. Lipreading from color video. *IEEE Transactions on Image Processing*, 6(8):1192–1195, 1997.

[3] T. Coianiz, L. Torresani, and B. Caprile. 2d deformable models for visual speech analysis. *NATO Advanced Study Institute: Speechreading by Man and Machine*, 150:391–398, 1995.

[4] S. Lankton and A. Tannenbaum. Localizing region-based active contours. *IEEE Transactions on Image Processing*, 17(11):2029–2039, 2008.

[5] M. Li and Y. M. Cheung. Automatic lip localization under face illumination with shadow consideration. *Signal Process*, 89(12):2425–2434, 2009.

[6] J. J. Lien, T. Kanade, J. F. Cohn, and L. Ching-Chung. Automated facial expression recognition based on facs action units. In *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pages 390–395, 1998.

[7] A. W. C. Liew, S. H. Leung, and W. H. Lau. Lip contour extraction from color images using a deformable model. *Pattern Recognition*, 35(12):2949–2962, 2002.

[8] P. Liu and Z. Wang. Visual information assisted mandarin large vocabulary continuous speech recognition. In *Proc. of International Conference on Natural Language Processing and Knowledge Engineering*, pages 72–77, 2003.

[9] J. Luettin, N. Thacker, and S. Beet. Visual speech recognition using active shape models and hidden markov models. In *Proc. of ICASSP*, volume 2, pages 817–820, 1996.

[10] I. Matthews, T. F. Cootes, J. A. Bangham, S. Cox, and R. Harvey. Extraction of visual features for lipreading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):198–213, 2002.

[11] F. Solina, P. Peer, B. Batagelj, S. Juvan, and J. Kovac. Color-based face detection in the “15 seconds of fame” art installation. In *Proc. Mirage, Conf. Computer Vision/Computer Graphics Collaboration for Model-Based Imaging, Rendering, Image Analysis and Graphical Special Effects,Rocquencourt, France, Mar.10-11*, pages 38–47, 2003.

[12] A. Yezzi, A. Tsai, and A. Willsky. A fully global approach to image segmentation via coupled curve evolution equations. *Journal of Visual Communication and Image Representation*, 13:195–216, 2002.

[13] X. Zhang, R. M. Mersereau, M. Clements, and C. C. Broun. Visual speech feature extraction for improved speech recognition. In *Proc. of ICASSP*, volume 2, pages 1993–1996, 2002.