A ROBUST LIP TRACKING ALGORITHM USING LOCALIZED ACTIVE CONTOURS AND DEFORMABLE MODELS

Xin Liu and Yiu-ming Cheung

Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China
{xliu,ymc}@comp.hkbu.edu.hk

ABSTRACT

Lip tracking is crucial to the success of a lip reading recognition system. This paper presents a robust lip tracking algorithm using localized active contours and deformable models. The proposed approach utilizes a combined semi-ellipse as the initial evolving curve, through which a separation of the original lip image into lip and non-lip regions can be found. Moreover, the dynamic parameter selection of local regions and a 16-point deformable model are obtained to achieve the finial tracking results. The proposed approach is adaptive to the movement of the lips from frame to frame, and also robust against the appearance of the teeth. Experiments show the promising results of our algorithm.

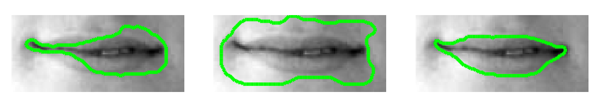
***Index Terms*—** Lip tracking, localized active contour, deformable model

1. INTRODUCTION

In recent years, lip tracking has received wide attention in the community because of its potential applications in areas such as lipreading, audio-visual speech recognition and facial expression analysis. Nevertheless, it is a non-trivial task to track the lip movements due to the large variations caused by different speakers, noise, low contrast between lip and skin, teeth effect and so forth.

In the last decade, a few techniques have been proposed to realize the lip tracking with the focus on segmentation of lip regions or extraction of lip contours. Essentially, point-based approach and region-based approach are two main approaches for tracking lips from frame sequences. In the point-based method [1], a set of characteristic points are detected through the low level spatial cues such as color and edges, as well as a priori knowledge of the lip structure. However, such a method is somewhat sensitive and ambiguous to the initial position of feature points along the lip edges. In general, the region-based method can make the tracking results more robust and realistic. Typical examples include deformable template (DT) [2] and active contour model (ACM) [3]. The DT algorithm employs a cost function to partition a color lip image into lip and non-lip regions via a parametric model. Generally, the tracking performance of this method may deteriorate when there exists a poor contrast between lip and surrounding skin regions. The conventional ACM algorithm allows an initial closed curve to deform via minimizing a global energy, such that an object contour is obtained. However, this approach is sensitive to the choice of parameters and uneven illuminations. Recently, some researchers attempt to combine the merits of the above approaches, papers [4][5] have shown the desired tracking results in their application domain. Nevertheless, their performance will be affected by the teeth appearances. Further, these methods usually involve iterative complexity or stochastic optimization, which is quite time-consuming.

Moreover, almost all the region-based approaches utilize the global statistical characteristics. When images have heterogeneous statistics or complex components, it is found that the localized active contour models (LACM) [6] can generally achieve a better segmentation result, e.g., as shown in Fig.1(a) and (c). Nevertheless, this model is dependent on the appropriate selection of correlative parameters, otherwise, its performance will deteriorate as shown in Fig.1(b).



(a) (b) (c)

Fig. 1. Lip contour extraction with uneven illuminations. (a) Conventional ACM based extracting result, (b) LACM based extracting result with improper parameters, (c) LACM based extracting result with proper parameters.

In this paper, we propose a robust lip tracking algorithm using LACM and deformable model. We find a combined semi-ellipse as the initial evolving curve fitted in LACM to extract the lip contour of the first frame. Subsequently, we utilize the dynamic selections of local radius and a 16-point deformable model to achieve the final tracking results. Experimental results show the efficacy of our algorithm.

2. OVERVIEW OF LACM

This section overviews the framework in LACM [6], which assumes that the foreground and background regions are locally different. This framework utilizes the evolving curve to split the local regions into the local-interior and local-exterior , through which a group of local energies are constructed. The advantage of this framework is that the complex appearances of objects can be successfully segmented with localized energies when the corresponding global energies fail.

Let I denote a pre-specified image defined on the domain Ω, parameters u and v are expressed as independent spatial variables to represent a single point, individually. C denotes a closed curve represented as the zero level set of a signed distance function ϕ, i.e., $C = \{u\,|\phi(u) = 0\}$ [6]. The interior of C is specified by the following approximation of the smoothed Heaviside function:

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

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 in the following is used to specify the area adjacent to the curve.

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

Along the curve C, the characteristic function $\mathcal{B}(u, v)$ marked the local regions in terms of a radius parameter r can be described as follows:

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

Let $\mu_{in}(u)$ and $\mu_{out}(u)$ represent the intensity mean in local interior and exterior regions localized by $\mathcal{B}(u, v)$ at a point u, respectively.

$$\mu_{in}(u) = \frac{\int_{\Omega_v} \mathcal{B}(u, v) \cdot \mathcal{H}\phi(v) \cdot I(v)dv}{\int_{\Omega_v} \mathcal{B}(u, v) \cdot \mathcal{H}\phi(v)dv}, \tag{4}$$

$$\mu_{out}(u) = \frac{\int_{\Omega_v} \mathcal{B}(u, v) \cdot (1 - \mathcal{H}\phi(v)) \cdot I(v)dv}{\int_{\Omega_v} \mathcal{B}(u, v) \cdot (1 - \mathcal{H}\phi(v))dv}. \tag{5}$$

Subsequently, a localized region-based energy formed from the global energy [7] is obtained:

$$F = -(\mu_{in}(u) - \mu_{out}(u))^2. \tag{6}$$

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

$$\mathrm{E}(\phi) = \int_{\Omega_u} \delta\phi(u) \int_{\Omega_v} \mathcal{B}(u, v) \cdot F(I(v), \phi(v))dvdu + \lambda \int_{\Omega_u} \delta\phi(u)\, \|\nabla(u)\| du. \tag{7}$$

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

$$\frac{\partial\phi}{\partial t}(u) = \delta\phi(u) \int_{\Omega_v} \mathcal{B}(u, v) \cdot \nabla_{\phi(v)} F(I(v), \phi(v))dv + \lambda\delta\phi(u) div\left(\frac{\nabla\phi(u)}{|\nabla\phi(u)|}\right) \|\nabla\phi(u)\|. \tag{8}$$

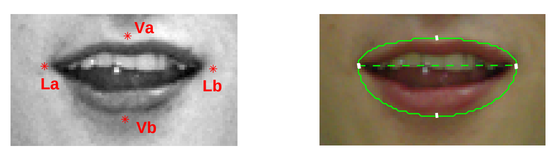
It is noteworthy that almost all the region-based segmentation energies can be put into this framework.

3. THE PROPOSED ALGORITHM

Our proposed lip tracking algorithm mainly includes the two steps: (a) lip contour extraction for the first frame, (2) lip tracking in the following frames.

3.1. Lip Contour Extraction

In LACM, the initial evolving curve C and the radius r of the local region are two crucial parameters. As the uneven illuminations and the teeth appearance always exist during the speech, improper parameters such as far away evolving curve, large radius may cause inaccurate result as shown in Fig.1(b).



(a) Lip corner dots (b) Combined semi-ellipse

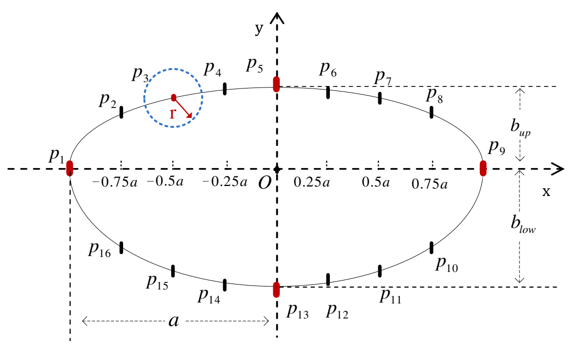
Fig. 2. A combined semi-ellipse around the lip.

Empirical studies have found that a lip shape can be approximatively surrounded by a combination of two semi-ellipses, which can be used as the initial evolving curve fitted in LACM. According to the previous work [4] [8], the primary lip corner dots have been successfully detected. We denote the left corner, right corner, up corner and down corner as La, Lb, Va and Vb, respectively. Let (x_c, y_c) be the origin center of the combined semi-ellipse, through which the mathematical equations can be described in the following:

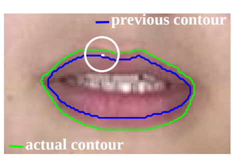
$$x_c = \frac{1}{2}(La_x + Lb_x), y_c = \frac{1}{2}(La_y + Lb_y),$$

$$\theta = \arctan\left(\frac{Lb_y - La_y}{Lb_x - La_x}\right),$$
$$a = \frac{1}{2}((Lb_x - La_x)^2 + (Lb_y - La_y)^2)^{\frac{1}{2}},$$
$$b_{up} = ((Va_x - x_c)^2 + (Va_y - y_c)^2)^{\frac{1}{2}},$$
$$b_{low} = ((Vb_x - x_c)^2 + (Vb_y - y_c)^2)^{\frac{1}{2}},$$
$$X = (x - x_c) \cdot \cos\theta + (y - y_c) \cdot \sin\theta,$$
$$Y = (y - y_c) \cdot \cos\theta - (x - x_c) \cdot \sin\theta,$$
$$\frac{X_{up}^2}{a^2} + \frac{Y_{up}^2}{b_{up}^2} = 1, \frac{X_{low}^2}{a^2} + \frac{Y_{low}^2}{b_{low}^2} = 1, \tag{9}$$

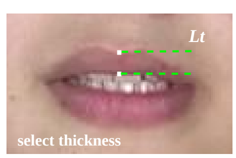
where a is the semi-major axes, b_{up} and b_{low} are the up and low semi-minor axes, respectively. θ is the inclined angle, which is positively defined in the counter-clockwise direction. Consequently, we let the combined semi-ellipse be the initial evolving curve fitted in LACM. Meanwhile, as a rule of thumb, $r = \frac{b_{up}}{2}$ is appropriate to extract the lip contour in this step.



(a) A 16-point lip model



(b) Opening mouth



(c) Upper lip thickness

Fig. 3. Proper parameters in our algorithm.

In addition, the extracted lip contours usually exist unsmoothness, as well as the slightly lack of the geometric lip shapes. Therefore, we employ a 16-point geometric deformable model [5] with cubic spline interpolation to model the lip shape as shown in Fig.3(a) , which is always physically meaningful in tracking applications.

3.2. Lip Tracking

As speaking usually includes many frames in a second, which can be regarded as a group of continuous sequences. Additionally, a lip shape of one frame changes a little compared with the adjacent one. After accurately extracted the lip contour of the first lip frame, we can use it as the initial evolving curve fitted in LACM to the next frame .

It is noteworthy that, the lip movements, especially in the process of opening a mouth, the lip contour of previous one may inside the current one. As is shown in Fig.3(b), to avoid the effects of teeth appearance, the dynamic parameter selection is proposed in LACM. We can easily compute the middle thickness Lt of the upper lip by the variations of pixel value along the line segment OP_5 as shown in Fig.3(c). Therefore, we employ the dynamic parameter r_i, which can be fitted in LACM as the local radius for tracking the lip movements, i.e.,

$$r_i = \frac{Lt_{i-1}}{2}, i \geq 2, \tag{10}$$

where Lt_{i-1} expresses the middle thickness of the upper lip in the previous one. We let N denote the total number of points on the evolving curve. The pseudocode for lip tracking algorithm is given as follows:

Input: $I_i \in \Omega, C_{i-1}$(*The previous lip contour*).
Output: C_i (*The tracking contour*).
Lip image preprocessing;
while *minimize the local energy is not met* **do**
  for $(j = 1; j \leq N; j++)$ **do**
    Let C_{i-1} be the initial evolving curve;
    Assign proper r to the $\mathcal{B}(u, v)$;
    $\delta\phi(u_j) = \delta\phi(u_j) + \frac{\partial\phi}{\partial t}(u_j)$;
  end
  if *all* $E(\phi(u_j)) < \varepsilon$, $j \in [1, \cdots, N]$ **then**
    $C_i = \{u \,|\phi(u) = 0\}$;
    Obtain the tracking result;
  else
    Go back to the beginning;
  end
end

Algorithm 1: The tracking algorithm

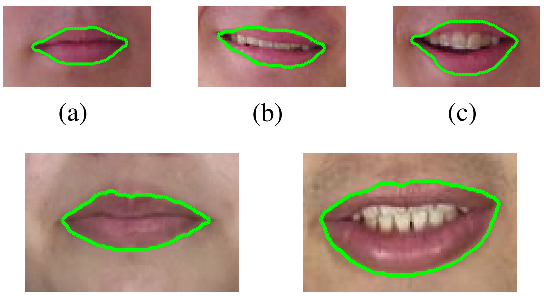
4. EXPERIMENTAL RESULT

The algorithm has been implemented on an Intel® Core$^{\text{TM}}$2 Quad Q9450 2.66 GHz machine and applied with Matlab 7.0 image processing toolkit. We project the RGB lip images into the gray-level space, each lip image is performed with a 3×3 mean filter and a contrast stretching adjustment. In our experiments, we set the parameter λ is equal to 0.3.

4.1. Experiment 1

We have applied our lip contour extraction approach to the 200 frontal face images from the CVL face database and 300 face images from our laboratory database.

Examples of lip contour extraction are shown in Fig.4. It can be seen that the accurate lip contours can be extracted using the proposed algorithm. In our experiments, more than 95% of the test database have a satisfactory result. We have also examined the unsatisfactory ones and found that they are all have the very poor contrast between the lip and skin region, or have obvious beard effects around the lips.



(a) (b) (c)

(d) (e)

Fig. 4. LACM based extracting results. (a)(b)(c) from the CVL database and (d)(e) from our laboratory database.

4.2. Experiment 2

We have performed the proposed lip tracking algorithm on a large number of face sequences captured from 10 speakers, who were speaking English and Chinese digits (0-9) in a uniform illuminance environment. Each second records 20 colorized face frames and the located lip image of size 116×72 from the face sequences.

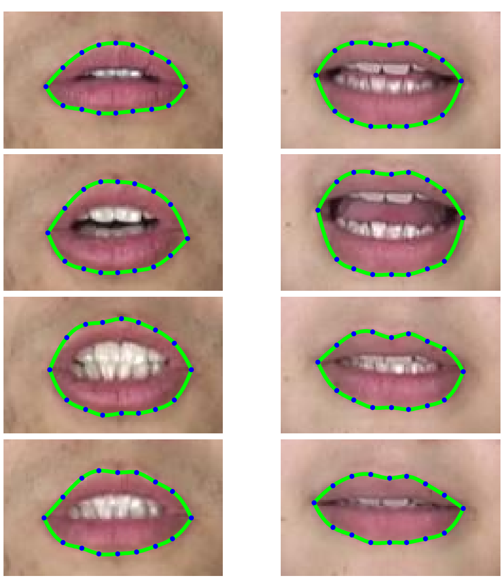


Fig. 5. Tracking result using the proposed method.

Table 1. Computing time of the proposed algorithm.

Algorithm Step	Extracting	Tracking
Iteration [average]	26	5
Computing time[average]	0.374s	0.073s

Tracking results are shown in Fig.5 and Table 1. It is found that the proposed algorithm has a promising tracking result, which is robust against the teeth appearance. Meanwhile, the use of a 16-point deformable model to describe a lip shape is physically meaningful. In addition, the computing time of tracking one lip frame is less than the extracting process. When there exists a large lip sequences, it is effective to utilize the previous lip contour as the initial evolving curve of the proceeding one, which can reduce a large amount of computing time. From the results, our approach is feasible and effective.

5. CONCLUSION

In this paper, we have proposed an algorithm to track the lip movements using localized active contours and deformable models. This algorithm is adaptive to the lip movements, as well as robust against the appearance of teeth effect, it is very suitable for applications that require a high level of accuracy such as lip reading or speech recognition.

6. REFERENCES

[1] M.U. Ramos Sánchez, J. Matas, and J. Kittler, "Statistical chromaticity-based lip tracking with b-splines," in *Proc. of ICASSP*, 1997, vol. 4, pp. 2973–2976.

[2] A.W.C Liew, S.H. Leung, and W.H Lau, "Lip contour extraction from color images using a deformable model," *Pattern Recognition*, vol. 35, no. 12, pp. 2949–2962, 2002.

[3] G.I. Chiou and J.N Hwang, "Lipreading from color video," *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1192–1195, 1997.

[4] N. Eveno, A. Caplier, and P.Y. Coulon, "Accurate and quasi-automatic lip tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 5, pp. 706–715, 2004.

[5] S.L. Wang, W.H. Lau, and S.H. Leung, "Automatic lip contour extraction from color images," *Pattern Recognition*, vol. 37, no. 12, pp. 2375–2387, 2004.

[6] S. Lankton and A. Tannenbaum, "Localizing region-based active contours," *IEEE Transactions on Image Processing*, vol. 17, no. 11, pp. 2029–2039, 2008.

[7] 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*, vol. 13, pp. 195–216, 2002.

[8] M. Li and Y.M. Cheung, "Automatic lip localization under face illumination with shadow consideration," *Signal Process*, vol. 89, no. 12, pp. 2425–2434, 2009.