A ROBUST LIP TRACKING ALGORITHM USING LOCALIZED COLOR 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 lipreading recognition system. This paper presents a robust lip tracking algorithm using localized color active contours and deformable models. The proposed approach utilizes a combined semi-ellipse as the initial evolving curve and computes the localized energies in color space for evolving such that a separation of the original lip image into lip and non-lip regions can be found. Moreover, the dynamic radius selection of the local region is presented with a 16-point deformable model (Wang et al. 2004) to achieve the lip tracking. The proposed approach is adaptive to the movement of the lips from frame to frame, and robust against the appearance of the teeth and tongue. Experiments have shown the promising results.

Index Terms— Lip tracking, localized color active contour, localized energy, 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, teeth or tongue effect and so forth.

In the last decade, a few techniques have been proposed to realize lip tracking with the focus on segmentation of lip regions or extraction of lip contours. Basically, the point-based and region-based methods 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. By contrast, the region-based method can make the tracking results more robust. Typical

examples include deformable template (DT) [2], Active Contour Model (ACM) [3], Active Shape Model (ASM) and Active Appearance Model (AAM) [4]. The DT algorithm employs a cost function to partition a lip image into lip and non-lip regions via a parametric model. In general, the tracking performance of this method may deteriorate if a lip shape is irregular or 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 regional energy such that an object contour is obtained. However, this approach is sensitive to the parameter initialization and uneven illuminations. The ASM approach employs a set of landmark points to describe the lip shape, and these points are controlled within a few modes derived from a large training data. Often, it is laborious to perform a training process to determine the lip shape. So does the AAM algorithm, which is an extension of ASM algorithm incorporating the eigenanalysis. Latter, some researchers attempt to combine the merits of above approaches. Papers [5][6] have shown the desired tracking results in their application domain. Nevertheless, their performances are usually suffered from the appearance of teeth and tongue.

Until recently, when images have heterogeneous statistics or complex components, it is found that the localized active contour models (LACM) [7] can achieve a better segmentation result as shown in Fig. 1(a) and (c). Nevertheless, this model is susceptible to the appropriate selection of correlative parameters. Otherwise, its performance will deteriorate as shown in Fig. 1(b). In addition, this model does not consider the color information, which actually provides additional significant information to improve the detection performance.
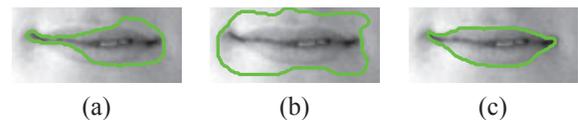
Fig. 1. (a) Conventional ACM based extracting result, (b) LACM based extracting result with the improper parameters, (c) LACM based extracting result with the proper parameters.

In this paper, we extend the LACM to the localized color active contour model (LCACM), and find a combined semiellipse as the initial evolving curve embedded in LCACM to

extract the lip contour of the first frame. Subsequently, the dynamic radius selection of local region is presented with a 16-point deformable model [6] to achieve lip tracking. Experiments have shown the efficiency of the proposed algorithm.

2. LCACM FRAMEWORK

This section introduces the LCACM framework extended from the work [7] provided that the foreground and background regions are locally different in color space. The advantage of this framework is that the complex appearances of color 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. $\mathcal{C}$ denotes a closed curve represented as the zero level set of a signed distance function ϕ, i.e., $\mathcal{C} = \{u \,|\, \phi(u) = 0\}$ [7]. The interior of $\mathcal{C}$ is specified by the following approximation of the smoothed Heaviside function:

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

Similarly, the exterior $\mathcal{C}$ 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:

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

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

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

Paper [7] lists three well-known examples of the energy function for representing the regional energies. In this section, we utilize mean separation energy for extension only, although the underlying techniques are surely applicable to other energies as well. Let $\mu_{in}(u)$ and $\mu_{out}(u)$ represent the vector-valued mean intensities in local interior and exterior regions localized by $\mathcal{B}(u,v)$ at a point u, respectively. Given D-dimensional measurements of a color space, i.e., $\mu_{in}(u) = \{\mu_{in}^1(u), \cdots, \mu_{in}^D(u)\}$, $\mu_{out}(u) = \{\mu_{out}^1(u), \cdots, \mu_{out}^D(u)\}$ and $I(v) = \{I_1(v), \cdots, I_D(v)\}$, the localized versions of the mean intensities $\mu_{in}^k(u)$ and $\mu_{out}^k(u)$ within the k channel of the color space can be formulated as follows:

$$\mu_{in}^k(u) = \frac{1}{A_{in}(u)}\int_{\Omega_v}\mathcal{B}(u,v)\cdot\mathcal{H}\phi(v)\cdot I_k(v)dv \quad (4)$$

$$\mu_{out}^k(u) = \frac{1}{A_{out}(u)}\int_{\Omega_v}\mathcal{B}(u,v)\cdot(1-\mathcal{H}\phi(v))\cdot I_k(v)dv \quad (5)$$

where $A_{in}(u) = \int_{\Omega_v}\mathcal{B}(u,v)\cdot\mathcal{H}\phi(v)dv$ and $A_{out}(u) = \int_{\Omega_v}\mathcal{B}(u,v)\cdot(1-\mathcal{H}\phi(v))dv$ represent the areas of the local interior and local exterior regions, respectively. Hence, a localized region-based energy in color space is obtained:

$$F_{MSC} = -\frac{1}{2}\cdot\sum_{k=1}^{D}(\mu_{in}^k(u) - \mu_{out}^k(u))^2. \quad (6)$$

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 arc length of the curve is penalized and weighted by a parameter λ, and the final energy $E(\phi)$ is given as follows:

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

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

$$\frac{\partial\phi}{\partial t}(u) = \delta\phi(u)\int_{\Omega_v}\mathcal{B}(u,v)\cdot\delta\phi(v)\cdot\sum_{k=1}^{D}(\mu_{in}^k(u) - \mu_{out}^k(u))$$
$$\cdot\left(\frac{(I_k(v) - \mu_{in}^k(u))}{A_{in}(u)} + \frac{(I(v) - \mu_{out}^k(u))}{A_{out}(u)}\right)dv$$
$$+\lambda\delta\phi(u)div(\frac{\nabla\phi(u)}{|\nabla\phi(u)|})\,\|\nabla\phi(u)\|. \quad (8)$$

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

3. THE PROPOSED ALGORITHM

The proposed lip tracking algorithm mainly includes the two phases: (a) lip contour extraction for the first frame, and (b) lip tracking in the subsequent frames.

3.1. Lip Contour Extraction

In LCACM, the initial evolving curve $\mathcal{C}$ 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 may lead to the inaccurate result as shown in Fig. 1(b).
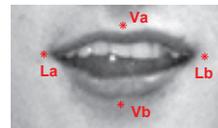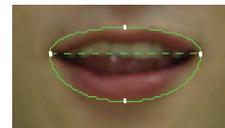

(a) Lip corner dots   (b) Combined semi-ellipse

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 utilized as the initial evolving curve embedded in LCACM. According to the previous work [5] [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 as follows:

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

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. Hence, we let the combined semi-ellipse be the initial evolving curve embedded in LCACM for lip contour extraction. As a rule of thumb, $r = \frac{b_{up}}{2}$ is appropriate to extract the lip contours in this phase. Please note that the value of r should become smaller if one opens the mouth widely.
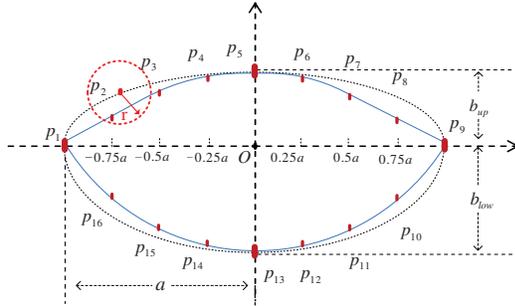
(a) Initial evolving curve with the 16-point lip model

(b) Upper lip thickness　　(c) Opening mouth
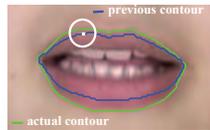
Fig. 3. Proper parameters in the proposed algorithm.

In addition, the extracted lip contours may not be smooth and are slightly lack of the geometric lip shape constraint as well. Therefore, we employ a 16-point geometric deformable model [6] with cubic spline interpolation connected to model the lip shape as shown in Fig. 3(a), which is always physically meaningful in tracking applications.

3.2. Lip Tracking

Often, the speech utterances are always comprised of many consecutive lip sequences in one second. In addition, a lip shape of one frame changes a little compared with the neighboring one. After accurately extracting the lip contour of the current lip frame, we can utilize it as the initial evolving curve embedded in LCACM to track 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 be inside the current one. As shown in Fig. 3(c), to avoid the effects of teeth and tongue appearance, the dynamic radius selection for lip tracking is proposed. 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(b). Therefore, we employ the dynamic parameter r_i, which can be embedded in LCACM as the local radius for tracking the lip movements, i.e., $r_i = \frac{Lt_{i-1}}{2}, i \geq 2$, where Lt_{i-1} expresses the middle thickness of the upper lip in the previous one. Let N_t denote the total number of points within the area adjacent to the evolving curve $\mathcal{C}$ at the t-th iteration. The lip tracking algorithm is given as follows:

Algorithm 1: The tracking algorithm

Input: , I_i, $\mathcal{C}_{i-1}$ % *The previous lip contour.*
Output: $\mathcal{C}_i$ % *The tracking contour .*
Lip image preprocessing, i.e., noise remove. Set $t = 0$;
Let $\mathcal{C}_{i-1}$ be the initial evolving curve;
for $(j = 1; j \leq N_t; j + +)$ **do**
　Compute the middle thickness Lt_{i-1} of I_{i-1};
　Assign proper r to $\mathcal{B}(u, v)$;
　Compute the localized energy $E(\phi(u))$;
　$\delta\phi(u_j) = \delta\phi(u_j) + \frac{\partial\phi}{\partial t}(u_j)$;
end
if *all* $E(\phi(u_j)) < \varepsilon, j \in [1, \cdots, N_t]$ **then**
　$\mathcal{C}_i = \{u \,|\, \phi(u) = 0\}$;
　Employ a 16-point deformable model to describe the current lip shape;
　Obtain the tracking result, break;
else
　$t = t + 1$; Go back to the **for** loop;
end

---

4. EXPERIMENTAL RESULTS

We implemented the proposed algorithm on an Intel® Core™2 Quad Q9450 2.66 GHz machine and applied with Matlab 7.0 image processing toolkit. As the Euclidean metrics and distances are perceptually uniform in CIE-1976 color space, we therefore projected the RGB lip images into the CIELAB color space. The L channel was performed with a 3×3 mean filter to reduce noise. In our experiments, we set the parameters $\lambda = 0.3$ and $\varepsilon = 1.3$.

Experiment 1: We have applied our lip contour extraction approach to the 200 frontal face images from the CVL face

database[1] 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. Meanwhile, we have also examined the unsatisfactory ones and found that they have either the very poor contrast between the lip and skin region, or obvious beard effects around the lips.
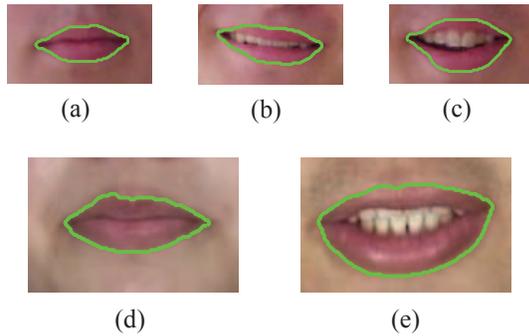
Fig. 4. LACM based contour extraction results: (a)(b)(c) from the CVL database, and (d)(e) from our laboratory database.

Experiment 2: We have performed the proposed lip tracking algorithm on a large number of face sequences captured from 10 speakers, who spoke English and Chinese digits (0-9) in a uniform illuminance environment. 30 colorized face frames and the located lip image of size 116×72 from the face sequences were recorded per second.
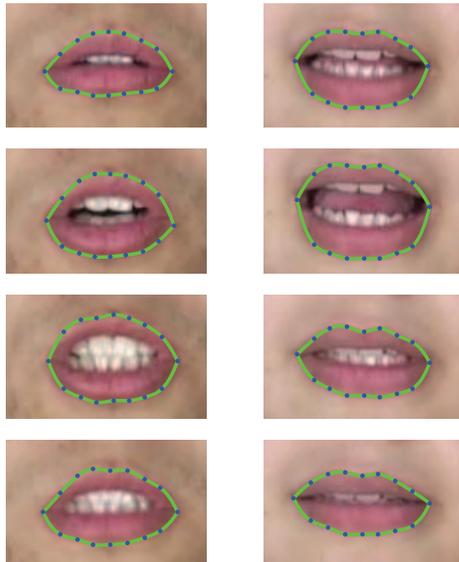
Fig. 5. Tracking result using the proposed method.

Table 1. Computing time of the proposed algorithm.

Algorithm Step	Extracting	Tracking
Iteration [average]	24	5
Computing time[average]	0.695s	0.103s

Tracking results are shown in Fig. 5 and Table 1. It is found that the proposed algorithm has achieved a promising tracking result, which is robust against the appearance of teeth and tongue. Furthermore, the utilization 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. In particular, when there exists a long lip sequences, it is effective to utilize the previous lip contour as the important parameter to track the current one. It is expected that such an operation can reduce a large amount of computing time.

5. CONCLUSION

In this paper, we have proposed a robust lip tracking algorithm using localized color active contours and deformable models. This approach is adaptive to the lip movements, and also robust against the appearance of teeth and tongue. Hence, it provides a promising way for lip tracking.

References

[1] M.U. Ramos Sánchez, J. Matas, and J. Kittler, "Statistical chromaticity-based lip tracking with b-splines," in *Proceedings of ICASSP'1997*, 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] 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*, vol. 24, no. 2, pp. 198–213, 2002.

[5] 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.

[6] 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.

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

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