Pattern Recognition 45 (2012) 3336–3347

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A local region based approach to lip tracking

Yiu-ming Cheung [a,*], Xin Liu [a], Xinge You [b]

[a] Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China
[b] Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, China

ARTICLE INFO

Article history:
Received 11 May 2011
Received in revised form
9 February 2012
Accepted 20 February 2012
Available online 5 March 2012

Keywords:
Lip tracking
Localized color active contour model
Semi-ellipse
Local region
Deformable model

ABSTRACT

Lip tracking has played a significant role in a lip reading system. In this paper, we present a local region based approach to lip tracking, which consists of two phases: (i) lip contour extraction for the first lip frame, and followed by (ii) lip tracking in the subsequent lip frames. Initially, we construct a localized color active color model provided that the foreground and background regions around the object are locally different in color space. In the first phase, we find a combined semi-ellipse around the lip as the initial evolving curve and compute the localized energies for curve evolution such that the lip image is separated into lip and non-lip regions. Then, we utilize a 16-point deformable model (Wang et al., 2004 [20]) with geometric constraint to achieve lip contour extraction. In the second phase, we present a dynamic selection of the radius of local regions associated with the extracted lip contour of the previous frame to realize lip tracking. The proposed approach not only adapts to the lip movement, but it is also robust against the appearance of teeth, tongue and black hole. Extensive experiments show the efficiency of the proposed lip tracking algorithm in comparison with the existing methods.

1. Introduction

Lip contour tracking (simply called lip tracking hereinafter) has received wide attention in recent years because of its potential applications in a variety of areas such as audio–visual speech recognition (AVSR) [1], lip reading [2,3], facial expression analysis [4], human computer interfaces [5] and so forth. Although various visual tracking methods have been developed in the literature, e.g., see [6], these methods are usually utilized to track the object positions, which may not be suitable for determining the variations of the lip contours. In fact, it is a non-trivial task to track the lip movements accurately due to its elastic shape and non-rigid motion, the large variations caused by different speakers, lighting conditions, low contrast between the lip and skin, teeth or tongue effect, and so forth.

In the past years, a few techniques have been proposed towards lip tracking with the focus on segmentation of lip regions or extraction of lip contours, which can be roughly classified into two categories: the edge-based approaches and the region-based approaches. The former basically utilizes the low level spatial cues such as edge and color information to track the lip movement. For instance, Zhang et al. [7] applied hue and edge information to achieve the mouth localization and segmentation. Eveon et al. [8] detected six key points, through which the fitting shapes connecting these points were obtained according to the edge information and color cues. In general, these two techniques work well under a desired environment, but their performances may deteriorate if the lips are glossy or their exists image noise. Moreover, Kass et al. [9], Delmas et al. [10] and Freedman et al. [5] introduced the applications of active contour model (ACM, i.e., snake) to detect the edge of the lip boundary via gradient descent technique. Unfortunately, this type of active contours often converge to the wrong result when the lip edges are indistinct or the lip is very similar to the skin region. Subsequently, Barnard et al. [11] integrated the edge-based ACM with 2D pattern matching technique to drive the energy minimizing spline onto the expected lip contours. However, such a method just employs a combination of two semi-elliptical shapes to model the lip shape, which may not fit the actual lip boundary quite well.

In contrast, the region-based approaches mainly utilize the regional statistic characteristics to realize lip tracking. Typical examples include deformable template (DT) [12–14], region-based ACM [15,16], active shape model (ASM) [17–19], and active appearance model (AAM) [2]. The DT algorithm utilizes a regional cost function to partition a lip image into the lip and non-lip regions via a parametric template, which represents the lip shape properly. The pioneering work introduced by Yuille [12] shows a lip template specified by a set of parameters, and these parameters are altered via an energy minimizing process so that the lip template can match the lip boundary gradually. Later, Liew et al. [13] addressed a different lip template and extended Yuille's

* Corresponding author. Tel.: +852 34115155.
E-mail addresses: ymc@comp.hkbu.edu.hk (Y.-m. Cheung), xliu@comp.hkbu.edu.hk (X. Liu), youxg@hust.edu.cn (X. You).

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2012.02.024

work by introducing a new cost function to realize lip contour extraction in color images, while Tian et al. [14] utilized a symmetrical DT to model the lip shape and formulated the color distribution inside the closed mouth region as a Gaussian mixture to regularize the DT. In general, the tracking performance of this kind of methods will be degraded if a lip shape is evidently irregular or when the mouth opens widely. The region-based ACM algorithm featuring on minimizing a regional energy function always outperforms the edge-based ACM for lip images with weak edges or without edges. For instance, Chiou et al. [15] modified the original ACM by adding eight radial vectors within the lip region to regularize the active contours driving to the lip boundary. Wakasugi et al. [16] applied the separability of regional color intensity distributions with ACM to achieve lip contour extraction. Nevertheless, it has been found that these methods often suffer from the complex components in oral cavity and are highly dependent on the parameter initialization. The ASM approach adopts a set of landmark points to describe the lip shape, and these points are controlled within a few modes derived from a training data set. For example, Luettin et al. [17] applied a set of manually labeled points with ASM to train the possible lip shapes. Sum et al. [18] presented an optimization procedure from a point-based model using ASM for extracting the lip contours. Nguyen et al. [19] integrated multi-features of lip regions with ASM to learn lip shapes. The AAM algorithm proposed by Matthews et al. [2] is an extension of ASM algorithm incorporating the eigenanalysis in gray-level case. Often, the ASM and AAM are both quite laborious to establish a training data set with manually cautious calibration and perform a training process to determine the lip shapes. Meanwhile, these methods may not be able to provide a good match to those lip shapes that are quite distinct from the training data. It is therefore unsuitable for the robust lip tracking applications from a practical viewpoint.

In recent years, lip image analysis in color space, e.g., CIELAB, CIELUV and HSV, has received much attention as the color can provide additional significant information that is not available in gray-level cases. Wang et al. [20] generated probability map of lip region in color space via fuzzy clustering method incorporating shape function (FCMS) and developed an iterative point-driven optimization scheme to fit the lip boundary based on pre-generated probability map. Subsequently, Leung et al. [21] further extended the above work with an elliptic shape function to segment the lip region in color space. Similar and related works can be found in [22,23]. It is found that this kind of methods can significantly simplify the detection and location of the lip regions. Nevertheless, as the distributions of skin, tongue and lip may overlap and diversify among different speakers, it may make such a method inaccurate and unstable to achieve lip segmentation or lip contour extraction, particularly in the case of mouth opening widely. Meanwhile, the implementation of these methods often suffers from the appearance of tongue or black hole as shown in Fig. 1, although multiple pre-processing procedures can reduce the teeth effect.

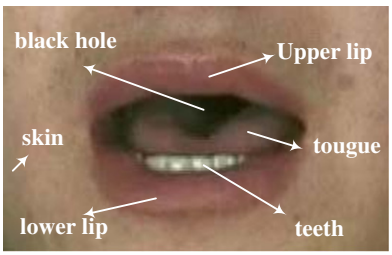

Fig. 1. A lip region incorporated the appearance of teeth, tongue and black hole in oral cavity.

More recently, Eveno et al. [24] attempted to combine the merits of the above-stated approaches and proposed a jumping snake with a parametric model composed of four cubic curves to achieve lip tracking. It is effective in most cases, but which is highly dependent on pre-and-post-processing techniques and adjustment process to make the model match the lip shape appropriately. Differing from the above region-based approaches, Jian et al. [25] addressed a modified attractor-guided particle filtering framework to track the lip contours. Unfortunately, such a method needs to segment a set of representative lip contours manually as the shape priors in advance. Furthermore, Ong et al. [26] proposed a learnt data-driven approach via linear predictors to track the lip movements, but which needs a data set composed of different types of lip shapes in advance. Further, this method, as well as the one in [25], involves the complicated iterative learning to match the lip shape, whose computation is time-consuming.

Thus far, almost all the region-based approaches involve the globally statistical characteristics. Subsequently, their performance may deteriorate upon the appearance of teeth, tongue or black hole. Until very recently, when object in an image has heterogeneous statistics or complex components, it is found that the localized active contour model (LACM) [27], which utilizes the local statistical characteristics, can generally achieve a better segmentation result as shown in Figs. 2 and 3(d). Nevertheless, this model highly depends on the appropriate selection of correlative parameters. Often, the improper parameters, e.g., ulterior evolving curve with small local radius or proper evolving curve with large local radius, could lead to erroneous extractions as shown in Fig. 3(c). In addition, Ref. [27] does not consider the prior knowledge about color information, which actually provides more information to improve the extraction performance, especially when the images are shadowed, shaded and highlighted [28,29].

In this paper, we present a local region based approach to lip tracking with two phases: (i) lip contour extraction for the first lip frame, and followed by (ii) lip tracking in the subsequent lip frames. Initially, we introduce a new kind of active contour model, namely *localized color active color model* (LCACM), provided that the foreground and background regions around the object are locally different in color space. In the first phase, we find a combined semi-ellipse around the first lip image as initial evolving curve and compute the localized energies for curve evolution such that the lip image is separated into lip and non-lip regions. Then, we utilize a 16-point deformable model [20] with geometric constraint to achieve lip contour extraction. In the second phase, we present a dynamic selection of the radius of local regions associated with the extracted lip contour of the previous frame to realize lip tracking. The proposed approach is adaptive to lip movement, and robust against the appearance of

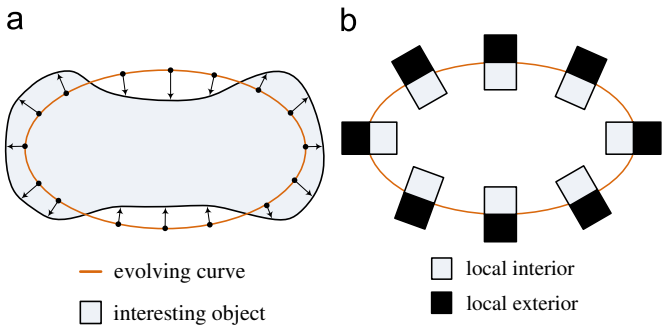

Fig. 2. Graphical representation of the active contour model: (a) evolving curve with diverging directions along the arrow; (b) the description of local interior and local exterior region.

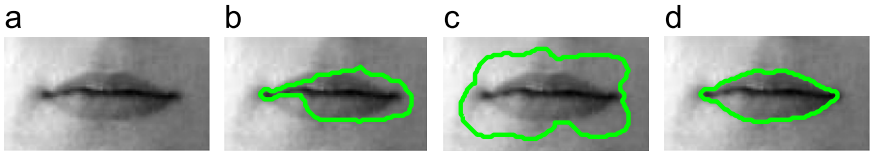

Fig. 3. (a) Original lip image with some uneven illuminations and noise effects; (b) conventional region-based ACM extraction result; (c) LACM based extraction result with the improper parameters; (d) LACM based extraction result with the proper parameters.

the teeth, tongue and black hole. Experimental results have shown the promising results.

The remainder of this paper is organized as follows: Section 2 will overview the LACM and its extension to the LCACM. Section 3 goes into the details of describing the proposed approach, in which the illumination equalization, the appropriate selection of parameters including the initial evolving curve, local radius and lip model, are presented. In Section 4, experimental results are conducted to compare the proposed approach with the existing methods. Finally, we draw a conclusion in Section 5.

2. Localized color active contour model

This section introduces the framework of LCACM extended from the work proposed by Lankton [27,30] provided that the foreground and background regions around the object are locally different in color space. As shown in Fig. 2(b), given a proper initial evolving curve, the local regions centered at each of the points along the curve can be split into the local interior and local exterior, respectively. Accordingly, a set of localized energies can be computed. By minimizing those energies, the evolving curve can gradually converge to the boundary of the object. The advantage of this framework is that the objects associated with complex appearances or intensity inhomogeneities can be successfully segmented using the localized energies, while the corresponding global energies may fail.

2.1. Overview of the LACM

Let I denote a pre-specified image defined on the domain Ω, and $\mathcal{C}$ denote a closed curve represented as the zero level set of a signed distance function (SDF) ϕ, i.e., $\mathcal{C}=\{u|\phi(u)=0\}$ [27]. The interior of $\mathcal{C}$ is specified by the following approximation of the smoothed Heaviside function:

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

Similarly, the exterior $\mathcal{C}$ can be defined as $(1-\mathcal{H}\phi(u))$. The derivative of $\mathcal{H}\phi(u)$, which is a smoothed version of the Dirac delta function:

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

is utilized to specify the area adjacent to the curve $\mathcal{C}$. For simplicity, parameters u and v are utilized as the two independent spatial variables, each of which represents a single point within an image. 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:

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

By ignoring the complex appearances of an image that may arise outside the local region, the contributions along the evolving curve within the local region are considered. In addition, an energy function is defined in terms of a generic force function F, which is a generic internal energy metric to measure the local adherence to a given model at each point along the curve [30]. Hence, the evolutionary energy $E(\phi)$ can be formulated as follows:

$$E(\phi)=\int_{\Omega_u}\delta\phi(u)\int_{\Omega_v}\mathcal{B}(u,v)\cdot F(I(v),\phi(v))\,dv\,du. \tag{4}$$

In Eq. (4), the purpose of the multiplication with the Dirac function $\delta\phi(u)$ in the outer integral over u is utilized to capture a much broader range of objects. It ensures that the curve will not change topology by spontaneously developing new contours, although it still allows contours to be split and merged. Finally, to keep the curve smooth, a regularization term is added into the localized energies. Specifically, the arc length of the curve is penalized and weighted by a parameter λ. The resulting energy $E(\phi)$ is formulated as follows:

$$E(\phi)=\int_{\Omega_u}\delta\phi(u)\int_{\Omega_v}\mathcal{B}(u,v)\cdot F(I(v),\phi(v))\,dv\,du + \lambda\int_{\Omega_u}\delta\phi(u)\|\nabla\phi(u)\|\,du. \tag{5}$$

By taking the first variation of energy F with respect to ϕ, the following evolution equation can be obtained [30]:

$$\frac{\partial\phi}{\partial t}(u)=\delta\phi(u)\int_{\Omega_v}\mathcal{B}(u,v)\cdot\nabla_{\phi(v)}F(I(v),\phi(v))\,dv + \lambda\delta\phi(u)div\left(\frac{\nabla\phi(u)}{|\nabla\phi(u)|}\right)\|\nabla\phi(u)\|. \tag{6}$$

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

2.2. An extension to LCACM

By taking into account the color information, the framework of LACM can be extended to the color space in case I represents a color image, which provides additional significant information that is unavailable in gray-level space. Ref. [27] lists three well-known examples of the energy function for representing the regional energies, i.e., uniform modeling (UM) energy [31], mean separation (MS) energy [32], and histogram separation (HS) energy [33]. In this Sub-section, we extend MS energy only to color space and embed it into the framework of LACM. It should be noted that the underlying techniques are surely applicable to the UM and HS energies as well.

For simplicity, the localized equivalents of $\mu_{in}(u)$ and $\mu_{out}(u)$ marked by $\mathcal{B}(u,v)$ at a point u represent the vector-valued mean intensities of local interior and local exterior regions, respectively. Given D-dimensional measurements of a color vector, i.e.,

$$\mu_{in}(u)=\{\mu_{in}^1(u),\ldots,\mu_{in}^D(u)\}, \tag{7}$$

$$\mu_{out}(u) = \{\mu^1_{out}(u), \ldots, \mu^D_{out}(u)\}, \tag{8}$$

$$I(v) = \{I_1(v), \ldots, I_D(v)\}, \tag{9}$$

the localized versions of the mean intensities $\mu^k_{in}(u)$ and $\mu^k_{out}(u)$ within the k channel of the color space can be formulated as follows:

$$\mu^k_{in}(u) = \frac{1}{A_{in}(u)} \int_{\Omega_v} \mathcal{B}(u,v) \cdot \mathcal{H}\phi(v) \cdot I_k(v)\, dv, \tag{10}$$

$$\mu^k_{out}(u) = \frac{1}{A_{out}(u)} \int_{\Omega_v} \mathcal{B}(u,v) \cdot (1-\mathcal{H}\phi(v)) \cdot I_k(v)\, dv, \tag{11}$$

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 formed from MS energy [32] is obtained:

$$F_{MSC} = -\frac{1}{2} \cdot \sum_{k=1}^{D} (\mu^k_{in}(u) - \mu^k_{out}(u))^2. \tag{12}$$

By substituting the derivative of F_{MSC} into Eq. (6), the local region-based flow can be computed through the following level set evolution:

$$\begin{aligned}\frac{\partial \phi}{\partial t}(u) = {} & \delta\phi(u) \int_{\Omega_v} \mathcal{B}(u,v) \cdot \delta\phi(v) \cdot \sum_{k=1}^{D} \Big((\mu^k_{in}(u) - \mu^k_{out}(u)) \\ & \cdot \left(\frac{(I_k(v) - \mu^k_{in}(u))}{A_{in}(u)} + \frac{(I(v) - \mu^k_{out}(u))}{A_{out}(u)} \right) \Big) dv \\ & + \lambda \delta\phi(u) div\left(\frac{\nabla\phi(u)}{|\nabla\phi(u)|} \right) \|\nabla\phi(u)\|. \end{aligned} \tag{13}$$

Note that the optimal energy for evolution can be obtained when $\mu_{in}(u)$ and $\mu_{out}(u)$ are the most different at each point u along the contour. In the evolutionary process, the value of the SDF $\phi(u)$ is updated iteratively until the evolving curve converges to the object boundary. Consequently, the object can be segmented through the zero level set $\{u | \phi(u) = 0\}$. This framework can be effectively utilized for segmenting the color objects associated with heterogeneous statistics or complex appearances.

3. The proposed lip tracking algorithm

The proposed lip tracking algorithm consists of two phases: (i) lip contour extraction for the first lip frame, and followed by (ii) lip tracking in the subsequent lip frames. First, we introduce an effective illumination equalization method to reduce lighting asymmetry. Then, we find a combined semi-ellipse contiguous to the lip region as the initial evolving curve for evolution such that the lip image can be segmented into lip and non-lip regions. Subsequently, we utilize a deformable model with the geometric constraint to achieve lip contour extraction. Finally, we present a dynamic selection of the radius of local regions associated with the extracted lip contour of the previous frame to realize lip tracking.

3.1. Illumination equalization

Illumination is one of the most significant factors that affect the appearance of an image. It often leads to heterogeneous intensities due to the different albedos of the object surface and the shadows cast from different illumination directions, which can be regarded as the uneven background, illumination asymmetry, and also known as non-uniform intensity distribution. Often, during the video capture of the lip motions, the unbalanced lighting conditions falling on different directions may cause the uneven illuminations frequently, which always lead to intensity heterogeneous. Hence, illumination equalization has played an important role in image analysis and processing. Liew et al. [13] has introduced an effective way to reduce the effects of uneven illumination provided that the illumination falls along the vertical direction. This method just adopts the luminance value of single point along the boundary, which may be susceptible to the image noise. Under such circumstances, we therefore propose an improved illumination equalization method via the analysis of local regions along the image boundary, which is more robust against noise and adaptive to the multifarious illumination directions.

We first consider two regular illumination directions: (1) the horizontal direction as shown in Fig. 4(a), and (2) the vertical direction as shown in Fig. 4(b). Given a located lip image of size $m \times n$, let $L(i,j)$, $\hat{L}(i,j)$ represent the luminance value before and after illumination equalization, respectively. Note that the uneven illumination can be regarded as the linear along its direction. Instead of utilizing the intensity value of a single boundary point, the mean value within a local region of size $(2p+1) \times (2q+1)$ is employed in our proposed method. Consequently, the luminance value adjusted by illumination equalization can be mathematically formulated as follows (see Appendix A for details):

Illumination direction (1):

$$\hat{L}(i,j) = \begin{cases} L(i,j) + \dfrac{(n-2j+1) \cdot (r(p) - l(p))}{2(n-1)}, & i \in [1,p), \\ L(i,j) + \dfrac{(n-2j+1) \cdot (r(m-p) - l(m-p))}{2(n-1)}, & i \in (m-p,m], \\ L(i,j) + \dfrac{(n-2j+1) \cdot (r(i) - l(i))}{2(n-1)}, & i \in [p,m-p]; \end{cases} \tag{14}$$

Illumination direction (2):

$$\hat{L}(i,j) = \begin{cases} L(i,j) + \dfrac{(m-2i+1) \cdot (b(q) - t(q))}{2(m-1)}, & j \in [1,q), \\ L(i,j) + \dfrac{(m-2i+1) \cdot (b(n-q) - t(n-q))}{2(m-1)}, & j \in (n-q,n], \\ L(i,j) + \dfrac{(m-2i+1) \cdot (b(j) - t(j))}{2(m-1)}, & j \in [q,n-q], \end{cases} \tag{15}$$

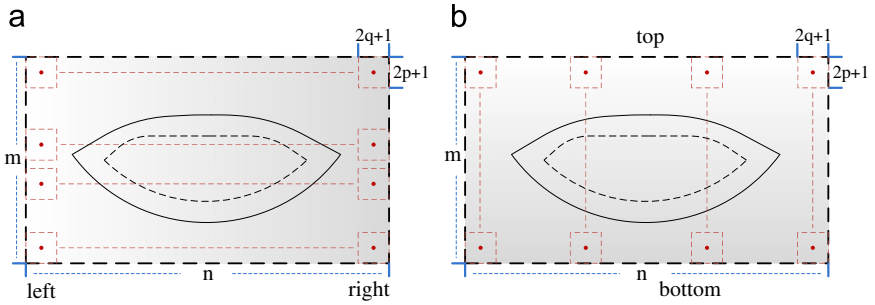

Fig. 4. Two regular illumination directions: (a) horizontal direction; (b) vertical direction.

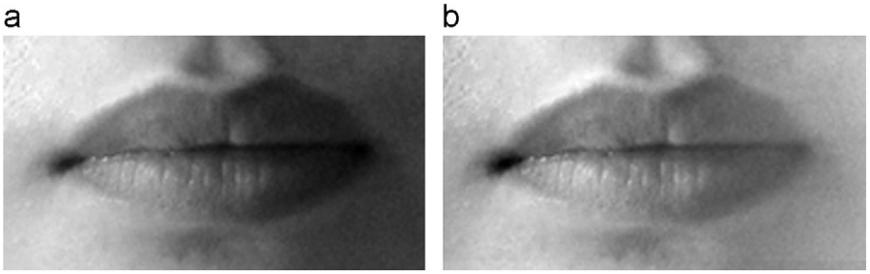

Fig. 5. Examples of illumination equalization: (a) original lip image with uneven illumination effect; (b) lip image after the adjustment via illumination equalization.

where $l(i)$ and $r(i)$ denote the mean intensity of left and right borders within a local region of size $(2p+1)\times(2q+1)$ at the i-th row, respectively. Similarly, $t(j)$ and $b(j)$ denote the mean intensity of top and bottom borders at the j-th column individually. The size of local region can be optionally set at 3×3, 3×5 or 5×3 according to the image scale requirement. To demonstrate the effectiveness of the proposed illumination equalization method, Fig. 5(a) and (b) show an example of the lip image before and after illumination equalization, where the uneven illumination is along the horizontal direction. It can be clearly observed that the uniform illumination can be obtained and the dark part of the right region has been significantly reduced. Further, if the illumination direction is not the previous-stated one, e.g., diagonal direction, we can perform the illumination equalization with illumination direction (a) and (b).

3.2. Lip contour extraction

Lip contour extraction is of crucial importance to the lip tracking system. However, the opening mouth incorporates the components like teeth, tongue and black hole, which always causes the complex statistical characteristics in the whole mouth region. In LCACM, the appropriate selection of the parameters such as the initial evolving curve $\mathcal{C}$ and local radius r will play an important role in determining the statistical characteristics for lip contour extraction. Intuitively, the initial evolving curve that contiguously encircles the lips will reduce the evolutionary process undoubtedly.

In the last decade, some researchers have successfully employed an elliptic shape function to model the lip shapes [21]. Evidently, almost all the lips can be encircled within an elliptic region. Nevertheless, in case the mouth opens widely, some marginal parts of the elliptic region may be far away from the lip boundary. As investigated by Mark et al. [11] and shown in Fig. 6(b), the biological lip shapes of the upper and lower lip are usually different, which can be modeled by a combination of two semi-elliptical shapes. Such a combined semi-ellipse can be better contiguous to the lip boundary compared with a single elliptic shape, which can therefore be utilized as the initial evolving curve embedded into LCACM framework for lip contour extraction.

According to Refs. [34,35], the primary lip corner dots can be successfully detected through the intensity variations and color cues. As shown in Fig. 6(c), the upper lip usually has the three geometric corner points due to the Cupidon's bow, i.e., dip point, left peak point and right peak point. To construct the upper semi-ellipse, only one upper corner dot is needed. Therefore, we utilize the middle value between the left peak and right peak to represent the upper point for constructing the upper semi-ellipse. From the practical viewpoint, the lip corner dots need not exactly fit the geometric position of lip structure. Instead, as shown in Fig. 6(a), an approximate position within a local region is enough. Therefore, we label the left corner, right corner, upper corner and lower corner points as *La*, *Lb*, *Va* and *Vb*, and let (x_c,y_c) be the

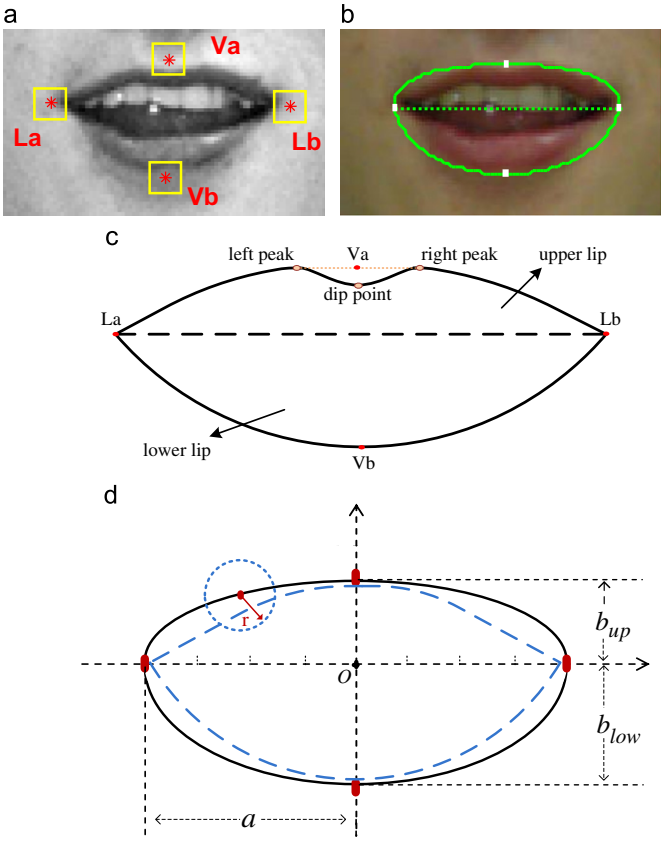

Fig. 6. The construction of the combined semi-ellipse: (a) lip corner dots detection; (b) a combined semi-ellipse around the lip; (c) standard lip model with a dip point and two peak points; (d) geometric description of a combined semi-ellipse.

origin center of the combined semi-ellipse, whose mathematical equations are formulated as follows:

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

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

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

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

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

$$\begin{bmatrix}X\\Y\end{bmatrix}=\begin{bmatrix}\cos\theta & \sin\theta\\-\sin\theta & \cos\theta\end{bmatrix}\cdot\begin{bmatrix}x-x_c\\y-y_c\end{bmatrix},$$

$$\frac{X_{up}^2}{a^2}+\frac{Y_{up}^2}{b_{up}^2}=1, \quad \frac{X_{low}^2}{a^2}+\frac{Y_{low}^2}{b_{low}^2}=1, \tag{16}$$

where a is the semi-major axes, b_{up} and b_{low} are the upper and lower semi-minor axes, respectively. θ is the inclined angle, which is positively defined at the counter-clockwise direction. Consequently, as shown in Fig. 6(b) and (d), such a combined semi-ellipse will be utilized as the initial evolving curve embedded in LCACM for lip contour extraction.

Further, the radius of local region marked by $\mathcal{B}(u,v)$ is another crucial parameter in LCACM to be considered when computing localized energies. As introduced in [27], the radius of the local region should be chosen based on the scale of the object of interest and the proximity of the surrounding clutter. As the pre-specified evolving curve is contiguous to the lip boundary, by a rule of thumb, it is effective to set $r=b_{up}/2$ for lip contour extraction in most cases, which is robust against the clutter such as the teeth, tongue or black hole. Note that the value of r should become smaller if the lip image incorporates the mouth opening widely. Although the selected local interior region may contain a very small fraction of the complex components of the oral cavity, it cannot affect the extraction performance while the large part may fail.

3.3. Lip model

In general, the extracted lip contours using above-stated method may be lack of geometric constraint occasionally. In the past years, some researchers have adopted different models to keep the geometric shape of lip contours. For example, Ref. [14] adopts four key points with two parabolas to model the outer lip contour, whereas Ref. [24] employs six key points with the cubic curves connected to describe the lip shape. In this paper, we adopt a 16-point geometric deformable model proposed in [20] to model the lip shapes, which is more flexible and physically meaningful in comparison with the less points based lip models. As shown in Fig. 7, the lip shape can be partitioned into three parts, in which the point set: $\{p_1, p_2, p_3, p_4, p_5\}$ represents the part of upper-left lip, while $\{p_5, p_6, p_7, p_8, p_9\}$ and $\{p_1, p_{16}, p_{15}, p_{14}, p_{13}, p_{12}, p_{11}, p_{10}, p_9\}$ describe the part of upper-right lip and lower lip, respectively. Its high flexibility and variability can model the lip shapes even for quite asymmetric or variational mouths.

Nevertheless, it is found that the dip point p_5 in this 16-point model may not be well obtained to represent the geometric position in some cases, which is influenced by the clutter around the dip point due to the convex structure of lips. Hence, p_5 is always constrained by the points p_4 and p_6, i.e., the vertical coordinate value of point p_5 is not more than points p_4 and p_6. The constrained condition can be formulated as

$$p_5(y)=\begin{cases} p_5(y) & \text{if } p_5(y)\le \min\{p_4(y),p_6(y)\},\\ \min\{p_4(y),p_6(y)\} & \text{otherwise.}\end{cases} \tag{17}$$

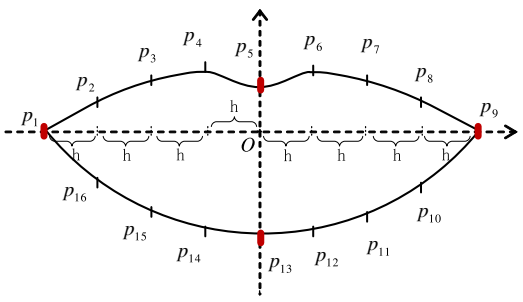

Fig. 7. A 16-point lip model introduced by Wang et al. [20].

Moreover, we employ the cubic spline interpolation (CSI) connecting the key points to piecewise approximate the lip shape [36,37], which always offers true continuity and smoothness between segments (see Appendix B for the details). Such an operation is physically meaningful in the vision, which can be well utilized for both lip contour extraction and lip tracking applications.

3.4. Lip tracking

Given a sequence of lip motion frames, a lip shape of one frame only changes a little compared with the neighboring one. Hence, as shown in Fig. 8, after extracting the lip contour of the previous lip frame, we let it be the initial evolving curve embedded in LCACM to track the lip contour of the current frame.

Nevertheless, as shown in Fig. 8(f), the extracted lip contour of the previous frame may be inside the current one, especially in the process of opening a mouth. Under the circumstances, an inappropriate selection of radius, i.e., large radius, will make the evolution curve diverging to an inaccurate direction. Therefore, the dynamic selection of the radius of the local region is addressed for lip tracking.

Algorithm 1. The lip tracking algorithm.

Input:
Lip frame image I_{i-1}, $I_i \in \Omega$.
The extracted contour $\mathcal{C}_{i-1}$ of lip image I_{i-1}.
Pre-processing:
Lip image pre-processing, i.e., noise remove, illumination equalization.
Let $\mathcal{C}_{i-1}$ be the zero level set of a SDF ϕ, i.e., $\mathcal{C}_{i-1}=\{u|\phi(u)=0\}$.
Set $t=0, temp=3, \varepsilon=1.3, \eta=0.05$.
Begin:
Compute the middle thickness Lt_{i-1} of I_{i-1} via Eq. (6).
while $temp>\eta$ **do**
  Specify the points adjacent to the evolving curve of number N_t via Eq. (2).
  for $j=1; j\le N_t; j++$ **do**
    Assign proper r to the $\mathcal{B}(u_j,v)$ via Eq. (18).
    Compute the localized energy $E(\phi(u_j))$ via Eq. (12).
    Compute $\frac{\partial\phi}{\partial t}(u_j)$ via Eq. (13).
    $\phi(u_j)=\phi(u_j)+\frac{\partial\phi}{\partial t}(u_j)$.
  end for
  $temp=\max|\frac{\partial\phi}{\partial t}(u_j)|, j\in[1,\ldots,N_t]$.
  $t=t+1$ % *Iteration times.*
end while
$\mathcal{C}_i=\{u|\phi(u)=0\}$. % *The tracking contour of current lip frame.*
end.
Output:
Obtain the tracking result $\mathcal{C}_i$ with model and geometric constraint via Eq. (17).

To the best of our knowledge, the middle thickness of the upper lip is usually smaller than the lower one, which may change during the speaking. Accordingly, we take this middle-upper lip thickness Lt as an effective parameter in our algorithm. As shown in Fig. 8(c), according to the pre-specified lip model, Lt can be approximately estimated along the line segment op_5 via the intensity variations processed with a low-pass filter [34]. Meanwhile, Lt is also constrained by Eq. (18) simultaneously, where the position of central point o is computed through the coordinate value of points p_1 and p_9, i.e., $o(x)=(p_1(x)+p_9(x))/2$, $o(y)=(p_1(y)+p_9(y))/2$. Therefore, we

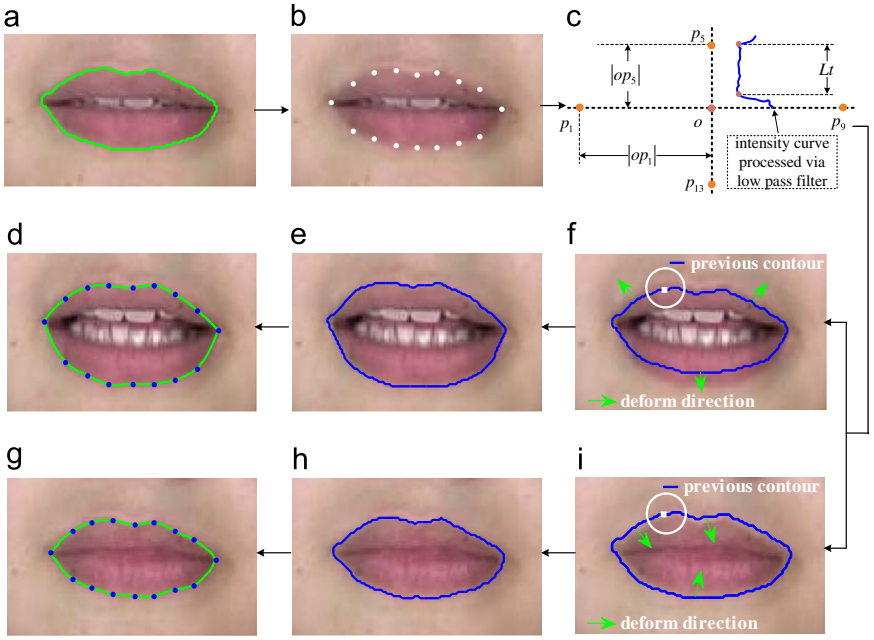

Fig. 8. (a) Lip contour extraction of the previous lip frame; (b) 16 key points obtained according to the lip model; (c) computation of the parameter Lt; (f and i) current lip frame with mouth opening and closing, respectively; (e and h) lip contour extraction; (d and g) geometric constraint with lip model.

let the dynamic parameter r_i be

$$\begin{cases} r_i = \dfrac{Lt_{i-1}}{2}, & i = 2, \ldots N_f, \\ Lt_j \le \min\{|(op_5)_j|, \frac{1}{2}|(op_1)_j|\}, & j = 1, \ldots \cdots N_f - 1, \end{cases} \tag{18}$$

where Lt_{i-1} represents the middle-upper lip thickness of the previous frame and N_f denotes the total number of frames. The value of r_i can be well utilized in LCACM as the local radius for tracking the current lip contour. The lip tracking algorithm is given in Algorithm 1, where η is a small threshold value that determines the stop condition of the evolving process. Similar to the case of lip contour extraction phase, the local region with the dynamic radius selection may also contain a very small fraction of the complex components of oral cavity, which, however, may not affect the tracking performance.

4. Experiments

The proposed lip contour extraction and lip tracking methods have been implemented and tested on a large number of lip images and videos. The experiments ran on an Intel® Core™2 Quad Q9450 2.66 GHz machine with Matlab R2007b image processing toolkit. In our experiments, we projected the RGB lip images into the CIELAB color space upon the fact that the Euclidean distances are perceptually uniform in CIE-1976 color space while non-uniform in RGB space [29]. We employed a 3×3 mean filter to reduce the noise effect and executed the illumination equalization in L channel. MS energy was employed for computing the regional energy, and the parameter λ was set at 0.3.

4.1. Experiment 1

We have applied the proposed lip contour extraction method to the 200 frontal face images from the CVL face database [38] and 500 face images from our laboratory database. To measure the difference between the extracted contour and real contour, we manually drew the lip contours to present the comparisons of contour matching performance in terms of the mean square error (MSE):

$$MSE = \frac{1}{N_p} \sum_{i=1}^{N_p} \sqrt{(x_i - x_i^*)^2 + (y_i - y_i^*)^2}, \tag{19}$$

where x_i, y_i are the coordinate values of the lip contour point i obtained by the proposed method, x_i^*, y_i^* are the actual coordinate values obtained by precisely manual annotation, and N_p denotes the total number of the selected contour points. In addition, the extraction performance tested on each database was measured by n_c/n_t, where n_t is the total number of the test database and n_c is the correctly extracted number with the value of MSE less than a pre-defined threshold Δc. In this paper, according to the image scale of the test database, parameter Δc is set at 2 and 3 for each test database, respectively. We set N_p at 16 and let the positions of the annotated points be the same as the lip model presented in Section 3.3.

Two snapshot results of lip contour extraction are shown in Figs. 9 and 10, in which the original lip images associated with the opening mouth incorporate the appearance of teeth and tongue, and also involve some certain light reflections acted on the lips. The scale of these two lip images are of size 101×146 and 126×160, respectively. It can be seen that the lip contour extraction results obtained by Kass et al. [9] usually obtained an inaccurate result when the lip edges are indistinct in grey level case. The other existing methods, e.g., Liew et al. [13] aiming at minimizing a cost function with a deformable model, Werda et al. [39] employing the color and geometric based model, Wakasugi et al. [16] utilizing the separability of multi-dimensional distributions with globally statistic characteristics, are susceptible to the complex appearance in the oral cavity and unable to make a good match of the real lip boundary, especially in extracting very irregular lip shapes. Another way to achieve lip contour extraction via lip segmentation using FCMS [21] is shown in Figs. 9(d) and 10(d), respectively. Although the teeth can be easily detected due to its striking difference from the surroundings, the membership distributions may fail to reflect the lip region accurately due to the effect of tongue and black hole in oral cavity, lighting reflection and other clutters around the lips. In addition, it is difficult

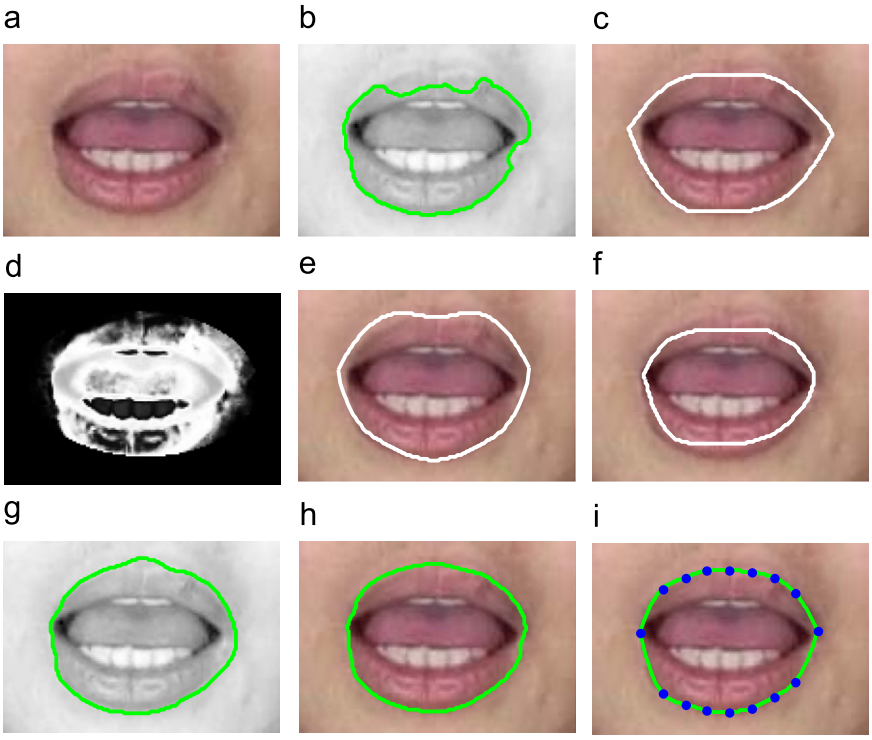

Fig. 9. (a) Original RGB lip image; (b) lip contour extraction obtained by Kass et al. [9]; (c) lip contour extraction obtained by Liew et al. [13]; (d) membership distribution obtained by Leung et al. [21]; (e) lip contour extraction obtained by Werda et al. [39]; (f) lip contour extraction obtained by Wakasugi et al. [16]; (g) lip contour extraction obtained by LACM [27] with proper parameters in grey level case; (h) lip contour extraction obtained by LCACM; (i) lip contour extraction obtained by the proposed method.

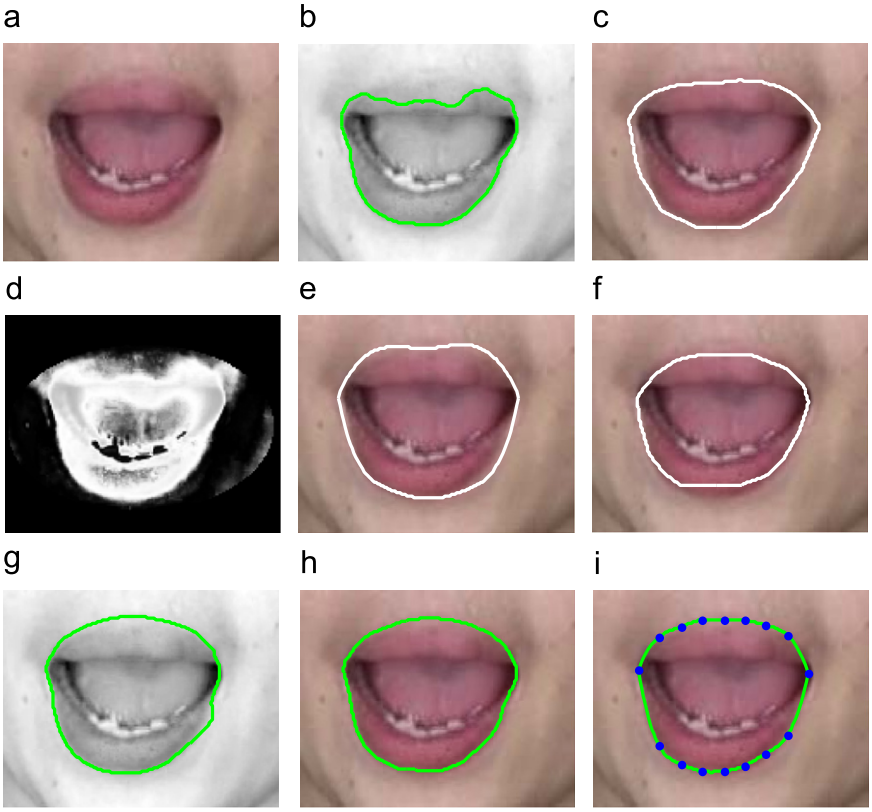

Fig. 10. (a) Original RGB lip image; (b) lip contour extraction obtained by Kass et al. [9]; (c) lip contour extraction obtained by Liew et al. [13]; (d) membership distribution obtained by Leung et al. [21]; (e) lip contour extraction obtained by Werda et al. [39]; (f) lip contour extraction obtained by Wakasugi et al. [16]; (g) lip contour extraction obtained by LACM [27] with proper parameters in grey level case; (h) lip contour extraction obtained by LCACM; (i) lip contour extraction obtained by the proposed method.

to draw a satisfactory lip contour along this line. By contrast, as shown in Table 1, the proposed lip contour extraction method has achieved a more satisfactory result with the smallest value of contour matching performance. This means that the proposed approach can well match the actual lip boundaries in comparison with the above-mentioned methods. It should be pointed out that the proposed method need not any image pre-processing technique to reduce the teeth or tongue effect. The analysis of the statistical characteristics in color space can provide more significant information that is unavailable in grey-level case, which can make the extracting performance more robust as shown in Figs. 9(g,h) and 10(g,h), respectively. Furthermore, the utilization of a 16-point deformable model with geometric constraint to describe a lip shape is physically meaningful in the vision.

More experimental examples are shown in Fig. 11. It can be clearly observed that the lip contours with opening mouth, rotation or deformation, can be accurately extracted. As shown in Table 2, the average extraction performance obtained by Kass et al. [9] was just 76.7%, which often failed to match the actual lip boundary appropriately due to the low contrast along the lip edges. The other methods, i.e., Liew et al. [13], Leung et al. [21], Werda et al. [39] and Wakasugi et al. [16], are all often degraded their performance especially in extracting very irregular lip shapes. Further, it can be found that these methods aiming at investigating the edge information or global statistical characteristics are somewhat sensitive to the uneven illuminations and susceptible to the complex appearance in oral cavity as well. In contrast, the proposed method aiming at local region analysis can successfully avoid the complex appearance in oral cavity such that the extracted lip contours can well match the real lip boundary. The average extraction performance is reached up to 96.1%. From the experimental results, the proposed lip contour extraction method is tolerant to the uneven illumination, rotation, deformation, the appearance of teeth, tongue and black hole.

Simultaneously, we have also investigated the unsatisfactory results (i.e., 3.9% of the test database), and found that they all have the very poor contrast between the lip and surrounding skin regions, or have obvious mustaches and beards around the lip regions.

4.2. Experiment 2

We have conducted the proposed lip tracking algorithm on a large number of speaking videos collected by our laboratory in a relatively uniform illumination environment. Image acquisition was performed using an HD-capable camera (SONY HDR-CX110) with the capturing frame-rate at 30 fps and the scale size of located lip images was set at 72×116.

Three representative groups of tracking results are shown in Figs. 12–14, respectively. It can be observed that the lip contours can be successfully tracked using the proposed algorithm in the process of mouth opening and closing. As shown in Fig. 13, although the thickness of the lip is changing over the lip movements, the proposed approach with the dynamic selection of radius can effectively prevent the complex components that may appear in the local region such that the accurate tracking results are obtained. We compared the proposed method with four existing approaches (i.e., [14,11,20,24]) via the tracking performance, which is measured by the average contour matching performance $\tilde{MSE}$ within a group frames, i.e.

$$\tilde{MSE} = \frac{1}{N_f \cdot N_p} \sum_{j=1}^{N_f} \sum_{i=1}^{N_p} \sqrt{(x_i - x_i^*)^2 + (y_i - y_i^*)^2}, \tag{20}$$

where N_f denotes the total number of the tracking frames.

As shown in Table 3, the values of tracking performance $\tilde{MSE}$ obtained by the proposed algorithm are smaller than the other four existing methods. This means that the tracked contours using the proposed method can well match the actual lip boundaries. Meanwhile, the utilization of a 16-point deformable model with geometric constraint to model the lip shape is physically meaningful. Comparatively speaking, the method proposed by Tian et al. [14] often suffered from the complex components in oral cavity and failed to well match the non-symmetrical lip shapes. Barnard et al. [11] just employed a combination of two semi-elliptical shapes to model the lip shape, which cannot fit the actual lip boundary quite well especially in very irregular lip shapes. Wang et al. [20] had shown the better extraction and

Table 1
The comparison of lip contour matching performance.

Data set	Counter matching performance (MSE)					
	Kass et al. [9]	Liew et al. [13]	Leung et al. [21]	Werda et al. [39]	Wakasugi et al. [16]	**Our method**
Fig. 9	4.0857	2.6972	3.7735	2.7522	2.9051	**1.7327**
Fig. 10	4.2762	2.7904	4.0927	2.8112	3.1086	**1.8768**

Table 2
The comparison of lip contour extraction performance.

Data set	Extraction performance ($(n_c/n_t) \times \%$)					
	Kass et al. [9] (%)	Liew et al. [13] (%)	Leung et al. [21] (%)	Werda et al. [39] (%)	Wakasugi et al. [16] (%)	**Our method** (%)
CVL [38]	79.5	92.5	89	91	90.5	**97**
Our lab	75.6	88.4	81.6	87.6	85.8	**95.8**
Average	76.7	89.57	83.71	88.57	87.14	**96.1**

Fig. 11. The experimental results of lip contour extraction from CVL face database using the proposed algorithm.

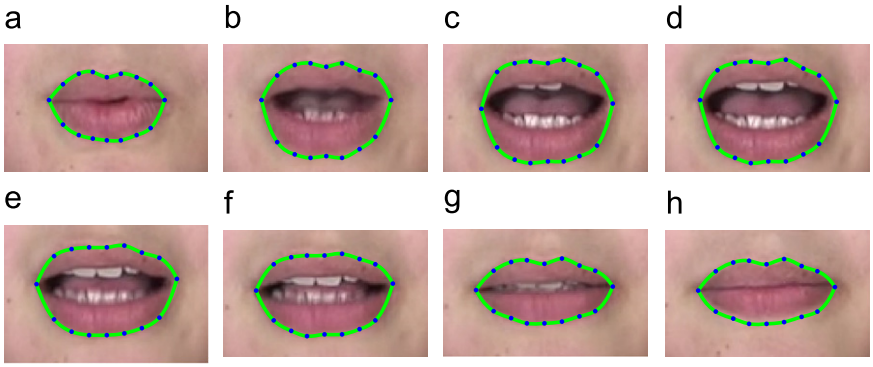

Fig. 12. Lip tracking results in speaking the English digital "1": (a) Frame 3, (b) Frame 7, (c) Frame 12, (d) Frame 15, (e) Frame 18, (f) Frame 21, (g) Frame 26, and (h) Frame 29.

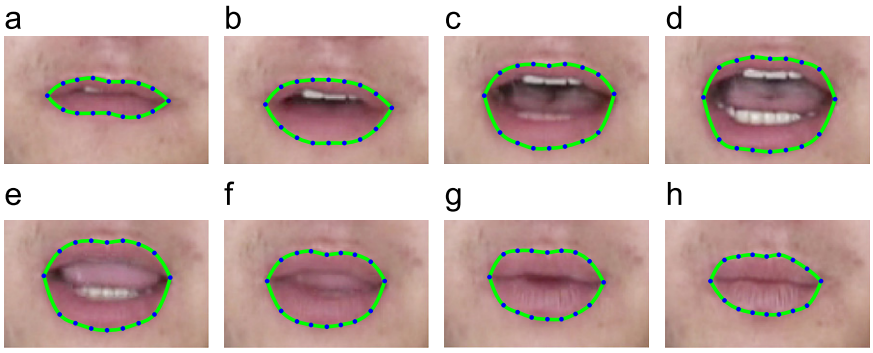

Fig. 13. Lip tracking results in speaking the English digital "5". (a) Frame 3, (b) Frame 8, (c) Frame 13, (d) Frame 16, (e) Frame 19, (f) Frame 22, (g) Frame 27, and (h) Frame 29.

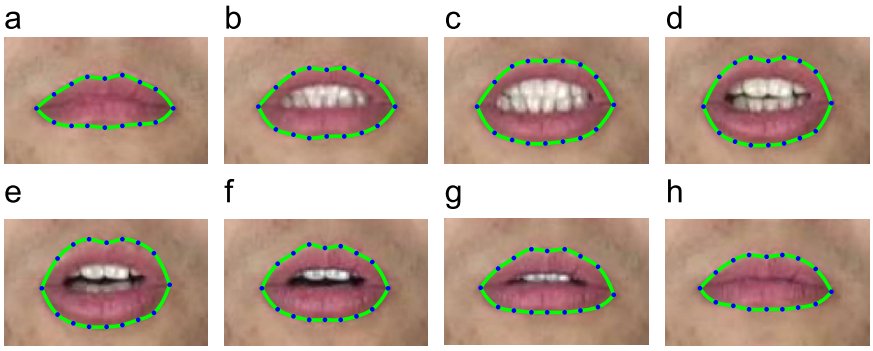

Fig. 14. Lip tracking results in speaking the Chinese digital "9". (a) Frame 1, (b) Frame 6, (c) Frame 11, (d) Frame 15, (e) Frame 20, (f) Frame 21, (g) Frame 26, and (h) Frame 30.

Table 3
The comparison of lip tracking performance.

Method	Tracking performance [$\widetilde{MSE}$]				
	Tian et al. [14]	Barnard et al. [11]	Wang et al. [20]	Eveno et al. [24]	**Our method**
Group 1	3.874	3.272	2.232	2.463	1.892
Group 2	4.109	3.316	2.153	2.535	1.954
Group 3	4.235	3.513	2.359	2.775	2.171

tracking results when the probability maps are well obtained. However, the probability maps may not be accurately generated due to the clutter around the lips and the visibility of the tongue or black hole such that the performance is deteriorated. Eveno et al. [24] introduced a parametric model composed of four cubic curves associated with some adjustment process to fit the lip boundary. Such a method is effective in most cases, but which may fail to well match the lip boundary in case some fractions of lip edges are difficult to differentiate.

Further, we randomly selected 10 groups of lip frames to evaluate the average computation time of the proposed lip tracking algorithm in comparison with the above four approaches. For lip contour extraction phase, the approach proposed by Tian et al. [14] needs the assistance of manually locating the mouth region while the approach presented by Barnard et al. [11] also requires to manually select the lip region of interest. We followed the instructions of the method [24] and manually selected a single point located above the upper lip region to perform the lip contour extraction. As listed in Table 4, the computation time obtained by the proposed method is 0.695 s, which is smaller than the result obtained by the method [20] and is comparable to the result obtained by the approach [24]. For the lip tracking phase, the average computation time of tracking one frame by the proposed approach is 0.103 s, which is less than the

Table 4
The comparison of the computation time.

Method	Computation time (s)				
	Tian et al. [14]	Barnard et al. [11]	Wang et al. [20]	Eveno et al. [24]	**Our method**
First frame	Manual	Manual	1.232	0.623	0.695
Tracking frame	0.097	0.089	0.133	0.171	0.103

values generated by the methods in [20,24], because the method [20] needs to compute the probability map every frame while the method in [24] requires a bit more pre-and-post-processing techniques and adjustment process to fit the lip boundary. Although the average computation time of tracking one lip frame is a bit higher than the results obtained by Tian et al. [14] and Barnard et al. [11], the proposed method can well match the actual lip boundary with a better tracking performance as shown in Table 3. It is effective to utilize the extracted contour of the previous lip frame to track the current one. Such a way can significantly reduce a large amount of computation time when there exists a long lip sequence. Meanwhile, the computation time is closely related to the region scale. The proposed lip tracking algorithm would generally take less time when the scale of the test lip images is smaller.

5. Conclusion

In this paper, we have introduced the framework of LCACM, through which the color objects incorporating complex appearances or intensity inhomogeneities can be effectively segmented. Accordingly, we have presented a two-phase local region-based approach to lip tracking. Being adaptive to the lip movements, the proposed approach features: (1) less pre-processing steps such as teeth removal, training data capture and training process, and (2) performance robustness to the appearance of teeth, tongue and black hole. The experimental results have shown its promising results in comparison with the existing methods. Nevertheless, in case the mustache effect around the lip region becomes noticeable, the proposed approach, as well as the other existing ones, needs an image pre-processing to detect the region of mustache and mask it out. We leave this study elsewhere as a future work.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments and insightful suggestions, and thank the volunteers for video database capturing. The work described in this paper was supported by the Research Grant Council of Hong Kong SAR with Project No.: HKBU 210309, the Faculty Research Grant of Hong Kong Baptist University with the Project Code: FRG2/09-10/098, FRG2/10-11/056 & FRG2/11-12/067, the NSFC under grants 61172136 and 60973154, the Program of International Science and Technology Cooperation (Grant No. 2011DFA12180), Ph.D. Programs Foundation of Ministry of Education, China (Grant No. 20110142110060), and the Hubei Provincial Natural Science Foundation under Grant 2010CDA006, China.

Appendix A. Illumination equalization

Illumination direction (1): Let $l(i)$ and $r(i)$ denote the mean intensity of left and right borders within a local region $(2p+1)\times(2q+1)$ at the i-th row:

$$l(i)=\frac{1}{(2p+1)(2q+1)}\sum_{k=-p}^{p}\sum_{l=-q}^{q}L(i+k,l),$$

$$r(i)=\frac{1}{(2p+1)(2q+1)}\sum_{k=-p}^{p}\sum_{l=n-2q}^{n}L(i+k,l).$$

We can obtain the new illuminance value:

$$\begin{aligned}\hat{L}(i,j)&=L(i,j)+\frac{l(i)-r(i)}{n-1}(j-1)+\frac{r(i)-l(i)}{2}\\&=L(i,j)+\left(\frac{1}{2}-\frac{j-1}{n-1}\right)\cdot(r(i)-l(i))\\&=L(i,j)+\frac{(n-2j+1)\cdot(r(i)-l(i))}{2(n-1)}.\end{aligned}$$

Illumination direction (2): Let $t(j)$ and $b(j)$ denote the mean intensity of top and bottom borders within a local region $(2p+1)\times(2q+1)$ at the j-th column:

$$t(j)=\frac{1}{(2p+1)(2q+1)}\sum_{k=-p}^{p}\sum_{l=-q}^{q}L(k,j+l),$$

$$b(j)=\frac{1}{(2p+1)(2q+1)}\sum_{k=m-2p}^{m}\sum_{l=-q}^{q}L(k,j+l).$$

We can obtain the new illuminance value:

$$\begin{aligned}\hat{L}(i,j)&=L(i,j)+\frac{t(j)-b(j)}{m-1}(i-1)+\frac{b(j)-t(j)}{2}\\&=L(i,j)+\left(\frac{1}{2}-\frac{i-1}{m-1}\right)\cdot(b(j)-t(j))\\&=L(i,j)+\frac{(m-2i+1)\cdot(b(j)-t(j))}{2(m-1)}.\end{aligned}$$

Appendix B. Cubic spline interpolation

Cubic Spline Interpolation (CSI) is an effective method that offers true continuity between the segments. For a data set x_i of $n+1$ points, it can construct a cubic spline with n piecewise cubic polynomials between the data points. If

$$S(x)=\begin{cases}S_0(x), & x\in[x_0,x_1],\\ S_1(x), & x\in[x_1,x_2],\\ \vdots & \vdots\\ S_{n-1}(x), & x\in[x_{n-1},x_n]\end{cases}$$

represents the spline function interpolating the function f, it requires that:

- The cubic polynomial matches the interpolating property:
  $S(x_i)=f(x_i).$
- The splines need to join up:
  $S_{i-1}(x_i)=S_i(x_i),\quad i=1,\ldots,n-1.$
- In order to make the interpolation as smooth as possible, the first and second derivatives should be continuous, i.e.,
  $S'_{i-1}(x_i)=S'_i(x_i),\quad S''_{i-1}(x_i)=S''_i(x_i),\quad i=1,\ldots,n-1.$

For the practical applications, there are two primarily standard choices as follows:

- Natural cubic spline:

$$S''(x_0)=0, \quad S''(x_n)=0.$$

- Complete cubic spline:

$$S''(x_0)=f'(x_0), \quad S''(x_n)=f'(x_n).$$

References

[1] X. Lei, C. Xiu-Li, F. Zhong-Hua, Z. Rong-Chun, J. Dong-Mei, A robust hierarchical lip tracking approach for lipreading and audio visual speech recognition, in: Proceedings of the International Conference on Machine Learning and Cybernetics, vol. 6, 2004, pp. 3620–3624.

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

[3] P. Delmas, M. Lievin, From face features analysis to automatic lip reading, in: Proceedings of the International Conference on Control, Automation, Robotics and Vision, vol. 3, 2002, pp. 1421–1425.

[4] Y.I. Tian, T. Kanade, J.F. Cohn, Recognizing action units for facial expression analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2) (2001) 97–115.

[5] D. Freedman, M.S. Brandstein, Contour tracking in clutter: a subset approach, International Journal of Computer Vision 38 (2) (2000) 173–186.

[6] H. Yang, L. Shao, F. Zheng, L. Wang, Z. Song, Recent advances and trends in visual tracking: a review, Neurocomputing 74 (18) (2011) 3823–3831.

[7] X. Zhang, R.M. Mersereau, Lip feature extraction towards an automatic speechreading system, in: Proceedings of the IEEE International Conference on Image Processing, vol. 3, 2000, pp. 226–229.

[8] N. Eveno, A. Caplier, P. Y. Coulon, Key points based segmentation of lips, in: Proceedings of the IEEE International Conference on Multimedia and Expo, vol. 2, 2002, pp. 125–128.

[9] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, International Journal of Computer Vision 1 (4) (1988) 321–331.

[10] P. Delmas, P. Coulon, V. Fristot, Automatic snakes for robust lip boundaries extraction, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 6, 1999, pp. 3069–3072.

[11] B. Mark, H. Eun Jung, O. Robyn, Lip tracking using pattern matching snakes, in: Proceedings of the Asian Conference on Computer Vision, 2002, pp. 23–25.

[12] A.L. Yuille, P.W. Hallinan, D.S. Cohen, Feature extraction from faces using deformable templates, International Journal of Computer Vision 8 (2) (1992) 99–111.

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

[14] Y. Tian, T. Kanade, J. Cohn, Robust lip tracking by combining shape, color and motion, in: Proceedings of the Asian Conference on Computer Vision, 2000, pp. 1040–1045.

[15] G.I. Chiou, J.N. Hwang, Lipreading from color video, IEEE Transactions on Image Processing 6 (8) (1997) 1192–1195.

[16] T. Wakasugi, M. Nishiura, K. Fukui, Robust lip contour extraction using separability of multi-dimensional distributions, in: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp. 415–420.

[17] J. Luettin, N.A. Thacker, S.W. Beet, Visual speech recognition using active shape models and hidden Markov models, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, 1996, pp. 817–820.

[18] 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, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, 2001, pp. 1485–1488.

[19] Q.D. Nguyen, M. Milgram, T.H. Nguyen, Multi features models for robust lip tracking, in: Proceedings of the International Conference on Control, Automation, Robotics and Vision, 2008, pp. 1333–1337.

[20] S. Wang, W. Lau, S. Leung, Automatic lip contour extraction from color images, Pattern Recognition 37 (12) (2004) 2375–2387.

[21] S.H. Leung, S.L. Wang, W.H. Lau, Lip image segmentation using fuzzy clustering incorporating an elliptic shape function, IEEE Transactions on Image Processing 13 (1) (2004) 51–62.

[22] A.W.C. Liew, S.H. Leung, W.H. Lau, Segmentation of color lip images by spatial fuzzy clustering, IEEE Transactions on Fuzzy Systems 11 (4) (2003) 542–549.

[23] R. Rohani, S. Alizadeh, F. Sobhanmanesh, R. Boostani, Lip segmentation in color images, in: Proceedings of the International Conference on Innovations in Information Technology, 2008, pp. 747–750.

[24] N. Eveno, A. Caplier, P.Y. Coulon, Accurate and quasi-automatic lip tracking, IEEE Transactions on Circuits and Systems for Video Technology 14 (5) (2004) 706–715.

[25] P. Narayanan, S. Nayar, H.-Y. Shum, Y.-D. Jian, W.-Y. Chang, C.-S. Chen, Attractor-guided particle filtering for lip contour tracking, in: Proceedings of the Asian Conference on Computer Vision, vol. 3851, 2006, pp. 653–663.

[26] E. Ong, R. Bowden, Robust lip-tracking using rigid flocks of selected linear predictors, in: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2008.

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

[28] L. Yang, P. Meer, D.J. Foran, Unsupervised segmentation based on robust estimation and color active contour models, IEEE Transactions on Information Technology in Biomedicine 9 (3) (2005) 475–486.

[29] D. Cremers, M. Rousson, R. Deriche, A review of statistical approaches to level set segmentation: integrating color, motion and shape, International Journal of Computer Vision 72 (2) (2007) 195–215.

[30] S. Lankton, Localized statistical models in computer vision. Georgia Institute of Technology.

[31] O. Michailovich, Y. Rathi, A. Tannenbaum, Image segmentation using active contours driven by the Bhattacharya gradient flow, IEEE Transactions on Image Processing 16 (11) (2007) 2787–2801.

[32] A. Yezzi, A. Tsai, A. Willsky, A fully global approach to image segmentation via coupled curve evolution equations, Journal of Visual Communication and Image Representation 13 (1–2) (2002) 195–216.

[33] S.C. Zhu, A. Yuille, Region competition: unifying snakes, and Bayes/MDL for multiband image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (9) (1996) 884–900.

[34] M. Li, Y.M. Cheung, Automatic lip localization under face illumination with shadow consideration, Signal Processing 89 (12) (2009) 2425–2434.

[35] X. Liu, Y.M. Cheung, M. Li, H. Liu, A lip contour extraction method using localized active contour model with automatic parameter selection, in: Proceedings of the IEEE International Conference on Pattern Recognition, 2010, pp. 4332–4335.

[36] H. Hsieh, H. Andrews, Cubic splines for image interpolation and digital filtering, IEEE Transactions on Speech, and Signal Processing 26 (6) (1978) 508–517.

[37] H.E. Cetingul, Y. Yemez, E. Engin, A.M. Tekalp, Discriminative analysis of lip motion features for speaker identification and speech-reading, IEEE Transactions on Image Processing 15 (10) (2006) 2879–2891.

[38] F. Solina, P. Peer, B. Batagelj, S. Juvan, J. Kovac, Color-based face detection in the 15 seconds of fame art installation, in: Proceedings of the Conference on Computer Vision/Computer Graphics Collaboration for Model-Based Imaging, Rendering, Image Analysis and Graphical Special Effects, Mirage, 2003, pp. 38–47.

[39] S. Werda, W. Mahdi, A. Ben Hamadou, Colour and geometric based model for lip localisation: application for lip-reading system, in: Proceedings of the International Conference on Image Analysis and Processing, 2007, pp. 9–14.

Yiu-ming Cheung (SM'06) received the Ph.D. degree from the Department of Computer Science and Engineering, The Chinese University of Hong Kong in 2000. He joined the Department of Computer Science at Hong Kong Baptist University in 2001, and then became an Associate Professor in 2005. His current research interests are in the fields of machine learning and information security, particularly the topics on clustering analysis, blind source separation, neural networks, nonlinear optimization, watermarking and lip-reading. He is the founding Chairman of IEEE (Hong Kong) Computational Intelligence Chapter. Currently, he is also the Associate Editor of Knowledge and Information Systems, as well as the guest co-editor and editorial board member of the several international journals.

Xin Liu received the B.S. and M.S. degrees from Hubei University, Wuhan, China, in 2004 and 2009, respectively. He is currently the PhD Student at the Department of Computer Science in Hong Kong Baptist University, Hong Kong. His research interests include image processing, computer vision, and pattern recognition.

Xinge You (M'08–SM'10) received the B.S. and M.S. degrees in mathematics from the Hubei University, Wuhan, China, in 1990, and the Ph.D. degree in computer science from the Hong Kong Baptist University, Hong Kong, in 2000 and 2004, respectively. He is presently a Professor in the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, China. His current research interests include wavelets and its application, signal and image processing, pattern recognition, machine learning, and computer vision.