A Real-time Lip Localization and Tacking for Lip Reading

Yao WenJuan
School of Electronic and Information Engineering
South China University of Technology
Guangzhou, China
e-mail: wenjuan.yao@gmail.com

Liang YaLing
School of Electronic and Information Engineering
South China University of Technology
Guangzhou, China
e-mail: ylliang@scut.edu.cn

Du MingHui
School of Electronic and Information Engineering
South China University of Technology
Guangzhou, China
e-mail: ecmhdu@163.com

***Abstract*—Most automatic speech recognition systems have concentrated exclusively on the acoustic speech signal, and therefore they are susceptible to acoustic noise. The benefits from visual speech cues have motivated significant interest in automatic lip-reading, which aims at improving automatic speech recognition by exploiting informative visual features of a speaker's mouth region, which means speaker lip motion stands out as the most linguistically visual feature. In this paper, we present a new improved robust lip location and tracking approach, aims at improving the lip-reading accuracy. Lip regions of interest are detected by a new method, combining with Intel Open source (OpenCV). In this new method, we analyze the distribution relationship between faces, eyes and mouth, and then the mouth region can be easily located. It can be proved as an effective method for lip tracking. In the subsequent step, color space is transferred to Lab from RGB color space, and a component of Lab color space is used for extracting lip segmentation and tracking lip region more accurately and efficiently from video sequences of a speaker's talking face in different lighting conditions, and with different lip shapes and head poses. Extensive experiments show that our proposed method can achieve superior performance to other similar lip tracking approaches, and then can be effectively integrated in lip-reading or visual speech recognition systems.**

Keywords- lip tracking; OpenCV; a component

I. INTRODUCTION

It is well known that visual speech information through lip-reading is very useful for human speech recognition[l]. Hearing-impaired people use lip-reading as a primary source of information for speech communication. Even for those with normal hearing, seeing the speaker's lip motion is also proven to significantly improve intelligibility, especially under adverse acoustic conditions.

A complete lip reading system can consist of lip location, lip tracking, lip movement extraction, and lip reading.

Lip location, from a face image, is the primary and critical part that its accuracy can affect the performance of following lip tracking, lip movement extraction and lip reading. In previous system, people use camera to capture only lip region, or lip region is marked manually, but this is not a practical and automatic method and will add limitation and difficult to lip-reading application. Later, lip location is usually based on face detection.

There have been some methods to locate lip region after face detection. One is to roughly locate the mouth region (region of interest) according to distribution feature of the mouth in face region. This method usually considers half size of the face width as the width of the mouth region, and one third of the face height as the height of the mouth region. The advantage of this approach is that it is a simple and fast mouth location method, but it will loose effectiveness for those images with different head poses and lip shapes.

Another approach is called the gray projection. In this approach, image is projected to horizontal and vertical axis, mouth region is defined by valleys of horizontal and vertical curve. The mouth region can be easily defined in this method, however, the accuracy can be easily affected by bad lighting conditions, low discrimination in lip and skin color, and beard around mouth.

Lip tracking is precise lip segmentation based on lip location. Historically, there have been two main approaches in lip tracking from image sequences. The first method is called the color-based approach. In this approach, different color space is proposed, for example: RGB [1], HSV. Red Exclusion is an effective approach to extract the lip, in this approach, mouth is tracked by using G and B color component, but it is only useful for white person. Another color-based approach [2] is proposed to extract the lip region by analyzing the color distribution of lip and skin. But the drawback is it is only useful for a specified skin color, yellow skin or white skin and does not take the beard and tooth into consideration.

The second type is known as the model-based method. A lip model is described by a set of parameters (e.g. height and width of lips). These parameters are calculated from a cost function minimization process of fitting the model onto a captured image of the lip. The active contour model, the deformable geometry model, and the active shape model are examples of such methods widely used in lip tracking and feature extraction. The advantage of this approach is that lip shapes can be easily described by low order dimensions and it is invariant under rotation, translation, or scaling. However, this method requires an accurate model initialization to ensure that the model updating process converges.

II. REAL-TIME LIP TRACKING APPROACH

Our real-time lip tracking approach includes several steps, first step is face detection and lip location, where an OpenCV technique is used to detect face and eye region, and then locate mouth region with a rectangle. Then technique of a component of Lab color space is followed to precisely extract the lip. Finally, lip tracking is executed.

A. Lip region location using OpenCV

Lip detection or segmentation is a very difficult problem due to the low grayscale variation around the mouth. Chromatic or color pixel based features, especially red domination of lips, have been adopted by most researchers to segment lips from the primarily skin background. However, color representation can he influenced by background lights, and red blobs in the speaker's clothing can cause segmentation failures. In this paper, we present a fast lip region detection method using OpenCV technique.

1) Theoretical basis: OpenCV is an open resource for image processing and computer vision developed by Intel. It can be used by us for Secondary development.

We analyzed two databases, aims to present the distribution relationship between face, eyes, and mouth. One is the cr face database with different lip shapes, different lighting conditions, and different face poses of one person (shown in table 1). Another is a dual-mode video database. We analyzed 3 different person, each person has 100 image sequences, with different lighting conditions or with beard (shown in table 2).

Table 1. 213 faces, different in lip shapes, varying lighting conditions

213 faces, different in lip shapes, varying lighting conditions					
High-level lighting conditions, frontal faces 49		Low-level Lighting conditions, frontal faces 49		Low-level Lighting conditions, side faces 115	
Closed lips	Opened lips	Closed lips	Opened lips	Closed lips	Opened lips
8	41	24	25	24	91

Table 2. 300 faces of 3 person, with different lighting conditions or with beard

300 faces of 3 person, with bad lighting conditions, different lip shapes, or with beard					
Male faces with bad lighting condition 100		Male faces with beard 100		Female faces with good lighting condition 115	
Closed lips	Opened lips	Closed lips	Opened lips	Closed lips	Opened lips
62	38	67	33	74	26

After the experimental analysis, it is found that:
(1) The mouth is always in the scope of the distance of two eyes.
(2) Half size of the distance from eye to face bottom can totally contain the mouth region.
(3) The mouth is horizontally parallel with the eye, and rotate with the rotation of the eye.
This is illustrated in following figure 1:

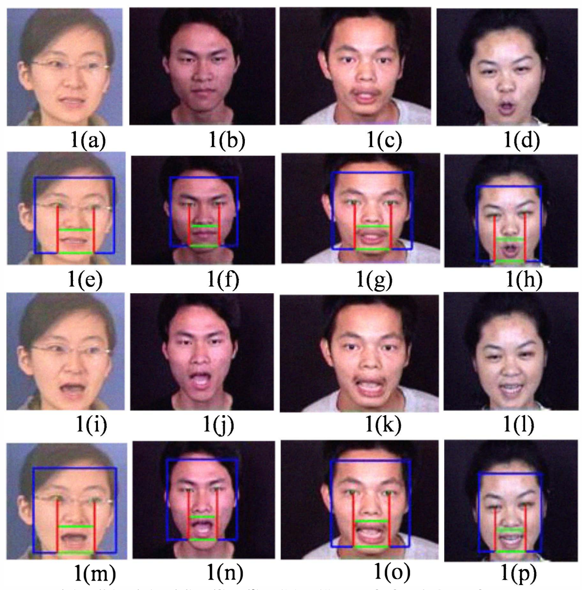

Figure 1(a), (b), (c), (d), (i), (j), (k), (l): Original face image
Figure 1(e), (f), (g), (h), (m), (n), (o), (p):
Distribution relationship between face, eyes, and mouth

2) Lip region location: Based on above analysis, in this paper, we propose a fast lip region detection method using OpenCV. In this algorithm, we use the combination of Adaboost and haar feature to detect the face and eyes with the classifier of OpenCV. Finally, we can locate the mouth region using its distribution relationship with face and eyes. The OpenCV algorithm is as follows:

(1) Detect the face region of the image (figure 2(b)).

(2) Detect the eye region and locate the eye ball (figure 2(c))

(3) Locate the mouth region by following rules:

a. The distance of the two eyes can be considered as the width of the mouth region.

b. And the height of the mouth region equals to half size of the distance from eye to face bottom (figure 2(d)).

c. Rotate the lip to horizontal parallel if there is rotation of eyes (figure 2(e)).

d. Normalize the mouth region size to an appropriate size.

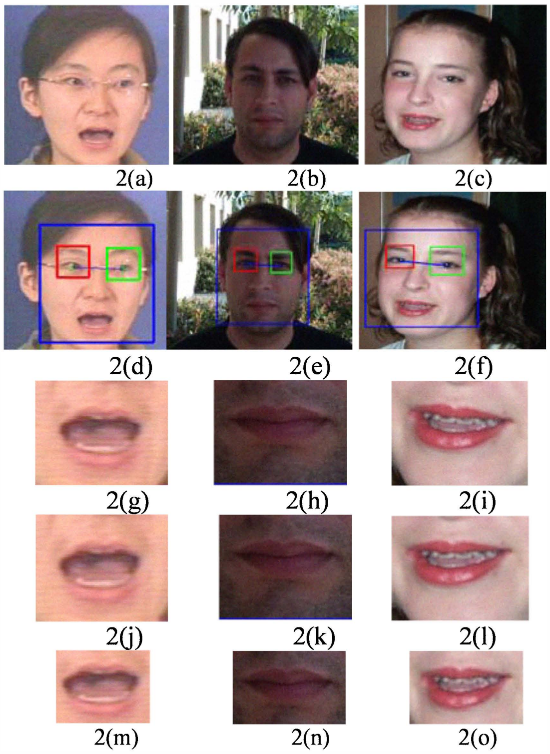

Figure 2. Procedure of lip region location

Figure 2(a), (b), (c): original face image

Figure2 (d), (e), (f): face and eye detection

Figure2 (g), (h), (i): lip region

Figure2 (j), (k), (l): lip region after rotation

Figure2 (m), (n), (o): normalized lip region

The lip location method described above is a robust method that does not affected by face poses, lip shapes, or even lighting conditions. It can avoid the drawback of normalization distortion from some method which achieves the lip tracking by looking for mouth corner [3].

B. *Lip tracking using a component of Lab color space*

After the mouth region location, a precise lip tracking should be followed to make preparation for following lip reading. It is a challenging problem to maximize the discrimination between lip and non-lip colors due to their similar color.

1) Color Space Transformation: At present, lip tracking method mainly includes the two branches: color-based method and Contour-based method. Contour-based method does not perform good result for those lips without clear contour.

Regarding color-based method, many methods have been proposed for segmenting the lip region that based on image intensity or color. Among color based lip segmentation methods are Red Exclusion [1], Mouth-Map [4], Pseudo hue [5].

Theoretically, pseudo hue method gives better color contrast, but we found that it is only useful in performing coarse segmentation which is not adequate for our purpose. And the RE method [1] has been approved as a simple and effective method, but the drawback is it has good performance only for white people.

In the paper [6], a new method using a component of Lab color space is proposed. This method is not only useful for different skin color, is also useful for mouth region with tooth or beard.

In this paper, we will present the a component method for lip tracking.

At first, the color space of mouth region image must be transferred to Lab color space from RGB.

The color space transformation from RGB to Lab is as follows (formula (1), (2)):

$$\begin{matrix} X \\ Y \\ Z \end{matrix} = \begin{pmatrix} 0.4125 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{pmatrix} * \begin{matrix} R \\ G \\ B \end{matrix} \quad (1)$$

$$\begin{aligned} L &= 116 * f(Y1) - 16 \\ a &= 500 * (f(X1) - f(Y1)) \\ b &= 200 * (f(Y1) - f(Z1)) \end{aligned} \quad (2)$$

2) Selection of the threshold: After the Lab color space transformation, we have the maximized contrast in a component between lip color and non-lip color. A threshold selection is followed to extract the complete lip segment.

We tried to select the mean value of a component as the threshold, but it may not be able to achieve the best performance regarding to the unknown proportion of the tooth and beard.

In this paper, we adopt the value of mean a component plus by Variance of a component as the threshold.

3) Lip tracking: According to the research of the color component in separating lip and non-lip [6], a component of Lab color space is approved to have the best separation ability. The lip and non-lip can be discriminated by setting threshold. The algorithm is as follows:

(1) Transfer color space from RGB to Lab

(2) Calculate the mean value of a component: *ameans*

(3) Calculate the variance of a component: *astd*

(4) Threshold is calculated as: θ = *ameans* + *astd*

(5) For all pixels of the mouth region, judge their pixel value with formula (3):

$$a(i, j) > \tilde{\theta} \qquad (3)$$

Where $a(i, j)$ represent the a component value of (i, j). It can be considered as lip pixel if meet above formula, or non-lip pixel.

III. EXPERIMENTS

In order to test the performance of our proposed robust procedure, we use 2000 normalized lip images over 69 people with different lip shapes, lighting conditions, and different head poses.

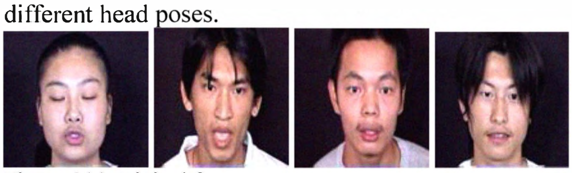

Figure 3(a) original faces

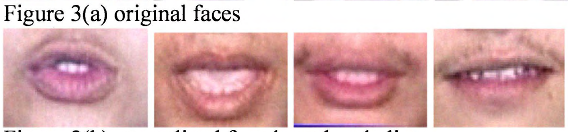

Figure 3(b) normalized female and male lips

Figure 3(c) lip extraction by a component

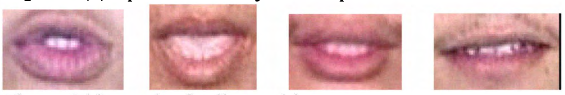

Figure 3(d) results for lip tracking

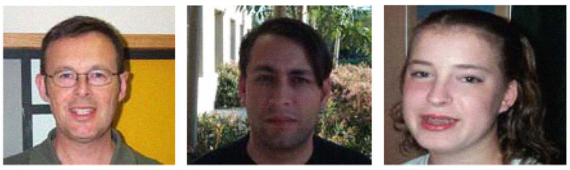

Figure 3(e) original faces

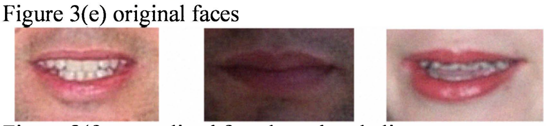

Figure 3(f) normalized female and male lips

Figure 3(g) lip extraction by a component

Figure 3(h) results for lip tracking

Figure 3. Lip tracking results of female and also male lips

Overall, 98% of the lips can be accurately tracked, Fig.3 shows some examples of such images. As is approved, a component of Lab have better separation ability comparing to other color space component. We also show that our proposed lip location method has successfully improved our lip tracking performance under different lightning conditions, different lip shapes, and different head poses.

IV. CONCLUSION

In this paper, we describe a new robust approach to improve lip localization and tracking. The first part of our proposed algorithm is lip location based on OpenCV. From experimental results, our proposed method can successfully detect the lip region. Results from the lip location accuracy allowed more accurate lip region segmentation. In the second part, a new method called a component of Lab color space is proposed to accurately extract lip shape and track lip region. Overall, our implemented approach has shown high reliability and is able to perform robustly under various conditions.

REFERENCES

[1] Lewis.T.W, Powers.D.M., “Lip Feature Extraction Using Red Exclusion”, Proc. Selected papers from Pan-Sydney Workshop on Visual Information Processing, pp.61-67, 2000.

[2] Zhang Zhi-wen, Shen Hai-bin, “Lip detecting algorithm based on chroma distribution diversity”, zhejiang university journal, 2008,42(8):pp1355-1359

[3] Xie Lei, Cai Xiu-Li, Fu Zhong-Hua, “A robust hierarchical LIP tracking approach for lipreading and audio visual speech recognition”, Proc. Int. Conf. Mach.Learning Cybernetics, 2004, 6:3620-3624.

[4] R.L.Hsu, M .Abdel, A.K. Jain, “Face Detection in Color Images”, IEEE Trans. on Pattern Analysis & Machine Intelligence, 2002.

[5] A.Hulbert and T.Poggio, “Synthesizing a Color Algorithm From Examples”, Science, vol. 239, pp. 482-485, 1998.

[6] Liang Ya-ling, Du Ming-hui, “Lip Extraction Based on a *component of Lab Color Space*”.