Automatic Segmentation of Color Lip Images Based on Morphological Filter

Meng Li
Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China

Abstract

This paper addresses the problem of lip segmentation in color space that is a crucial issue to the success of a lip-reading system. We present a new segmentation approach to lip contour extraction by taking account of the color difference between the lip and skin in the color spaces. Firstly, we obtain a lip segment sample via the distinction between the lip and skin in 1976 CIELAB color space. Secondly, we establish a probability model in HSV color space and make use of the segment sample so that the membership of lip and non-lip region is calculated. Thirdly, we employ a morphological filter to obtain the lip counter candidate based on the two memberships. Finally, we extract the lip contour via convex hull algorithm with the prior knowledge of the mouth shape. Experiments show the efficacy of the proposed approach in comparison with the existing lip segmentation methods.

1. Introduction

In the past decade, the lip segmentation has received considerable attention from the community because of its wide applications in audio-visual speech recognition, biometric person identification, lip synchronization, human expression recognition, and so forth [1, 2, 3, 4]. In general, lip segmentation is a non-trivial task because the color difference between the lip and the skin regions is not so noticeable sometimes. In particular, it becomes more challenging when the illuminations in the environment are complex.

In the literature, a few image segmentation techniques have been proposed. One class of methods is based on the clustering with color features [1, 5, 6] provided that the number of clusters (e.g. the clusters of skin and lip) is known in advance. Unfortunately, the hair, moustache and the visibility of teeth and tongue in the mouth opening generally require that the number of clusters is selected adaptively. Consequently, such a method is unable to operate fully automatically [7]. Another class of widely-used methods is model-based ones, such as Active Shape Model, Active Appearance Model, Active Contour Model (Snake), and so forth [1, 8, 9, 10, 11]. Specifically, they build a deformable model for lip by learning the patterns of variability from a training set of correctly annotated images. The shape of model can be adjusted by a parameter set so as to match and locate the lip in test images. Empirical studies have shown the success in their application domain, but they need to label some landmarks manually for training. Recently, some automatic segmentation approaches to lip images have been developed. For example, [12] and [13] utilize a color transformation or color filter to enlarge the difference between the lip and skin. Paper [7] utilizes the multi-scale wavelet to detect the edge of lip.

In this paper, we will present a new method for the automatic segmentation of lip images provided that the lower part of a face (i.e. the part between nostril and chin) has been available. The proposed method employs the color transformations so as to enlarge the difference between the lips and the skin like the existing methods, but the new method extracts a segment of lip only rather than the whole lip contour. Specifically, we firstly obtain a lip segment sample via the distinction between the lip and skin in 1976 CIELAB color space. Secondly, we establish a probability model in HSV color space and make use of the segment sample so that the membership of lip and non-lip region is calculated. Thirdly, we employ a morphological filter to obtain the lip counter candidate based on the two memberships. Finally, we extract the lip contour via convex hull algorithm with the prior knowledge of the mouth shape. Since our method is based upon the color diversity between skin and lip region rather than "absolute" color [5], its performance is stable and accurate on the images with different color temperatures and the testers with different skin or lip colors. Experiments have shown the efficacy of the proposed approach in comparison with the existing lip segmentation methods.

The remainder of this paper is organized as follows. We describe the calculation of lip membership in Section 2, and show the lip contour extraction in Section 3. In Section 4, we will conduct the experiment to empirically compare the proposed method with the existing ones. Finally, we draw a conclusion in Section 5.

2 Lip Membership Based on Color Space Transformation

It is desirable to work in a color space (a sample of original image is illustrated in Figure 1), in which the lip color (i.e. relative red) out of others can be highlighted. Since the value of a^* channel in 1976 CIELAB color space can determine the color component between red/magenta and green, i.e. the small values indicate green while the large values indicate magenta, we therefore convert the source image into 1976 CIELAB color space and normalize the a^* component to the range of $[0, 255]$, denoted as I_{a^*}. Furthermore, we utilize Eq.(1) as proposed in [12] to convert the source image to the range of $[0, 255]$ with equalization, denoted as $I_{G/R}$:

$$I_{G/R} = \begin{cases} 256 \times \frac{G}{R} \; R > G \\ \\ 255 \qquad otherwise. \end{cases} \tag{1}$$

Let $I_{sub} = I_{a^*} - I_{G/R}$.[1] Subsequently, we can establish a Gaussian model for I_{sub} based on the gray-level value for each non-zero pixel with the mean $\hat{\mu}_{sub}$ and the standard deviation $\hat{\sigma}_{su}$. The candidate of lip segments can be obtained by

$$I_{candidate} = \begin{cases} 0 \qquad I_{sub} \leq \hat{\mu}_{sub} - 2\hat{\sigma}_{sub} \\ \\ I_{G/R} \; otherwise. \end{cases} \tag{2}$$

In Eq. (2), it is found that most non-black points are in the lip region. Hence, to eliminate those non-black parts outside the lip region, we utilize Eq.(3) and Eq.(4) to calculate the gravity center, denoted as (g_x, g_y):

$$g_x = \frac{\sum_{x=1}^{col} \sum_{y=1}^{row} x I_{candidate}(x, y)}{\sum_{x=1}^{col} \sum_{y=1}^{row} I_{candidate}(x, y)} \tag{3}$$

$$g_y = \frac{\sum_{x=1}^{col} \sum_{y=1}^{row} y I_{candidate}(x, y)}{\sum_{x=1}^{col} \sum_{y=1}^{row} I_{candidate}(x, y)}, \tag{4}$$

where row and col denote the vertical and horizontal size of the image in pixel, respectively. We extract the nearest non-black part to the gravity center, which corresponds to the lip segment as shown in Figure 2. Please note that it is enough to extract a part of lip rather than the whole lip region because the extracted lip segment is used for sample data so as to establish a probability model.

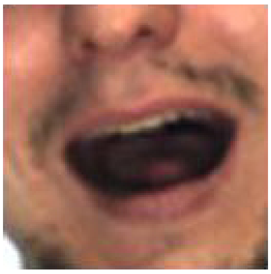
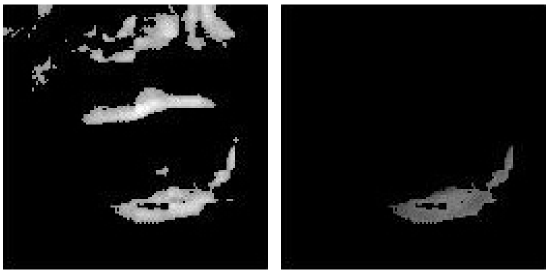


Figure 1. The original lip image, which is the source image of the subsequent figures.



(a) (b)

Figure 2. A sample image of (a) $I_{candidate}$, and (b) the extracted lip segment. In $I_{candidate}$, most non-black parts correspond to lip segments in the source image. The extracted lip segment is a part of the lip.

Subsequently, pixels in source lip image belonging to the lip segment area extracted above are converted into HSV color space. For each pixel, we let:

$$C_1 = H \cdot cos(2\pi \cdot S) \tag{5}$$

$$C_2 = H \cdot sin(2\pi \cdot S). \tag{6}$$

As a result, we obtain a 2-dimensional sample vector denoted as $C_{Seg} = (C_1, C_2)$.

We establish a probability model as follows:

$$M_{lip} = \frac{1}{2\pi\sqrt{\hat{\Sigma}}} \cdot exp(-\frac{(C_{Src} - \hat{\mu})\hat{\Sigma}^{-1}(C_{Src} - \hat{\mu})^T}{2}) \tag{7}$$

[1] In this paper, all equations are employed in positive area. That is, as long as a result is negative, it will be set at 0 automatically.

where C_{Src} is a 2-dimensional value for arbitrary pixel in source lip image. The parameters $\hat{\mu}$ and $\hat{\Sigma}$ can be estimated via the following equations:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} C_{Seg}^{i}}{n} \tag{8}$$

$$\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(C_{Seg}^{i} - \hat{\mu})(C_{Seg}^{i} - \hat{\mu})^{T}. \tag{9}$$

Based on the probability model of Eq. (7), we can calculate the lip membership for each pixel in a source image.

Similarly, based on the black area of $I_{candidate}$, we can also establish a probability model to calculate the non-lip membership denoted as $M_{non-lip}$. Figure 3 gives an example of the two membership maps, in which the high membership corresponds to the light area, and vice versa. Moreover, considering the convenience of visibility, we project the membership from $[0, 1]$ to $[0, 255]$, i.e., the gray scale.

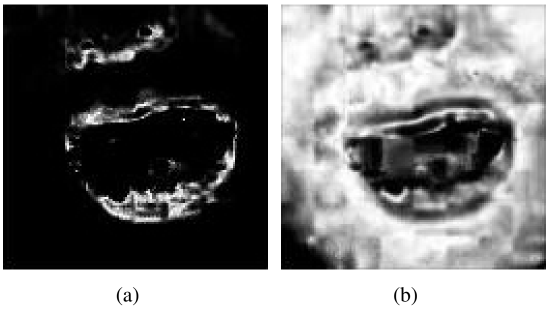


Figure 3. A sample image of (a) lip membership map, (b) non-lip membership map. The high membership corresponds to the light area, and vice versa.

3 Lip Counter Extraction

Obtain a mask by letting

$$Mask = 255 - M_{non-lip} - I_{G/R}. \tag{10}$$

Moreover, the lip membership is considered as marker. Thus, the morphological reconstruction operation proposed in [14] can be employed. Figure 4 (a), (b) and (c) illustrate the marker, mask and result, respectively.

We utilize a gray-level threshold selection method proposed in [15] to transform the reconstruction result into a binary image denoted as B_{RT} (see Figure 4 (d)), and mark the biggest continued foreground block by B_{lip_1}. In the case of mouth closing, B_{lip_1} can represent the whole lip region accurately. However, in most cases of mouth opening, the blocks corresponding to upper and lower lips are usually separate. It is hard to extract the whole lip region via selecting the biggest block. Thus, some refinements are needed.

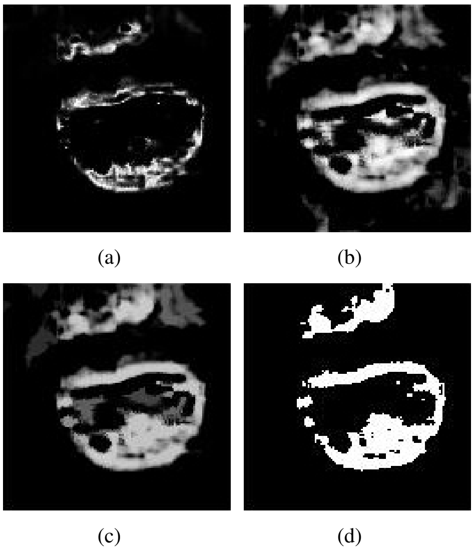


Figure 4. A sample image of (a) marker, (b) mask, (c) the result of morphological reconstruction, and (d) the result of gray-level threshold processing.

Considering the primary reason for discontinuity between upper and lower lip is that the teeth and tongue are eliminated via the above steps. Hence, we utilize the following equation

$$I_{TTM} = I_{G/R} - norm(I_{G/R} - (I_{G/R} - I_{a^*})) \tag{11}$$

to transform the image $I_{G/R}$ to I_{TTM}, where $norm(.)$ denotes the normalization in [0,255]. Subsequently, we can obtain the region covering the teeth, tongue and some parts of oral cavity approximately. Please note that, as we stated in Section 2, the computation is employed in the positive area, i.e. each negative results are set at zero, thus $I_{G/R} - (I_{G/R} - I_{a^*})$ is not equal to I_{a^*}.

We further transform I_{TTM} into a binary image denoted as B_{TTM} by the threshold selection method. Then, the morphological closing is employed to $B_{RT} \cup B_{TTM}$ by performing a 5×5 structuring element operation. We select the biggest foreground block denoted as

B_{lip_2} in the closing operation result. Hence, the binary image $B_{lip_1} \cup B_{lip_2}$ can represent the whole lip region even in the case of mouth opening. Furthermore, we can utilize the morphological opening with 3×3 structuring element so as to make the edge more smooth. The result is denoted as B_{lip}.

For the foreground pixels in B_{lip}, the corresponding positions, i.e. the row and column indices, are recorded and compose an $M \times 2$ matrix, denoted as P, where M denotes the number of foreground pixels in B_{lip}. We calculate the eigenvectors and eigenvalues of the covariance matrix of P. Subsequently, we can obtain an ellipse whose position and inclination are defined by the eigenvectors, and the length of major/minor axis are defined by the 1.5 times eigenvalues, respectively. Consequently, two horizontal lines crossing the highest and lowest point of the ellipse are obtained. The continued objects on the outside of the two lines are masked out from the stage of lip contour extraction.

Finally, the quickhull algorithm proposed in [16] is employed to draw the counter of lip (e.g. see Figure 5).

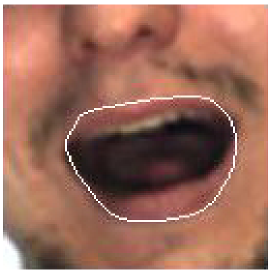


Figure 5. The final result of lip counter extraction.

4 Experimental Results

To demonstrate the performance of the proposed approach in comparison with the existing methods denoted as: Liew03 proposed in [5], and Guan08 in [7]. We utilized the four databases to test the robustness in different capture environments: (1) AR face database (126 people with 26 images for each) [17], (2) CVL face database (114 persons with 7 images for each) [18], (3) GTAV face database (44 persons with 27 images for each), (4) a database established by ourselves, including 19 persons (10 male and 9 female) with 15 pictures per person corresponding to different mouth shapes. We randomly selected 900 images in total (400 images from AR database, 200 images from CVL database, 200 images from GTAV database, 100 images from our database) and manually segmented the lip to serve as the ground truth. Moreover, in AR database, the images with the feature number 11, 12, 13, 24, 25, 26 (wearing scarf which covers the whole mouth) are not used for this experiment. Some segmentation results can be found in Figure 6.

Algorithm	Liew03	Guan08	Proposed
average OL, %	80.73	45.10	**89.27**
average SE, %	20.15	55.21	**9.32**

Table 1. The segmentation results across the four databases.

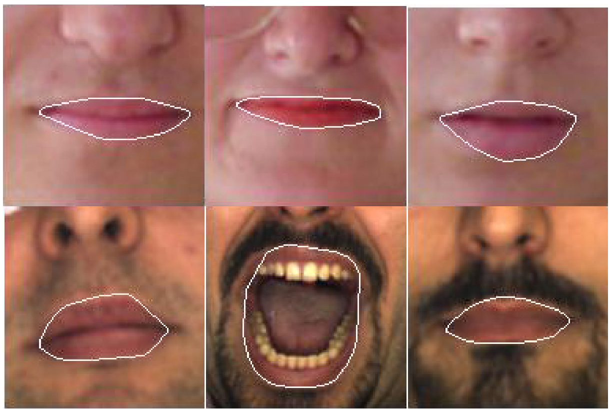


Figure 6. Some samples of lip counter extraction in different databases.

Two measures defined in [5] are used to evaluate the performance of the algorithms. The first measure determines the percentage of overlap (OL) between the segmented lip region A_1 and the ground truth A_2:

$$OL = \frac{2(A_1 \cap A_2)}{A_1 + A_2} \times 100\%. \tag{12}$$

The second measure is the segmentation error (SE) defined as

$$SE = \frac{OLE + ILE}{2 \times TL} \times 100\%, \tag{13}$$

where OLE is the number of non-lip pixels classified as lip pixels (i.e. outer lip error), ILE is the number of lip-pixels classified as non-lip ones (inner lip error), and TL denotes the number of lip-pixels in the ground truth.

Table 1 shows the segmentation results on the four different databases. It can be seen that the proposed method outperforms the Liew03 and Guan08 in both of the two measurements.

5 Conclusion

In this paper, we have proposed a new approach to automatic lip segmentation via the probability model in color space and morphological filter. This approach features the high stability of lip segmentation and robust performance against the different capture environment and different skin color (white and yellow). Experiments have shown the promising result of the proposed approach in comparison with the existing methods.

References

[1] I. Matthews, T.F. Cootes, and J.A. Bangham. Extraction of visual features for lipreading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:198–213, 2002.

[2] W. Gao, Y. Chen, R. Wang, S. Shang, and D. Jiang. Learning and synthesizing mpeg-4 compatible 3-d face animation from video sequence. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(11):1119–1128, 2003.

[3] G. Potamianos, C. Neti, J. Luettin, and I. Matthews. Audio-visual automatic speech recognition: An overview. In G. Bailly, E. Vatikiotis-Bateson, and P. Perrier, editors, *Issues in Visual and Audio-Visual Speech Processing*. MIT Press, 2004.

[4] H.E. Cetingul, Y. Yemez, E. Erzin, and A.M. Tekalp. Discriminative analysis of lip motion features for speaker identification and speechreading. *IEEE Transaction on Image Processing*, 15(10):2879–2891, 2006.

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

[6] W.C. Liew S.H. Leung S.L. Wang, W.H. Lau. Robust lip region segmentation for lip images with complex background. *Pattern Recognition*, 40(12):3481–3491, 2007.

[7] Y.P. Guan. Automatic extraction of lips based on multi-scale wavelet edge detection. *Computer Vision, IET*, 2(1):23–33, 2008.

[8] J. Luettin, N.A. Thacker, and S.W. Beet. Speechreading using shape and intensity information. In *Proceedings of IEEE International Conference on Spoken Language Processing*, pages 58–61, Philadelphia, USA, 1996.

[9] B. Dalton, R. Kaucic, and A. Blake. Automatic speechreading using dynamic contours. In D.G. Stork and M.E. Hennecke, editors, *Speechreading by Humans and Machines: Models, Systems, and Applications*, Berlin, 1996.

[10] D. Chandramohan and P.L. Silsbee. A multiple deformable template approach for visual speech recognition. In *Proceedings of IEEE International Conference on Spoken Language Processing*, pages 50–53, Philadelphia, USA, 1996.

[11] S. Dupont and J. Luettin. Audio-visual speech modeling for continuous speech recognition. *IEEE Transactions on Multimedia*, 2(3):141–151, 2000.

[12] M. Lievin and F. Luthon. Nonlinear color space and spatiotemporal mrf for hierarchical segmentation of face features in video. *IEEE Transactions on Image Processing*, 13(1):63–71, 2004.

[13] N. Eveno, A. Caplier, and P.Y. Coulon. A new color transformation for lips segmentation. In *Proceedings of the 4th IEEE Workshop on Multimedia Signal Processing*, pages 3–8, Cannes, France, 2000.

[14] L. Vincent. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Transactions on Image Processing*, 2(2):176–201, 1993.

[15] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[16] C.B. Barber, D.P. Dobkin, and H.T. Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483, 1996.

[17] A.M. Martinez and R. Benavente. The ar face database. *CVC Technical Report No.24*, June 1998.

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