MULTI-THRESHOLD LIP CONTOUR DETECTION

Panagiota Spyridonos[a], Aggelos Fares Saint[b], Aristidis Likas[b], Georgios Gaitanis[c], Ioannis Bassukas[c]

[a] Department of Medical Physics, Faculty of Medicine, School of Health Sciences, University of Ioannina,University Campus, 45110 Ioannina, Greece
Tel.: +302651007782, Fax.: +302651007854
Email: pspyrid@cc.uoi.gr
[b]Department of Computer Science & Engineering, University of Ioannina, University Campus, 45110 Ioannina, Greece.
Tel.: +302651008810, Fax.: +302651008882
Email: {asaint,arly}@cs.uoi.gr
[c]Department of Skin and Venereal Diseases, Faculty of Medicine, School of Health Sciences, University of Ioannina, University Campus, 45110 Ioannina, Greece.
Tel.: +302651007425, Fax.: +302651007031
Email: {ggaitan,ibassuka}@cc.uoi.gr

ABSTRACT

In this work, we propose a novel, multi-threshold method for lip contour extraction from high-resolution static lip images acquired in an uncontrolled environment and emphasizing on the contour details. The introduced method is a "Divide and conquer" approach. We broke the problem of lip contour extraction into two base sub-problems of locating the upper and lower lip contours using a novel threshold selection algorithm and combining them into the solution for the original lip contour. The method was evaluated on a set of lip images taken from healthy subjects as well as on a set acquired from elder people with degraded lip shapes, and diagnosed with solar cheilosis. Our approach is a low complexity algorithm for robust and accurate lip contour extraction revealing shape details, which might be of value in diverse applied fields such as automated lip biometric and monitoring patients with affected lip contour.

Index Terms— *Lip detection, Lip contour, Multi-threshold, threshold selection*

1. INTRODUCTION

Automatic lip contour detection has become an essential issue in face image analysis. Applications such as human-machine interaction, emotion analysis, person identification and bimodal speech recognition require an efficient and fully automated lip contour detection technique. The development of a robust and accurate approach is an elaborate task due to anticipated inter-person variability, the presence or absence of teeth, the tongue, moustaches or beards, low contrast between lip and skin, high deformable level of lips, illumination conditions, and so forth.

A number of lip segmentation methods based on different methodologies have been proposed. Early approaches utilized basic image processing techniques, such as threshold specific color channels to obtain the lip region [1],[2]. Such approaches are of low computational complexity, but are vulnerable to complexion diversity and variability in the illumination conditions.

Active contours (AC), also known as snakes, have been commonly used for lip segmentation [3],[4] but are mainly dependent on curve initialization and on finding proper values for their parameters. Modified AC models alleviate the inherent problems of the proper initialization and parameter selection but still fail in cases where either facial hair is present such as mustache and/or beard or the contrast between the lip and the surrounding skin region is poor [5],[6].

A significant category of studies has utilized model-based approaches such as deformable templates, active shape models and active appearance models where a priori information about the lip shape is used [7],[8]. Parametric models have been extended and modified by several researchers to handle lip deformation and to improve robustness and accuracy of lip detection [9],[10],[11]. However, model selection is often a compromise between good performance -especially for noisy conditions such as the presence of a beard, uneven illumination, poor contrast of lip and skin- and algorithmic complexity.

Colour clustering has also been used and various modifications of the classical fuzzy c-means algorithm have been proposed to enhance the performance of lip detection. Treating lip segmentation as a two-class segmentation problem, fuzzy c-means clustering incorporating spatial restrictions has improved the separation of the lip and background pixels that otherwise were inseparable due to color similarity [12]. However, the assumption of two class problem breaks down in the presence of facial hair, and in the presence of teeth and/or tongue, leading thus in poor lip segmentation. The automatic selection of the proper number of clusters is of vital importance for robust lip segmentation, however, the corresponding approaches are computationally intensive [13],[14].

Other techniques make use of the Markov random field model to add spatial continuity and to improve the robustness of lip segmentation [15],[16].

All the above methods document varied degrees of accuracy in lip segmentation in their field of application. Yet, there is no clear answer which algorithm is optimum, mainly for two reasons: first the accuracy requirements are different for different applications, and second, there are no benchmark databases and quantitative methods to evaluate their performance [17]. Additionally, the majority of the existing lip localization techniques have been evaluated in low-resolution lip images aiming to determine the lip area and without taking into account the details of lip contour.

In this work, we present a novel, multi-threshold method for lip contour extraction from static lip images acquired in an uncontrolled environment. The proposed method follows a "Divide and conquer" approach. We broke the problem into two base sub-problems of locating the upper and lower lip contour respectively. The extraction of each contour was achieved using a novel threshold selection method and the extracted contours were subsequently combined to provide whole outer lip contour. Since face and mouth detection techniques are well established, the proposed method begins with pre-cropped face images that include only the mouth area. To the best of our knowledge, our work is the first that aims to extract detailed lip contours from high-resolution static images. The latter might be of value in diverse applied fields as in automated lip biometric [18] and monitoring patients with affected lip contour as in solar cheilosis which is the most common precancerous lesion of the lips due to chronic sun exposure [19].

2. LIP EXTRACTION METHOD

2.1. Color transformation

The RGB model is inefficient for lip segmentation since RGB values are sensitive to illumination intensity, and there is significant overlap between lip and skin in all three channels.

It has been demonstrated that correct color transformation can increase segmentation accuracy up to three times [20]. In the present work, images were transformed from RGB to YIQ color space. In [21], it has been observed that the lips are brighter in the Q channel suggesting Q channel for lip detection. In [20] 33 color transforms for lip segmentation have been evaluated. Based on Otsu's discriminant metric the Q component was ranked first.

An important observation is that the lip-skin differentiation based on Q component is not affected by the presence of facial hair. In [22] the Q component has been used to cluster lip and non-lip pixels, and it has been shown, that facial hair pixels are not biasing the color centroid of the non-lip cluster.

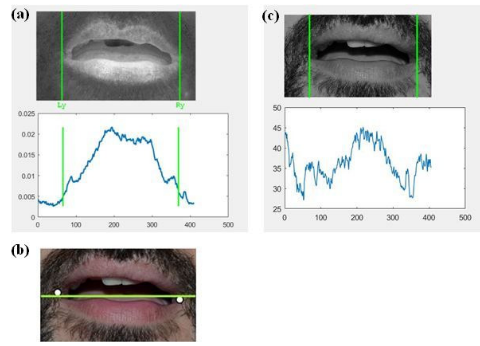

Figure 1: (a) Q chromaticity lip image and the corresponding plot of column-wise standard deviation. Vertical lines are the most left and the most right detectable abrupt changes and correspond to the lip corner columns, $\boldsymbol{L_y}, \boldsymbol{R_y}$ **(b)** Lip corner detection (white dots) and image split into upper and lower part. **(c)** In the presence of facial hair around the lips, column wise standard deviation of the image gray matrix fails to highlight abrupt changes that correspond to lip corner colums.

2.2. Lip Corners Location

Lip corners location is a prerequisite step, for the subsequent lip contour detection because the corners will be used to split the image into upper and lower part. Our approach is a modification of a method proposed by Liu et al. [6]. Assume a lip image matrix of size mxn pixels, and(L_x, L_y), (R_x, R_y) the image coordinates of the left and right lip corner respectively. Image is converted to YIQ color space and the Q chromaticity is retained. The column wise standard deviation of Q matrix forms the signal: $std = std_1, \dots, std_n$ from which L_y and R_y are computed as the most left and the most right detectable changes respectively. Abrupt changes of the std signal were automatically estimated using an optimal detection method of change points [23]. An example is given in Figure 1a. Since in Q chromaticity lips are brighter than surrounding skin, the values of L_x, R_x are computed through the coordinate with the maximum Q value of $Q(i, L_y)$, and $Q(i, R_y)$. Finally, the image row $\left(\frac{(L_x+R_x)}{2}, j\right), j = 1, \dots n$ is used to split the image into the upper and lower lip part (Figure 1b).

Worth to mention that in [6] authors have detected changes in column wise standard deviation of image gray-matrix. However, such approach fails in the presence of facial hair around the lips (Figure 1c). In our approach the use of Q chromaticity successfully tackles this problem.

2.3. Base sub-problem solution for upper and lower lip contour extraction

Our working hypothesis is that there exist effective threshold values of Q chromaticity that adequately separate skin pixels from upper (lower) mouth area including the upper (lower) lip regions. Such threshold values $\mathrm{T_{upper}}$ and $\mathrm{T_{lower}}$ are separately estimated for each image part, by examining the contours produced by several candidate thresholds of chromaticity Q and selecting the best one according to a novel separation criterion.

More specifically, given an image Q (upper or lower part) and a threshold T, the processing steps were the following:

i. We threshold image Q using T to form the binary image, and then find the 4-connected foreground object with the maximum area.

ii. Morphological filters (closing with disk structuring element of radius=10 and opening with structuring element of radius =5) are applied to this object to remove possible artifacts

iii. The most left and right points (P_1, P_2) of the binary object are detected

iv. The exterior boundary ($Contour$) of the object is traced clockwise starting from P_1

v. The part of $Contour$ corresponding to outer lip contour ($cont$) is retained for evaluation (Figure 2).

$$cont = \begin{cases} Contour(P_1{:}P_2), for\ the\ upper\ lip\ contour \\ Contour(P_2{:}P_1), for\ the\ lower\ lip\ contour \end{cases}$$

vi. Assuming l pixels forming an extracted $cont$, we define:

$$Q^{up} = \{q_1^{up}, q_2^{up} \dots q_l^{up}\} \tag{1}$$

and $$Q^{down} = \{q_1^{down}, q_2^{down} \dots q_l^{down}\} \tag{2}$$

Q values of pixels that lie at a vertical distance w bilaterally to the $cont$ pixels.

vii. We compute the average (D) of pairwise absolute differences between the elements of Q^{up} and Q^{down}:

$$\mathrm{D} = \sum_{\mathrm{i}=1}^{l} \sum_{\mathrm{j}=1}^{l} \frac{|\mathrm{q_i^{up}} - \mathrm{q_j^{down}}|}{l^2} \tag{3}$$

In steps vi-vii the evaluation of a contour is made by computing the absolute average difference (D) of the Q values of a set of pixels above the contour (Q^{up}) to a set of points below the contour (Q^{down}), both at vertical distance w bilaterally to the contour pixels. The larger the difference D, the more discriminative the contour is expected to be.

Steps i-vii are repeated for a set of candidate threshold values:

$$thresholds = [Q_{min}{:}\,0.001{:}\,Q_{max}]$$

where Q_{min} and Q_{max} are the minimum and maximum value of Q and the lip contour with the maximum D is selected (Figure 3).

2.4. Extraction of continuous lip contour and quantitative evaluation

To combine the extracted upper and lip contours and form the continuous whole lip outer contour, the missing points around the lip corners must be computed. This is achieved by interpolating two second-degree polynomials that fit each extracted lip contour, in a least-squares sense, with the constraint to pass from the lip corners. A moving average filter of span equal to 7 was used to smooth the final contour (Figure 4).

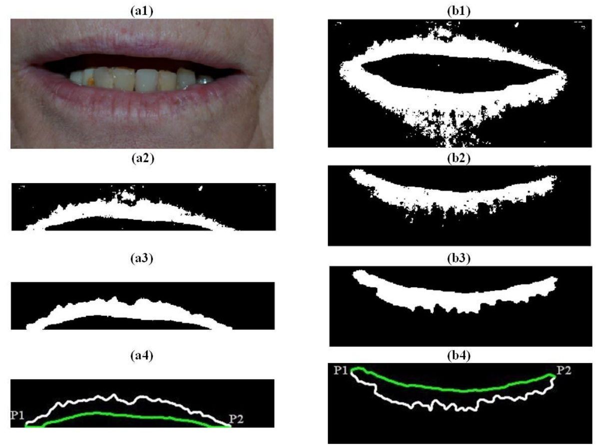

Figure 2: Lip image **(a1)** and the corresponding result of applying Otsu's threshold in Q chromaticity **(b1)**. In this example a threshold value that can adequately distinguish between skin and lip pixels does not exist. Splitting the image into lower and upper lip part, optimal threshold values of Q chromaticity can be found, $\mathbf{T_{upper}}$ and $\mathbf{T_{lower}}$ that separate skin pixels from mouth area including the upper and lower lip regions **(a2-b2)**. Binary images are processed to remove possible artifacts **(a3-b3)**. $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ are the most left and right points of the binary object. Tracing the exterior boundary of the binary object clockwise starting from $\boldsymbol{P_1}$, upper lip contour is extracted by taking the contour points between $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ (**a4**-white line) whereas lower lip contour is extracted by taking the contour points between $\boldsymbol{P_2}$ and $\boldsymbol{P_1}$ (**b4**-white line)

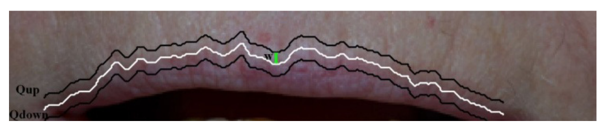

Figure 3: Upper lip contour (white line) resulted by maximizing the normalized sum of pairwise $\mathbf{Q}$ intensity Euclidean distances between lip ($\mathbf{Q^{up}}$) and skin pixels ($\mathbf{Q^{down}}$) that lie at a vertical distance $\mathbf{w}$ to contour pixels

To have an objective view of the algorithm performance and to fine-tune its parameters (the vertical distance w and the set of candidate threshold values), quantitative evaluation of lip contour extraction was carried out using ground truths. The average version of Hausdorff distance (H_{avg}) was employed to evaluate the "closeness" of the manually built ground truth to the contour found by the proposed method. The Hausdorff distance is a favorable measure used in computer vision to determine the degree of resemblance between two objects that are superimposed on one another [24].

The evaluation of the method was performed for both the upper and lower lip as well as for the whole lip contour (Figure 5a-c). For comparison purposes, the Otsu's method, k-means clustering with k=2, and active contours (Chan-Vese method) were tested, which are of similar computation time to our approach.

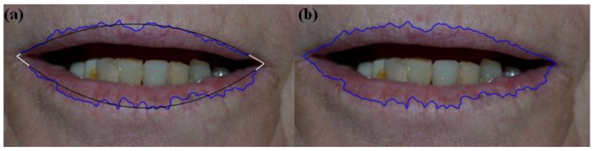

Figure 4: **(a)** Missing points around the lip corners were interpolated (white points) by the values of second-degree polynomials that fit the extracted lip contours and are constrained by the lip corners (dark curves) **(b)** The final lip contour after smoothing.

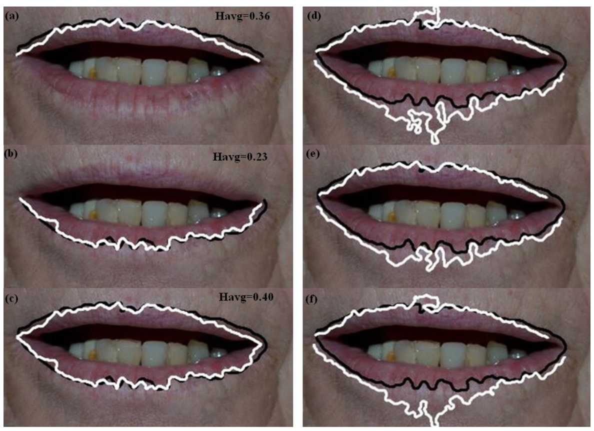

Figure 5: Hausdorff distance estimates the "closeness" of the estimated lip contour (white line) to the ground truth (dark line) for the upper lip **(a)**, lower lip **(b)** and overall lip contour **(c)**. "Divide and conquer" approach for Otsu's method **(d)**, active contours **(e)**, and k-means **(f)**

3. RESULTS AND DISCUSSION

For each lip sub-image, a grid search over the values of chromaticity Q and different values of distance w was performed to extract the upper and lower lip contours. Optimal results were obtained for both healthy and solar cheilosis (SC) lip images, for distance w equal to 7 (Table 1).

The proposed method outperformed the k-means and active contours algorithms and resulted in more accurate contours than Otsus's thresholds (Table 2). A qualitative example is illustrated in Figure 5(d-f). Analyzing the estimated and the Otsu's thresholds (Ot) retrospectively, it came up that optimal threshold values are in Otsu's neighborhood. This observation reduces dramatically the search for optimal thresholds which is constrained in the set:

$$Th = [Ot - k \cdot step, Ot + k \cdot step] \quad (4)$$

where the best values for k and $step$ were 16 and 0.001 respectively.

Splitting the problem of lip contour detection in upper and lower part, we found that the main difficulty in detecting accurate lip contours, in healthy subjects, originates from the upper lip. The latter is mainly due to the lower chromaticity contrast between upper lip and skin. In patients with SC (which are also elder people) the accuracy to detect both upper and lower lip contour is even more challenging. The results verify that our multi-threshold approach can handle more efficiently both the low contrast between upper lip and skin and the affected lips of older adults (Table 2).

Table 1: Upper and lower lip contour quantitative evaluation using Hausdorff metric

N=57 Healthy subjects		
W	Upper Lip	Lower Lip
3	0.595	0.361
5	0.491	0.324
7	0.417	0.298
9	0.482	0.329
N=30 Solar cheilosis subjects		
w	Upper Lip	Lower Lip
3	0.701	0.594
5	0.607	0.601
7	0.581	0.585
9	0.607	0.597

Table 2: Comparison results (Havg)

N=57 Healthy subjects		
	Upper Lip	Lower Lip
Multi-threshold	0.417	0.298
Otsu	0.677	0.337
Active Contours (Chan-Vese)	0.754	0.316
K-means (K=2)	0.781	0.325
N=30 Solar cheilosis subjects		
	Upper Lip	Lower Lip
Multi-threshold	0.581	0.585
Otsu	0.596	0.611
Active Contours (Chan-Vese)	1.374	1.063
K-means (K=2)	1.465	0.999

A limitation of the proposed method is that splitting the image into upper and lower lip using lip corners fails in the case where the face expression deforms the lips in a way that lip corners are located lower than lower lip part.

Note that the proposed method was evaluated using our lip database, where each cropped lip image was around 270x350 pixels since our aim was to extract detailed lip contours from high-resolution images. This is a notable difference with respect to several other works, where the evaluation was on lip area extraction using available face databases [12],[13],[14],[16] with very low image resolution (clipped lip is about ~40x70 pixels). While low image resolution is not a burden to extract ground truths for lip area extraction, it is difficult to accurately specify the ground truth lip contours in such images and use them in our experiments. In future work, we plan to evaluate our method on a bigger lips data set and to expand our approach to extract the inner lip contour investigating different chromaticity channels.

REFERENCES

[1] T. Wark, S. Sridharan, and V. Chandran, "An approach to statistical lip modelling for speaker identification via chromatic feature extraction," in *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*, vol. 1, pp. 123–125.

[2] X. Zhang and R. M. Mersereau, "Lip feature extraction towards an automatic speechreading system," in *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, pp. 226–229.

[3] P. S. Aleksic, J. J. Williams, Z. Wu, and A. K. Katsaggelos, "Audio-Visual Speech Recognition Using MPEG-4 Compliant Visual Features," *EURASIP J. Adv. Signal Process.*, vol. 2002, no. 11, pp. 1213–1227, 2002.

[4] P. Delmas, N. Eveno, and M. Lievin, "Towards robust lip tracking," in *Object recognition supported by user interaction for service robots*, vol. 2, pp. 528–531.

[5] S. Stillittano, V. Girondel, and A. Caplier, "Lip contour segmentation and tracking compliant with lip-reading application constraints," *Mach. Vis. Appl.*, vol. 24, no. 1, pp. 1–18, Jan. 2013.

[6] X. Liu, Y. Cheung, M. Li, and H. Liu, "A Lip Contour Extraction Method Using Localized Active Contour Model with Automatic Parameter Selection," in *2010 20th International Conference on Pattern Recognition*, 2010, pp. 4332–4335.

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

[8] P. Gacon, P.-Y. Coulon, and G. Bailly, "Statistical Active Model for Mouth Components Segmentation," in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 2, pp. 1021–1024.

[9] K. S. Jang and K. S. Jang, "Lip contour extraction based on active shape model and snakes," *INT. J. COMP. SCI. Netw. Secur.*, pp. 148--153, 2007.

[10] Quoc Dinh Nguyen and M. Milgram, "Multi features Active Shape Models for lip contours detection," in *2008 International Conference on Wavelet Analysis and Pattern Recognition*, 2008, pp. 172–176.

[11] W. Lirong, W. Jianlei, X. Jing, and S. Yi, "Lip Contour Modeling Based on Active Shape Model," in *2012 Fifth International Conference on Intelligent Networks and Intelligent Systems*, 2012, pp. 298–301.

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

[13] S.-L. Wang, W.-H. Lau, A. W.-C. Liew, and S.-H. Leung, "Robust lip region segmentation for lip images with complex background," *Pattern Recognit.*, vol. 40, no. 12, pp. 3481–3491, Dec. 2007.

[14] Y. Cheung, M. Li, Q. Peng, and C. L. P. Chen, "A Cooperative Learning-Based Clustering Approach to Lip Segmentation Without Knowing Segment Number," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 1, pp. 80–93, Jan. 2017.

[15] M. Lievin and F. Luthon, "Nonlinear Color Space and Spatiotemporal MRF for Hierarchical Segmentation of Face Features in Video," *IEEE Trans. Image Process.*, vol. 13, no. 1, pp. 63–71, Jan. 2004.

[16] Yiu-ming Cheung, Meng Li, Xiaochun Cao, and Xinge You, "Lip Segmentation under MAP-MRF Framework with Automatic Selection of Local Observation Scale and Number of Segments," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3397–3411, Aug. 2014.

[17] A. W.-C. Liew and S. Wang, *Visual speech recognition : lip segmentation and mapping*. Medical Information Science Reference, 2009.

[18] M. Choraś, "The lip as a biometric," *Pattern Anal. Appl.*, vol. 13, no. 1, pp. 105–112, Feb. 2010.

[19] D. D. Picascia and J. K. Robinson, "Actinie cheilitis: A review of the etiology, differential diagnosis, and treatment," *J. Am. Acad. Dermatol.*, vol. 17, no. 2, pp. 255–264, Aug. 1987.

[20] A. D. Gritzman, D. M. Rubin, and A. Pantanowitz, "Comparison of colour transforms used in lip segmentation algorithms," *Signal, Image Video Process.*, Jan. 2014.

[21] S. Werda, W. Mahdi, and A. Hamadou, "Lip localization and viseme classification for visual speech recognition," *Int. J. Comput. Inf. Sci.*, vol. 5, no. 1, pp. 62–75, Jan. 2007.

[22] P. Spyridonos, G. Gaitanis, M. Tzaphlidou, and I. D. Bassukas, "Spatial fuzzy c-means algorithm with adaptive fuzzy exponent selection for robust vermilion border detection in healthy and diseased lower lips.," *Comput. Methods Programs Biomed.*, vol. 114, no. 3, pp. 291–301, May 2014.

[23] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal Detection of Changepoints With a Linear Computational Cost," *J. Am. Stat. Assoc.*, vol. 107, no. 500, pp. 1590–1598, Dec. 2012.

[24] D. P. Huttenlocher, G. a. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, 1993.