Lip Contour Extraction Scheme Using Morphological Reconstruction Based Segmentation

R. Nath, F. S. Rahman, S. Nath, S. Basak, S. I. Audin, and S. A. Fattah[1]

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
E-mail: [1]fattah@eee.buet.ac.bd

Abstract - **In this paper, a two stage lip contour extraction scheme is proposed, where at the first stage, pixel intensity variation based information is used for obtaining a preliminary estimation of lip and in the second stage, morphological reconstruction based segmentation is employed to detect the final lip region. From a given video frame, first the mouth region is extracted by using block threshold based binary conversion and some morphological operation. Analysing the variation of RGB pixel intensity pattern, intensity ratio based lip region detection is performed, which provides an accurate estimate of lower lip region. However, because of critical shape of the upper lip, further processing is carried out by using morphological opening by reconstruction. Finally, polynomial curve fit is performed to obtain the lip contour. From extensive experimentation on several real-life images from audio-visual clips, it is found that the proposed method offers high level of accuracy in image.**

Index Terms – Morphology, video frame, RGB color space, lip contour, lip detection.

I. Introduction

Lip detection plays significant role in numerous applications, such as audio visual speech recognition, emotion detection from facial image, speaker identification, facial gesture identification and biometric security [1]–[4]. The purpose of lip detection is to extract spatial variation of lip region in audio-visual data. Different approaches of lip contour detection can be broadly categorized into two major classes: the model-based approach and the image based approach. In the model-based approaches, a lip model is prepared and described by a set of model parameters [5]–[7]. Models reported in [6] and [7] require a set of landmark points that can represent shape changes. These models focus on the changes between consecutive images and the features extracted by these models demonstrate good continuity. The model proposed in [7] is able to resolve fine contour details but the shape constraints are difficult to incorporate. Although model based approaches are robust in nature, because of involvement of cost function minimization, they are computationally expensive. On the other hand, image based techniques mainly exploit the differential characteristics in RGB components between lip and skin color [8]–[10]. These methods involve less computational burden and offer low loss of information. In [8], a pseudo hue is proposed as a ratio of RGB values for lip detection. In [9], it is shown that suppressing blue color may improve segmentation quality, as blue color plays a subordinate role in the lip region. Color clustering is also used which works on the assumption that there are only two classes: lip and skin, which may not be true if there is facial hair. In [10], fuzzy color clustering is used where color information and spatial distance between pixels in an elliptical shape function are combined. Markov random field is also used for color based segmentation to make segmentation more robust. There has been very few works on lip detection using morphology [12]–[13]. In [12], dilation followed by opening is employed in time difference images for lip localization. A major drawback of this method is that the background and other facial part except lip are considered as constant. The method proposed in [13] employs only dilation to detect lip corner, which fails to provide accurately overall shape and contour of lips.

The objective of this paper is to develop an efficient lip detection algorithm. First, the mouth region of a facial image is extracted by performing some morphological operation on binary converted image. Next, depending on the ration of pixel intensity values in R and G planes, a preliminary estimate of lip region is obtained. Finally, morphological reconstruction based algorithm, which mainly consists of morphological opening by reconstruction, is employed to obtain accurate lip estimation. Polynomial curve fitting is performed to obtain the lip contour. Experimental results are presented to investigate the lip detection performance of the proposed scheme.

II. Proposed Method

A. Mouth Region Extraction

From given RGB video data, frame by frame analysis is carried out for lip detection. The first task is to identify the region of interest (ROI), here the mouth region that contains the lip. For this purpose, RGB video frame is converted to grayscale image. It is observed that the skin region intensity is greater than that of other portions of the face including lip, nose, eyes and hair. In order to classify pixels into two classes, foreground and background, the grayscale image is divided into non-overlapping blocks of pixels and weighted average of intensity values of each block is computed. Considering a threshold value (T_b) for the block, each block is classified into one of the two classes and the pixel inside the block are labeled as foreground or background pixels. In this way the grayscale image is converted into binary image. In selecting the threshold value, care must be taken such that the lip region be identified as foreground.

Next some morphological operations are performed on the binary image to precisely detect the mouth portion in the whole image. The main idea here is to obtain an image without the lip region and then by subtracting this without lip image from the initial binary image, one may get the lip region. In this regard following two steps are performed:

- Opening by reconstruction: using the binary image as mask image and marker image resulting from erosion of the binary image.

- Closing operation: here the complement of output from opening operation is used as mask image.

Here the objective is to filter out the foreground lip region and foreground region having smaller area than lip. Thus, the disk shaped structural element need to have a radius such that the connected component cannot contain the lip. By choosing a suitable radius, lip and smaller areas (foreground) are turned in to background. Subtracting the binary image from the image obtained after reconstruction, one can get an image where lip region pixels are labeled as one class. There may be some regions having areas smaller than that of the lip region, which will contain pixels similar to lip region. However, it is expected that the choice of lip neighborhood should be such that it converts lip and the area smaller than lip from foreground to background. Searching the highest horizontal length formed by white pixels, mouth region of interest can be now detected. A rectangular ROI containing the lip region is selected. In Fig. 1, all steps involved in the extraction of ROI are shown. It is clearly observed that following the steps described in this section can provide a precise ROI.

B. Preliminary Lip Detection Using Pixel Intensity Ratio

Analysing different RGB lip images, it is observed that R and G values of lip pixels exhibit distinguishable characteristics in comparison to those of non-lip pixels. In Figs. 2(a) and (b), pixel intensity values of R, G, and B planes along horizontal line through the skin region and lip region are plotted, respectively. The blue component values remain almost same in skin and lip region and thus it will play subordinate role in detecting lip. It is clearly observed that pixel intensity ratio in R and G planes significantly differs in case of lip and skin. Clearly the R/G intensity ratio is higher in lip region. Considering a threshold R/G value ($T_{R/G}$), pixels in the ROI can be classified into two classes.

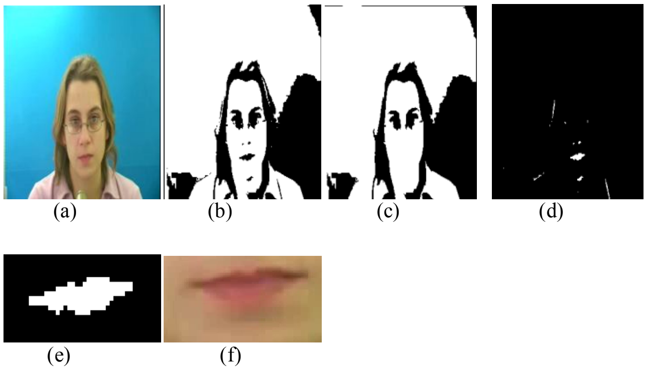

Fig.1 ROI selection. (a) RGB video frame, (b) Binary image obtained based on thresholding non-overlapping blocks. (c) Reconstructed image obtaine from the image in (b) by using morphological operation. (d) Imgae obtained by subtracting image in (b) from the image in (c). (e) Rectangular ROI around the highest horizontal length in white region of the image in (d). (f) RGB ROI corresponding to the ROI in (e).

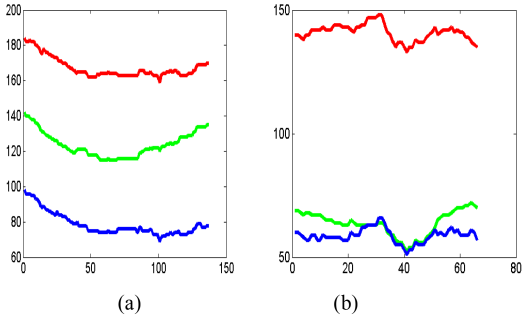

Fig.2 RGB component intensity values in (a) skin region and (b) lip region.

C. Gray Scale Morphological Operation

In order to get precise lip region, morphological operation based processing is employed in the ROI obtained in sub-section II-A. This processing scheme consists of two steps:

- Opening by reconstruction
- Conversion to binary from gray scale image

It is to be mentioned that opening morphology is basically an erosion followed by dilation while morphology opening by reconstruction is an erosion followed by reconstruction. The grayscale ROI image of mouth region is used as mask image for reconstruction. The marker image of reconstruction results from the erosion of grayscale mouth region image. In this way opening by reconstruction is performed, which makes the lip area more compact in the region. The radius of the disk shaped structural element is chosen such that the skin area remains as white and the white feature inside lip (e.g. teeth) cannot be contained by structural element. After morphological opening the skin and lip area will be more distinctive having no white feature inside the lip. The grayscale image is then converted to binary based on a threshold value (*Tg*) the and a perfect upper lip shape is achieved. In Fig. 4, all steps involved in the proposed morphological based lip region detection are presented. One can easily observe that the resulting binary image efficiently extract the lip region, especially the upper

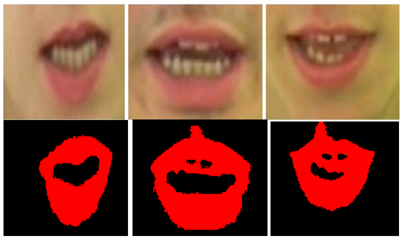

Fig.3: Lip region detected by R/G ratio for different persons

lip is accurately extracted in comparison to the lip detection accuracy obtained by using the pixel intensity ration based method described in sub-section II-B. However, it is to be noted that the level of accuracy obtained in estimating the lower lip is not very satisfactory and it is comparatively better in the case of pixel intensity ration based method.

D. Total Lip Contour

In the proposed lip detection scheme, the upper lip region is extracted by using the morphological based approach and the lower lip region is extracted from the pixel intensity ration based method. Combination of these two extracted lip regions is finally considered as the desired lip region. The purpose of combined approach is to accurately locate the total lip contour. The reason behind this joint approach is that the information, missing in one approach, can be complemented by the other one. In Fig.5, the lip region detection by using the proposed two stages is shown. Along with the original image, in Fig. 5(b), the lip region detected by using pixel intensity ration approach is shown and in Fig. 5(c), the lip region detected by using morphological reconstruction is shown.

It is observed that upper lip shape is better in images after doing reconstruction than the images after R by g threshold and the lower lip shape is better in the images after R by G threshold. As the upper lip shape is quite complicated than that of the lower one, it is divided into upper left and upper right region and polynomial curve fitting for these two regions are employed separately. Then the polynomial curve fit is also performed for the lower lip region. The total lip contour for different persons is shown in Fig. 6. It is clearly observed that the lip contour obtained by the proposed method is very accurate both in case of lower lip and upper lips for every case no matter the speaker is uttering various sounds or in silent mode.

III. Experimental Result

In this section, experimental results are presented considering a widely used publicly available audio visual database GRID. The database contains videos of 34 speakers speaking different sentences. In the simulation results presented in this section, video clips are taken from 8 speakers. For each speaker, 9 different videos are selected, 3 videos containing utterance /a/, 3 containing utterance /u/ and another 3 are silence. For the purpose of experimentation, from each video one image is selected. The lip detection task is carried out independently on 72 images.

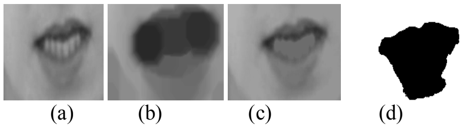

Fig.4 Lip region detected by proposed morphological based operatio. (a) Gray scale image of mouth region of interest, (b) erosion, (c) morphological reconstruction, (d) binary of (c).

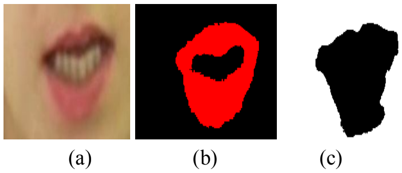

Fig.5 Lip region detection by using the proposed two stages. (a) Original image, (b) lip region detected by using pixel intensity ration approach, and (b) lip region detected by using morphological reconstuction.

These images are selected in such a way that they contain both illumination and shape variations. In determining the mouth ROI, 5×5 non-overlapping blocks are considered, the threshold (T_b) average intensity of non-overlapping blocks is empirically chosen as 120 and the radius of the disk shaped structural element is taken 30. However, further investigation is necessary for selecting a suitable threshold value based on statistical analysis, which would be a potential future work. Finally the size of the rectangular ROI is chosen as 120×220. While detecting lip contour, radius of disk shaped structural element is taken 20. In morphological reconstruction operation, Matlab built-in function “imreconstruct” is used [14].

In pixel intensity ration based approach, the threshold value of R/G pixel intensity ratio ($T_{R/G}$) is taken as 2. In Fig. 7, effect of variation of $T_{R/G}$ on the accuracy of lip region detection is shown. It is clearly observed from this figure that $T_{R/G} = 2$ provides more accurate estimation of lip regions. In Fig. 8, histogram representation of R/G ratio is shown for the two cases, lip and skin regions. It is seen that in the lip region, most of the pixels correspond to pixel intensity ration $T_{R/G} = 2$ and in the skin region, it corresponds to $T_{R/G} = 1$. The ground truth for some lips is shown in Fig. 9. For the purpose of performance measurement of the proposed method, following five different performance indices are considered:

$$\text{Sensitivity, Se} = \frac{\text{True pos.}}{\text{False Neg.} + \text{True pos.}}$$

$$\text{Specificity, Sp} = \frac{\text{True Neg.}}{\text{False Pos.} + \text{True Neg.}}$$

$$\text{Positive Predictive Value, Ppv} = \frac{\text{True Pos.}}{\text{False Pos.} + \text{True Pos.}}$$

$$\text{Negative Predictive Value, Npv} = \frac{\text{True Neg.}}{\text{False Neg.} + \text{True Neg.}}$$

$$\text{Accuracy, Acc} = \frac{\text{True Neg.} + \text{True Pos.}}{\text{False Neg.} + \text{True Neg.} + \text{True Pos.} + \text{False Pos.}}$$

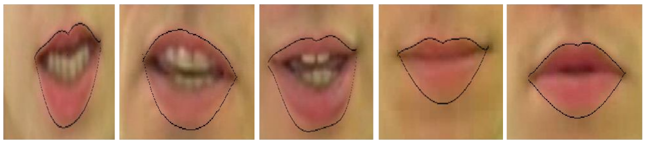

Fig.6 Lip contour extracted by the proposed method for different persons.

Here true positive indicates the event where an original lip pixel is correctly detected as lip pixel and false positive corresponds to non-lip pixel being detected lip pixel. Similarly, true negative indicates the event where an original lip pixel is detected as non-lip pixel and false negative corresponds to non-lip pixel being detected as non-lip pixel. The experimental results, obtained by the proposed method, are shown in Table I in terms of all five indices mentioned above. The average accuracy for all 72 number of tests is achieved as 95.46% .

IV. Conclusion

It is found that the ration of pixel intensity values in R and G planes can provide a preliminary estimate of lip regions. However, using this approach a very satisfactory performance can only be obtained for lower lip region. For accurate estimation of the upper lip region, morphological reconstruction based algorithm is introduced, which is found to be very effective, even in estimating complicated upper lip structures. In order to obtain the complete lip contour, polynomial curve fitting is employed. Experimental results show that the proposed method can provide a high level of accuracy with speaker and illumination variation in terms of sensitivity, specificity, positive predictive value, negative predictive value, and overall accuracy.

References

[1] S. W. Chin, Li-Minn. Ang and Kah Phooi. Seng, “Lips detection for audio-visual speech recognition system,” *International Symposium on Intelligent Signal Processing and Communications Systems*, pp. 1-4, 2008.

[2] M. Rizon, M. Karthigayan, S. Yaacob and R. Nagarajan, “Japanese face emotions classification using lip features,” *Geometric Modeling and Imaging*, pp. 140-144, 2007.

[3] P. Singh and V. Laxmi, “Speaker identification using optimal lip biometry,” *5th IAPR International Conference on Biometrics Compendium, IEEE*, pp. 472-477, 2012.

[4] J. Raheja, L. Shyam, R.,Gupta, J. Kumar and U. Prasad, “Facial Gesture Identification using Lip Contours,” *Second International Conference on Machine Learning and Computing*, pp. 3-7, 2010.

[5] L. Wang, J. Wang, J. Xu and Y. Sun, “ Lip Contour Modeling Based on Active Shape Model,” *Fifth International Conference on Intelligent Networks and Intelligent Systems*, pp. 298-301,2012.

[6] L. Nayoung, H. Chuck, R. Ada, H. Terry and T. Carrel-Ann, “Facial Landmark Extraction for Lip Tracking of Patients with Cleft Lip using Active Appearance Model,” *Communication in Computer and Information Science*, vol. 174, pp. 350-354, 2011.

[7] G.I. Prajapati and N.M. Patel, “DToLIP: Detection and tracking of Lip Contours from human facial images using Snake’s method,” *Int. Conference on Image Information Processing*, pp. 1-6, 2011.

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

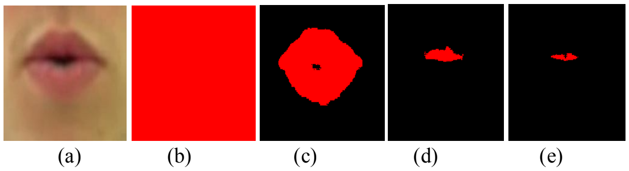

Fig.7 effect of variation of R/G threshold ($T_{R/G}$) on the accuracy of lip region detection. (a) Mouth ROI, (b) $T_{R/G} = 1$, (c) $T_{R/G} = 2$, (d) $T_{R/G} = 3$, and (e) $T_{R/G} = 4$.

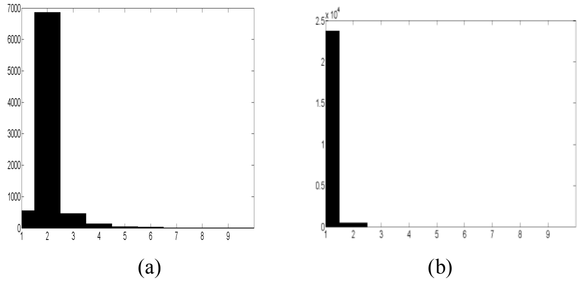

Fig.8 (a) R and G pixel’s intensity ratio in lip region (b) R and G pixel’s intensity ratio in skin region.

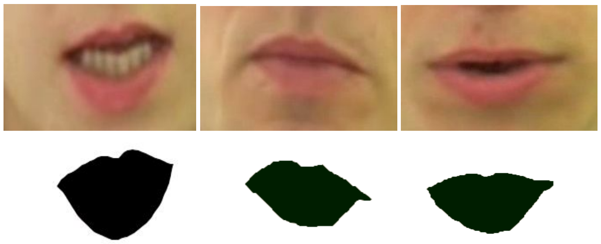

Fig.9 Ground Truth for Different Persons

TABLE I
LIP CONTOUR DETECTION RESULTS

Lip sound	Se	Sp	Ppv	Npv	Acc
/a/	94.76%	95.82%	91.88%	97.15%	95.43%
/u/	92.02%	96.85%	91.52%	97.27%	95.65%
silence	92.28%	95.95%	89%	97.77%	95.3%

[9] U. Canzlerm and T. Dziurzyk, “Extraction of Non Manual Features for Video based Sign Language Recognition,” *Proceedings of IAPR Workshop*, pp. 318-321, 2002.

[10] S. H. Leung S. L. Wang, and W. H. Lau, “Lip image segmentation using fuzzy clustering incorporating an elliptic shape function,” *IEEE Transactions on Image Processing,* vol.13, no.1, pp.51-62, 2004.

[11] S. Lucey, S. Sridharan and V. Chandran, “Initialised eigenlip estimator for fast lip tracking using linear regression,” *Proceedings in 15th International Conference on Pattern Recognition*, vol.3, pp.178-181, 2000.

[12] W. N. Lie and H. C. Hsieh, “Lips detection by morphological image processing,” *Signal processing Proceedings, 1998. ICSP ’98.*, vol. 2, pp. 1084-1087,1998.

[13] A. Das and D. Ghosha, “Extraction of time invariant lips based on Morphological Operation and Corner Detection Method,” *International Journal of Computer Application*, Volume 48, no. 21, pp. 7-11, 2012.

[14] L. Vincent, “Morphological grayscale reconstruction in image analysis: Application and efficient algorithms,” *IEEE Transaction on Image Processing*, vol. 2, pp. 176-201, 1993.