

**RESEARCH PAPER**

Available Online at [www.jgrcs.info](http://www.jgrcs.info)

## LIP CONTOUR DETECTION TECHNIQUES BASED ON FRONT VIEW OF FACE

Prof. Samir K. Bandyopadhyay

Sr Member IEEE, Professor,

Dept of Computer Science & Engineering,

University of Calcutta, 92 A.P.C. Road,

Kolkata – 700009, India

skb1@vsnl.com

**Abstract:** Lip contour detection and tracking is the most important pre-requisite for computerized speech reading. Several approaches have been proposed for lip tracking after lip contour is accurately initialized on first frame. Detection and tracking of the lip contour is an issue in speech reading. A relatively large class of lip reading algorithms are available based on lip contour analysis. In these cases, lip contour extraction is needed as the first step. By lip contour extraction, we usually refer to the process of lip contour detection in the first frame of an audio-visual image sequence. Obtaining the lip contour in subsequent frames is usually referred as lip tracking. While for lip contour tracking there are well developed techniques and algorithms to perform this task automatically, in the case of lip contour extraction in the first frame the things are different. This is a much more difficult task than tracking, due to the lack of a good a-priori information in respect to the mouth position in the image, the mouth size, the approximate shape of the mouth, mouth opening etc. In this paper we propose a solution to automatic lip contour detection if front view of face is available. The proposed method has been tested on a database containing face images of different people and was found to have maximum success rate of 85%.

**Key Words:** Lip contour, Level set evolution, Lip segmentation, and Speech reading

### INTRODUCTION

Lip boundary extraction is an important problem that has been studied to some extent in the literature [1, 2, 3, 4]. Lip segmentation can be an important part of audio-visual speech recognition, lip-synching, modelling of talking avatars and facial feature tracking systems. In audio-visual speech recognition, it has been shown that using lip texture information is more valuable than using the lip boundary information [5, 6]. However, this result may have been partly due to inaccurate boundary extraction as well, since lip segmentation performance was not independently evaluated in earlier studies. In addition, it is possible to use lip segmentation information complementary to the texture information. Lip boundary features can be utilized in addition to lip texture features in a multi-stream Hidden Markov model framework with an appropriate weighting scheme. Thus, we conjecture it is beneficial to use lip boundary information to improve accuracy in AVSR. Once the boundary of a lip is found, one may extract geometric or algebraic features from it. These features can be used in audio-visual speech recognition systems as complementary features to audio and other visual features.

The visual appearance of the human mouth holds a lot of information about the individual it belongs to. It is not only a distinct part of each person's look the lip shape also serves as mean of expressing our emotions. Moreover, the lips' motion indicates if the person is talking and even allows conclusions about what is being uttered. Localizing the exact lip boundaries in an image or video is demanded. Valuable information for various applications with human

computer interaction and in automated surveillance is required in many commercial applications.

In recent years, problems in the automatic speech recognition (ASR) have cropped up and drawn the attention of researchers [1]-[3]. With the presence of noise as in real world circumstances, the ASR rate could be dramatically reduced. The ASR system would be able to provide an appreciable performance only under a certain controlled environment. With the inspiration of lips-reading capability from the impaired society and the limitation of the noise robust techniques, the audio-visual speech recognition (AVSR) has become a research trend and is growing rapidly [4].

Prior to the ASM part, a number of steps are required in order to obtain more information on what is in the image. There are three steps:

- In the face detection step, the face regions in the image frame are detected and localized.
- In each face frame, an eye detector is executed to find both eye positions.
- A smaller part of the face frame containing only the mouth is found, using the eye locations as indicators to the mouth's position. This mouth-frame is the region-of-interest (ROI).

In this paper, we propose an effective method for extracting lip contour. The lip shape is represented as a set of landmark

points and the lip deformation is modeled by a statistically deformable model based ASM. In the traditional ASM, each landmark point is moved independently to the best matching point with its local profile model, so it deforms the lip shape to implausible one and may cause many errors for locating a correct lip contour.

Speech perception is multi-modal in nature, that is, it involves information from more than one sensory modality. With the development of human computer interaction, lip reading technology has become a topic focus in the multimode technologic field. However, detecting and locating lip accurately are very difficult because lip contours of different people, varied luminance conditions, head movements and other factors. Based on the methods of detecting and locating lip we proposed the methods which are based on the lips colour extracted lip contour using the adaptive chromatic filter from the facial images.

It is not sensitive to illumination, but appropriate chromatic lip filter is given by analysing the entire face colour and clustering statistics of lip colour. It is proposed the combinable method which is pre-processing the face image including rotating the angle of face and improving image contrast in this paper and the lip region is analysed clustering characteristics for the skin colour and lip colour, obtained adaptive chromatic filter which can prominent lips from the facial image. This method overcomes the varied illuminate, incline face. The experiments showed that it enhanced detection and location accurately through rough detecting lip region. It lays a good foundation for extraction the lip feature and tracking lip subsequently. Lip contour detection is then performed on isolated lip region using the level set evolution technique for image segmentation. The proposed method was applied on a database having 180 front-view face images of males and females from different regions. Experiment results reveal that the proposed method can detect lip contour in real world face images with a maximum success rate of 85%.

## REVIEW WORKS

Recently, there is an increasing requirement for a system to track and locate human lip[1, 2]. Human lip has much more information than any other face features, so the lip information could be used in image coding [2]. To improve the performance of speech recognition, the lip information is used together with the acoustic signal [3, 4]. The information is also be applied to the graphic animation systems, which need it for generating the lip shape of the speaker[2, 4]. Gradient based techniques [5, 6] for edge detection of lip often fail due to the poor contrast between lip and surrounding skin region. For methods using colour information to build a parametric deformable model for the lip contour, these require optimization technique to refine estimates of contour model to the human lip[7, 8]. Many papers have described the applications of active contour model (snake) for lip boundary detection[9, 10]. The snake

methods are able to resolve fine contour details but shape constraints are difficult to incorporate.

A variety of lip localization methods have been described in the literature throughout the last 15 years. Popular approaches base on colour and intensity thresholding to segment the lips from the rest of the face [11, 12, 13, 14]. Usually the lips are then located by fitting a shape model around the segmented mouth, where many techniques were investigated. Another popular method is the use of snakes in combination with mouth corner feature detection [6, 7]. Also, shape templates have been used in order to localize the lip contours [8]. Another approach is to classify the areas in an image according to the horizontal and vertical intensity profiles, with special consideration of the different casting of shadows in the mouth area [9].

There are several publications that specifically focus on real-time lip tracking. They often use the same methods as mentioned above, maybe as simplified or speed-up variants. For example [10, 11, 12] use the same colour segmentation technique as described above. The colour segmentation based approaches are often lacking robustness to changes of lighting and speakers, but in particular to facial hair. An interesting solution to this was proposed by Petajan et al. [13], where the nostrils' openings were used to determine the approximate mouth location and to estimate the facial hair. A simpler approach was proposed by Yang et al., which only searched for six characteristic points on the lip with characteristic corner features [14]. In a more recent paper Jang et al. propose the use of Gaussian mixture models (GMM) as a replacement for the GLDM [12]. Although the overall detection quality only improves slightly, the placement of the inner lip contour was significantly improved by this mean.

Lip feature extraction, or lip tracking, is complicated by the same problems that are encountered with face detection, such as variation among persons, lighting variations, etc. However, lip feature extraction tends to be more sensitive to adverse conditions. A moustache, for example, can be easily confused to be an upper lip. The teeth, tongue, and lack of a sharp contrast between the lips and face can further complicate lip feature extraction.

Recent techniques use knowledge about the lip's colour or shape to identify and track the lips. Indeed, colour differentiation is an effective technique for locating the lips. A study by [5] showed that, in the hue saturation value colour space, the hue component provides a high degree of discrimination. Thus, the lips can be found by isolating the connected area with the same lip colour. Obviously, colour discriminating techniques will not work for gray scale images. Techniques that use information about the lip's shape include active contour models [13], shape models [14], and active appearance models [8]. Unfortunately, these techniques also require a large amount of storage, which is unattractive from a hardware perspective. In Section IV, we propose a lip feature extraction technique, which makes use of the contrast at the contour of the lips. This technique works well on gray scale images and can be easily implemented on hardware.

The proposed lip feature extraction technique uses the contrast between the lips and the face to locate the four corners of the mouth. The position of the four corners in turn gives an estimate of the mouth's height and width. Notice that the left and right corners of the mouth are where the contrast is highest. The left corner of the mouth is located by searching from the leftmost column of the search area toward the middle. In each column, we take the pixel with the highest contrast and compare its contrast with a threshold.

If the contrast is greater than the threshold, that pixel is considered the left corner and we stop the search. If it is not, we continue to the next column. The threshold can be made a tenable parameter to compensate for different lighting conditions. The right corner is located in a similar way, resulting in the two points. To locate the top of the lips, the proposed technique traces along the edge of the mouth by, starting at the left corner of the mouth, following neighbouring points with the highest contrast. The search is terminated midway between the left and right corners of the mouth.

The bottom of the lips can be found in a similar manner. An example of the search paths traced is shown in following algorithm1, the resulting points denoting the width and height of the lips are shown in Fig. 1. It is shown Region of Interest. Note that the top of the lips indicated falls on the outside of the upper lip and the bottom of the lips indicated falls on the inside of the lower lip. This does in no way affect the ability for other systems to make use of the lip motion information provided as only the motion of the lips is important and not the lips' absolute position.

We found that this technique works better on faces that are larger than  $20 \times 20$  pixels. We found that the face must be at least  $80 \times 80$  pixels for this technique to work well. As such, the hardware implementation detects faces using a  $20 \times 20$  search window, but performs lip motion extraction on faces that are at least  $80 \times 80$  pixels.

## PROPOSED METHOD

A relatively large class of lip reading algorithms is based on lip contour analysis. Different authors tried different procedures to solve the extraction of a good lip contour in the initial frame. Of course, the goal would be to solve this task automatically; approaches like region-based image segmentation and edge detection have been proposed. These methods work quite well in profile images and also in frontal images where the speaker wears lipstick or reflective markers. However, in the frontal images without any marking of the lips, the above-mentioned techniques unfortunately fail; and these images are the most used for speech reading. The problem of automatic extraction of the lip contour becomes even harder in the gray-level images, where the chromatic information differentiating between lips and skin is no longer present. Usually these images have a low contrast, so region-based segmentation and edge detection algorithms fail to provide good results.

The first task involves locating the Region of Interest (ROI). This will be done by manually marking the probable nearest points in the left, right, above and below of the lip region in such a manner so that the rectangle comprising the aforementioned points as the edge points encloses the area of the face where lip is located. Figure 2 showed Lip shape model and Intensity profile.

For the proper extraction of outer lip contour, the following conditions must be assumed:

- a. The manually selected points should lie inside but very near of the boundary of the lip region.
- b.. The distance between two adjacent points should be kept almost the same.
- c.. The total number of manually selected points should be almost the same for the upper and lower part as well as for the left and right part of the lip.
- d. The ridges of the contour should be marked.

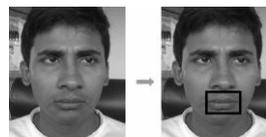


Fig 1 Locating Region of Interest (ROI)

### Algorithm 1

- Step 1. Scan the Image Array Horizontally from left-most pixel to right-most pixel from first row to last row.
- Step 2. Take the first pixel intensity value as a reference value.
- Step 3. Compare intensity of subsequent pixels with the reference value. If the value is same continue to next pixel.
- Step 4. If the value differs, change the value of reference value to the pixel intensity value and mark the pixel black.
- Step 5. If the last row and column pixel is not reached then Goto Step3

### Algorithm 2

- Step 1. Scan the image from the right side of the image to locate the rightmost pixel of the Lip region.
- Step 2. Draw a vertical line along this pixel from top to bottom.
- Step 3. Draw a horizontal line parallel to the top margin passing through the rightmost pixel on the right vertical line to the left base line.
- Step 4. Scan the edge map from the right side to left, on the obtained rectangle, from the first row.
- Step 5. Obtain a pixel that is black indicating an edge path, traverse the pixel path by considering all the surrounding pixels in a clockwise priority and consider the pixel with the highest priority.

- Step 6. The pixels that surrounded the edge pixel, but are of lower priority are stored in a Backtrack Stack to be used only if the traversal process reaches a dead end.
- Step 7. If a dead end is reached, pop out from the Backtrack stack a lesser priority pixel and continue with the traversal process.
- Step 8. Store the pixels traversed in a Plotting List to be used later for drawing the boundary.
- Step 9. Traversal continues to the next pixel till it reaches the left baseline or the bottom of the rectangle.
- Step 10. If the bottom of the rectangle is reached the path is discarded, the plot list is erased and continue from Step5.

Else the path indicated by the Plotting List is plotted on another image indicating edge

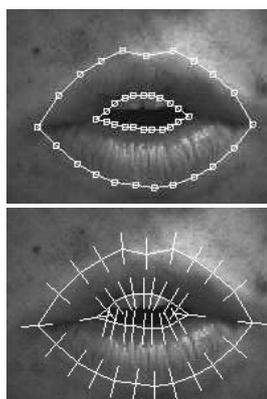


Fig. 2 Lip shape model and Intensity profile

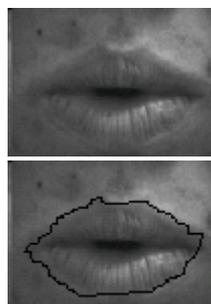


Fig. 3 Contour finding after the algorithm 2

## CONCLUSIONS

For testing the performance of the proposed lip contour extraction algorithm, we used mouth images from the two most used databases in speech reading experiments: Tulips1 [11] and M2VTS [12]. The evaluation of the quality of the results was done visually. In this paper, we propose a method for extracting lip contour. The proposed method was tested to many samples of various shapes and the result showed that it extracted correctly lip shapes that were not extracted by a traditional ASM. The better performance may

be obtained by defining the more global information on lip shape.

## REFERENCES

- [1] Paul Kuo, Peter Hillman and John Hannah, "Improved Lip Fitting and Tracking For Model-Based Multimedia and Coding", International Conference on Visual Information Engineering Conference, Glasgow, UK, pp. 251-258, 2005.
- [2] Mohammad Sadeghi, Josef Kittler and Kieron Messer, "Segmentation of Lip Pixels For Lip Tracker Initialization", International Conference on Image Processing, ICIP, IEEE, Greece, 2001.
- [3] Rainer Stiefelhagen, Jie Yang, Alex Waibel, "A Model based Gaze Tracking System", Proc. of IEEE International Joint Symposia on Intelligence and Systems, pp. 304-310, Rockville Maryland, 1996.
- [4] Nicolas Eveno, Alice Caplier, Pierre-Yves Coulon, "Accurate and quasi-automatic lip tracking", IEEE Trans. Circuits Syst. Video Technology, vol. 14, no. 5, pp. 706-715, 2004.
- [5] C. Neti, G. Potamianos, J. Luetten, I. Matthews, H. Glotin, and D. Vergyri, Large-vocabulary audio-visual speech recognition: A summary of the Johns Hopkins Summer 2000 Workshop, Proc. Works. Multimedia Signal Process. (MMSP), pp. 619-624, Cannes, France, 2001.
- [6] G. Potamianos, C. Neti, G. Gravier, A. Garg, and A.W. Senior, Recent advances in the automatic recognition of audio-visual speech, Invited, Proceedings of the IEEE, vol. 91, no. 9, pp. 1306-1326, 2003.
- [9] Ara V.Nefian, Luhong Liang, Xiaobo Pi, Xiaoxing Liu, and Kevin Murphy, Dynamic Bayesian networks for audio-visual speech recognition, Eurasip Journal on Applied Signal Processing 2002, Vol. 2002, Issue 1, pp1274-1288
- [10] Trent W.Lewis, and David M.W.Powers, Audio-visual speech recognition using red exclusion and neural networks, Journal of Research and Prac. In Info. Tech., Vol.35, No.1, 2003, pp41-63.
- [11]J. R. Movellan. "Visual Speech Recognition with Stochastic Networks," in Advances in Neural Information Processing Systems, (G. Tesauro, D. Toruetszky, and T. Leen, Eds.), Vol 7, MIT Press, Cambridge, MA, 1995
- [12] S. Pigeon and L. Vandendorpe. "The M2VTS multimodal face database," in Lecture Notes in Computer Science: Audio- and Video- based Biometric Person Authentication (J. Bigun, C. Chollet and G. Borgefors, Eds.),vol. 1206, pp. 403-409, 1997
- [13] Robert Kaucic, Barney Dalton, and Andrew Blake, Real-time lip tracking for audio-visual speech recognition applications, Proc. Of the 4th Euro. Conf. on Comp. Vis., Vol 2, pp376-387, Springer-Verlag, 1996.
- [14] XiaoZheng Zhang, Charles C. Broun, Russell M. Mersereau, and Mark A. Clements, Automatic speechreading with applications to human-computer interfaces, Eurasip Journal on Applied Signal Processing, Vol. 2002, Issue 11, pp 1228-1247.