Extraction of Time Invariant Lips based on Morphological Operation and Corner Detection Method

Alak Das
Department of Electronics and Communication Engineering
National Institute of Technology, Agartala
Agartala, India-799055

Dibyendu Ghoshal
Department of Electronics and Communication Engineering
National Institute of Technology, Agartala
Agartala, India-799055

ABSTRACT
The study presents a robust yet simple technique for the extraction of lips from human face. Harris corner detection is applied on a morphologically pre-processed human face and the location and dimension of the lips along the horizontal and vertical directions are obtained. The lip dimension thus found is compared with a predetermined dimension of the same obtained through 'data cursor' operator prevailing in Matlab toolbox. The dimension obtained from the present technique is accepted only when they fall within a certain meaningful latter. The proposed technique is found to yield 88% of average correct rate of lip extraction from human faces of ORL Face Dataset.

General Terms
Image processing, Harris corner detector

Keywords
Lip extraction, morphological operation, Harris corner detector, Matlab toolbox.

1. INTRODUCTION
Lip extraction from human face images has been found to be a useful phenomenon in human face recognition and in the field of facial expression recognition. This has found vast application in which visual information obtained from the mouth region can improve the performance of the vision system [1-3]. Further, it has been established and demonstrated that there are various real life applications like audio-visual person identification. It has been well recognized in scientist community that observation on the lips of the speaker while articulation provide better speech understanding and intelligibility to the audience. In the present scenario of image understanding, recognition and analysis, there is tremendous need for efficient and automated yet accurate method, characterized by high level of robustness to the speaker in spite of variation in image characteristics and illumination level. But the weak skin color contrasts between the lip area and the rest part of the face have made the problem of lip extraction cum detection a difficult one. Various methods towards lip detection have been proposed in gray scale as well as colored images [4-9]. There are a number of analyses such as template models founded on dynamic contours; active shape models [10], deformable templates [11], speaker's lip segmentation algorithm and feature extraction [12] are found in different journals and conference proceedings. Color face images are also dealt with where a color video sequence of speaker's face is acquired by mounting a special micro camera under natural lighting condition and without any particular make-up [13]. Some studies are either based on lip segmentation directly from color space [4, 14] and this type of algorithm are often used a color transformation or color filter to enlarge the gap between the lips and skin. The weak color contrast between the lip and face skin has rendered this type of approach almost inefficient due to the inability of outlining the lip boundary because lip shape is highly variable. The variation comes from individual appearance (spatiotemporal variation). Clustering of the mouth area utilizing color features is found to be another solution [1, 2, 5, 6]. Lip image clustering is usually performed under the assumptions that the number of clusters involving lip and face skin is given. Model based detection/recognition techniques, like deformable templates [11], active shape models [10], snakes [15], generally use a set of features. All these methods are extensive in implementation and require detailed mathematical application.

The present study is based on familiar morphological dilation operation and a well known corner detection algorithm [16]. This process neither requires a detailed hardware nor any involved mathematical technique. It is also found to be insensitive to the color distinction between the lips and adjoining face skin. The processing time for the execution of the algorithm has been found to be small. The present study has yielded satisfactory outcome for detection cum extraction of human lips from the static image of human.

2. MATHEMATICAL BACKGROUND OF PROPOSED METHOD

2.1 Lip Region Extraction
The grayscale image has been used for the extraction of human lip region and detection human lip accurately. To obtain the exact location of human lip, the lip region should be extracted from the given input image. Most observation proves that upper and lower regions between both eyes and lips intensity are high than eyes and lips. So there is an obvious intensity changes in the upper and lower regions of both of eyes and lips. Otsu has proposed [17] a method to convert the grayscale image to binary image on the basis of thresholding technique. The Otsu's method is first applied to convert the input grayscale image to a binary image which assumes that the image to be threshold contains two classes of pixels (e.g. foreground and background). This Otsu's method searches the threshold that minimizes the intra-class variance. Mathematically, it is defined as a weighted sum of variances of the two classes as below,

$$\partial_{\omega}^{2}(t) = w_1(t)\partial_1^2(t) + w_2(t)\partial_2^2(t)$$

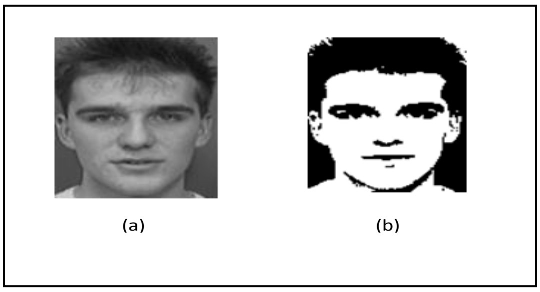

Fig 1: grayscale image and (b) Binary image

Weights ω_i are the probabilities of the two classes separated by a threshold t and ∂^2 variances of these classes. The of minimizing the intra-class variance is the same as maximizing inter-class variance one defined as,

$$\partial_b^2(t) = \partial^2 - \partial_\omega^2 = \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)]^2$$

Where ω_i and μ_i are class probabilities and class means respectively. The threshold t will be obtained when ∂_b gets its maximum. The input grayscale image and binary image followed Otsu's method are shown in Figure 1.

After converting into binary image from input image by Otsu's method, a horizontal projection profile of the binary image is defined as the vector of the sums of the pixel intensities over each row [17]. It has been observed from the horizontal projection that lower intensities are present on the regions of hair, eyes, lips and chin. Mostly, human lips are present in the lower half part of any human face. By considering lower half part of the human face, the lips have the lower intensity than other half part of face. For the lower half part of human face, it is claimed that minimum point of horizontal projection profile is related to lip region. There are two maximum points of horizontal projection profile (lower half part of image) which presents the lower and upper region of lips. Based on the above mentioned from on the lip region can be extracted from the human face image detect in order to go exact location of lips. The horizontal projection profile of binary image and result of lip region extraction are shown in Figure 2.

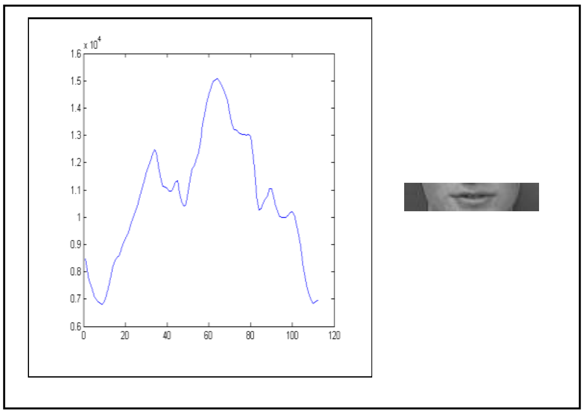

Fig 2: Horizontal projection of binary image and extracted eye region

2.2 Morphological Operation

Morphological operation has been used as means to identify and extract meaningful image descriptors based on properties of form or shape within the image. In this operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbours[20]. The two most basic morphological operations are dilation and erosion. In this paper, dilation has been applied to gradually enlarge the boundaries of region foreground pixels. Thus area of foreground pixels grows in size while holes within those regions become smaller. Hence, it also has been used to create unique background of lip region image. As a result, the accurate minimum corners can be detected on the lip region. In work, dilation operation has been applied on gray scale images. The dilation operator takes two pieces of data as inputs. The first is the image which is to be dilated. The second is a set of coordinate points known as a structuring element (kernel). The structuring element determines the precise effect of the dilation on the input image. Mathematically, dilation is defined in terms of set operation. The dilation of A by B, denoted $A \oplus B$, is defined as,

$$A \oplus B = \{z \mid (\hat{B})_z \cap A \neq \varnothing\}$$

Where, Ø is the empty set, z is pixel element, A is object and B is the structuring element. In words, the dilation of A by B is the set consisting of all the structuring element origin locations where the reflected and translated B overlaps at least some portion of A.

2.3 Creation of Corner on lip region

A corner can be defined as points for which there are two dominant and different edge directions in a local neighbourhood of the point. The main advantages of a corner detector is its ability to detect the same corner in multiple similar images, under conditions of different lighting, translation, rotation and other transforms.

The corner detection block finds the corners in the image based on the pixels that have the largest corner metric values. A simple approach to corner detection in images is using correlation, but this gets computationally very expensive and suboptimal. Harris Corner Detector is one of the important tools to analyze the corner points in images. It is based on the autocorrelation of image intensity values or image gradient values. The gradient covariance matrix M is given by,

$$M = \begin{pmatrix} A & C \\ C & B \end{pmatrix}$$

Where A, B and C are as follows,

$$A = (I_x)^2 \otimes w$$

$$B = (I_y)^2 \otimes w$$

$$C = (I_x I_y)^2 \otimes w$$

I_x and I_y are the gradients of the input image, I in the X and Y direction respectively. The symbol $\otimes$ denotes a convolution operation. The coefficients have been used for separable smoothing filter parameter to define a vector of filter coefficients here. The block multiplies this vector of coefficients by its transpose to create a matrix of filter coefficients *w*.

The Harris corner detection method avoids the explicit computation of the eigenvalues of the sum of squared differences matrix by solving for the following corner metric matrix *R*.

$$R = AB - C^2 - k(A+B)^2$$

The variable k corresponds to the sensitivity factor. The value of k has to be determined empirically, and in this paper we have used the value 0.04. On the basis of R the pixels are classified as follows:

$R > 0$: Corner pixel, $R \sim 0$: pixel in flat region, $R < 0$: Edge pixel.

The detected corner using harris corner detection operation before and after dilation operation has been shown in Figure 3.

2.4 Distance measurement of two lip corners using Data cursor

After detecting corners on lip regions, we first need to measure maximum horizontal and vertical distance of corners on lips. For this purpose, manual distance calculation has been done on lips region of different classes. We have been used the 'data cursor' operator of Matlab, which gives the (x, y) coordinate values of the lip corner pixel, on selection of the pixel.

A Data Cursor is a small black square with a white border that you interactively position on a graph in data cursor mode. When you click a graphic object such as a line on a graph, a data tip appears. Data tips are small text boxes or windows that float within axes that display data values at data cursor locations. Data tips list *x*-, *y*- and (where appropriate) *z*-values for one data point at a time. An example of the process is shown in Figure 4.

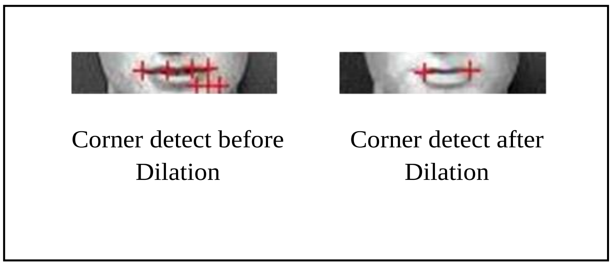

Fig 3: Corner detection after and before Dilation operation

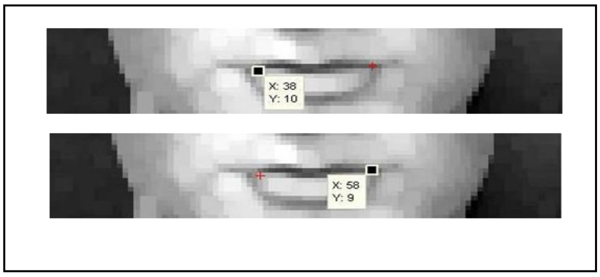

Fig 4: Manual corner Coordinate Measurement using Data Cursor

3. SYSTEM OVERVIEW

The complete system and implementation steps are described with a block diagram shown in Figure 5. At first, we have been applied Otsu's method to convert binary image from input grayscale human face image for find out the maximum intensity region of human face region. In order to find out maximum intensity, the horizontal projection profile has been developed which is defined as the vector of the sums of the pixel intensities over each row. After getting the horizontal projection profile of the binary image, we consider the lower half part of that image to extract the lip region. Because in most of the human face images, lips are present in the lower half part of the image. The lip region extraction has been found from the lower half part of the human face image based on the maximum intensity of upper and lower region of lips. After extraction of lip region, we have been applied morphological dilation operation before corner detector method to ensure that no unwanted corners will be detected on that lip region with background. After that, Harris corner detector which is analyze the corner points in images based on the autocorrelation of image intensity values has been used to detect the corners on the lip region images. After corner detection, we have done manual analysis to find out the possible range of distances between the detected lip corners and used it as the predefined distance. For this purpose, we

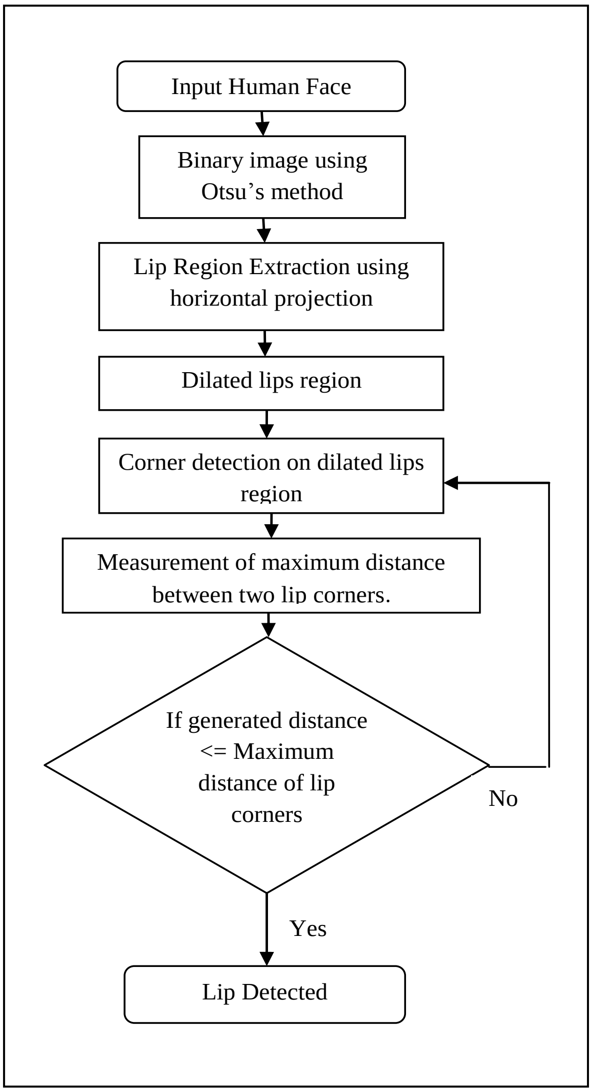

Fig 5: Block Diagram of System

have used 'Data Cursor' operator in Matlab environment which gives the (x, y) coordinate values of the lip corner pixel, on selection of the pixel. Then, we have searched for the lip corners by measuring the distances between the detected corners and comparing them with the predefined distance. After the separation of the lip corners from the other corners, the exact lip was detected.

4. EXPERIMENTAL RESULTS AND DISCUSSION

This work has been simulated using MATLAB R2009b in a machine of the configuration 3.00GHz Intel(R) Core(TM) 2Duo Processor and 3.00GB of Installed Memory. We analyze the performance of our proposed technique using the ORL face database.

4.1 ORL Face Database

There are 10 different visual images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses) [19]. All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The size of each image is 92×112 pixels, with 256 grey levels per pixel.

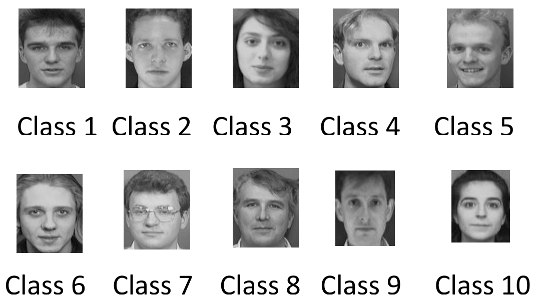

Fig 6: Image samples from ORL Face Database

4.2 Results

In this section, we present our experimental results to assess the performance of the proposed method. For our experiment, 10 different classes of human face from ORL Face Database has been taken in where each class consists of 10 frontal human face images. One image sample per class from 10 different classes are shown in Figure 6. At first, lip region extraction is performed on human face including background using horizontal projection profile after convert into binary image from gray scale image. Then the dilation operation has been applied to remove unwanted lighting object of background by thickening pixel of lip region. After dilation operation, Harris corner detection algorithm has been applied

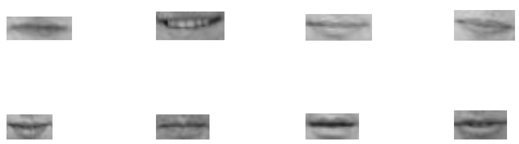

Fig 7: Some Results of Lip Detection

to detect corners on lip region. Because of removing unwanted lighting object, a small number of corners can be detected on lip region. After detecting the corners on lip region, the automatic row-column distance measurement using predetermined distance has been used to detect the exact lip of human face. The predetermined distance between two corners has been calculated using 'Data Cursor' operator of MATLAB. Our experimental results on ORL Face Database shows that lip detection has been found successfully on total no of 88 images among 100 images (10 images per class), which implies that average lip detection rate is 88%. All the results obtained from 10 different classes of proposed dataset have been reported in Table 1. Some lip detection results from 100 images has been shown in Figure 7. The lip detection rate over 10 different classes of human face from proposed dataset also has been shown by graph in Figure 8.

Table 1. Table captions should be placed above the table

Classes used	No of Images taken	No of Detected Image	Detection Rate	Average Detection Rate
Class 1	10	10	100%	88%
Class 2	10	7	70%	
Class 3	10	10	100%	
Class 4	10	9	90%	
Class 5	10	7	70%	
Class 6	10	10	100%	
Class 7	10	8	80%	
Class 8	10	10	100%	
Class 9	10	9	90%	
Class 10	10	8	80%	

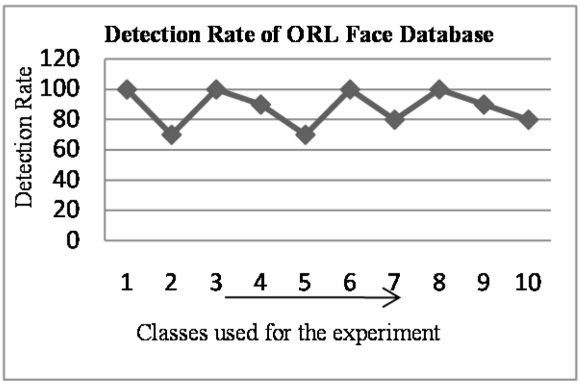

Fig 8: Graph of Lip Detection Rate

5. CONCLUSION

The present study to extract lips from still images of the real life face images has been found to yield 88% rate of detection. Moreover the proposed method is insensitive to color distinction between lips and face skin. The method has been found to be simple yet robust and can be extended for color images also.

In future works, we will try to ameliorate our approach and make it work for smile detection, face recognition and facial expression recognition using other facial expression databases of human face.

6. ACKNOWLEDGMENTS

The authors thankfully acknowledge constant inspiration from Prof (Dr) P.K Bose, Director, National Institute of Technology, Agartala, Jirania-799055, Tripura (west), India to carry out the research work. Shri Alak Das acknowledge the fellowship provided by the National Institute of Technology, Agartala, Jirania-799055, Tripura (west), India for the research work.

7. REFERENCES

[1] Sadeghi, M., Kittler, J., and Messer, K. 2001. Spatial clustering of pixels in the mouth area of face images. Proc. 11th Int. Conf. Image Analysis and Processing, Palermo. April 2001, 36-41.

[2] Matthews, I., Cootes, T. F., Bangham, J. A., Cox, S., and Harvey, R. 2002. Extraction of visual features for lipreading. IEE Trans. Pattern Anal. Mach.Intell. 198-213.

[3] Lucey, S., Sridharan, S., and Chandran, V. 2001. A suitability metric for mouth tracking through chromatic segmentation. Proc. Int. Conf. Image Processing. October. 258-261. Greece.

[4] Eveno, N., Caplier, A., and Coulon, P. Y.2000. A new color transformation for lips segmentation. Proc. 4th IEEE Workshop on Multimedia Signal Processing. 3-8. France.

[5] Wang, S. L., Lau, W. H., Liew, W. C., and Leung, S. H. 2007. Robust lip region segmentation for lip images with complex background. Pattern Recognit. 3481-3491.

[6] Liew, W. C., Leung, S. H., and Lau, W. H. 2003. Segmentation of color lip images by spatial fuzzy clustering. IEEE Trans. Fuzzy Syst. 542-549.

[7] Werda, S., Mahdi, W., and Hamadou, A. B. 2007. Colour and geometric based model for lip localisation: application for lip-reading system. Proc. 14th Int. Conf. Image Analysis and Processing. 9-14. Italy.

[8] Dargham, J. A., and Chekima, A. 2006. Lips detection in the normalized RGB colour scheme. Proc. 2nd IEEE Int. Conf.Information and Communication Tech. 1546-1551 Syria.

[9] Prakash, S., Jayaraman, U., and Gupta, P. 2009. A Skin-Color and Template Based Technique for Automatic Ear Detection. Proc. Int. Conf., Advances in Pattern Recognition. 213-216. India.

[10] Shdaifat, I., Grigat, R., and Langmann, D. 2003. Active shape lip modeling.,Proc. Int. Conf. Image Processing. 875-878. Spain.

[11] Liewa, A. W. C., Leung, S. H., and Lau, W. H. 2000. Lip contour extracting using a deformable model. Proc. Int. Conf. Image Processing. 255-258. Canada.

[12] Wark, T., and Sridharan, S. 1998. A syntactic approach to automatic lip feature extraction for speaker identification. Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. 3693-3696. ISA.

[13] Chiou, G., and Hwang, J. 1997. Lipreading from color video. IEEE Trans. Image Process. 1192-1195.

[14] Gao, W., Chen, Y. Q. Wang, R., Shan, S. G., and Jiang, D. L. 2003. Learning and synthesizing MPEG-4 compatible #-D face animation from video sequence. IEEE Trans. Circuits Syst. Video Technol. 1119-1128.

[15] Delmas, P., Coulon, P. Y., and Fristot, V. 1999. Automatic snakes for robust lip boundaries extraction. Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing. 3069-3072. USA.

[16] Harris, C., and Stephens, M. J. 1988. A combined corner and edge detector. 4th Alvey Vision Conference. 147-151. UK.

[17] Ji, Z., Su, Y., Wang, J., and Hua, R. 2009. Robust sea-sky- line detection based on horizontal projection and hough transformation. IEEE Trans.

[18] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[19] Gonzalez, R. C., Woods, R. E., and Eddins, S. L. 2006. Digital Image Processing Using Matlab. Published by Dorling Kindersley(India) Pvt. Ltd. ISBN 81-7758-898-2.