I.J. Intelligent Systems and Applications, **2012, 8, 30-36**
Published Online July 2012 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2012.08.04

Design and Implementation of Face Recognition System in Matlab Using the Features of Lips

Sasikumar Gurumurthy
School of Computer Science and Engineering, VIT University, Vellore, Tamil Nadu, India
g.sasikumar@vit.ac.in

B.K.Tripathy
School of Computer Science and Engineering, VIT University, Vellore, Tamil Nadu, India
tripathybk@vit.ac.in

Abstract— Human Face Recognition systems are an identification procedure in which a person is verified based on human traits. This paper describes a fast face detection algorithm with accurate result. Lip Tracking is one of the biometric systems based on which a genuine system can be developed. Since the uttering characteristics of an individual are unique and difficult to imitate, lip tracking holds an advantage of making the system secure. We use pre- recorded visual utterance of speakers has been generated and stored in the database for future verification. The entire project occurs in four different stages in which the first stage includes obtaining face region from the original image, the second stage includes mouth region extraction by background subtraction, the third stage includes key points extraction by considering the lip centroid as origin of co-ordinates and the fourth stage includes storing the obtained feature vector in the database. The user who wants to be identified by the system provides the new live information, which is then compared with the existing template in the database. The feedback provided by the system will be 'a match or a miss-match'. This project will increase the accuracy level of biometric systems.

Index Terms— Biometric, Tracking, Centroid, Identification, Origin, Co-ordinates

I. Introduction

The main objective of this system is to identify the speaker using the lip motion vectors which is a more reliable process. It also aims at conducting biometric speaker identification experiments using an audio-visual database and to analyse performance analysis of open-set speaker identification system which is done by Equal error rate (EER) figure. The success of a lip based speaker identification system eventually depends how much of the obtained precision, that is useful for discrimination, is then included in the reduced low dimensional feature set [8]. Lip boundary has to be tracked over time and only motion of lip boundary pixels are taken into account. Before the lip-motion feature extraction, each face image frame is aligned using a 2D parametric motion estimator. For every two consecutive face images global head motion parameters are calculated using hierarchical Gaussian image pyramids and 12- parameter quadratic motion model. Then the face images are warped according to these calculated parameters. After this alignment, the motion vectors from the lip frames of size 128 × 80 are extracted using hierarchical block-matching technique. Lip contour extraction is done by parametric model fitting. The log-ratio of the speaker likelihoods and the world class likelihood results in a stream of log-likelihood ratios that are used in the speaker identification system.

A. Goals

A quasi-automatic system to extract and analyse robust lip-motion features are presented for the open-set speaker identification problem. To be concluded that the utilization of the grid points for motion vector computation is better than using only lip contour points. Accurate and robust lip motion information is an asset to improve the performance of unimodal (i.e. speech-only) systems, which are mostly corrupted by noise in real-life [4, 5].

B. Motivation

The extracted lip shape information can explicitly be included and exploited in the feature set. However, as stated before and demonstrated in our experiments, robustness issue in lip contour tracking is still an unsolved problem. Noisy motion vectors are mostly eliminated at the cost of disregarding some useful motion information around the lip. The lip boundary has to be tracked over time and only motion of lip boundary pixels are taken into account. It is quite natural to assume that lip movement would also characterize the identity of an individual as well as what the individual is speaking. Showing the improved performance of speech-lip fused systems over those of speech-only systems.

II. Problem Definition

Lip motion feature extraction and speaker identification are scenarios in our system. In lip motion

feature extraction firstly lip contour tracking is performed and then lip motion representation is done. There are a number of approaches such as splines, active contours, and parametric models in the literature in order to represent and extract the lip contour. Classical active contours and splines suffer from complex parameter tuning and they are mostly unable to perfectly fit to the characteristic lip parts such as Cupidon's bow because of the erroneous gradient information due to illumination differences [7, 9].

A. Software Context

The design and implementation provides support for the next generation of information performing biometric speaker identification systems. The temporal characterization of the lip motion modality is performed using MATLAB. Word-level continuous-density HMM structures are built for the speaker identification task. Each speaker in the database is modelled using a separate HMM and is represented with the feature sequence that is extracted over the lip stream while uttering the secret phrase.

B. Proposed System

The proposed system can select the best lip motion features for biometric open-set speaker identification. The best features are those that result in the highest discrimination of individual speakers in a population. We first detect the face region in each video frame. The lip region for each frame is then segmented following registration of successive face regions by global motion compensation. The initial lip feature vector is composed of the 2D-DCT coefficients of the optical flow vectors within the lip region at each frame. The discriminate analysis is composed of two stages, at first stage, the most discriminate features are selected from full set of DCT coefficient of a single lip motion frame by using a probabilistic measure that maximizes the ratio of intra-class and inter-class probabilities [1, 4]. At second stage, the resulting discriminative feature vectors are interpolated and concatenated for each time instant within a neighbourhood, and further analysed by LDA to reduce dimension, this time taking into account temporal discrimination information.

C. Process Model

A project work, where the period of completion is confined, it is advisable to use the Rapid Application Development (RAD) Process model. But to proceed with our project work, the best Process Model to our assumption i.e. ITERATIVE PROCESS MODEL (EVOLUTIONARY) has been chosen which is more adaptable with our project. Once the face detection and mouth region detection is constructed, speaker identification is performed with the use of lip motion features strategies once the performance evaluation stage is completed, and if proper efficiency with identification technique is not gained then the algorithm has to be rewritten to get paper efficiency. Those are one of the main reasons to choose the Iterative Process Model where modifications can be done in any stage thus RAD cannot be used for the system development.

D. Feasibility Study

The analyst does study to evaluate the likelihood of the usability of the system to the organization. The feasibility team ought to carry initial architecture and design of the high-risk requirements to the point at which we can answer the question like, if the requirements pose risk that would likely make the project infeasible.[4,5] They have to check if the defects were reduced to a level matching the application needs. The analyst has to be sure that the organization has the resources needed in the requirements may be negotiated. The following feasibility technique has been in this project.i.Economic Feasibility, ii.Technical Feasibility.

III. Overview of Proposed System

We propose to use a much more flexible model made of cubic curves. The algorithm uses edge information and key points position for the segmentation, the key points ensure a good accuracy for the model position, and edge information is used to find the best shape between the key points. It allows a robust and accurate detection of the outer tip boundary the method is divided into three stages: i) Mouth region localization, ii) Key point's extraction, iii) Model fitting. In the first and second steps, mouth region and key points are found by using "hybrid edges", which combine color and intensity information. In the third step, the cubic polynomial models are fitted using key points position and "hybrid edges". Lip boundary is divided in 5 different parts. Each one of them is described by a polynomial curve that fits to the edge. It gives to the global model enough flexibility to reproduce the flexibility of very difficult lip shapes.

A. Module –I: Face Detection

In this phase we detect face region from input video, extracting it into frames. To extract face region we perform lighting compensation on image, then extract skin region and remove all the noisy data from image region. We now find skin color blocks from the image and then check face criterions of the image. In lighting compensation we normalize the intensity of the image, when extracting skin region we apply threshold for the chrominance and then we select the pixels that are satisfying the threshold to find the color blocks.[2,6] The skin color blocks are identified based on the measure properties of image regions in image. Height and width ratio is computed and minimal face dimension constraint is implemented. Crop the current region, existence and localization of mouth then compute vertical mouth histogram.

B. Module –II: Mouth Region Detection

Mouth region localization is done by subtracting of present frame with previous frame, morphing takes place. Skin and lip color analysis is performed. Sorting the areas in descending order and obtain the major

difference region, then find the bounding box for the region we find the centroid of the region and perform this for all the frames. We take mean of all the centroids and extract the mouth region. We calculate all left and right limits, middle boundary, up and down limits. Detection of the key points the three upper points and the mouth corners and the lower points [10]. It is done through two step process. I. Upper edge localization, II. Detection of the key points on this edge.

C. Module –III: Lip Region Extraction

In lip region extraction phase firstly lip color analysis is performed. The polynomial model and fine tuning of corners positions are calculated. We propose a much more flexible model made of 5 independent curves. Each one of them describes a part of the lip boundary. Between P, and P, Cupidon's bow, is drawn with a broken line and the other parts are approximated by 4 cubic polynomial curves y1, they are flexible enough to reproduce the specificity of very different mouth shapes. For each one of them 4 parameters has to be estimated. This ensures a fast and stable convergence for the fitting process. The 4cubic's are fitted by using an edge criterion. The method to find the mouth corners is based on a local criterion.[9] We only consider the value of Luminance along the line Ldn. Most of the time, it is a good cue because the difference of luminance between lips and skin is usually high. In that case the estimation of corners position is reliable because the transition interval, where luminance increases is narrow. However, in particular conditions this interval may be wide and the estimation is coarse.

D. Architectural Model

The architectural style used for developing the lip motion features extraction for speaker identification is PIPES AND FILTER, as output of one process serves as an input for the next process. We move on to face region detection, mouth region localization in this phase we perform skin and lips color analysis, detection of mouth with help of middle boundary, left and right limits, upper and lower limits after calculating this we then go for detection of key points. Key points gives importance cues about lip shape, calculate the three upper points. We calculate mouth corners and the lower points and then lip contour extraction is performed by the polynomial model and fine tuning of corners positions. [10, 5] After lip extraction we go for speaker identification. The extracted lip motion features is tested against various angles to test the robustness and efficiency of the extracted lip features.

IV. Design

The detailed design of the system describes the data and information flow in the system and the way the system works. Design is process of applying the various techniques and principles for the purpose of defining a device, a process or a system in sufficient detail to permit its physical realization. This design specification is vital as it guides the builders to build the system effectively. In case of contour processing, after interpolating both of the motion vectors on *x* and *y* directions to vectors of maximum allowable length in the database the first *Cmax* 1D-DCT coefficients of the motions vectors are combined with possible concatenation of the lip shape parameters.

A. Internal Software Design

In internal software design first we move on to face region detection, mouth region localization in this phase we perform skin and lips color analysis, detection of mouth with help of middle boundary, left and right limits, upper and lower limits after calculating this we then go for detection of key points. Key points gives importance cues about lip shape, calculate the three upper points. The three upper points are located on the Cupidon's bow, near the *ym* level line. They are found through a two steps process: i) Upper edge localization ii) Detection of the key paints on this edge We calculate mouth corners and the lower points and then lip contour extraction is performed by the polynomial model and fine tuning of corners positions. After lip extraction we go for speaker identification.

B. Database Description

Database used in the phases of face region detection, mouth region detection and lip region extraction is MATLAB, for speaker identification we need to calculate centroid of lip and compare with previous values of users that we stored in the database. The database also holds the details of others angles and histogram analysis of motion vectors for lip contour tracking and is periodically checked whether the access frequency has crossed its threshold value in order to compare lip motion features.

We used sequences of different speakers to test our algorithm. They were acquired under natural non uniform lightning conditions and without any particular make-up. Images of the sequence are RGB (8bits/color/ pixel) and contain the region of the face spanning from chin to nostrils. Moreover, we consider that light comes from above, and that the head can be turned so long as the comers of the mouth are visible.

We see that the obtained lip shapes are very realistic and fit to the edges, the model is able to reproduce the specificity of very different speaker's lips. Moreover, the method is robust even in challenging cases such as bearded speaker, non-uniform lighting or if teeth or tongue are visible. In the case of a turned head, the segmentation can still be achieved with accuracy as long as the two comers are visible. In many cases, this first estimation is quite rough.

However, the coarse-to-fine process enables an accurate adjustment of position. For the moment, our method is implemented under MATLAB. Computation time is about 1.5 second for a l00x100 mouth region image.

If a human operator has to find the comers, he implicitly uses the global shape of mouth. He follows

the upper and lower edges extend them when they are becoming indistinct, and finally put the corner where they intersect. We propose an adjustment process that works the same way. Detection is the first step in lips parameterization for later identification. [5, 6]. We have only work with the area around the lips. Because of all pictures were taken with the seated people and at the same distance, this area is selected in all picture in the same position, i.e., we suppose that the lips area inside a predefined block of 230 by 400 pixels.

A multiplayer neural network (MLP) is used for the geometric height and width lip envelope description. The MLP used was trained with back propagation algorithm and a hidden layer containing 120 neurons. Outputs are independently normalized in the range.

C. Detailed Design

Software design is a process with problem solving and solution planning as its main focus. After the definition of the purpose and the analysis of requirements the developers plan the system with the analyzed information in mind. While constructing or developing the system there are a few aspects to consider such as the reliability, accuracy, maintainability, usability and the like which form the non-functional requirements of the system.

V. System Implementation

The implementation methodology outlines key activities that should be considered and planned for when developing or implementing an underwriting system. It begins with the design and development activities involved in obtaining policy details, as a specification, provides a guideline to develop and implement underwriting system.

A. Face Region Detection

In this phase input video is converted to frames and face region is detected as output. This phase has color space transformation and lighting compensation module, high frequency noisy removal module, find out the skin color blocks and height to width ratio detection module.

Step1 Color Space Transformation and Lighting Compensation: In order to apply to the real-time system, we adopt skin- color detection as the first step of face detection we select this transform to detect human skin. However, the luminance of every image is different. It results that every image has different color distribution. Therefore, our lighting compensation is based on luminance to modulate the range of skin-color distribution. First, we compute the average luminance Y_{aveg} of input image.

$$Y_{aveg} = \sum Y_{ij} \qquad (1)$$

where Y_{ij} = 0.3R + 0.6G + 0.1B,

Y_{ij} is normalized to the range (0,255), and i, j are the indices of pixels. According to Y_{aveg}, we can determine the compensated image C_{ij} by following equations:

$$R'_{ij} = (Rij).t \qquad (2)$$

$$G'_{ij} = (Gij).\tau \qquad (3)$$

$$C_{ij} = \{R'_{ij}, G'_{ij}, B'_{ij}\} \qquad (4)$$

where $t = \begin{cases} 1.4, & Y_{aveg} < 64; \\ 0.6, & Y_{aveg} > 192 \ and \\ 1, & \text{otherwise.} \end{cases}$

Note that we only compensate the color of R and G to reduce computation. Due to chrominance (Cr) can represent human skin well, we only consider Cr factor for color space transform to reduce the computation. **Cr** is defined as follows:

$$C_r = (0.5)R' - (0.419)G' - (0.081)B' \qquad (5)$$

In Eq. (5) we can see that R'and G' are important factors due to their high weight. Thus, we only compensate R and G to reduce computation. According to Cr and experimental experience, we define the human skin by a binary matrix:

$$S_{ij} = \begin{cases} 0, & 10 < Cr < 45; \\ 1, & \text{otherwise}; \end{cases} \qquad (6)$$

where "0" represents the white point and "1" represents the black point.

Step2 High Frequency Noise Removal: In order to remove high frequency noise fast, we implement a low pass filter by a 5×5 mask. First, we segment Sij into 5×5 blocks, and calculate how many white points in a block. Then, every point of a 5 × 5 block is set to white point when the number of white points is greater than half number of total points. On the other hand, if the number of black points is more than a half, this 5 × 5 block is modified to a complete black block. Although this fast filter will bring block effect, it can be disregarded due to that our target is to find where human skin is.[1,9]

Step3 Find out skin color blocks: After performing the low pass filter, there are several skin color regions may be human face will be in Si,j. In order to mark these regions, we store four vertices of rectangle for every region. First, we find the leftmost, rightmost, upmost, and lowermost points. By these four points, we create a rectangle around this region. Thus, we can get several skin-color blocks called candidate blocks to detect facial feature.

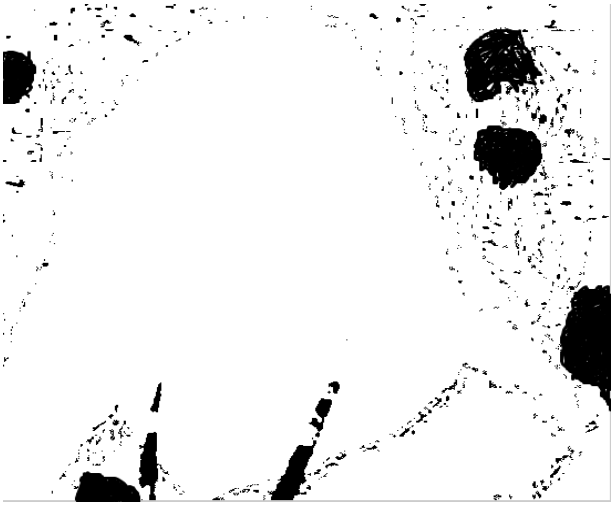

Fig. 1 Skin color Detection block with Noise.

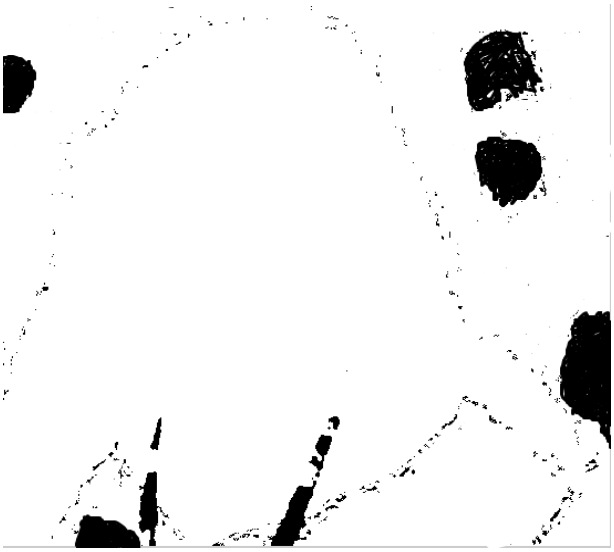

Fig. 2 Skin color Detection block after Noise Removal.

Step4 Height to Width Ratio Detection: After the step of face localization, we can get several regions which may be human face. Then, the feature of height to width ratio, mouth, and eyes are detected sequentially for every candidate block. Because any of these three detections can reject the candidate blocks, low computation module has high priority to process. Height to width ratio is a very fast and simple detection. Let the size of candidate block is h×w [10, 7]. We define that if the height to width ratio (h : w) is out of range between 1.5 and 0.8, it should be not a face and this candidate block will be discarded. Note that the range is determined by experiments. If the ratio is between 1.5 and 0.8 may be a face, the block should be processed by the following two detections.

B. Mouth Region Detection

After determining the height to width ratio for the candidate blocks, morphing takes place based on the image properties of image regions we identify the color blocks in the range, then we go for sorting the areas in the descending order and obtain the major difference region from this region we can find the bounding box for the region. The bonding box plays a vital role in the mouth region detection. Now we can calculate the centriod of the region then we perform this for all the frames, obtain centroid for all frames and take mean of all the centroids calculated, we now can extract the mouth region.

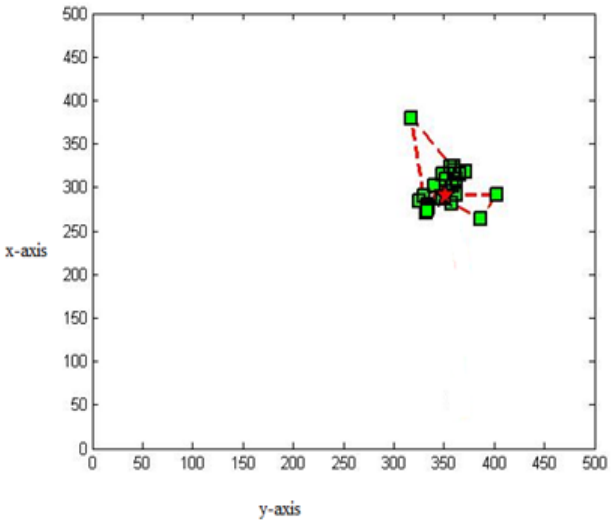

Fig. 3 Graph for Centroid for Background Subtraction.

In the above figure 5.1.2.1, we calculate the centroids for all frames and take mean of all centroids. The red circle region gives the mean of centroids and thus the mouth region is extracted.

C. Lip Region Detection

In RGB space, skin and lip pixels have quite different components. For both, red is prevalent. Moreover, there is more green then blue in the skin color mixture and for lips these two components are almost the same, where R(x,y) and G(x,y) are respectively the red and the green components of the pixel (x,y).[4,8] Unlike usual hue, the pseudo hue is bijective. It is higher for lips than for skin.

Once the image has got the lip shape, we select lip as the biggest object inside the image. Extremas of these detected local maxima pixels will be defined as the left and the right corners of the mouth (XRCorner, YRCorner) and (XLCorner, YLCorner). Get the perimeter of the lip region and find the (x,y) of that perimeter.

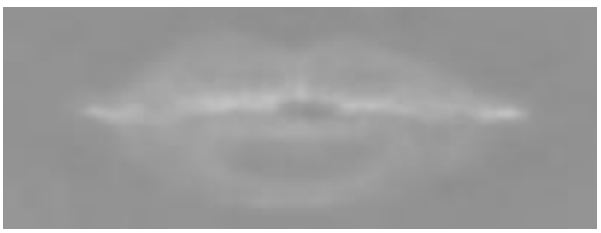

Fig. 4 Pseudo-Hue of Lip Region

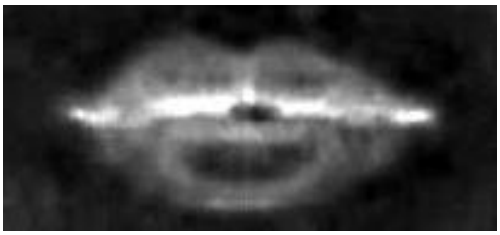

Fig. 5 Adjusted Lip Intensity Values

Fig. 6 Binary Image

Fig. 7 After Filling Hole in Region

Fig. 8 Maximum area of object

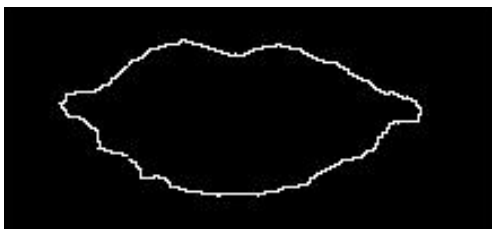

Fig. 9 Perimeter of Lips

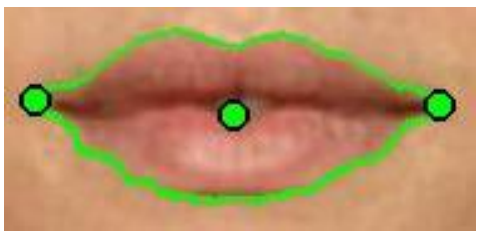

Fig. 10 Corner point and centroid of Lips

VI. Conclusion

In this project, we have presented an automatic, robust and accurate lip segmentation method. Then five points is fitted to the outer lip boundary. Its high flexibility enables very accurate and realistic results. It makes this method very suitable for applications which require a high level of precision such as lip reading. We introduced a new biometric identification system based on lip shape biomeasures, a field in which little research has being done. This is considered as good result and encourage for its use combined with other biometrics systems. A quasi-automatic system to extract and analyse robust lip-motion features is presented for the open-set speaker identification problem. Therefore, if available, accurate and robust lip motion information is an asset to improve the performance of unimodal (i.e. speech-only) systems, which are mostly corrupted by noise in real-life. The proposed lip features can be used in conjunction with audio to improve the performance of the multimodal speaker identification systems. Person authentication can be realized by solely using visual lip features, the uses of shape-based lip features do not warrant acceptable performance.

VII. Future Enhancement

Further enhancements will primarily investigate on different human emotions, even if the person move his lips in different emotions, the performance of the overall speaker identification system using one specific secret phrase (i.e. password) and features fused with corresponding speech modality. The future work will address the robustness of the proposed scheme against noise. Another direction of the future work will be to improve the feature extraction phase via lip tracking and motion estimation. A new visual feature representation incorporating the outer lip contour and inner mouth features is introduced to perform recognition experiments.

Acknowledgment

The authors would like to thank the anonymous reviewers for their careful reading of this paper and for their helpful comments. This work was supported by the author's research guide Dr.B.K.Tripathy.

References

[1] Yu-Ting Pai, Shanq-Jang Ruan, Mon-Chau Shie and Yi-Chi Liu, "A Simple And Accurate Color Face Detection Algorithm In Complex Background", Low Power Systems Lab, Department of Electronic Engineering, National Taiwan University of Science and Technology, No.43, Sec.4, pp. 1545 – 1548.

[2] Nicolas Eveno, Alice Caplier and Pierre-Yves Coulon, "A Parametric Model for Realistic Lip Segmentation", 7th International Conference on

Control, Robotics and Vision, December 2002, pp. 1426 – 1431.

[3] H. E. Cetingul, Y. Yemez, E. Erzin and A. M. Tekalp, "Robust Lip-Motion Features For Speaker Identification", Multimedia, Vision and Graphics Laboratory, pp. 509 – 512.

[4] [4] Enrique Gomez, Carlos M. Travieso, Juan C. Briceno and Miguel A. Ferrer, "Biometric Identification System Using Lip Shape", pp. 39 – 42.

[5] Q.Yuan, W. Gao, and H. Yao, "Robust Frontal Face Detection In Complex Environment", in 16th International Conference on Pattern Recognition, 2002. Proceedings, August 2002, vol. 1, pp. 25–28.

[6] S. Gundimada, Li Tao, and V. Asari, "Face detection technique based on intensity and skin color distribution", in 2004 International Conference on Image Processing, October 2004, vol. 2, pp. 1413–1416.

[7] J. Kovac, P. Peer, and F. Solina, "Illumination Independent Color Based Face Detection", in Proceedings of the 3rd International Symposium on Image and Signal Processing And Analysis, September 2003,vol. 1 , pp.510-515.

[8] R. L. Hsu, M. Abdel-Mottaleb and A. K. Jain, "Face Detection In Color Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, May 2002, vol. 24, no. 5, pp. 696–706.

[9] H.A. Rowley, S. Baluja, and T.Kanade,"Neural network-based face detection," IEEE Transactions on pattern Analysis and Machine Intelleigence,vol.20,pp.23-38,jan 1998.

[10] Chris Boehnen and Trina Russ. Afast multi-modal approach to facial feature detection. In Proc. IEEE Workshops on Appplication of Computer Vision, pages135-142, Breckenridge, Co USA, 2005.

SASIKUMAR Gurumurthy is a Assistant professor (Senior) in the school of computing sciences and engineering, VIT University, at Vellore, Tamil Nadu, India, has published more than 38 technical papers in international journals/ proceedings of international conferences. He is having more than 6 years of teaching experience. He is a member of international professional associations like CSI, IAENG and AIRCC and is a reviewer of AIRCC international journals. Also, he is in the editorial board of AIRCC. His current research directions include detecting technique and signal processing, intelligence computation and soft computing.

B.K Tripathy is a senior professor in the school of computing sciences and engineering, VIT University, at Vellore, India, has published more than 155 technical papers in international journals/ proceedings of international conferences/ edited book chapters of reputed publications like Springer and guided 12 students for PhD. so far. He is having more than 30 years of teaching experience. He is a member of international professional associations like IEEE, ACM, IRSS, CSI, IMS, OITS, OMS, IACSIT, IST and is a reviewer of around 21 international journals which include IEEE, World Scientific, Springer and Science Direct publications. Also, he is in the editorial board of at least 11 international journals. His current research interest includes Fuzzy sets and systems, Rough sets and knowledge engineering, Granular computing, soft computing, Data clustering, Database anonymisation techniques, bag theory, list theory and social network analysis.

How to cite this paper: Sasikumar Gurumurthy,B.K.Tripathy, "Design and Implementation of Face Recognition System in Matlab Using the Features of Lips", International Journal of Intelligent Systems and Applications(IJISA), vol.4, no.8, pp.30-36, 2012. DOI: 10.5815/ijisa.2012.08.04

