



Lip Password Based Personal Authentication

Mule Sandip S., Prof. H. B. Mali.

ENTC Department Pune University, Pune,
Maharashtra, India

Abstract— *The human face is an attractive biometric identifier. Lip password is composed of the password embedded in the lip and underlying characteristics of lip. In the speaker verification; target speaker saying wrong password or correct password will be rejected or detected respectively. This is carried out by visual feature extraction and lip segmentation where former presents the extracted representative visual features while latter aims to separate visibly distinguishable units of each password element. This biometric information can be interpreted as being genetic or behavioural depending on whether static or dynamic features are extracted. So the proposed project work performs favourably compared with traditional authentication system. The improvement obtained suggests that there is enough discriminative information in the mouth-region to enable its use as a primary biometric. Initially we focus on a local feature of the lip and analyses it for its relevance and influence on person recognition. In depth study is carried out with respect to various steps involved, such as detection, evaluation, normalization and the applications of the human lip. The feature extraction technique is to accurately retrieve features from the image. We implement and compare three robust feature detection methods, they are, Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF) and Oriented FAST and Rotated BRIEF (ORB). These methods give significance of ORB over SIFT and SURF algorithms. Use of ORB for personal authentication provides rotation invariance of both train image and query image.*

Keywords— *Biometric Identifier, SIFT, SURF, ORB.*

I. INTRODUCTION

Biometric person recognition has gained vast interest in the scientific community due to several developments in the past few decades. Speaker verification has received considerable attention in the community because of its attractable applications such as financial transaction authentication; secure access control, security protection, and human-computer interfaces. It aims at verifying a claimed speaker using pre-stored information, whereby the speaker will be either accepted as a target speaker or rejected as an impostor under a certain matching criterion. Lip reading has been shown to improve speech recognition accuracy in noisy environments and also is a component technology for multimedia phones for hard-of-hearing people. So the speaker verification by taking into account some video information, e.g., the still frames of face, has shown an improved performance over acoustic only systems. Automatic person recognition is a challenging pattern recognition problem, and several evaluation campaigns have been carried out to evaluate various systems, some biometric recognition techniques such as fingerprint are mature enough for real world applications, but for the rest the present market is quite restricted. Face detection aims to determine the position of a single face in an image. Most of the person authentication methods assume that the location of face in an image is known. Similarly, face tracking algorithms often assume that the initial location of the face is known. Segmentation of a face region from a still image or video is the first step in an automatic person authentication system. The variability in scale or size of the face, orientation, facial expression and lighting conditions complicates the face localization task.

Now a day's human voice recognition is also carried out for security applications. Its performance would be degraded dramatically in the environment corrupted by the background noise or multiple talkers. So we need such a well-designed system which should not be harmful to human body as well as not tedious. Hence we proposed a LIP password based personal authentication.

Here we have to design and develop a robust and secured a lip password based personal authentication using programmable logic in real time environment for person identification. Objectives of the project are given below:

- Authentication should be done basis on LIP only.
- The modality of lip should be completely insensitive to the background noise.
- The acquisition of lip should be somewhat insusceptible to the distance.
- Authentication should be simply applicable to a speech impaired person.

The performance of the existing lip based speaker verification system is degraded due to the two main reasons: 1) The principal feature components representing each lip frame are not always sufficient to distinguish the biometric properties between different speakers. And 2) the traditional lip modelling approaches.

II. LITURATURE REVIEW

Here we will provide a brief review on various previous works on lip password based authentication systems. Biometrics are automated methods of recognizing a person based on a physiological or behavioural characteristic such as Biometric recognition technology relies upon the physical characteristics of an individual, such as fingerprints, voiceprint, pattern of the iris of the eye and facial pattern, in identifying an individual, offering positive identification that is difficult to counterfeit[4]. Xin Liu and Yiu-ming Cheung [1] proposed the Personal authentication is based on LIP only. It is composed of a password embedded in the lip movement and the underlying characteristic of lip. Subsequently, a lip-password protected speaker verification system aiming at holding a double security is established. That is, the claimed speaker will be verified by both of the password information and the underlying behavioural biometrics of lip simultaneously. Accordingly, the target speaker saying the wrong password or an impostor who knows the correct password will be detected and rejected.

The primary and important task in person authentication is to extract features representing the person-specific information [2] in the speech signal and face image. It is known that human beings use high level features such as style of speech and verbal mannerisms (for example, use of particular words and idioms) to recognize speakers from their voice. Intuitively, it is clear that these features contain important speaker specific information. Difficulty arises due to limitations of the existing feature extraction techniques. Current speaker recognition systems use segmental features (which characterize the vocal tract) to represent the speaker-specific information. These features show significant variations across speakers, but they also show considerable variation for a single speaker over a period of time. H.E. Cetingul, Y. Yemez, E. Engin, and A. M. Tekalp determine “Discriminative Analysis of Lip Motion Features for Speaker Identification and Speech-Reading”, gives explicit lip motion information, instead of or in addition to lip intensity and/or geometry information, for speaker identification and speech-reading within a unified feature selection and discrimination analysis framework. But the principal feature components representing each lip frame are not always sufficient to distinguish the biometric properties between different speakers; hence it is quite tedious complex method to implement.

During the past decade, a few techniques such as neural Network (NN), GMM and HMM have been developed for lip motion based applications [2]. In general, the successful achievement of lip motion based speaker verification lies in a closer investigation of the physical process of the corresponding lip motion activities, which always contain strong temporal correlations between the adjacent observed frames. Hence, among these methods, the HMM has been the most popular methodologies because its underlying state structure can successfully model the temporal variations in lip motion activities. Sunil S. Morade and B. Suprava Patnaik , "Automatic Lip Tracking and Extraction of Lip Geometric Features for Lip Reading “have described LIP Tracking scenario[4].LIP Once a region of interest of mouth is located, algorithms can be used for lip contour estimation. Here three methods of active contour for boundary extraction are compared. One of the most common methods of active contour models is snakes. Second method is based on Region scalable fitting energy method (RCFE). Third method is localized active contour model (LACM) which depends upon local (region) information. To our knowledge, in this work both RCFE and LACM methods are used for the first time for lip tracking.

In the proposed method we overcome the drawback present in existing system by using single camera and Speaker verification. Figure 1 show our system is designed by using ARM 32- bit micro controller which supports different features and algorithms for the development of automotive vision systems. Here the camera is connected to ARM controller. The camera will capture the lip movement when we trying to pronounce. If the lip movement is matched with the previously recorded movement and voice then only verification will confirm.

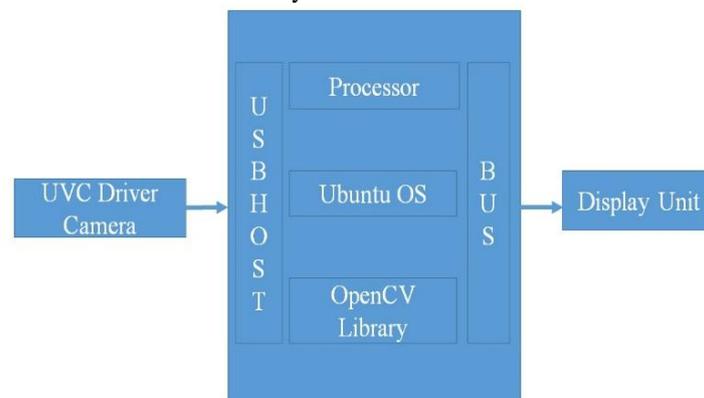


Figure 1: Block Diagram of Overall System

In this method we have to stand in front of camera and make the lip movement as per password and speaking the password. The OpenCV algorithms in our ARM development board match the received lip movement from camera and speaker verification with the previously recorded.

III. SYSTEM IMPLEMENTATION

Initially face recognition based authentication systems detects the face in the scene and it is cropped, followed by normalization for translation scale. This normalized face image is then fed to the face recognition module in order to verify the identity of the person. The main focus of this work is to track the lips in the given face image and then extract suitable features from the lips to perform the differentiation task. Tracking lip motion in image sequences accurately and

robustly is especially difficult because lips are highly deformable, and they vary in shape, colour and size in relation to surrounding features of individuals.

The main content of our approach (1)Feature detection and Descriptor (2)Feature Matching. Complete work is divided into two phase called "Training phase" and "Testing phase". The workflow works in 3 steps.

1. Local Features are extracted from input image.
2. These extracted features are then described with the help of the feature descriptor algorithm and known as feature vector.
3. In the last step of matching several stored feature vector from the database are compared with feature vectors of the input image of testing phase with the help of Euclidean distance. The object is getting recognized with the maximum number of matched feature vector.

A. Hardware Requirements:

The Raspberry pi-2 is the main hardware unit with different module such as the micro SD card, Power adapter, HDMI cable, USB cable. The Raspberry Pi Camera Module attaches to Raspberry Pi by way of one of the small sockets on the board upper surface. This interface uses the dedicated CSI interface.

B. Software Requirements:

The Raspberry Pi uses the Linux kernel. Raspberry Pi has associated with ARM coded Linux drivers. Proposed system uses an OpenCV where OpenCV library have compiled on Raspberry Pi-2 Module. Python programming language is used for system implementation.

IV. RESULTS AND ANALYSIS

A. Results:

Image has been captured by camera and received by AP Raspberry Pi. Figure 2 shows the output window with SIFT Algorithm where person identified with their respective LIP features; when we run the file asift.py on command prompt. Using sift flann we get affine sampling 48: 48. here we get 1474 features from lip image1 and 13241 lip features from image2 matching, as it is crossed threshold hence authenticates homographic estimation so results shows the "Person Identified". Figure 3 shows the output window with SIFT Algorithm where person not identified with their respective LIP features; when we run the file asift.py on command prompt. Using sift flann we get affine sampling 48: 48. here we get 921 features from lip image1 and 11414 lip features from image2 matching, as it is not crossed threshold and 37 matches found only hence It does not authenticate homographic estimation so results shows the "Person NOT Found".



Figure 2: Output with SIFT Algorithm when "Person Identified"

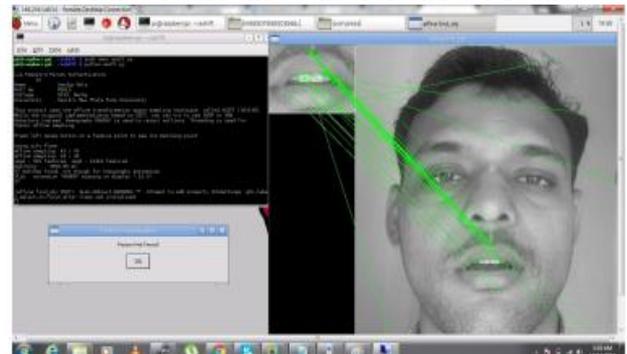


Figure 3: Output with SIFT Algorithm when "Person not found"

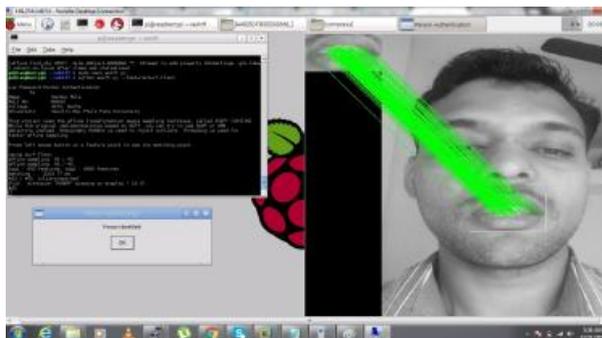


Figure 4: Output with SURF Algorithm when "Person Identified"

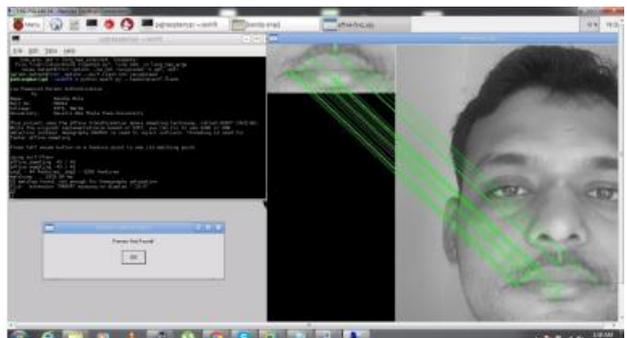


Figure 5: Output with SURF Algorithm when "Person not Found"

Figure 4 shows the output window with SURF Algorithm where person identified with their respective LIP features; when we run the file ashift.py on command prompt. Using surf flann we get affine sampling 48: 48. here we get 642 features from lip image1 and 6606 lip features from image2 matching, as it is crossed threshold hence authenticates homographic estimation so results shows the "Person Identified". Figure 5 shows the output window with SURF Algorithm where person not identified with their respective LIP features; when we run the file asift.py on command

prompt. Using surf flann we get affine sampling 48: 48.here we get 44 features from lip image1 and 3256 lip features from image2 matching, as it is not crossed threshold and 32 matches found only hence It does not authenticate homographic estimation so results shows the "PERSON NOT FOUND". The image has been captured by camera and received by AP Raspberry Pi.

Figure 6 shows the output window with ORB Algorithm where person identified with their respective LIP features; when we run the file ashift.py on command prompt. Using ORB flann we get affine sampling 48: 48. here we get 3955 features from lip image1 and 16996 lip features from image2 matching, as it is crossed threshold hence authenticates homographic estimation so results shows the "Person Identified". Figure 7 shows the output window with ORB Algorithm where person not identified with their respective LIP features; when we run the file asift.py on command prompt. Using surf flann we get affine sampling 48: 48.here we get 474 features from lip image1 and 16437 lip features from image2 matching, as it is not crossed threshold and 14 matches found only hence It does not authenticate homographic estimation so results shows the "Person NOT Found".



Figure 6: Output with ORB Algorithm when "Person Identified"



Figure 7: Output with ORB Algorithm when "Person NOT Found."

B. Analysis:

Comparing to SIFT and SURF on the same data, for the same number of features (roughly 1000), and the same number of scales, Obtained times shown in table 1 These times were averaged over 30 640x480 images from the Pascal dataset. ORB is an order of magnitude faster than SURF and over two orders faster than SIFT.

Table I: Time/Frame Comparison

| Detector | ORB | SURF | SIFT |
|------------|------|-------|--------|
| Time/Frame | 15.3 | 217.3 | 5228.7 |

The below table 2 show that the ORB provides the maximum number of good matches points as compared to SIFT and SURF algorithm. The time required for feature extraction and feature matching is less as compared to SIFT and SURF.

Table II: The Feature Extraction Comparison

| Algorithm | Object 1 Keypoint detected | Object 2 keypoint detected | Good Matches | Feature Extraction time(ms) | Feature matching time(ms) |
|-----------|----------------------------|----------------------------|--------------|-----------------------------|---------------------------|
| SHIFT | 269 | 455 | 28 | 206.86 | 1168.74 |
| SURF | 641 | 842 | 43 | 86.82 | 1892.26 |
| ORB | 454 | 500 | 87 | 52.92 | 73.61 |

V. CONCLUSION

A System is developed using a Raspberry Pi processor acts as the controller. Pi cluster have feature high processing speed, low cost and power efficient architecture Use of Raspberry Pi has solved the cost and bulkiness of the previous work. A speaker will be verified by both of the password embedded in the lip and the underlying behavioral biometrics of lip simultaneously.

Object detection, and video copy detection with the use of utilization of SIFT and also Observed time were averaged over 24 640x480 images. ORB is an order of magnitude faster than SURF and over two orders faster than SIFT. Then system can easily use for the personal authentication schemes. This system may be used in Security applications, Commercial applications, Industrial applications and Bio-medical applications.

REFERENCES

[1] "Learning Multi-Boosted HMMs for Lip Password Based Speaker Verification" Xin Liu, Member, IEEE, and Yiu-ming Cheung, Senior Member, IEEE Trans. on Information Forensic and Security, vol:9, No:2 Feb. 2014.
 [2] "Robust Biometric Person Identification Using Automatic Classifier Fusion of Speech, Mouth, and Face Experts" by Niall A. Fox, Member, IEEE, Ralph Gross, Member, IEEE, Jeffrey F. Cohn, Member, IEEE, and Richard B. Reilly, Senior Member, IEEE, IEEE Trans. on multimedia, Vol.9, No:4, June 2007.

- [3] H. E. Cetingul, Y. Yemez, E. Engin, and A. M. Tekalp, "Discriminative analysis of lip motion features for speaker identification and speech reading", IEEE Trans. Image Process, vol. 15, no. 10, pp. 2879–2891, Oct. 2006. R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
- [4] "Automatic Lip Tracking and Extraction of Lip Geometric Features for Lip Reading", by Sunil S. Morade and B. Suprava Patnaik International Journal of Machine Learning and Computing, Vol. 3, No. 2, April 2013.
- [5] An article on "ASIFT: A new framework for fully Affine Invariant Image Comparison", by Jean-Michel morel & Guoshen Yu.
- [6] Usman Saeed. A thesis on "Lip Analysis for Person recognition", Signal and Image processing. Telecom Paris Tech, 2010. English.
- [7] An Article on "ORB: an efficient alternative to SIFT or SURF", Rublee Vincent Rabaud Kurt Konolige Gary Bradski Willow Garage, Menlo Park, California.
- [8] An article on "Face Components Detection using SURF Descriptors and SVMs", by Donghoon Kim & Rozenn Dahyot, Trinity College Dublin, Ireland.