ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

Facial Expression and Visual Speech Based Person Authentication Using Local Binary Pattern Histogram

S. Saravanan[1], S. Palanivel[2], M. Balasubramanian[3]

Research Scholar, Department of CSE, Annamalai University, TamilNadu, India[1]

Professor, Department of CSE, Annamalai University, TamilNadu, India[2]

Assistant Professor, Department of CSE, Annamalai University, TamilNadu, India[3]

ABSTRACT: This paper is an important milestone towards automatic person authentication. In this work, from a video source, video frames with pose free face is automatically sensed. Then mouth region is localized automatically. Then person identification is done using local binary pattern histogram of mouth region. Person identification performance while using neutral face, smile expression and visual speech are compared. A big custom dataset of 180 videos of 30 persons are used. The result gives the conclusion that for person identification visual speech performs well with F1 score of 0.95, followed by neutral face, 0.93 and then smile expression, 0.87. It also concludes that the mouth component of the face itself is efficient enough to identify persons.

KEYWORDS: Person identification, Local binary pattern histogram, Neutral, Smile expression, Visual speech, Accuracy, F1 score.

I. INTRODUCTION

Image processing done on digital images using algorithms and computer is called digital image processing. A closely related topic of digital image processing is computer vision which involves capturing and processing of images, mostly from videos and take decision based on the processing result. Object recognition is an important part of computer vision. An important application side of object recognition is person authentication. Person authentication can be of two types. It can be either person verification or person identification. Person verification is a one to one comparison, whereas person identification is a one to many comparisons. Along with precision and recall, quality of person identification system is measured in many work using accuracy and F1 Score [1]. Performance of person verification is usually measured using equal error rate [2]. Seven basic facial expressions are anger, disgust, fear, happiness, neutral, sadness and surprise [3]. Among these we use only the neutral and smile expressions for our research comparison in addition to visual speech. Visual speech is detecting visually the mouth region speech activity [4]. Yaw, roll and tilt are poses of faces which degrade the performance of person authentication system. In our research an extension of the automatic pose free face detector used in [5], hereafter called as extended pose free face detector is used. Using local binary pattern histogram, person identification performances while using neutral, smile facial expression and visual speech are compared with our custom dataset. Using methodology used in [5], person verification performance comparison while using neutral, smile facial expression and visual speech using auto associative neural network is checked with our above dataset. Auto associative neural network when trained with features of facial and visual speech has the potentiality to learn the features to be compared later while testing [6]. The comparative performances of above both person authentication while using neutral, smile facial expression and visual speech are compared while using auto associative neural network and local binary pattern histogram. The objective of the work is to find out the best among neutral face, smile facial expression and visual speech on which better automatic person authentication systems can be built upon.

Section II explains about related works. Section III explains about proposed work. Section III A explains about extended pose free face detector and localization of mouth region. Section III B gives the pseudo code for extended pose free face detector and localization of mouth region. Section III C explains about the custom dataset. Section III D explains about the person identification system. Section III E explains about circular local binary pattern histogram. Section III F explains about accuracy, precision, recall and F1 score. Section IV is analysis of results. Section V is conclusions and future work.

II. RELATED WORK

To avoid problems due to scaling, relative size of nose can be used as threshold. Wrong rejection in selecting appropriate images from videos may reduce the time efficiency of the verification system, but not the perfection [5]. Number of researches on face recognition based on a portion of face is very less. Like occlusion and pose variation, illumination variation also disturbs the accuracy of a face recognition system. Face alignment is an important step in face recognition systems and mostly they use eyes or some specific landmarks in the face like in Active Shape Model and Active Appearance Model. Other than frontal view in face recognition needs generally generating frontal views using landmark localization, which will be highly disturbed by occlusion. Near infrared images can also be used for recognition, but quality of the images needed is very high and less tolerant to poses. One half of the face or component of faces like ear, nose, mouth and eye can be used for recognition. Similar to this, face region can be divided manually into many small regions and can be used for recognition by combining the results to manage occlusion [7].

Procrustes analysis can be used to manage pose variations while determining landmark in face images. Perfection in acquiring landmark in face images is mainly disturbed by pose variations in addition to lightings. Many alignment algorithms fail to manage pose variances, even in its starting phase. Holistic face detection algorithms were not perfect in managing elusive shape changes. In such situations parts of face based systems can perform better. Landmark detection based on deformable part models and structure output support vector machines works faster with more perfection. Uniformity is not followed in between manual land marking [8].

Success rate of face recognition methods which uses appearances reduces a lot when the poses of the face changes. Videos give more images for training which can help to increase the accuracy of face recognition even when there is difference in the poses of the image to be probed from the trained. More vertices in model fitting may give more accuracy in face recognition, but needs more computing resource and hence cannot be used in real time systems. When eyes are used for recognition, they should be very near and nose exhibit much similarity, but a speaking mouth shows good variances between persons. In face normalization, chin area can be excluded as the feature points from the chin will not align well. Moreover chin area was not able to show enough variances to enhance recognition. Increase in number of images of training enhances the recognizing result accuracy, even when there are variations in poses. If landmarks are labelled manually, errors may be high and it cannot be completed in stipulated time also. In some situations, same landmark may be noted differently by different researches, which increases the complication further. Detection of mouth automatically was not successful [9].

Obstructed areas of the face are identified and image without obstruction is created and then used for recognition. The starting alignment in spite of variations in pose mainly decides the quality of result in such recognition systems. Many researchers use predetermined threshold for removing obstructions in face recognition, but it will not give good results for all face databases. To identify and remove the object obstructing the face, principal component analysis method was used by using face average for comparison. In this research, 200000 epochs were needed in the neural network training to get the desired result [10].

Numerous face recognition systems which perform well in standard databases, fails miserably in real life situations, as they are not capable of handling variations in lighting and alignment along with obstructions at the same time. Holistic approach face recognition methods may be fast and not so complicated also, but its quality falls rapidly when the face image encounters occlusion. 2D techniques gives better results when pose variation exists. Model based techniques increases the complexity of the system. Increase in the number of images helps to overcome difficulties that arise due to lighting variations. Sparse representation and classification algorithm gives good results for restricted environment images and fails in real life situations as it always assumes perfect fitting of images. No publicly accessible face images with pose have sampled with high concentration. Real world lighting conditions is absent in Multi PIE face dataset. Inclusion of forehead usually reduces accuracy as it may be obstructed frequently by hair. Two bottom nook of the face image can be excluded as many times it contains content which are behind the face. Many algorithms fail to provide good results, when alignment fails. Local Binary Patterns can perform well even when lighting varies. Outer edges of the eyes are usually used for alignment [11].

No existing method is able to manage pose variations perfectly in face recognition task. Nearly all face recognition systems are not tested under real world situations. Usually 3D image generating methods are highly time consuming and not perfect to adopt in real world situations [12].

In recognition of faces, portion based method perform much better than holistic approach methods [13]. Recognition based on local binary pattern histogram are not perplex, which helps the system to work quickly [14]. Compared with methods which handle faces as a whole, local binary pattern method, which handles faces as combination of components, can perform well when change of poses and light exists [15].

III. PROPOSED WORK

A. *Extended Pose Free Face Detector and Localization of Mouth Region:*

In this, from the video source, video frames with pose free faces are detected. In addition the mouth region is also detected using which our proposed work continues. From the video source, face detection is done with extended set of haar like features [16], which is an improved form of [17]. Nose is detected using as in [18] using pattern, knowledge and temporal coherence. Video frames detected with multiple face or multiple nose in the two previous steps are filtered [5]. Detector which utilizes Deformable Part Models using structured output Support Vector Machine [19] is utilized to identify landmarks of the face. From the detected landmarks, nose tip and lip corners are used to localize the mouth region. Nose width is assigned as the width of the mouth region. Half of mouth region width is assigned as the height of the mouth region. As relative quantities are used, even in scaling, same area of the face will be marked as mouth region for a particular person. Nose tip is used to calculate the mouth region horizontal centre. The centre of the vertical line between a lip corner and the nose tip is assigned as the mouth region top. From the mouth region top and mouth height, mouth region bottom is identified. From the mouth region horizontal centre and mouth region width, mouth region both sides are identified. Thus the mouth region is completely localized. Video frames are rejected if any one lip corner is above or below the mouth region, if one lip corner is inside and the other outside the mouth region horizontally, if the absolute difference of right and left side horizontal distance of the lip corner from the mouth region is greater than 5% of nose width, if the absolute difference of right and left side vertical distance of the lip corner from the nose tip is greater than 10% of nose width as like in [5]. Above pose free face detector gives perfect pose free faces, but it has some rare images when the face is in a little far off distance from the camera, where face is detected correctly, but nose is detected wrongly by detecting the entire face, which leads to wrong mouth region detection. To eliminate this, the pose free face detector is extended as extended pose free face detector. In extended pose free face detector, video frame is initially accepted only if the face width is greater than double the width of nose, which increases its performance accuracy further. Sample automatically localized extracted mouth regions of thirty persons during neutral face are shown in Fig. 1.

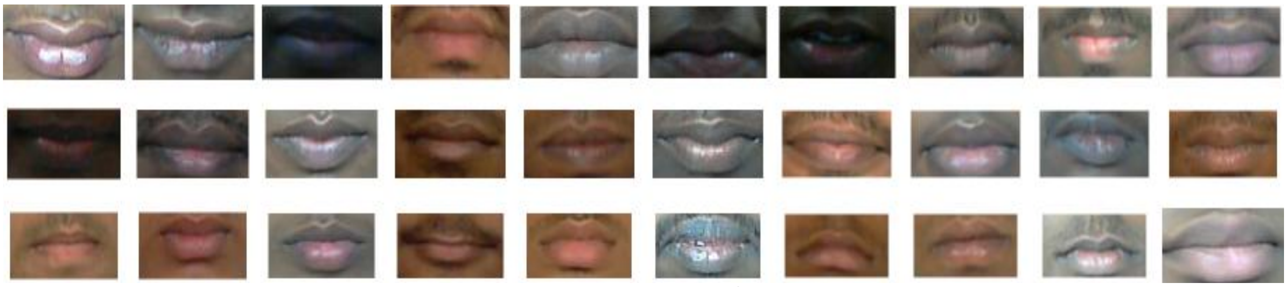
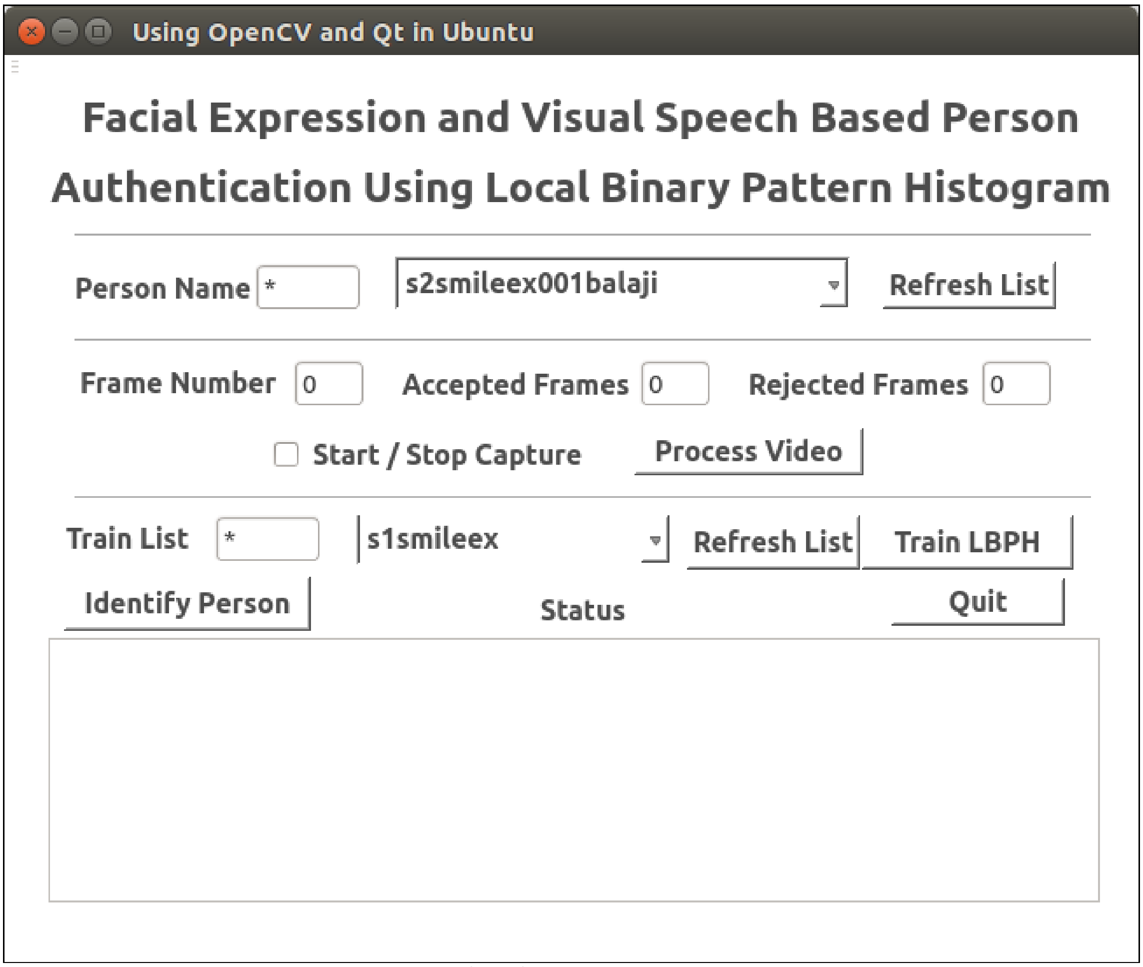


Fig. 1. Automatically Localized Extracted Mouth Region of Thirty Persons During Neutral Face

10.15680/ijircce.2015.0301061



Facial Expression and Visual Speech Based Person Authentication Using Local Binary Pattern Histogram

Person Name * s2smileex001balaji

Frame Number 0 Accepted Frames 0 Rejected Frames 0

□ Start / Stop Capture

Train List * s1smileex

Fig. 2. User Interface of the Person Authentication System

B. *Pseudo code for Extended Pose Free Face Detector and Localization of Mouth Region:*

```
Step 1: Detecting face and nose independently from video source.
Step 2: If (number of detected nose > 1 or number of detected face > 1)
                    The video frame is rejected
Step 3: If (2 x nose width > face width)
The video frame is rejected.
Step 4: Nose tip and two lip corners are identified automatically.
Step 5: Nose tip and lip corners are used to localize the mouth region as detailed above.
Step 6: If (any one lip corner is above or below the mouth region)
                    The video frame is rejected
Step 7: If (one lip corner is inside and the other outside the mouth region horizontally)
                    The video frame is rejected
Step 8: If ((absolute difference (distances between right and left lip corner to mouth region)) > 5% of nose width)
                    The video frame is rejected
Step 9: If (absolute difference (vertical distances between right and left lip corner to nose tip) > 10% of nose width)
                    The video frame is rejected
Step 10: Remaining video frames are accepted as video frames with pose free faces
```

C. *Custom Dataset:*

For our dataset, videos including face are recorded with different poses of face. The persons are asked to exhibit poses, that is roll along with tilt and yaw. The videos are recorded at normal lighting condition with neutral, smile expression and visual speech. Two sessions of video recordings have been done with a gap of approximately twenty days. Thirty persons are recorded out of which four are females. Hence the database becomes very big containing 180 videos as it includes two sessions of 30 persons with neutral, smile expression and visual speech each. Length of each video is approximately two minutes. Resolution of the videos is 640x480 with 15 frames per second.

D. *Person Identification system:*

For this proposed person identification system using local binary pattern histogram, the mouth region images got as output from extended pose free face detector is used as input for person training and testing. Fig. 2 shows the user interface of the system. Ten mouth region images from the first recorded session are taken for each person for training to train in local binary pattern histogram model with label for each person. After the model is trained, ten mouth region images from the second recorded session are taken for each person for testing. Initially a threshold value of fifty is used in the model and all the ten images are tried for identification. If any image fails in identification from the database, the identification procedure is repeated with threshold value 60. The threshold value is increased each time in steps of ten, until a threshold value is reached, where all ten images are identified from the database. The process is repeated for all 30 persons for neutral, smile expression and visual speech independently and confusion matrix is prepared from the output. Using the confusion matrix data and the eq. (1), eq. (2), eq. (3) and eq. (4) accuracy, precision, recall and F1 score is calculated. The resultant values are shown in Table 1.

Table 1. Person Identification Performance while using Neutral Face, Smile Expression and Visual Speech

Person Identification	**Accuracy**	**Precision**	**Recall**	**F1 score**
Neutral Face	0.96	0.96	0.90	0.93
Smile Expression	0.94	0.85	0.88	0.87
Visual Speech	0.97	1.00	0.91	0.95

E. *Circular Local Binary Pattern Histogram:*

The input gray image is divided into smaller divisions from which the histogram features are generated and grouped to form a single feature vector called local binary pattern histogram. In local binary pattern, for each pixel, the intensity of its nearby pixel is considered. If pixel intensity is greater than one of its nearby pixels, then 1 is considered and 0 if smaller. Like that for each nearby pixel a 1 or 0 is generated to form a 8 digit binary number, if 8 nearby pixels are considered. Fig. 3 shows an example local binary pattern generation. So for a pixel there are 256 possibilities of Local Binary Pattern starting from 00000000 to 11111111. To manage under various scaling, the nearby pixels are selected from a round around the pixel under consideration with changing radius. If two pixels get covered by a point, then bilinear interpolation is employed to get the value. A local binary pattern is considered as uniform and circular if it has no changes of bit from 0 to 1, like 00000000 or 11111111 or two transitions like from 0 to 1, then from 1 to 0 or from 1 to 0, then 0 to 1, with first and last bit being the same. Non uniform patterns constitute approximately 10% of all the local binary patterns in textures. Hence while generating local binary pattern histogram, one independent bin is used for all non uniform patterns. All remaining uniform patterns use each one a bin. Circular local binary pattern is also called as extended local binary pattern. If the number of nearby points from the circle is 8 and radius is 2 pixels, then it is denoted as (8,2). Chi square minimum distance estimation is applied to compare histogram. This performs well even when poses and gray level varies. Prior to using local binary pattern on the image there is no need to normalize the intensity values which increases the efficiency [15]. If the size of each division is large, the length of the feature will be short. This short feature helps for faster performance. As the variation of size and other parameters does not vary the result much, time need not be wasted for finding the best parameters [14].

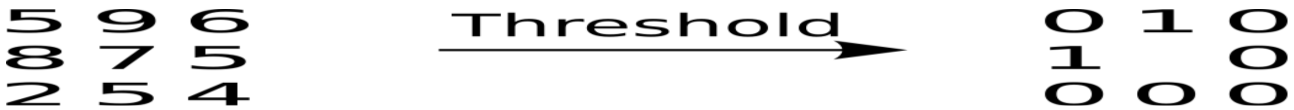


Fig. 3. Local Binary Pattern Generation

F. *Accuracy, Precision, Recall and F1 score:*

Accuracy, precision, recall and F1 score are generally used to compare the quality of classifiers. In the confusion matrix, true positives are number of persons correctly identified as the concern persons. False positives are number of persons falsely identified as the concern persons. True negatives are number of persons correctly identified as incorrect persons. False negatives are number of persons falsely identified as incorrect persons.

Accuracy is the proportion of sum of true positives and true negatives among the sum of true positives, true negatives, false positives and false negatives. It is calculated using the eq. (1).

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \qquad eq.\ (1)$$

Precision is the proportion of true positives among the sum of true positives and false positives. It is calculated using the eq. (2).

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \qquad eq.\ (2)$$

Recall is the proportion of true positives among the sum of true positives and false negatives. It is calculated using the eq. (3).

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \qquad eq.\ (3)$$

Usually, precision and recall scores are combined into a single measure, called F-measure [1], [20], [21]. The traditional F-measure or balanced F-score or F1 score is the harmonic mean of precision and recall. Two other commonly used F measures are the F_2 measure, which weights recall higher than precision, and the $F_{0.5}$ measure, which weights precision higher than recall. F1 score is calculated using the eq. (4).

$$F1\ score = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \qquad eq.\ (4)$$

IV. ANALYSIS OF RESULTS

In this work, while using neutral, smile expression and visual speech for person identification, visual speech shows comparatively higher accuracy, precision, recall and F1 score, followed by while using neutral face and performance is least while using smile expression. Similar result is recorded for person verification using auto associative neural network in [5], where the dataset used is of smaller size than used in this work. Same above method is operated on this dataset of 180 videos which also gives similar results. Fig. 4 shows equal error rate for auto associative neural network based person verification, while using neutral face as 0.37. Fig. 5 shows equal error rate for auto associative neural network based person verification, while using smile expression as 0.41. Fig. 6 shows equal error rate for auto associative neural network based person verification, while using visual speech as 0.36. Fig. 7 shows accuracy, precision, recall and F1 score for local binary pattern histogram based person identification, while using neutral face, smile expression and visual speech. In all the performance measures person identification performs well while using visual speech followed by neutral face.

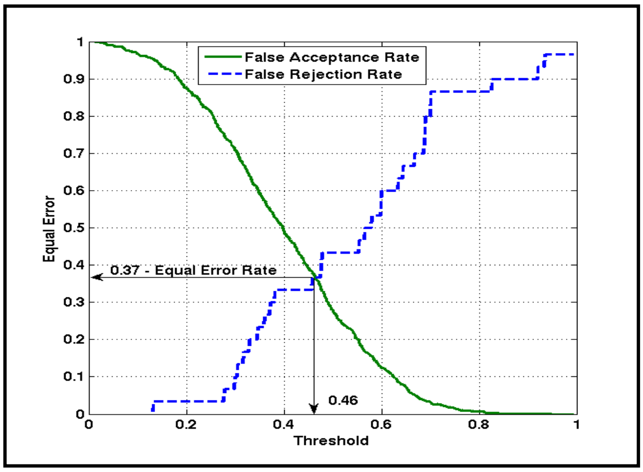


Fig. 4. Equal Error Rate while using Neutral Face

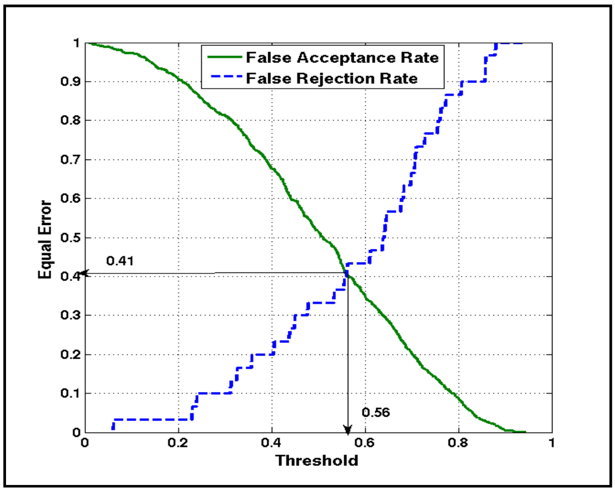


Fig. 5. Equal Error Rate while using Smile Expression

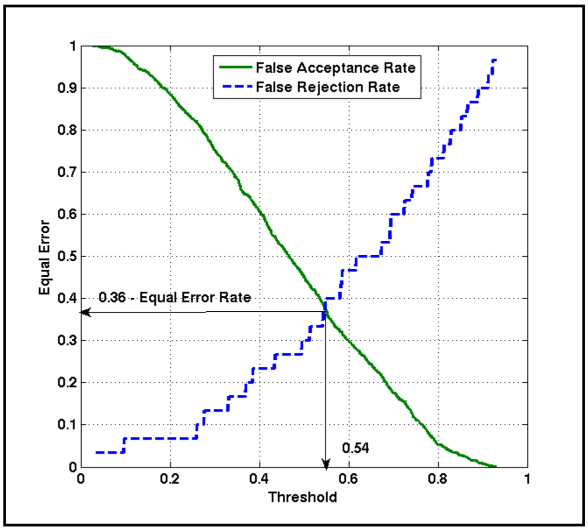


Fig. 6. Equal Error Rate while using Visual Speech

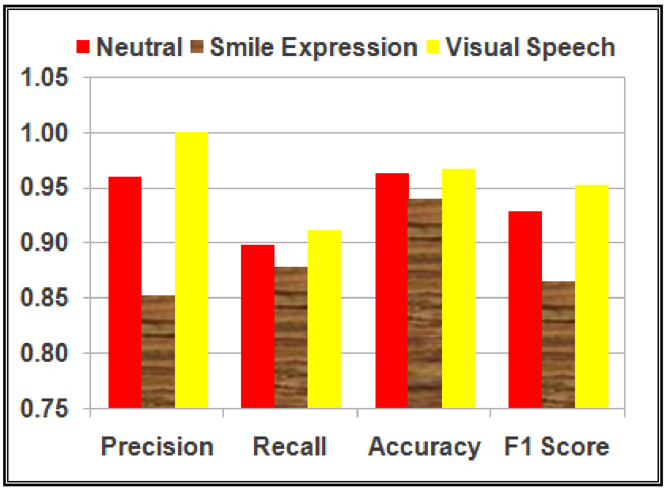


Fig. 7. Person Identification Performance Comparison

V. CONCLUSIONS AND FUTURE WORK

In this work it is concluded that for person authentication, whether it may be person verification or person identification, visual speech gives very good performance than neutral face and neutral face performs better than smile expression. It is also concluded that the mouth region component of the face itself is efficient in person authentication. Future work in this may be using of other components of face independently for person authentication. Authentication results from independent face components can be combined based on the absence of occlusion in face components.

REFERENCES

1. Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz, "Beyond Accuracy, F-score and ROC: a Family of Discriminant Measures for Performance Evaluation," in *Proceedings of the 19th Australian Joint Conference on Artificial Intelligence*, vol. 4304, pp. 1015-1021, Hobart, Australia, 4-8 December 2006.
2. M. Balasubramanian, S. Palanivel, and V. Ramalingam, "Fovea intensity comparison code for person identification and verification," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 8, pp. 1277-1290, December 2010.
3. Peiyao Li, Son Lam Phung, Abdesselam Bouzerdoum, and Fok Hing Chi Tivive, "Automatic Recognition of Smiling and Neutral Facial Expressions," in *Proceedings of the International Conference on Digital Image Computing: Techniques and Applications*, pp. 581-586, IEEE, Sydney, Australia, 1-3 December 2010.
4. Kate Saenko, Karen Livescu, Michael Siracusa, Kevin Wilson, James Glass, and Trevor Darrell, "Visual Speech Recognition with Loosely Synchronized Feature Streams," in *Proceedings of the 10th International Conference on Computer Vision*, pp. 1424-1431, IEEE, Beijing, China, 17-20 October 2005.
5. S. Saravanan, S. Palanivel, and M. Balasubramanian, "Facial Expression and Visual Speech based Person Authentication," *International Journal of Computer Applications*, vol. 100, no. 6, pp. 8-15, August 2014.
6. S. Palanivel, and B. Yegnanarayana, "*Multimodal person authentication using speech, face and visual speech*," Computer Vision and Image Understanding, vol. 109, no. 1, pp. 44-55, January 2008.
7. Shengcai Liao, and Anil K. Jain, "Partial Face Recognition: An Alignment Free Approach," in *Proceedings of the International Joint Conference on Biometrics*, pp. 1-8, IEEE, Washington, D.C., USA, 11-13 October 2011.
8. Xiang Yu, Junzhou Huang, Shaoting Zhang, Wang Yan, and Dimitris N. Metaxas, "Pose-free Facial Landmark Fitting via Optimized Part Mixtures and Cascaded Deformable Shape Model," in *Proceedings of the International Conference on Computer Vision*, pp. 1944-1951, IEEE, Sydney, Australia, 3-6 December 2013.
9. Hua Gao, HazÄ±m Kemal Ekenel, and Rainer Stiefelhagen, "Pose Normalization for Local Appearance-Based Face Recognition," in *Proceedings of the 3rd International Conference on Advances in Biometrics*, vol. 5558, pp. 32-41, Alghero, Italy, 2-5 June 2009.
10. Parama Bagchi, Debotosh Bhattacharjee, and Mita Nasipuri, "Robust 3D Face Recognition in Presence of Pose and Partial Occlusions or Missing Parts," *International Journal in Foundations of Computer Science and Technology*, vol. 4, no. 4, pp. 21-35, July 2014.
11. Andrew Wagner, John Wright, Arvind Ganesh, Zihan Zhou, Hossein Mobahi, and Yi Ma, "Towards a Practical Face Recognition System: Robust Alignment and Illumination by Sparse Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 372-386, February 2012.
12. Dong Yi, Zhen Lei, and Stan Z. Li, "Towards Pose Robust Face Recognition," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3539-3545, IEEE, Portland, Oregon, USA, 23-28 June 2013.

13. Bernd Heisele, Purdy Ho, Jane Wu, and Tomaso Poggio, "Face recognition: component-based versus global approaches," *Computer Vision and Image Understanding*, vol. 91, no. 1-2, pp. 6-21, July 2003.
14. Timo Ahonen, Abdenour Hadid, and Matti Pietikainen, "Face Recognition with Local Binary Patterns," in *Proceedings of the 8th European Conference on Computer Vision*, vol. 3021, pp. 469-481, Prague, Czech Republic, 11-14 May 2004.
15. Timo Ahonen, Abdenour Hadid, and Matti Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, December 2006.
16. Rainer Lienhart, and Jochen Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection," in *Proceedings of the International Conference on Image Processing*, vol. 1, pp. 900-903, IEEE, Rochester, New York, USA, 22-25 September 2002.
17. Paul Viola, and Michael Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp. I-511-I-518, IEEE, Kauai, Hawaii, USA, 08-14 December 2001.
18. M. Castrillon, O. Deniz, C. Guerra, M. Hernandez, "ENCARA2: Real-time detection of multiple faces at different resolutions in video streams," *Journal of Visual Communication and Image Representation*, vol. 18, no. 2, pp. 130-140, April 2007.
19. Michal Uricar and Vojtech Franc and VÃ¡clav HlavÃ¡c, "Detector of Facial Landmarks Learned by the Structured Output SVM," in *Proceedings of the 7th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Application*, vol. 1, pp. 547-556, Rome, Italy, 24-26 February 2012.
20. Minqing Hu and Bing Liu, "Mining and Summarizing Customer Reviews," in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168-177, Seattle, Washington, USA, 22-25 August 2004.
21. Marina Sokolova, Vivi Nastase, Mohak Shah and Stan Szpakowicz, "Feature Selection for Electronic Negotiation Texts," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, pp. 518-524, Borovets, Bulgaria, 21-23 September 2005.