A Multi-Biometric Recognition System Based On Deep Features of Face and Gesture Energy Image

Onur Can KURBAN, Tülay YILDIRIM
Department of Electronics and Communications
Engineering Yildiz Technical University,
Istanbul, Turkey
{ockurban, tulay}@yildiz.edu.tr

Ahmet BİLGİÇ
Tubitak Bilgem Information Technologies Institute,
Kocaeli, Turkey
ahmet.bilgic@tubitak.gov.tr

Abstract*—Nowadays, with the increasing use of biometric data, it is expected that systems work robustly and they can give successful results against difficult situations and forgery. In face recognition systems, variables such as direction of light, facial expression and reflection makes identification difficult. With biometric fusion, both safe and high performance results can be achieved. In this work, Eurocom Kinect Face dataset and BodyLogin Gesture Silhouettes dataset are used to create a virtual dataset and they were fused with score level. For face database, VGG Face deep learning model was used as feature extractor and energy imaging method was used for extracting gesture features. Afterwards the features reduced by principal component analysis and similarity scores were produced with standard deviation Euclidean distance. The results show that face recognition achieved a high performance with deep learning features under different light and expression conditions, however, multi-biometric results have reached higher genuine match rate (GMR) performance and lower false acceptance rate (FAR). As a result of this process, gesture energy imaging can be used for person recognition and for multi biometric data.

Keywords — biometrics; face recognition; multi biometric; convolutional neural network; gesture recognition; gesture energy imaging.

I. INTRODUCTION

Genuine data recognition is one of the important process in the field of biometry. Face, fingerprint, iris are the most commonly used biometric data because of their uniqueness in every individual. Unfortunately, some of these characteristics can be easily intercepted: the face is public available and fingerprint may remain on a surface unwittingly. As a result, once a biometric information is stolen, it's counterfeit can be produced and it becomes difficult to reliably prove one's identity [1]. Therefore, multi biometric methods are preferred in order to increase both safety and accuracy. Especially soft and behavioral biometric data are specific to the person, difficult to imitate and changeable according to the wishes of the person [2]. In this way, the user's information is more robust and recognition performance can be higher. In recent years, work has been increasing in this context, especially on behavioral and soft biometrics. Features such as gait [3], pose based identification [4], full body analysis [5] are used to improve security and identification.

In this study, we produced a virtual multi biometric database with face biometric and gesture biometric to perform multi biometry. For this purpose, Eurocom Face dataset [6] was used because it provides different facial expressions and lighting conditions. Convolutional Neural Network (CNN) based methods are preferred in order to get better features and also to reduce different variation effects including pose, lighting and facial expressions in images [7]. In CNN, there is a flow from simple features to complex features. Class predictions are obtained as from the CNN outputs. Filters used in convolution layers are updated to minimize error rate during training. As the filters are updated, the features obtained in the layer outputs are also changed. During training phase, this iterative improvement allows obtaining more discriminating features [8,9,10]. For the gesture data, BodyLogin gesture silhouettes dataset [11] was used. It contains frames of gesture from start to finish. Gesture energy images (GesEI) was produced with all frames.

The size of these face and gesture feature sets has been reduced by PCA. Then similarity matrices were generated by standard deviation Euclidean distance method. These matrices were combined with sum rule fusion and acceptance analysis was performed.

In this study, literature and summary are given in the first section. In the second section, dataset and preprocessing operations are summarized, methods and classifiers explains in third section. Then, in the last section, conclusions and discussions are given.

II. DATASETS AND PREPROCESSING

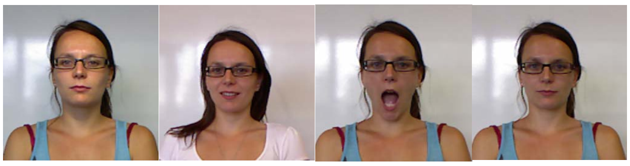
This section refers to the dataset and preprocessing operations on the dataset.

A. Datasets

In this study, the EURECOM Kinect Face Database [6] is used. The dataset consists of pictures of 52 people, 14 female and 38 male, under different poses, face expressions and lighting conditions, taken with two different session between them using the Kinect camera. Within the scope of the study, the first 40 people in the EURECOM dataset are used with the images of light on, natural, smile and open mouth as shown in Fig. 1.

For gesture analysis, BodyLogin Gesture Silhouettes dataset [5] is used. The Dataset contains two gesture types performed by 40 different users (13 women and 27 men) in the range 18-33 ages. Each subject was asked to perform two unique short gestures, each approximately three seconds long, each with 20 samples. First is drawing an "S" shape with both hands as shown in Fig. 2.

978-1-5090-5795-5/17/$31.00 ©2017 IEEE



(a) (b) (c) (d)

Fig. 1. Different facial expressions and lighting conditions a)Light on, b)Smile, c)Open mouth, d)Neutral

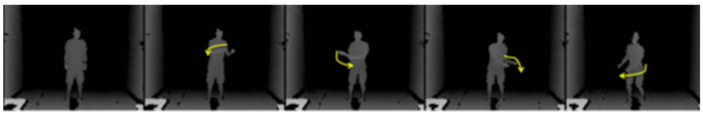


Fig. 2. Obtaining S gesture silhouette frames [5]

Although simplistic, this gesture is harder for authentication because it is similar to all users. The Second is “user defined gesture”. The user chooses his/her own gesture with no instruction. Although potentially complex, this gesture is unique for most users and easier for authentication. In this case, higher performance can be achieved with user defined gestures to identify a person. However, using different gestures can cause some difficulties for authentication. With user defined gesture in the person identification, the clusters are decided according to the movement in general. That is, system try to distinguish the type of gesture performed from other types of gestures. This will lead to an increase in the rate of false accepted rate (FAR) and increase counterfeiting. When users perform the same gesture, the recognition cluster contains differences of user behavior. So this will lead to a decrease in the FAR and it will prevent counterfeiting [1,4]. For this reason, S gesture have been chosen in this work.

B. *Preprocessing*

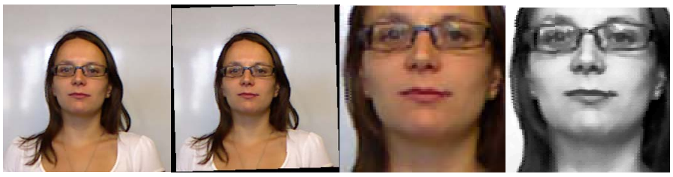
Normalization operations are required to remove variation effects in the images. While the non-biometric features on input data decrease, the performance of the recognition algorithms increases. The landmark points on the face are utilized when preprocessing operations are implemented. With Dlib C ++ [11] algorithm, position coordinate of 68 points on face is obtained. In Figure 3. (a) and Figure.3. (b), the landmark points are marked.

It is necessary to place the eyes in the same position horizontally. Thus, pose correction and scaling operation are applied. The rotation angle is calculated using eye coordinates for pose correction. The scaling factor γ is obtained using the maximum distance between eyes (1). In the database containing M pictures, for image at index k in the database, the scaling factor for the image and the inter-eye distance value D_k are used to calculate the scaling factor S_k (2). The image is scaled using S_k.

$$\gamma = \text{maks}(D_1, D_2, \ldots D_m) \qquad (1)$$

$$S_k = \frac{\gamma}{D_k} \qquad (2)$$

After pose and scaling normalization, region of interest (ROI) from each image in the database is extracted to be equal sized. The



a) b) c) d)

Fig. 3. a) Original image b) Pose correction and scaling c) Region of interest d) Histogram normalization

Fig. 3 shows the original, pose normalized and region of interest image. ROI window size is determined by considering average size of the faces in the database.

C. *Gesture Energy Imaging (GesEI)*

Each user perform S gesture approximately three seconds long and these gestures were recorded at 30 fps. Approximately 90 silhouette frames were obtained. The energy imaging process used in the gait analysis was applied to these silhouette frames.

Firstly, some preprocessing was performed on silhouette images. These are dimension normalization (ensures that all silhouettes have the same height), and horizontal alignment. In this way, equal sized images are obtained. If each binary image is defined as a two-dimensional variable Bt (x,y) bounded at time t, the energy image can be defined as follows [12].

$$GesEI(x,y) = \frac{1}{N}\sum_{t=1}^{N} B_t(x,y) \qquad (3)$$

With these two biometric data, a virtual data set for each person was randomly generated. For the multi modal system, each subject is enrolled by one face and one gesture sample. Each sample was randomly selected to increase the system's robustness. During verification, we assume that the test subject will provide one face and one gesture sample as shown in Fig. 4.

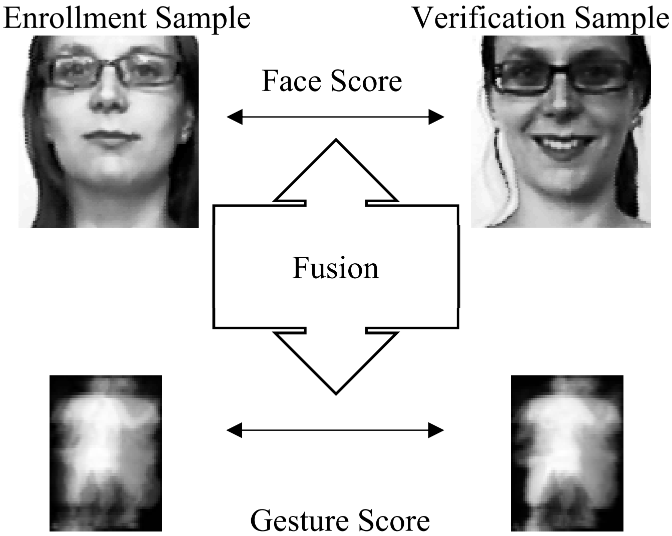


Fig. 4. Virtual multi modal dataset

III. Method And Classification

A. *Deep Learning*

Deep learning methods have begun to take part in many tasks, such as image classification, pedestrian sensing, generic visual recognition, face recognition, and large data analysis, along with the progress in processing power and graphics processors. Deep learning is a set of algorithms in machine learning that try to model high-level abstractions of data using model architectures of non-linear transformations. CNN model is one of the deep learning architects for computer vision. CNN is a special type of feed forward network, as shown in Fig. 5. The main characteristic of this method is that it can learn features at various levels, providing a more abstract representation of the input data [13].

In this study, VGG Face model [10], which is a deep learning model, is used as a feature extractor.

B. *Principle Component Analysis*

This analysis is used to determine the best transformation to express the current data with less variables. The variables obtained after the transformation are called the principle components. The first major component variance is the largest and the other major components are sorted by gradually decreasing the variance [14].

C. *Standard Deviation Euclidean Distance*

Euclidean distance is used to find differences between objects. In standard deviation Euclidean distance, each coordinate difference between rows in X and Y is scaled by dividing by the corresponding element of the standard deviation. With this method can be produced a similarity matrix (SM) that contains similarity or dissimilarity scores. High similarity values represent genuine comparisons and low similarity values represent imposter comparisons or vice versa for dissimilarity [15].

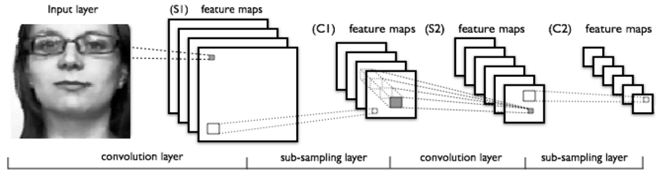


Fig.5. A sample of deep features extracting

In this study, we fused face and S gesture modalities with score level fusion technique. The face features obtained from VGG face deep learning model and S gesture features obtained from energy images were firstly applied to PCA process for features reduction. The similarity matrices was created by the standard deviation Euclidean distance method from the reduced features. SM is a square matrix and each matrix entry holds the comparison score of two biometric templates. These two similarity matrices of face and gesture fused the similarity scores using sum fusion rule. Also for this method, Z-score normalization technique was implemented. The Z-score normalization is to take the difference of the average from each variable value and divide the obtained difference by the standard deviation and can be defined as follows.

$$S^* = \frac{S - S^I_{mean}}{S^I_{SD}} \tag{4}$$

As a result of these methods, performance analyzes for both face and gesture were examined. Later, the biometric data were fused with the sum rule (No-normalization and Z-score normalization) and the performance analysis was examined. The results are explained in the last section.

IV. Results And Discussion

As a result of the applied operations are shown in Fig. 6. The graphs of FAR and FRR for EURECOM face, and Body Login S gesture. EER results are shown in Table. 1.

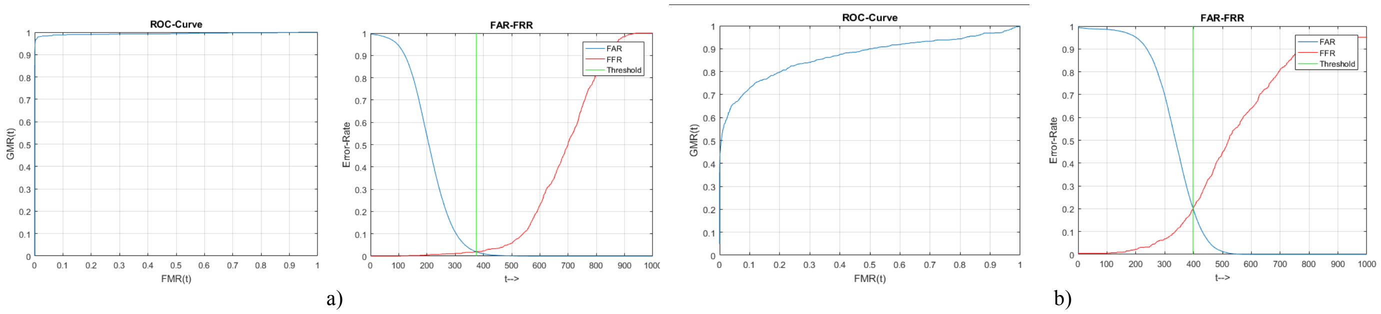
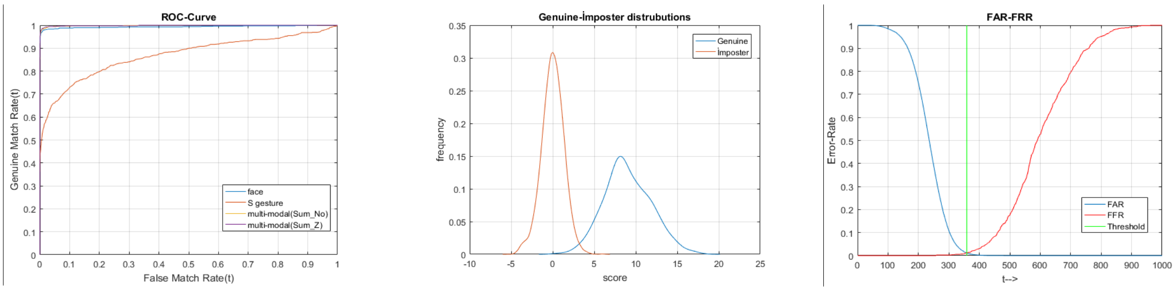


Fig. 6. a) EURECOM Face ROC curve and FAR-FRR graphs, b) BodyLogin S gestures ROC curve and FAR-FRR graphs

TABLE.1 EQUAL-ERROR RATE RESULTS

Dataset	Face	S Gesture	Multi Modal Sum Rule	Multi Modal Sum Rule-Z normalization
EER	0.0194	0.2157	0.0144	0.0128



a) b) c)

Fig. 7. Score level fusion results a) Roc curve, b) Genuine-imposter distributions, c) FAR-FRR graph

As seen in Fig. 6. (a), the features obtained from the convolutional neural network are reduced by PCA and provide a very high performance when classified with standard deviation Euclidean distance. S gesture, which is also described as a difficult gesture for authentication, energy images were created. Features of images were reduced by PCA and then classified by the same method.

The results show that the S gesture can be used for authentication with this method. These two biometrics are fused as a score level by sum rule to increase both safety and robustness. The results of multi-biometric performance, genuine-imposter distributions and FAR-FRR graph are shown in Fig. 7.

With this method, it is seen that a higher performance and more reliable system have been achieved comparing to the face recognition with the convolutional neural network. The FAR and EER values have also been reduced. Biometric systems that fused with a gesture appear to be more reliable and more suitable for aliveness analysis. The obtained FAR values show that print-attack or forgery that may occur in face recognition can be eliminated by gesture data.

In addition to this work, next studies will be carried out to improve the gesture authentication performance and analyze the different gestures. It will also be examined which gesture will provide better results for user comfort and system security.

REFERENCES

[1] J. Wu, J. Konrad, and P. Ishwar, “Dynamic timewarping for gesture-based user identification and authentication with kinect,” IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 26-31 May 2013,Canada.

[2] J.A.Unar, Woo C.S., Almas A., “A Review Of Biometric Technology Along With Trends And Prospects”, Pattern Recognition, vol.47, Issue 8, pp. 2673–2688, 2014.

[3] B. Dikovski, G. Madjarov, and D. Gjorgjevikj, “Evaluation of different feature sets for gait recognition using skeletal data from Kinect”, 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 26-30 May 2014, Croatia.

[4] A. Sinha, K. Chakravarty, “Pose Based Person Identification Using Kinect, IEEE International Conference on Systems, Man, and Cybernetics (SMC)”, 13-16 Oct. 2013, Lausanne.

[5] J. Wu, P. Ishwar, and J. Konrad, “Silhouettes versus Skeletons in Gesture-Based Authentication with Kinect”, 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 26-29 Aug. 2014, Seoul.

[6] R. Min, N. Kose and J. L. Dugelay, “Kinectfacedb: a kinect database for face recognition”, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44(11):1534-1548, 2014.

[7] G. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S.Z. Li, and T. Hospedales, “When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition”, In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp142-150, 2015.

[8] Y. Taigman, M. Yang, M. A. Ranzato and L. Wolf, , “Deepface: closing the gap to human-level performance in face verification”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701-1708, 2014.

[9] Y. Sun, X. Wang and X.Tang, “ Deep learning face representation from predicting 10,000 classes”,In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1891-1898, 2014.

[10] O. M. Parkhi, A. Vedaldi and A.Zisserman, “Deep face recognition” In BMVC, vol.1(3), pp.6, 2015.

[11] V. Kazemi, J. Sullivan, “One millisecond face alignment with an ensemble of regression trees”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1867-1874, 2014.

[12] J. Han and B. Bhanu, ”Individual recognition using gait energy image,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28(2), pp .316–322, 2006.

[13] H.A. Perlin, H.S. Lopes, “Extracting human attributes using a convolutional neural network approach”, Pattern Recognition Letters, vol. 68, pp. 250-259, 2015.

[14] W.J. Krzanowski, “Between-groups comparison of principal components”, Journal of the American Statistical Association, vol. 74(367), pp. 703-707, 1979.

[15] S. I. Audin, S. Nath, S. Basak, F. S. Rahman, R. Nath, and S. A. Fattah, “A human action recognition scheme based on spatio-temporal variation of region of interest in horizontal and vertical direction”, 23-24 May 2014, Dhaka.