Advances in Manufacturing Technology XXXV
M. Shafik and K. Case (Eds.)

doi:10.3233/ATDE220607

Cyber Security Biometric Solution Using Automatic Matching and Deep Learning

Sulaiman ALSHEBLI[1], Mahmoud SHAFIK and Fatih KURUGOLLU
University of Derby, UK.

Abstract. Biometrics is an alternative solution to the old means of identity verification, such as access cards. However, monomodal biometric systems suffer from multiple limitations, such as the noise introduced by the sensor and non-universality. Multi-biometrics allows us to overcome these problems and thus obtain better performance. Multimodal biometrics, using deep learning, has recently gained interest over single biometric modalities. In this work, we propose a deep learning model for persons identification/verification using Face and Iris traits. The features of the iris and the face are extracted utilizing the DenseNet121 and FaceNet models, and these features are merged using a feature-level fusion scheme. We also proposed a new automatic matching technique to verify the person's identity. The results presented in this paper show interest in deep learning approaches for face and iris recognition, especially when models are pre-trained. The results also highlighted the interest in the proposed DenseNet121-FaceNet model and the automatic matching method compared to the standard threshold selection according to the Equal Error Rate point.

Keywords. Multimodal biometrics, DeneseNet121, FaceNet, automatic matching, face and iris recognition.

1. Introduction

Biometrics is seen as an indispensable solution to the ongoing security, fraud, and terrorism problem and is seen by governments as an excellent security solution. Biometric features are an alternative solution to the old means of identity verification. The advantage of these biometric characteristics is that they are universal, measurable, and unique: no two people can have the same characteristic. They are also permanent, meaning they do not vary over time [1].

Multimodal biometric systems enable improved recognition performance by combining several sources of information. They also solve the problem of non-universality of certain biometrics and offer a high degree of flexibility since biometric traits that are unusable or not preferred in some individuals can be compensated for by other biometric modalities. They limit the possibilities of fraud since they provide additional protection, making it more difficult to obtain and reproduce several characteristics simultaneously.

This work aims to develop a multimodal biometric identification system based on two biometrics, namely the iris and the face. The iris is a biometric modality judged

[1] Corresponding Author. s.alshebli1@unimail.derby.ac.uk

among the most reliable, with a very rich texture and stability over time. Also, the biometric modality of the face is natural, non-intrusive, and less expensive.

Cybersecurity has encountered many challenges, especially when intruders have learned how to use advanced technologies and sensitive substances to build fake identification tools such as a fake finger, fake face, and printout of iris on the cosmetic lens to spoof the system and breach the security. Over the years, several researchers have tried to find solutions to these problems through multimodal biometrics using machine learning or deep learning.

The proposed models have been based on combining at least two characteristics. Among the works are those based on face and voice [2] face and fingerprint [3], face and palm print [4], fingerprint and iris [5], or face and iris [6].

Recently, deep learning has provided great results in multimodal biometrics systems [1], [3], [6], [7]. In addition, the limitations of classical machine learning algorithms, particularly those associated with feature extraction techniques, have been overcome by deep learning algorithms.

Omara et al. [8] have proposed a deep learning multimodal biometric system that uses the face and ears. The deep features extracted for the face and ear images are based on VGG-M Net and Discriminant Correlation Analysis (DCA). They are used for fusion and reduction of the vector size, then SVM is used for the classification.

Yang et al. [9] proposed a new multimodal biometric recognition model based on stacked methods ELMs (Stacked Extreme Learning Machines) and Canonical Correlation Analysis (CCA) methods. The model, which has a symmetrical structure, has a high potential for multimodal biometrics.

Tiong et al. [10] proposed a multi-deep learning network for the facial recognition system by combining the periocular and the facial characteristics. In addition, they further improved the recognition accuracy by combining the textural and multimodal features.

Umer et al. [11] combined the periocular and iris features for a person's biometric recognition. They deployed different deep learning-based CNN frameworks such as ResNet-50, VGG-16, and Inception-v3 for feature extraction and classification. They demonstrated that combining the features from various traits improves the system's performance.

Leghari et al., [12] proposed a multimodal system based on the feature-level fusion of fingerprint and online signature. The authors proposed using an early (before the fully connected layer) and a late (after the fully connected layer) feature fusion scheme to combine the proposed CNN architectures of the fingerprint and the online signature.

It is known that when using a machine learning approach to recognize biometrics, the biometrics images require specialized feature extraction algorithms depending on the biometric type. Sometimes the images need several pre-processing stages, while in the deep learning approach, the deep learning network extracts the features from the images automatically. Therefore, the performance of deep learning approaches is generally better than the performance of the machine learning approaches. However, while effective, these deep learning-based approaches are very computationally expensive and time-consuming [7], [13]. Studies deploying deep learning algorithms in multimodal biometric systems [14] begin experimentation by applying region detection methods before entering data into a deep learning model. The use of region detection methods requires first selecting a suitable technique for a particular trait. Also, the process can be time-consuming. It is also noticed that the physical biometric traits deliver better performance than behavioral biometric traits [7]. Moreover, the iris trait tends to increase the accuracy rate [14]. Based on what has been observed in previous studies, this work develops an identification multimodal biometric system that combines face and iris using a combination of some CNN architectures. Feature and decision level fusion were applied to determine the most effective approach. An end-to-end CNN algorithm will be used to extract features and recognize images.

2. Material and methods

This work proposes an efficient multimodal biometric system based on Deep Learning. Two approaches based on Convolutional Neural Network (CNN) using pre-training models are explored. The DenseNet121 and FaceNet architectures are used for iris and face recognition. These models were pre-trained on the ImageNet dataset [15]; fine-tuning is needed to use them with the datasets used in this work.

2.1. Deep learning models

2.1.1. Densenet

Huang et al. [16] proposed a so-called DenseNet architecture with dense connections, building a model in which all activation maps from the lower layers are transmitted to all the upper layers. The model is divided into several dense blocks to avoid the explosion of the number of parameters and activations. The dense connections' presence in each block allows the gradient to propagate immediately from the upper to the lower layers, thus applying an implicit form of deep supervision [17]. In addition, a convolutive transition layer is applied between two blocks to reduce the number of planes and is followed by max-pooling to reduce spatial dimensions. This architecture obtains comparatively better results than the ResNet based on the validation of the ILSVRC 2012. But, like ResNet, if the number of parameters of DenseNet architectures is low, these modifications are costly in the memory space required to store activations and intermediate gradients.

2.1.2. Facenet

The FaceNet model was developed by researchers at Google [18]. This model consists of a batch input layer and a deep CNN followed by L2 normalization, which results in face embedding. The authors used the Inception ResNet v1 backbone and triplet-loss during training. This model allows the direct extraction of vectors of 128 elements (rather than extracting them from an intermediate layer of a model), representing the embedding of the input images and then being used as the basis for the classification training systems. FaceNet is built on the idea of inception, using a 1×1 convolution and pooling layers in parallel to remove possible redundant parameters and make the recognition model lighter.

2.2. Datasets

Two public datasets widely used in the state-of-the-art are utilized. The first one is the IIT Delhi Iris dataset [19], which contains 2240 iris images collected from 224 subjects using JIRIS, JPC1000, digital CMOS camera and saved in bitmap (*.bmp) format with a resolution of 320x240 pixels.

The second is the VGG-Face dataset proposed by the VGG group [20]. The dataset consists of 2,622 identities for a total of 982,803 images. Each identity has an associated text file containing URLs for images and corresponding face detections.

The two previously cited datasets are grouped for the multimodal recognition experiment (face and iris), like in the first experiments. Since the iris dataset contains only 224 subjects, the same amount of data from the VGG-Face dataset is randomly selected. The irises do not belong to the people on the faces, but this combination is necessary to test the different algorithms on multimodal classification (face and iris). The final used dataset contains 224 folders representing individuals. Each folder contains two sub-folders named" Face" and" Iris," which contain the images of faces and irises, respectively.

3. Proposed method

The macroscopic cycle of a biometric identification process can be broken down into two main stages: enrollment and identification. Person enrollment is the initial phase of creating the biometric template and storing it in conjunction with a declared identity. Then, the physical characteristics (face and iris images) are transformed into a template representative of the person using feature extraction techniques. Identification consists of identifying a person using their physical characteristics within a previously registered population. Finally, depending on the distance, if it is less than a threshold, the person is identified; else, the person is unrecognized.

3.1. Feature extraction techniques

Feature extraction is a crucial and indispensable step for classification and allows one to consider only relevant elements by optimally describing the image via image-specific features. This work presents a feature-level fusion of the multimodal iris and face recognition system (see Figure 1). Two different architectures: DenseNet121 for the iris and FaceNet for the face, have been trained using transfer learning. The Total fine-tuning approach for transfer learning is adopted. All layers of each network are re-trained on the pre-processed images. As the weights are initialized with the pre-trained network's values and refined, this new learning is faster.

The two obtained models are used as feature extractors. These models will extract feature vectors from each training image and store them in a database. This database will be used later to decide whether a test image belongs to an authorized individual or not.

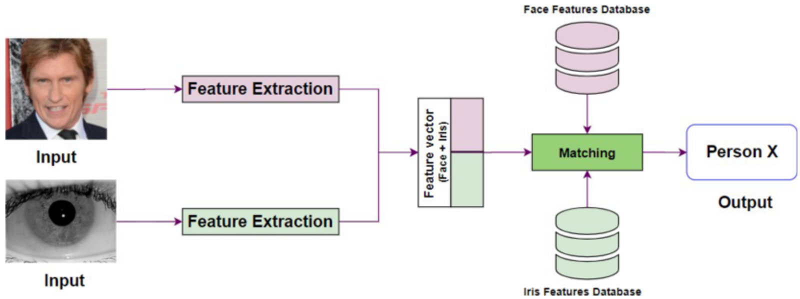

Figure 1. Face and iris feature-level fusion.

3.2. Automatic matching

To improve the system's performance (correct recognition rate) and minimize the false acceptance rate (FAR) and false rejection rate (FRR), the distance threshold must be set correctly. Generally, we chose a compromise between the FAR and FRR error, so the threshold represents the best compromise between the Accuracy, FAR, and FRR. Most of the methods proposed in state-of-the-art focus on improving the feature extraction approach and neglect the setting and improvement of the decision threshold.

In this work, we introduce a new automatic matching technique. The proposed method could be used in a unimodal or multimodal system. Its objective is to ensure the best compromise between FAR and FRR. This technique consists of a neural network model that allows people's automatic classification into two classes: recognized or unrecognized. If the person is authorized, it will be authenticated with the person's identity closest to it in the database of training features (FDB-train). Otherwise, the person is not recognized and will not be authenticated. To train the proposed model, a feature extraction on the validation dataset and on another set containing unauthorized persons is performed using the already trained convolution network (see Figure 2).

This database is named Feature Database Validation (FDB-val). After the feature extraction step, the distances between each feature vector of the FDB-val feature database and each vector of the FDB-train database is calculated. The results will be stored in a new database (distance database) which will be used to train a multilayer neural network (MLP). Each row of this database will contain the minimum distance (Dmin), the average distance (Dmean), and the maximum distance (Dmax) between each feature vector of the FDB-val database and the whole FDB-train database. The last column represents the class: 0 if the person is not authorized, and one if the person is authorized.

Let F be the feature extraction function using a deep learning model. It will have as input an image I and returns a vector that represents it. Let SM denotes a similarity measurement operation using Euclidean distance. For a given image I, the resulting distance between I and the training features database FDB-train can be expressed as a distance-vector D_I:

$$D_I = SM(F(I), FDB_{train}) \qquad (1)$$

The idea is to perform an automatic classification based on the distance. Indeed, an unauthorized person will necessarily be far from the authorized persons. Thus, the system will have the ability to learn from the data to distinguish between these two classes.

During the test phase, the new image follows the same process: feature extraction, distance calculation with the learning feature base (FDB-train), and classification with the multilayer neural network (MLP). Once the decision is made, if this person is considered authorized, the match will be found by taking the closest person on the FDB-train database.

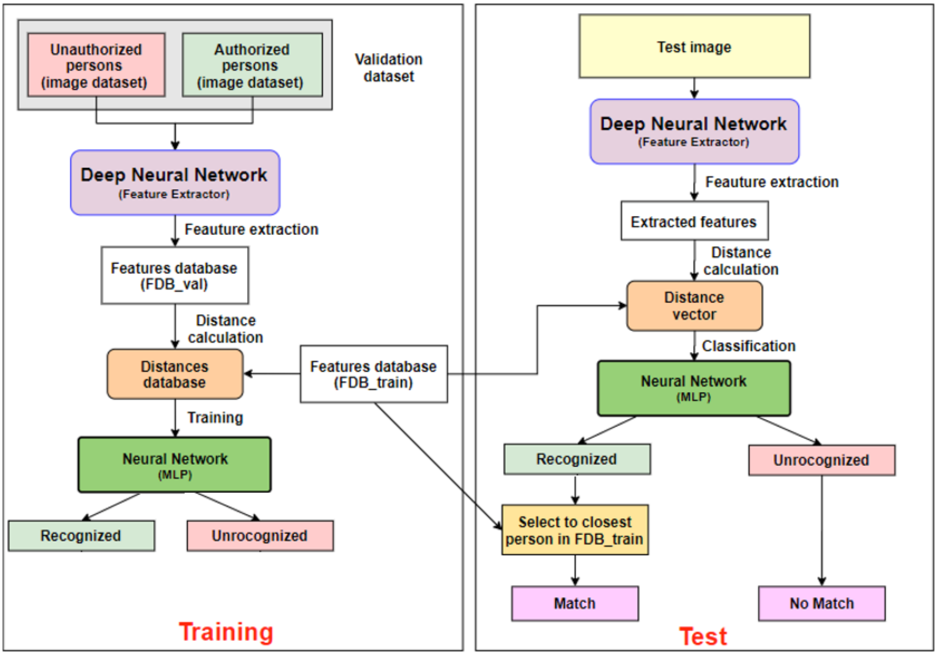

Figure 2. Automatic matching technique train and test process.

4. Results and discussion

All experiments of this work were performed on a local machine (desktop computer with an i7 4820k processor, 64GB of ram, and two GTX1070 8GB graphics cards). The face training was performed on the VGG-Face database [20], and the iris training on the IIT Delhi Iris dataset [19]. Each database is divided into three parts (60% for training, 20% for validation, and 20% for testing).

Since the iris database in our possession contains only 224 classes, and for comparison purposes, only 224 files from the VGG-Face database are used to have a multimodal database containing the iris and face images for each person.

Only 448 images per modality are kept for the test to evaluate the fusion. In addition, sixty face images and 60 iris images considered unauthorized persons are also added. These data are randomly selected from the rest of the VGG-Face database for the face images and from the MMU1 Iris dataset[2] for the iris images (see Table 1).

[2] Kaggle dataset, Multimedia University, Last Accessed (May 2022). https://www.kaggle.com/datasets/naureenmohammad/mmu-iris-dataset

Table 1. Used face and iris datasets

	# training images	# validation images	# test images
Face	10386	3410	448 + 60
Iris	1344	448	448 + 60

During the training, data augmentation, which will allow the models to be more robust and generalize better, is performed. In addition, several transformations such as shifts, flips, zooms, and more are executed.

All the models were trained during 100 epochs using batches of size 32 and the Stochastic Gradient Descent (SGD) method to reduce the loss function. In the feature-level fusion, each recognition model for each biometric feature (face and iris) performs a feature extraction independently. The fusion merges the two feature vectors generated by each recognition model. The fusion result will be considered a single feature and used to find the match in the database.

Table 2. A comparative study with techniques from the state-of-the-art

Metric	Leghari et al., [12]	Alay and Al-Baity[13]	Our work
Accuracy (%)	97,09	95,091	**99.78**
Precision (%)	97,11	94,67	**99.56**
Recall (%)	96,67	94,67	**99.33**
FRR (%)	0,45	6,25	**0.22**
FAR (%)	**0**	**0**	**0**

Table 2 summarizes the performance obtained by each technique. All methods are implemented in the same environment and using the same training and testing data. The obtained results by the DenseNet121-FaceNet model by adopting feature-level fusion and using automatic matching are very satisfactory and exceed those obtained in state-of-the-art. The proposed approach also ensured 99.78% accuracy, 99.56% precision, and 99.33% recall while keeping FAR at 0.22% and FRR at 0%. On the same test dataset, other approaches were less performed. Such performance will allow security systems to guarantee access to all authorized individuals and reject all unauthorized ones.

According to Table 2, the proposed method gives good results compared to those that used the EER threshold. This can be explained by the fact that the proposed MLP learned how to make a difference between them automatically instead of choosing a threshold variable from the validation dataset and the unrecognized dataset.

5. Conclusion

The results presented in this chapter have shown interest in deep learning approaches in face and iris recognition. In particular, the latest work on extracting iris and face features by DenseNet121 and FaceNet models and their fusion in an approach based on feature-level fusion opens exciting perspectives. Indeed, combining the strengths of feature extraction using deep architectures and automatic matching using multilayer perceptron gives promising results.

The results also highlighted the interest in the proposed automatic matching method compared to the standard threshold selection according to the EER point.

References

[1] S. Guennouni, A. Mansouri, and A. Ahaitouf. Biometric systems and their applications. *In G. L. Giudice and A. Catala, editors, Visual Impairment and Blindness*, ch. **20**. IntechOpen, Rijeka, (**2020**). https://doi.org/10.5772/intechopen.84845.

[2] A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, vol. **38**, no. 12 (**2005**), 2270–2285. https://doi.org/10.1016/j.patcog.2005.01.012.

[3] T. Shikhar, C. Bhavya, J. Rupav and S. Smriti 'Multimodal Biometric System Using Deep Learning Based on Face and Finger Vein Fusion'. *Journal of Intelligent & Fuzzy Systems*, vol. **42**, no. 2 (**2022**), 943 – 955. https://doi.org/10.3233/JIFS-189762.

[4] T. Ali Alghamdi. Evaluation of Multimodal Biometrics at Different Levels of Face and Palm Print Fusion Schemes. *Asian Journal of Applied Sciences*, vol. **9** (**2016**), 126-130. https://doi.org/10.3233/JIFS-189762.

[5] M. Elhoseny, E. Essa, A. Elkhateb, A. E. Hassanien, and A. Hamad. Cascade multimodal biometric system using fingerprint and iris patterns. *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics* (**2017**), 590– 599. https://doi.org/10.1007/978-3-319-64861-3_55.

[6] B. Ammour, L. Boubchir, T. Bouden, and M. Ramdani. Face–iris multimodal biometric identification system. *Electronics*, vol. **9** no. 1 (**2020**). 1-18. https://doi.org/10.3390/electronics9010085.

[7] S. Arora, M. P. S. Bhatia, and H. Kukreja. A multimodal biometric system for secure user identification based on deep learning. *Proceedings of Fifth International Congress on Information and Communication Technology,* (**2021**), 95–103, Singapore. https://doi.org/10.1007/978-981-15-5856-6_8.

[8] I. Omara, G. Xiao, M. Amrani, Z. Yan, and W. Zuo. Deep features for efficient multi-biometric recognition with face and ear images. *Ninth International Conference on Digital Image Processing* ICDIP vol. 10420 (**2017**), 68–73. https://doi.org/10.1117/12.2281694.

[9] J. Yang, W. Sun, N. Liu, Y. Chen, Y. Wang, and S. Han. A novel multimodal biometrics recognition model based on stacked ELM and CCA methods. *Symmetry*, vol. **10**, no. 4, (**2018**). https://doi.org/10.3390/sym10040096.

[10] L. Tiong, S. Kim, and Y. Ro. Implementation of multimodal biometric recognition via multi-feature deep learning networks and feature fusion. *Multimedia Tools and Applications*, vol. **78**, (**2019**), 22743–22772. https://doi.org/10.1007/s11042-019-7618-0.

[11] S. Umer, A. Sardar, B. C. Dhara, R. Rout, and H. Pandey. Person identification using fusion of iris and periocular deep features. *Neural Networks*, vol. **122** (**2019**), 407–419. https://doi.org/10.1016/j.neunet.

[12] M. Leghari, S. Memon, L. D. Dhomeja, A. H. Jalbani, and A. A. Chandio. Deep feature fusion of fingerprint and online signature for multimodal biometrics. *Computers*, vol. **10**, no. 2 (**2021**). https://doi.org/10.3390/computers10020021.

[13] N. Alay and H. H. Al-Baity. Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits. *Sensors*, vol. **20**, no. 19 (**2020**). https://doi.org/10.3390/s20195523.

[14] A. Al-Waisy, R. Qahwaji, and et. al. A multi-biometric iris recognition system based on a deep learning approach. *Pattern Anal Applic*, vol. **21** (**2018**), 783–802. https://doi.org/10.1007/s10044-017-0656-1.

[15] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and F.-F. Li. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, vol. **115**, no. 3 (**2015**), 211–252. http://dblp.uni-trier.de/db/journals/ijcv/ijcv115.html#RussakovskyDSKS15.

[16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. I*n 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* (**2017**), 2261–2269. https://doi.org/10.1109/CVPR.2017. 243.

[17] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu. Deeply-Supervised Nets. *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, *Proceedings of Machine Learning Research*, vol. **38** (**2015**), 562–570. http://proceedings.mlr.press/v38/lee15a.html.

[18] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* (**2015**). https://doi.org/10.1109/cvpr.2015.7298682.

[19] A. Kumar and A. Passi. Comparison and combination of iris matchers for reliable personal authentication. *Pattern Recognition*, vol. 43, no. 3 (2010), 1016–1026. https://doi.org/10.1109/cvpr.2015.7298682.

[20] Omkar M. Parkhi, Andrea Vedaldi and Andrew Zisserman. Deep Face Recognition. Proceedings of the British Machine Vision Conference (BMVC), (**2015**), 1-41.