2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS) | 978-1-7281-6921-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICECOCS50124.2020.9314403

Student's attendance management using deep facial recognition

Tarik Hachad
Laboratory of Information Modelling and Communication Systems, Faculty of sciences, Ibn Toufail University, Kenitra, Morocco
tarik.hachad@uit.ac.ma

Abdelalim Sadiq
Laboratory of Information Modelling and Communication Systems, Faculty of sciences, Ibn Toufail University, Kenitra, Morocco
a.sadiq@uit.ac.ma

Fadoua Ghanimi
Laboratory of Technological Information and Modeling Faculty of sciences Ben M'sick, University Hassan II, Casablanca, Morocco
ghanimi_fadoua@yahoo.fr

Abstract-. Managing student attendance is a repetitive and time-consuming task for both teachers and school administrators. For this reason, we thought of automating this task by deploying the recent advances of machine learning. In this article, we propose an attendance management system based on facial detection and recognition. The classroom is continuously photographed using a camera. An in-depth analysis is applied to the captured images to detect and extract the facial features of the students. Next, a pattern recognition model predicts their identities. The results of the experiment validate the proposed architecture. The process of marking the students' attendance is maintained without any human intervention.

Keywords: Face Recognition; Face Detection; Automatic Attendance; Deep Learning

I. INTRODUCTION

Facial recognition (FR) is an area of computer vision that is concerned with the analysis of the face image features in order to identify the person. It is one of the most reliable biometric techniques for identifying people. This method is used in several industrial, security, and data analysis fields [1]. The progress this technique has known in the last recent years has led us to use it in the automation of student attendance management. Manually managing attendance in classrooms with large numbers of students becomes a monotonous and time-consuming task. For this purpose, we have thought of using the FR to set up a reliable system for the automation of the attendance management by recognizing the students' faces.

The proposed system consists of a high-resolution digital camera placed at the top of the board to capture the audience in the classroom. The images captured by the camera are then sent to an in-depth analysis unit which will detect the faces present in the images and extract their features. Finally, a pattern recognition model will accurately predict the identities of the people in the image.

Our article is structured as follows. The FR existing methods especially deep learning methods and faces datasets are presented in sections II and III. In section IV, we discuss the proposed system. We validate the proposed system using experimental results in section V. Section VI ends our document with some conclusions.

II. STATE OF THE ART

The literature on FR shows a variety of approaches and methods for achieving FR with different performance levels and under varying conditions. Among the many methods proposed in the literature, we distinguish two families, those known as ancients, which require the definition and extraction of features like Local Binary Patterns (LBP)[2] , Histogram Of Oriented Gradient (HOG) [3], Scale Invariant Feature Transform (SIFT)[4] and Speeded-Up Robust Features (SURF) [5]and those based on deep learning.

In this work, we mainly focus on deep learning architectures. Unlike the old methods, Deep FR methods have the particularity of using a CNN features extractor. The learning function is obtained by the combination of several linear and non-linear operators. In 2014 Taigman et al implemented the DeepFace[6] using a deep CNN trained to classify faces for a large dataset of 4 million facial images. The network architecture is based on the assumption that once the alignment is completed, the location of each facial region is fixed at the pixel level[6]. And so, it becomes possible to learn from the raw RGB values of the pixels, without the need to apply additional layers of convolutions. It also uses a siamese network architecture [7], where the same CNN is applied to pairs of faces to obtain descriptors that are then compared using the Euclidean distance[8].Training serves to minimize the distance between the pairs of faces of the same person and to maximize the distance between the pairs of faces of two different people.

978-1-7281-6921-7/20/$31.00 ©2020 IEEE

DeepFace and DeepID, adopted a cross-entropy based SoftMax loss for feature learning [9][10]. Unfortunately, the SoftMax loss was not able by itself to learn large margin features, which forced the researchers to explore more discriminative loss functions to improve the results. However, in DeepID2 [9][10], the authors combine the face identification (SoftMax) and verification (contrastive loss) supervisory signals to learn a discriminating representation, and apply a joint Bayesian to obtain a robust embedding space. The next version DeepID2+ [10][11] increased the dimension of hidden representations and added a supervision to early convolutional layers, while DeepID3 [10][12] further introduced VGGNet and GoogleNet to their work.

Later, google implemented a deep learning architecture FaceNet [13]. It consists of convolutional layers based on GoogleNet[14] inspired inception models. FaceNet output is a 128-dimensional vector embedding that represents the most important features of the face. The FaceNet model is already trained in a triplet loss function based on LMNN[15] to capture similarities between faces. Thus, the 128-dimensional embedding returned by the FaceNet model, groups faces in classes. Consequently, the Euclidean distance between two vectors embedding of two images of the same person is much smaller than that of two different people.

III. FACES IMAGE DATASETS

Literature offers several data sets to make easy the training task of FR models. These datasets have been developed as the needs of the FR have evolved to identify more constraints. Moreover, the development of deep FR methods has created a need for large training datasets. Table 1 summarizes the most commonly used datasets.

IV. SYSTEM OVERVIEW

We used a new method to automate attendance management based on deep facial recognition. This method has given results that approximate those of the human being. The system captures the audience using a camera placed in front of the classroom. The images are then processed to detect and extract all the faces they contain. Finally, facial recognition is carried out to determine the identity of people in the images. Our system consists of two parts as shown in Figure 1. The first part is responsible for the acquisition and collection of data for enrollment and model training. The second part ensures the tasks of recognition and management of attendance. The technical details of implementing each part are discussed in the following.

The processes of recognition and management of attendance is performed by the following algorithm.

Algorithm 1 Pseudo Code of Proposed System

1. Image acquisition
2. Apply MTCNN (For Face Detection)
3. Extract the ROI in Rectangular Bounding Box
4. Convert to RGB scale, and Resize to 160x160
5. Apply FaceNet model to get face embedding / Apply Distance (Triplet Loss)
6. Classification with SVM
7. Post-processing

- **Data acquisition collection**

Since we have to generate our facial recognition model from the individual face images. The composition of a dataset containing a number of images (30 images/individual with different face poses) of each individual is necessary in order to improve the

Table 1: The most common FOR datasets FR used for training

Datasets	Publishing date	Photos	Subjects	Available
MS-Celeb-1MChallenge 3 [16]	2018	6,8M	180K	Public
VGGFace2 [17]	2017	3.31M	9,131	Public
MS-Celeb-1M (Challenge 2) [18]	2016	1.5M	20K	Public
IJB-A [19]	2015	25,809	500	Public
VGGFace[8]	2015	2.6M	2,622	Public
Facebook [6]	2014	4.4M	4K	Private
CASIA WebFace[20]	2014	494,414	10,575	Public
CelebFaces+ [9]	2014	202,599	10,177	Private
LFW [21]	2007	13K	5k	Public

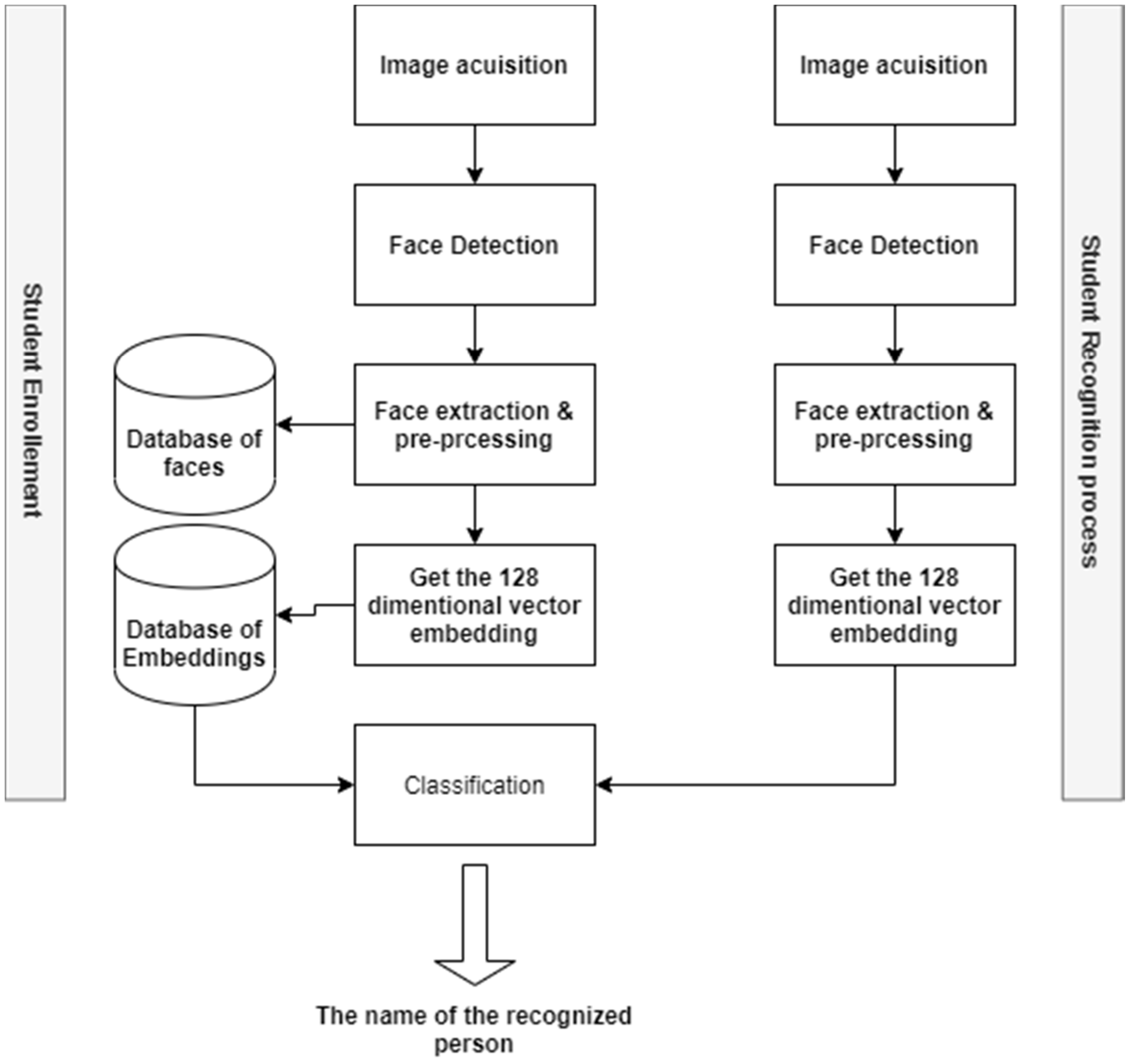

Figure 1: Attendance management system architecture

performance of our learning model. This stage consists in capturing the image of each individual and extracting the region of interest, in our case, it is the face, then a pre-treatment is carried out so that the images conform to the entries of the CNN that we used. For this, the images are converted to RGB mode, resized to a size of 160 pixels and then stored in a database. The dataset is divided into two parts:75% for training and 25% for test and validation.

- **Face Detection**

Face detection is the first step in a FR system. The choice of an appropriate face detection algorithm is crucial for the proper functioning of the FR system. Indeed, before executing FR it is necessary to detect and extract the faces. There are different methods for detecting faces such as skin color methods[22], methods based on features searching (Viola Jhones)[23], and machine learning based methods. Among all these methods, we choose to work with a deep learning network (MTCNN)[24] which offers very high detection rate under the constraint of different face poses.

- **Image pre-processing**

To meet the expectations of our CNN, the detected faces are extracted and must be pre-treated. This preprocessing phase involves resizing the face image to a square input faces with the shape 160×160 [13].

- **Features extraction and classification**

FaceNet is a deep neural network published in 2015 by researchers from Google Schroff et al[13]. It is used to extract features from an image of a person's face. FaceNet takes as input an image of a person's face and generates 128-dimensional embedding vector which represents the most important features of a face.

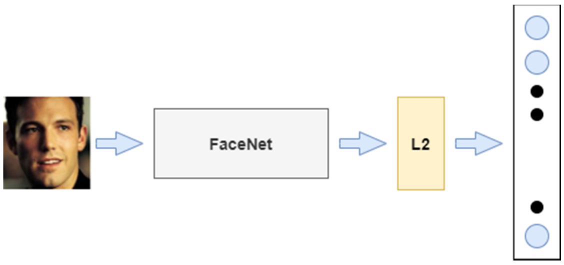

Figure 2: FaceNet model Structure.

The network consists of an input layer and a deep CNN followed by L2 normalization, which outputs a 128-dimensional face embedding vector.

- **Triplet Loss**

The embedding vector incorporates the important features of a face image. It is represented by a function $f(x) \epsilon\ \mathbb{R}^d$ that embeds an image x into a Euclidean space of dimension d. The triplet loss[15] ensures that the distance between an image x_i^a (anchor) of a specific person is minimal with all other images x_i^p(positive) of this same person and which is maximum with any image x_i^n(negative) of any other person. α is a margin applied between positive and negative pairs, the triplet loss procedure is shown in the Figure 3.

$$\left\|x_i^a - x_i^p\right\|_2^2 + \alpha < \|x_i^a - x_i^n\|_2^2$$

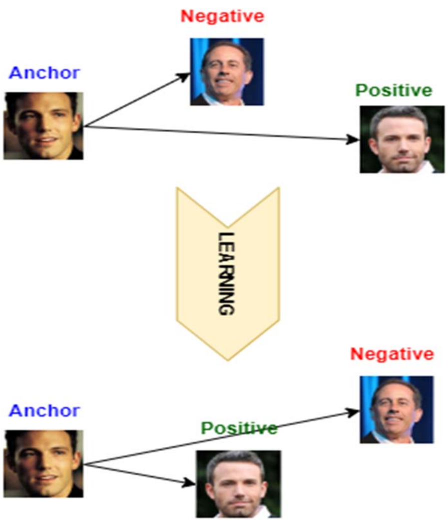

Figure 3:The Triplet Loss minimizes the distance between images of the same person, and maximizes the distance between images of two different people.

- **Classification**

Classification is the final stage of the FR, it is used to classify a person according to the embedding of his face. Since this is a pattern recognition problem, Support Vector Machine (SVM) is a logical choice for us. Indeed, SVM is very effective in separating the embedding vectors of faces[25].

V. EXPERIMENTAL RESULTS

In order to demonstrate the efficiency of our system, we carried out a test on images of faces of 7 celebrities downloaded from the Internet. To have a meaningful representation of the anchor positive distances, it needs to be ensured that a minimal number of exemplars of any one identity is present in each mini-batch[13]. We used 30 faces for each identity divided into 23 images for training and 7 for test and validation. The training and testing parts are randomly sampled. The system achieved an accuracy of 97.959% on the test set. The Figure 4 shows the performance of our classification model.

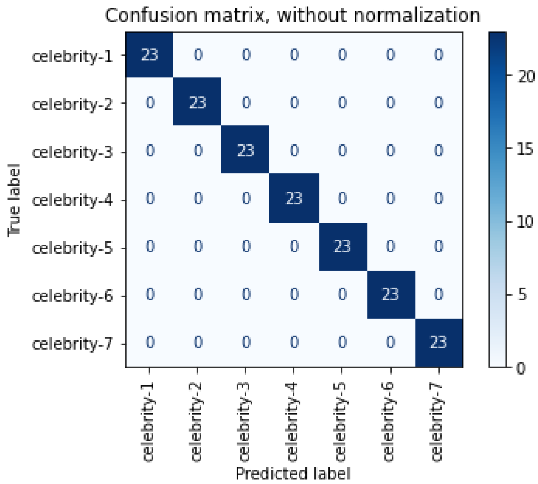

Figure 4: Confusion matrix of the face recognition system based FaceNet on the celebrities' dataset

The classification system has been trained to distinguish between faces of the same person and faces of different people. The obtained confusion matrix summarizes the results of testing the algorithm for further inspection. All correct predictions are located in the diagonal of the confusion matrix as highlighted in Figure 4 . For prediction errors, they are represented by null values outside the diagonal.

Therefore, our system is able to recognize the faces of the students present in the classroom and saves their identity name in a file with the date and time. This facilitates the task of the teachers and constitutes a log of students' attendance usable for the statistics of the institute. The FR module can also be reused in other school management systems as a face image identification tool.

VI. CONCLUSION

This paper proposes an attendance management system based on facial recognition. Face detection and recognition are performed by convolutional neural network models MTCNN and FaceNet, respectively. Based on the results, it can be concluded that the proposed architecture presents a good solution for managing the attendance of students in classrooms.

REFERENCES

[1] A. K. Jain and S. Z. Li, *Handbook of face recognition*, vol. 1. Springer, 2011.

[2] M. A. Rahim, M. S. Azam, N. Hossain, and M. R. Islam, "Face recognition using local binary patterns (LBP)," *Glob. J. Comput. Sci. Technol.*, 2013.

[3] C. Huang and J. Huang, "A fast HOG descriptor using lookup table and integral image," *arXiv Prepr. arXiv1703.06256*, 2017.

[4] Y. Feng, X. An, and X. Liu, "The application of scale invariant feature transform fused with shape model in the human face recognition," in *2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, 2016, pp. 1716–1720.

[5] R. Verma and R. Kaur, "Enhanced character recognition using surf feature and neural network technique," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, pp. 5565–5570, 2014.

[6] Y. Taigman, M. Yang, M. 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*, 2014, pp. 1701–1708.

[7] I. Melekhov, J. Kannala, and E. Rahtu, "Siamese network features for image matching," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 378–383.

[8] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," 2015.

[9] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Advances in neural information processing systems*, 2014, pp. 1988–1996.

[10] M. Wang and W. Deng, "Deep Face Recognition: {A} Survey," *CoRR*, vol. abs/1804.0, 2018.

[11] W.-S. T. WST, "Deeply learned face representations are sparse, selective, and robust," *Perception*, vol. 31, pp. 411–438, 2008.

[12] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv Prepr. arXiv1502.00873*, 2015.

[13] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[14] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[15] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, no. Feb, pp. 207–244, 2009.

[16] "Ms-celeb-1m challenge 3." [Online]. Available: http://trillionpairs.deepglint.com.

[17] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 67–74.

[18] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *European conference on computer vision*, 2016, pp. 87–102.

[19] B. F. Klare *et al.*, "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1931–1939.

[20] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv Prepr. arXiv1411.7923*, 2014.

[21] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments," in *Workshop on Faces in "Real-Life" Images: Detection, Alignment, and Recognition*, 2008.

[22] Z. Jun, H. Jizhao, T. Zhenglan, and W. Feng, "Face detection based on LBP," in *2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, 2017, pp. 421–425.

[23] K. Vikram and S. Padmavathi, “Facial parts detection using Viola Jones algorithm,” in *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2017, pp. 1–4.

[24] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks,” *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[25] S. Guo, S. Chen, and Y. Li, “Face recognition based on convolutional neural network and support vector machine,” in *2016 IEEE International Conference on Information and Automation (ICIA)*, 2016, pp. 1787–1792.