





Face Recognition Based on Harris Detector and Convolutional Neural Networks

Assaad Oussama Zeghina^(✉) , Oussama Zoubia^(b) , and Ali Behloul ^(b) 

LaSTIC Laboratory, Department of Computer Science, Batna-2 University, Batna, Algeria
assaad.o.zeghina@gmail.com, zoubia.oussama@hotmail.com,
a.behloul@univ-batna2.dz

Abstract. Facial recognition has always been a field of continuous development and research due to its usage in different areas such as security and robotics. It has gained even more popularity and interest by the researchers with the recent advancements in artificial intelligence and deep learning, which improved the robustness of facial recognition systems. In this paper, we focus on facial recognition using deep learning on small data sets with a limited number of individuals, for that we propose a local features based facial recognition approach that combines the robustness of feature extraction of CNN with the Harris corner detector. The experimental results of our proposed method surpassed the results of classical methods (LBP, Eigen Face, and Fisher Face) as well as recent works on Georgia Tech Face Database and AR Face Database and proved its efficiency and its robustness in different conditions including illumination variation, face pose variation, changes in facial expressions and face occlusions.

Keywords: Facial recognition · Deep learning · Convolutional neural networks · Harris corners detector

1 Introduction

Facial recognition is a biometric identification process that consists of identifying a person based on digital images of his face. It is widely used in various fields such as security and robotics. The research on facial recognition can be traced back to the 1960s, and it is still continuous, it gained more popularity recently due to the advancements in computer hardware and artificial intelligence. Despite the achievements made in this area, it still faces major challenges caused by the variations that the images of an individual's face could contain, such as the variation in light, the face pose variation, the variation in age, changes in facial expressions and face occlusions [1].

The main steps of a facial recognition system are the face detection, the extraction and classification of facial features. The face detection is the process of identifying the location of a face within an image. Object detection algorithms such as the Viola Jones algorithm [2] are used for face detection, and recently deep learning based methods have also been used like Faster R-CNN and YOLO [3, 4]. The extraction and the classification

of facial features step consists of extracting the important facial features, which are used to distinguish between the individuals. Three approaches for feature extraction can be used: global approaches which consider the entirety of the face images to extract global features, local approaches extract local face features, the last approach is the hybrid approach which is a combination of the two approaches [5].

Classical facial recognition methods use simple handcrafted features to describe the content of the image, machine learning algorithms are then employed for the classification. Recent years have seen a great development in deep learning which spread its uses to facial recognition. The convolutional neural networks (CNN) are the most used type of neural networks in facial recognition.

In this paper, a facial recognition method based on the detection of regions of interest and the convolutional neural network is proposed. The main focus of this method is to achieve high recognition rates on small data sets with limited number of subjects.

The paper is organized as follows: In Sect. 2, the state-of-the-art methods for facial recognition are explored, including classical methods, a brief overview of CNN and deep face recognition. In Sect. 3, our proposed deep learning based facial recognition method is presented. The description of the data sets as well as the results and discussions of the experiments are presented in Sect. 4. The last section is dedicated to the conclusion.

2 Related Work

Many researches were conducted to improve the robustness of the facial recognition methods. The classical methods were based on techniques such as edges and contours, Gabor filters [6] are an example of these methods, which have been successfully applied in many image processing tasks including face recognition. The Local Binary Pattern (LBP) [7] is another method that can be used for facial recognition, it is a powerful texture descriptor that is invariant against the change of illumination. Other variants of the LBP were proposed which achieved better recognition rates [8–10]. The Eigen face [11] and the Fisher face [12] are two other facial recognition methods which are based on dimensionality reduction algorithms, the Eigen face is based on PCA (principal component analysis) and the Fisher face is based on LDA (Linear Discriminant Analysis).

Recently, deep learning based methods particularly convolutional neural networks have gained much popularity in various fields, including facial recognition. A CNN architecture is composed of different types of layers: a convolution layer which is the core component of the CNN is used to extract the features from an image and return feature maps, it consists of a combination of convolution operations and activation functions. For each layer, the convolutions are calculated between the feature maps of the previous layer and a set of kernels whose weights are learned during the training, followed by an activation function applied on the resulting feature maps. The convolution layer is usually followed by a pooling layer. The pooling layer is used to reduce the dimensionality of the feature map and retain only the most important features. The input image is divided into a set of windows of the same size, each window is down-sampled by outputting its maximum or average value and discarding all the other values. The fully connected layers are a multilayer perceptron which takes a flattened vector from previous layers and outputs a class for the input image. The number of output nodes of the last fully

connected layer is the same as the number of classes [22]. In facial recognition, the convolution layers are used for the automatic features extraction from the face images, while the fully connected layers are used for the classification.

Deep face [13] was the first proposed CNN model for facial recognition, it was developed by Facebook AI research in 2014, mainly composed of nine layers, containing more than 120 million parameters. It achieved an accuracy of 97.35% on the LFW face data set when trained on the SFC (social face classification data set) which contains over 4.4 million images. FaceNet [14] is another CNN model that was developed in 2015, composed of 22 layers with more than 140 million parameters, trained on more than 200 million images, it achieved an impressive accuracy of 99.63% on the LFW data set. Another CNN model is the VGG-face [15] with 22 layers and more than 138 million parameters, it was trained on 2.6 million images, achieving an accuracy of 98.95% on the LFW data set. DeepID [16] is another popular CNN model for facial recognition reaching a high accuracy of 97.45% on the LFW data set when it was trained on 0.2 million images, this CNN model has more than 101 million parameters. These CNN models are complex and have high number of parameters, and they were trained on massive face data sets. A lot of work has been done on shallow CNN models with small number of parameters, which showed good results on fairly small data sets [17–19].

3 Proposed Method

Our method is composed of three main modules: the regions of interest extraction module, the convolutional neural network and finally the decision module (Fig. 1).

A points of interest detection algorithm [20] is used mainly for two reasons: firstly, because these algorithms are generally very fast and efficient. Secondly, the detected points of interest in a facial region in an image have a high probability of being detected in the same facial regions in other images of the same person (Fig. 2). For each image a maximum of 28 regions of size 32×32 pixels are extracted, the minimum distance between the centers of any two regions is set to 20 pixels.

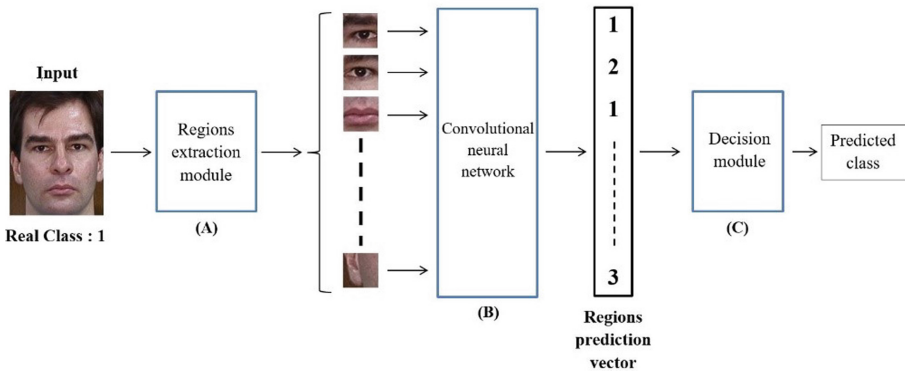


Fig. 1. The architecture of the proposed method.

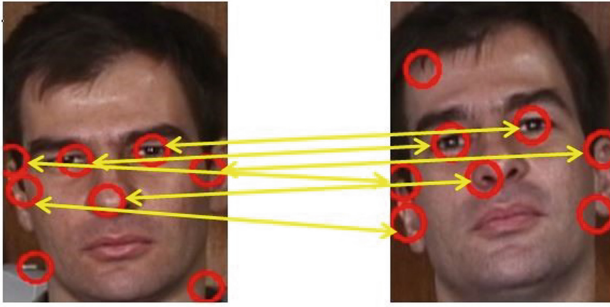


Fig. 2. Points of interest detected in a region are detected in the same regions in other images of the same person, 5 out of 8 points of interest detected in both images appear in similar regions.

The regions of interest are passed to CNN which returns a class for each region. The choice to use a CNN as a classifier is due to its efficiency in image classification tasks, particularly in facial recognition [13–19]. In case of small databases, a shallow neural network may achieve slightly higher recognition rates than a large neural network [32], for that reason we chose a shallow CNN model composed of 10 layers with 4 blocks of convolution and pooling layers followed by a fully connected layer of 512 nodes, and finally a softmax classifier. All blocs contain a batch normalization layer between the convolution and the pooling layers except for the first block. The batch normalization layers were used to prevent overfitting. We used filters of shape 3×3 for all the convolution layers, the first convolution layer employs 32 filters, where the second and the third contain 64 filters each, the last convolution layer contains 128 filters as shown in Fig. 3. The model was trained over 25 epochs using Adam optimizer and categorical cross entropy as a loss function. The number of layers and parameters as well as the number of epochs were chosen using a grid search technique.

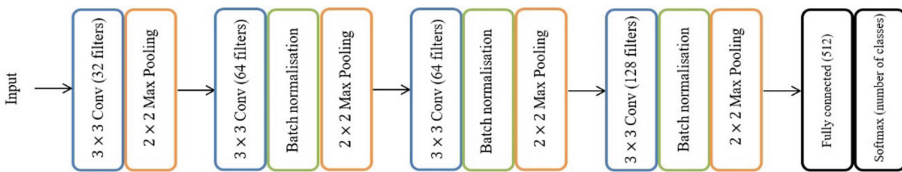


Fig. 3. The proposed CNN architecture.

The CNN outputs of each of the regions are assembled into a “regions prediction vector”, and each element of this vector corresponds to the predicted class of a region by the CNN, the decision module takes this vector and returns the class with the most occurrences, which is the predicted class of the original input image.

4 Experiments and Results

In order to evaluate the results of our proposed method, we compared its recognition rate to those obtained with LBP, Eigen face, Fisher face, a CNN model similar to the one used in our method as well as the results of recent works on Georgia Tech Face Database and AR Face Database.



Fig. 4. Sample images of two subjects (male and female) from the Georgia Tech Face Database.

The first data set is Georgia Tech Face Database¹ which contains images of 50 people taken in sessions between 06/01/99 and 15/11/99 at different times at the Georgia Institute of Technology's Image and Signal Processing Center. Each individual in the database is represented by 15 color JPEG images, and the images of each subject are taken under different conditions, such as variation in exposure, variation in brightness, different facial expressions (as shown in the Fig. 4). The average size of the faces of these images is 150×150 pixels. We used a k-fold cross validation to divide this data set into training and testing where 80% of the images are used for training, and 20% are used for testing.

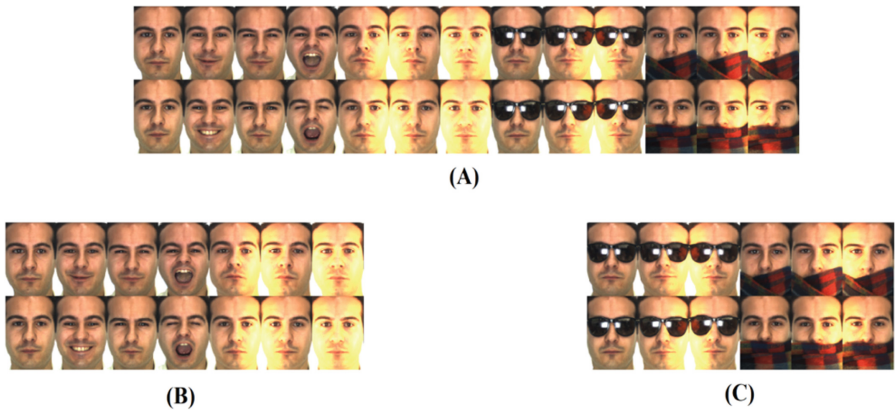


Fig. 5. (A) represents all the images of an individual (26 images), the images are divided into 2 sets, (B) which contains only the fully visible face images used for training, (C) which contains the face images with occlusions used for testing.

¹ Georgia Tech Face Database, https://www.anefian.com/research/face_reco.htm.

The second data set is AR Face Database [21] which contains more than 4.000 color images corresponding to 126 faces of people (70 men and 56 women). Images show frontal faces with different facial expressions, lighting conditions and occlusions (sunglasses and scarves). In this work, we selected 100 subjects (50 men and 50 women) where each subject has 26 images. This data set is used mainly to test the robustness of our method against occlusions. For the training subset we chose to include only the images of fully visible faces where each subject has 14 images (Fig. 5. B), the rest of the images, which are images of individuals with glasses and scarves are used for testing (Fig. 5. C).

Table 1. The average recognition rates of classical methods and our method on the Georgia Tech Face Database.

Method	LBP	Eigen face	Fisher face	CNN	Our method
Fold 1	77.33%	76.66%	86.66%	95.33%	95.33%
Fold 2	72.66%	77.33%	84.66%	95.33%	94.66%
Fold 3	74.66%	80%	87.33%	94%	99.55%
Fold 4	74.66%	87.33%	86%	95.66%	98%
Fold 5	74.66%	84%	85.33%	95.33%	99.55%
Average	74.79%	81.06%	85.99%	95.13%	97.41%

Table 2. The average recognition rates of recent works and our method on the Georgia Tech Face Database.

Method	ILRVCS [23]	sgFKNN [24]	CNN [18]	CG-RF [25]	SAPFR [26]	Our method
Recognition rate	90%	79.57%	94.8%	95.1%	95.75%	97.41%

Table 1 shows the performance of different methods, as well as our method, we can notice that the LBP achieved the lowest recognition rate compared to other methods. Eigen Face performed slightly better than LBP, surpassed by the Fisher Face. CNN achieved a better recognition rate compared to previous methods, which proves its robustness against different variations in the face images. Compared to recent works on this data set, our method surpassed them as shown in Table 2.

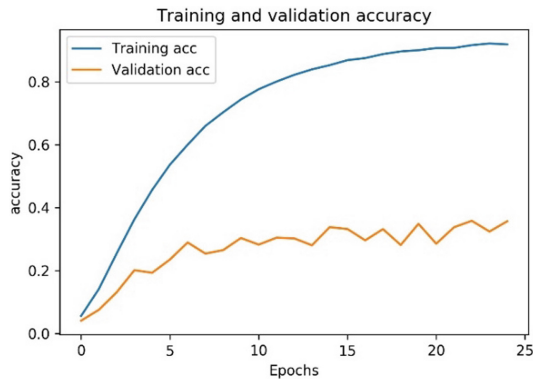
Table 3. The recognition rates of classical methods and our method on the AR Face Database.

Method	LBP	Eigen face	Fisher face	CNN	Our method
Recognition rate	24.83%	23.66%	51%	16.41%	94.72%

Table 4. The average recognition rates of recent works and our method on the AR Face Database.

Method	OA-LGBPMS [27]	SSRC1 [28]	VGG-Face [29]	DLSL [30]	SR-NMR [31]	Our method
Recognition rate	83.54%	92.99%	59.09%	89.11%	86.35%	94.72%

The AR Face Database is the database on which our proposed method shows its potential and its robustness against occlusions. Table 3 and Table 4 present the recognition rates of classical methods and recent works on this data set, which used the same approach that we used to divide the data set into training and testing as well as the recognition rate of our method. We can observe that our method surpassed both the classical methods and the recent works, which confirms its effectiveness against face occlusions.

**Fig. 6.** History of training and testing accuracy of the CNN of our proposed method on the AR Face Database.

In Fig. 6, we notice that the testing accuracy of the CNN (the classification of the regions of interest) is very low (around 30%) but the face image recognition achieves high rates. The low accuracy of the CNN is caused by the misclassification of the occluded regions of the faces, since these regions have low similarity to the regions used for training, their classification is inconsistent which reduces their effect as noise.

5 Conclusion

The integration of deep learning into facial recognition has already proven to show an improvement in the recognition rates of facial recognition systems. In this paper, we propose a facial recognition method based on shallow convolutional neural networks and Harris corner detection algorithm, which showed great performance when tested on Georgia Tech Face Database and AR Face Database, and proved its robustness against pose variation, illumination variation, change in facial expressions and particularly face

occlusion. In addition, we obtained better results than state-of-the-art methods. The proposed facial recognition approach can be useful in recognizing facial identity even with facial occlusion, and could be extended to explore larger databases.

References

1. Malikovich, K.M., Ugli, I.S.Z., O'ktamovna, D.L.: Problems in face recognition systems and their solving ways, In: International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, pp. 1–4 (2017)
2. Wang, Y.Q.: An analysis of the Viola-Jones face detection algorithm. *Image Process. On Line* **4**, 128–148 (2014)
3. Jiang, H., Learned-Miller, E.: Face detection with the faster R-CNN. In: 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, pp. 650–657 (2017)
4. Yang, W., Jiachun, Z.: Real-time face detection based on YOLO. In: 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII), Jeju, pp. 221–224 (2018)
5. Parisa Beham, M., Mohamed Mansoor Roomi, S.: A review of face recognition methods. *Int. J. Pattern Recogn. Artif. Intell.* **4**(27), 1356005 (2013)
6. Bhuiyan, A.A., Liu, C.H.: On face recognition using Gabor filters. *World Acad. Sci. Eng. Technol.* **28**, 51–56 (2007)
7. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 2037–2041 (2006)
8. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.* **19**(6), 1635–1650 (2010)
9. Hasanul Kabir, Md., Ahmed, F.: Face recognition with directional ternary pattern (DTP). In: International Conference on Graphic and Image Processing (ICGIP 2012) (2013)
10. Yang, W., Wang, Z., Zhang, B.: Face recognition using adaptive local ternary patterns method. *Neurocomputing* **213**, 183–190 (2016)
11. Slavković, M., Jevtić, D.: Face recognition using eigenface approach. *Serb. J. Electr. Eng.* **9**, 121–130 (2012)
12. Anggo, M., Arapu, L.: Face recognition using fisherface method. *J. Phys. Conf. Ser.* **1028**, 012119 (2018)
13. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: closing the gap to human-level performance in face verification. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
14. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
15. Parkhi, O.M., Vedaldi, A., Zisserman, A., et al.: Deep face recognition. In: BMVC, vol. 1, p. 6 (2015)
16. Sun, Y., Wang, X., Tang, X.: 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)
17. Yan, Y., Li, C., Lu, Y., Zhou, F., Fan, Y., Liu, M.: Design and experiment of facial expression recognition method based on LBP and CNN. In 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, pp. 602–607 (2019)
18. Coşkun, M., Uçar, A., Yildirim, Ö., Demir, Y.: Face recognition based on convolutional neural network. In: 2017 International Conference on Modern Electrical and Energy Systems (MEES), pp. 376–379 (2017)

19. Chen, J., Zhang, Z., Yao, L., Li, B., Chen, T.: Face recognition using depth images base convolutional neural network. In International Conference on Computer, Information and Telecommunication Systems (CITS), Beijing, China, pp. 1–4 (2019)
20. Shi, J., Tomasi, C.: Good features to track. In: IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA (1994)
21. Martinez, A.M., Benavente, R.: The AR Face Database. CVC Technical Report #24, June 1998
22. Yamashita, R., Nishio, M., Do, R.K.G., et al.: Convolutional neural networks: an overview and application in radiology. *Insights Imaging* **9**, 611–629 (2018)
23. Huang, W., Wang, X., Zhu, Y., Zheng, G.: Improved LRC based on combined virtual training samples for face recognition. *Int. J. Pattern Recogn. Artif. Intell.* **30**, 1656006 (2016)
24. Kasemsumran, P., Auephanwiriyakul, S., Theera-Umpon, N.: Face recognition using string grammar fuzzy K-nearest neighbor. In: 8th International Conference on Knowledge and Smart Technology (KST), Chiangmai, pp. 55–59 (2016)
25. Chark, S.Y., Noor, N.Mohd.: Integrating complete Gabor filter to the random forest classification algorithm for face recognition. *J. Eng. Sci. Technol.* **14**, 859–874 (2019)
26. Li, Z.M., Li, W.J., Wang, J.: Self-adapting patch strategies for face recognition. *Int. J. Pattern Recogn. Artif. Intell.* **34**, 2056002 (2019)
27. Rui, M., Hadid Abdenour, Dugelay Jean-Luc: Efficient detection of occlusion prior to robust face recognition. *Sci. World J.* **2014**, 519158 (2014)
28. Ou, W., You, X., Tao, D., Zhang, P., Tang, Y., Zhu, Z.: Robust face recognition via occlusion dictionary learning. *Pattern Recogn.* **47**, 1559–1572 (2014)
29. Ghazi, M.M., Ekenel, H.K.: A comprehensive analysis of deep learning based representation for face recognition. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 102–109 (2016)
30. Liao, M., Gu, X.: Face recognition based on dictionary learning and subspace learning. *Digit. Signal Process.* **90**, 110–124 (2019)
31. Chen, Z., Wu, X.J., Kittler, J.: A sparse regularized nuclear norm based matrix regression for face recognition with contiguous occlusion. *Pattern Recogn. Lett.* **125**, 494–499 (2019)
32. Peng, M., Wang, C.Y., Chen, T., Liu, G.Y.: NIRFaceNet: A convolutional neural network for near-infrared face identification. *Information* **7**, 61 (2016)