EasyChair Preprint

№ 8231

Real Time Face Detection and Identification from Video Sequences Combining LBP Algorithm and Convolutional Neural Network

Zouhair Mbarki, Besma Miladi, Chiraz Jabeur Seddik, Maryem Fadhly and Hassene Seddik

June 10, 2022

Real time face detection and identification from video sequences combining LBP algorithm and convolutional neural network

Zouhair Mbarki, Besma Miladi, Chiraz Jabeur Seddik, Maryem Fadhly and Hassene Seddik

RIFTSI Research Laboratory, ENSIT, University of Tunis, Tunisia 1008, Tunis

***Abstract*— in this work, we propose an algorithm for face detection and recognition in real time applications. The proposed algorithm combines the local binary pattern (LBP) with the convolutional neural network (CNN) and it is split into two steps: The first step is the face detecting from the video sequence and the second one is the face identification after features extraction operation using the LBP algorithm. These two steps are performed using a convolutional neural network which is a tending type of neural network based on deep learning. Extensive experiments on several test datasets are conducted to evaluate the proposed method. In fact simulation results are very interesting and show the efficiency of the suggested method.**

Keywords: face detection, face identification, real time, LBP, deep learning, convolutional neural network

I. INTRODUCTION

The field of biometrics is very interesting and complex at the same time. With biometrics, we try to distinguish between individuals, using often very advanced mathematical tools, which force us to work in a context of great diversity. This diversity is also found in the considerable number of algorithms that have been proposed in facial recognition. In fact the human face is a sophisticated multidimensional structure that can convey a lot of information about the individual, including expression, feeling and facial features [1]. Indeed, in the last few decades, face recognition becomes an extensively studied problem and the features analyzing related to facial information become a challenging task that requires lot of time and efforts [2]. As well as face identification and recognition plays an important role in several domains such as intelligent security [3], robotics manufacturing [4], clinical psychology [5], multimedia [6] and automotive security [7]. Recently and with the swift advances in processing power and memory, real-time video processing for computer vision tasks is within reach. This progression has allowed developing a new algorithm for face detection and identification, among the most prominent, the convolutional neural networks (CNN) which has greatly improved the performance of computer vision tasks [8]. Despite the good results given by the use of CNN, the change of face posture is still one challenge of face recognition system in practical application. To mitigate, this drawback, more studies are presented. In fact, Masi et al proposed an algorithm to calculate the posture distribution of the training data and establish two CNNs models which correspond to the frontal face and the profile face respectively [9]. In the same context, Liao et al suggested a partial face recognition localization method with multi-key-point descriptors to represent align-free faces in which the descriptors' size was determined by image content and face image [10]. In view of the great performance of deep learning methods in various identification and recognition tasks, this study aims to intensively investigate the use of the convolutional neural network (CNN) combined by the local binary pattern(LBP) to detect faces and recognize individuals in real time. The proposed algorithm is splitted into two parts. In this sense, the first part consists in locating the faces using a CNN network based on deep learning. This detection method, which works in real time, has provided good results which guarantee more robustness and reliability compared to other famous detection methods such as: Viola and Jones, HOG, Haar stunts. The second one, focuses on the task of identifying already localized faces by using the features extracted by the LBP algorithm followed by a deep learning CNN. In fact, this recognition method gave significant results and a fairly significant learning rate compared to other methods in the literature.

The paper is organized as follows. In the next section, we will briefly present the state-of-the-art of face detection and identification methods. In the third section, we will break down our proposed method. Afterwards, we will present our obtained results and finally we conclude.

II. ART OF STATE METHODS

A. Face detection methods

Recently, with the rapid development algorithm and technology, face detection techniques have made great progress. In fact, many CNN-based methods have been proposed in recent years. These methods achieve better performance due to powerful discriminative capability. In [11], authors proposed to localize faces, a pre-trained AlexNet with convolutional structure which adapts to different input sizes. In the same context, a CNN is designed to improve face detection accuracy [12]. The CNN proposed is used where the aim is to classify faces/non-faces and regress face bounding boxes simultaneously. Recently, many researches on effective face detection in mobile device applications have been proposed. In fact, a CNN cascade model is presented [13][14][15]. It consists of several steps and each step alone is a CNN based binary classifier that classifies an input patch as face or non-face categories.

B. Face identification methods

Many techniques based on CNN models have been proposed in recent years. In fact, Taigman et al proposed a Deep face recognition method [16]. The proposed method can achieve higher identification effect. On the other hand, in [17], a deep residual network is proposed by He et al. The proposed model, in addition to being solve the accuracy degradation problem, it can enhance the image feature expression ability. In other work [18], authors proposed a feature recalibration method. The proposed technique can

determine automatically by learning the importance of each feature, then enhance useful features and remove unimportant ones. In addition, some researchers suggested a face identification model with the basis of Center Loss function. The proposed algorithm had excellent generalization ability by increasing the inter-class distance by bringing the sample closer to each category [19]. In the same context, Liu et al in [20] proposed an angular softmax loss function. The aim of the proposed method is to maximize the intra-class distance and to minimize the inter-class distance. In other work, cosine additive angular margin algorithm is introduced [21]. The suggested technique can facilitate the training procedure and improve the face recognition rate.

III. THE PROPOSED METHOD

The proposed work consists in designing a real time face detection and identification technique (FDI). In fact, the proposed method is splitted into two steps. The first step is the face detection from video sequence and the second one is the face recognition. The face recognition is performed by comparing the face detected during the first step to a reference picture called the identity picture. Both steps are realized in real time using a convolutional neural network (CNN). The features extraction in the second step is performed using the local binary pattern (LBP). The proposed scheme is shown in figure 1.

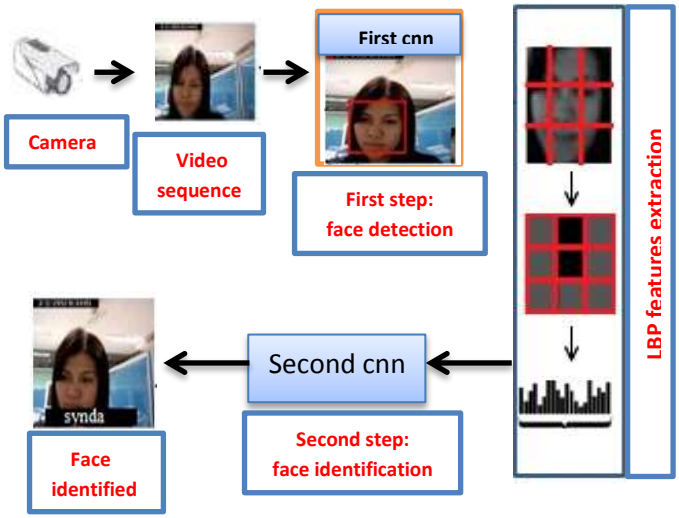

Figure 1. Diagram of the proposed method

A. *The proposed CNN*

The CNN was originally proposed by Yann LeCun. This type of network was chosen mainly since it's implicitly incorporates a feature extraction phase and it has been used successfully in many applications. CNNs are known for their robustness to low input variations and the low pretreatment rate necessary for their operation. With CNN, the learning rate reached 97% for face detection and 98% for face recognition. The architecture of a CNN is based on several deep neural networks consisting first of a succession of convolution and aggregation layers (pooling), dedicated to an automatic extraction of characteristics, while the second part, which is composed of fully connected neuron layers, is dedicated to classification.

Each cell of the convolutional layers is connected to a set of cells which are grouped in a rectangular neighborhood on the previous layer. Local receiver fields allow the extraction of basic characteristics. The layers are called "convolutional layers" because the weights are shared, and because each cell in the layer performs the same linear combination (before applying the sigmoid function) which can be seen as a simple convolution. Then, these characteristics are combined with the next layer in order to detect higher level characteristics.

The convolution layer C^i (network layer i) is parameterized by its number N of convolution maps M_j^i (j ∈ {1, N}), the size of the convolution kernels $K_x \times K_y$ (often square) and the connection diagram to the previous layer L^{i-1}. In fact, each convolution map M_j^i is the result of a sum of convolution of the maps of the previous layer M_j^{i-1} by its respective convolution kernel. In addition a bias b_j^i is then added and the result is passed to a non-linear transfer function Φ (x) as following.

$$\varphi(x) = 1.7159 \tanh(2/3\,x) \quad (1)$$

In the case of a map completely connected to the previous layer maps, the result is then calculated by:

$$M_j^i = \Phi\left(b_j^i + \sum_{n=1}^{N} M_n^{i-1} * K_n^i\right) \quad (2)$$

Between each two phases of feature extraction, the network reduces the resolution of the feature map with a means of subsampling using the max-pooling operation. In fact, the output of the max-pooling layer is given by the maximum activation value within the input layer for different non-overlapping $K_x \times K_y$ size regions. The max pooling layers is mathematically expressed as follows:

$$F = D_{\max}^{1,1}(P) \quad (3)$$

Where, P is the input vector, $D_{\max}^{1,1}$ is the maximum pooling of the l ×l block size and F is the output pooling face.

The reduction essentially leads to two goals: to decrease the size of the layer and to bring robustness compared to the weak distortions. The architecture of the convolutional neural network and the max polling operation are shown in figure 2 and figure 3.

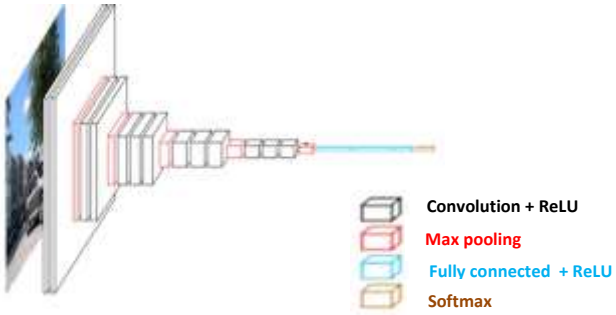

Figure 2. Architecture of a basic convolutional neural network

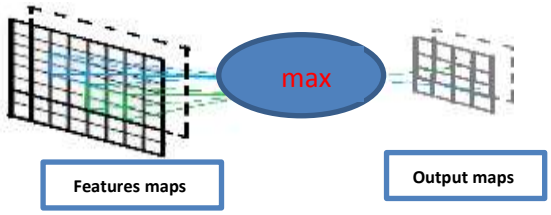

Figure 3. Max-pooling operation

The proposed CNN is composed by:

- 4 layers of Convolution each followed by a layer of Batch Normalization
- 2 layers of Max pooling
- A Dropout layer to avoid over-learning the model
- A fully connected layer for classification

In fact, the input layer is a 100 × 100 size image. The first convolution is performed by 32 filters of size 3×3 with an activation function "re-read". The result is a set of convoluted cards of size 98 × 98. Similarly, the second convolution is done by 32 filters of size 3×3 with an activation function o "re-read". The result is a set of convolutional cards of size 96×96. On the other hand the first max-pooling is a 2×2 filter with stride equal to 2. The result is a set of size 48×48 cards.

The third and fourth convolutions are performed by 64 filters of size 3x3 with an activation function of the "re-read" type. Finally, the second max pooling is a 2×2 filter with stride equal to 2 and the number of classes is equal to 2 and the activation function is the "softmax" function.

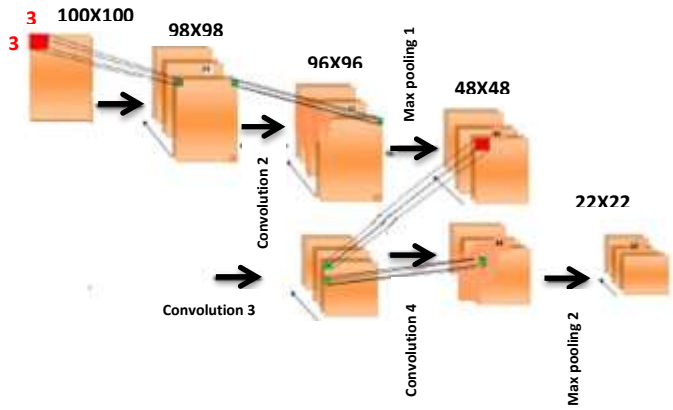

Figure 5. CNN classification phase

B. *LBP features extraction algorithm*

The LBP algorithm is presented for the first time in 1949 and has since been found to be a powerful tool for features extraction. It' is a simple efficient operator which divides a face in a dataset into M×N equally sized cells. Then, for each of these cells a local binary pattern histogram is computed. The histogram of the labeled image fl(x, y) can be defined as:

$$H_i = \sum_{x,y} I\{f_l(x,y) = i\}, i = 0,...,n-1 \quad (4)$$

Furthermore, using the computed histogram for each of the cells, every level of spatial information is encoded such as the eyes, nose and mouth. As well as, this spatial encoding allows weighting the resulting histograms from each of the cells differently, giving more discriminative power to more distinguishing features of the face. The spatially enhanced histogram is defined as:

$$H_{i,j} = \sum_{x,y} I\{f_l(x,y) = i\} I\{(x,y) \in R_j\}, i = 0,...,n-1, j = 0,...,m-1 \quad (5)$$

In figure 6, an example of the weighting scheme for each of the cells is shown:

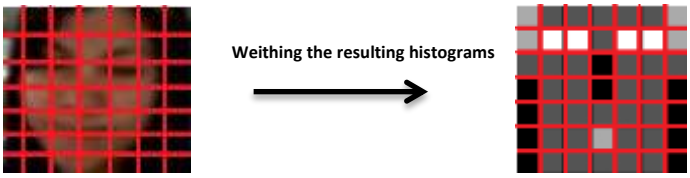

Figure 6. The original face image (left) and the weighting scheme (right)

Finally, the weighted M×N, LBP histograms are concatenated together to form the final feature vector. The complete LBP algorithm is shown in figure 7.

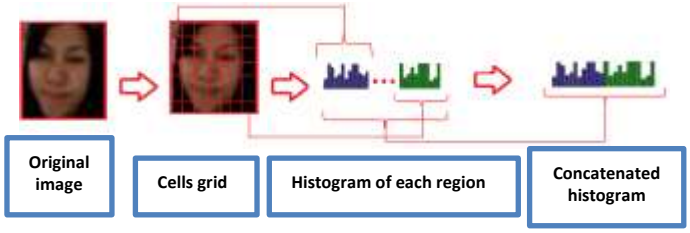

Figure 7. The LBP algorithm

C. *Database presentation*

The database used contains 2 people, where each person is represented by a set of 350 different images. Indeed, to improve performance and thus exploit the full power of the algorithm, images are acquired under different conditions:

- The positioning and orientation of the faces.
- Changing the lighting.

So, the database contains:

- 2 classes of people
- 620 images sized 100x100 grouped into 2 classes, of which,600 images will be used for the model training (300 for each person) and 20 images will be used for the test and validation (10 for each person).

IV. SIMULATION RESULTS

A. *Face detection*

To evaluate the performance of the proposed CNN in face detection, a video sequence filmed by a Huawei Y7 Smartphone with HD 1920 * 1080 resolution is used. This sequence contains 2 peoples that the system must detect them. The proposed algorithm is compared to the Multi-task Cascaded Convolutional Networks (MTCNN) and the Histograms of Oriented Gradients (HOG) algorithms.

The obtained results are given in the following figures and tables.

Figure 8. Frontal face detection

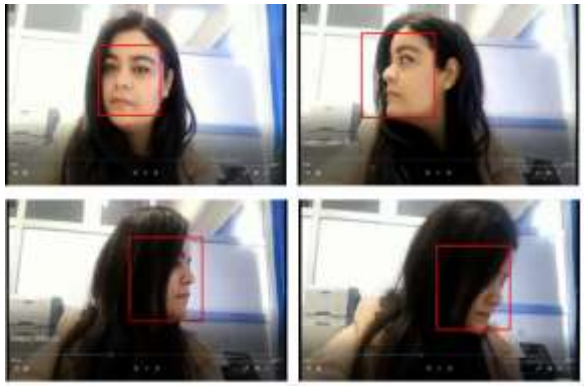

Figure 9. Face detection with change of position

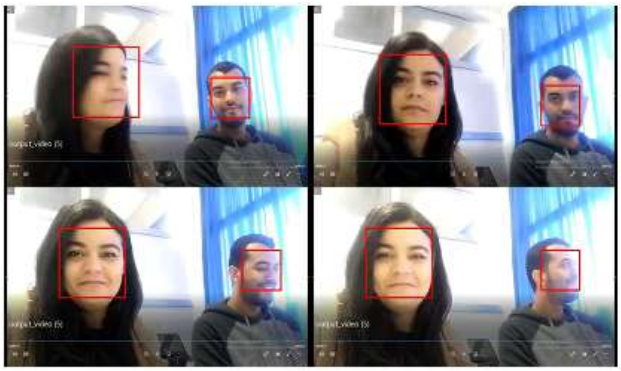

Figure 10. Simultaneous detection of 2 peoples

TABLE I. COMPARISON OF THE FACES DETECTED WITH HOG, MTCNN AND CNN DETECTION METHODS.

HOG Detection	MTCNN Detection	CNN Detection
		

TABLE II. COMPARISON OF EXECUTION TIMES FOR HOG, MTCNN AND CNN DETECTION METHODS

	Detection time with HOG	Detection time with MTCNN	Detection time with CNN
Figure 1 :	4.01	4.54	3.20
figure 2 :	4.28	11.57	2.72
figure 3 :	2.38	4.11	2.37
Figure 4 :	5.99	2.63	11.32

According to this comparison in table1 and table 2, it is found that detection with the CNN method is much more effective than the other methods according to the number of faces detected and according to the execution time. This is why our application is based on the choice of the CNN network.

A. Face Recognition

After extracting the faces already detected in our video during the first part, they must be identified during this second part. The database used composed by two persons which named Meryam Fadhly and Houssem Edd. The obtained results are given in the following images.

Figure 11. Detection and identification of person with change of position (1 person)

Figure 12. Detection and identification of person with change of position (2 persons)

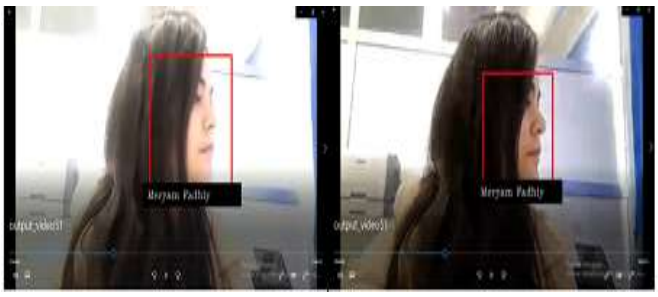

Figure 13. Detection and identification of a person with variation in brightness

The visual examination of the obtained results shows that the proposed method can detect and recognize persons exactly in different position and with variation in brightness.

TABLE III. COMPARISON OF THE LEARNING RATES TO THE EPOCH NUMBER.

Epoch number	Learning rate
50	25.66%
70	45.56%
100	66.77%
150	97%
200	97%

The obtained results from table 3 proved that the number of epochs has a big influence on the learning rate. In fact, the performance of the CNN method increases with the increase in the number of epoch so that it stabilizes when it reaches the fixed learning rate.

V. CONCLUSION

In this work we presented a rapid and efficient technique of face detection and recognition based on deep neural network. The proposed method is composed of two main parts: first, facial detection by CNN method, and then, face identification based on CNN combined by the LBP algorithm. We have thus shown the integration between the two parts to obtain a whole robust real time face detection and recognition application. The obtained results show that the proposed technique can detect and recognize people in real time and under different conditions, with 97% learning rate for face recognition and 98% for face identification. The obtained results are motivating since the processing time is low which make the implementation of the proposed algorithm on a real device is possible.

References

[1] Serign Modou Bah, Fang Ming IEEE Conference on Computer Vision and Pattern Recognition, « An improved face recognition algorithm and its application in attendance management system» Array , vol 5, (2020), p 100014.

[2] Heming Zhang a et al "Fast face detection on mobile devices by leveraging global and local facial characteristics" Signal Processing: Image communication, vol78, (2019), pp1–8.

[3] R. Wang , B. Fang , Affective computing and biometrics based HCI surveillance system, in: Proceedings of the International Symposium on Information Sci- ence and Engineering, 2008, pp. 192–195 .

[4] W. Weiguo, M. Qingmei, W. Yu, Development of the humanoid head portrait robot system with flexible face and expression, in: Proceedings of the 2004 IEEE International Conference on Robotics and Biomimetics, 2004, pp. 757–762, doi: 10.1109/ROBIO.2004.1521877 .

[5] M.H. Su , C.H. Wu , K.Y. Huang , Q.B. Hong , H.M. Wang , Exploring microscopic fluctuation of facial expression for mood disorder classification, in: Proceed- ings of the International Conference on Orange Technologies, 2017, pp. 65–69 .

[6] M.B. Mariappan, M. Suk, B. Prabhakaran, Facefetch: a user emotion driven mul- timedia content recommendation system based on facial expression recog- nition, Proceedings of the 2012 IEEE International Symposium on Multime- dia(2012) 84–87.

[7] S.A. Patil , P.J. Deore , Local binary pattern based face recognition system for automotive security, in: Proceedings of the International Conference on Signal Processing, Computing and Control, 2016, pp. 13–17 .

[8] Fenggao Tang et al, « An end-to-end face recognition method with alignment learning» Optik - International Journal for Light and Electron Optics ,vol 205, (2020), p 164238.

[9] I. Masi, S. Rawls, G. Medioni," Pose-aware face recognition in the wild," Conference on Computer Vision and Pattern Recognition (2016), pp 4838–4846.
[10] S. Liao, A.K. Jain, S.Z. Li, "Partial face recognition: alignment-free approach," IEEE Trans. Pattern Anal. Mach. Intell. Vol 35 (5), (2013), pp 1193–1205.
[11] S.S. Farfade, M.J. Saberian, L.-J. Li," Multi-view face detection using deep convolutional neural networks, "in: 5th ACM on International Conference on Multimedia Retrieval, ACM, (2015), pp. 643–650.
[12] D. Wang, J. Yang, J. Deng, Q. Liu, Facehunter:" A multi-task convolutional neural network based face detector", Signal Process., Image Commun. Vol 47 (2016) pp 476–481.
[13] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks", IEEE Signal Process. Lett. Vol 23 (10), (2016), pp 1499–1503.
[14] H. Li, Z. Lin et al " A convolutional neural network cascade for face detection," in: IEEE Conference on Computer Vision and Pattern Recognition,(2015), pp. 5325–5334.
[15] J. Deng, X. Xie, "Nested shallow CNN-Cascade for face detection in the wild," in: IEEE International Conference on Automatic Face & Gesture Recognition (FG),IEEE, (2017), pp. 165–172.
[16] Y. Taigman, M. Yang, M. Ranzato, "Deepface: closing the gap to human-level performance in face verification", IEEE Conference on Computer Vision and Pattern Recognition (2014), pp 1701–1708.
[17] K. He, X. Zhang, S. Ren," Deep residual learning for image recognition", IEEE Conference on Computer Vision and Pattern Recognition (2016) pp 770–778.
[18] J. Hu, L. Shen, S. Albanie," Squeeze-and-excitation networks", IEEE Conference on Computer Vision and Pattern Recognition (2017) , pp 4324–4335.