Real-time Face Recognition based on Deep neural network methods to solve occlusion problems

Salma Annagrebah
RITM-ESTC / CED-ENSEM,
University Hassan II, Casablanca,
Morocco,
salma.annagrebah@gmail.com

Pr.Abderrahim Maizate
RITM-ESTC / CED-ENSEM,
University Hassan II, Casablanca,
Morocco,
maizate@hotmail.com

Pr.Larbi Hassouni
RITM-ESTC / CED-ENSEM,
University Hassan II, Casablanca,
Morocco,
lhassouni@hotmail.com

***Abstract*—with massive amounts of crime around the world, the government must guarantee high security by using different methods to track and identify people in public places. Facial recognition is considered as the most effective tech for analyzing and identifying individuals' videos scenes or during authentication systems, but it has many difficulties. To address this issue, this paper presents all my comprehensions and my critical synthesis about methods and algorithms applied in face recognition and detection to solve occlusion problems.**

Keywords—Facial recognition, CNN, Face detection, occlusion, Deep Learning.

I. INTRODUCTION

Nowadays, it has been very complicated to maintain our information and physical property secured.
Oftentimes compagnies systems encountered many kinds of crimes scenes as security breaches and identities falsification.
As processes are continuously changing and improving especially in deep learning techniques, it has become very easy to verify a true individual identity by using the different methods applied in biometric technology.

However, these methods requires a good quality of image, full faces , optimal lighting and for that, different projects appeared such as FaceNet that returns a 128 dimensional vector embedding for each face. It has also achieved a new record accuracy of 99.63% using a database of 2D faces captured in an uncontrolled environment.
In 2014, researchers from Facebook published an article called "DeepFace : Closing the Gap to Human-Level Performance in Face Verification" to prevent the problems caused by changing of poses.
Moreover, facial recognition systems could be tempered with occlusion obstacles with the use accessories such as sunglasses, facemask, changing the hair style and others objects that partially occlude the covert monitoring of post.

The rest of the paper is organized as follows : The related Works are introduced in section II. Then section III exposes a synthesis of different methodologies used in deep learning for solving occlusion in facial recognition. Finally, a conclusion and discussion about the different problematics encountered.

II. RELATED WORKS

There are several factors that result in difficulties of face detection and recognition. The critical one is facial occlusion such as face covered by masks, scarfs or caps.

The obstructed faces can be regarded as the face images needed for restoration by removing the occlusions on the face. The figure 1 (below) illustrate some occulted faces that can impact on face recognition.

Figure 1. Example of image showing occlusion faces

To address this issue and remove those small occulted objects. There is a recent approach that propose a new framework that identify a person over a disguised face.
To address this issue, different methods have been developed to remove small occulted objects on the faces.
There is a recent approach that uses Disguised Face Identification (DIC) Framework using the Spacial Fusion Convolutional Network.The DIC analyses the shape of the face then extract 14 different key-points that are related to different face regions. After that, the detected points are connected to form a structure then used by the proposed face identification approach to perform a classification [1].
To increase the training of deep convolution network, the related work proposed two different datasets.
A dataset is a collection of data that contain a number of images, texts, signals and can apply on its different deep learning techniques.

The CNN method employed takes as input an image and returns the pixel that correspond every key-point.
After that, the Disguised Face Identification concept create a heat-map that synthesizes each key-point separately.
Finally, the comparison is made between a disguised face inserted as an input with five non-disguised different person face.
The classification is considered accurate if τ is the minimum for the analysis between the disguised image and non-disguised image of the same person.
The correlation is calculated according to the equation below:

$$\tau = \sum_{i}^{\cdot} | \theta i - \varphi i |$$

The τ is considered as the similarity, θi represents the orientation of the key point of the disguised image, and φi stands for the corresponding angles in the non-disguised image.

978-1-7281-0003-6/19/$31.00 ©2019 IEEE

The performance of the spacial fusion network in the Disguised Face Identification framework is better for simple datasets as opposed for the complex one with complex backgrounds.
The effect is important in the eye region, while for the nose and lips region key-points is similar for both simple and complex background datasets.
Regarding multiple faces from the given image, they use a viola Jones face detector to first locate multiple faces then proposed DIC framework is used on each face to extract the net structure of each one of them. To prove their point, they analyzed the keypoint detection classification performance of two different faces in an image in a simple and a complex dataset, the resulting performances are 80% for the simple categories and 50% for the complex one. They also redone the experience with three faces in the image and the results are 76% and 43% respectively.

III. Face Recognition and DEEP LEARNING-state of the art

Facial recognition is considered as a biometric methods able for identifying or verifying a person by making a comparison and an analyze of the template representing the person's facial contours.
Before proceeding with face recognition we first must detect the face of the person and extract the facial features from the picture.

Face detection is a computer technology that determines the locations and sizes of human faces in digital images, which is a key technology in face information processing.[2]. There are many differences between face recognition and detection. For the detection concept we need to find if there is some face in the image, but in recognition, we want to knows whose face it is.

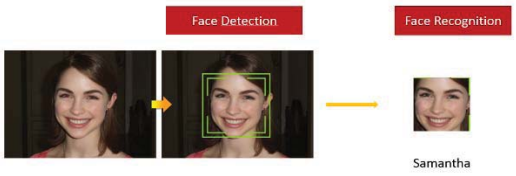

Figure.2. Detection and Recognition System

1. *Appearance-Based Face Recognition approach*

The appearance-based face recognition approach receives an important attention from several researchers such as in the biometrics domain or pattern and image detection. This approach is based on representing the global face instead of the local representation of the entire image for face identification.
One of the many algorithms adapted by face recognition in dimension reduction and features extraction is the Principal Component Analysis.
Introduced in 1901 by Karl Pearson, this method is considered as a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components[3].
By using PCA, it is not required to have a knowledge of geometry or specific feature of the face, and then the most

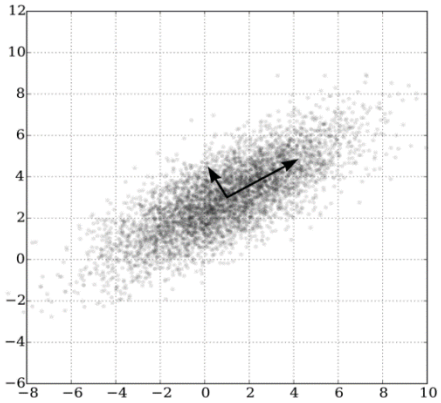

Figure3. Principal Sketch of PCA

disadvantage in this approach is that's applicable only for frontal face.
PCA is for finding patterns in data of high dimension. In addition, when we compressed data you can reduce the number of dimensions without loss of information.

2. *3D based Face recognition*

There is some kind of approach in facial recognition that uses 3D sensors to capture information about the shape of a face. It inherits advantages from traditional 2D facial recognition and allow recognizing facial even under dim light and in variant facial position.
The information collected are then used to identify distinctive points on the surface of the face, such as the contour of the eye, nose, and chin.
By using 3D-based approach, the detailed information can potentially be used for generating more robust, discriminative, and view-independent features under facial occlusion, especially for those caused by head pose changes, or missing parts.[4]
The 3D-based is considered as faster and need less memory. It can derive the facial features and make a comparison of different facial data contained in the system.
Furthermore, the 3D model can minimize the problem of head pose, illumination and keep all information's about the face.

3. *Template based approach*

A template is an illustration of the different features that compose the face. The method uses the relation between the pattern of the input image and the defined pattern of the face or its features.
Template based method is simple to implement but it's working only for frontal face without any occlusion. Also, face must be in the same size that as predefined templates.

4. *Haar-cascade features approach*

Haar-cascade features-based are attributes extracted from images used in pattern recognition. A method trains a machine learning for detecting objects and faces in a picture. A Haar-cascade needs to be train with various positive and negative pictures (The positive and negative images are those images that are include and not include human face).

978-1-7281-0003-6/19/$31.00 ©2019 IEEE

The objective is to extract the combination of these features that represents a face.

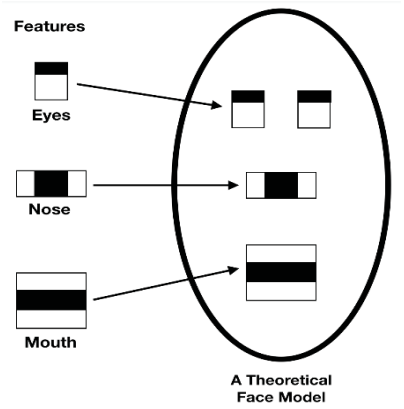

Figure.4. Features Extraction

5. *Convolutional Neural Network*

In recent year, Deep learning has benefit of learning patterns that are more abstract progressively and automatically from raw image pixels in a multiple layer architecture rather than using hand-engineered features [4].

Additional to that, there are different kinds of deep learning models which are mostly used in face recognition.

According to this paper, we are going to focus by explaining the Convolutional Neural Networks (CNN). The CNN are a type of Neural Networks that is the most used in pattern recognition and classification because of his high accuracy. Each layer takes a multi-dimensional array of numbers as input and generates another multi-dimensional array of numbers as output. The major asset of CNN compared to the others algorithms is that in doubt it detects the most imported features without any human assistance. The similar example is by giving many pictures of objects as cats or cars, the algorithm is capable for learning the different keys that compose the image and for each to return the corresponding output [16].

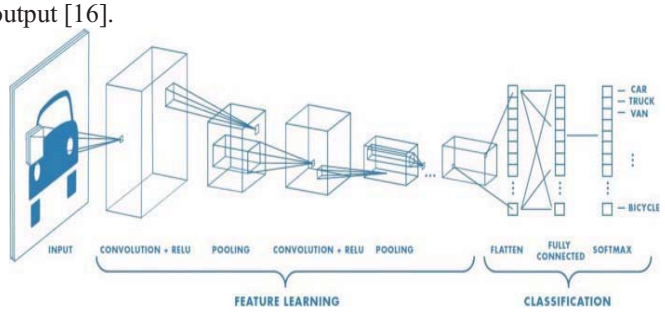

Figure.5. CNN to reduce the component

The convolutional neural network architecture is composed of multiple layers as described bellow:

- CONVOLUTION layer computes the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume.
- POOLING LAYERS reduce the dimensions of the data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. The layer average pooling uses the average value from each of a cluster of neurons at the prior layer.
- FULLY CONNECTED layers connect every neuron in one layer to every neuron in another layer. It is in principle the same as the traditional multi-layer perceptron neural network (MLP).

Receptively, in a fully connected layer, each neuron receives input from every element of the previous layer.

In a convolutional layer, neurons receive input from only a restricted subarea of the previous layer. In a convolutional layer, the receptive area is smaller than the entire previous layer.

- Weights : Each neuron in a neural network computes an output value by applying some function to the input values coming from the receptive field in the previous layer. The function that is applied to the input values is specified by a vector of weights and a bias (typically real numbers). Learning in a neural network progresses by making incremental adjustments to the biases and weights.

6. Face Database

There is several databases used to study recognition face and divided into three categories: Image processing, Natural Language Processing and Audio/Speech Processing.
In this paper, we are not going to focus on all image data processing:

- **MNIST**: The most popular dataset of handwritten digits. Its contains 60,000 images,text.
- **MUCT:** database contains 3.755 faces with 76 manual landmarks.
- Face94: contains 153 images each with a resolution of 180*200 pixel
- **ImageNet:** Is a large visual database designed for visual object recognition software research. It is a dataset of images that are organized according to the WordNet hierarchy. The WordNet database contains 100,000 phrases and ImageNet has provided around 1000 images on average to illustrate each phrases.
- **CelebFaces Attributes Dataset (CelebA)** is a large-scale face attributes dataset with more than 200K celebrity images, each with 40 attribute annotations. The images in this dataset cover large pose variations and background clutter. CelebA has large diversities, large quantities, and rich annotations, including: 10,177 number of identities.
- **Labeled Faces in the Wild (LFW)**: contains 2927 face images into Hair/Skin/Background labels.

  The dataset can be employed as the training and test sets for the following computer vision tasks: face attribute recognition, face detection, and landmark (or facial part) localization.

IV. DISCUSSION AND ANALYSIS

Not all proposed algorithms used in detection and recognize face are the best for solving occlusion problems.

978-1-7281-0003-6/19/$31.00 ©2019 IEEE

We made a comparison of all those approaches that exist and the result is described in the figure bellow.

Technique	Advantage	Disadvantage
PCA	• Reduces the dimension of the Image • Simple,Fast &Robust • Insensitive to variation in face position, facial expression.	• PCA is a less sensitive to different training data set.
EigenFace	• Easy and Fast to Implement	• Sensitive to lighting condition
CNN	• High performance • Fast at learning model	• Need a large dataset • Cost effective
Haar-like Features	• Robust in illumination changes • Allows any detection	• Sensitive to rotation and angle changes of objects
Template matching	• Simple to implement	• Frontal face shouldn't be occulted

Table2. Analysis of different methods used in face recognition

Based on precedent results [6], the convolutional neural networks achieve a very good results in image analysis with accuracy of 98.3 % and it is considered as the best techniques applied in face recognition.
To get best performance, some researches proposed two-stage of convolutional neural network. The first CNN detects the head from a person's upper body image while the second CNN distinguishes which facial part is occluded from the head image [14].
However, the Haar-cascades based feature is good in detecting face and body in images but it's not achieved good result in complex posing of the face.

V. CONCLUSION

This paper aims at reviewing the multiple approaches used to recognize persons' face in common and uncommon conditions, which is considered as one of the biggest challenges in face recognition technology.
We propose in our future researches to include a real implementation of the CNN algorithms with images to analyze from a dataset. In addition to that it would be interesting to vary the weights in CNN for more effective applying multi-task learning.
.

REFERENCES

[1] Disguised Face Identification (DFI) with Facial KeyPoints using Spatial Fusion Convolutional Network

[2] Wang, Y. Q. (2014). An analysis of the Viola-Jones face detection algorithm. Image Processing On Line, 4, 128-148

[3] Kassambara, A. (2017). Practical Guide To Principal Component Methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra (Vol. 2). STHDA.

[4] Zhang, Ligang, et al. "Facial Expression Analysis under Partial Occlusion: A Survey." *ACM Computing Surveys (CSUR)* 51.2 (2018): 25.

[5] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[6] KAMENCAY, Patrik, BENČO, Miroslav, MIŽDOŠ, Tomáš, et al. A new method for face recognition using convolutional neural network. 2017.

[7] SHONE, Nathan, NGOC, Tran Nguyen, PHAI, Vu Dinh, et al. A deep learning approach to network intrusion detection. IEEE Transactions on Emerging Topics in Computational Intelligence, 2018, vol. 2, no 1, p. 41-50.

[8] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[9] B. Dong and X. Wang, "Comparison deep learning method to traditional methods using for network intrusion detection," in *Proc. 8th IEEE Int. Conf. Commun. Softw. Netw.*, Beijing, China, Jun. 2016, pp. 581–585.

[10] R. Min and J. L. Dugelay, "Inpainting of sparse occlusion in face recognition,"in Proc. ICIP, 2011, pp. 1425–1428.

[11] Li, Yong, et al. "Patch-Gated CNN for occlusion-aware facial expression recognition." *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018.

[12] Balaban, Stephen. "Deep learning and face recognition: The state of the art." *Biometric and Surveillance Technology for Human and Activity Identification XII*. Vol. 9457. International Society for Optics and Photonics, 2015.

[13] Farfade, Sachin Sudhakar, Mohammad J. Saberian, and Li-Jia Li. "Multi-view face detection using deep convolutional neural networks." *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*. ACM, 2015.

[14] Xia, Yizhang, Bailing Zhang, and Frans Coenen. "Face occlusion detection using deep convolutional neural networks." International Journal of Pattern Recognition and Artificial Intelligence 30.09 (2016).

[15] Kamencay, Patrik, et al. "A new method for face recognition using convolutional neural network." (2017).

[16] Kamencay, Patrik, et al. "A new method for face recognition using convolutional neural network." (2017).

978-1-7281-0003-6/19/$31.00 ©2019 IEEE