The Visual Computer (2023) 39:239–254
https://doi.org/10.1007/s00371-021-02324-x

ORIGINAL ARTICLE

Deep 3D-LBP: CNN-based fusion of shape modeling and texture descriptors for accurate face recognition

Sahbi Bahroun[1] · Rahma Abed[1] · Ezzeddine Zagrouba[1]

Accepted: 1 October 2021 / Published online: 30 October 2021

Abstract
The key challenge of face recognition is to develop an effective feature representations for reducing intrapersonal variations while enlarging interpersonal differences. In this paper, we show that the face recognition accuracy may be enhanced with the combination of a 3D model-based alignment, and an LBP descriptor constructed on the 3D mesh. First, 3D face data are reconstructed from 2D images that aim to normalize the input image. Then, shape and texture features on the mesh are extracted using the mesh local binary patterns: mesh-LBP. With the use of the extracted 3D features and a simple CNN architecture, much higher accuracy rates can be achieved. We achieve the accuracy of 99.59% on the widely used labeled Faces in the Wild dataset. On YouTube Faces dataset, the proposed method achieves 94.97%, despite using a small training dataset.

Keywords 3D morphable model · Mesh-LBP · CNN · Face recognition

1 Introduction

Video cameras are extremely cheap and easily integrated into today's mobile and static devices such as surveillance cameras, police body cameras, laptops, smartphones, and Google Glass. In front of the large amount of video data, several researches have focused on techniques that enhance video indexing and retrieval of videos in large datasets based on analyzing the video structure. Furthermore, video structure analysis leads to segmenting the video into a number of structural elements that have semantic contents, including shot boundary detection [1, 2], key frame extraction [3], and scene segmentation [4].

Indeed, video-based face recognition systems aim to scan environments in a discreet manner and without feeling peoples under surveillance [5]. In other words, without having a physical contact with an acquisition device (i.e. a fingerprint, an iris scanner...). This is why face recognition from the video has become one of the major areas of interest in the field of biometric and security and rapidly overcome the image-based methods [6].

However, face images in video suffer from several issues which may could not lead to higher results in an unconstrained environment. Recently, two surveys have classified these issues into two different categories. The first survey aims to classify face recognition issues into intrinsic and extrinsic factors [7]. Intrinsic components are the physical characteristics of the human face such as aging, facial expression, and plastic surgery, whereas extrinsic factors are the ones responsible for changing the appearance of the face, such as occlusion, low resolution, noise, illumination and pose variation.

The second proposes to classify them using a face recognition covariates [8], in which, the covariate is a variable that affects the intra- or the inter-class variation. The image covariates could be classified into two categories: controlled and uncontrolled covariates. The controlled covariates include the properties of an image that can be controlled by the user or an application (i.e., pose, illumination, facial expression, occlusion, resolution). The uncontrolled covariates are the inheritable properties of a person's face, such as the effects of aging, age group, race, and gender.

✉ Rahma Abed
rahma.abed@etudiant-isi.utm.tn

Sahbi Bahroun
sahbi.bahroun@isi.utm.tn

Ezzeddine Zagrouba
ezzeddine.zagrouba@uvt.tn

[1] Laboratoire LIMTIC, Institut Supérieur d'Informatique, Université de Tunis El Manar, 2 Rue Abou Rayhane Bayrouni, 2080 Ariana, Tunisia

To overcome these issues, two alternatives are proposed. The first is to reproduce a new face image, neutral and frontal, using a set of face images taken in different views [9]. This technique is known as face frontalization [10]. Different techniques have been used such as 2D/3D texture mapping [11], statistical modeling [12], and deep learning-based methods [13–15].

The second technique is to extract robust features for the recognition. Besides, feature-based methods aim to extract a discriminate representation of faces based on one or various face feature extractor. Several feature extraction techniques are currently being used such as the scale-invariant feature transform (SIFT) [16–19], the histogram of oriented gradients (HOG) [20–23], the local binary pattern (LBP) [22, 24, 25], the local Gabor binary patterns (LGBP) [26], and the local phase quantization (LPQ) [27].

While methods based on facial image reconstruction or neutralization can be used for several tasks related to facial images analysis or with any facial recognition system, their complexity always remains a challenge given the large dataset required for training. In addition, generating neutral face images require complex models and longer learning process. Whereas, by using face features, we can combine descriptors from several face feature extractor to construct a robust descriptor.

In this paper, we focus on face recognition under uncontrolled conditions using face features. In fact, a face feature extractor provides a face description that represents well face image and could improve recognition accuracy. For this aim, we explore the possibility of combining several face features methods in order to obtain significant and robust face descriptor against facial expression, pose and illumination.

The main contribution of this paper is to fuse shape and texture information into a single face feature used to train a neural network in order to extract robust face representation for more efficient face recognition. To achieve this goal, we propose to use both 3D data and LBP feature since this descriptor is invariant to illumination variations. In addition, LBP could be used in several fields such as texture analysis, face detection and recognition and facial expression analysis [28]. In other words, we build a face recognition systems under variation in pose, illumination and expression by the use of a 3D model-based alignment, an LBP descriptor constructed on the 3D mesh to obtain a face representation that combines shape and texture LBP local histograms, and a CNN model for efficient facial recognition based on descriptor rather that an entire face image.

The rest of this paper is organized as follows. Section 2 outlines the related works in video-based face recognition. Section 3 devotes to introduce the proposed method. Section 4 discusses and analyzes the obtained experimental results. The conclusion is reported in Sect. 5.

2 Related work

Face recognition is considered as one of the most complex systems in the field of pattern recognition because of the several constraints caused by face image appearance changes. The various face recognition techniques have achieved remarkable success in well-controlled environments. However, these techniques tend to fail in real-world scenarios. So far, most of face recognition techniques have not achieved the level of accuracy that can be achieved in controlled recognition environments due to these problems [29]. Therefore, several researches have been made in order to improve the recognition accuracy either by the use of large datasets for training deep models or by enhancing the face image before proceeding to the recognition stage. According to [30], three promising techniques can be identified for the further development of this area. The three techniques are 3D face recognition methods, multimodal fusion methods and deep learning methods.

For 2D face recognition, lighting and pose variations are two major unresolved problems. Since a slight illumination variation often causes significant changes in the facial image, variation in pose can also obscure facial detail, both of these could degrade the performance of facial recognition systems [31, 32]. This is why 3D face recognition has been widely studied in order to overcome these challenges and obtain higher accuracy [33–35].

For the multimodal facial recognition, sensors have been developed with a higher ability to acquire not only two-dimensional information about texture, but also about the facial shape (i.e., three-dimensional information). Therefore, studies have merged the two types of 2D and 3D information in order to take advantage of each of them and to obtain a hybrid system that improves recognition as the only modality. In addition, these 3D face data could also be acquired by a reconstruction process from 2D face images. Contrary to the higher cost of these devices, 2D facial image acquisition mechanisms (such as surveillance cameras and webcams) are more economical and common for such applications [36].

On the other hand, the use of deep learning (DL) techniques leads to build high-level abstractions, by modeling multiple processing layers. Although several deep neural networks are used in face recognition, convolutional neural network (CNN) is the most popular [37]. Autoencoder (AE) and its variants [38] also gained much attention especially in unsupervised learning. Recently, generative adversarial networks (GAN) have been increased rapidly and are widely used for image reconstruction fields [39].

Face recognition systems usually start by detecting faces from the input image. Then, the FR system extracts a face feature vector that we could name it signature [40]. The recognition will be based on a comparison between the input face image signature and all signature in our dataset.

Although 2D face recognition research made significant progresses in recent years, its accuracy is still highly dependent on environment and human conditions. With the evolution of 3D face acquisition hardware and 3D face reconstruction techniques, a new path is emerging for face recognition that could surpass the drawbacks of 2D methods. In fact, the 3D face data provide geometric information that could improve the recognition accuracy under some conditions that are difficult to deal with using 2D technologies [41].

From another side, the use of 3D-assisted for 2D face recognition has been attracting increasing attention because it can be used for pose-invariant face matching. This requires fitting a 3D face model to the input image, and using the fitted model to align the input and reference images for matching. As 3D facial shapes are intrinsically invariant to pose and illumination, the fitted shape also provides an invariant representation that can be used directly for recognition.

Nevertheless, these methods have been replaced by deep learning methods, in particular those based on CNNs. Face recognition systems based on CNNs have become a standard because of the significant improvement in accuracy achieved compared to other methods. However, this improvement still needs to be improved in cases of crowded environments.

The main advantage of deep learning methods is their ability to be trained with large amounts of data in order to learn a robust face representation. In this manner, and rather than designing robust features, CNNs are able to learn them from training data. Despite of the availability of several large-scale face in the wild datasets into the public domain and powerful hardware equipment, these models require a long time in the learning phase [42].

In this paper, we study the techniques that use the multimodal 2D/3D. These techniques take benefit of the 3D face texture and the 2D face image descriptors (or the face feature extractor extended to deal with 3D data) in order to improve the recognition rate with low costs. We present, in the end of this section, a brief summary of the most popular deep learning-based methods.

2.1 2D/3D-based face recognition

Multimodal methods try to combine multiple processing paths (typically in 2D and 3D) into a coherent architecture to solve critical aspects of individual methods. The 3D techniques are used as an intermediate step for 2D face recognition. In other words, they perform a pre-processing step by reconstructing a new image more suitable for recognition rather than the original one.

In [43], the authors have detected salient points in 3D faces by maximum and minimum curvatures estimated in the 3D Gaussian scale space. Then, the local region around each salient point is described by three quantities: the histogram of the mesh gradient (HoG), the histogram of the shape index (HoS) and the histogram of the gradient of the shape index (HoGS). Berretti et al. [44] used the meshDOG as a detector to capture the local information of the face surface. After the keypoints detection, different local descriptors are extracted at each keypoint and are then used to compare faces during the match. Abbad et al. [45] proposed 3D face recognition based on geometric and local shape descriptors to overcome the challenges of different facial expressions. They had applied four different steps to solve the problem: First step was to model 3D face, second step was feature extraction, third step was to find out geometric information on the 3D surface in terms of curves and fourth step was to find out feature vectors on each scale. Deng et al. [46] employed features extracted from different features based on local covariance operators. Therefore, feature concatenation is a process of merging a set of features to obtain a unique and powerful operators, which contribute more characterization and useful information. Zhang et.al [47] propose a data-free method for 3D face recognition using generated data from Gaussian Process Morphable Models (GPMM).

Recently, Guosheng et al. [48] have addressed the problem of 3D-assisted 2D face recognition when the input image is subject to degradation or exhibits intrapersonal variations not captured by the 3D model. They learn a subspace spanned by perturbations caused by the missing modes of variation and image degradation, using 3D face data reconstructed from 2D images rather than 3D capture. The experiments show that this method achieves very competitive face recognition performance. Koppen et al. [49] propose a Gaussian mixture 3D morphable face model (GM-3DMM) that models the global population as a mixture of Gaussian subpopulations, each with its own mean, but shared covariance. This model is constructed using Caucasian, Chinese and African 3D face data. Finally, Liang et al. [36] use Mugshot face images for identity recognition. In fact, Mugshot face images consist of 2D frontal and pro le face images of each person. The frontal and pro le face images provide complementary information of a face and are thus believed to be useful for pose-robust face recognition.

2.2 Deep learning-based face recognition

Deep neural networks have also been applied in the past to face detection [50], face alignment [51] and face verification [52]. In the unconstrained domain, Huang et al. [53] used as input LBP features with convolutional deep belief networks and they showed improvement when combining with traditional methods. Zhenyao et al. [54] employ a deep network to “warp” faces into a canonical frontal view and then learn CNN that classifies each face as belonging to a known identity. For face verification, principal component analysis (PCA) on the network output in conjunction with an ensemble of support vector machines (SVMs) is used.

Taigman et al. [55] proposed a multistage approach that aligns faces to a general 3D shape model. A multi-class network is trained to perform the face recognition task on over four thousand identities. The authors also experimented with a Siamese network where they directly optimize the L1-distance between two face features. Schroff et al. [56] published the best performance on controlled environment. Their method, called FaceNet, learns a mapping from face images to a compact Euclidean space where distances correspond to a measure of face similarity. FaceNet uses a deep convolutional network trained to directly optimize the embedding itself, rather than an intermediate bottleneck layer as in previous deep learning approaches. Parkhi et al. [57] combined very deep convolution neural network and the triplet embedding in order to build a robust face recognition system named VGG-faces. Wu et al. [58] propose Light CNN frameworks with reduced parameters and time to learn a 256-D compact embedding on the large-scale face data with massive noisy labels. Other works aim to design powerful loss functions: Wen et al. [59] propose a center loss function. This function is used to learn a center for deep features in each class and penalized the distances between the deep features and their corresponding class centers. Yeung et al. [60] introduced a constrained triplet loss layer (CTLL) to enhance the deep model to specify further distinguishable clusters between different subjects by placing extra constraints on images of the same person while putting margins on images of different persons. However, other works focus on fusing loss functions in order to enhance face recognition systems. Fredj et. al [61] propose to learn a deep face representation from large-scale data that contain massive noisy and occluded faces. The proposed deep face model uses a fusion of softmax and center loss in order to improve the final classification.

Based on the above study of the state of the art, the following conclusions could be drawn:

- The use of face feature extraction and /or 3D methods has enhanced facial recognition rate, but this improvement remains still limited in crowded environments.
- The use of deep models tends to achieve high recognition rates but they require a long learning process and require powerful hardware and large datasets.
- The image reconstruction process is useful for recognition. However, such process requires complex models for learning (e.g., GANs).

Based on these conclusions, we propose a method based on the fusion of feature descriptors and 3D model with a neural network. Unlike method that extract several feature from the input face image [18, 19, 62–65], we use a 3DMM for attenuating the pose variation effect. After that, we propose to extract face feature from a 3D face data using the mesh-LBP. Indeed, by using the mesh-LBP, we obtain a robust descriptor against pose, illumination and facial expression variation, which is not as expensive as generating new face image from a 3D model as used in [48, 49, 66]. Nevertheless, the use of LBP on 3DMM aims to obtain a robust face representation that will prove a huge improvement in the mentioned difficulties. Then, the obtained features will be fed into a neural network that could allow to classify the faces. In our method, we use raw images as our underlying representation. We also provide a new CNN architecture through the use of the locally layer. This network will be trained on a very large labeled dataset.

3 Proposed method

The proposed method is three stage fold: First, we perform face detection and landmarks location. Then, a generic 3D face model is used to match 2D images. This is accomplished by modeling the difference in the texture map of the 3D aligned input and reference images. After that, we use the mesh-LBP [67] as a face feature extractor. Finally, the obtained descriptors are used for training a CNN architecture basically on fusing pose, texture and shape information and recognize faces. The complete flowchart of our method (named Deep 3D-LBP) is illustrated in Fig. 1.

Deep 3D-LBP uses 3D data are provided by 3D Morphable Face Models (3DMM) as an intermediate step for 2D face recognition. These 3DMM consist on a generative model of the shape and appearance of the face, essentially based on two concepts: First, all faces are in dense point-to-point correspondence. Second, the facial separation of shape and color and disentangling them from external factors such as lighting and image capturing conditions [68]. In the following subsections, we detail each step of our method.

3.1 Face and face features detection

We use the Dlib face detector [69] to detect and crop faces from the images. In fact, Dlib detects faces using histograms of oriented gradients (HoG) [70] trained with structural support vector machines (SVM)-based training algorithm. For the landmark localisation, the dlib library implements Khazemi and Sullivan's algorithm in order to detect precisely a 68-point face feature set using ensembles of regression trees [71]. Figure 2 gives an example of the Dlib face detector workflow. First, the face detection and location are shown in Fig. 2a. Then, the landmark detection occurs, which can be founded in Fig. 2b. The landmark localisation aims to detect the important facial structures from a face image. We note that the main facial areas to be labeled are mouth, right eye and eyebrow, left eye and eyebrow, nose and jaw.

The facial landmark detector used in the dlib is an implementation of the One Millisecond Face Alignment with an

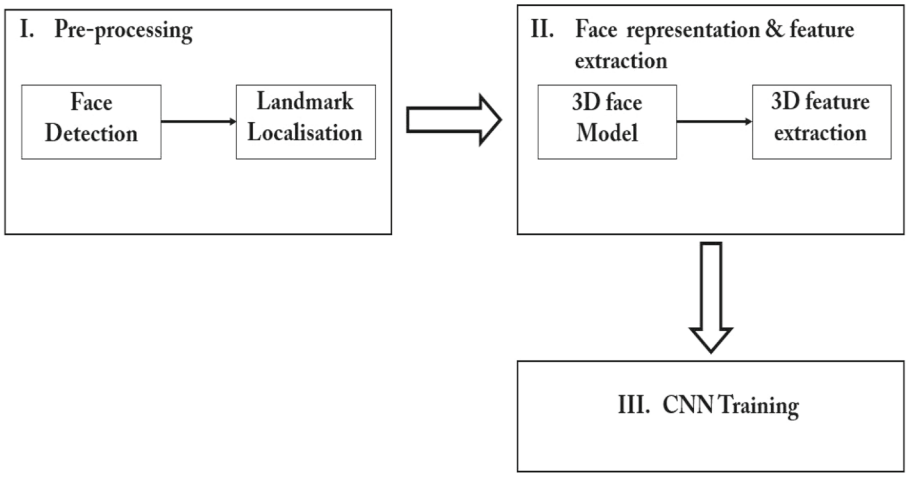
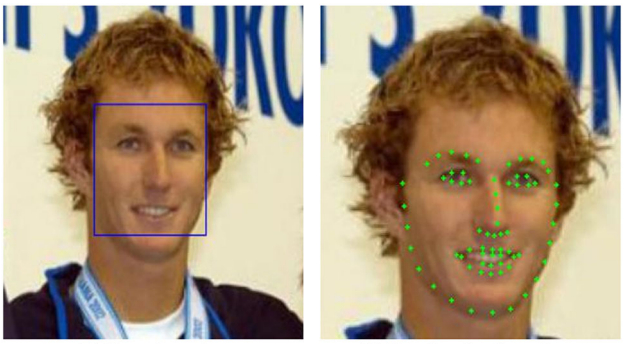


Fig. 1 Outline of the proposed face recognition method



(a) Face detection (b) Landmarks detection

Fig. 2 Output of the DLIB detector

Fig. 3 The 68 facial landmarks configuration [71]

Ensemble of Regression Trees [71]. The pre-trained facial landmark detector estimates the location of a 68 (x, y) coordinates that correspond to the facial structures, as shown in Fig. 3. From the obtained map composed of 68 landmark points, we can identify the following face features: jaw points [1–14, 16–18], right brow points [18{22], left brow points [23{27], nose points [28{36], right eye points [37{41], left eye points [44–49], mouth points [50–64, 66–70].

3.2 3D Morphable model

In this work, the 3D model used is surrey face model [72]. This 3D morphable model is a PCA shape model and a PCA color model. Each mode have a different resolution level, and accompanying metadata, like a 2D texture representation and landmark annotations. The open source library provided includes methods to the pose and the shape of a model and perform face frontalization.

The first component is pose (camera) fitting. Given a set of 2D landmark locations and their known correspondences in the 3D Morphable Model, the goal is to estimate the pose of the face (or the position of the camera, which in this case is the identical problem). It assumes an affine camera model and implements the Gold Standard Algorithm of Hartley and Zisserman [73] which finds a least squares approximation of a camera matrix given a number of 2D—3D point pairs.

The second component consists of reconstructing the 3D shape using the estimated camera matrix. It implements a simple shape-to-landmarks fitting similar to the algorithm from Aldrian and Smith [74]. The pose estimation and shape fitting steps can be iterated if desired to refine the estimates. The pose estimation can make use of the shape estimate (instead of using the mean face) in order to re ne the face pose. The shape estimate can in turn use the refined camera matrix to improve the shape fitting. The shape estimation is as fast as the pose estimation: Each of them involves only solving a small linear system of equations and runs in the order of milliseconds.

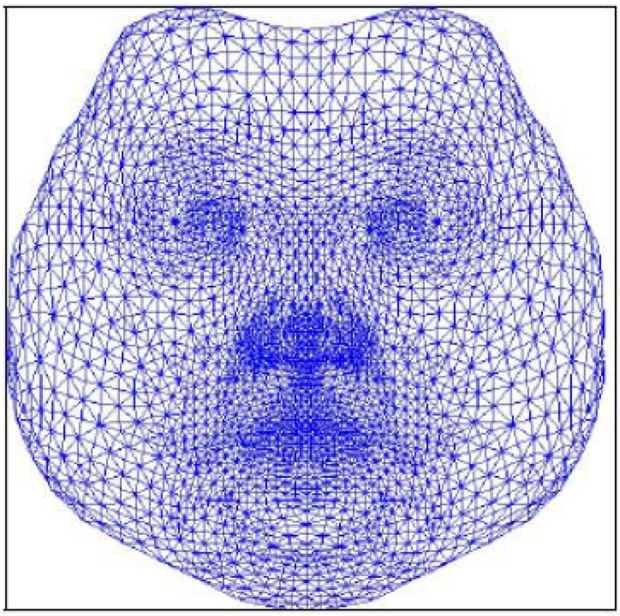


Fig. 4 Texture representation in the form of an isomap [68]

After obtaining the pose and shape coefficients, there is a dense correspondence between mesh vertices and the face in the input image. We can then remap the texture into the model, store it, and re-render it in arbitrary poses (e.g., frontalise it). The texture can be then extracted and stored in the isomap, similar to that presented in Fig. 4.

Figure 5 shows an example of landmarks fitting for the input image (Fig. 5a), the resulting shape and camera model fitting is shown in Fig. 5c. In Fig. 5b, regions of self-occlusion are depicted as white spots; however, in the isomap, they are identified by the alpha channel.

3.3 Mesh-LBP

The mesh-LBP descriptor is used to fuse the geometric and appearance features extracted from 3D face models. In the standard LBP-based face representation [63], a 2D face image is divided into a grid of rectangular blocks, then histograms of LBP descriptors are extracted from each block and concatenated afterward to form a global description of the face. Thus, image partitioning is performed easily thanks to the natural ordering of image pixels. To extend this scheme to the face manifold, we need first to partition the facial surface into a grid of regions (the counterpart of the blocks in the 2D-LBP), compute their corresponding histograms, and then group them into a single structure. Since partitioning of the 2D mesh manifold is not straightforward, we rely on the idea of extracting a grid of fiducial points of the face with predefined position and then use their neighborhood regions as local supports for computing mesh-LBP.

The mesh-LBP operator at the facet f_c is defined in Eq. 1:

$$\text{mesh } LBP_m^r(f_c) = \sum_{k=0}^{m-1} s\left(h\left(f_k^r\right) - h(f_c)\right) \times \alpha(k) \tag{1}$$

$$S(x) = \begin{cases} 1 if x \geq 0 \\ 0 if x < 0 \end{cases}$$

We define by r as the number of ring facets. The parameter m is the number of facets uniformly spaced on the ring. In fact, r and m control, respectively, the radial resolution $r = 7$ and the azimuthal quantization $m = 12$. We refer by h, the scalar function defined on the mesh, either a geometric (e.g., curvature) or photometric (e.g., color or gray level) information. The (k) is a weighting function and can be used for the purpose of deriving different LBP variants.

In this work, we will consider two variants of $\alpha(k)$: First, $\alpha(k) = 2^k$, we obtain the mesh counterpart of the basic LBP operator. Second, $\alpha(k) = 1$, to obtain the sum of the digits that is equal to 1. We will refer to these two functions by α_2 and α_1, respectively. For the discrete surface function *h(f)*, and for this work, we experimented the mean curvature (H), the curvedness *(C),* the Gaussian curvature *(K)* and the shape index *(SI),* as shape descriptors, plus the gray level value *(GL)* as photometric characteristic of the facets.

The mesh-LBP will be calculated according to the following steps:

First, the plane formed by the nose tip and the two eyes inner-corner landmark points is initially computed. We used these three landmarks as they are the most accurately detectable landmarks on the face, and they are also quite robust to facial expressions. From these landmarks, we derive, via simple geometric calculation, an ordered and regularly spaced set of points on that plane. Afterward, the plane is tilted slightly, by a constant amount, to make it more aligned with the face orientation, and then, we project this set of points on the face surface, along the plane's normal direction. The outcome of this procedure is an ordered grid of points, which defines an atlas for the facial regions that will divide the facial surface. The grid contains 49 points forming 7×7 constellation as shown in Fig. 6a.

Once the grid of points has been defined, we extract a neighborhood of facets around each point of the grid. Each neighborhood can be defined by the set of facets confined within a geodesic disk or a sphere, centered at a grid point (Fig. 6b).

3.4 CNN Architecture and Training

We train our CNN in order to classify the face descriptor image created using mesh-LBP. In our work, we deal with on a small neural network since we are dealing with images of face descriptors rather than images of faces. The learn-

Fig. 5 : An example result of the landmark fitting [68]

Fig. 6 The face descriptor image construction

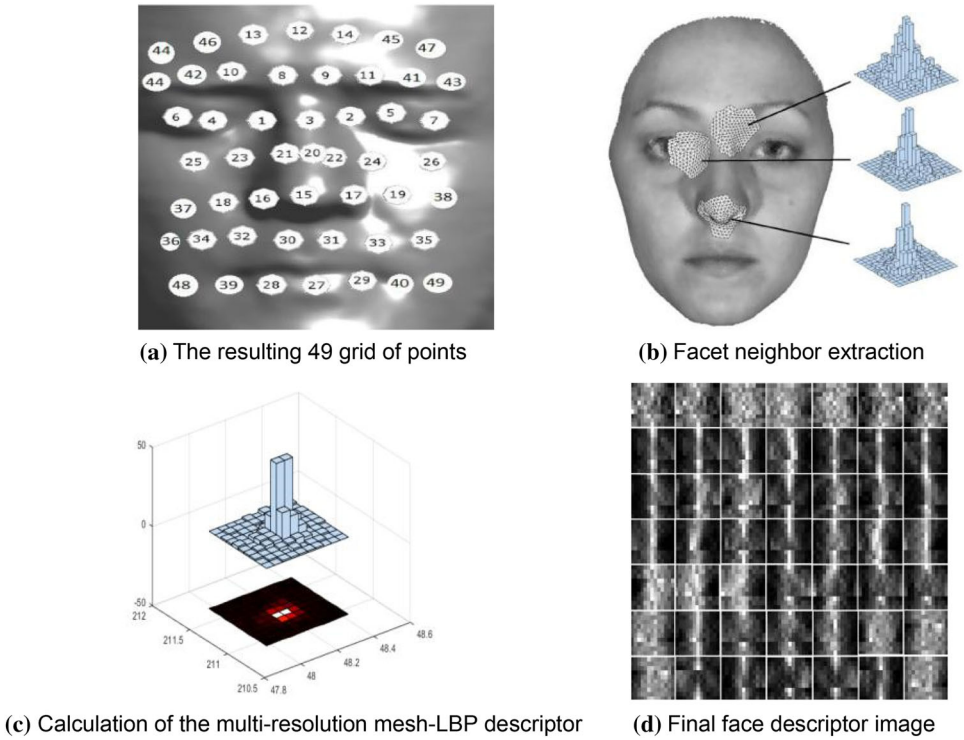
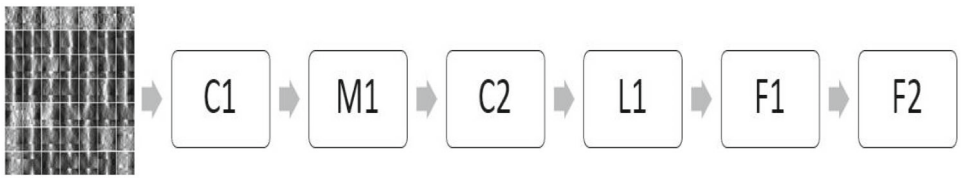


Fig. 7 Outline of the Deep 3D-LBP's CNN architecture. We denote that C is referred to a Convolution layer, M for Max Poling, L for Locally connected layer and F for fully connected layer



ing step is to reinforce the ability of this network to predict and classify facial images. The proposed CNN, presented in Fig. 7, is composed of two convolution layers (C), two fully connected layer (F), max-pooling layer (M) and a locally connected layer (L).

The main difference between the convolution layer and locally connected layer is that the filter in a convolutional layer is common among all output neurons (pixels). In other words, we used a single filter to calculate all neurons (pixels). While, in locally connected layer, each neuron has its own filter. This type of layer lets the network able to learn different types of feature for different regions of the input. In fact, several researchers have benefited from this property especially for face verification tasks [64]. For example,

areas between the eyes and the eyebrows exhibit very different appearance and have much higher discrimination ability compared to areas between the nose and the mouth.

The use of local layers does not affect the computational burden of feature extraction, but does affect the number of parameters subject to training [75]. That means that the number of parameters will be multiplied by the number of output neurons, which could increase the number of parameters in our network. However, in a smaller CNN architecture like ours, we could avoid such issues.

The size of the face descriptor image is 91 × 91 pixels. These images are fed to our CNN. The first convolutional layer (C1) has 32 filters with size 11 × 11. The resulting 32 feature maps are then fed to a 3 × 3 max-pooling layer (M1) with a stride of 2, separately for each channel, followed by another convolutional layer (C2) with 16 filters of size 9 × 9. The subsequent layers (L1) are a locally connected layer composed of 16 filter.

Finally, the last two layers, F5 and F6, are fully connected layers. These layers are able to capture correlations between distant face features. The output of the first fully connected layer (F1) in the network is used as our raw face representation feature vector throughout this paper. The output of the last fully connected layer F2 is fed to a K-way softmax (where K is the number of classes) which produces a distribution over the class labels. If we denote by ok the k-th output of the network on a given input, the probability assigned to the k-th class is the output of the softmax function (Eq. 2):

$$p_k = \exp o_k \Big/ \sum_{h}^{k} o_h \tag{2}$$

It is important to mention the use of the ReLU [76] activation function after the convolution, locally connected and fully connected layer (except the last one L6). In addition, we use the cross-entropy loss in order to maximize the probability of the correct class (face id).

We train our architecture with around 500.000 images from the CASIA-WebFace [77], which contains 494,414 images of 10,575 subjects collected from the Internet. Therefore, we implement the standard back-propagation on feed forward nets by stochastic gradient descent (SGD) with momentum (set to 0,9). We have set an equal learning rate for all trainable layers to 0.01, which was manually decreased, each time by an order of magnitude once the validation error stopped decreasing, to a final rate of 0.0001. We initialized the weights in each layer from a zero-mean Gaussian distribution with = 0.01, and biases are set to 0.5. As a first experiment, we are working on face descriptor image, we use a smaller batch size of 200, and we train the network for 10 epochs over the whole data.

4 Experimental Results

In this section, we first introduce the datasets used in the experiment process. Then, we present the evaluation protocol adopted and the metrics used to validate our method. After that, we focus on the objective evaluation of the proposed method against several challenges present in an uncontrolled environment. Lastly, we provide and analyze the utility of the main components of our system which will be presented as an ablation study. More details are presented in the following paragraphs.

4.1 Datasets

In this evaluation, we use four datasets:

- The CMU Multi-PIE face dataset [78]: It contains more than 750,000 images of 337 people recorded in up to four sessions over the span of five months. Subjects were imaged under 15 viewpoints and 19 illumination conditions while displaying a range of facial expressions.
- The Bosphorus dataset [79]: It contains 4666 scans of 105 subjects scanned in different poses, action units, and occlusion conditions. The dataset is divided in multiple subsets corresponding to neutral and expressive scans (the six fundamental expressions are considered, namely anger, disgust, fear, happy, sad, surprise), scans with Action Units, scans with rotations, and scans with occlusions.
- The LFW dataset [80]: It consists of 13,323 web photographs of 5749 celebrities which are divided into 6,000 face pairs in 10 splits. Performance is measured by mean recognition accuracy using A) the restricted protocol, in which only same and not same labels are available in training; B) the unrestricted protocol, where additional training pairs are accessible in training; and C) an unsupervised setting in which no training whatsoever is performed on LFW images.
- The YTF dataset [81]: It collects 3425 YouTube videos of 1595 subjects (a subset of the celebrities in the LFW). These videos are divided into 5000 video pairs and 10 splits and used to evaluate the video-level face verification.

4.2 Evaluation protocol

Our evaluation has three axes. Firstly, we evaluate the effectiveness of our method for recognizing faces while varying pose and illumination conditions using the Multi-PIE face dataset. Second, we evaluate the robustness of our method against facial expression variations by measuring the recognition rate under six facial expression variations from the Bosphorus dataset. Finally, we test our method in a crowded environments.

For the face recognition in the presence of pose variation (PFR) and combining pose and illumination variations (PIFR) using the Multi-PIE Dataset, two settings (Setting-I and Setting-II) are used for PFR and PIFR, respectively.

- Setting-I: We used a subset in session 01 consisting of 249 subjects with 7 poses and 20 illumination variations. The images of the first 100 subjects constitute the training set. The remaining 149 subjects form the test set. In the test set, the frontal images under neutral illumination work as the gallery and the remaining are probe images.
- Setting-II: We used the images of all the 4 sessions (01{04) under 7 poses and only neutral illumination. The images from the first 200 subjects are used for training and the remaining 137 subjects for testing. In the test set, frontal images from the earliest session work as gallery, and the others are probes.

To evaluate the performance of our system, we use the following metrics:

1. Accuracy: The accuracy defines the total number of correct predictions returned within the face recognition compared to all predictions during test.
2. The receiver operating characteristics (RoC) curve: the ROC curves aim to summarize the trade-off between the true positive rate (TPR) and false positive rate (FPR) for a predictive model. In other words, it tells us how good the model is for distinguishing a given classes, in terms of the predicted probability.

A ROC curve plots the TPR on the y-axis versus the FPR on the x-axis. We note that the true positive rate describes how good the model is at predicting the positive class when the actual outcome is positive, while the false positive rate (FPR), also referred as the false alarm rate, summarizes how often a positive class is predicted when the actual outcome is negative.

$$\mathrm{TPR} = \frac{\mathrm{True\ \ Positives}}{\mathrm{True\ \ Positives} + \mathrm{False\ \ Negatives}} \tag{3}$$

$$\mathrm{FPR} = \frac{\mathrm{False\ \ Positives}}{\mathrm{True\ \ Positives} + \mathrm{False\ \ Negatives}} \tag{4}$$

4.3 Pose and illumination-invariant face recognition

Pose and illumination-invariant face recognition is a challenging problem that has taken high interest from researchers [82]. We start by comparing our method with several state-of-the-art methods for pose-invariant face recognition. Then, for both pose and illumination invariant face recognition.

4.3.1 Pose-invariant face recognition PFR

In this test, we compare our Deep 3D-LBP against state-of-the-art method. We could classify the method into two classes: 2D and 3D methods. 2D methods deal with the pose and illumination with the use of the image pixels or image features for recognition.

However, the 3D methods model the variations in pose based on an analysis-by-synthesis approach [83]. In other words, these methods aim at matching a 3D face model to an annotated 2D image as input. We note that the annotation essentially consists of the facial landmarks detection. The overall results are summarized in Table 1.

Overall, the results show that most of 3D methods can obtain higher accuracy rates than the 2D methods, especially in the presence of a large pose variation ($\pm 45°$). Nevertheless, SPAE [85], which is characterized by a robust nonlinear modeling capability, works more effectively than other 2D methods and provides better results than some 3D methods such as Asthana [88] and MDF [89]. But this remains limited, and it can be noted that other 3D methods (HPEN [66], and ESO [83]) outperform its results with a large margin of difference. HPEN [66], which generates a neutral and frontal face image based on the 3DMM and identity preserving 3D transformation and uses the PCA for classification, performs well only in small pose variation. Besides, the U-3DMM can model both pose and facial shape rather than pose only by [88, 89]. We present the result of the U-3DMM using high-dimensional Gabor feature (HDF) [91] and PCA coefficients. U-3DMM (HDF) works much better than U-3DMM (PCA) due to the ability of the HDF feature to capture both global and local facial information.

However, our proposed method outperforms both the 2D and 3D methods. This is due to the ability of our method to model and learn both pose variation and illumination variation. To summary, our method is more reliable than other methods since it deals with the limitations of the latter. We use 3D data from 3DMM to tackle the problems of wide variation in pose such as [48, 63, 89], and LBP feature to be more robust against illumination and expression. Moreover, the use of CNNs for training allows us to have high accuracy rates, rather than use of other classifiers methods [48, 83]. In addition, the face frontalization [66] improves the result but it seems limited especially in higher degrees which could be caused by the classifier used (PCA).

4.4 Pose and illumination-invariant face recognition PIFR

The result presented in Table 2 compares our Deep 3D-LBP against other methods while varying illumination and pose.

The results in Table 2 show that the use of the subspace methods [92] offers the worst results. In addition, we could

Table 1 Recognition rate (%) on various poses under neutral illumination on the Multi-PIE Dataset [78]

	Method and date	− 45	− 30	− 15	+ 15	+ 30	+ 45
2D	DAE [63], 2009	69.0	81.2	91.0	91.9	86.5	74.3
	GMA [62], 2012	75.0	74.5	82.7	92.6	87.5	65.2
	MRFs [84], 2013	86.3	89.7	91.7	91	89	85.7
	SPAE [85], 2014	84.9	92.6	96.3	95.7	94.3	84.4
	RFG [86], 2014	86.4	91.2	96.0	96.1	90.90	85.4
	SF-VF + LBP [87], 2020	91.43	93.88	91.14	90.91	92	87.14
	asthana [88], 2011	74.1	91.0	95.7	95.7	89.5	74.8
3D	MDF [89], 2012	78.7	94.0	99.0	98.7	92.2	81.8
	PAF [90], 2013	84	99	99.33	99.67	99.67	98.33
	HPEN + PCA [66], 2015	88.5	95.4	97.2	98	95.7	89
	U-3DMM + (PCA) [48], 2016	91.2	95.7	96.8	96.9	95.3	90.9
	U-3DMM + (HDF) [48], 2016	96.5	98.4	99.2	98.2	98.9	97.9
	ESO + LPQ [83], 2017	91.7	95.3	96	96.7	95.3	90.3
	Deep 3D-LBP	97.4	99.5	99.5	99.7	99.0	96.7

Table 2 Recognition rate (%) on the Multi-PIE dataset [78] across pose and illumination variations

Method and date		− 45	− 30	− 15	+ 15	+ 30	+ 45
Subspace learning	Li [92], 2011	63.5	69.3	79.7	75.6	71.6	54.6
Deep learning	DNN-RL [64], 2013	67.1	74.6	86.1	83.3	75.3	61.8
	MVP [93], 2014	84.9	92.6	96.3	95.7	94.3	84.4
	DNN-CPF [94], 2015	73	81.7	98.4	89.5	80.4	70.3
	LNFF-LRA [95], 2017	77.2	87.7	94.9	94.8	88.1	76.4
	HPN [96], 2017	71.3	78.8	82.2	86.2	77.8	74.3
3D	U-3DMM, 2016 [48]	73.1	86.9	93.3	91.3	81.2	69.7
	ESO-3DMM, 2016 [83]	80.8	88.9	96.7	97.6	93.3	81.1
	GM-3DMM [49], 2018	84.3	89.4	97.4	99	96.8	92
	Deep 3D-LBP	97.4	99.5	99.5	99.7	99.0	96.7

notice that our method outperforms both the deep learning and 3D-based methods. Moreover, it is obvious that the use of a 3DMM is well adapted to solve the problems caused by the extreme variations in pose and illumination. For example, we can notice that the maximum value obtained by the DL-based methods is 96.3% and 95.7% with a pose variation of -15 ° and + 15 °, respectively. Nevertheless, for the 3D-based methods, GM-3DMM [49] obtains 97.4% and 99% in -15 ° and + 15 °, respectively.

Using the deep 3DLBP, we are able to obtain much more interesting results, and this is more notable in the large variations in installation (97.4%, 96.7% obtained in—45 ° and + 45 °, respectively, by our method against 84.9 and 92 as a maximum value obtained for all the state-of-the-art methods).

We may also notice the difference taking the example of U-3DMM [48] from Tables 1 and 2, in which, U-3DMM offers more interesting results while varying pose (96.5% in -45 ° rather than for variation in illumination and pose (73.1% in − 45 °). On the other hand, our method remains reliable and gives better results in both cases.

4.5 Facial expression-invariant face recognition

To include more challenging cases, we tested our method on the Bosphorus dataset, which presents huge variation in facial expressions. Table 3 provides a comprehensive comparison between our method and recently published methods.

Table 3 provides a comprehensive comparison between our method and the published methods. We conclude that our solution is efficient against the state-of-the-art methods. Several methods achieve the rate of 100% when the face is neutral. But, this accuracy decreases when the expressions change. For instance, Zhang et al. [104] achieve higher results by observing emotional statements Neutral and Happy (100% and 96.23%, respectively). However, there is a significant decrease when it concerns an emotion of ANGER (81.6%), DISGUST (79.71%) and FEAR (88.5%). However, our method keeps higher results even in those categories.

Table 3 Recognition rate (%) across facial expressions on the Bosphorus dataset [79]

Method and date	Neutral	Anger	Disgust	Fear	Happy	Sad	Surprise
Li et al. [43]	100	88.7	76.8	92.9	95.3	95.5	98.6
Berretti et al. [44]	97.9	85.9	81.2	90	92.5	93.95	91.5
Li et al. [97]	100	97.18	86.96	98.57	98.11	100	98.59
Azazi et al. [98]	81.25	82.5	90	86.25	97.5	67.5	83.75
Lei et al. [99]	98.96	94.12	88.24	98.55	98.08	96.08	96.92
Deng et al. [100]	100	95.8	92.8	97.7	95.3	98.5	98.6
Hariri et al. [101]	87.5	86.25	85.25	81	93	79.75	90.5
Abbad et al. [45]	100	95.77	88.41	81.41	88.68	96.97	92.96
Zhang et al. SPS:refid::bib99[99]	100	81.69	79.71	88.57	96.23	90.91	95.77
Deng et al. [46]	100	97.2	94.2	97.1	96.2	98.5	98.6
Liang et al. [102]	100	94.37	85.51	97.14	69.23	98.59	98.52
Atik et al. [103]	98.68	78.87	81.16	80	97.17	95.45	95.77
Deep 3D-LBP	100	97.18	96.75	100	97.63	98.88	100

Table 4 Face verification on the LFW dataset

Method and year	Accuracy (%)
LFW-3D [105], 2015	93.62
LFW-HPEN [66], 2015	96.25
FF-GAN[106], 2017	96.42
DED-GAN [107], 2020	97.52
CAPG-GAN [108], 2018	99.37
DA-GAN [109], 2020	99.56
joint-Res [110], 2018	98.03
Light CNN [111], 2018	98.13
M2FPA [112], 2019	99.41
FR-CNN [61], 2020	99.2
Deep 3D-LBP	99.59

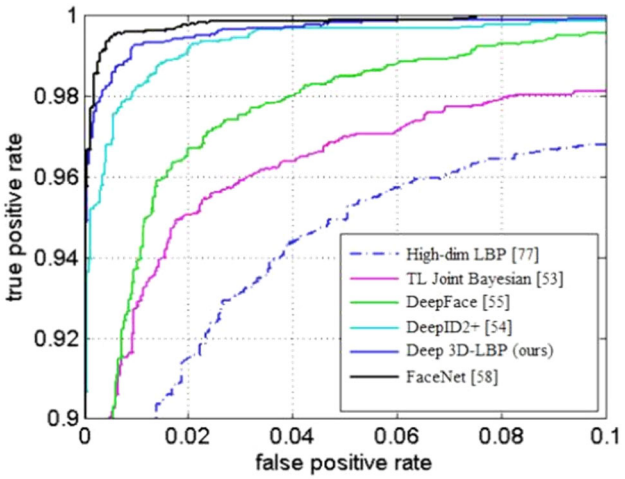


Fig. 8 ROC for face verification on LFW

4.6 LFW verification

We evaluate our model on LFW using the standard protocol for unrestricted, labeled outside data. Thus, we learn an SVM on top of the 2-distance vector following the restricted protocol, i.e., where only the 5400 pair labels per split are available for the SVM training. Results are presented in Table 4, in which, we compare our method against deep face recognition methods.

It is important to note that the first three methods aim to generate a new neutral and frontal image for the recognition [66, 105–107, 109]. In spite of its efficiency, this is still limited. The deep methods [61, 110–112] use large datasets and very deep models and offer results that are competitive to our methods. To conclude, it can be mentioned that the use of robust and reliable features with a learning process even if the network seems small could lead to better results and outperform even very deep models. We plot the ROC curve, and the results are illustrated in Fig. 8.

Considering the ROC curve, we observe that the best results are those closer to the value $y = 1$, especially the blue (Deep 3D-LBP) and black curves (FaceNet [113]). Regarding the accuracy obtained using the faceNet (99.63%) [113] and our method are very close to each other. On the other hand, it should also be mentioned that we use a simple neural network compared to the one composed of 25 FaceNet networks.

4.7 YouTube Faces dataset verification

We further validate our Deep 3D-LBP on the YouTube Faces dataset (YTF). The image quality of YouTube video frames is generally worse than that of web photographs, mainly due to motion blur or viewing distance. The results are summarized in Table 5.

Our method, although the small learning process, allows us to achieve higher learning rates rather than [59, 109,

Table 5 Face verification on YouTube Faces

Method and year	# img	# network	Accuracy (%)
DeepFace, 2014 [75]	4 M	3	91.4
DeepID2 + ,2015 [109]	0.3	25	93.2
Webface, 2017 [114]	0.49	1	92.24
normface, 2015 [115]	1.5	1	94.72
CD-loss, 2019 [116]	0.49	1	92.45
Center loss, 2016[59]	0.7	1	94.9%
Cosface, 2018[117]	5 M	1	94.6%
IVR-FR, 2021[118]	0.8 M	1	94%
Deep 3D-LBP	0.5	1	94.97

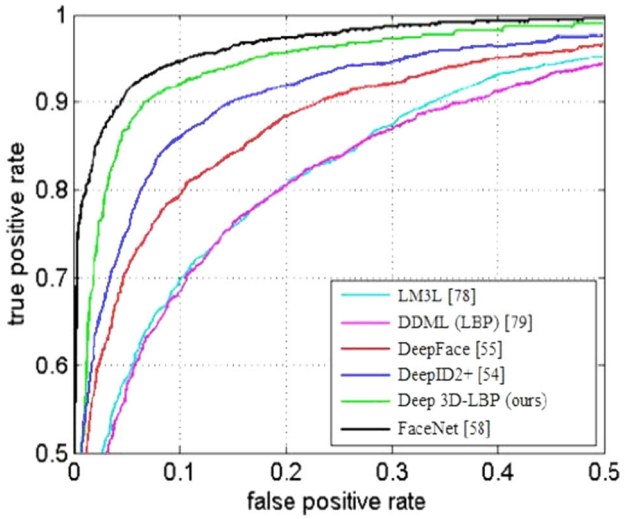


Fig. 9 ROC for face verification on YTF

114, 116] that use also smaller datasets for training. Despite the fact that our results remain limited and do not reach the best rates like [57, 113, 119], the previous tests proved the effectiveness of our method for classification task. But improvement is always possible.

The results are also presented in the ROC curve shown in Fig. 9. Indeed, the margin between FaceNet and our method is larger than the other case (Fig. 8). Otherwise, this margin is produced due to the complexity of the environment and problems of the YTF dataset images.

4.8 Ablation study

To assure the effectiveness of our method, we will evaluate, separately, each component used in our method again the proposed Deep 3D-LBP system, with the same database that are already used in for evaluation process.

We start by analyzing the 3DMM's utility. In fact, the use of 3DMM is designed to attenuate the pose variation's effect. In a preliminary experiment, we will compare the results

Table 6 Recognition rate (%) across poses on multi-PIE

Method	− 45	− 30	− 15	+ 15	+ 30	+ 45
3DMM	97.4	99.5	99.5	99.7	99.0	96.7
Deep 3D-LBP	97.4	99.5	99.5	99.7	99.0	96.7

Table 7 Recognition rate (%) averaging pose and illuminations on multi-PIE

Method	− 45	− 30	− 15	+ 15	+ 30	+ 45
3DMM	74.1	91.0	95.7	95.7	89.5	74.8
Deep 3D-LBP	97.4	99.5	99.5	99.7	99.0	96.7

provided using only 3DMM against the corresponding ones obtained using Deep 3D-LBP for face recognition under pose variation.

From Table 6, we can definitely notice that the results are identical, so we can conclude that for images showing a pose variation, 3DMM is simpler and more efficient. But these results varied under different conditions. Let us consider the effects of varying both pose and illumination.

On the one hand, we notice from Tables 6 and 7 that the results decrease when we consider the illumination constrains. On the other hand, the results obtained by our method remain the same. This improvement is due to the use of the Mesh-LBP features which are also robust to illumination variations, and also to the learning process that use a set of varied images, which further improves the result obtained.

Subsequently, we test our method when varying facial expression. In this context, we evaluate the results obtained by the Mesh-LBP as a robust feature to facial expression and the proposed Deep 3D-LBP system to assess the usefulness of combining robust face features with training. The results are shown in Table 8.

The results are similar when the expressions have no influence on the human face characteristics. But in other cases, we clearly visualize the difference. As in the case of the expressions "Disgust" or "Happy," the evolution is about 11% and 9%, respectively.

As a last test, we compare Mesh-LBP and Deep 3D-LBP using the Bosphorus dataset while varying pose. The results are shown in Fig. 10.

Mesh-LBP method achieves comparable results to our method with a small rotation degree (10 °). However, it decreases for a rotation of 20 °, and higher.

5 Conclusion

Face recognition in crowed environment is considered as a challenging problem due to the several issues could be caused in the process of acquiring facial images. Among the

Table 8 Recognition rate (%) across facial expressions on Bosphorus

Method	Neutral	Anger	Disgust	Fear	Happy	Sad	Surprise
Mesh-LBP	100	97.18	85.51	98.57	88.68	96.97	97.18
Deep 3D-LBP	100	97.18	96.75	100	97.63	98.88	100

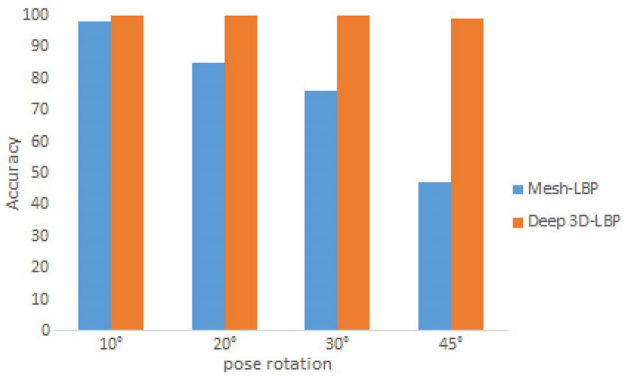


Fig. 10 Recognition rate (%) across poses on Bosphorus

factors that should be considered in a facial recognition system is the feature extraction or face description. Indeed, the underlying face descriptor would need to be invariant to pose, illumination, expression, and image quality. In addition, short descriptors are preferable, and if possible, sparse features.

From the research that has been carried out in this paper, it is possible to conclude that coupling a 3D model-based alignment, an LBP descriptor constructed on the 3D mesh with a feed forward CNN model can effectively learn from many examples to overcome the drawbacks and limitations of previous methods. The findings of our research are quite convincing, and thus the following conclusions can be drawn: The results obtained indicate that our method outperforms the state-of-the-art methods on several points: Pose-invariant face recognition, pose and illumination face recognition and facial expression-invariant face recognition. However, the main limitation of the experimental result concerns face recognition in videos.

Clearly, further research will be required to improve our results and to work on the remaining issues. We suggest, as a future work, to generate a face image for those images having a weak quality or suffering from several problems (pose, resolution, and illumination) in order to improve its quality and make the recognition more accurate.

Declarations

Conflict of interest Sahbi Bahroun declares that he has no conflict of interest. Rahma Abed declares that he has no conflict of interest. Ezzeddine Zagrouba declares that he has no conflict of interest.

References

1. Chakraborty, S., Thounaojam, D. M., Sinha, N.: A shot boundary detection technique based on visual colour information. Multimed. Tools Appl. **80** 1–16 (2020)
2. Singh, A., Thounaojam, D.M., Chakraborty, S.: A novel automatic shot boundary detection algorithm: robust to illumination and motion effect. Signal, Image and Video Processing **14** 1–9 (2019)
3. L. Ferreira, L. A. da Silva Cruz, P. Assuncao, Towards key-frame extraction methods for GAN video: a review, EURASIP J. Image Video Process. (1). 28 (2016).
4. Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., Garcia-Rodriguez, J.: A survey on deep learning techniques for image and video semantic segmentation. Appl. Soft Comput. **70**, 41–65 (2018)
5. Hassaballah, M., Aly, S.: Face recognition: challenges, achievements and future directions. IET Comput. Vision **9**(4), 614–626 (2015)
6. Anil, J., Suresh, L.P.: Literature survey on face and face expression recognition. In: International Conference on Circuit, Power and Computing Technologies (ICCPCT), IEEE, pp. 1–6 (2016)
7. Anwarul, S., Dahiya, S.: A comprehensive review on face recognition methods and factors affecting facial recognition accuracy. In: Proceedings of ICRIC 2019, Springer, pp. 495–514 (2020)
8. Abdurrahim, S.H., Samad, S.A., Huddin, A.B.: Review on the effects of age, gender, and race demographics on automatic face recognition. Vis. Comput. **34**(11), 1617–1630 (2018)
9. Yin, Y., Jiang, S., Robinson, J.P., Fu, Y.: Dual-attention GAN for large-pose face frontalization. arXiv preprint https://arxiv.org/abs/2002.07227.
10. Cao, J., Hu, Y., Zhang, H., He, R., Sun, Z.: Towards high fidelity face frontalization in the wild. Int. J. Comput. Vis. **128** 1–20 (2019)
11. Ferrari, C., Lisanti, G., Berretti, S., Del Bimbo, A.: Effective 3D based frontalization for unconstrained face recognition. In: 23rd International Conference on Pattern Recognition (ICPR). IEEE, pp. 1047–1052 (2016)
12. Sagonas, C., Panagakis, Y., Zafeiriou, S., Pantic, M.: Robust statistical face frontalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3871–3879 (2015)
13. Huang, R., Zhang, S., Li, T., He, R.: Beyond face rotation: global and local perception GAN for photorealistic and identity preserving frontal view synthesis. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2439–2448 (2017)
14. Zhang, S., Miao, Q., Zhu, X., Chen, Y., Lei, Z., Wang, J., et al.: Pose-weighted GAN for photorealistic face frontalization. In: 2019 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 2384–2388 (2019)
15. Wang, S., Zou, Y., Min, W., Wu, J., Xiong, X. (2021). Multi-view face generation via unpaired images. The Vis. Comput., 1–16.
16. Mahamdioua, M., Benmohammed, M.: Automatic adaptation of SIFT for robust facial recognition in uncontrolled lighting conditions. IET Comput. Vision **12**(5), 623–633 (2018)
17. Arya, K., Rajput, S.S., Upadhyay, S.: Noise-robust low-resolution face recognition using SIFT features. Comput. Intel. Theor. Appl. Future Dir. 645–655 (2019)

18. Sushama, M., Rajinikanth, E.: Face recognition using DRLBP and SIFT feature extraction. In: International Conference on Communication and Signal Processing (ICCSP) 994–999 (2018).
19. Gupta, S., Thakur, K., Kumar, M.: 2d-human face recognition using SIFT and SURF descriptors of face's feature regions. Vis. Comput. **37** 1–10 (2020).
20. Wang, Y., Li, M., Zhang, C., Chen, H., Lu, Y.: Weighted-fusion feature of MB-LBPUH and HoG for facial expression recognition. Soft. Comput. **24**(8), 5859–5875 (2020)
21. Voronov, V., Strelnikov, V., Voronova, L., Trunov, A., Vovik, A.: Faces 2d-recognition and identification using the hog descriptors method. In: Conference of Open Innovations Association, FRUCT, no. 24, FRUCT, pp. 783–789 (2019)
22. Nhat, H.T.M., Hoang, V.T.: Feature fusion by using lbp, hog, gist descriptors and canonical correlation analysis for face recognition. In: 2019 26th International Conference on Telecommunications (ICT), IEEE, pp. 371–375 (2019)
23. Xu, J., Xue, X., Wu, Y., Mao, X.: Matching a composite sketch to a photographed face using fused HOG and deep feature models. Vis. Comput. **37**(4), 765–776 (2021)
24. Zhang, H., Qu, Z., Yuan, L., Li, G.: A face recognition method based on LBP feature for CNN. In: IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, pp. 544–547 (2017)
25. Shi, L., Wang, X., Shen, Y.: Research on 3d face recognition method based on lbp and svm, Optik **220**, 165157 (2020)
26. Ren, X., Guo, H., Di, C., Han, Z., Li, S.: Face recognition based on local Gabor binary patterns and convolutional neural network. In: International Conference in Communications, Signal Processing, and Systems, Springer, pp. 699–707 (2016)
27. Zhang, B., Liu, G., Xie, G.: Facial expression recognition using LBP and LPQ based on Gabor wavelet transform. In: 2016 2nd IEEE International Conference on Computer and Communications (ICCC), IEEE, , pp. 365–369 (2016)
28. Chengeta, K., Viriri, S.: Facial expression recognition: A survey on local binary and local directional patterns. In: International Conference on Computational Collective Intelligence, Springer, pp. 513–522 (2018)
29. Oloyede, M.O., Hancke, G.P., Myburgh, H.C.: A review on face recognition systems: recent approaches and challenges. Multimedia Tools and Applications **79**(37), 27891–27922 (2020)
30. Kortli, Y., Jridi, M., Al Falou, A., Atri, M.: Face recognition systems: a survey, Sensors **20**(2), 342 (2020)
31. Cheng, Y., Jiao, L., Cao, X., Li, Z.: Illumination-insensitive features for face recognition. Vis. Comput. 33 (11) (2017) 1483–1493.
32. Ahmed, S.B., Ali, S.F., Ahmad, J., Adnan, M., Fraz, M.M.: On the frontiers of pose invariant face recognition: a review. Artif. Intell. Rev. **53**(4), 2571–2634 (2020)
33. Napoleon, T., Alfalou, A.: Pose invariant face recognition: 3d model from single photo. Opt. Lasers Eng. **89**, 150–161 (2017)
34. Juefei-Xu, F., Luu, K., Savvides, M.: Spartans: Single-sample periocular-based alignment-robust recognition technique applied to non-frontal scenarios. IEEE Trans. Image Process. **24**(12), 4780–4795 (2015)
35. Ding, C., Tao, D.: Robust face recognition via multimodal deep face representation. IEEE Trans. Multimedia **17**(11), 2049–2058 (2015)
36. Liang, J., Tu, H., Liu, F., Zhao, Q., Jain, A.K.: 3d face reconstruction from mugshots: Application to arbitrary view face recognition. Neurocomputing **410**, 12–27 (2020)
37. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al.: Recent advances in convolutional neural networks. Pattern Recogn. **77**, 354–377 (2018)
38. Finizola, J.S., Targino, J.M., Teodoro, F.G., Lima, C.A.: Comparative study between deep face, autoencoder and traditional machine learning techniques aiming at biometric facial recognition. In: 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, pp. 1–8 (2019)
39. Guo, G., Zhang, N.: A survey on deep learning based face recognition. Comput. Vis. Image Underst. **189**102805 (2019)
40. Chihaoui, M., Elke , A., Bellil, W., Ben Amar, C.: A survey of 2d face recognition techniques. Computers **5** (4) 21 (2016)
41. Zhou, S., Xiao, S.: 3d face recognition: a survey. HCIS **8**(1), 35 (2018)
42. Trigueros, D. S., Meng, L., Hartnett, M.: Face recognition: from traditional to deep learning methods. arXiv preprint https://arxiv.org/abs/1811.00116.
43. Li, H., Huang, D., Lemaire, P., Morvan, J.-M., Chen, L.: Expression robust 3d face recognition via mesh-based histograms of multiple order surface differential quantities 3053–3056 (2011)
44. Berretti, S., Werghi, N., Bimbo, A.D., Pala, P.: Matching 3d face scans using interest points and local histogram descriptors. Comput. Graph. **37**(5), 509–525 (2013)
45. Abbad, A., Abbad, K., Tairi, H.: 3d face recognition: Multi-scale strategy based on geometric and local descriptors. Comput. Electr. Eng. **70**, 525–537 (2018)
46. Deng, X., Da, F., Shao, H., Jiang, Y.: A multi-scale three-dimensional face recognition approach with sparse representation-based classifier and fusion of local covariance descriptors. Comput. Electric. Eng. **85**, 106700 (2020).
47. Zhang, Z., Da, F., Yu, Y.: Data-free point cloud network for 3d face recognition, arXiv. arXiv- 1911 (2019)
48. Hu, G., Yan, F., Chan, C.-H., Deng, W., Christmas, W., Kittler, J., Robertson, N. M.: Face recognition using a unified 3d morphable model. In: European Conference on Computer Vision, pp. 73–89. Springer (2016)
49. Koppen, P., Feng, Z.-H., Kittler, J., Awais, M., Christmas, W., Wu, X.-J., Yin, H.-F.: Gaussian mixture 3d morphable face model. Pattern Recogn. **74**, 617–628 (2018)
50. Kumar, A., Kaur, A., Kumar, M.: Face detection techniques: a review. Artif. Intell. Rev. **52**(2), 927–948 (2019)
51. Bodini, M.: A review of facial landmark extraction in 2d images and videos using deep learning. Big Data and Cognitive Computing **3**(1), 14 (2019)
52. Taskiran, M., Kahraman, N., Erdem, C.E.: Face recognition: past, present and future (a review). Digital Signal Process (2020) 102809
53. Huang, G.B., Lee, H., Learned-Miller, E., Learning hierarchical representations for face verification with convolutional deep belief networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
54. Zhu, Z., Luo, P., Wang, X., Tang, X.: Recover canonical view faces in the wild with deep neural networks. arXiv
55. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE conference on computer vision and pattern recognition
56. Schro, F., Kalenichenko, D., Philbin, J., Facenet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp. 815–823
57. Parkhi, O.M., Vedaldi, A, Zisserman, A.: Deep face recognition (2015)
58. Wu, X., He, R., Sun, Z., Tan, T.: A light CNN for deep face representation with noisy labels. IEEE Trans. Inf. Forensics Secur. **13**(11), 2884–2896 (2018)
59. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: European Conference on Computer Vision, Springer, pp. 499–515 (2016)
60. Yeung, H. W. F., Li, J., Chung, Y.Y.: Improved performance of face recognition using CNN with constrained triplet loss layer. In: 2017

International Joint Conference on Neural Networks (IJCNN), IEEE, pp 1948–1955 (2017)
61. Fredj, H.B., Bouguezzi, S., Souani, C., Face recognition in unconstrained environment with CNN. Springer, pp. 1–20 (2020).
62. Sharma, A., Kumar, A., Daume, H., Jacobs, D.W.: Generalized multi view analysis: a discriminative latent space, pp. 2160–2167 (2012)
63. Bengio, Y.: Learning deep architectures for ai
64. Zhu, Z., Luo, P., Wang, X., Tang, X.: Deep learning identity-preserving face space, pp. 113–120 (2013)
65. Chu, Y., Zhao, L., Ahmad, T.: Multiple feature subspaces analysis for single sample per person face recognition. Vis. Comput. **35**(2), 239–256 (2019)
66. Zhu, X., Lei, Z., Yan, J., Yi, D., Li, S.Z.: High- fidelity pose and expression normalization for face recognition in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 787–796 (2015)
67. Werghi, N., Tortorici, C., Berretti, S., Del Bimbo, A.: Boosting 3d lbp-based face recognition by fusing shape and texture descriptors on the mesh. IEEE Trans. Inf. Forensics Secur. **11**(5), 964–979 (2016)
68. Egger, B., Smith, W.A., Tewari, A., Wuhrer, S., Zollhoefer, M., Beeler, T., Bernard, F., Bolkart, T., Kortylewski, A., Romdhani, S., et al.: 3D morphable face models past, present, and future. ACM Trans. Gr. **39**(5), 1–38 (2020)
69. King, D.E.: Dlib-ml: A machine learning toolkit, J. Mach. Learn. Res. 1755–1758.
70. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition
71. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
72. Huber, P., Hu, G., Tena, R., Mortazavian, P., Koppen, W., Ratsch, W.C.M., Kittler, J.: A multi-resolution 3D morphable face model and fitting framework. In: Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 79–86 (2016)
73. Hartley, R.I., Zisserman, A.: Multiple view geometry in computer vision, 2nd edn. Cambridge University Press, Cambridge
74. Aldrian, O., Smith, W.A.P.: Inverse rendering of faces with a 3D morphable model. IEEE Trans. Pattern Anal. Mach. Intell. **35**(5), 1080–1093
75. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
76. Dahl, G.E., Sainath, T.N., Hinton, G.E.: Improving deep neural networks for lvcsr using rectified linear units and dropout. In: International Conference on Acoustics, Speech and Signal Processing 28 (5)
77. Yi, Dong, et al.: Learning face representation from scratch. arXiv preprint arXiv:1411.7923 (2014)
78. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. Image Vis. Comput. **28**(5), 807–813 (2010)
79. Savran, A., Alyuz, N., H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, L. Akarun, Bospho-rus database for 3d face analysis, in: European workshop on biometrics and identity management, Springer, 2008, pp. 47–56.
80. Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments (2008)
81. Wolf, L., Hassner, T., Maoz, I.: Face recognition in unconstrained videos with matched background similarity. In: CVPR 2011. IEEE, pp. 529–534 (2011)
82. Ding, C., Tao, D.: A comprehensive survey on pose-invariant face recognition. ACM Trans. Intell. Syst. Technol. **7**(3), 1–42 (2016)
83. Hu, G., Yan, F., Kittler, J., Christmas, W., Chan, C.H., Feng, Z., Huber, P.: Efficient 3d morphable face model fitting. Pattern Recogn. **67**, 366–379 (2017)
84. Ho, H.T., Chellappa, R.: Pose-invariant face recognition using markov random fields. IEEE Trans. Image Process. **22**(4), 1573–1584 (2012)
85. Kan, M., Shan, S., Chang, H., Chen, X.: Stacked progressive auto-encoders (SPAE) for face recognition across poses , pp. 1883–1890 (2014)
86. Kafai, M., An, Le., Bhanu, B.: Reference face graph for face recognition. IEEE Trans. Inf. Forensics Secur. **9**(12), 2132–2143 (2014)
87. Petpairote, C., Madarasmi, S., Chamnongthai, K.,: 2D pose-invariant face recognition using single frontal-view face database, Wireless Personal Communications, 1–17 (2020)
88. Asthana, A., Marks, T. K., Jones, M. J., Tieu, K. H., Rohith, M.: Fully automatic pose-invariant face recognition via 3d pose normalization, pp. 937–944 (2011)
89. Li, S., Liu, X., Chai, X., Zhang, H., Lao, S., Shan, S.: Morphable displacement field based image matching for face recognition across pose, pp. 102–115 (2012)
90. Yi, D., Lei, Z., Li, S. Z,: Towards pose robust face recognition, pp. 3539–3545 (2013)
91. Chen, D., Cao, X., Wen, F., Sun, J.: Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3025–3032 (2013)
92. Li, A., Shan, S., Gao, W.: Coupled bias-variance tradeoff for cross-pose face recognition. IEEE Trans. Image Process. **21**(1), 305–315 (2011)
93. Zhu, Z., Luo, P., Wang, X., Tang, X.: Deep learning multi-view representation for face recognition, arXiv preprint https://arxiv.org/abs/1406.6947.
94. Yim, J. , Jung, H., Yoo, B., Choi, C., Park, D., Kim, J.: Rotating your face using multi-task deep neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 676–684 (2015)
95. Deng, W., Hu, J., Wu, Z., Guo, J.: Lighting-aware face frontalization for unconstrained face recognition. Pattern Recogn. **68**, 260–271 (2017)
96. Ding, C., Tao, D.: Pose-invariant face recognition with homography-based normalization. Pattern Recogn. **66**, 144–152 (2017)
97. Li, H., Huang, D., Morvan, J.-M., Wang, Y., Chen, L.: Towards 3d face recognition in the real: a registration-free approach using ne-grained matching of 3d keypoint descriptors. Int. J. Comput. Vision **113**(2), 128–142 (2015)
98. Azazi, A., Lut , S. L., Venkat, I., Fernandez-Martinez, F.: Towards a robust affect recognition: automatic facial expression recognition in 3D faces. In: Expert Systems with Applications, Vol. 42, pp. 3056–3066. Elsevier (2015)
99. Lei, Y., Guo, Y., Hayat, M., Bennamoun, M., Zhou, X.: A two-phase weighted collaborative representation for 3d partial face recognition with single sample. Pattern Recogn. **52**, 218–237 (2016)
100. Deng, X., Da, F., Shao, H.: Efficient 3d face recognition using local covariance descriptor and riemannian kernel sparse coding. Comput. Electr. Eng. **62**, 81–91 (2017)
101. Hariri, W., Tabia, H., Farah, N., Benouareth, A., Declercq, D.: 3D facial expression recognition using kernel methods on Riemannian manifold. In: Engineering Applications of Artificial Intelligence, Vol. 42, pp. 25–32. Elsevier (2017)
102. Liang, Y., Liao, J.-C., Pan, J,: Mesh-based scale-invariant feature transform-like method for three-dimensional face recognition

under expressions and missing data. J. Electron. Imaging, Vol. 29, International Society for Optics and Photonics, p. 053008 (2020)
103. Atik, M.E., Duran, Z.: Deep learning-based 3D face recognition using derived features from point cloud. In: Innovations in Smart Cities Applications Volume 4: The Proceedings of the 5th International Conference on Smart City Applications, Springer International Publishing, pp. 797–808 (2021)
104. Zhang, Z., Da, F., Yu, Y.: Data-free point cloud network for 3d face recognition (2019). https://arxiv.org/abs/1911.04731
105. Hassner, T., Harel, S., Paz, E., Enbar, R., Effective face frontalization in unconstrained images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4295–4304 (2015)
106. Yin, X., Yu, X., Sohn, K., Liu, X., Chandraker, M.: Towards large-pose face frontalization in the wild. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3990–3999 (2017)
107. Hu, C., Feng, Z., Wu, X., Kittler, J.: Dual encoder-decoder based generative adversarial networks for disentangled facial representation learning. IEEE Access **8**, 130159–130171 (2020)
108. Hu, Y., Wu, X., Yu, B., He, R., Sun, Z.: Pose-guided photorealistic face rotation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8398–8406 (2018)
109. Yin, Yu, et al. Dual-attention GAN for large-pose face frontalization. 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020). IEEE, 2020
110. Zhang, Y., Shang, K., Wang, J., Li, N., Zhang, M.M.Y.: Patch strategy for deep face recognition, vol. 12, pp. 819–825. IET (2018)
111. Wu, X., He, R., Sun, Z., Tan, T.: A light CNN for deep face representation with noisy labels, vol. 13, IEEE, pp. 2884–2896 (2018)
112. Li, P., Wu, X., Hu, Y., He, R., Sun, Z.: M2fpa: a multi-yaw multi-pitch high-quality dataset and benchmark for facial pose analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10043–10051 (2019)
113. Mian, A.S., Bennamoun, M., Owens, R.: An efficient multimodal 2d–3d hybrid approach to automatic face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **29**(11), 1927–1943 (2007)
114. Yi, D., Lei, Z., Liao, S., Li, S.Z,: Learning face representation from scratch, pp. 2892–2900 (2014)
115. Wang, F.. Xiang, X., Cheng, J., Yuille, A. L.: Normface: L2 hypersphere embedding for face veri cation, pp. 1041–1049 (2017)
116. Zhang, M.M.Y., Shang, K., Wu, H.: Deep compact discriminative representation for unconstrained face recognition, Vol. 75, pp. 118–127. Elsevier (2019).
117. Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., and Zhou, J., Li, Z., Liu, W.: Cosface: large margin cosine loss for deep face recognition, pp. 5265–5274 (2018)
118. Kim, M., Hong, J., Kim, J., Lee, H. J., Ro, Y. M.: Unsupervised disentangling of viewpoint and residues variations by substituting representations for robust face recognition. IEEE, pp. 8952–8959 (2021)
119. Cao, Z., Yin, Q., Tang, X., Sun, J.: Face recognition with learning-based descriptor. In : IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE **2010**, 2707–2714 (2010)
120. Chang, K.I., Bowyer, K.W., Flynn, P.J.: An evaluation of multimodal 2D+3D face biometrics. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 619–624 (2005)

Sahbi Bahroun received his Ph.D. From the higher school of communication at the University of Carthage, Tunis. He is an assistant professor in the high institute of computer science (ISI) University of Tunis El Manar. His main activity is focused on intelligent imaging and computer vision.

Rahma Abed is a Ph.D. student in image and video processing within LIMTIC Laboratory at University of Tunis El Manar. She received the Master's degree in Intelligent Imaging and Artificial Vision in 2018. Her current research interests include computer vision and face recognitions and applications.

Ezzeddine Zagrouba received his HDR from FST/University Tunis ElManar and his Ph.D. and engineering degree from the Polytechnic National Institute of Toulouse (ENSEEIHT/INPT) in France. He is a Professor at the Higher Institute of Computer Science (ISI). He is Vice President of Virtual University of Tunis and the Director of LimTic Research Laboratory at ISI. His main activity is focused on intelligent imaging and computer vision and he is vice president of the research association ArtsPi.