



A comparative study of human facial age estimation: handcrafted features vs. deep features

SE. Bekhouche^{1,2} · F. Dornaika^{2,3}  · A. Benlamoudi⁴ · A. Ouafi⁵ · A. Taleb-Ahmed⁶

Received: 16 May 2019 / Revised: 5 June 2020 / Accepted: 29 June 2020 /
Published online: 17 July 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

In recent times, the topic of human facial age estimation attracted much attention. This is due to its ability to improve biometrics systems. Recently, several applications that are based on the demographic attributes estimation have been developed. These include law enforcement, re-identification in videos, planned marketing, intelligent advertising, social media, and human-computer interaction. The main contributions of the paper are as follows. Firstly, it extends some handcrafted models that are based on the Pyramid Multi Level (PML) face representation. Secondly, it evaluates the performance of two different kinds of features that are handcrafted and deep features. It compares handcrafted and deep features in terms of accuracy and computational complexity. The proposed scheme of study includes the following three main steps: 1) face preprocessing; 2) feature extraction (two different kinds of features are studied: handcrafted and deep features); 3) feeding the obtained features to a linear regressor. In addition, we investigate the strengths and weaknesses of handcrafted and deep features when used in facial age estimation. Experiments are run on three public databases (FG-NET, PAL and FACES). These experiments show that both handcrafted and deep features are effective for facial age estimation.

Keywords Age estimation · Handcrafted features · Deep features · Support vector regression

1 Introduction

Humans live a certain period of time, and with the progress of time, the human facial appearance beside other parts shows some remarkable changes due to the aging progression. According to Berry et al. [8], we can recognize this progress in some different facial appearances between the different age groups such as the infants and young children have larger pupils, children's lips are redder and proportionately larger than lips of adults and

✉ F. Dornaika
fadi.dornaika@ehu.eus

baby's nose is typically small, wide, and concave. Predicting the age is a difficult task for humans and it is more difficult for computers, although the accurate age estimation is very important for some applications. Recently, several applications that exploit the exact age or the age group have emerged. The person's age information can lead to higher accuracy in establishing the user's identity for the traditional biometric identifiers which can be used in access control applications [23].

In this paper, we introduce an automatic age estimation scheme in facial image. The proposed scheme contains three phases: face registration, descriptor extraction and age estimation. The goal of face registration or alignment is to detect faces in images, normalize the 3D or 2D pose of each detected face, and then produce a cropped face image. This preprocessing step can be crucial since the processing stages rely on it. Thus, the first stage can influence the performance of the whole estimation process. The preprocessing phase can be challenging since it encounters many variations affecting face images. The extraction stage computes a set of features from the cropped face. These features can be given by either shallow texture descriptor or deep neural network. In the last phase, we fed the extracted features to a regressor to estimate the age.

The main contributions of this paper can be summarized in the followings:

- We provide extensive comparison of handcrafted-feature-based and deep-learning-based approaches methods.
- Extending some handcrafted-feature-based methods that are based on Pyramid Multi-Level face representation.
- A study of the computational cost of each method.

The remaining of the paper is organized as follows: In Section 2, we summarize the existing techniques of facial age estimation. We introduce our approach in Section 3. The experimental results are given in Section 4. In Section 5, we present the conclusion and some perspectives.

2 Background and related work

Facial age estimation is an important task in the domain of facial image analysis. It aims to predict the age of a person basing on his or her facial features. The predicted age can be an exact age (years) or age group (year range) [43]. Predicting age is a difficult task for humans and it is more difficult for computers, although accurate age estimation is very important for some applications. Most of the classic age estimation methods are reviewed in [14], and both classic and deep learning methods are covered and reviewed in [3]. In the literature, few approaches are studying facial age progression or facial age synthesis compared to the facial age estimation studies such as [51–54, 56]. From a general overview, the facial age estimation approaches can be categorized based on the face image representation and the estimation algorithm. The current methods for age estimation can be divided into two-dimensional (2-D) and three-dimensional (3-D) methods based on the dimensionality of the processed samples. We focus on 2-D images in this work. It is possible to further divide 2-D age estimation approaches into many categories. By adopting a simple categorization, age estimation approaches can be divided into three main types: anthropometric-based, handcrafted-feature-based and deep-learning-based approaches [5]. For this reason, we will focus on these three categories.

The anthropometry-based approaches mainly depend on measurements and distances of different facial landmarks. Kwon and Lobo [28] proposed an age classification method that classifies input images into one among three age groups: babies, young adults, and senior adults. Their method is based on craniofacial development theory and skin wrinkle analysis. The main theory in the area of craniofacial research is that the appropriate mathematical model to describe the growth of a person's head from infancy to adulthood is the revised cardioid strain transformation [2]. Hu et al. [22] take an image pair of the same person and derived the age difference by using Kullback-Leibler divergence is employed.

Handcrafted-feature-based approaches are one of the most popular approaches for facial age estimation since a face image can be viewed as a texture pattern. These approaches have been used in many computer vision applications due to their strengths such as fast and easy implementation, suitable for real-time applications, and low computational cost. On the other hand, they are vulnerable to profile faces and wild poses and are considered classic approaches. Many texture features have been used like Local Binary Pattern (LBP), Histograms of Oriented Gradients (HOG), Biologically Inspired Features (Bif), Binarized Statistical Image Features (BSif), and Local Phase Quantization (LPQ). LBP and its variants were also used by many works like in [4, 15, 48, 60]. Bif and its variants are widely used in age estimation works such as [18, 45]. Also, Guo et al. [18] investigated biologically inspired features comprised of a pyramid of Gabor filters in all positions in facial images and used either Support Vector Machine (SVM) or SVR with Radial Basis Function (RBF) kernels for evaluation. Liu et al. [33] propose an ordinal deep feature learning (ODFL) method to learn feature descriptors for face representation directly from raw pixels. Motivated by the fact that age labels are chronologically correlated and age estimation is an ordinal learning problem. Some researchers used multi-modal features. For instance, the work presented in [4] proposed an approach that used LBP and BSif extracted from Multi-Block face representation. Lanitis et al. [29] were the first to use Active Appearance Models (AAMs). Yang and Ai [59] used a real AdaBoost algorithm to train a strong classifier by composing a sequence of the local binary pattern (LBP) histogram features. They conducted experiments on gender, ethnicity and age classifications. Lu et al. [37] proposed a local binary feature learning method (CS-LBFL) to learn a face descriptor that is robust to local illumination. However, these methods aim to seek simple feature filters, so that they are not powerful enough to exploit the nonlinear relationship of face samples in such cases that facial images are exposed to large variances of diverse facial expressions and cluttered background. In [39], the authors classify the input face image into one of the demographic classes, then estimate age within the identified demographic class.

Deep learning approaches mainly use Convolutional Neural Networks (CNN) which is a type of feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex. Some approaches train the networks from scratch such as [30] and others do transfer learning such as [44]. Deep learning approaches are considered the best in achieving good and stable results. Also, their implementation is suited for real-time applications which are very important nowadays. Moreover, they are more immune to facial poses compared to classical approaches, yet there is a main downside which is the high computational cost because of the need for the graphics processing units (GPUs). In [20], the authors used VGGFace network to extract the deep features and the kernel Extreme Learning Machines to predict the age. Hu et al. [22] proposed deep architecture with multi-label loss function, their proposed multi-label loss

function composed of three losses which designed to drive their CNN to understand the age progressively. Gurpinar et al. [20] used Kernel Extreme Learning Machines to classify the age estimation and for features, the authors used pretrained deep CNN by using the features from a deep network that is trained on face recognition. Liu et al. [32] to estimate the age, they propose a group-aware deep feature learning (GA-DFL) approach. The same authors to improve the performance they designed a multi-path CNN to capture age-informative appearances from different scale information. Shen et al. [50] propose two Deep Differentiable Random Forests methods, for age estimation. Both methods connect split nodes to the top layer of the CNNs and deal with non homogeneous data by jointly learning input-dependent data partitions at the split nodes and age distributions at the leaf nodes. The name of the methods is Deep Label Distribution Learning Forest (DLDLF) and Deep Regression Forest (DRF). Dornaika et al. [12] to get better performance on age estimation, they used robust loss function for training deep network regression.

3 Proposed approach

In this section, we present the different stages of our approach that estimates the human age based on facial images. Our approach takes a face image as input. A face preprocessing stage is first applied to the image in order to obtain a cropped and aligned face. In the second stage, a set of features is extracted across a texture descriptor or a pretrained CNN. Finally, these features will be fed to a linear SVR in order to predict the age. Fig. 1 illustrates an overview of the proposed facial age estimation approaches.

We give the pseudocode of the facial age estimation pipeline algorithms starting by input image to output age. Algorithm 1 summarizes the different stages of the proposed facial age estimation approach using handcrafted features. Algorithm 2 summarizes the different stages of the second approach based on deep features.

Algorithm 1 Facial age estimation based on handcrafted features.

Input:

- A face image
- PML parameters: pyramid level and type of image descriptor

Output: Predicted age

```

1 for each face image do
2   - Convert it into a gray-scale image
3   - Detect face region
4   - Detect face landmarks
5   - Compensate the in-plane rotation and rectify the scale
6   - Crop the face
7   - Compute the PML representation
8   - Extract the image features in each block
9   - Concatenate the extracted features from different division of each pyramid level
10  - Estimate the age using the learned SVR regressor;
11 end

```

Algorithm 2 Facial age estimation based on deep features.

Input:

- A face image
- A pretrained CNN: last fully connected layer

Output: Predicted age

```

1 for each face image do
2   - Detect face region
3   - Detect face landmarks
4   - Compensate the in-plane rotation and rectify the scale
5   - Crop the face
6   - Feed the cropped face into the pretrained CNN
7   - Extract the deep features from the last fully connected layer
8   - Estimate the age using the learned SVR regressor;
9 end
    
```

3.1 Face preprocessing

Firstly, we apply the cascade object detector that uses the Viola-Jones algorithm [57] to detect people’s faces. This face detector uses the histogram of oriented gradients (HOG) features and a cascade of classifiers trained using boosting. Then, we detect the eyes of each face using dlib’s face landmark detector [26], which is an implementation of Kazemi et al. [25] that uses an ensemble of regression trees (ERT) to estimate the face’s landmark positions directly from a sparse subset of pixel intensities. To rectify the 2D face pose in the original image, we apply a 2D transformation based on the eyes center landmarks. Therefore, we align the face by rotating clockwise the face by an angle θ around the image center. Unlike the work described in [6], the cropping parameters are set as follows: $k_{side} = 0.9$, $k_{top} = 1.3$ and $k_{bottom} = 1.9$. These parameters are multiplied by a rescaled inter-ocular distance in order to obtain the side margins, the top margin and the bottom margin. The

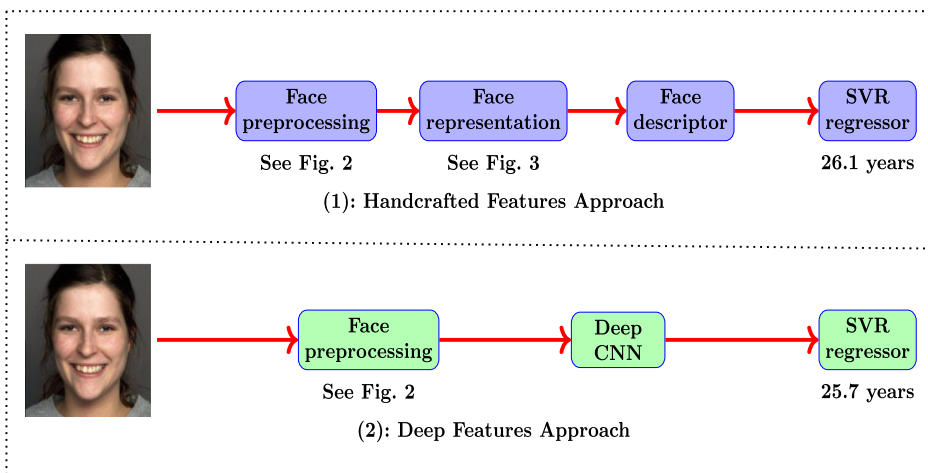


Fig. 1 The general structure of the proposed facial age estimation approaches

aligned and cropped images are resized to 224×224 . Figure 2 illustrates the steps of the face preprocessing stage.

3.2 Feature extraction

The feature extraction stage has been the most studied topic among the remaining stages due to its effective role in the age estimation systems performance. In our approach, we studied two different kinds of feature extraction methods. The first one is based on handcrafted features or texture features and the second one is based on deep features that are extracted using pretrained networks [11]. The main codes used for generating the features can be available upon request.

3.2.1 Handcrafted features

Refer to the attributes derived using generic purpose texture descriptors that use the information present in the image itself. In our case, we used three types of texture descriptors LBP, LPQ and BSIF on a Pyramid Multi-Level (PML) face representation.

Local Binary Pattern (LBP) is a very efficient method for analyzing two dimensional textures. It used the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. LBP was used widely in many image-based applications such as face recognition. The face can be seen as a composition of micro-patterns such as edges, spots and flat areas which are well described by the LBP descriptor [1].

In the following, we will present the LPQ and BSIF descriptors.

- Local phase quantization (LPQ): LPQ was originally proposed by Ojansivu and Heikkila [40]. LPQ is a texture descriptor based on the application of STFT. It uses the short-term Fourier transform STFT 2-D computed over a rectangular $M - by - M$ neighborhood N_x centered at each pixel position x of the image $f(x)$ defined by the formula (1).

$$F(u, x) = \sum_{y \in N_x} f(x - y)e^{-j2\pi u^T y} = w_u^T f_x \tag{1}$$

where w_u is the basis vector of the 2 - D DFT at frequency u (a 2D vector), and f_x is another vector containing all M^2 image samples from N_x .

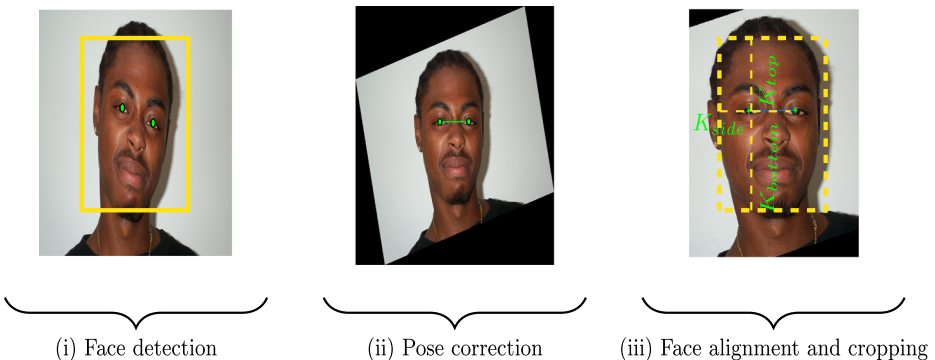


Fig. 2 Face preprocessing stage

In LPQ only four complex coefficients are considered, corresponding to 2-D frequencies: $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u_3 = [a, a]^T$, $u_4 = [a, -a]^T$, where a is a sufficiently small scalar. For each pixel the vector obtained is represented by the following formula:

$$F_x = [F(u_1, x), F(u_2, x), F(u_3, x), F(u_4, x)] \quad (2)$$

The phase information in the Fourier coefficients is recorded by observing the signs of the real and imaginary parts of each component in $F(x)$. This is done using a scalar quantization defined by this formula:

$$q_j = \begin{cases} 1 & \text{if } g_j \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

where g_j is the j th component of the vector $G(x) = [\text{Re}\{F(x)\}, \text{Im}\{F(x)\}]$. The resulting eight binary coefficients q_j represent the binary code pattern. This code is converted to decimal numbers between 0-255. From that, the LPQ histogram has 256 bins.

- Binarized statistical image feature (BSIF): The BSIF is an image texture descriptor proposed by Kannala and Rattu [24]. The idea behind BSIF is to automatically learn a fixed set of filters from a small set of natural images, instead of using hand-crafted filters such. The set of filters is learned from a training set of natural image patches by maximizing the statistical independence of the filter responses.

Given an image patch I of size $L \times L$ pixels and a linear filter W_k of the same size, the filter response S_k is obtained by:

$$S_k = \sum_{i,j} W_k(i, j) I(i, j) = W_k'^T I' \quad (4)$$

where W_k' and I' are vectors of size $L \times L$ (vectorized form of the 2D arrays W_k and I). The binarized feature b_k is obtained by:

$$b_k = \begin{cases} 1 & \text{if } S_k \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

The filters W_k are learnt using independent component analysis (ICA) by maximizing the statistical independence of S_k . The number of histogram bins (N_{bins}) obtained by the BSIF descriptor is calculated using this formula:

$$N_{bins} = 2^{N_f} \quad (6)$$

where N_f is the number of the filters used by BSIF.

3.2.2 Pyramid multi-Level (PML) representation

The Pyramid Multi-Level (PML) representation adopts an explicit pyramid representation of the original image. It preceded the image descriptor extraction. This pyramid represents the image at different scales. For each such level or scale, a corresponding Multi-block representation is used. PML sub-blocks have the same size which is determined by the image size and the chosen level. In our work, we use 7 levels based on [6] observation. Figure 3 illustrates the PML face representation adopting three levels.

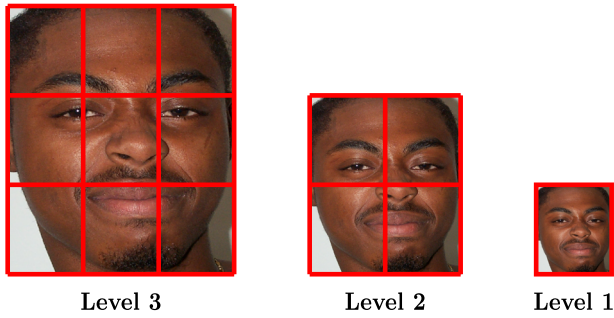


Fig. 3 PML face representation adopting three levels.

3.2.3 Deep features

Refer to descriptors that are often obtained from a CNN. These features are usually the output of the last fully connected layer. In our work, we used the VGG16 architecture [55] as well two variants of this architecture. We extract the deep features from the layer FC7 (fully connected layer) of this architecture, and the number of these features is 4096.

VGG16 [55] is a convolutional neural network that is trained on more than one million images from the ImageNet database. The network has an image input size of 224×224 and it is 16 layers deep and can classify images into 1000 object categories. It has approximately 138 million parameters, which makes it expensive to evaluate and use a lot of memory. As a result, the network has learned rich feature representations for a wide range of images. VGGFace [42] is inspired by VGG16, it was trained to classify 2,622 different identities based on faces. DEX-IMDB-WIKI [44] was fine-tuned on the IMDB-WIKI face database which has more than 500K images, yet it is also based on VGG16. Figure 4 presents the CNN architecture of VGG-16 as deep descriptor.

ResNet-50 [21] is a convolutional neural network that was trained on the ImageNet database. It is based on a residual learning framework where layers within a network are reformulated to learn a residual mapping rather than the desired unknown mapping between the inputs and outputs. It is similar in architecture to VGG16 network but with the additional identity mapping capability (see Fig. 5). VGGFace2 [9] is a variant of ResNet-50 which was trained on VGGFace2 database for facial recognition. ResNet-50 and VGGFace2 features have 2048 dimensions extracted from the global average pooling layer.

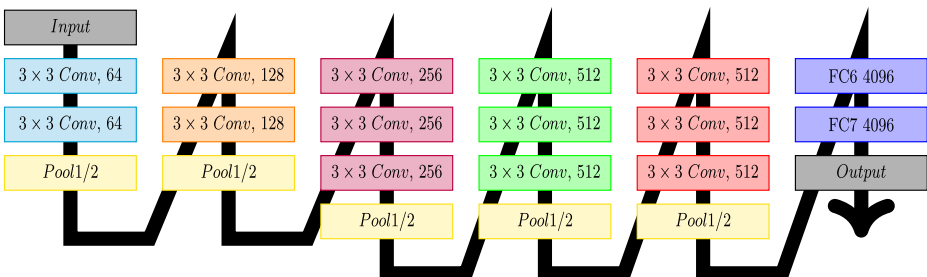


Fig. 4 VGG-16 CNN architecture

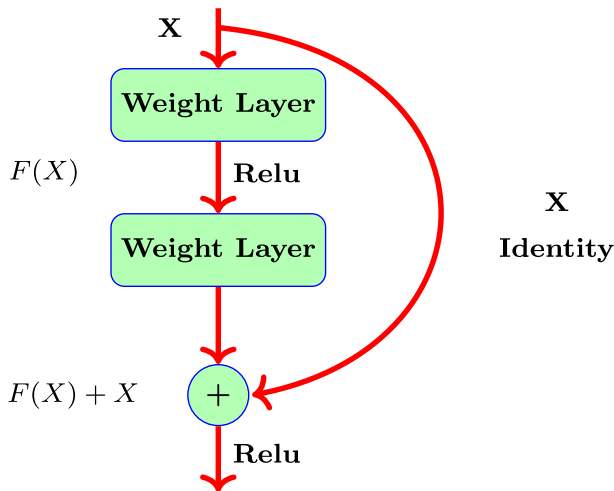


Fig. 5 ResNet residual block diagram with identity mapping

In our empirical study, we use the following deep features: VGG16, VGGFace, DEX-IMDB-WIKI, ResNet-50 and VGGFace2.

3.3 Age estimation

Once the facial image features are extracted, we need to predict the age from the extracted features. The proposed method estimates the person's age using a linear support vector machines regressor (SVR) with a ridge penalty based on L_2 norm, and optimizes the objective function using dual stochastic gradient descent (SGD) which reduces the computing time. Knowing that no feature selection was performed. This regressor was tuned to find the best configuration for its hyper-parameters that controls the trade off between a large margin and a small loss by trying to minimize the k-fold cross-validation loss.

4 Experiments and results

To evaluate the performance of the proposed approach, we use FG-NET, PAL and FACES databases. The performance is measured by the Mean Absolute Error (MAE) and the Cumulative Score (CS) curve.

The MAE and CS are two different indicators that are used for evaluating the performance of facial age estimation. The Mean Absolute Error (MAE) gives a global indicator about the performance in the sense that it summarizes the prediction errors over all tested images. It cannot quantify the number of successful predictions if a tolerance in age prediction error is used. On the other hand, the CS quantifies the number of test images that got a prediction error (years) that is smaller than a given threshold (this image is considered as an image with a correct prediction). In ideal cases, where the prediction coincides with the ground-truth age, this number should be equal to the total number of tested images and for

any value of the error threshold (tolerance) T . The MAE is the average of the absolute errors between the ground-truth ages and the predicted ones. The MAE equation is given by:

$$MAE = \frac{1}{N} \sum_{i=1}^N |p_i - g_i| \tag{7}$$

where N , p_i , and g_i are the total number of samples, the predicted age, and the ground-truth age respectively. The CS reflects the percentage of tested cases where the age estimation error is less than a threshold. The CS is given by :

$$CS(T) = \frac{N_{e \leq T}}{N} \% \tag{8}$$

where T , N and $N_{e \leq T}$ are the error threshold (years), the total number of samples and the number of samples on which the age estimation has an absolute error no higher than the threshold, T . Thus, the CS gives the percentage of the tested samples that are correctly predicted within the tolerance T .

4.1 FG-NET

The FG-NET [41] aging database was released in 2004 in an attempt to support research activities related to facial aging. Since then a number of researchers used the database for carrying out research in various disciplines related to facial aging. This database consists of 1002 images of 82 persons. On average, each subject has 12 images. The ages vary from 0 to 69. The images in this database have large variations in aspect ratios, pose, expression, and illumination. The Leave One Person Out (LOPO) protocol has been used due to the individual’s age variation in this database, each time a person’s images are put into a test set whereas the other persons’ images are put in a train set. Figure 6 shows the cumulative score curves for the eight different features. We can observe that the PML-LPQ descriptor performs better than the deep features in terms of CS which can be viewed as an indicator of the accuracy of the age estimators.

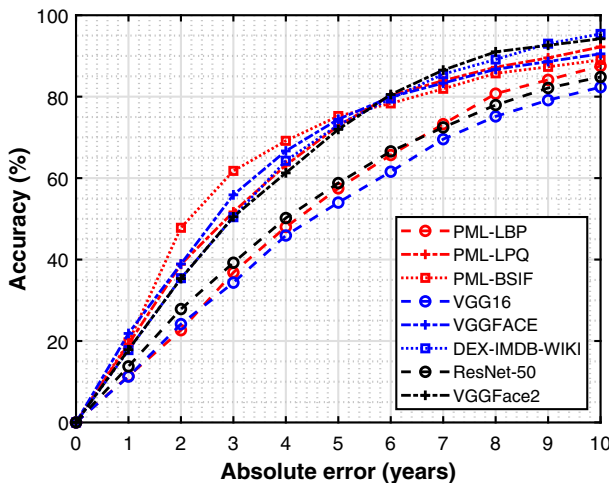


Fig. 6 Cumulative scores obtained by the proposed approach on the FG-NET database

Table 1 illustrates the MAE of the proposed approach as well as of that of some competing approaches. From this table, we can observe that the best deep-features DEX-IMDB-WIKI and the best handcrafted features PML-LPQ outperform some of the existing approaches. Moreover, we can see that the DEX-IMDB-WIKI features give the best results, followed closely by the VGGFace2 and PML-LPQ features. Based on the CS curve of Fig. 6 and Table 1, we can see that the MAE and the CS are two different indicators that do not always highlight the same best approach. Since the CS curve treats prediction errors by a gradually increasing (or decreasing) tolerance, it may happen that, when two methods are compared, a method having a worse MAE could provide a CS curve (or part of it) that is better than that of the method with a good MAE. Thus, when the tolerance is increased (or decreased), it is possible that the method with a worse MAE can be more accurate in prediction adopting that large (or small) tolerance, than the method having a good MAE. The case above can be seen in the FG-NET database results. In Fig 6, the CS curves of the two features PML-BSif (MAE = 4.48) and DEX-WIKI-IMDB (MAE = 3.74) are depicted. Despite the fact that DEX-WIKI-IMDB is globally better than PML-BSif, the latter has a better CS curve for small tolerances between one and five years.

Table 2 depicts the CPU time (in seconds) of the feature extraction stage and the training phase associated with 1002 images of FG-NET database. The experiments were carried out on the Alienware Aurora R8 workstation (Intel Core i9-9900K Processor, 16 Cache, 3.60 GHz, 64GB RAM, 2 × GPU GeForce RTX 2080, Windows 10). The handcrafted features significantly outperform the deep features in terms of CPU time execution of the feature

Table 1 Comparison with existing approaches on FG-NET database

	Approach	MAE
Published works	SVR+BSif+LBP (2014) [4]	6.34
	CS-LBFL (2015) [37]	4.43
	CS-LBMFL (2015) [37]	4.36
	AAM+GABOR+LBP (2016) [16]	4.87
	SWLD (2017) [17]	5.85
	GA-DFL + OHRanker (2017) [32]	3.93
	LSDML (2018) [34]	3.92
	DRF (2018) [49]	3.84
	M-LSDML (2018) [34]	3.74
	Deep_SRC+HSVR (2019) [31]	4.65
	LDMP (2019) [46]	4.60
	H-GPR (2019) [47]	4.41
	SSMCC (2019) [10]	4.25
	MA-SFV2 (2020) [35]	3.81
	Hand	PML-LBP
PML-BSif		4.48
PML-LPQ		4.10
Deep	VGG16	5.89
	ResNet-50	5.61
	VGGFace	4.30
	VGGFace2	3.88
	DEX-IMDB-WIKI	3.74

Table 2 CPU time (in seconds) of extracting features and training stages on the FG-NET database

	Feature Extraction	SVR training
PML-LBP	22.50	59.09
PML-LPQ	32.16	123.55
PML-BSif	28.05	145.25
VGG16	72.63	34.30
ResNet-50	144.57	10.84
VGGFace	72.65	39.02
VGGFace2	145.45	11.69
DEX-IMDB-WIKI	71.21	29.05

extraction stage knowing that the deep features are computed using the GPU instead of the CPU. On the other hand, the CPU time associated with the regressor training with the deep features is smaller than that of the regressor training using the handcrafted features. This is due to the fact that the size of the deep features is smaller than that of the handcrafted features. It is worthy noting that if dimensionality reduction or feature selection are applied on the features before the regression phase, the cost of the training of the latter will be the same for all types as long as the size of the final features is the same.

4.2 PAL

The Productive Aging Lab Face (PAL) database from the University of Texas at Dallas [38] contains 1,046 frontal face images from different subjects (430 males and 616 females) in the age range from 18 to 93 years old. The PAL database can be divided into three main ethnicities: African-American subjects 208 images, Caucasian subjects 732 images and other subjects 106 images. The database contains faces having different expressions. For the evaluation of the approach, we conduct 5-fold cross-validation, our distribution of folds is selected based on age, gender and ethnicity. Figure 7 shows the cumulative score curves

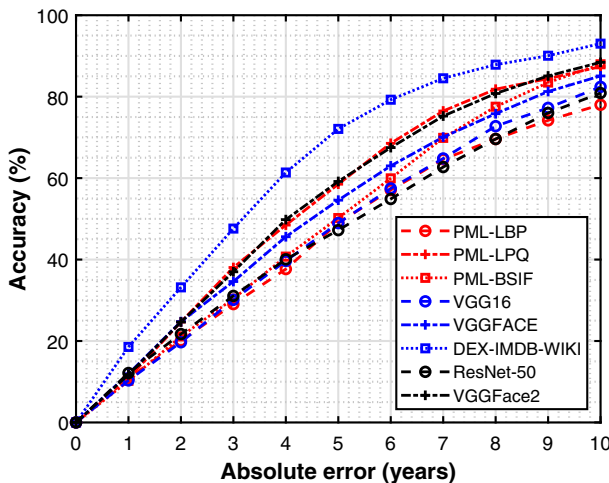


Fig. 7 Cumulative scores obtained by the proposed approach on the PAL database

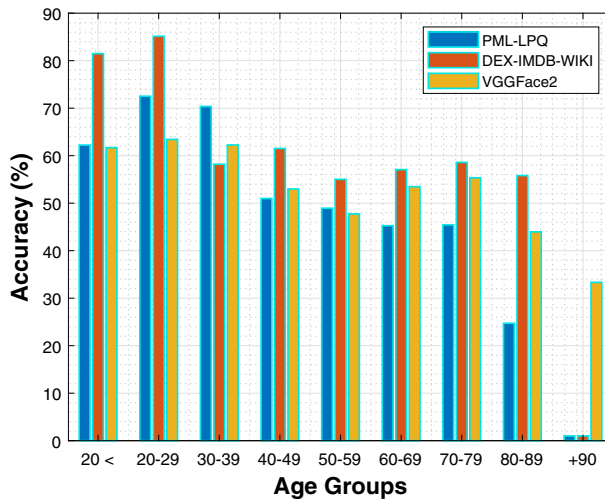


Fig. 8 Age estimation accuracy (CS) of different age groups when the absolute error is less than five years, i.e., the threshold T is set to five

for the eight different features. We can appreciate a change in the performance compared to FG-NET performance curve. It can be seen that the DEX-IMDB-WIKI features outperform the best handcrafted features approach which is PML-LPQ.

Figure 8 illustrates the accuracy of the age estimation associated with nine different age groups for the PAL dataset using the two best deep features and the best handcraft feature.

Table 3 Comparison with existing approaches on the PAL database

	Approach	MAE
Published works	Bif (2012) [19]	8.93
	Bif+MFA (2012) [19]	6.05
	SVR+BSif+LBP [4]	6.25
	CS-LBFL (2015) [37]	5.79
	CS-LBMFL (2015) [37]	5.26
	AAM+GABOR+LBP (2016) [16]	5.38
	SWLD (2017) [17]	6.68
	PML+BSif+LPQ (2017) [7]	5.00
	LS-SVM (2018) [36]	5.26
	TSE (2019) [39]	4.49
	Hand	PML-LBP
PML-LPQ		5.30
PML-BSif		5.43
Deep	ResNet-50	6.24
	VGG16	5.96
	VGGFace	5.51
	VGGFace2	5.11
	DEX-IMDB-WIKI	3.98

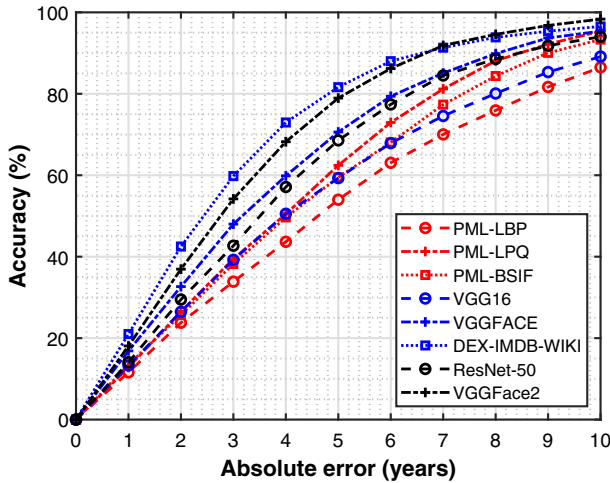


Fig. 9 Cumulative scores obtained by the proposed approach on the FACES database

These results were obtained by using the CS (2) with $T = 5$ (the used tolerance is five years) and for nine different subsets of test images that correspond to nine age groups.

We can observe that the used features (handcrafted and deep) are fairly robust to the variance of age groups in the PAL database.

Table 3 illustrates the MAE of the proposed approach as well as of that of some existing approaches. These results show that the DEX-IMDB-WIKI features outperform most of the existing approaches on the PAL database.

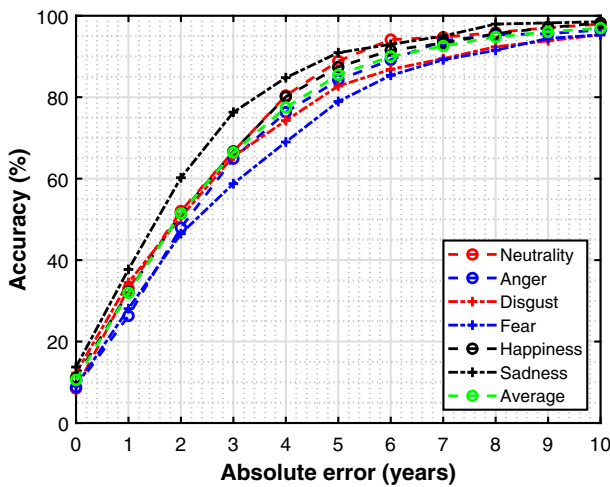


Fig. 10 Cumulative scores obtained by the DEX-IMDB-WIKI features for different expressions on the FACES database

Table 4 Comparison with existing approaches on FACES database

	Approach	Neutrality	Anger	Disgust	Fear	Happiness	Sadness	Average
Published works	Bif (2012) [19]	9.50	13.26	13.23	12.65	10.69	10.78	11.68
	Bif+MFA (2012) [19]	8.14	10.96	10.73	12.24	10.32	10.66	10.51
	LS-SVM (2018) [36]	5.97	8.21	8.17	8.25	6.77	7.07	7.41
	CS-LBFL (2015) [37]	5.06	6.94	7.15	6.32	6.53	6.27	6.38
	DeepRank+ (2018) [58]	4.61	6.48	7.50	5.90	5.92	5.30	5.95
	CS-LBMFL (2015) [37]	4.84	5.50	5.70	6.10	5.85	4.98	5.49
	DLF (2018) [27]	-	-	-	-	-	-	5.18
	LSDML (2018) [34]	3.88	3.87	4.41	5.10	3.49	4.09	4.14
Hand	PML-LBP	4.87	5.30	5.49	6.01	5.44	5.16	5.38
	PML-BSIf	4.01	4.63	4.51	4.99	4.66	4.67	4.58
	PML-LPQ	4.00	4.32	4.20	4.67	4.49	4.34	4.34
Deep	VGG16	4.44	5.61	4.62	5.73	5.06	4.27	4.95
	ResNet-50	3.81	4.40	3.99	4.44	4.22	3.85	4.12
	VGGFace	3.07	4.50	4.19	4.16	3.53	3.69	3.86
	VGGFace2	2.73	3.45	3.56	3.51	3.13	3.20	3.64
	DEX-IMDB-WIKI	2.96	3.36	3.34	3.66	3.04	2.56	3.15

4.3 FACES

This database [12] consists of 2052 images from 171 subjects. The ages vary from 19 to 80 years. For each subject, there are six expressions: neutral, disgust, sad, angry, fear, and happy. The database encounters large variations in facial expressions bringing an additional challenge for the problem of age prediction. Figure 9 shows the cumulative score curves for the eight different features. The results considered all images in all expressions. We can observe that the DEX-IMDB-WIKI features outperform all the other features. It is followed by the PML-LPQ features.

Figure 10 illustrates the cumulative score curves for the different facial expressions of the FACES database when using the DEX-IMDB-WIKI features. This figure shows that the neutral and sadness expressions correspond to the most accurate age estimation. In other words, among all tested expressions the neutral and sadness expressions are the ones that lead to the best age estimation.

Table 4 presents comparison with some existing approaches. This table confirms the idea that the neutral expression is the expression that provides the most accurate age estimation compared to other facial expressions.

5 Conclusion

This paper presents a study about using handcrafted and deep features for facial age estimation. Using a small number of images, the results showed that the handcrafted features sometimes gave better results than the deep features. Thus, it confirms that deep-learning-based approaches are not necessarily the best ones specially for low-cost less-accurate real-time visual analysis applications, so handcrafted-based approaches become more

appealing for this case. However, for applications requiring an accurate age prediction and robustness to facial poses, the deep-learning-based approaches are far better.

As it can be seen, the main strength of the handcrafted-based approaches is their relatively cheap computational cost associated with the training and testing phases. This allows them to be easily deployed on devices having limited hardware resources. Their main limitation is their possible dependency on an accurate face detection and alignment. On the other hand, despite the good accurate age prediction provided by the deep features, their main limitation is the expensive computational cost associated with the training and testing stages.

As future work, we envision the adaptation and fine-tuning of some recent CNN architectures on face databases and creating new CNNs from scratch. We also envision studying the effect of deep features extraction level on the performance of the facial age estimation by evaluating the difference between the extraction of low-level and high-level features from pretrained networks.

References

1. Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: Application to face recognition. *IEEE Trans Pattern Anal Mach Intell* (12) 2037–2041
2. Alley ST (1988) Applied aspects of perceiving faces resources for ecological psychology Lawrence Erlbaum associates
3. Angulu R, Tapamo JR, Adewumi AO (2018) Age estimation via face images: a survey. *EURASIP J Image Video Process* 2018(1):42
4. Bekhouche S, Ouafi A, Taleb-Ahmed A, Hadid A, Benlamoudi A (2014) Facial age estimation using bsif and lbp. In: *Proceeding of the first International Conference on Electrical Engineering ICEEB'14*
5. Bekhouche SE (2017) Facial Soft biometrics: Extracting demographic traits. PhD thesis, Faculté des sciences et technologies
6. Bekhouche SE, Dornaika F, Ouafi A, Taleb-Ahmed A (2017) Personality traits and job candidate screening via analyzing facial videos. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp 10–13
7. Bekhouche SE, Ouafi A, Dornaika F, Taleb-Ahmed A, Hadid A (2017) Pyramid multi-level features for facial demographic estimation. *Expert Syst Appl* 80:297–310
8. Berry DS, McArthur LZ (1986) Perceiving character in faces: the impact of age-related craniofacial changes on social perception. *Psychol Bull* 100(1):3
9. Cao Q, Shen L, Xie W, Parkhi OM, Zisserman A (2018) Vggface2: A dataset for recognising faces across pose and age. In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, IEEE, pp 67–74
10. Dong Y, Lang C, Feng S (2019) General structured sparse learning for human facial age estimation. *Multimed Syst* 25(1):49–57
11. Dornaika F, Arganda-Carreras I, Belver C (2019) Age estimation in facial images through transfer learning. *Mach Vis Appl* 30(1):177–187
12. Dornaika F, Bekhouche S, Arganda-Carreras I (2020) Robust regression with deep cnns for facial age estimation: an empirical study. *Expert Syst Appl* 112942:141
13. Ebner NC, Riediger M, Lindenberger U (2010) Faces—a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behav Res Methods* 42(1):351–362
14. Fu Y, Guo G, Huang TS (2010) Age synthesis and estimation via faces: a survey. *IEEE Trans Patt Anal Mach Intell* 32(11):1955–1976
15. Günay A, Nabiyeve VV (2008) Automatic age classification with lbp. In: *Computer and Information Sciences, 2008 ISICIS'08. 23rd International Symposium on*, IEEE, pp 1–4
16. Günay A, Nabiyeve VV (2016) Age estimation based on hybrid features of facial images. In: *Information Sciences and Systems 2015*. Springer, New York, pp 295–304
17. Günay A, Nabiyeve VV (2017) Facial age estimation using spatial weber local descriptor. *Int J Adv Telecommun Electrotechnics Signals Syst* 6(3):108–115

18. Guo G, Mu G, Fu Y, Huang TS, Human age estimation using bio-inspired features (2009). In: Computer Vision and Pattern Recognition, 2009. CVPR IEEE Conference on, IEEE, 2009, pp 112–119
19. Guo G, Wang X (2012) A study on human age estimation under facial expression changes. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, pp 2547–2553
20. Gurpinar F, Kaya H, Dibeklioglu H, Salah A (2016) Kernel elm and cnn based facial age estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 80–86
21. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
22. Hu Z, Wen Y, Wang J, Wang M, Hong R, Yan S (2016) Facial age estimation with age difference. *IEEE Trans Image Process* 26(7):3087–3097
23. Jain AK, Li SZ (2011) Handbook of face recognition, vol 1, Springer, New York
24. Kannala J, Rahtu E (2012) Bsf: Binarized statistical image features. In: Pattern Recognition (ICPR), 2012 21st International Conference on, IEEE, pp 1363–1366
25. Kazemi V, Sullivan J (2014) One millisecond face alignment with an ensemble of regression trees. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp 1867–1874
26. King DE (2009) Dlib-ml: a machine learning toolkit. *J Mach Learn Res* 10:1755–1758
27. Kotowski K, Stapor K (2018) Deep learning features for face age estimation: Better than human? In: International Conference: Beyond Databases, Architectures and Structures. Springer, New York, pp 376–389
28. Kwon YH, da Vitoria Lobo N (1999) Age classification from facial images. *Comput Vision Image Understand* 74(1):1–21
29. Lanitis A, Draganova C, Christodoulou C (2004) Comparing different classifiers for automatic age estimation. *IEEE Trans Syst Man Cybernet Part B (Cybernetics)* 34(1):621–628
30. Levi G, Hassner T (2015) Age and gender classification using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp 34–42
31. Liao H (2019) Facial age feature extraction based on deep sparse representation. *Multimed Tools Appl* 78(2):2181–2197
32. Liu H, Lu J, Feng J, Zhou J (2017) Group-aware deep feature learning for facial age estimation. *Pattern Recogn* 66:82–94
33. Liu H, Lu J, Feng J, Zhou J (2017) Ordinal deep learning for facial age estimation. *IEEE Trans Circ Syst Video Technol* 29(2):486–501
34. Liu H, Lu J, Feng J, Zhou J (2018) Label-sensitive deep metric learning for facial age estimation. *IEEE Trans Inform Forensics Secur* 13(2):292–305
35. Liu X, Zou Y, Kuang H, Ma X (2020) Face image age estimation based on data augmentation and lightweight convolutional neural network. *Symmetry* 12(1):146
36. Lou Z, Alnajjar F, Alvarez JM, Hu N, Gevers T (2018) Expression-invariant age estimation using structured learning. *IEEE Trans Pattern Anal Mach Intell* 40(2):365–375
37. Lu J, Liong VE, Zhou J (2015) Cost-sensitive local binary feature learning for facial age estimation. *IEEE Trans Image Process* 24(12):5356–5368
38. Minear M, Park DC (2004) A lifespan database of adult facial stimuli. *Behav Res Methods Instruments Comput* 36(4):630–633
39. nadhir Zighem M-E, Ouafi A, Zitouni A, Ruichek Y, Taleb-Ahmed A (2019) Two-stages based facial demographic attributes combination for age estimation. *J Vis Commun Image Represent* 61:236–249
40. Ojansivu V, Heikkilä J (2008) Blur insensitive texture classification using local phase quantization. In: International conference on image and signal processing. Springer, New York, pp 236–243
41. Panis G, Lanitis A, Tsapatsoulis N, Cootes TF (2016) Overview of research on facial ageing using the fg-net ageing database. *Iet Biometrics* 5(2):37–46
42. Parkhi OM, Vedaldi A, Zisserman A et al (2015) Deep face recognition. In: BMVC, vol 1, p 6
43. Petra G (2013) Introduction to human age estimation using face images. *J Slovak Univ Technol* 21:24–30
44. Rothe R, Timofte R, Gool LV (2015) Dex: Deep expectation of apparent age from a single image. In: IEEE International Conference on Computer Vision Workshops (ICCVW)
45. Sai P-K, Wang J-G, Teoh E-K (2015) Facial age range estimation with extreme learning machines. *Neurocomputing* 149:364–372
46. Sawant M, Addepalli S, Bhurchandi K (2019) Age estimation using local direction and moment pattern (ldmp) features. *Multimed Tools Appl* 78(21):30419–30441
47. Sawant MM, Bhurchandi K (2019) Hierarchical facial age estimation using gaussian process regression. *IEEE Access* 7:9142–9152
48. Shan C (2010) Learning local features for age estimation on real-life faces. In: Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis, ACM, pp 23–28

49. Shen W, Guo Y, Wang Y, Zhao K, Wang B, Yuille A (2018) Deep regression forests for age estimation. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition
50. Shen W, Guo Y, Wang Y, Zhao K, Wang B, Yuille AL (2019) Deep differentiable random forests for age estimation. *IEEE Trans Pattern Anal Mach Intell*
51. Shu X, Tang J, Lai H, Liu L, Yan S (2015) Personalized age progression with aging dictionary. In: Proceedings of the IEEE international conference on computer vision, pp 3970–3978
52. Shu X, Tang J, Lai H, Niu Z, Yan S (2016) Kinship-guided age progression. *Pattern Recogn* 59:156–167
53. Shu X, Tang J, Liu L, Niu Z, Yan S (2015) What shall i look like after n years? In: Proceedings of the 23rd ACM international conference on Multimedia, pp 789–790
54. Shu X, Xie G-S, Li Z, Tang J (2016) Age progression: Current technologies and applications. *Neurocomputing* 208:249–261
55. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
56. Tang J, Li Z, Lai H, Zhang L, Yan S et al (2017) Personalized age progression with bi-level aging dictionary learning. *IEEE Trans Pattern Anal Mach Intell* 40(4):905–917
57. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the IEEE Computer Society Conference on, volume 1, pages I–511–I–518 vol. 1, p 2001
58. Yang H-F, Lin B-Y, Chang K-Y, Chen C-S (2018) Joint estimation of age and expression by combining scattering and convolutional networks. *TOMCCAP* 14(1):9–1
59. Yang Z, Ai H (2007) Demographic classification with local binary patterns. In: International Conference on Biometrics. Springer, New York, pp 464–473
60. Ylioinas J, Hadid A, Pietikäinen M (2012) Age classification in unconstrained conditions using lbp variants. In: Pattern recognition (icpr), 2012 21st international conference on, IEEE, pp 1257–1260

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

SE. Bekhouche^{1,2} · F. Dornaika^{2,3}  · A. Benlamoudi⁴ · A. Ouafi⁵ · A. Taleb-Ahmed⁶

SE. Bekhouche
s.bekhouche@univ-djelfa.dz

A. Benlamoudi
benlamoudi.azeddine@univ-ouargla.dz

A. Ouafi
a.ouafi@univ-biskra.dz

A. Taleb-Ahmed
taleb@uphf.fr

¹ Department of Electrical Engineering, University of Djelfa, Djelfa, Algeria

² University of the Basque Country UPV/EHU, San Sebastian, Spain

³ IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

⁴ Faculté des Nouvelles Technologies de l'information et de la communication, Laboratoire de Génie Électrique (LAGE), University of Kasdi Merbah Ouargla, Ouargla, 30000, Algeria

⁵ Laboratory of LESIA, University of Biskra, Biskra, Algeria

⁶ IEMN DOAE UMR CNRS 8520, UPHF, 59313, Valenciennes, France