Facial age estimation using tensor based subspace learning and deep random forests

O. Guehairia, F. Dornaika, A. Ouamane, Abdelmalik Taleb-Ahmed

Information Sciences 609 (2022) 1309–1317

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Facial age estimation using tensor based subspace learning and deep random forests

O. Guehairia [b,c], F. Dornaika [a,d,e,*], A. Ouamane [c], A. Taleb-Ahmed [f]

[a] Henan University, Henan Key Lab of Big Data Analysis and Processing, Kaifeng, China
[b] Laboratory of LESIA, University of Biskra, Biskra, Algeria
[c] Laboratory of LI3C, University of Biskra, Biskra, Algeria
[d] University of the Basque Country UPV/EHU, San Sebastian, Spain
[e] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[f] Institut d'Electronique de Microélectronique et de Nanotechnologie (IEMN), UMR 8520, Université Polytechnique Hauts de France, Université de Lille, CNRS, 59313 Valenciennes, France

ARTICLE INFO

Article history:
Received 3 February 2022
Received in revised form 22 July 2022
Accepted 23 July 2022
Available online 29 July 2022

Keyword:
Age estimation
Deep features
Multi-linear whitened principal component
Tensor exponential discriminant analysis
Tensor based subspace
Deep random forests

ABSTRACT

Recently, the estimation of facial age has attracted much attention. This letter extends and improves a recently developed method (Guehairia et al., 2020) for fusing multiple deep facial features for age estimation. This method was based on deep random forests. We propose a new pipeline that integrates tensor-based subspace learning before applying DRFs. Deep face features of a training set are represented as a 3D tensor. Multi-linear Whitened Principal Component (MWPCA) and Tensor Exponential Discriminant (TEDA) are used to extract the most discriminative information. The tensor subspace features are then fed into DRFs to predict age. Experiments conducted on five public face databases show that our method can compete with many state-of-the-art methods.

1. Introduction

Face age estimation is one of the biometric technologies that have evolved greatly in recent years [22,40,32,9,23,34,37]. This technology can be used variously in the real world applications within various domains, such as Human Computer Interaction (HCI), security and management applications. Forensic Investigations[10,15]. The aging process of faces works according to some well-known aging modes. When children grow up, their shape changes significantly due to the growing skull. On the other hand, the aging process in adulthood can be identified by changes in facial skin texture as the physical appearance changes. These are described by deep wrinkles, skin that peels off with time, and spots on all parts of the face [14]. In addition to the complexity of facial features and age, the aging process can also be impacted by other factors such as gender, genes, race, health status, and life circumstances. In the field of image-based age estimation, the acquisition of facial images is considered a challenge due to its difficulty and tediousness. In the field of image-based age estimation, the acquisition of facial images is considered a challenging process due to its difficulty and tediousness. It is dilemmatic to find out the requirements for a comprehensive research, because the public age data of faces may have some limitations, such as unbalanced representation of age, gender, and ethnicity. These difficulties show that age estimation research faces major obstacles.

* Corresponding author at: Henan Key Lab of Big Data Analysis and Processing, Henan University, Kaifeng, China.
E-mail address: fadi.dornaika@ehu.eus (F. Dornaika).

https://doi.org/10.1016/j.ins.2022.07.135
0020-0255/

Although face age estimation technology faces many challenges, there are many different applications in the fields of close observation and inspection systems, information management, human–computer interaction, and entertainment, etc.

Modern Convolutional Neural Networks (CNNs) are becoming more extensive and complicated [13,5]. Despite their success in estimating the age of faces in images, they still have some weaknesses: (i) the choice of architecture of the network, (ii) high computational cost, (iii) limited portability between different datasets, and (iv) the need for a very large dataset for training.

In this letter, we propose a new approach that can solve many of the above problems.

In our proposed approach, we use pre-trained CNN models to extract features from facial images. These features are provided by multiple networks and used as input features to our estimator. The latter consists of tensor transformations and deep random forests.

So far, the subspace transformation is the most commonly used technique for dimension reduction. Recently, several dimension reduction algorithms have been proposed that are suitable for feature extraction [35,39]. Principal component analysis (PCA) [1] and linear discriminant analysis LDA [30] are commonly used. They are linear subspace techniques. Essentially, an image face is a matrix of $m \times m'$ pixels treated as a 1-D feature vector of size $m \times m'$. Unfortunately, this method loses the position information of the pixels. Recently, multilinear subspace techniques based on tensor analysis of data in high-dimensional spaces have been considered as a remarkable multi-linear technique [27]. These approaches allow the preservation of the important information about the face structure. Multilinear transformations analyze the multifactorial structure of facial images over different index numbers.

The common linear subspace methods PCA and LDA are extended to Multilinear PCA (MPCA) [27] and Multilinear Discriminant Analysis (MDA) [36], which allow to manipulate the mathematical tensors. The high tensor order (i.e. >2) is presented in a normal form to show the set of face images without collapsing the original structure and correlation of the data [28] [8]. In [29], the authors propose a new application of an adopted MPCA, Multilinear Whitened Principal Component Analysis (MWPCA), which can solve the problem of small sample size in high-dimensional space and improve the hard discrimination obtained by classical MPCA. Multilinear Variational Analysis MDA has also been extended to Tensor Exponential Discriminant Analysis TEDA to improve the discriminant data contained in the null space of the within-class scatter matrix of each tensor mode. TEDA increases the distance between samples belonging to multiple classes by distance-diffusion mappings. To our knowledge, there are no studies on age estimation from face images using MWPCA, TEDA, and Deep Random Forest (DRF) tensor projection methods.

The main contributions of this work are as follows:

- We propose a multiview feature fusion [20] that improves the performance of our method previously proposed in [17]. Our proposed method also leverages the techniques in [29].
- We fuse the deep features using Whitened Principal Component Analysis (MWPCA) and Tensor Exponential Discriminant Analysis (TEDA), respectively.
- Once the facial image features are represented in the tensor subspace, the final age is estimated using our new Deep Random Forests (DRF) [17].

2. Building Blocks of the Proposed Method

In this section, we describe the main modules of our pipeline. The latter consists of two parts (see Fig. 1). The first part consists of Multilinear Whitened PCA (MWPCA) followed by Tensor Exponential Discriminant Analysis (TEDA) [29]. The second part performs regression by classification using deep random forests [17] that map features of the tensor space to a predicted age.

2.1. Multilinear whitened PCA (MWPCA):

In [29], the authors introduce MWPCA. MWPCA is an extension of MPCA to improve data representation in tensor space. To achieve this, the training tensor dataset is centered by subtracting the average tensor from the training sample in a preprocessing step. Then, in an initialization step, the covariance matrix and its eigenvalue decomposition are computed. This allows for the whitening of each tensor, which consists of normalizing each eigenvector by the square root of its corresponding eigenvalue. In this way, the data are less correlated and their variation is uniform in all directions. After initialization, the whitening of each mode of the tensor patterns is performed by an iterative local optimization step of the projection matrices until the maximum number of iterations is reached or the difference of the projected tensors between two consecutive iterations becomes smaller than a predefined threshold. The process is performed with the set of tensor samples $\mathbf{A}_i \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_m}$ $(i = 1, \ldots, n)$, the number $n_{(k)}$ of selected eigenvectors for each k-mode, the itr_{max}, which is the maximum number of iterations and the threshold η. MWPCA generates the projection matrices of the modes as well as the projected tensor $\widetilde{\mathbf{A}}_i \in \mathbb{R}^{I'_1 \times I'_2 \times \cdots I'_m}$ which is the new representation of the original tensor $\mathbf{A}_i$.

In [29], the authors introduce MWPCA. MWPCA is an extension of MPCA to improve data representation in tensor space. To achieve this, the training tensor dataset is centered by subtracting the average tensor from the training sample in a preprocessing step. Then, in an initialization step, the covariance matrix and its eigenvalue decomposition are computed. This

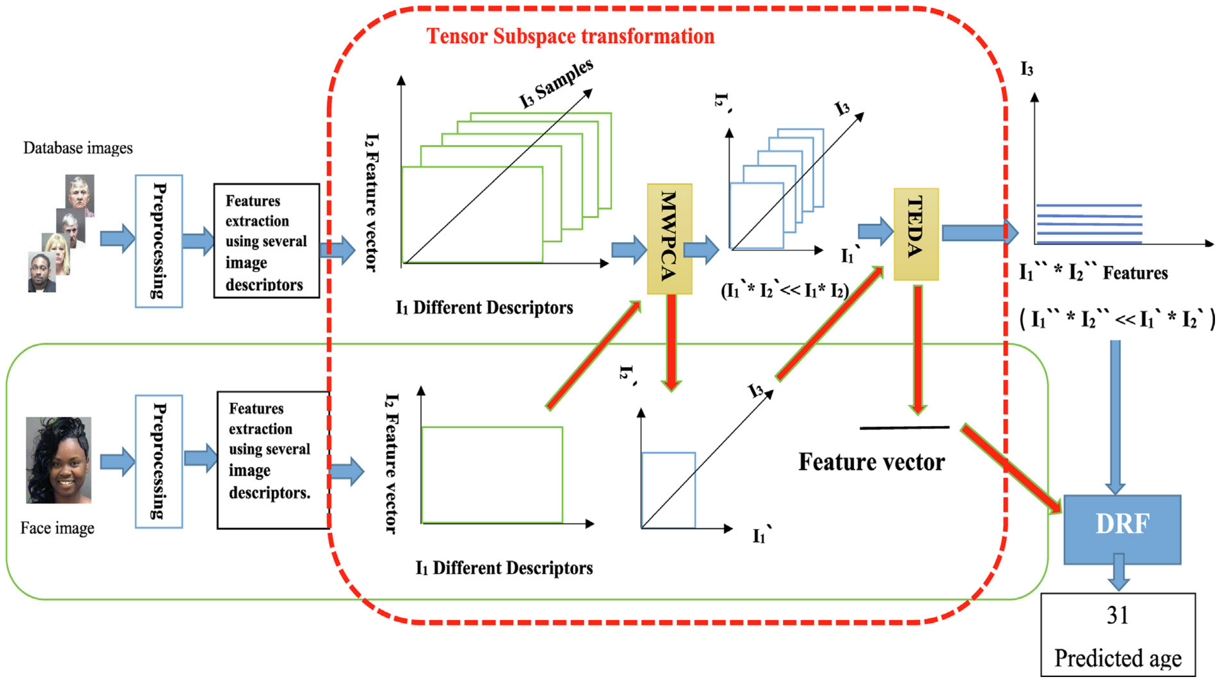

Fig. 1. Illustration of the proposed architecture. The model is given by the MWPCA, TEDA, and DRFs. Each test image is fed into this pipeline (red arrows) to obtain the associated age.

allows for the whitening of each tensor, which consists of normalizing each eigenvector by the square root of its corresponding eigenvalue. This makes the data less correlated and has a uniform variance in all directions. After initialization, the whitening of each mode of the tensor patterns is performed by an iterative local optimization step of the projection matrices until the maximum number of iterations is reached or the difference of the projected tensors between two consecutive iterations becomes smaller than a predefined threshold. The process is performed with the set of tensor samples $\mathbf{A}_i \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_m}$ $(i = 1, \ldots, n)$, the number $n_{(k)}$ of selected eigenvectors for each k-mode, the $\mathrm{itr}_{\max}$, which is the maximum number of iterations and the threshold η. MWPCA generates the mode projection matrices as well as the projected tensor $\widetilde{\mathbf{A}_i} \in \mathbb{R}^{I'_1 \times I'_2 \times \cdots I'_m}$ which is the new representation of the original tensor $\mathbf{A}_i$.

2.2. Tensor Exponential Discriminant Analysis (TEDA):

For the TEDA, presented in [29], the input is a tensor generated by the previous MWPCA, which is defined as $\widetilde{\mathbf{A}} \in \mathbb{R}^{I'_1 \times I'_2 \times \cdots I'_m \times n}$ of the n training samples belonging to L classes, each class $l = 1, \ldots, L$ has a tensor $\widetilde{\mathbf{A}_l}$ containing n_l samples, the itr_{max} is the maximum number of iterations and the final lower dimensions $I''_1 \times I''_2 \ldots I''_m$.

From these inputs, TEDA estimates the projection matrices U_k. TEDA includes the null space of the within-class scatter matrices of each tensor mode. In addition, TEDA increases the distance between samples belonging to different classes via distance-diffusion mappings.

2.3. Deep Random Forests for age estimation [17]:

Our previous method [17] targets the age estimation problem and predicts age based on a single facial image. It performs regression by classification. In this subsection, we briefly describe its principle.

The goal of the method RF is to generate multiple predictors before combining their different predictions, rather than trying to obtain an optimized method all at once. For more details on random forests, see [7]2. In [17], we put into practice a new approach that can solve the problem of age estimation from facial images (Deep Random Forests). It has shown adequate performance compared to the state of the art. It consists of ensembles of random forests. The ensembles of Random Forests form a cascade structure by forming more than one layer. An ensemble of Random Forests forms one layer in this structure. The feature vector is received from the first layer as a given input. A class probability distribution is generated by each forest in the same layer. An L-dimensional class vector is the output of each forest if there are L classes to predict. By concatenating the original input vector with the generated class vectors of each forest (from the previous layer), the input vector for subsequent layers is obtained.

3. Proposed approach

In this section we will be elaborating on the proposed architecture. It consists of two main steps: the dimension reduction of the tensor and Deep Random Forest for the final age estimation. These two main parts are performed during a training phase, where the first part (tensor transformations) is first determined and then the second part (DRFs) is estimated. We assume that each face image has several types of deep descriptors. These feature vectors form a 2D matrix, which is considered as the feature matrix of a face image. Thus, each face image is represented by a 2D tensor. A training set of face images can be represented by a 3D tensor. Then, the 3D tensor is successively treated by MWPCA and TEDA techniques to reduce the dimensions while preserving the essential components. The output from TEDA is used as input to the Deep Random Forests method for the final estimation. In the next section, we will present the implementation details of the proposed method.

We collect several types of descriptors for each facial image. The purpose of using these different feature types is to utilise the different types of information to improve the age estimation process. After extracting the different feature vectors of the face image, we reshape them into a 2D matrix of size $I_1 \times I_2$, where I_1 is the number of feature vectors used and I_2 is the dimension of the feature vectors.

The optimal multi-linear projection matrices are estimated in the training phase. After the model is computed, each new face image can be projected in the testing phase by the above-mentioned tensor transformations.

The training $\mathbf{X} \in \mathbb{R}^{\mathrm{I}_1 \times \mathrm{I}_2 \times \mathrm{I}_3}$ is constructed using the feature vectors extracted from the preprocessed face images of the training database. The feature vectors can be of different types. However, they should have the same dimension I_2.

The modes of the tensor $\mathbf{X} \in \mathbb{R}^{\mathrm{I}_1 \times \mathrm{I}_2 \times \mathrm{I}_3}$ are:

- I_1 denotes the number of the descriptors.
- I_2 denotes the dimension of the feature vector.
- I_3 denotes the number of training face samples.

The transformations of the input 3D tensor $\mathbf{X}$ are estimated based on the MWPCA/TEDA methods. The I_1 and I_2 modes are projected into another subspace. Hence, a new reduced tensor $\mathbf{Y} \in \mathbb{R}^{\mathrm{I}''_1 \times \mathrm{I}''_2 \times \mathrm{I}_3}$, where $\mathrm{I}'_1 \times I''_2 < \mathrm{I}_1 \times \mathrm{I}_2$. For further processing, these transformed features are reshaped (flattened) into a vector of dimension $\mathrm{I}''_1 \times I''_2$. The same procedure is applied consistently upon the test samples. Having obtained the reduced tensor, we invoke the second learning module, namely the deep random forests proposed in [17,16], where an ensemble of random forests interacts in the form of a cascade structure. The input is a feature vector. This is processed by multiple random forests. This collection of RFs is considered as the first layer, and the Deep Random Forest consists of multiple layers.

Each RF of the same layer generates a class probability vector. The generated class probability vectors are concatenated with the input vector to form the input vector of the next layer. In this way, a new feature vector with more information is created. The dimension of the new vector (in the first layer) is given by:

$$Dim = D_1 + (F \times L) \tag{1}$$

where D_1 denotes the original feature size, F the number of forests, and L the number of classes.

The output of the first layer is the input for the second layer, up to the last layer (the number of layers is a user-selected parameter). In the last layer, the vectors of probabilities are averaged to obtain a final vector of class probabilities. The final estimated class is the average of the N largest class probabilities. This is suitable for the age estimation problem. For more details, see [17].

4. Experiments and implementation details

We used five datasets to test the performance of our proposed architecture: MORPH II (with 55,608 images and 5-fold random partitioning), FG-NET (with 1002 images and LOPO protocol), PAL (1,046 images with 5-fold random partitioning), LFW+ (with 15,699 images and 5-fold cross validation), APPA-REAL (real age labels of 7591 images and 5-fold cross validation).

4.1. Pre-processing

We used the Ensemble of Regression Trees (ERT) algorithm [21]to locate the landmarks on the face. colorblueThis latter is considered a good algorithmfor face feature detection. The landmarks are used to align the 2D face image based on the eye coordinates. After the 2D alignment, the face region is cropped.

4.2. Feature extraction

We used the pre-trained models IMBD-WIKI and DEXchalearn to extract deep facial features [31]. We extract the last two fully connected layer vectors of the above pre-trained models FC6 and FC7 of the pre-processed input images with size

224×224. For each input face image, the FC6 and FC7 vectors of both models are later extracted as mentioned above to create a 2D matrix feature of size 4096×4.

4.3. Evaluation Metric

To evaluate the performance of the proposed age estimation method, we used the Mean Absolute Error (MAE). It is one of the most well-known indicators for evaluating the performance of age estimators in the literature. MAE calculates the average of the absolute error between the predicted age and the actual age. It is given by:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|p_t - g_t| \tag{2}$$

where n is the number of images tested, p_t is the predicted age of image t, and g_t is the actual age of that image.

4.4. Implementation

After obtaining the feature vectors from the pre-trained model, we extract them to create a matrix of 4096×4. We then extract the matrices of all the training samples to form a tensor of order 3. The training data from each database is used to estimate the projection matrices for the subspace projection. There are two matrices for the MWPCA method and two matrices for the TEDA method. The dimension was determined automatically by keeping 97% of the energy of the eigenvalues. To monitor the convergence of the 3rd order tensor projection, the maximum number of iterations is empirically set to 16. For the MWPCA algorithm, we set the convergence threshold to 10^6, as done in [29]. For the TEDA method, we changed the class labels by combining the closest labels (ages) into a single class, which can be considered as grouping the ages. The range of years in each group is the class width, which is considered as a hyperparameter in our work. The values are selected from the set $\{1, 2, 3, 4, 5\}$, resulting in the class widths in the set $\{1, 2, 3, 4, 5\}$ in years. Note that the DRF part uses the normal labels (i.e., the class is indicated by a year) and uses two levels (including the decision level). The other settings of the DRFs are similar to those described in [17].

4.5. Results

Tables 1 and 2 show an ablation study of our proposed architecture, which generally consists of three modules. Each time we excluded one module to observe its impact on the overall architecture. In each table, we used one of the mentioned space reduction methods (MWPCA or TEDA) with multiple decision parts: Deep Random Forest (DRF), Random Forest (RF) and Support Vector Machine (SVM) for age classification.

In Table 1, we eliminated the MWPCA module and used the TEDA approach with DRF, RF, and SVM for the final classification. TEDA has the potential to decrease the intra-class distance and increase the inter-class distance. This process may take more than one iteration. In the same table, we give the number of iterations to show its impact. In Table 2, we eliminated the TEDA module and used the MWPCA approach with DRF, RF, and SVM for the final classification. In Tables 1 and 2, the input representation for the decision part is created from the concatenation or fusion of the projected original tensor. The fusion used is the arithmetic mean of all tensor vectors.

From Tables 1 and 2 it can be seen that TEDA is generally better than MWPCA, but the proposed overall architecture, i.e. the combination of all modules (MWPCA + TEDA + DRF), gives the best results, as can be seen from Table 3 using the same dataset PAL, discussed in more detail below.

Fig. 2 shows all experimental results obtained with the five databases. Subplots (a), (b), (c), (d), and (e) show the MAE in years as a function of two hyperparameters: the number of highest probabilities and the class width in the TEDA method. The X axis corresponds to the number of highest probabilities used by the last layer of DRFs. The Y axis corresponds to the class width (in years) in the TEDA method. Each integer on this axis indicates the number of years used to form the age groups in the method TEDA. For example, a value of 2 means that the age groups are formed from classes with a width of two years.

In Fig. 2, we can see that TEDA generally performs better when the class width is set to 2 years. We also note that, in general, the best performance is obtained when the number of highest probabilities is greater than one. In (a) we see that the

Table 1
MAE in years obtained by using the TEDA module only with different classifiers. Results correspond to the PAL dataset.

	Nbr Iteration	TEDA + DRF	TEDA + RF	TEDA + SVM
PAL	8	2.95	3.26	3.1
	10	2.93	3.21	3.02
	12	2.74	3.09	2.92
	16	2.71	3.13	2.91
	18	2.73	2.98	2.84

Table 2
MAE in years obtained by using the MWPCA module only with different classifiers. Results correspond to the PAL dataset.

	MWPCA + DRF Concatenation	MWPCA + DRF Fusion	MWPCA + RF Concatenation	MWPCA + RF Fusion	MWPCA + SVMConcatenation	MWPCA + SVM Fusion
PAL	3.06	2.91	3.14	3.05	3.20	3.18

Table 3
MAE in years obtained by using the two modules (MWPCA + TEDA) with different classifiers. Results correspond to the PAL dataset.

	MWPCA + TEDA + DRF	MWPCA + TEDA + RF	MWPCA + TEDA + DRF
PAL	2.39	2.87	2.74

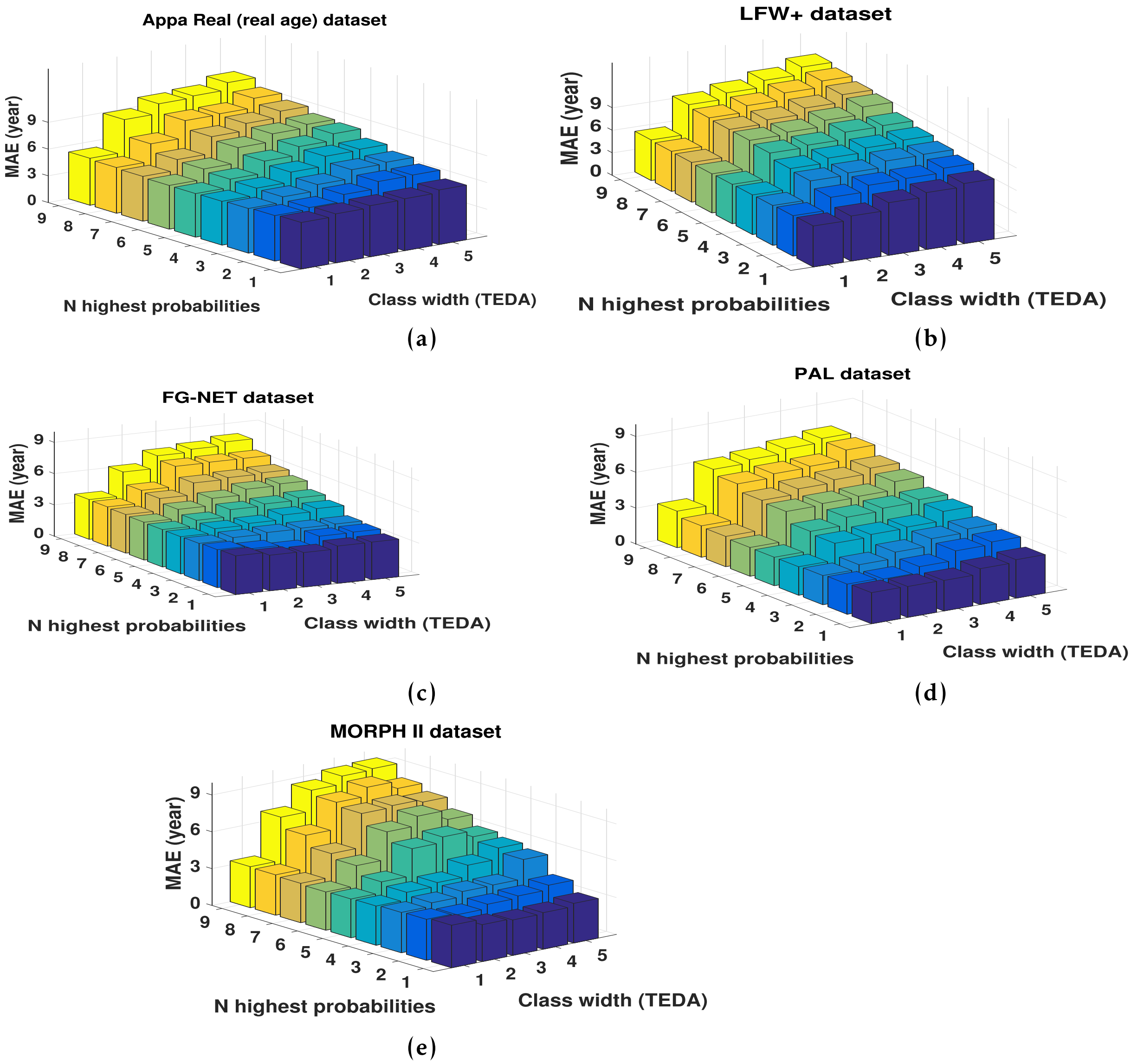

Fig. 2. MAE (years) of the proposed method as a function of two hyper-parameters: (i) the number of highest probabilities used by the last layer in the Deep Random Forest and (ii) the class width **(in years)** used by the TEDA method.

Table 4
Comparison of our method with some state-of-the-art methods using five datasets FG-NET, PAL, LFW+, APPA-REAL (REAL AGE) and MORPH II in terms of MAE (years).

Method	Database				
	FG-NET	**PAL**	**LFW+**	**APPA REAL (real age)**	**MORPH II**
Liu et al. [24]	3.93	/	/	/	/
LSDML [25]	3.92	/	/	/	3.08
DRFs [33]	3.85	/	/	/	**2.17**
Gunay and Nabiyev [18]	/	5.40	/	/	/
Bekhouche et al. [6]	/	5.00	/	/	/
Dornaika et al. [12]	/	3.79	/	/	3.67
(DMTL) Hun et al.[19]	/	/	**4.50**	/	3.0
Structured learning [26]	3.89	/	/	/	/
Agustsson et al. [4]	/	/	/	5.46	3.25
Agustsson et al. [4] (Residual DEX)	/	/	/	5.35	2.68
Olatunbosun et al. [3]	3.56	/	/	5.31	2.72
Guehairia et al. [17]	3.65	2.73	5.82	5.25	3.98
Zhang et al. [38]	3.14	/	/	/	**2.15**
Dagher et al. [11]	2.97	/	/	/	2.94
Proposed Approach	**3.05**	**2.39**	5.21	**4.92**	2.89

best MAE for the dataset APPA-REAL (real age) was 4.89. This was achieved with a number N of probabilities equal to 5 and the original labels. For the dataset LFW+, the best MAE was 5.21, obtained with a number N of probabilities equal to 6. For the FG-NET data set, the best MAE was 3.05, obtained with a number N of probabilities equal to 2 and a class step of 2 years. For the dataset PAL, the best MAE was 2.39 obtained with the original labels and a number of probabilities equal to 4. For the dataset MORPH II, the best MAE was 2.89, obtained with a number of probabilities $N = 2$ and a class width of 2 years.

In Table 4 we have compared our method with respect to MAE with some state-of-the-art methods. This table contains the results for the databases FG-NET, PAL, LFW+, APPA REAL (real ages), and MORPH II. Our work outperforms most of the state-of-the-art methods. The proposed method outperformed our previous DRF method. This is due to the use of tensor transformations applied to the deep features.

4.6. Discussion and analysis of the results

In this section we will discuss the results presented in the previous section. As shown in Table 4, the proposed approach outperforms the state-of-the-art for the datasets PAL and APPA-REAL. For the remaining datasets, the results of the proposed method are consistent with the state of the art, but with some advantages. For the FG-NET database, the MAE of our proposed method is 3.05 years, which is better than the MAE of [17], which is 3.65 years, although the work of [17] uses feature fusion and enrichment. In addition, our method is better than the work of Zhang et al. [38] whose MAE is 3.14 years and the work of Olatunbosun et al. [3] with MAE of 3.56 years. The work of Dagher et al. [11] slightly outperforms our work in terms of MAE by 0.08 years.

For the PAL database, we obtained a MAE of 2.39 years, which is the best MAE compared to [17,12,6], whose MAEs for the same database are 2.73, 3.79, and 5.0, respectively.

LFW + is a famous database constructed by Hun et al. in [19]. The work presented in [19] has a MAE of 4.5 years. The MAE of our method using the same database is 5.21 years. As far as we know, the team that created LFW + has achieved the best result. There are many reasons for this. These include the training phase requiring more auxiliary attributes, the proposed deep multi-task learning network (DMLT), and the modified layer with batch normalization layer at the end of each convolutional layer. Their result outperforms ours by 0.70 years in terms of MAE. Nevertheless, our work offers several advantages in terms of complexity, either in time or space, and simplicity of the building blocks of our work.

For the dataset APPA-REAL (REAL-AGE), our method provides the best result compared to the state of the art. The MAE of our method is 4.92 years and actually outperforms our previous work in [17] by 0.33 years.

For the MORPH II dataset, the MAE of our proposed method is 2.89 years. This result outperforms several methods, such as the work of Dagher et al. [11], Agustsson et al. [4], and Guehairia et al. [17]. However, the best result is reported by Zhang et al. [38] with MAE equal to 2.15 years reported. The authors of the latter paper proposed many paradigms for the age distribution. They restricted the age distribution to a reasonable number of adjacent ages. They also explored different age distributions to improve the performance of the proposed learning model. They used CNN and the improved learning model for age estimation.

5. Conclusion

We have improved our previous DRF method for facial age estimation. The current work combines two completely different methods: the first based on tensor subspace learning and the second on deep random forests performing regression by classification. The performance obtained with five public face databases is very promising. The proposed method outperformed several competing methods that rely on end-to-end deep solutions. These results pave the way for further research on enriching face descriptors and integrating feature selection paradigms into the main components of the proposed pipeline.

CRediT authorship contribution statement

O. Guehairia: Software, Validation, Resources, Investigation, Data curation, Writing - original draft, Writing - review & editing. **F. Dornaika:** Conceptualization, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Supervision. **A. Ouamane:** Resources, Validation, Supervision, Formal analysis, Investigation. **A. Taleb-Ahmed:** Resources, Investigation, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Abdi, L.J. Williams, Principal component analysis. WIREs, Computational Statistics 2 (4) (2010) 433–459.
[2] M.N. Adnan, R.H. Ip, M. Bewong, M.Z. Islam, Bdf: A new decision forest algorithm, Information Sciences 569 (2021) 687–705.
[3] O. Agbo-Ajala, S. Viriri, A lightweight convolutional neural network for real and apparent age estimation in unconstrained face images, IEEE Access 8 (2020) 162800–162808.
[4] E. Agustsson, R. Timofte, S. Escalera, X. Baro, I. Guyon, and R. Rothe. Apparent and real age estimation in still images with deep residual regressors on appa-real database. In 12th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2017. IEEE, 2017.
[5] S. Bekhouche, F. Dornaika, A. Benlamoudi, A. Ouafi, A. Taleb-Ahmed, A comparative study of human facial age estimation: handcrafted features vs. deep features, Multimedia Tools and Applications 79 (2020) 26605–26622.
[6] S.E. Bekhouche, A. Ouafi, A. Taleb-Ahmed, A. Hadid, and A. Benlamoudi. Facial age estimation using bsif and lbp. arXiv preprint arXiv:1601.01876, 2016.
[7] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32.
[8] B. Cai, G.-F. Lu, Tensor subspace clustering using consensus tensor low-rank representation, Information Sciences (2022).
[9] Y. Cao, D. Berend, P. Tolmach, G. Amit, M. Levy, Y. Liu, A. Shabtai, Y. Elovici, Fair and accurate age prediction using distribution aware data curation and augmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 3551–3561.
[10] P.K. Chandaliya, N. Nain, Childgan: Face aging and rejuvenation to find missing children, Pattern Recognition 129 (2022) 108761.
[11] I. Dagher, D. Barbara, Facial age estimation using pre-trained cnn and transfer learning, Multimedia Tools and Applications 80 (13) (2021) 20369–20380.
[12] F. Dornaika, I. Arganda-Carreras, C. Belver, Age estimation in facial images through transfer learning, Machine Vision and Applications (2018) 1–11.
[13] F. Dornaika, S. Bekhouche, I. Arganda-Carreras, Robust regression with deep cnns for facial age estimation: An empirical study, Expert Systems with Applications 141 (2020) 112942.
[14] Y. Fu, G. Guo, T.S. Huang, Age synthesis and estimation via faces: A survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (11) (2010) 1955–1976.
[15] T. Grubl and H.S. Lallie. Applying artificial intelligence for age estimation in digital forensic investigations. arXiv preprint arXiv:2201.03045, 2022.
[16] O. Guehairia, A. Ouamane, F. Dornaika, A. Taleb-Ahmed, Deep random forest for facial age estimation based on face images, in: 2020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP), 2020, pp. 305–309.
[17] O. Guehairia, A. Ouamane, F. Dornaika, A. Taleb-Ahmed, Feature fusion via deep random forest for facial age estimation, Neural Networks 130 (2020) 238–252.
[18] A. Günay, V.V. Nabiyev, in: Age estimation based on hybrid features of facial images In Information Sciences and Systems 2015, Springer, 2016, pp. 295–304.
[19] H. Han, A.K. Jain, F. Wang, S. Shan, X. Chen, Heterogeneous face attribute estimation: A deep multi-task learning approach, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (11) (2018) 2597–2609.
[20] A.F. Kamara, E. Chen, Z. Pan, An ensemble of a boosted hybrid of deep learning models and technical analysis for forecasting stock prices, Information Sciences 594 (2022) 1–19.
[21] V. Kazemi, J. Sullivan, One millisecond face alignment with an ensemble of regression trees, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1867–1874.
[22] S. Kim, H. Kim, E.-S. Lee, C. Lim, J. Lee, Risk score-embedded deep learning for biological age estimation: Development and validation, Information Sciences 586 (2022) 628–643.
[23] Y. Lin, J. Shen, Y. Wang, and M. Pantic. Fp-age: Leveraging face parsing attention for facial age estimation in the wild. arXiv preprint arXiv:2106.11145, 2021.
[24] H. Liu, J. Lu, J. Feng, J. Zhou, Group-aware deep feature learning for facial age estimation, Pattern Recognition 66 (2017) 82–94.
[25] H. Liu, J. Lu, J. Feng, J. Zhou, Label-sensitive deep metric learning for facial age estimation, IEEE Transactions on Information Forensics and Security 13 (2) (2018) 292–305.
[26] Z. Lou, F. Alnajar, J.M. Alvarez, N. Hu, T. Gevers, Expression-invariant age estimation using structured learning, IEEE Transactions on pattern analysis and machine intelligence 40 (2) (2018) 365–375.
[27] H. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Mpca: Multilinear principal component analysis of tensor objects, IEEE Transactions on Neural Networks 19 (1) (2008) 18–39.
[28] H. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, A survey of multilinear subspace learning for tensor data, Pattern Recognition 44 (7) (2011) 1540–1551.
[29] A. Ouamane, A. Chouchane, E. Boutellaa, M. Belahcene, S. Bourennane, A. Hadid, Efficient tensor-based 2d+3d face verification, IEEE Transactions on Information Forensics and Security 12 (2017) 2751–2762.
[30] Y. Pang, S. Wang, Y. Yuan, Learning regularized lda by clustering, IEEE Transactions on neural networks and learning systems 25 (12) (2014) 2191–2201.
[31] R. Rothe, R. Timofte, L. Van Gool, Deep expectation of real and apparent age from a single image without facial landmarks, International Journal of Computer Vision 126 (2–4) (2018) 144–157.
[32] N. Sharma, R. Sharma, N. Jindal, Face-based age and gender estimation using improved convolutional neural network approach, Wireless Personal Communications (2022) 1–20.

[33] W. Shen, Y. Guo, Y. Wang, K. Zhao, B. Wang, A. Yuille, Deep regression forests for age estimation, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018, pp. 2304–2313.
[34] Q. Tian, M. Cao, H. Sun, L. Qi, J. Mao, Y. Cao, J. Tang, Facial age estimation with bilateral relationships exploitation, Neurocomputing 444 (2021) 158–169.
[35] S. Wang, S. Yan, J. Yang, C. Zhou, X. Fu, A general exponential framework for dimensionality reduction, IEEE Transactions on Image Processing 23 (2) (2014) 920–930.
[36] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, H. Zhang, Multilinear discriminant analysis for face recognition, IEEE Transactions on Image Processing 16 (1) (2007) 212–220.
[37] T. Yi, Estimation of human age by features of face and eyes based on multilevel feature convolutional neural network, Journal of Electronic Imaging 31 (4) (2022) 041208.
[38] H. Zhang, Y. Zhang, X. Geng, Practical age estimation using deep label distribution learning, Frontiers of Computer Science 15 (3) (2021) 1–6.
[39] T. Zhang, B. Fang, Y.Y. Tang, Z. Shang, B. Xu, Generalized discriminant analysis: A matrix exponential approach, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 40 (1) (2010) 186–197.
[40] Q. Zhao, J. Dong, H. Yu, S. Chen, Distilling ordinal relation and dark knowledge for facial age estimation, IEEE Transactions on Neural Networks and Learning Systems 32 (7) (2021) 3108–3121.