



A framework for facial age progression and regression using exemplar face templates

Ali Elmahmudi¹ · Hassan Ugail¹

Published online: 17 August 2020
© The Author(s) 2020

Abstract

Techniques for facial age progression and regression have many applications and a myriad of challenges. As such, automatic aged or de-aged face generation has become an important subject of study in recent times. Over the past decade or so, researchers have been working on developing face processing mechanisms to tackle the challenge of generating realistic aged faces for applications related to smart systems. In this paper, we propose a novel approach to try and address this problem. We use template faces based on the formulation of an average face of a given ethnicity and for a given age. Thus, given a face image, the target aged image for that face is generated by applying it to the relevant template face image. The resulting image is controlled by two parameters corresponding to the texture and the shape of the face. To validate our approach, we compute the similarity between aged images and the corresponding ground truth via face recognition. To do this, we have utilised a pre-trained convolutional neural network based on the VGG-face model for feature extraction, and we then use well-known classifiers to compare the features. We have utilised two datasets, namely the FEI and the Morph II, to test, verify and validate our approach. Our experimental results do suggest that the proposed approach achieves accuracy, efficiency and possess flexibility when it comes to facial age progression or regression.

Keywords Age progression and regression · Average face · Template-based face generation · Face similarity

1 Introduction

With the emerging use of high functional computational techniques, automated human face analysis has become a topic of immense interest. In this regard, face recognition and face verification, human emotion recognition and age synthesis are some of the prominent application areas [1,2]. In fact, computer-based face recognition itself has many challenges and depends on factors such as ethnicity, the quality and the age of the input photograph and facial expressions [3–5]. In this sense, for example, the task of face recognition centred on the concept of ageing still poses problems, especially since people may appear unrealistically older compared to a probe photograph [6,7].

As far as the ageing of the face is concerned, lifestyle- and health-related factors are known to affect the process of physical ageing. Hence, face ageing is complex and therefore raises significant challenges for computer-based models to create accurate and realistic-looking aged or de-aged faces [8, 9]. As people age, the physical morphology of the face does change [10]. This change depends on many factors. Though it is known that all human faces follow the same general pattern of changes—for example, loss of baby fat from the young age to the appearing of prominent wrinkles at an older age—the rate of these changes is measurable with ethnicity and specific lifestyles [11,12]. Figure 1 shows the typical set of aged face images a modern computer algorithm would generate, given a single frontal face image of an individual as input.

Many algorithms have been introduced in the literature to address the problem of ageing, and most of them rely on strategies which can simulate the effects of ageing on facial images [13]. The Cartoon technique to exaggerate age, for example, was reported by Burt et al. to simulate the effects of ageing of faces [14]. In their method, they compute the average faces of various ages in order to synthesise ages with an

✉ Hassan Ugail
h.ugail@bradford.ac.uk

Ali Elmahmudi
a.a.m.elmahmudi@bradford.ac.uk

¹ Centre for Visual Computing, University of Bradford, Bradford, UK

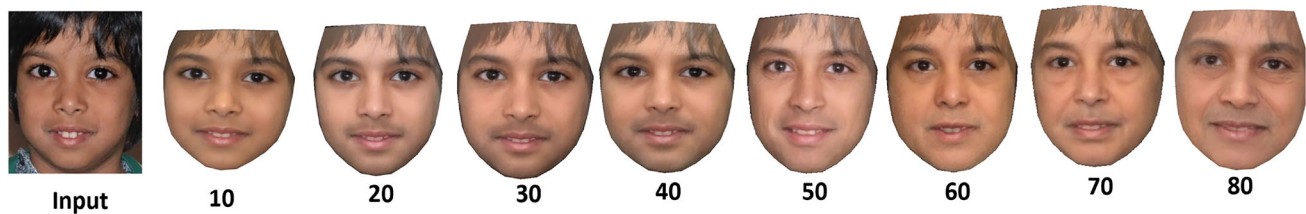


Fig. 1 Given the frontal view of a single image of a child, an ageing algorithm can generate his faces with varied ages

input image to produce new faces. On the other hand, principal component analysis (PCA) was used by Changseok with a 3D face shape model for extracting the components of age change from 3D face to which a test face was added in order to synthesise the output faces at various ages [15,16]. Young et al. addressed the changes in faces along with the ageing effects and demonstrated that several parts of the face—for instance, nose, mouth and eyes—as well as finding some proportions of differences between these parts and wrinkles can be utilised in the algorithmic form to simulate the ageing effect [17,18].

The key objective of the proposed work is to develop and build a technique which addresses the age progression and regression of facial images based on the corresponding template images computed using different ethnicities as well as gender. The main contributions of this work are,

- the development of an efficient template-based formulation to generate specific ages from a given facial image,
- the deployment of a face ageing algorithm with two key parameters—based on the shape and texture characteristics of the input face—to efficiently generate the aged faces,
- and to propose a method based on computer-based face recognition to test and verify the accuracy of the computer-generated aged faces.

The rest of this paper is organised as follows. In Sect. 2, we discuss the recent and relevant literature on the topic of computer-based face ageing. In Sect. 3, we discuss the methodology we have proposed, and in Sect. 4, we present our experiments and the results. Finally, in Sect. 5, we conclude the paper.

2 Literature review

Automatic age generation is a topic of importance with many real-life applications. As such, researchers in the past have suggested various approaches to address the challenge. Recently, deep neural networks, such as the use of generative adversarial networks (GANs) [19] for age synthesis, have become somewhat prominent. The focus of most of

these techniques is simulation based whereby facial data are utilised for constructing generative models which are then utilised to synthesise age—for either progression or regression.

An automatic face ageing method recently proposed is based on the development of a person-specific facial ageing system using constrained regression [20]. This method consists of face features extracted by a colour-based active appearance model (AMM) and then applying regression to generate a face image of a given age. Experiments were conducted on the HQFaces dataset and the Dartmouth Children's Faces database, and the results generated were reliable estimates of the input faces. In 2017, ConvNet features for age estimation were used for facial age estimation by Bukar and Ugail [21]. The method based on extracting features from an input image by using the VGG-face model [22] and partial least squares regression (PLS) was applied to reduce the dimensions of extracted features as well as the redundant information. Two different databases, namely FGNET-AD [23] and Morph II [24], were used as part of the experiments, and the results reported were comparable to other algorithms.

Similarly, Riaz et al. introduced a new method based on 3D gender-specific ageing model, which produced simulated faces at a given age automatically from an input face [25]. The model was constructed with the help of different datasets. Their own comparative analysis of the method with other methods as well as with the ground-truth faces has demonstrated the accuracy of their technique.

Other than that, simulation of ageing on faces based on super-resolution in the tensor space and AAMs was proposed by Wang et al. [26]. Through this method, they can simulate the effects of an adult face by means of super-resolution and AAM. The method also accounts for to reduce the blurring effects which result from a normalisation process of the input face. To verify the accuracy, the FGNET [23] database was used, and the experimental results show that the results of their ageing simulation were adequate.

Similarly, the Personalised Age Progression with ageing Dictionary was proposed by Shu et al. [27]. The main goal of this method was to produce rendered faces in a personalised way. Their approach relied on two stages, namely offline and online. During the offline stage, short-term ageing image pairs were collected from available datasets, and an ageing

dictionary was trained. And, during the online stage, the researchers rendered an aged face for an input face within an age group determined by the computation of the nearest neighbour. Then, the resulting aged face was used as an input to the algorithm. This process is repeated until all the desired aged faces are generated. To test their methodology, they used the Cross-Age Celebrity [28] and Morph ageing databases [24] in their experiments. The results demonstrated some advantages of the proposed method compared with others.

Recent GAN-based work [19] was introduced by Zhang et al. [29] for age regression and progression. The approach is referred to as Conditional Adversarial Autoencoder network (CAAE). They use the convolutional encoder in order to map an input face to a latent vector [30] and then to project the resulting vector to a face manifold. This vector conserves features of a personalised face, and an age condition controls the regression and progression. The system was trained on a large dataset called the UTKFace dataset [31], and it was evaluated through different databases such as Morph and CACD [28]. The results indicate that the system can generate faces in a more realistic manner, and it has a degree of flexibility too.

From what has been presented above, it can be seen that some of the facial age generation techniques produce aged facial images in a way that depend on the features of the face and ignore key areas such as the forehead as indicated in [20]. Besides, GANs-based approaches [19] that use generative algorithms for creating new faces of varying age can give good results if the test faces are part of the training database. However, the results in such cases appear to be rather unsatisfactory when using faces outside the training set. This essentially limits the practical applicability of such systems.

Therefore, it appears that a flexible and computationally less complex method that produces reliable results is much needed.

3 Proposed methodology

The methodology we have proposed here for facial age progression and regression is intended for overcoming some of the key challenges in such systems that currently exist. One key objective we strive to achieve here is the development of a flexible and lightweight method for generating realistic aged faces. The proposed framework is based on face templates, which are built by extracting information on the age, gender, colour and texture characteristics from a number of faces corresponding to the principal ethnic groups. Ethnicity-based face templates can play a vital role in generating realistic faces, by combating the artefacts that arise

from modern and commonly available techniques such as GANs-based ageing systems.

The proposed system consists of two key parts. The first part is the mathematical method for building and generating the proposed face templates. It uses an average face—for a given ethnicity, age and gender—considering a sufficient number of faces for the corresponding category. In the second part, the generated templates are applied to the target faces for age generation with two key control parameters, based on the colour and texture of the face. Finally, as part of our methodology, we also propose a framework for verifying the accuracy of the generated faces through similarity comparison by means of standard face recognition. To compute and verify face similarities, we use a method based on the state-of-the-art CNNs.

3.1 Building the ageing templates

To construct the person-specific ageing templates, we use the concept of the average face—based on a given ethnicity, age and gender. A similar technique to what we propose here is also presented in [20]. We generate person-specific ageing templates for five specific ethnicities, namely Middle Eastern—Arabic, Southeast Asian—Indian, African—Black, Caucasian—White and Eastern—Chinese, and with nine age gaps from age 10 to 80 years with increments of 10 years for both the genders.

3.1.1 Data collection

The required data for creating the ethnicity-specific templates were collected in four phases. Firstly, an Arab educational institution in Bradford, UK, was approached, and participants were recruited for photography. The participants consisted of male and female kids and teachers, with an average age of 10–15 years for the kids and 31–53 for the teachers. In the second phase of image collection, colleagues from some Arab countries consented to send images of themselves. Thirdly, students from the University of Bradford were recruited. The fourth and final stage of data collection consisted of downloading readily available images from the Internet, again of various races and ages.

3.1.2 Ageing template

All the images that were gathered were categorised into groups based on the corresponding race, gender and age. In addition, since all input faces were of different dimensions originally, it was necessary to normalise them and bring each image to the same reference frame. The method of generating the templates for ageing is as follows.

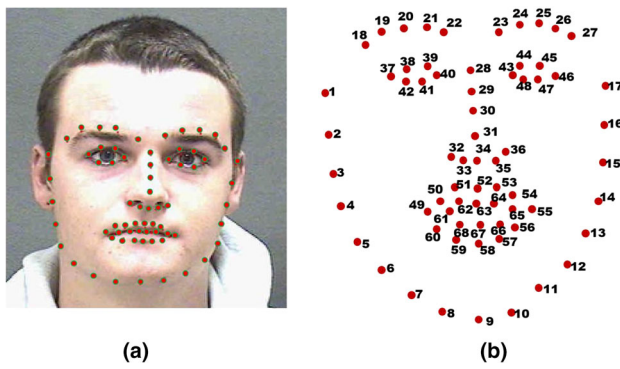


Fig. 2 Identification of facial landmarks using Dlib. **a** Facial landmarks. **b** The position and order of 68 points on the face

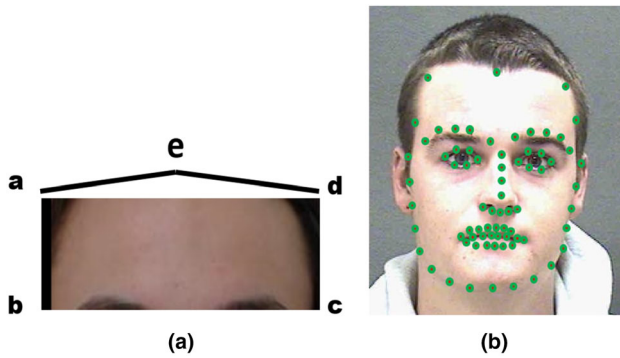


Fig. 3 Identification of additional landmarks to include the forehead. **a** The five facial landmarks to cover the forehead area. **b** The resulting facial landmarks after adding five additional points on the face

1. Detection of the facial features: For face landmark detection, Dlib algorithm is used [32]. Dlib uses a pre-trained model to estimate the position of 68 facial landmark points (x, y) on the face, as shown in Fig. 2. However, the forehead provides information about a person’s age, and Dlib’s facial feature points do not cover that region of the face. Therefore, we added five extra points based on the information from the Dlib algorithm, as 1, 2, 3, 4 and 5, as shown in Fig. 2. It is assumed that the forehead is rectangular in shape, and hence we identify the forehead area using 5 landmark points, as shown in Fig. 3.

After computing the five points to identify the forehead section of the face, the number of facial landmarks increases to 73 points in total, as shown in Fig 3b.

To find the coordinates of the point $a(x, y)$ —as seen in Fig. 3a, first we compute the height of the forehead. This is approximated using the length of nose D , and we apply the following equation,

$$a_x = P_x^{19} - D_{nose}, \quad a_y = P_y^{19} - C, \quad (1)$$

where P_x^{19} and P_y^{19} are x and y values at the point 19 and C is a constant to normalise the distance and $D_{nose} = |P_x^{28} - P_x^{31}|$. Similarly, the points b, c, d and e are computed using Eqs. 2, 3, 4 and 5, respectively, i.e.

$$b_x = P_x^{18} - \text{round}[(D_{nose}/2)] + 1, \quad b_y = P_y^{18} - C, \quad (2)$$

$$c_x = P_x^{27} - \text{round}[(D_{nose}/2)] + 1, \quad c_y = P_y^{27} - C, \quad (3)$$

$$d_x = P_x^{26} - D_{nose}, \quad d_y = P_y^{26} - C, \quad (4)$$

$$e_x = \text{round}[(P_x^{21} + P_x^1)/2], \quad e_y = P_y^{21} - C_1, \quad (5)$$

where C_1 are the normalised distances.

2. Generating the templates: As discussed earlier, the templates for given ages are considered for the five principal ethnicities, namely 1. Middle Eastern—Arabic, 2. South-east Asian—Indian, 3. African—Black, 4. Caucasian—White and Eastern and 5. Chinese. We consider templates in age increments of 10 years—i.e. ages 10, 20, ..., 80 years—for both females and males for each of the five ethnicities.

Consider, for example, the face images I_n of Middle Eastern males at the age of 70 years, where n is the number of images. Suppose we want to generate a template for this age category for the ethnicity. In the first step, images are preprocessed to resize them to the same size and remove the backgrounds I_i^p , where $i = 1, 2, \dots, n$. Then, the facial landmarks P_i are extracted for all the images I_i^p by the method discussed earlier. Before computing an average, all the images are aligned to the main shape by using the generalised procrustes analysis (GPA) [20,33], using Eq. 6, such that

$$AI_i = \text{GPA}(I_i^p, M_s), \quad i = 1, 2, \dots, n, \quad (6)$$

where M_s is the mean shape and computed using Eq. 7, such that

$$M_s = \frac{1}{n} \sum_{i=1}^n P_i. \quad (7)$$

Now, we can compute the template by warping the aligned faces AI_i to the mean shape M_s and then computing the average, using Eq. 8, such that

$$\text{template} = \frac{1}{n} \sum_{i=1}^n \text{warp}(AI_i, M_s), \quad (8)$$

where warp is inferred as the spatial transformation [34]. The resulting template face is shown in Fig. 4a.

3. Wrinkle map: Wrinkles play an important role in simulating realistic-looking faces as they age. In simple terms, a

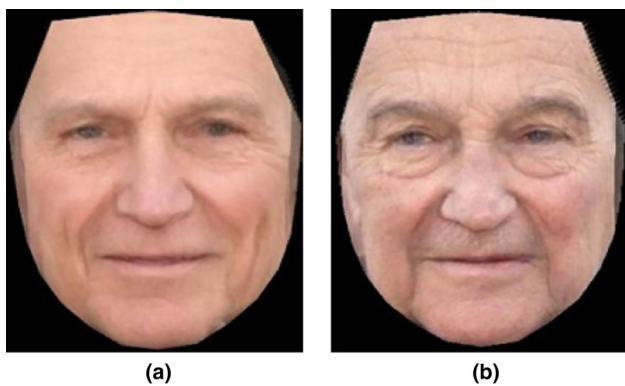


Fig. 4 An example template face. **a** The template face for a male of middle eastern origin at age 80 years. **b** The template face after adding wrinkles

Table 1 List of sample face template codes

| Ethnicity | Gender | Age | Generated code |
|-----------|--------|-----|----------------|
| 1 | 1 | 10 | 1110 |
| 2 | 1 | 20 | 2120 |
| 3 | 2 | 30 | 3230 |
| 4 | 2 | 40 | 4240 |
| 5 | 2 | 90 | 5290 |

The first digit represents the ethnicity, the second digit is gender and the last two digits represent an age

wrinkle map W_m is an image with high-quality wrinkles that can be added to ageing templates. In this step, we add the wrinkle maps W_m through Eq. 9 such that

$$N_t = \text{warp}(W_m, \text{template}). \tag{9}$$

Figure 4b shows an example face template image with wrinkles—for an 80-year-old male of middle eastern ethnicity.

- At the final step, all the templates are coded with specific labels based on ethnicity, gender and age. For example, 4280 stands for 4—White ethnicity), 2—female and 80—age 80 years. Table 1 shows some samples of the generated codes. Figure 5 shows example templates for various ethnicities, ages and gender.

3.2 Computing age progression or regression

Once the templates have been generated, we can then utilise them to either progress or regress a facial image to a given age. To do this, we utilise image morphing with cross-dissolve [37], as discussed below.

Suppose we have an input face image I_{in} and we want to age it to a 60-year-old middle eastern male. We invoke the corresponding template, i.e. T_{1160} . First, we obtain the

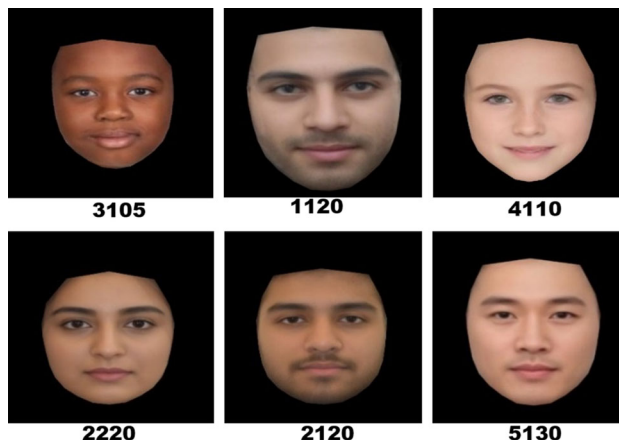


Fig. 5 Examples of generated templates with numeric descriptions; e.g. the code 3105 denotes a black male of age 5 years, the code 1120 denotes a middle eastern male of age 20 years and 4280 refers to a white female of age 80 years

| α_{colour} | α_{shape} | | | | | |
|-------------------|------------------|-------|--------------|------------|-------|--------------|
| | 0.25 | | | 0.5 | | |
| | Similarity | | Ground truth | Similarity | | Ground truth |
| SSIM | CS | SSIM | | CS | | |
| 0.25 | | 64.78 | 71.39 | | 65.65 | 81.57 |
| 0.5 | | 73.14 | 73.02 | | 74.68 | 86.15 |

Fig. 6 An example from the tests carried out to determine the optimal values for the two parameters, α_{shape} and α_{colour} . Here, a face is taken and aged to 80 years with various choices on the values for α_{shape} and α_{colour} . The similarities are determined using the Cosine Similarity (CS) and the Structural Similarity Index (SSIM). The results for the aged face of 80 years by using different values of α_{shape} and α_{colour} . As a result, the optimal parameter values that should be considered are determined to be $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$

corresponding landmark points for I_{in} and T_{1160} using the modified Dlib algorithm discussed earlier. We refer to these points as P_{in} and P_t , respectively. Then, we generate an intermediate warping field I_{wp} by using interpolation as in Eq. 10, such that

$$I_{wp} = \alpha_{shape} \times P_{in} + (1 - \alpha_{shape}) \times P_t, \tag{10}$$

where α_{shape} is a parameter to control the degree of shape such that $0.25 \leq \alpha_{shape} \leq 0.75$.

Then, the average between P_{in} and P_t is computed and used to find the corresponding Delaunay triangulations DT [38]. To avoid any ghosting effects in the resulting image,

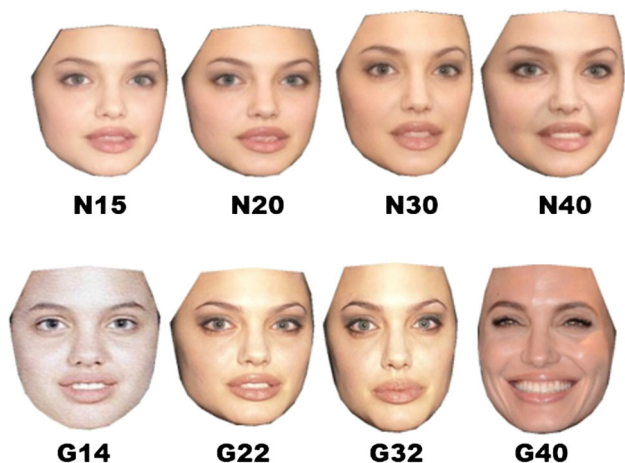


Fig. 7 Comparison of aged faces with the corresponding ground-truth faces for Angelina Jolie. The first row shows the generated faces from age 15 to 40 years with the naming convention NXX, where N indicates the face is computed and XX refers to the age in years. For example, N15 indicates a newly generated face at age 15 years. In the second row, real faces for the ground truth are presented, which were collected from the Internet. The naming convention used here is GXX, where G indicates that image is a ground truth and XX corresponds to the age in years. For instance, G14 indicates a real face of age 14 years

Table 2 Percentage of similarities between ground-truth and generated faces for different ages, for Angelina Jolie

| Methods of comparison | Age range and similarity | | | |
|------------------------------|--------------------------|-------|-------|-------|
| | 15 | 20 | 30 | 40 |
| Cosine similarity (features) | 76.41 | 90.63 | 85.38 | 71.59 |
| Structural similarity | 62.02 | 96.08 | 85.33 | 68.36 |
| IMG-online | 73.46 | 96.46 | 84.03 | 68.93 |

we warp I_{in} and T_{1160} into I_{wp} by applying an affine transformation function AT [39,40], such that

$$I_{in}^w = AT(I_{in}, P_{in}, I_{wp}, DT), \tag{11}$$

$$T_{1160}^w = AT(T_{1160}, P_t, I_{wp}, DT), \tag{12}$$

where I_{in}^w and T_{1160}^w are the warped images.

Finally, we apply the method of cross-dissolving [37] to the warped images I_{in}^w and T_{1160}^w to obtain an aged face I_{aged} , using Eq. 13, such that

$$I_{aged} = \alpha_{colour} \times I_{in}^w + (1 - \alpha_{colour}) \times T_{1160}^w, \tag{13}$$

where α_{colour} is a parameter to control the degree of colour such that $0.25 \leq \alpha_{colour} \leq 0.75$.

3.3 Method of verification with the ground truth

Once the facial age progression and regression algorithm is in place, it is vital to test the accuracy of the generated faces

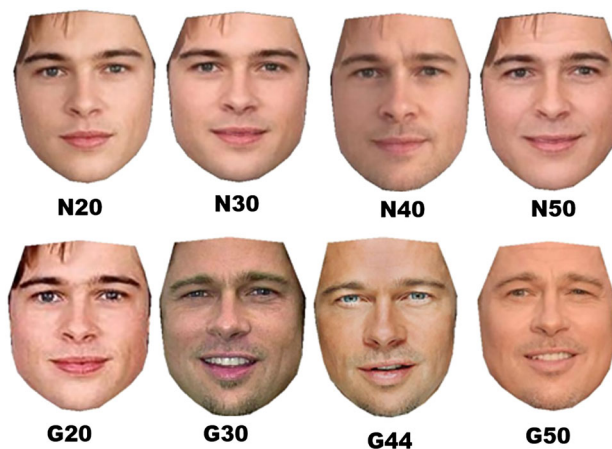


Fig. 8 Comparison between the aged faces and the ground-truth faces for Brad Pitt. The first row shows the aged faces from age 20 to 50 years. The second row shows the ground-truth faces, which were obtained from the Internet

Table 3 Percentage of similarities between the ground-truth and generated faces, for Brad Pitt

| Methods of comparison | Age range and similarity | | | |
|------------------------------|--------------------------|-------|-------|-------|
| | 20 | 30 | 40 | 50 |
| Cosine similarity (features) | 85.82 | 75.83 | 72.37 | 73.61 |
| Structural similarity | 80.43 | 75.57 | 83.72 | 83.36 |
| IMG-online | 80.14 | 82.28 | 76.56 | 78.63 |

Table 4 Summary of the facial similarity results between the generated faces and the ground truth, for all the faces in the FEI dataset

| Dataset | CS | SSIM |
|---------|---------|-------|
| FEI | 85.161% | 78.77 |



Fig. 9 Sample images from the FEI face dataset

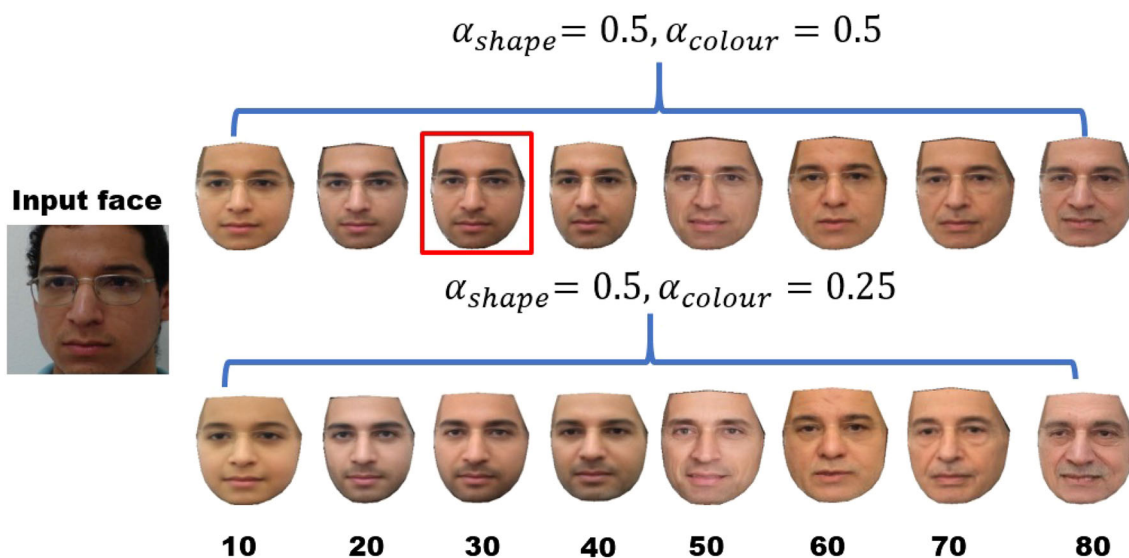


Fig. 10 Sample aged faces with different parameter values for an individual. The ages considered are between 10 years and 80 years old. The generated face shown in the red box in the first row is the one with the maximum similarity; i.e. both the input face and the generated face, in this case, are at age 30 years

with the corresponding faces of the ground truth. In order to evaluate the accuracy of our method for face ageing, discussed above, we compare face similarities between the aged faces and the corresponding faces of ground truth. There are various approaches suggested for face recognition and classification on real faces as in [41,42]. The verification method we have adopted here is based on the use of state-of-the-art CNN-based face recognition approach [43].

Due to the low number of images available per subject, here we have utilised the VGGF model [22] which is widely used for face recognition tasks. The VGGF model was developed by Oxford Visual Geometry Group [22]. This model was trained on a large database which consisted of about 2.6M faces of more than 26K individuals. The model contains 38 training layers. In our case, we utilised the layer 34 for feature extraction because it is widely reported to be the layer that provides the most classification accuracy.

The extracted facial features—from both a ground-truth image and an aged image—are represented as a vector of dimensions 4096 for each face considered. All these vectors can then be used for training the classifiers such as the cosine similarity CS [44], decision trees [45], *k*-nearest neighbours *K*-NN [46] and linear support vector machines (SVMs) [47].

Before we discuss the experiments and their results, it is worth mentioning about the choice of the two parameters α_{shape} and α_{colour} . In order to understand the best choice for these parameters, we ran a number of preliminary experiments in which both the parameters were tested for possible values between $0.24 < \alpha_{shape}, \alpha_{colour} < 0.76$. As a result, based on our observations and computing the similarities—using cosine similarity (CS) [44] and Structural Similarity

Index (SSIM) [48])—between the ground truth and the aged faces, we found the optimal values α_{shape} and α_{colour} is 0.5. We have illustrated this in the example shown in Fig. 6 where we have taken a subject, aged him to 80 years by the various choices of the values for the α_{shape} and α_{colour} and compared the resulting face images with the ground truth. As can be observed in that figure, the highest similarity percentage is recorded at $\alpha_{shape} = 0.5$ and $\alpha_{colour} = 0.5$, which is recorded to be 86.15%.

In Table 3, it can be observed that the similarity measures for all the generated ages when compared with the ground truth are well above 70%. The highest similarity percentage is for age 20, which is 85.82% by using CS, and the lowest value obtained is 72.37% for age 40 years.

3.3.1 Sample tests

Before we embarked upon large-scale experiments to test the accuracy and efficiency of our methodology, we decided to run small-scale tests in which we wanted to compare the results of our aged faces with the corresponding ground truth. For this purpose, we did some comparative analysis of our generated faces with the ground-truth facial images for two celebrities, namely Angelina Jolie and Brad Pitt. We computed the face similarity matrices between the generated faces and real faces with the corresponding ages of the two celebrities. Once new faces are generated, we use the VGGF model described above and two other methods to measure the similarities between the facial images.

In the first approach, which is based on a feature map, a total of 4096 features, using the VGGF model, are extracted

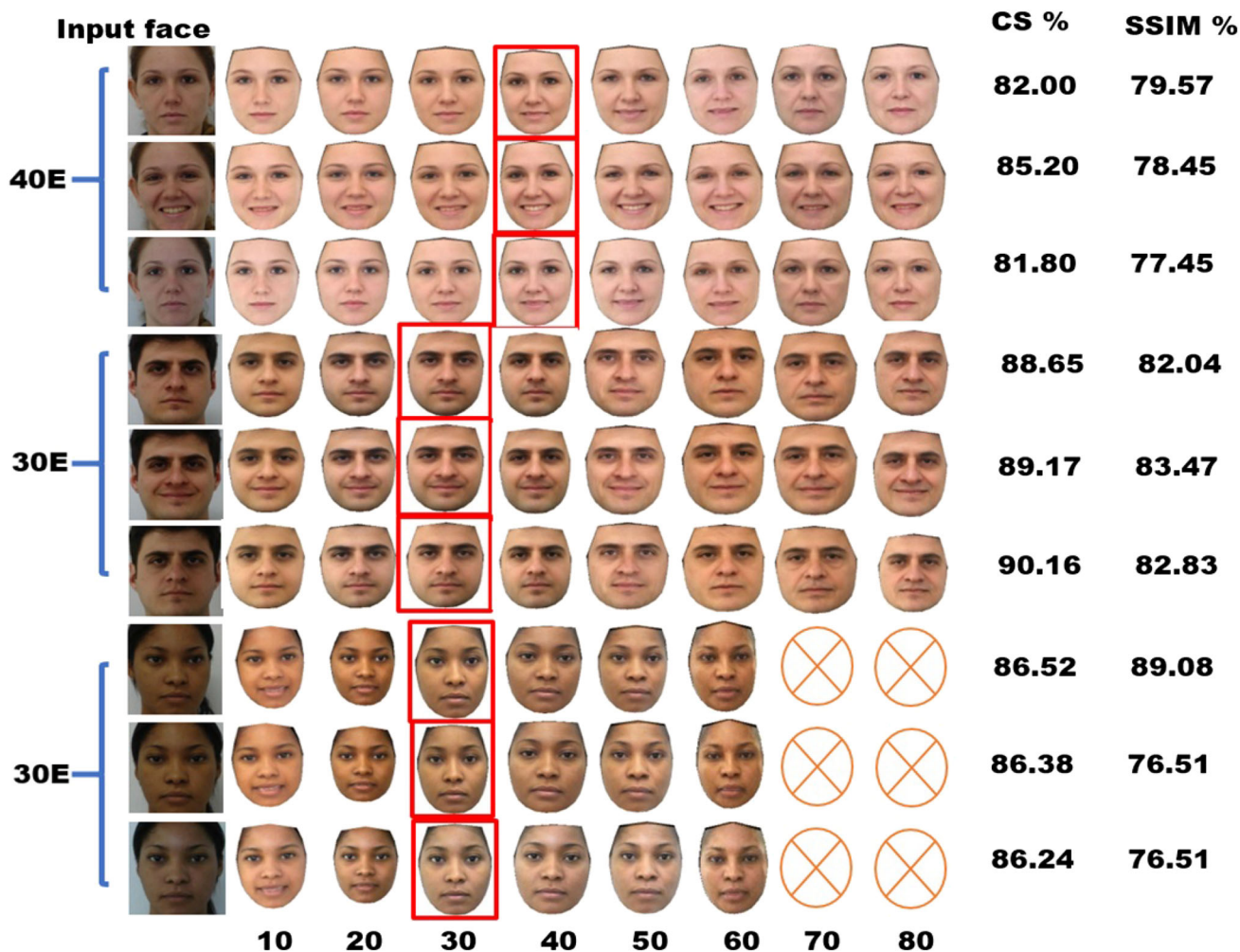


Fig. 11 Examples of aged face images—for individuals from the FEI dataset. The first column represents the input images with estimated age, i.e. 30E means the estimated age is 30 years. The next 8 columns show the aged faces from 10 to 80 years. The faces inside red boxes correspond to the those with matching ages between the input and the

generated face. Note crosses indicate that we do not have the templates for the black females between 70 and 80 years. Finally, the last two columns show the percentage similarities between input faces and the aged faces for the same ages

for all the facial images by using the convolutional layer 34 in the VGGF. These feature parameters are then passed to the cosine similarity (CS) classifier to compute the percentage similarity. In the second approach, we used the Structural Similarity Index (SSIM) [48] to compute the similarity between the ground truth and the new faces. For our final approach, we used the image map method in which an online Web portal was used to identify the similarity between two facial images, which we refer to as IMG-online (IMG) [49].

In Fig. 7, we show the face images generated for four different ages for Angelina Jolie. It can be observed in Table 2 that the highest similarity measure obtained when compared to the ground truth with the aged faces is for the age 20 years which is 90.63% for CS, 96.08% for SSIM and 96.46% for IMG. In contrast to that, the lowest similarity measure

recorded is for the age of 15 years, which were 76.41% for CS, 62.02% for SSIM and 68.93% for IMG.

In the second example, the face images of Brad Pitt were used to evaluate the proposed method. Firstly, four different aged images were generated using our proposed approach. Figure 8 shows the gendered faces and the corresponding faces of ground truth. Similar to the previous example, to compute the facial similarities, we extracted the features for all images by using the VGGF. We then compared the aged images with the corresponding ground truths based on the three approaches discussed earlier.



Fig. 12 Sample face images from the Morph II face dataset

4 Experiments and results

For performance evaluation, experiments were conducted using two different public-domain face databases (FEI [50] and Morph II [24]) to generate faces of different ages, sex and races from the generated templates. In addition, the optimal values for the shape and colour parameters were estimated through comparative studies with other similar work reported in the literature.

In Table 4, we summarise the facial similarity results between the generated faces and the ground truth for all the faces the FEI dataset. Note the results reported in Table 4 are for the values $\alpha_{\text{shape}} = 0.5$ and $\alpha_{\text{colour}} = 0.5$.

4.1 Using the FEI dataset

FEI is a Brazilian facial dataset consisting of 200 faces of students and staff of both the male and female sex [50]. Each participant had 14 images captured, and the resolution of all the images is 640×480 pixels. All the facial images are in colour and are taken against a neutral background. The ages of the individuals are between 19 and 40 years and consisted of faces with facial expressions and types of various poses. Figure 9 shows some sample images from the FEI dataset.

For each of the experiments, using the FEI dataset, three front face images were selected for each subject—totalling to 600 facial images. The faces are then age progressed and regressed using the methodology described earlier. From the experimental results, by setting the two parameters (α_{shape} and α_{colour}) into different values, we found that for $\alpha_{\text{shape}} = 0.5$ and $\alpha_{\text{colour}} = 0.5$, our method consistently produced the best aged face. In Fig. 10, we show a sample of aged faces with different parameter values for an individual whereby the ages considered are between 10 years and 80 years with different parameters.

Thus, the proposed method can be utilised to generate face images at various ages which are both ethnicity and gender specific. For a rigorous evaluation of our age regression and progression method, we also performed the K -fold cross-validation by taking $K = 3$ —which means that for each subject we used three different original images to produce the aged faces, as shown in Fig. 11.

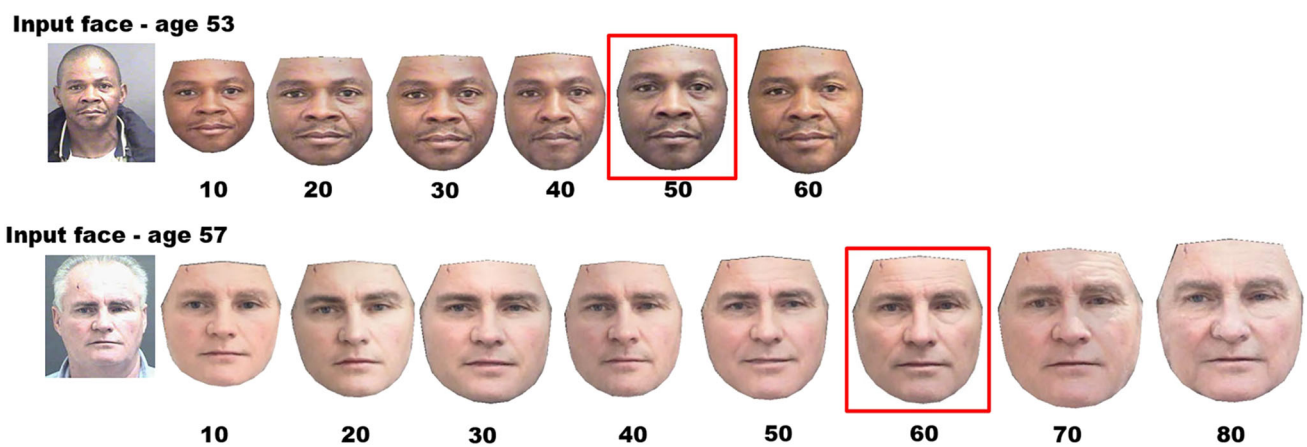


Fig. 13 Age generation results; the single two images on the left are the real images. The faces on the top row were generated using our the proposed approach for the ages 10 to 60 years. Images in the second

row represent the generated faces between 10 and 80 years. The face image enclosed in the red box, in each case, corresponds to the age of the input faces

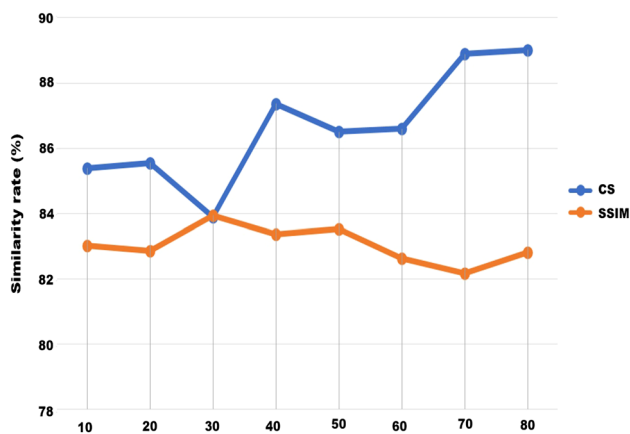


Fig. 14 Comparison of matching between the generated faces and ground-truth images in Morph II faces

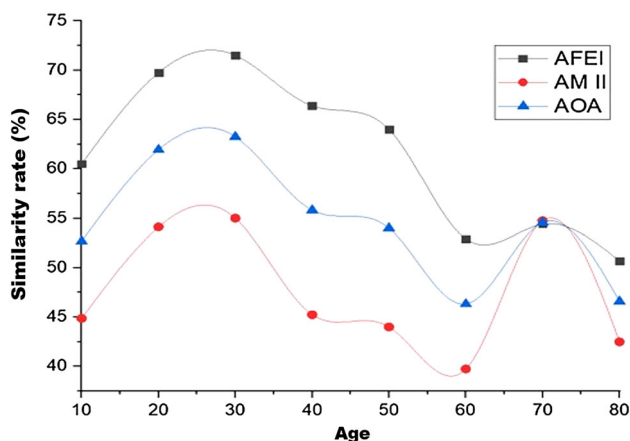


Fig. 15 Summary of the results from one-to-many face similarity matching tests on the FEI and Morph II datasets. The results reported here are for the average similarity rates—i.e. the average for the FEI dataset (AFEI), the average for the Morph II dataset (AMII) and the overall average (AOA)—for the generated faces in the age ranges from 10 to 80 years, for $\alpha_{\text{shape}} = 0.5$ and $\alpha_{\text{colour}} = 0.5$

The faces in the FEI dataset are not recorded with the corresponding age of the individuals. Thus, an Internet application (How-Old.net) [51] was used to estimate the ages for each of the faces. Furthermore, similarities between the aged faces and the corresponding faces of ground truth for the same ages were computed by using the CS and SSIM, as shown in the last two columns in Fig. 11. As one can see, there is a similarity match between the aged faces and the corresponding ground truth. Note a similarity match of 70% from the CNN face recognition algorithm means the faces considered are an identity match. Since all the aged faces from our method, show similarity values higher than 70% indicating the accuracy of our results, it verifies the identity of the individuals.

4.2 Using the Morph II dataset

Similar to using the FEI dataset, experiments were repeated on the faces from the Morph II face dataset. This dataset contained roughly 55,000 faces of 13,000 subjects and was collected over four years. It contains faces with a range of ethnicities, gender, and it consists of face images of individuals of ages between 16 to 77 years. The quality of images in this dataset is generally poor, particularly because the brightness contrast of some faces is very high. As a result, some of the prominent features of the face in some of the faces are poorly represented. After carefully analysing all the images in the dataset, we selected images corresponding to 200 individuals through which we conducted our experiments. Figure 12 shows sample face images from the Morph II face dataset.

In order to generate new ages, we selected subjects with the most number of available images. We then applied the methodology described above, again by using the same setting for the two parameters ($\alpha_{\text{shape}} = 0.5$ and $\alpha_{\text{colour}} = 0.5$). Figure 13 shows some examples of the aged faces. The first row shows the aged faces of a black male resulting from the input of his real face at the age of 53 years whereby it was progressed and regressed to generate aged faces between 10 and 60 years. Similarly, Figure 13 shows the aged faces of a white male whereby the input was his real facial image at 57 years. Again, by utilising the input facial image, it was then progressed as well as regressed to generate aged faces between 10 and 60 years. Moreover, in Figure 14, we can see how the generated ages are very close to the ground-truth faces when two different classifiers were applied for matching.

4.3 One-to-many similarity matching trials

In the previous two experiments, we showed that there is an excellent match between the aged facial images and the corresponding faces of ground truth when the similarity matching face recognition is conducted on a one-to-one basis. In this experiment, we extend it so that similarity matching can be conducted on one-to-many basis; i.e. given an aged face, we wanted to know the similarity figure for it when we compare it with all the available images in the entire dataset. We conducted this experiment for both the FEI and Morph II datasets.

For classification purposes, in this experiment, we have utilised the CS, K -nearest neighbour (KNN) and decision tree (DT) classifiers. We looked at the classification results individually for each classifier and reported the results based on the average recognition rates for all the classifiers considered.

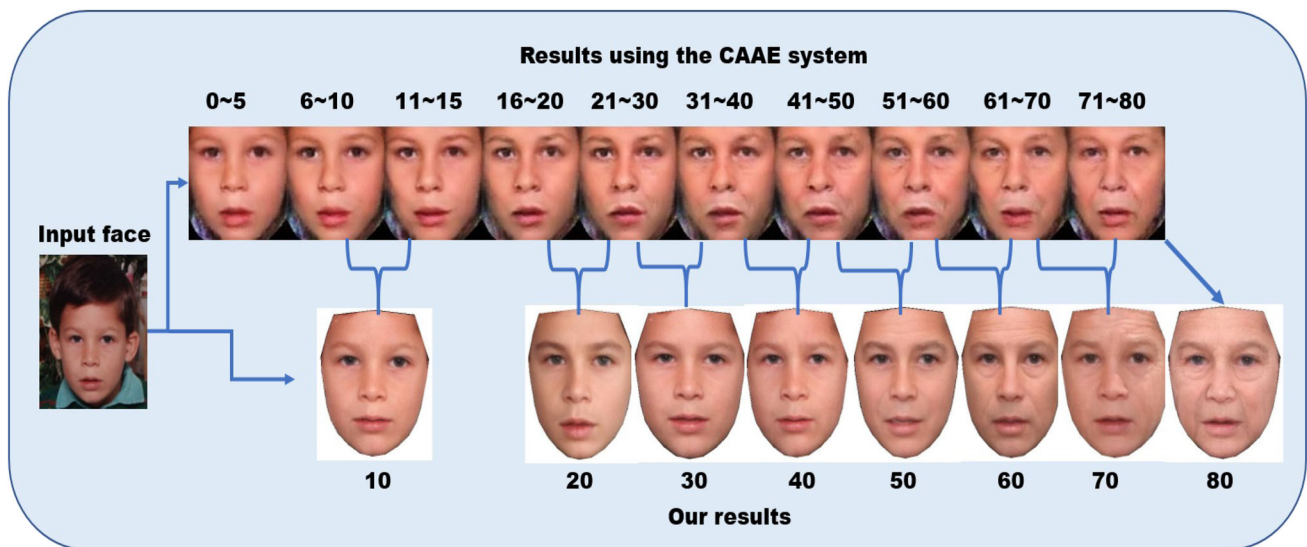


Fig. 16 Comparison of our method with the method based on the CAEE system. The first row shows aged faces from an individual for ten age groups—the results from prior work [29]. The second row shows the results from our method for the same individual in age groups of 10 to 80 years

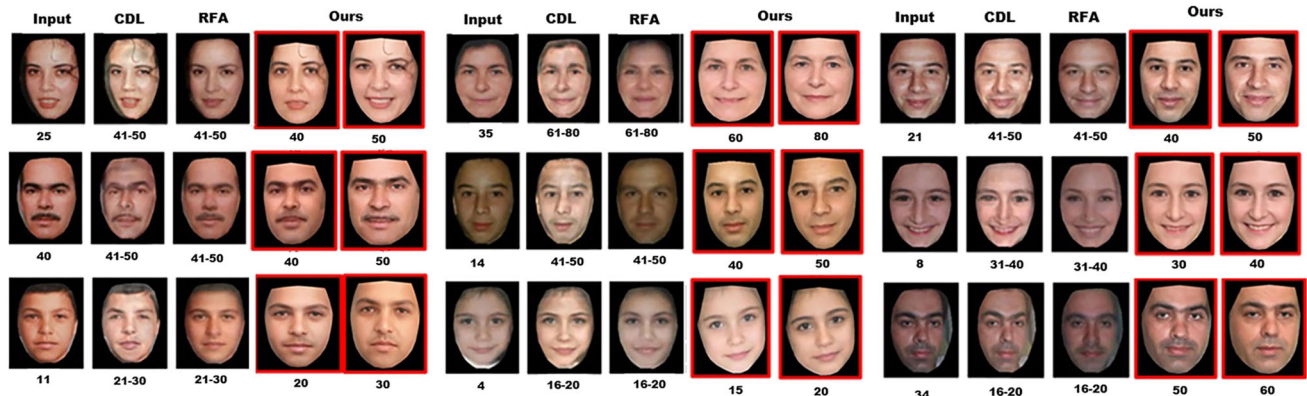


Fig. 17 Comparison of our proposed method with some of the state-of-the-art methods in the field. The first column in each block represents the input images and the second and the third columns the results from

CDL [35] and RFA [36], respectively. The last two columns in each block with red squares show the results from our method

For the experiment on the FEI dataset, as previously discussed, we selected three images per subject for age generation for ages from 10 to 80 years in increments of 10 years. The remaining images of the individuals were utilised for training faces for the recognition process. In all the experiments, we separated the test images into groups representative of their age ranges. That allowed us to test face images of each age group separately and also allowed us to identify the individual recognition rates for various age groups.

Thus, we carried out the recognition process under four different scenarios of classification for each age group. For the images in the FEI dataset, the recognition rate using the CS classification significantly outperformed the

rest of the classifiers, reaching between 93% and 96% for the ages of 20 up to 50. However, we observed that for the age of 60 years, the percentage of recognition decreased to about 75%, which is still significant.

Similarly, for the images in the Morph II dataset, we followed the same approach as above. For this experiment, we selected 200 subjects from the Morph II dataset and also selected three images for age progression and regression for the age groups from 10 to 80 years.

Based on the one-to-many face recognition results of this experiment, the most challenging ages for similarity classification appear to be distributed in the very young and the very old age groups; i.e. for age 10 and from the

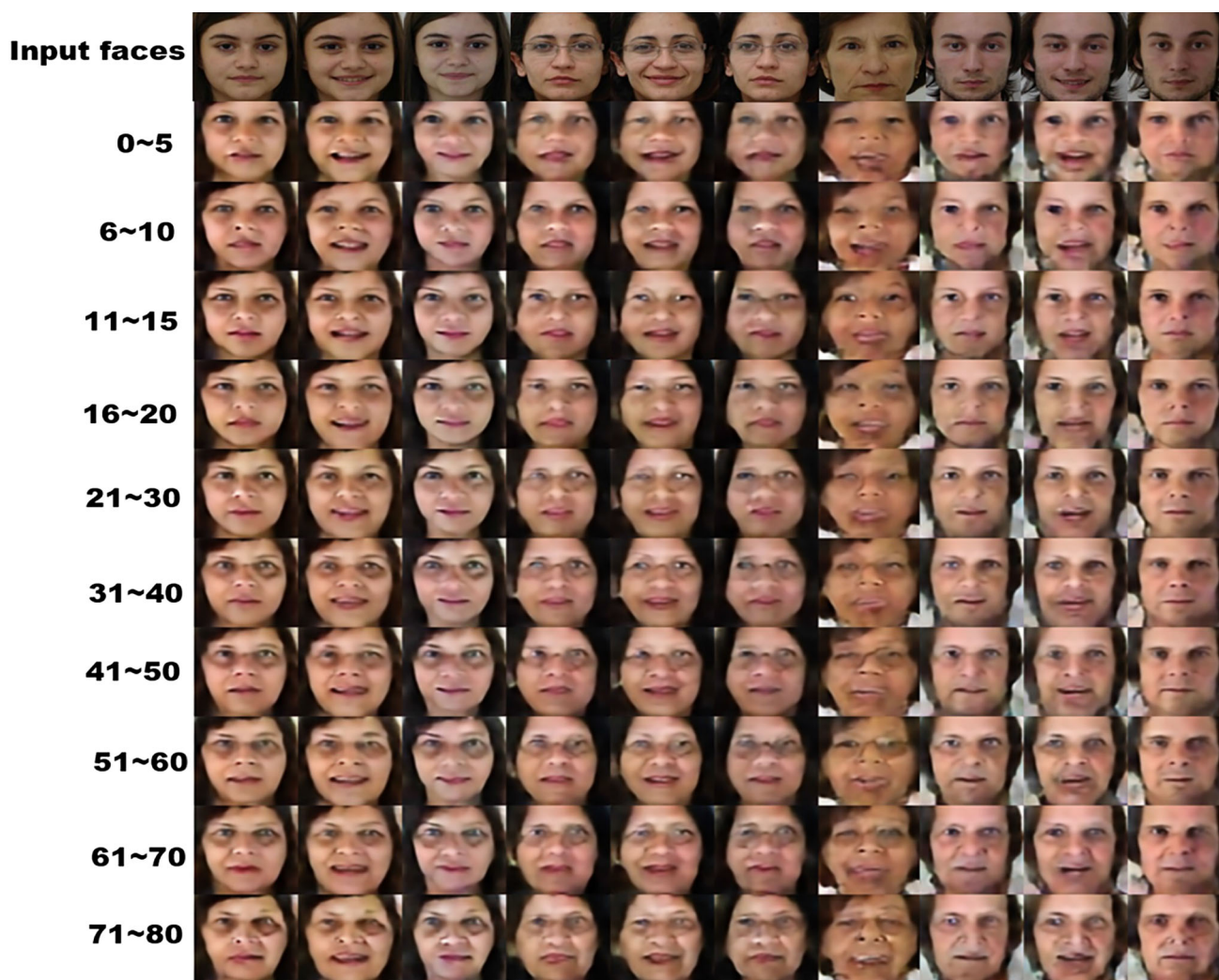


Fig. 18 Results after using the CAAE system to age selected some faces taken from the FEI dataset. The first row shows the input faces. The rest of the rows show the aged faces in the different age groups

ages of 50 through to 80, the similarity classification rates are relatively poor, as shown in Fig. 15. The main reason for this is that in the Morph II dataset, the subjects are between the ages of 30 to 50 years old. Therefore, the dataset does not have subjects with very young and very old ages. Additionally, the images in the Morph II dataset are generally of poor quality and therefore would have contributed to the overall rate of recognition negatively. Thus, the average of overall (AOA) indicates that the rate of recognition for $\alpha_{\text{shape}} = 0.5$ and $\alpha_{\text{colour}} = 0.5$ in both datasets has the best outcome of around 68% at the age 30 years, and for the worst case, it is roughly 47% at the age 60 years.

4.3.1 Comparison with the most recent work

Finally, we also rigorously compared our method with some of the most recent work in the literature. For comparison, we have selected the use of GANs [29], the Recurrent Face Aging (RFA) framework [36] and Coupled Dictionary Learning (CDL) [35]. All these methods are reported to be examples of state of the art on age progression and regression.

In this experiment, firstly, we investigated the effectiveness of our method by applying it to the method mentioned in [29]. The advantage of our proposed approach is the ability to progress and regress a given face image efficiently with the choice of two control parameters. In Fig. 16, it can be seen that our method generates aged faces which are more realistic and plausible.

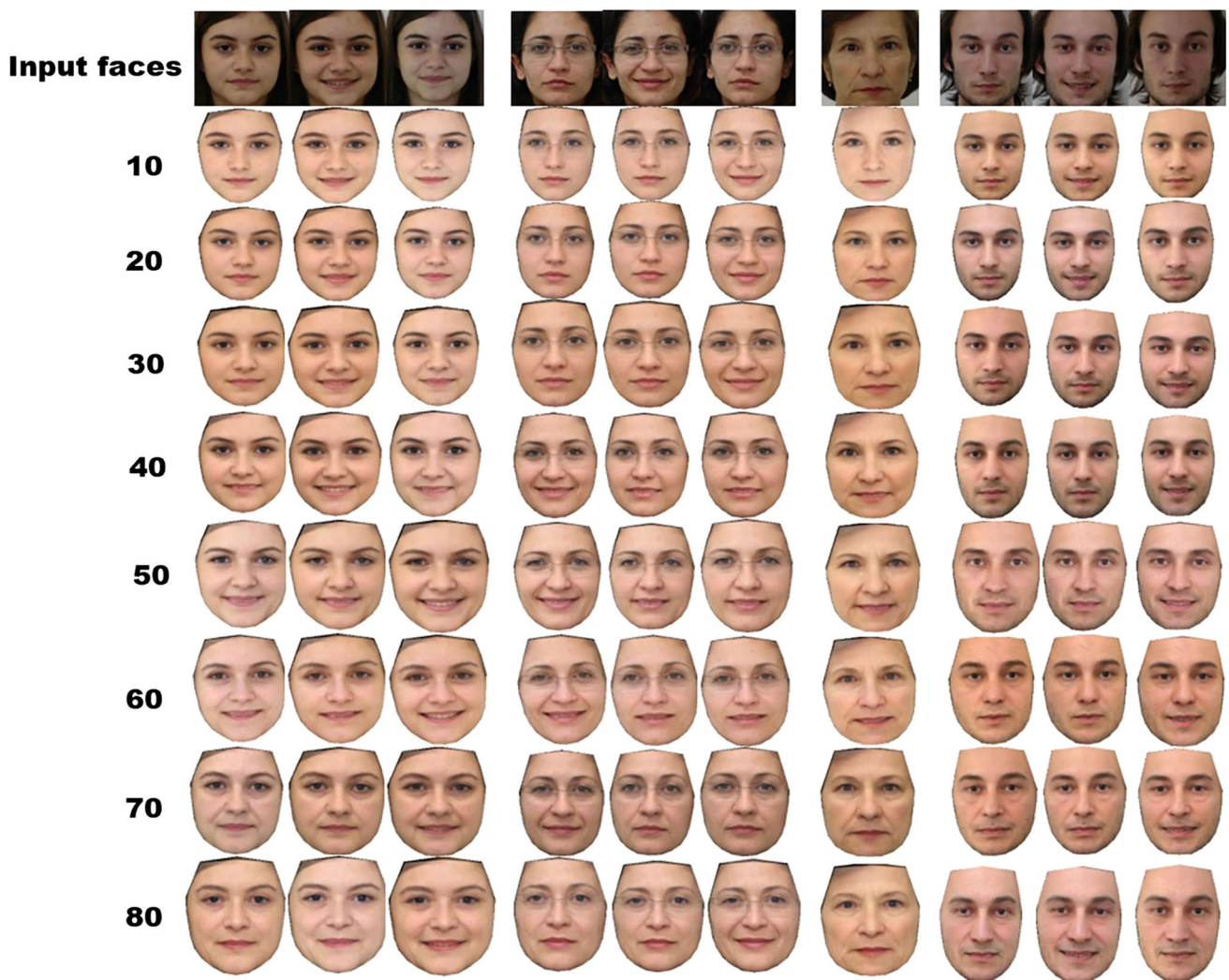


Fig. 19 An example of aged faces using our method—for a selection of facial images from the FEI dataset. The first row shows the input face, and the remaining rows show the resulting images for the age groups from 10 to 80 years

Additionally, we compared our method with RFA and CDL. In this case, images from the FGENT age dataset were selected in order to make the comparison. Our results in Fig. 17 show they are more plausible and more realistic.

Furthermore, we also made a comparison between our method and that utilised in the CAAE system, which uses the UTKFace dataset [31] containing 23,000 images. We have used the CAAE system to age images from samples taken from the FEI dataset. We then compared the resulting images with our method. Figure 18 shows the aged facial images generated by the CAAE system. Figure 19 shows the aged images generated using our method for the same input images. As one can clearly see, our method can produce aged face images which are not only realistic but also more plausible.

5 Conclusion

The proposed approach addresses the problem of computer-assisted facial age progression and regression. The criteria we have subscribed to while searching for a solution to this problem are to design a method which is efficient, computationally lightweight and yet provides us with accurate results. We address this problem by adopting a methodology for creating person-specific ageing templates. The templates are ethnicity, gender and age specific. The templates are based on the formulations of an average face for the corresponding ethnicity, gender and for the predefined range of ages.

We conducted experiments and tested the proposed method using two publicly available face datasets, namely the FEI and the Morph II. We utilised these datasets not only to show that a diverse range of facial images can be generated

using our proposed method but also to verify the accuracy of our results when compared to the images of ground truth. The accuracy of the aged faces was verified through measures of facial similarity between aged faces and the corresponding images of ground truth. This was undertaken using standard CNN-based face recognition with the use of classifiers such as cosine similarity, structural similarity and K -nearest neighbours.

Additionally, we also benchmarked our method with the existing state-of-the-art methods such as those based on GANs, RFA and CDL. Based on the extensive experimentation we have carried out, we can confidently claim that the proposed method for age progression and regression is efficient, lightweight yet accurate when compared to the present state-of-the-art in the field.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Gogić, I., Manhart, M., Pandžić, I.S., Ahlberg, J.: Fast facial expression recognition using local binary features and shallow neural networks. *Vis. Comput.* **36**(1), 97–112 (2020). <https://doi.org/10.1007/s00371-018-1585-8>
- Liu, X., Zhou, F.: Improved curriculum learning using SSM for facial expression recognition. *Vis. Comput.* <https://doi.org/10.1007/s00371-019-01759-7>
- Jilani, S.K., Ugail, H., Bukar, A.M., Logan, A., Munshi, T.: A machine learning approach for ethnic classification: the British Pakistani face, In: International Conference on Cyberworlds (CW), vol. 2017, pp. 170–173 (2017). <https://doi.org/10.1109/CW.2017.27>
- Pereira, Tde Freitas, Anjos, A., Marcel, S.: Heterogeneous face recognition using domain specific units. *IEEE Trans. Inf. Forensics Secur.* **14**(7), 1803–1816 (2019). <https://doi.org/10.1109/TIFS.2018.2885284>
- Jin, Y., Lu, J., Ruan, Q.: Coupled discriminative feature learning for heterogeneous face recognition. *IEEE Trans. Inf. Forensics Secur.* **10**(3), 640–652 (2015). <https://doi.org/10.1109/TIFS.2015.2390414>
- 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). <https://doi.org/10.1007/s00371-017-1428-z>
- Chu, Y., Zhao, L., Ahmad, T.: Multiple feature subspaces analysis for single sample per person face recognition. *Vis. Comput.* **35**(2), 239–256 (2019). <https://doi.org/10.1007/s00371-017-1468-4>
- Fredj, H.B., Bouguezzi, S., Souani, C.: Face recognition in unconstrained environment with CNN. *Vis. Comput.* <https://doi.org/10.1007/s00371-020-01794-9>
- Dehshibi, M.M., Shanbehzadeh, J.: Cubic norm and kernel-based bi-directional PCA: toward age-aware facial kinship verification. *Vis. Comput.* **35**(1), 23–40 (2019). <https://doi.org/10.1007/s00371-017-1442-1>
- Wang, W., Cui, Z., Yan, Y., Feng, J., Yan, S., Shu, X., Sebe, N.: Recurrent face aging. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2378–2386. (2016). <https://doi.org/10.1109/CVPR.2016.261>
- Zhi, R., Liu, M., Zhang, D.: A comprehensive survey on automatic facial action unit analysis. *Vis. Comput.* **36**(5), 1067–1093 (2020). <https://doi.org/10.1007/s00371-019-01707-5>
- Liu, S., Sun, Y., Zhu, D., Bao, R., Wang, W., Shu, X., Yan, S.: Face aging with contextual generative adversarial nets, In: Proceedings of the 25th ACM International Conference on Multimedia, MM '17, Association for Computing Machinery, New York, pp. 82–90. (2017). <https://doi.org/10.1145/3123266.3123431>
- shu, x, Tang, J., Li, Z., Lai, H., Zhang, L., Yan, S.: Personalized age progression with bi-level aging dictionary learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 905–917 (2018)
- Burt, D.M., Perrett, D.I.: Perception of age in adult Caucasian male faces: computer graphic manipulation of shape and colour information. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **259**(1355), 137–143 (1995). <https://doi.org/10.1098/rspb.1995.0021>
- Choi, C.: Age change for predicting future faces. In: FUZZ-IEEE'99. 1999 IEEE International Fuzzy Systems. Conference Proceedings (Cat. No.99CH36315), vol. 3, pp. 1603–1608 (1999). <https://doi.org/10.1109/FUZZY.1999.790144>
- Shu, X., Tang, J., Lai, H., Niu, Z., Yan, S.: Kinship-guided age progression. *Pattern Recogn.* **59**, 156–167 (2016). <https://doi.org/10.1016/j.patcog.2015.12.015>
- Kwon, Y.H., da Vitoria Lobo, N.: Age classification from facial images. *Comput. Vis. Image Underst.* **74**(1), 1–21 (1999). <https://doi.org/10.1006/cviu.1997.0549>
- Shu, X., Xie, G.-S., Li, Z., Tang, J.: Age progression: current technologies and applications. *Neurocomputing* **208**, 249–261 (2016). <https://doi.org/10.1016/j.neucom.2016.01.101>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems*, vol. 27, pp. 2672–2680. Curran Associates Inc., Red Hook (2014)
- Bukar, A.M., Ugail, H., Connah, D.: Individualised model of facial age synthesis based on constrained regression. In: 2015 International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 285–290. (2015). <https://doi.org/10.1109/IPTA.2015.7367147>
- Bukar, A.M., Ugail, H.: Convnet features for age estimation (2017). <http://hdl.handle.net/10454/12860>
- Parkhi, O.M., Vedaldi, A., Zisserman, A., et al.: Deep face recognition. In: *bmvc*, vol. 1, p. 6 (2015)
- Cootes, T., Lanitis, A.: The fg-net aging database (2008). [Online]. <http://www.fgnet.rsunit.com/>. Accessed 07 July 2018
- Ricanek, K., Tesafaye, T.: Morph: a longitudinal image database of normal adult age-progression. In: 7th International Conference on Automatic Face and Gesture Recognition (FGR06), pp. 341–345 (2006). <https://doi.org/10.1109/FGR.2006.78>

25. Riaz, S., Park, U., Choi, J., Natarajan, P.: Age progression by gender-specific 3d aging model. *Mach. Vis. Appl.* **30**(1), 91–109 (2019). <https://doi.org/10.1007/s00138-018-0975-2>
26. Wang, Y., Zhang, Z., Li, W., Jiang, F.: Combining tensor space analysis and active appearance models for aging effect simulation on face images. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **42**(4), 1107–1118 (2012). <https://doi.org/10.1109/TSMCB.2012.2187051>
27. Shu, X., Tang, J., Lai, H., Liu, L., Yan, S.: Personalized age progression with aging dictionary. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), ICCV '15, IEEE Computer Society, Washington, pp. 3970–3978. (2015). <https://doi.org/10.1109/ICCV.2015.452>
28. Chen, B.-C., Chen, C.-S., Hsu, W.H.: Cross-age reference coding for age-invariant face recognition and retrieval. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *Computer Vision—ECCV 2014*, pp. 768–783. Springer, New York (2014). https://doi.org/10.1007/978-3-319-10599-4_49
29. Zhang, Z., Song, Y., Qi, H.: Age progression/regression by conditional adversarial autoencoder. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2017, pp. 4352–4360 (2017). <https://doi.org/10.1109/CVPR.2017.463>
30. Wen, Y., Li, Z., Qiao, Y.: Latent factor guided convolutional neural networks for age-invariant face recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2016, pp. 4893–4901 (2016). <https://doi.org/10.1109/CVPR.2016.529>
31. Kemelmacher-Shlizerman, I., Suwajanakorn, S., Seitz, S.M.: Illumination-aware age progression. In: *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2014, pp. 3334–3341 (2014). <https://doi.org/10.1109/CVPR.2014.426>
32. King, D.E.: Dlib-ml: a machine learning toolkit. *J. Mach. Learn. Res.* **10**, 1755–1758 (2009)
33. Larsen, R.: L1 generalized procrustes 2d shape alignment. *J. Math. Imaging Vis.* **31**(2), 189–194 (2008). <https://doi.org/10.1007/s10851-008-0077-2>
34. Goshtasby, A.: Piecewise cubic mapping functions for image registration. *Pattern Recogn.* **20**(5), 525–533 (1987). [https://doi.org/10.1016/0031-3203\(87\)90079-3](https://doi.org/10.1016/0031-3203(87)90079-3)
35. Shu, X., Tang, J., Lai, H., Liu, L., Yan, S.: Personalized age progression with aging dictionary. In: *The IEEE International Conference on Computer Vision (ICCV)* (2015)
36. Wang, W., Yan, Y., Cui, Z., Feng, J., Yan, S., Sebe, N.: Recurrent face aging with hierarchical autoregressive memory. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(3), 654–668 (2019). <https://doi.org/10.1109/TPAMI.2018.2803166>
37. Grundland, M., Vohra, R., Williams, G.P., Dodgson, N.A.: Cross dissolve without cross fade: preserving contrast, color and salience in image compositing. In: *Computer Graphics Forum*, vol. 25, pp. 577–586. Wiley Online Library (2006). <https://doi.org/10.1111/j.1467-8659.2006.00977.x>
38. Shewchuk, J., Dey, T.K., Cheng, S.-W.: *Delaunay Mesh Generation*. Chapman and Hall/CRC, London (2016). <https://doi.org/10.1201/b12987>
39. Dong, P., Galatsanos, N.P.: Affine transformation resistant watermarking based on image normalization. In: *Proceedings. International Conference on Image Processing*, vol. 3, pp. 489–492 (2002). <https://doi.org/10.1109/ICIP.2002.1039014>
40. Yu, G., Morel, J.: A fully affine invariant image comparison method. In: *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1597–1600 (2009). <https://doi.org/10.1109/ICASSP.2009.4959904>
41. Gao, Y., Ma, J., Yuille, A.L.: Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples. *IEEE Trans. Image Process.* **26**(5), 2545–2560 (2017). <https://doi.org/10.1109/TIP.2017.2675341>
42. Chen, J., Patel, V.M., Chellappa, R.: Unconstrained face verification using deep CNN features. In: *IEEE Winter Conference on Applications of Computer Vision (WACV)*, vol. 2016, pp. 1–9 (2016). <https://doi.org/10.1109/WACV.2016.7477557>
43. Elmahmudi, A., Ugail, H.: Deep face recognition using imperfect facial data. *Future Gener. Comput. Syst.* **99**, 213–225 (2019). <https://doi.org/10.1016/j.future.2019.04.025>
44. Sidorov, G., Gelbukh, A., Gómez-Adorno, H., Pinto, D.: Soft similarity and soft cosine measure: similarity of features in vector space model. *Comput. Sist.* **18**(3), 491–504 (2014). <https://doi.org/10.13053/CyS-18-3-2043>
45. Murthy, S.K.: Automatic construction of decision trees from data: a multi-disciplinary survey. *Data Min. Knowl. Disc.* **2**(4), 345–389 (1998). <https://doi.org/10.1023/A:1009744630224>
46. Zhang, Z.: Introduction to machine learning: k-nearest neighbors. *Ann. Transl. Med.* **4**(11), 218 (2016). <https://doi.org/10.21037/atm.2016.03.37>
47. Amarappa, S., Sathyanarayana, S.: Data classification using support vector machine (SVM), a simplified approach. *Int. J. Electron. Comput. Sci. Eng* **3**, 435–445 (2014)
48. Brunet, D., Vrscay, E.R., Wang, Z.: On the mathematical properties of the structural similarity index. *IEEE Trans. Image Process.* **21**(4), 1488–1499 (2012). <https://doi.org/10.1109/TIP.2011.2173206>
49. imgonline.com.ua. <https://www.imgonline.com.ua/eng/similarity-percent.php>
50. Thomaz, C.E., Giraldi, G.A.: FEI face database (2010). <https://fei.edu.br/~cet/facedatabase.html>. Accessed 23 July 2020
51. How-old.net. <https://www.how-old.net>. Accessed 23 July 2020

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



Ali Elmahmudi is a doctoral researcher at the Centre for Visual Computing in the Faculty of Engineering and Informatics at the University of Bradford, UK. He received his masters degree in medical image segmentation in 2007 from The Libyan Academy. His research interests include image and video processing, pattern recognition and machine learning.



Hassan Ugail is the director of the Centre for Visual Computing in the Faculty of Engineering and Informatics at the University of Bradford, UK. He has a first class BSc Honours degree in mathematics from King's College London and a PhD in the field of geometric design from the School of Mathematics at the University of Leeds. Professor Ugail's research interests include computer-based geometric and functional design, imaging and machine learning.