



# Fully automated age-weighted expression classification using real and apparent age

Nora Al-Garaawi<sup>1</sup> · Tim Morris<sup>2</sup> · Timothy F. Cootes<sup>3</sup>

Received: 30 January 2021 / Accepted: 13 December 2021 / Published online: 6 January 2022  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

## Abstract

After decades of research, automatic facial expression recognition (AFER) has been shown to work well when restricted to subjects with a limited range of ages. Expression recognition in subjects having a large range of ages is harder as it has been shown that ageing, health, and lifestyle affect facial expression. In this paper, we present a discriminative system that explicitly predicts expression across a large range of ages, which we show to perform better than an equivalent system which ignores age. In our system, we first build a fully automatic facial feature point detector (FFPD) using random forest regression voting in a constrained local mode (RFRV-CLM) framework (Cootes et al., in: European conference on computer vision, Springer, Berlin, 2012) which we use to automatically detect the location of key facial points, study the effect of ageing on the accuracy of point localization task. Second, a set of age group estimator and age-specific expression recognizers are trained from the extracted features that include shape, texture, appearance and a fusion of shape with texture, to analyse the effect of ageing on the face features and subsequently on the performance of AFER. We then propose a simple and effective method to recognize the expression across a large range of ages through using a weighted combination rule of a set of age group estimator and age specific expression recognizers (one for each age group), where the age information is used as prior knowledge to the expression classification. The advantage of using the weighted combination of all the classifiers is that more information about the classification can be obtained and subjects whose apparent age puts them in the wrong chronological age group will be dealt with more effectively. The performance of the proposed system was evaluated using three age-expression databases of static and dynamic images for deliberate and spontaneous expressions: FACES (Ebner et al., in Behav Res Methods 42:351–362, 2010) (2052 images), Lifespan (Minear and Park in Behav Res Methods Instrum Comput 36:630–633, 2004) (844 images) and NEMO (Dibeklioglu et al., in: European conference on computer vision, Springer, Berlin, 2012) (1,243 videos). The results show the system to be accurate and robust against a wide variety of expressions and the age of the subject. Evaluation of point localization, age group estimation and expression recognition against ground truth data was obtained and compared with the existing results of alternative approaches tested on the same data. The quantitative results with 2.1% error rates (using manual points) and 3.0% error rates (fully automatic) of expression classification demonstrated that the results of our novel system were encouraging in comparison with the state-of-the-art systems which ignore age and alternative models recently applied to the problem.

**Keywords** AFER · Automatic age-group estimation · Automatic FFPD-based RFRV-CLM · Age features effect on AFER

## 1 Introduction

Automatic Facial Expression Recognition (AFER) has become a crucial area of research in the field of computer vision and machine learning and has many applications such as in human–computer interaction and human behaviour

understanding [10, 36, 41, 43]. Though much progress has been made on the automatic recognition of facial expressions [31, 38], recently researchers in psychology have realized that human ageing has a significant effect on understanding facial expression [16, 18, 21, 22], and thus, it is considered to be one of the causes of poor performance in AFER systems [14, 18, 20, 29, 32, 34, 39, 42, 44].

Ageing causes changes in facial musculature and skin elasticity which distort the expression cues. Age-related structural changes can overlap with facial

✉ Nora Al-Garaawi  
noora.algaraawi@uokufa.edu.iq

Extended author information available on the last page of the article

expression-induced changes. For example, the fold between the cheek and upper lip can appear in both the happy expression of young people and the neutral expression of old people, and sagging eyelids in old people can appear similar to the sad expression of young people, as illustrated in Fig. 1. Therefore, providing an AFER system that can reduce the effect of that overlapping between the age and the expression features and can work across a large range of ages is necessary. Such a system can be an important aid for classifying the expressions of elderly people and so understanding their feelings to provide better care.

The first purpose of this paper is to reduce the negative effect of age appearance on the overall performance of the AFER system. Therefore, the *first contribution* of this paper is the creation of a new framework for the AFER system that can explicitly estimate the expression across a large range of ages and performs better than an equivalent system that ignores ages. This contribution is achieved by using a weighted combination of a set of age groups and age-specific facial expressions classifiers. The advantage of using a combination of all classifiers is that more information about the classification can be obtained, and subjects whose apparent age differs from their chronological age group are dealt with more effectively.

One way to build the AFER system is by using of an automatic facial feature point detector (FFPD) to detect the facial features points in order to use the detected points and their locations (x and y coordinates) to extract both the shape and texture features from the patch around each point or from the whole face image and then to use them for building an

expression classifier. Detecting the facial feature points is a challenging task owing to rigid face deformations (scaling, rotation, translation) and nonrigid face deformations such as age and expression as in the present work. Therefore, the *second contribution* of this paper is the creation of a fully automatic FFPD system that can work across a large range of ages and expressions which is used in this paper to both study the influence of ageing on the performance of FFPD and build the AFER system. To the best of our knowledge, this is the first study on analysing the sensitivity of the point localization task to age. The underlying hypothesis is that automatically finding the optimal position of facial feature points in the input image with small mislocalization error will help to obtain robust and accurate results for further analysis such as feature extraction and facial expression recognition.

Recently, apparent age (how old does a person look) has been used as a new measurement for the age estimation task. The only difference between real age and apparent age is that the label of the apparent age is provided by an assessor, whereas the label for the real age is the chronological age of the person [4, 9, 23, 37, 40, 45]. Inspired by this idea, the *third contribution* of this paper is to study the influence of using the apparent age on the overall performance of AFER system. The underlying assumption is that the advantages that are brought by the apparent age to the age estimation task might by consequence enhance the performance of the facial expression recognition task against the impact of ageing since both tasks are correlated.

## 2 Related works

Most of the previous AFER approaches have focused on recognizing emotions using subjects from a restricted age range and ignored the changes in facial appearance induced by age [31, 38]. This was done using a facial expression database of limited ages such as the JAFFE database [30], which contains only the expressions of 10 young Japanese women, and the CMU facial expression database [24], which mostly has images of young adults and some middle aged adults.

The recent review published by [18] suggests that age has a significant effect on the appearance of facial expression and plays an important role in the process of facial expression analysis. The lower accuracy obtained in the analysing of facial expressions of older people might be due to several factors, such as lower expressivity, age-related facial changes, less elaborated emotion schemas, etc [18]. In addition, the psychological study developed by [16] demonstrated that the characteristics of emotional faces are different in different age groups. Furthermore, facial expression signal produced by older people is not as clear as those produced by young people due to wrinkles



**Fig. 1** Illustration of the similarities between age and expressions appearances: the fold between the cheek and upper lip of old people (top left) appears similar to the happy expression of young people (top right), and the sagging eyelids in elderly people with a neutral expression (bottom left) appears similar to the sad expression of young people (bottom right). The images are from the FACES dataset [17]

and folds and hence, might affect the behavioural inference of other people [21]. Motivated by the effects of human ageing on the meaning of expression, psychobiology researchers have created three age-expression databases: FACES [17], LifeSpan [33], and NEMO [13]. These databases have allowed computer vision researchers to shift their focus from building an AFER system with high expression recognition accuracy to create a more generalized and effective system capable of recognizing expressions across a large range of ages, such as those developed by [3, 14, 20, 29, 42, 44].

Guo et al. [20] were the first to study and analyse the influence of age features on the performance of AFER. In their study, subjects were divided into four age groups and each expression in each age group was treated as a separate class. The authors manually labelled 31 fiducial points and applied Gabor filters at those points and trained a kernel SVM to recognize expressions from the extracted features. They then proposed removing ageing details before facial expression recognition by using smoothing methods. Experimental results on FACES and LifeSpan databases demonstrated that the age has a significant influence on the accuracy of facial expression classification. In spite of the good results obtained by the proposed method in [20], the authors relied on the manually placed points while the demand of any automatic system is to generate those points automatically which is a difficult task especially with the presence of non-rigid face deformations related to the age and expression. Furthermore, the authors in [20] extracted the texture features only where the micropatterns in skin texture that are important for age estimation and expression recognition can encode but are negatively affected by identity bias. Moreover, the authors eliminate the effect of age features on the performance of expression recognition by deleting the age information using image smoothing techniques before expression recognition. Using image smoothing techniques might lead to the loss of information related to age and expression

In [42], the authors manually labelled facial features points and extracted the geometric features at those points and trained a Dynamic Bayesian Networks (DBN) to recognize expressions from the extracted features. They then proposed to reduce the effect of age features on the performance of AFER by using age information as prior knowledge to the expression recognition, in which the age information is required during training only in order to avoid the error in age estimation and subsequently on the expression recognition. Experimental results on FACES and LifeSpan databases demonstrated that age has a significant influence on the accuracy of facial expression classification. In spite of the good results obtained by the proposed method, the authors relied on the manually placed points. Furthermore, the authors extracted the geometric features only. Using the

geometric feature only might be sensitive to registration error.

In [3] and [29], the authors studied the impact of age features on the performance of age and expression classification task. They proposed a graphical model with a latent layer between age and expression to learn the relationship between the two. For feature extraction, the local binary pattern (LBP) features were used. Multi-class support vector machine (MC-SVM) was then used for age and expression classification. Evaluation results using FACES, LifeSpan, and NEMO databases showed an improvement in the performance of age estimation when age and expression are jointly learnt, compared to the age estimation system that ignores expression. The results also showed an improvement in the facial expression recognition when the age is jointly learnt with expression in comparison with the facial expression recognition system which ignores age. In spite of the good results obtained by the proposed method, the authors extract the texture features only which are negatively affected by identity bias. Moreover, the authors eliminate the effect of age features on the performance of expression recognition by using age information as prior knowledge to the expression recognition, in which the age information is required during training only.

Dibeklioglu et al. [14] studied the usefulness of using age features along with the other features in distinguishing between posed and spontaneous smiles. Experimental results using the NEMO database proved that using age significantly helps to differentiate between posed and spontaneous expressions. Wang et al. [42] proposed a probabilistic model using a Bayesian network (BN) to classify expressions with the help of age features. For facial expression feature extraction, they manually labelled 18 fiducial points, and geometric features were extracted from which a BN was trained to recognize expressions. They then proposed using multiple BNs to capture the spatial information of expression patterns. Experimental results on FACES and LifeSpan databases demonstrated that the proposed model for using geometric features only has achieved comparable performance to the previous methods of using texture features only [3, 14, 20, 29]. Here, the authors also relied on the manually placed points and they extracted the geometric features only.

In [44], a deep multi-task learning model is proposed which consists of two parallel columns composed of ConvNet and ScatNet, two fully connected layers, and an output layer. ConvNet and ScatNet provide feature representations shared by the subsequent tasks. The multi-task learning formulation is employed to simultaneously learn to predict age and to classify expression. Experimental results on FACES and LifeSpan databases demonstrated that the proposed model of using deep features, where the features are learned from the training data, has successfully estimated the ages and recognized expressions with comparable or better

performance than previous methods that used features that are designed beforehand by experts. In spite of the good results obtained by the proposed method, the authors extract the deep features from a small dataset as in the FACES databases where there are only 114 images per expression per age which excludes the possibility of reaching a convincing conclusion since deep representation requires a huge amount of data to implement and train. That is why the authors used the pre-trained model on the MORPH dataset (with 55,134 images of age only) to extract the age and expression features. Different accuracy might be obtained if the DL model was trained on the FACES dataset only.

In addition, from the above literature, we can observe that all the previous studies are reliant on the real age (chronological age) for dividing the data into groups and then for training and testing the expression recognition classifiers; an error can occur if there is a difference between the real age (chronological age) and apparent age since recently there is a concern that the apparent age of the individual might be different from his/her real age [4, 9, 23, 37, 40, 45]. Table 1 summarizes the work that has been done on facial expression recognition across a large range of ages.

Motivated by these limitations, and to eliminate the effect of overlap of the age and expression features on the overall accuracy of AFER, this paper introduces a new system using a simple and effective method to recognize the expression automatically across a large range of ages. Our system uses a weighted combination of a set of age group and age-specific facial expression classifiers. It is an end-to-end automated system to detect facial expression features, extract shape and texture features as well as appearance features, for both age and expression, estimate the age group, and recognize the expression. To meet the demand of any automatic system, we present a fully automated facial expression points localization (FEL) system used to both analyse the sensitivity of facial feature point localization to age and to extract the age and expression features automatically. To the best of our knowledge, this is the first study on analysing the sensitivity of the point localization task to age. For compact representation, we study the use of shape and texture features, both individually and in combination (appearance), obtained from both the manual and automatic points to analyse the effect of age and to recognize expression since both shape and

texture features are subject to the ageing's effect. To model age effects on expression, we propose to combine the performance of an age group estimator with age-specific facial expression classifiers (one trained for each age group) in a single framework. The underlying hypothesis is that an age-specific facial expression classifier will outperform a general classifier. In the proposed system, the age information is required during training and testing. To avoid the error in the age estimation task due to differences between the real and apparent age of some individuals, we propose a simple and effective method for recognizing expressions across a large range of ages using a combination of age group classifier and age-specific facial expression classifiers. The advantage of using a combination of all classifiers is that more information about the classification can be obtained, and subjects whose apparent age differs from their chronological age group are dealt with more effectively.

The remainder of this paper is organized as follows. Section 3 gives a description of the framework of the proposed AFER system. The databases used in the current work are then described in Sect. 4. Section 5 reports the experimental design and results, and Sect. 6 presents conclusions and future works.

### 3 Methods

The proposed method estimates the age group and the expression automatically from a facial image. The stages of the process are (i) locate the facial feature points and then extract the age and expression features, (ii) estimate the age group (iii) recognize the expression given the age group. Figure 2 describes the stages of the proposed system. See text in Sects. 3.1, 3.2 and 3.3 for more details about stages 1, 2 and 3, respectively.

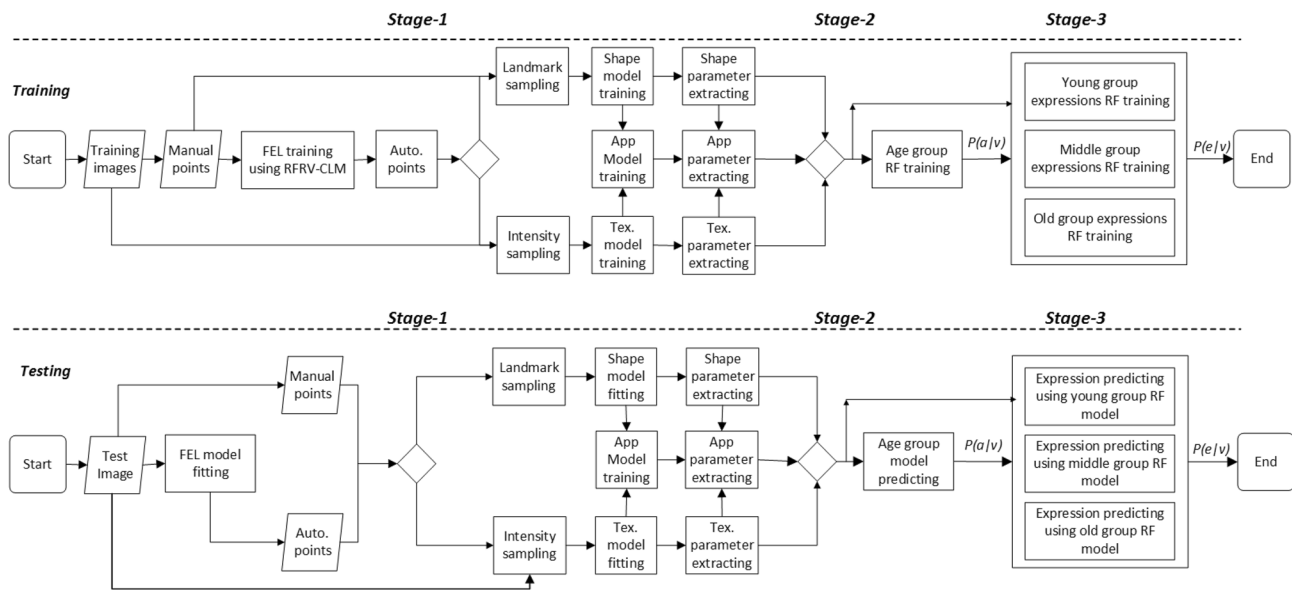
We show that an age-specific expression classifier works better than an age-agnostic one for a known age, using an estimate of the age helps to classify the expression across a large range of ages, and that using the weighted combination of all the age-specific expression classifiers helps to avoid errors related to the ageing effect and apparent age on the total performance of the AFER system.

#### 3.1 Stage 1: facial expression localization and feature extraction

We use the RFRV-CLM method [12, 26, 28] to build an automatic FEL system. RFRV-CLM has been applied successfully to automatic landmark points localization in face [12, 26, 28] and clinical [6–8, 27, 28] images. In this paper, we introduce, validate, and re-optimize the RFRV-CLM method for the problem of facial expression localization and test its ability to localize the face features under a large

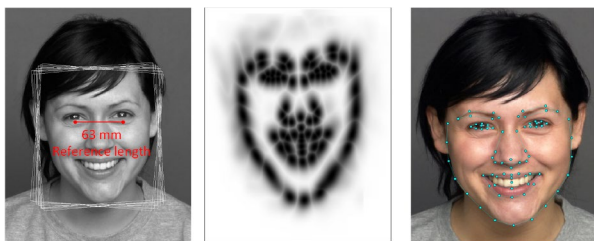
**Table 1** Recognition rate of recent studies on the present problem with the three databases

Refs.	Features	Classifier	FACES	Lifespan	NEMO
[20]	Gabor-filters	SVM	97.89	96.79	–
[42]	Geometric	DBN	95.74	96.53	–
[29]	LBP	MC-SVM	92.19	93.68	98.0
[44]	Deep	CNN	95.13	96.32	–



**Fig. 2** System overview showing our three stage system. The first stage is the automatic points detector and feature extractor to extract features vectors  $v$  containing both age and expression sent to age

group estimator (stage 2) to estimate the age group and to age-group-specific expressions classifiers (stage 3) to estimate the expression category



**Fig. 3** Fully automatic facial expression point detector system using RFRV-CLM: Global searcher to estimate the approximate position of the 2-points corresponding to the eyes’ centres (left), Local search to localize the 76-points around the face components (right), and superposition of model votes of the 76 points (middle). The local search uses the highlighted red points from the global model(left) to initialise its search

range of ages and expressions using a small amount of data per age per expression.

The proposed FEL system consists of global and local searches (models) to locate the facial feature points and then to extract the features. The global model is performed using a random forest voting method [28] to detect the approximate position of the eye centres. The local model uses the detected eye points to initialise the local search to locate all facial feature points using RFRV-CLM. Figure 3 shows the output of the global and local models.

During training, each image in the training set is labelled with  $n = 76$  feature points, a feature point  $l$  in an image is represented by  $(x_l, y_l)$  landmark point, where  $l = 1 \dots n$ , The resulting face shape vector  $x$  of length  $2n$  can be encoded as

$$x = (x_1, \dots, x_n, y_1, \dots, y_n)^T \tag{1}$$

A statistical shape model is trained by applying Principal Component Analysis (PCA) to the aligned facial shape vectors, creating a model with the form:

$$x = T(\hat{x} + P_s b_s; t) \tag{2}$$

where  $x$  represents the shape vector in the reference frame,  $\hat{x}$  represents the mean shape;  $P_s$  is a matrix of the set of eigenvectors corresponding to the highest eigenvalues, which describe different modes of variation;  $b_s$  is a set of parameter values of the shape model; and  $T_\theta(\cdot; t)$  applies a global similarity transformation with parameters  $t$ .

The RVRV-CLM manipulates the shape ( $b_s$ ) and pose ( $t$ ) parameters to best fit the points to the image in a coarse-to-fine manner [12]. The shape of the face is then encoded in the parameters  $b_s$ .

To represent the facial texture, we build a texture model [11]. For each example in the training set, we warp the face into a reference frame defined by the mean shape and then sample at regular positions to obtain a vector of intensities  $g$ . We normalise each vector and then apply PCA to obtain a texture model of the form

$$g = \hat{g} + P_g b_g \tag{3}$$

where  $b_g$  is a vector of weights on the modes  $P_g$  [11]. The texture of a new example can be encoded as the vector  $b_g$  which best fits such a model to the intensities from the sample.

To extract the appearance features in which the correlation between shape and texture features is learned, an appearance model [11] is built by applying another PCA to the concatenation of the shape  $\mathbf{b}_s$  and the texture  $\mathbf{b}_g$  parameters. The concatenation is performed in a weighted form to compensate for the difference in units and generate the concatenated appearance vector  $\mathbf{b}_a$  :

$$\mathbf{b}_a = \begin{pmatrix} \mathbf{W}\mathbf{b}_s \\ \mathbf{b}_g \end{pmatrix} \tag{4}$$

where  $\mathbf{W}$  is a diagonal matrix of weights to account for the difference in units between the shape and texture models and the  $\mathbf{W}$  is chosen to balance the total variation in shape and texture,

$$\mathbf{W} = \left( \frac{\text{TotalVar}(\text{texture})}{\text{TotalVar}(\text{shape})} \right)^{\frac{1}{2}} \mathbf{I} \tag{5}$$

Applying PCA on the concatenated vectors gives the appearance model:

$$\mathbf{b}_a = \mathbf{p}_c \mathbf{c} \tag{6}$$

where  $\mathbf{p}_c$  are the eigenvectors and  $\mathbf{c}$  is a vector of appearance coefficients describing both the shape and texture of the face

We then combine (fusion) the shape feature  $\mathbf{b}_s$  with the texture feature  $\mathbf{b}_g$  to create the feature  $\mathbf{d}$  as follows:

$$\mathbf{d} = (\mathbf{b}_s^T | \mathbf{b}_g^T)^T \tag{7}$$

The shape,  $\mathbf{b}_s$ , texture  $\mathbf{b}_g$ , appearance  $\mathbf{c}$  together with their combination  $\mathbf{d}$  are used as features vectors from which a random forest classifiers are trained to distinguish among age groups and expressions.

### 3.2 Stage 2: age group estimation

Suppose we have  $N$  labelled training images, where for each image  $I_i$  we have a feature vector  $v_i(\mathbf{b}_s, \mathbf{b}_g, \mathbf{c}, \mathbf{d})$  (from stage 2), an age index  $a_i$  indicating the age group and an expression index  $e_i$  indicating the expression label. Thus,  $a \in \{1, 2, 3\}$  for young, middle, and old age groups, respectively, and  $e \in \{1, 2, 3, 4, 5, 6\}$  for angry, disgust, fear, happy, neutral and sad.

We train a random forest [5] to estimate the probability that an example with feature vector  $v$  belongs to age group  $a$ ,  $p(a|v)$ . The most probable age group is given by

$$f_a(v) = \operatorname{argmax}_a p(a|v) \tag{8}$$

We combined data from three age and expression databases described in Sect. 3 to create a data set which contains subjects with ages ranging from 8 to 94 years. Figure 4 shows the age distribution. For the experiments below, we divided

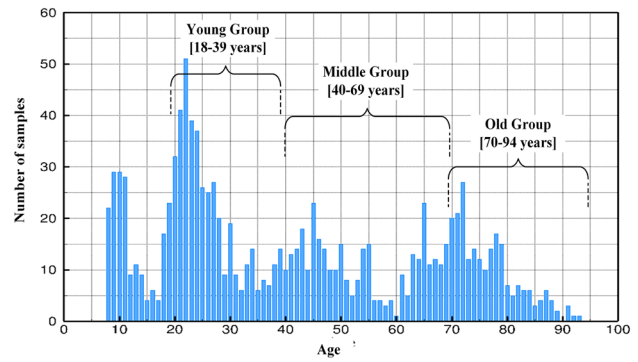


Fig. 4 Age distribution of three databases: FACES, Lifespan and NEMO

the data into three groups based on age: Young (18–39), Middle (40–69), Old (70–94). Subjects with ages from 8 to 17 were not included because there were too few samples.

### 3.3 Stage 3: expression classification

For each age group,  $a$ , we trained a separate random forest to estimate the probability of each expression,  $p(e|v, a)$ . One approach for classifying a new image (of unknown age) is to estimate the age group using Eq. (8) then use the appropriate expression classifier for that age, that we called the hard-level-based or real-age-based schema:

$$f_e(v) = \operatorname{argmax}_e p(e|v, f_a(v)) \tag{9}$$

An alternative is to weight by the probability, that we can call a soft-level-based or apparent-age-based schema:

$$f_e(v) = \operatorname{argmax}_e p(e|v), \tag{10}$$

where

$$p(e|v) = \sum_a p(a|v)p(e|v, a). \tag{11}$$

For the experiments reported in this paper, we trained three age-specific expression classifiers for young, middle, and old age groups, respectively.

The advantage of using the weighted combination of all classifiers is that more information about the classification can be obtained, and subjects whose apparent age puts them in the wrong chronological age group are dealt with more effectively.

## 4 Databases

**Data:** To assess the proposed system, we used three datasets: FACES [17], Lifespan [33] and NEMO [13]. Some example faces are shown in Fig. 5. The datasets are designed

for research into both age and expression estimation. (See Table 2).

The FACES dataset contains 171 people showing six expressions (anger, disgust, fear, happy, neutral and sad), totalling 2052 frontal face images.

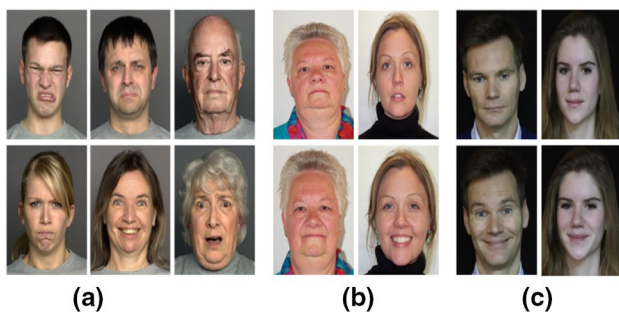
The Lifespan dataset contains face images from different ethnicities showing eight different expressions with different subsets sizes: neutral (N = 580), happiness (N = 258), surprise (N = 78), sadness (N = 64), annoyed (N = 40), angry (N = 10), grumpy (N = 9), and disgust (N = 7). We follow [20] and use the neutral and happy subset.

The NEMO dataset consists of 400 male and female subjects and 1240 videos. The age of the subjects ranges from 8 to 76 years. Each subject recorded several videos of their expression changing from neutral to happy, both spontaneously and deliberately. In this paper, each video is converted to frames and used as independent images.

*Landmarks* We manually annotated 2052 images from the FACES data with 76 landmark points, and we automatically detect the points of the Lifespan and NEMO datasets using a model trained on the FACES dataset. We will make the points available to other researchers for further studies.

## 5 Experiments

We performed a series of experiments to investigate the effect of age appearance on different components of the AFER system including landmark localization, age group estimation, and facial expression classification.



**Fig. 5** Sample face images depicting different expressions from three databases: **a** FACES, **b** Lifespan and **c** NEMO

**Table 2** Description of age-expression datasets: FACES, Lifespan, NEMO.Exam., Exp., Res. and Mod. are the examples number, expressions number, persons number, frame resolution and database modality, respectively

Data	Age	Exam.	Exp.	Per.	Res.	Mod.
FACES	19–31	969	6	58	335×419	Image
	39–55	672	6	56	335×419	Image
	69–80	684	6	57	335×419	Image
Lifespan	18–93	100	8	230	1760×1168	Image
NEMO	8–76	1240	2	400	1920×1080	Video

For the landmark localization, the accuracy was tested by comparing the locations of automatically detected points against the manual annotation (ground-truth) across each test image. Errors are given as the mean point-to-point error as a percentage of the inter-ocular distance (IOD) in humans (around 63 mm [15])

$$e_n = \frac{1}{nd_{IOD}} \sum_{i=1}^n |p_a^i - p_m^i| \tag{12}$$

where  $p_m^i$  is the manually annotated location of point  $i$ ,  $p_a^i$  is the automatically detected point, and  $d_{IOD}$  is the interocular distance, the distance between the centres of the eyes,  $|\hat{x}_{lefteye} - \hat{x}_{righteye}|$ . We further estimated the error in millimetres, using the average IOD length in humans of 63 mm. Each fiducial point detector is evaluated using a twofold cross-validation (training on 50% the data, testing on the other 50% then swapping the sets) to build an automatic FEL system. The automatic points for all datasets are found, and used to investigate the age effect on point localization.

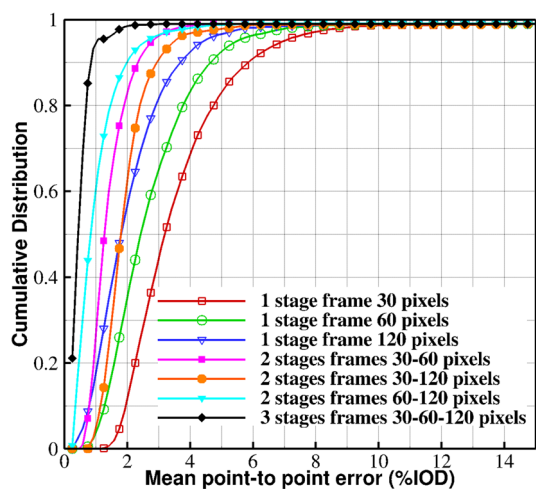
For age group estimation and expression classification, we report both the average of all classes and per-class (the average of the diagonal of the confusion matrix) classification accuracy between the ground truth label and the predicted label. Each classifier (random forests with 100 trees) is evaluated using 10-fold cross-validation.

### 5.1 Facial landmark localization

A series of twofold cross-validation experiments (subject independent) on the FACES dataset are performed to find the optimal RFRV-CLM parameters for facial expression recognition. The best quality of fits are found using a series of three stages (30-60-120) of increasing resolution of frame width as shown in Fig. 6 (see the supplementary material for more details).

We applied the resulting model trained on the FACES database to the Lifespan and Nemo datasets described in Sect. 4, to locate 76 points on each image. Qualitatively the results look satisfactory (see Fig. 8 for an examples of the automatic points).

The resulting automatic points of the three databases are used to extract shape and texture features for the age and expression estimation experiments described below.



**Fig. 6** Performance optimization of a coarse-to-fine multi stages RFRV-CLM on FACES dataset

## 5.2 Age effect on automatic landmark localization

In this experiment, we studied the sensitivity to age effects of automatic fiducial landmark localization. In other words, we investigated the performance of the automatic landmark location on faces from different age groups. We split the subjects in each age group into training (50%) and test (50%) sets. We trained three age-group specific RFRV-CLMs, one for each age group. We then tested each model on each of the age group test sets. We also built an age-agnostic (mixed) model by combining the training sets from all ages.

Figure 7a–c summarizes the results for each age-specific model. The results from the age-group-specific detectors show that for all models the errors increase as the target age increases. Each age-specific model works better on the younger test group than the older test group, suggesting that it is harder to locate features accurately on older faces than younger.

Figure 7d compares the performance of a model trained on all ages (age-agnostic) model to the age-specific point detectors. It shows that the age-agnostic model works almost as well as the specific models for each age range. We thus use the age-agnostic models for locating the points in subsequent experiments, as it avoids the requirement to know the target age.

## 5.3 Age effect on AFER

In this experiment, we studied the sensitivity of automatic facial expression classification to age appearances. We performed experiments with both the manually located points and those from the automatic system. In each case, we used feature vectors which were either the shape, texture appearance or the concatenation of shape and texture parameters.

We trained age-specific random forests to estimate the probability of each expression given a feature vector and age group,  $p(elt, a)$ , and age-agnostic RFs to estimate the probability of each expression for all age groups,  $p(elt)$ .

The performance of the age-specific and age-agnostic classifiers, for each type of feature, on each age group is summarized in Table 3 including the results where the landmark points were manually or automatically placed. The results in Table 3 show that (i) performance is best on the age-group for which the system was trained and degrades as the age difference increases, (ii) performance on the older group is worse than that on the young and middle-aged groups, (iii) using the concatenation of shape and texture features gives the best overall results, (iv) when using the correct age-group classifier, there is only a small loss of performance when using features from automatically placed points, rather than those from the manually placed points (Man–Auto lines).

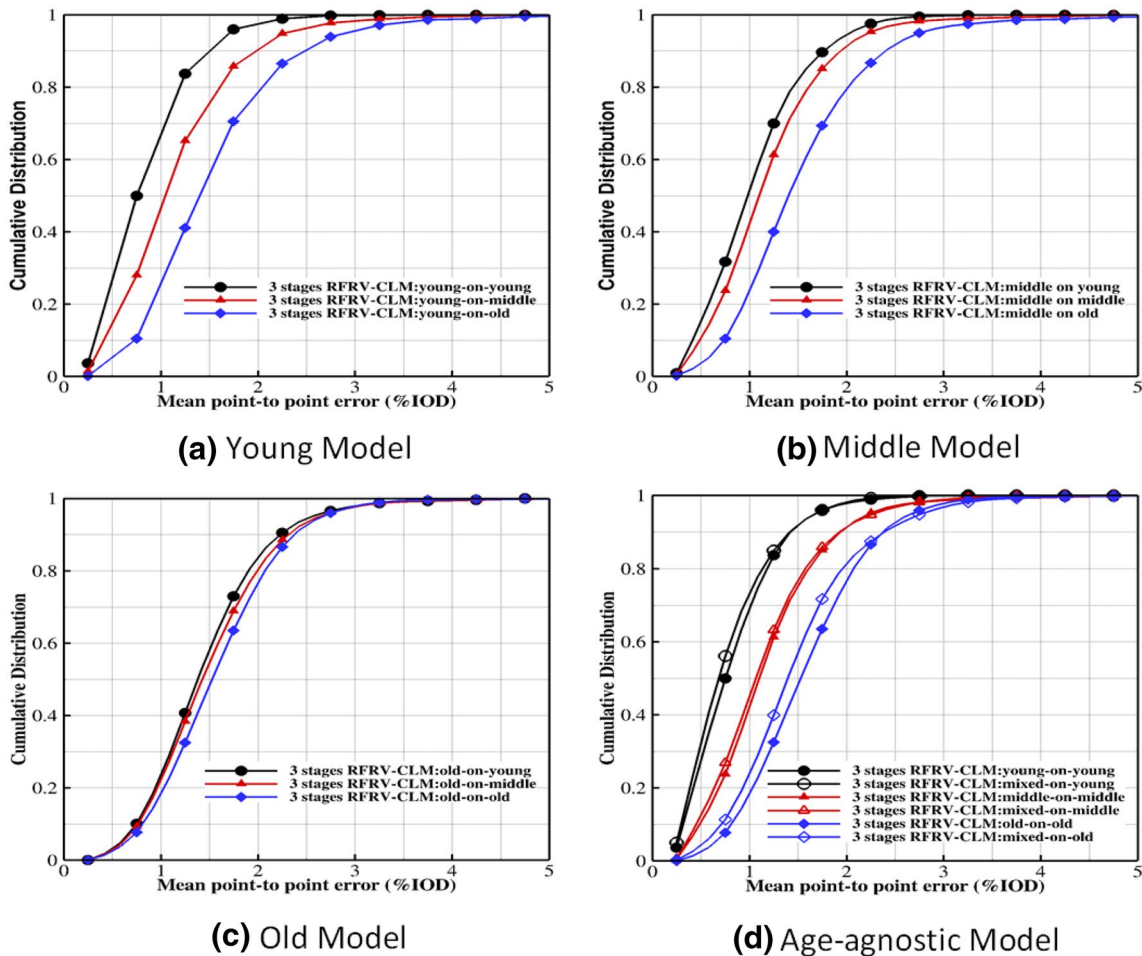
Thus, if we use the most appropriate age-specific classifier for each subject when three age facial expression classifiers are trained and tested using the same age range (each classifier is trained and tested on the same age range) (the age is known and selected manually by the user), we get improved performance, compared to using an age agnostic system.

We further evaluate the effect of the ageing on the system performance using the confusion matrices of automatic points and combined shape and texture features as described in Table 4 for the young age group, middle age group, old age group, and age-agnostic classifiers in each sub-table the training and testing data are from the same age limits. Results in the confusion matrix tables show that there is an age-related decline in automatic expression recognition which is mainly in negative emotions such as the sad expression (see the last row of each sub-table). We found that these results are consistent with those of psychological studies reported by [1, 2, 19, 25], where they try to analysis the relationship between emotional experiences and emotion recognition. The reason for that age-related differences in expression recognition is due to the age-related decrease in experiencing negative emotions in general. Indeed, the study reported by [35] demonstrated that most of the age-related difference in facial expression recognition is due to the lesser intensity of negative emotions in older adults, which will affect the facial expression appearance and hence, effect the automatic expression recognition.

## 5.4 Sensitivity of AFER to real and apparent age

Results in the previous section (see results in Table 3) suggest that the best performance would be achieved using age-group-specific classifiers if the age of the subject is known. When the subject's age is not known, it must be





**Fig. 7** CDFs of the mean point-to-point errors of the 76-point 3-stage RFRV-CLM age-group-specific detectors: **a** trained on the young age group data and tested on the young, middle, and old age groups data, **b** trained on the middle age group data and tested on the young, middle, and old age groups data, **c** trained on the old age group data and tested on all three age groups data, and **d** comparing age-group-specific detectors to the age-agnostic detector: the **black lines** represent the young group error when tested with the young-group model and age-agnostic model, the **red lines** represent the middle group error when tested with middle-group model and age-agnostic model, and the **blue lines** represent the old group error when tested with the old-group model and the age-agnostic model

estimated from the image. Thus, we trained a random forest to estimate the age-group given different feature types, and the performance is summarized in Table 5. The right of Table 5 shows that using combined shape and texture features gives the best overall performance, and that manual annotations lead to slightly more accurate results than automatic annotations. The left of Table 5 shows the confusion matrix for the age-group classifier using automatic points of the combined of shape and texture features. These results reveal that 6.8% of the young group and 9.6% of the old group were classified in the middle age group, and 12.3% and 16.2% of the middle group were classified in the young and old age-groups, respectively. This confusion among age groups might be caused either by failure of the classification algorithm or by the discrepancy between real and apparent age, thus taking the

apparent age into account might enhance the performance of the system.

*Real age* The results in Table 6 (fourth column) show the performance of expression classification when combining the age-group estimator with age-specific expression classifiers in a single model. In this hard-level-based or real-age-based schema, the age classifier is used to automatically select the age-specific expression classifier to use, as in Eq. 9. These results show that although the accuracy of the age group estimation is relatively low (84%)(see Table 5), its combination with age-specific expression classifiers helped to achieve comparable performance, (with 95.2% manually and 93.8% automatically; fourth column of Table 6), to the age-group specific models (with 96.4% manually and 93.2% automatically; third column of Table 6). The above results demonstrate that the

**Table 3** Expression classification results using manual and automated annotation for age-specific models and the age-agnostic model. S+T is the concatenation of both shape and texture feature vectors. Man–Auto is the difference between the manual-based and automatic-based performance

Group	Features	Test																
		Young			Middle			Old			All							
		Shape	Tex	App	S+T	Shape	Tex	App	S&T	Shape	Tex	App	S+T	Shape	Tex	App	S+T	
Train	Young	Manual	96.8±0.4	96.3±0.2	97.0±0.4	97.7±0.4	89.0±1.6	85.6±1.1	86.2±1.5	88.4±1.1	78.8±0.6	65.4±0.5	77.9±0.5	78.4±0.4	82.5±0.8	75.0±0.7	86±0.7	85.7±0.5
	Auto-	91.9±0.8	95.9±0.6	95.7±0.6	96.5±0.4	67.5±1.8	77.1±1.3	76.9±1.9	80.2±1.6	80.2±1.6	61.9±1.2	66.7±1.0	72.9±0.9	75.0±0.9	68.5±2.2	70.8±1.7	71.7±1.8	75.2±1.4
	matic	4.9	1.4	1.3	1.2	21.5	8.5	9.3	8.2	8.2	16.9	1.3	5	3.4	14	4.2	14.3	10.5
Middle	Manual	85.0±0.5	83.3±0.4	91.0±0.4	91.7±0.4	95.0±0.7	95.6±0.6	97.0±0.6	97.6±0.4	89.3±0.5	89.1±0.7	88.2±0.7	90.4±0.7	84.5±1.3	80.8±1.1	85.9±1.4	88.7±1.3	
	Auto-	85.6±0.7	80.0±0.6	80.9±0.6	84.3±0.6	87.1±0.5	93.3±0.4	92.2±0.8	94.7±1.1	75.0±0.6	84.5±0.9	86.6±0.8	89.3±0.8	68.5±1.7	75.8±0.7	70.8±0.9	75.0±0.9	
	matic	0.6	3.3	10.1	7.4	7.9	2.3	4.8	2.9	2.9	14.3	4.6	1.6	1.1	16	5.0	15.1	13.7
Old	Manual	81.7±0.5	73.3±0.3	83.9±0.5	85.0±0.4	88.9±0.4	87.5±0.3	87.8±0.5	89.9±0.3	89.2±1.0	91.1±0.9	92.5±0.9	93.8±1.5	77.8±1.1	71.2±0.9	80.9±0.9	83.9±1.2	
	Auto-	61.0±0.4	73.8±0.6	76.4±0.4	77.8±0.3	63.9±1.0	75.0±0.9	74.2±1.1	77.0±1.1	78.2±1.1	86.0±0.5	86.4±1.1	88.5±2.1	66.7±1.5	66.7±0.4	64.8±1.3	65.9±0.6	
	matic	20.7	0.5	13.6	7.2	25.0	12.5	13.6	12.9	12.9	11.0	5.1	6.1	5.3	11.1	4.5	16.1	18.0
All	Manual	89.0±0.4	87.3±0.4	86.5±0.5	88.1±0.4	86.4±0.6	88.9±0.4	90.0±0.5	91.1±0.6	87.5±0.9	79.8±0.7	85.7±0.8	87.5±0.9	89.0±0.6	91.0±0.5	93.4±0.8	96.4±0.8	
	Auto-	83.9±1.1	86.9±0.9	83.1±0.7	84.5±1.2	81.5±1.0	88.1±1.3	81.4±1.1	84.5±1.3	70.8	77.4±0.9	71.2±1.1	73.2±1.2	76.8±0.6	87.2±0.7	90.1±0.9	93.2±1.2	
	matic	5.1	0.4	3.4	3.6	4.9	0.8	8.9	6.6	6.6	16.7	2.4	14.5	14.3	12.2	3.8	3.3	3.3

**Table 4** Confusion matrices of age-specific and age-agnostic classifiers

(a) Young-age model						
Exp	Anger	Disgust	Fear	Happy	Neutral	Sad
Anger	94.6	2.4	0.0	0.0	2.1	0.9
Disgust	1.5	97.0	0.0	0.0	0.0	1.5
Fear	0.0	0.0	97.0	0.3	0.3	2.4
Happy	0.0	0.0	0.0	100	0.0	0.0
Neutral	0.3	0.0	0.6	0.0	99.1	0.0
Sad	0.9	1.2	2.4	0.0	4.2	91.4
(b) Middle-age model						
Exp	Anger	Disgust	Fear	Happy	Neutral	Sad
Anger	94.0	4.8	0.0	0.0	0.3	0.9
Disgust	3.0	94.3	0.0	0.3	0.0	2.4
Fear	0.0	0.9	98.2	0.3	0.6	0.0
Happy	0.0	0.0	0.0	99.7	0.0	0.3
Neutral	0.0	0.0	0.9	0.0	97.9	1.2
Sad	3.6	2.4	2.1	0.3	7.4	84.2
(c) Old-age model						
Exp	Anger	Disgust	Fear	Happy	Neutral	Sad
Anger	88.1	8.0	0.0	0.3	1.5	2.1
Disgust	5.7	86.9	1.5	0.6	1.2	4.2
Fear	1.2	0.0	94.3	0.0	1.2	3.3
Happy	0.0	1.2	1.5	96.7	0.0	0.6
Neutral	0.0	0.3	0.3	0.6	96.7	2.1
Sad	8.3	8.0	5.4	0.9	9.2	68.2
(d) Agnostic-age model						
Exp	Anger	Disgust	Fear	Happy	Neutral	Sad
Anger	89.0	6.8	0.3	0.0	1.2	2.7
Disgust	5.4	90.5	1.2	0.3	0.9	1.8
Fear	0.9	0.0	95.2	0.3	0.9	2.7
Happy	0.0	0.6	2.1	96.7	0.0	0.6
Neutral	0.0	0.6	0.3	0.3	96.7	2.1
Sad	7.4	6.5	3.9	1.5	8.6	72.0

**Table 5** Accuracy of age group estimation (right) and confusion matrix (left)

Features	Manual Points	Automatic Points	
Shape	81.3±0.6	76.0±0.2	
Tex	86.1±0.6	82.6±0.2	
App	84.9±0.5	80.7±0.9	
S+T	86.6±0.1	84.0±1.1	
Group	Young	Middle	Old
Young	91.2	6.8	2.0
Middle	12.3	71.5	16.2
Old	1.1	9.6	89.3

**Table 6** Summary results of four schemas of combining age estimator with age-group-specific expression classifiers

Group	Points	Age specific Known Real Age	Age specific Eq. 9 Estimated Real Age	Weighted Eq. 10 Estimated Real Age	Weighted Eq. 10 Estimated Apparent Age
Young	Manual	97.7 ±0.4	96.9 ±0.7	98.6 ±0.1	98.8 ±0.2
	Automatic	96.5 ±0.4	96.7 ±0.2	98.0 ±0.2	98.3 ±0.1
Middle	Manual	97.6 ±0.4	96.1 ±0.2	98.8 ±0.3	98.5 ±0.4
	Automatic	94.7 ±1.1	94.7 ±0.6	95.3 ±0.4	96.0 ±0.2
Old	Manual	93.8 ±1.5	92.7 ±0.4	95.9 ±0.5	96.3 ±0.4
	Automatic	88.5 ±2.1	90.0 ±0.9	93.8 ±0.2	93.7 ±0.3
Mean	Manual	96.4 ±0.8	95.2 ±0.4	97.8 ±0.3	<b>97.9 ±0.3</b>
	Automatic	93.2 ±1.2	93.8 ±0.6	95.7 ±0.3	<b>97.0 ±0.4</b>

models are sensitive to the age of the individual, and that expression can be better estimated if we know the subjects' approximate age and can use an appropriate age-specific expression classifier. The results also suggest that it might be better to analyse the effect of apparent age on the performance of AFER.

*Apparent age* In Eqs. 10 and 11, we propose to weight and combine the probability of the age group estimator with the all age-group-specific expression classifiers in order to have full use of all the age and expressions representations. The idea is to fully utilize the representations not only from one expression classifier, such as the results calculated using Eq. 9, but from all the age and expression classifiers in a soft-level-based or apparent-age-based schema. The results are summarized in Table 6 (fifth column), in which the performance of the age group estimator and age-group-specific expression classifiers are combined and weighted, and the expression is estimated using all expression classifiers in proportion to their age weights from the age estimator.

These results demonstrate that the performance using a probabilistic approach (Eqs. 10 and 11), avoiding a hard decision (Eq. 9), is better (fifth column) than the performance with a hard decision (fourth column), which is actually better than that achieved when the true age is known (third column). We assume that this difference is because the accuracy of the expression recognition system depends on an individual's apparent age, rather than their actual chronological age.

Each image in the database is labelled manually with the apparent age group by five assessors. We also used the age group estimator to label each image by the apparent age. Those labels are used to redivide the database into three age groups by moving some people among the groups. We then re-trained the age group estimator and the age-specific facial expression classifiers using the new groups and the apparent age group label, as in Eqs. 10 and 11. We train the age group and expression classifiers several times (the training was performed once using the assessor labels and once using the classifier labels).

Table 6 (sixth column) shows the overall mean performance of the soft approach when we train the age group estimator and the expression classifiers on age groups defined by the apparent age of the individuals (as estimated by five assessors and age group estimator), rather than their real age.

In summary, the results in Table 6 demonstrate that the models are sensitive to the age of the individual, and that expression can be better estimated if we know the subject's approximate age and can use an appropriate age-specific classifier. These results also demonstrate that is better to use apparent age in selecting the most appropriate expression classifier than chronological age.

## 5.5 Computational complexity

Results in Table 7 show the computation complexity of the current work of combining the age and expressions classifiers when using both the estimated age expression classifier (hard level) and when using the weighted age and expression classifiers (soft-level) based on both the real and apparent age. The time was recorded in milliseconds (ms) and was compared to the time of the method in [29] which recorded the time in seconds (for a simpler comparison it is converted to ms). Experiments were performed on a Dell Intel(R) Core(TM) i5-2400 CPU 3.10 GHz and 16 GB memory. The results demonstrate that our method with 0.16 ms per image

**Table 7** Computational complexity (in millisecond) of the proposed method

Method	Features	Test time(ms)
Age-agnostic	157.5	0.06
Age-specific (known real age)	157.5	0.14
Age-specific (estimated real age)	157.5	0.08
Weighted (estimated real age)	157.5	0.15
Weighted (estimated apparent age)	157.5	0.16
Independent learn [29]	–	7.00
Joint learn [29]	–	10.00

without feature extraction computation has lower test time than the method in [29] with 10 ms performed on a machine with 2 Intel(R) Xeon(R) CPU X5570 2.93 GHz and 64 GB memory. The results demonstrated also that the computation time of the soft-level schema of using all the age and expressions classifiers is comparable to the computation time of hard-level schema of using one expression classifier select by the age estimator.

### 5.6 Comparison to the state of the art

Table 8 compares our system to the results presented in [20] and [42]. We show the results for manually placed points, as [20] and [42] rely on manual points.

These results show that the results of our models of within age and weighted combinations (joint age-Exp) are comparable to the results in [20] and [42]. These results show also that our features are less sensitive to ageing with 87.3% performance (across all ages) compared to the method in [20] with 64.0%

Table 8 compares our results to the recent results presented in [29] and [44] who used BP features and a

convolution neural network (CNN), respectively. The mean results of the three datasets of our automated system are better than the results of the methods presented in these earlier results. Our method may perform better than the method used in [29] because in our method the texture features are combined with shape features (compact representation), while in [29] only texture features were used. Moreover, the reason that our results are better than the DL results of [44] might be because we train the model using representative datasets of both age and expressions deformations, while in [44], the authors used the pre-trained model on the MORPH dataset with 55,134 images of age only (Table 9).

### 5.7 Example results

Figure 8 shows some examples of automatic results from our system including points localization, age group estimation, and expression recognition on three different age and expressions datasets. These results show that, although the system makes some mistakes in the age group estimation, expressions are recognized correctly due to the weighted combination schema of different expression classifiers. For

**Table 8** Performance Comparison of Various Methods on expression recognition on FACES, Lifspan and NEMO datasets

Data	Across age			All ages			Within age			Joint age-exp		
	[20]	[42]	Ours	[20]	[42]	Ours	[20]	[42]	Ours	[20]	[42]	Ours
FACES	64.0	–	87.3	88.8	–	93.4	97.9	95.0	96.4	97.1	95.7	97.9
Lifspan	69.3	–	90.1	86.3	–	91.2	82.6	96.0	97.4	92.0	96.5	95.4
NEMO	–	–	92.3	–	–	93.5	–	–	95.0	–	–	99.6
Mean	66.65	–	89.9	87.55	–	92.7	90.25	95.5	95.0	94.55	96.1	97.63
Error	33.35	–	10.1	12.45	–	7.3	9.75	4.5	95.0	5.45	3.9	2.3

**Table 9** Performance comparison of various methods on expression recognition on FACES, lifespan and NEMO datasets

Data	Emotion	Independent learn age-exp [29]	Joint learn of age-exp [29]	Joint estimation age-exp [44]	Age-group specific of known age[Ours]	Weighted combination age-exp [Ours]	Weighted combination apparent [Ours]
FACES	Neutral	91.2	95.9	–	97.9	97.3	98.5
	Happy	99.2	99.4	–	98.8	98.9	99.1
	Anger	84.8	88.3	–	92.2	96.2	95.3
	Disgust	89.4	92.9	–	92.7	97.1	95.4
	Fear	92.4	94.1	–	96.5	90.0	97.1
	Sadness	83.1	83.0	–	81.3	94.8	89.0
	Average	90.1	92.2	95.1	93.2	95.7	96.0
Lifspan	Neutral	97.4	96.2	–	98.7	99.3	99.2
	Happy	85.9	88.1	–	85.3	91.3	91.6
	Average	91.7	92.2	96.3	92	95.3	95.4
NEMO	Neutral	97.9	98.1	–	95.1	99.4	100
	Happy	97.5	97.9	–	94.8	98.8	99.1
	Average	97.7	98.0	–	95.0	99.1	99.6
Mean		93.2	94.1	95.7	93.4	96.7	<b>97.0</b>
Error		6.8	5.9	4.3	6.6	3.3	<b>3.0</b>

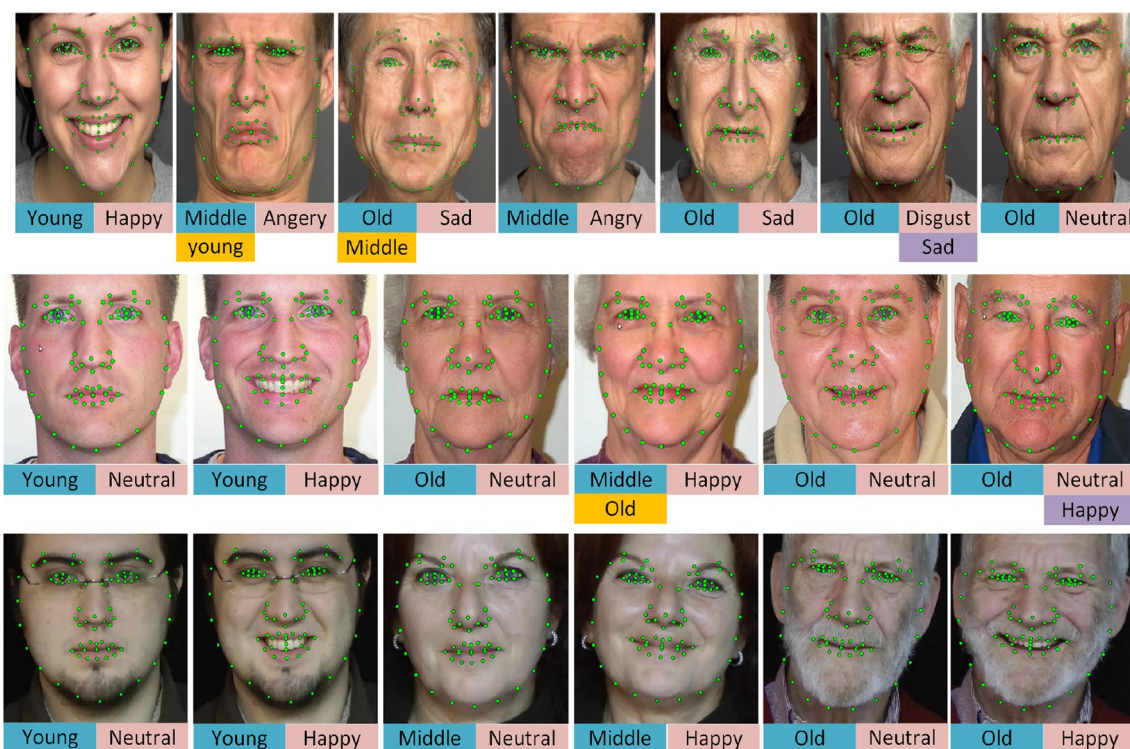
instance, in the second image in the first row (Fig. 8), the expression is correctly recognized as angry although the age group prediction is incorrect. Moreover, comparing our results to the those in [21], we found that the fifth image in the first row from the FACES data and fifth image in the second row from the lifespan data is classified correctly as sad and neutral expression, respectively, rather than misclassified as disgusted and happy, respectively, by [21].

## 6 Conclusion and future work

In this paper, first, we conducted extensive experiments to investigate the impact of ageing on the performance of automatic facial feature detection and expression recognition tasks. We have shown that the ageing has a significant effect on the accuracy of both tasks and the detector we used, based on RFRV-CLM, achieved satisfactory performance that can generalize well across a wide range of ages and expressions. It achieved a mean point-to-point error of less than 3.4% of the IOD (2.21 mm) on 99% of samples. That performance gain is due to applying the RFRV-CLM framework following the coarse-to-fine, multi-stage strategy. We also showed that using age-specific expression

classifiers gives better results than an age-agnostic classifier. Second, we have introduced a simple and effective system to recognize the expression, which is fully automatic, deals with individuals from a wide range of ages, and uses the age group information as prior knowledge to obtain better results of expression classification. We have shown that the system has a wide range of convergence and generalization across a large range of ages and expressions, and it is robust to outliers (due to the age features and apparent age). It identifies 97.0% of expressions using automatically found points, which is almost as effective as the system based on manual points (97.9%). Errors for the classification may be caused either by failure in locating landmarks accurately or by the failure of the classification algorithm. That performance gain is due to the idea of integrating the information over all ages through using weighted combination rules of a set of age group classifier and age-specific facial expression classifiers.

Finally, the results of our simple system (using a set of a sequential classifier on a limited range of age groups) are satisfactory and encouraging for more investigations about the problem. Therefore, for future works, we aim to revise the present work using joint optimization (inference /learning algorithms) for age segmentation and expression classification rather than using heuristic search for grouping the age



**Fig. 8** Example results of our system including point localization, age group estimation, and expression category recognition on three different age and expression datasets: FACES data (first row), LifeSpan data (second row), and NEMO data (third row). The predicated age

group and expression are shown on the blue and pink backgrounds, respectively. The ground truth is shown on the yellow and purple background for age and expression, respectively, if the system makes a mistake

of the subjects. We also plan to revise the present work using deep representation as it has shown remarkable performance in several computer vision and image processing researches to see its performance on the small size of the data per age per expression. Moreover, a study of the performance of the proposed approach on different factors with that effect on facial expressions such as the human race is an important aspect of future work.

## References

- Adolphs R, Tranel D (2004) Impaired judgments of sadness but not happiness following bilateral amygdala damage. *J Cogn Neurosci* 16(3):453–462
- Ahonen T, Hadid A, Pietikäinen M (2004) Face recognition with local binary patterns. *Comput Vision-ECCV 2004*:469–481
- Alnajar F, Lou Z, Álvarez JM, Gevers T, et al (2014) Expression-invariant age estimation. In: *BMVC*
- Antipov G, Baccouche M, Berrani SA, Dugelay JL (2016) Apparent age estimation from face images combining general and children-specialized deep learning models. In: *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 96–104
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Bromiley P, Adams J, Cootes T (2015) Localisation of vertebrae on dxa images using constrained local models with random forest regression voting. In: Yao J, Glocker B, Klinder T, Li S (eds) *Recent advances in computational methods and clinical applications for spine imaging*. Springer, Cham, pp 159–171
- Bromiley PA, Adams JE, Cootes TF (2015) Automatic localisation of vertebrae in dxa images using random forest regression voting. In: *international workshop on computational methods and clinical applications for spine imaging*, pp. 38–51. Springer
- Bromiley PA, Kariki EP, Adams JE, Cootes TF (2016) Fully automatic localisation of vertebrae in ct images using random forest regression voting. In: *international workshop on computational methods and clinical applications for spine imaging*, pp. 51–63. Springer
- Clapés A, Bilici O, Temirova D, Avots E, Anbarjafari G, Escalera S (2018) From apparent to real age: gender, age, ethnic, makeup, and expression bias analysis in real age estimation. In: *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 2373–2382
- Cohn JF, Kruez TS, Matthews I, Yang Y, Nguyen MH, Padilla MT, Zhou F, De la Torre F (2009) Detecting depression from facial actions and vocal prosody. In: *3rd international conference on affective computing and intelligent interaction and workshops, ACHI 2009*. pp. 1–7. IEEE
- Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. *IEEE Trans Pattern Anal Mach Intell* 23(6):681–685
- Cootes TF, Ionita MC, Lindner C, Sauer P (2012) Robust and accurate shape model fitting using random forest regression voting. In: *European conference on computer vision*, pp. 278–291. Springer
- Dibeklioglu H, Salah AA, Gevers T (2012) Are you really smiling at me? spontaneous versus posed enjoyment smiles. In: *European conference on computer vision*, pp. 525–538. Springer
- Dibeklioglu H, Salah AA, Gevers T (2015) Recognition of genuine smiles. *IEEE Trans Multimed* 17(3):279–294
- Dodgson NA (2004) Variation and extrema of human interpupillary distance. In: *stereoscopic displays and virtual reality systems XI*, vol. 5291, pp. 36–47. International society for optics and photonics
- Ebner NC, Johnson MK (2010) Age-group differences in interference from young and older emotional faces. *Cognit Emotion* 24(7):1095–1116
- Ebner NC, Riediger M, Lindenberger U (2010) Faces—a database of facial expressions in young, middle-aged, and older women and men: development and validation. *Beh Res Methods* 42(1):351–362
- Fölster M, Hess U, Werheid K (2014) Facial age affects emotional expression decoding. *Front Psychol* 5:30
- Goldman AI, Sripada CS (2005) Simulationist models of face-based emotion recognition. *Cognition* 94(3):193–213
- Guo G, Guo R, Li X (2013) Facial expression recognition influenced by human aging. *IEEE Trans Affect Comput* 4(3):291–298
- Hess U, Adams RB, Simard A, Stevenson MT, Kleck RE (2012) Smiling and sad wrinkles: age-related changes in the face and the perception of emotions and intentions. *J Exp Soc Psychol* 48(6):1377–1380
- Houstis O, Kiliaridis S (2009) Gender and age differences in facial expressions. *Eur J Orthodont* 31(5):459–466
- Huo Z, Yang X, Xing C, Zhou Y, Hou P, Lv J, Geng X (2016) Deep age distribution learning for apparent age estimation. In: *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 17–24
- Kanade T, Cohn JF, Tian Y (2000) Comprehensive database for facial expression analysis. In: *proceedings fourth IEEE international conference on automatic face and gesture recognition (Cat. No. PR00580)*, pp. 46–53. IEEE
- Lawrence AD, Calder AJ (2004) Homologizing human emotions
- Lindner C, Bromiley PA, Ionita MC, Cootes TF (2015) Robust and accurate shape model matching using random forest regression-voting. *IEEE Trans Pattern Anal Mach Intell* 37(9):1862–1874
- Lindner C, Cootes T (2015) Fully automatic cephalometric evaluation using random forest regression-voting. In: *proceedings of the IEEE international symposium on biomedical imaging (ISBI) 2015—grand challenges in dental X-ray image analysis—automated detection and analysis for diagnosis in cephalometric X-ray image*. Citeseer
- Lindner C, Thiagarajah S, Wilkinson J, Consortium T (2013) Fully automatic segmentation of the proximal femur using random forest regression voting. *IEEE Trans Med Imaging* 32(8):1462–1472
- Lou Z, Alnajar F, Alvarez JM, Hu N, Gevers T (2018) Expression-invariant age estimation using structured learning. *IEEE Trans Pattern Anal Mach Intell* 40(2):365–375
- Lyons M, Akamatsu S, Kamachi M, Gyoba J (1998) Coding facial expressions with gabor wavelets. In: *proceedings third IEEE international conference on automatic face and gesture recognition, 1998*, pp. 200–205. IEEE
- Martinez B, Valstar MF (2016) Advances, challenges, and opportunities in automatic facial expression recognition. In: Kawulok M, Celebi E, Smolka B (eds) *Advances in face detection and facial image analysis*. Springer, Berlin, pp 63–100
- Mary R, Jayakumar T (2016) A review on how human aging influences facial expression recognition (fer). In: Abraham A, Haqiq A, Muda AK, Gandhi N (eds) *Innovations in bio-inspired computing and applications*. Springer, Berlin, pp 313–322
- Minear M, Park DC (2004) A lifespan database of adult facial stimuli. *Beh Res Methods Instruments Comput* 36(4):630–633
- Nguyen DT, Cho SR, Shin KY, Bang JW, Park KR (2014) Comparative study of human age estimation with or without preclassification of gender and facial expression. *Sci World J*. <https://doi.org/10.1155/2014/905269>
- Phillips LH, Allen R (2004) Adult aging and the perceived intensity of emotions in faces and stories. *Aging Clin Exp Res* 16(3):190–199

36. Prkachin KM, Solomon PE (2008) The structure, reliability and validity of pain expression: evidence from patients with shoulder pain. *Pain* 139(2):267–274
37. Rothe R, Timofte R, Van Gool L (2015) Dex: deep expectation of apparent age from a single image. In: proceedings of the IEEE international conference on computer vision workshops, pp. 10–15
38. Sariyanidi E, Gunes H, Cavallaro A (2015) Automatic analysis of facial affect: a survey of registration, representation, and recognition. *IEEE Trans Pattern Anal Mach Intell* 37(6):1113–1133
39. Sawant MM, Bhurchandi KM (2019) Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging. *Artif Intell Rev* 52(2):981–1008
40. Uříčář M, Timofte R, Rothe R, Matas J, et al (2016) Structured output svm prediction of apparent age, gender and smile from deep features. In: proceedings of the 29th IEEE conference on computer vision and pattern recognition workshop (CVPRW 2016), pp. 730–738. IEEE
41. Vural E, Cetin M, Ercil A, Littlewort G, Bartlett M, Movellan J (2007) Drowsy driver detection through facial movement analysis. In: international workshop on human-computer interaction, pp. 6–18. Springer
42. Wang S, Wu S, Gao Z, Ji Q (2016) Facial expression recognition through modeling age-related spatial patterns. *Multimed Tools Appl* 75(7):3937–3954
43. Whitehill J, Bartlett M, Movellan J (2008) Automatic facial expression recognition for intelligent tutoring systems. In: IEEE computer society conference on computer vision and pattern recognition workshops, 2008. CVPRW'08, pp. 1–6. IEEE
44. Yang HF, Lin BY, Chang KY, Chen CS (2018) Joint estimation of age and expression by combining scattering and convolutional networks. *ACM Trans Multimed Comput Commun Appl (TOMM)* 14(1):1–18
45. Zhu Y, Li Y, Mu G, Guo G (2015) A study on apparent age estimation. In: proceedings of the IEEE international conference on computer vision workshops, pp. 25–31

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

## Authors and Affiliations

Nora Al-Garaawi<sup>1</sup>  · Tim Morris<sup>2</sup> · Timothy F. Cootes<sup>3</sup>

Tim Morris  
tim.morris@manchester.ac.uk

Timothy F. Cootes  
timothy.f.cootes@manchester.ac.uk

<sup>1</sup> Department of Computer Science, Faculty of Education for Girls, University of Kufa, Najaf, Iraq

<sup>2</sup> Department of Computer Science, The University of Manchester, Manchester, United Kingdom

<sup>3</sup> Division of Informatics, Imaging and Data Science, The University of Manchester, Manchester, United Kingdom