Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2019) 27: 2430 – 2443

doi:10.3906/elk-1811-70

A computational study on aging effect for facial expression recognition

Elena BATTINI SÖNMEZ*
Department of Computer Engineering, Faculty of Engineering and Natural Sciences, İstanbul Bilgi University, İstanbul, Turkey

Received: 12.11.2018 • **Accepted/Published Online:** 05.05.2019 • **Final Version:** 26.07.2019

Abstract: This work uses newly introduced variations of the sparse representation-based classifier (SRC) to challenge the issue of automatic facial expression recognition (FER) with faces belonging to a wide span of ages. Since facial expression is one of the most powerful and immediate ways to disclose individuals' emotions and intentions, the study of emotional traits is an active research topic both in psychology and in engineering fields. To date, automatic FER systems work well with frontal and clean faces, but disturbance factors can dramatically decrease their performance. Aging is a critical disruption element, which is present in any real-world situation and which can finally be considered thanks to the recent introduction of new databases storing expressions over a lifespan. This study addresses the FER with aging challenge using sparse coding (SC) that represents the input signal as the linear combination of the columns of a dictionary. Dictionary learning (DL) is a subfield of SC that aims to learn from the training samples the best space capable of representing the query image. Focusing on one of the main challenges of SC, this work compares the performance of recently introduced DL algorithms. We run both a mixed-age experiment, where all faces are mixed, and a within-age experiment, where faces of young, middle-aged, and old actors are processed independently. We first work with the entire face and then we improve our initial performance using only discriminative patches of the face. Experimental results provide a fair comparison between the two recently developed DL techniques. Finally, the same algorithms are also tested on a database of expressive faces without the aging disturbance element, so as to evaluate DL algorithms' performance strictly on FER.

Key words: Aging, dictionary learning, facial expression recognition, sparse representation-based classifier

1. Introduction

Humans are social creatures who communicate via body language. Since faces are the most expressive part of the body, facial expressions are widely used in psychological studies for research in several areas including perception, human development, social reasoning, and neuroscience. Previous studies in psychology suggested that faces are special types of visual stimuli, which are processed in a separate part of the brain [1], and that, whereas recognition of complex scenes is not affected by age [2], when compared with young individuals, older adults are impaired in the acknowledgment of unfamiliar faces [3]. In the parallel field of computer engineering, automatic facial expression recognition (FER) is an open and interesting issue with a wide range of applications, including human behavior understanding, human–computer interaction, marketing, and synthetic face animation.

To date, most of the studies done on automatic FER used pictures belonging to a small interval of ages.

*Correspondence: elena.sonmez@bilgi.edu.tr

This is mainly due to the type of stimulus presents in the popular FER databases, e.g., the Cohn Kanade (CK) [4], the Extended Cohn Kanade (CK+) [5], the Japanese Female Facial Expression (Jaffe) [6], and the MMI [7] datasets, which store faces of middle-aged women and men. On the contrary, the Facial Expressions in the Wild [8] dataset has a wide lifespan, but the age of the actor present in an image is not given, since the database was built for a different challenge.

In 2004, psychologists Minear and Park [9] realized that the lack of available datasets with expressive faces crossing all lifespan was limiting the generality of the research, and they introduced the CAL/PAL face database, which stores faces of 218 people divided into four groups with age ranges of 18–29, 30–49, 50–69, and 70-and-over years. Every subject acts one or more neutral and happy emotions for a total of 844 stimuli.

In 2008, the psychologists Ebner [10] showed that age matters in both human perception and expression of a stimulus. Ebner asked an equal number of young and old participants to rate a selection of faces from the CAL/PAL database [9], with regards to several dimensions such as attractiveness, likability, distinctiveness, goal orientation, energy, mood, and age. Analysis of the results showed that old faces were less attractive, less likable, and less distinctive than young ones. Considering the differences in judgment, as a function of the age of the participants, old individuals evaluated the faces in a more positive way. Overall, the results of this study stressed the necessity of using more appropriate stimulus material; that is, it underlined the need for face databases covering the whole lifespan.

In 2010, Ebner et al. [11] introduced the FACES database, which stores expressive faces made by 171 actors in the age range from 19 to 80 years old. The database is divided into three age groups: young, middle, and old. Each actor in FACES made the five basic facial expressions of happiness, sadness, disgust, fear, and anger, plus the neutral pose. Having validated the dataset, the authors of [11] ran several experiments; this work challenged the 'expression identification as a function of the age of the face' experiment of [11].

In the parallel and complementary field of engineering, an automatic FER system must be able to detect the correct expression of every person, regardless of the age of the actor. From a computation point of view, mixing emotional faces belonging to different ages increases the complexity of the problem since older people show expressions in a subtle way [12], and the presence of wrinkles in older faces can mislead the classifier.

In 2013, Guo et al. investigated the influence of aging in automatic FER [13]. They worked with both the CAL/PAL and the FACES databases; to compare their results across the two datasets, they imposed the four age-groups division of CAL/PAL to FACES. In all experiments, they used manually labeled facial landmarks (FLs).

In 2016, Algaraawi and Morris questioned how the changes in shape and texture information affect automatic facial expression recognition [14]. They worked with preannotated training images of the FACES database where key landmark points were manually labeled.

While manually labeled FLs are more precise, a fully automatic system requires an automatically detected FL, which is a fundamental preprocessing step that allows performing face alignment and addressing more complex issues such as subject identification, expression recognition, gender, and age estimation. In 2018, Johnston and Chazal published a review of automatic FL detectors [15]. Experiments run in this study used automatically detected FLs.

This paper challenges the FACES database using sparse coding (SC) as it is a very effective way to convert signals into a high-level representation of the data. Since SC exploits the property of signals to be sparse in an appropriate dictionary, one of its main challenges is dictionary learning (DL), i.e. the search for the best

dictionary to represent the given data. A survey on DL algorithms for face recognition was given by Xu et al. [16].

The aim of this work is twofold. First, it aims to present the first computational study of the FACES database with automatically annotated FLs, and then it investigates the performance of two recently proposed DL algorithms.

The obtained accuracy is compared against the performance of humans on FACES, documented in [11]; the analysis of similarity against human performance could be used in the engineering field to suggest future development of more robust algorithms. Comparison with the artificial systems proposed by Guo et al. [13] and Algaraawi and Morris [14] is possible but not logically correct because both papers used manually labeled FL; however, like them, we run mixed-age and within-age experiments using 10-fold cross-validation.

To summarize, the main contributions of this work are (1) to compare the performance of several sparse representation-based algorithms with the age disturbance element, (2) to propose a benchmark for the FER challenge on FACES with automatically detected FLs, (3) to spread FACES in the engineering community, and (4) to bridge the gap between researchers in psychology and the engineering field.

The rest of the paper is organized as follows: Section 2 presents the FACES database, Section 3 gives an overview of the used classifiers, Section 4 details the experimental setup, Section 5 presents the results obtained on FACES and compares them against the two benchmark papers, and Section 6 reports the performance of SC algorithms strictly on FER. Finally, conclusions are drawn in Section 7.

2. The FACES database

The FACES database was introduced in 2010 by Ebner et al. [11] with the aim of varying both expressions and ages of the faces. It stores expressive faces of 171 actors divided into three age groups: 58 young models, with an age range of 19–31; 56 middle-aged people, with an age range of 39–55 years; and 57 older actors, with an age range of 69–80 years. All models are Caucasian average-type subjects. Each actor of FACES makes one neutral face and the five basic facial expressions of happiness, sadness, disgust, fear, and anger; the surprise basic expression is missing. The FACES database is a good opportunity to investigate traditional computer vision and machine learning issues, as recommended in [17].

At recording time, each model was asked to take off jewelry and glasses, and to remove any makeup. After that, every picture was rated by two pretrained evaluators, who considered the position of the muscles in the faces, i.e. the presence of action units, to assign emotional labels and intensity values on a 3-point rating scale.

At picture selection time, for each emotion and from each model, Ebner et al. chose the two best acted expressions, in which both referees agreed on the type of facial expression; pronounced expressions were preferred to subtle ones.

Having 171 subjects, with 2 images per person and 6 expressions, the database stores a total of $171 \times 2 \times 6 = 2.052$ faces.

Figure 1 shows a middle-aged woman acting the five emotions plus the neutral face.

Figure 2 displays six actors of FACES who gave consent for publication; all subjects are displaying the happy emotion. Figure 2 also shows the big somatic difference among the models; other disturbance elements of the database are illumination, skin color, and wrinkles. Paying attention to the age disturbance element, since older faces have fewer muscles, the amount of action units in the expressive face is lower. As a result, working on FER with older faces is particularly difficult for three main reasons: (1) older subjects express emotion in a

Figure 1. From left to right: (top row) angry, disgusted, and fearful faces; (bottom row) happy, neutral and sad expressions.

subtle way [12], (2) the expressive face does not have all expected action units, and (3) the feature extraction stage can confuse wrinkles with action units.

Figure 2. From left to right: the 1st row shows happy faces of young woman, young man, middle-aged woman; the 2nd row displays the happy expressions of middle-aged man, old woman, and old man.

The first objective of Ebner et al. [11] was to validate their database by enrolling young, middle-aged, and older humans to rate every image in terms of perceived age and expression. All participants were asked 'Which facial expression does this person primarily show?' and 'How old is this person?' Faces were presented in a random order, and each participant rated only one set of the database, i.e. either set A or set B, not both. Each picture was rated by an average of 11 referees per age group. Results reported in [11] showed that expression identification is more difficult with older faces, the task becomes easier with middle-aged stimuli, and the best performance is reached with young faces. Furthermore, happiness was easier to detect than all other expressions, whereas sadness and disgust were the problematic ones.

3. Classification

This work uses the sparse coding (SC) method as a classifier; a comparison to the performance of other classifiers is part of our future work.

The past years have witness an increasing interest in SC algorithms due to their successful applications in several fields such as image denoising [18, 19], clustering [20], deblurring [21], and face classification [22–28]. SC is attractive since it attempts to represent the input signal in a high-level fashion, i.e. as a linear combination of some columns of the dictionary. As a consequence, DL, which is the design of ad hoc dictionaries capable of extrapolating the inner essence of the data, is one of the main challenges of SC.

In more detail, this paper compares the robustness and running time of the sparse representation-based classifier (SRC) [29], Fisher discrimination dictionary learning (FDDL) [30], and interclass sparsity-based discriminative least square regression (ICS_DLSR) [28] for the aging disturbance element.

3.1. Notation

This subsection briefly introduces basic notations that will be used throughout the paper. Let $x \in \mathbb{R}^N$ be the observed signal, and $X = D = [x_1, x_2, \ldots, x_T] \in \mathbb{R}^{N \times T}$ is the training matrix, called a dictionary, made up of T training samples belonging to C classes; that is, matrix D is a column matrix that aligns vectorized training samples; it can also be written as $D = [D_1, D_2, \ldots, D_C]$, where D_i is a submatrix aligning only samples of the ith class i, $\forall i = 1, ..., C$. Let $S = [s_1, s_2, \ldots, s_T] \in \mathbb{R}^{T \times T}$ be the corresponding matrix of sparse coefficients and X_i and S_i be, respectively, the submatrices of X and S restricted to the columns of the ith class, $\forall i = 1, ..., C$.

Let $l_i \in \mathbb{R}^C$ be the label of training sample x_i, having all zero values except for its ith element; the corresponding zero-one label matrix is $L = [l_1, l_2, \ldots, l_T] \in \mathbb{R}^{C \times T}$.

For a general vector $z \in \mathbb{R}^M$, z_i is a general element of z; the l_1 norm is calculated as $||z||_1 = \sum_1^M |z_i|$, and the l_2 norm is calculated as $||z||_2 = \sqrt{\sum_1^M z_i^2}$. For a general matrix $Z \in \mathbb{R}^{M \times T}$, $Z_{i,j}$ is a general element; Z^{-1} is the inverse matrix and Z^T is its transpose matrix. The l_1, $l_{2,1}$, and l_F, i.e. Frobenius norms, are calculated as $||Z||_1 = \sum_{i=1}^M \sum_{j=1}^T |z_{i,j}|$, $||Z||_{2,1} = \sum_{i=1}^M \sqrt{\sum_{j=1}^T z_{i,j}^2}$, and $||Z||_F^2 = \sum_{i=1}^M \sum_{j=1}^T z_{i,j}^2$, respectively.

Finally, α, β, and γ are regularization parameters, i.e. small positive values, used to balance the effects of terms of the cost function.

3.2. Sparse representation-based classifier (SRC)

SRC was first introduced in 2009 by Wright et al. [29]. Theoretically, SRC is based on compressive sensing theory [31–33].

Practically, SRC represents the input test image, x, as a linear combination of the training set, and, among all solutions, it considers the sparsest one, s. That is, every training sample is vectorized and aligned to form a column matrix, called a dictionary, D; training samples belonging to the same class are near each other; and, having C emotions, the dictionary is logically divided into C classes. Figure 3 gives the graphical representation of the SRC.

In the original SRC algorithm, classification requires the solution of the following equation:

$$argmin_s ||x - Ds||_2 + \gamma ||s||_1, \tag{1}$$

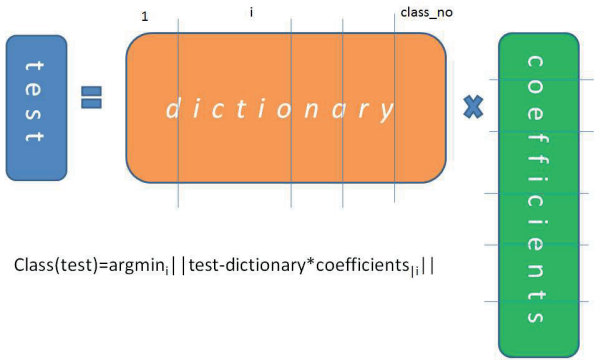

Figure 3. Graphical representation of SRC.

where $x \in \mathbb{R}^N$ is the vectorized test sample, $D \in \mathbb{R}^{N \times T}$ is the dictionary, and $s \in \mathbb{R}^T$ is the coefficient vector, the sparse representation of x calculated using l_1 minimization.

Having the sparse coefficient s, the SRC algorithm uses Eq. (2) to calculate the distance between every class and the test sample x:

$$residual_{class_i} ||x - Ds_i||_2, \tag{2}$$

where s_i is the coefficient vector restricted for $class_i$; that is, it has all zero entries except for the coefficients of $class_i$. In other words, the nonzero elements of the sparsest solution, s, select the corresponding column and allow for the calculation of the classes' error. Previous studies on SRC were presented in [34, 35].

Finally, the test sample is assigned to the class having minimum residual.

3.3. Fisher discrimination dictionary learning (FDDL)

FDDL was introduced in 2014 by Yang et al. [30]. It is based on the hypothesis that classification performance can be improved with the construction of a Fisher dictionary capable of increasing the interclass variation while decreasing the intraclass scatter. The cost function of FDDL is:

$$\begin{aligned} argmin_{D,S} \sum_{i=1}^{C} \left[||X_i - DS_i||_F^2 + ||X_i - D_i S_i^i||_F^2 + \sum_{j=1, j \neq i}^{C} ||D_j S_i^j|| \right] \\ + \alpha ||S||_1 + \beta \Big[tr\big(SC_W(S) - SC_B(S)\big) + \gamma ||S||_F^2 \Big], \end{aligned} \tag{3}$$

where S_i^j is the representation coefficient of X_i over D_j, tr() represents the trace of the matrix, and $SC_W(S)$ and $SC_B(S)$ are, respectively, the within-class and between-class scatter matrices. The intuition of Eq. (3) is as follows:

- The term $||X_i - DS_i||_F^2$ requires a dictionary D capable of representing class X_i, $\forall i = 1, ..., C$ well.
- Since D_i is the submatrix of D associated to the ith class, it is expected to represent S_i well; this constraint is express by the term $||X_i - D_i S_i^i||_F^2$.
- Following the same logic, atoms of the submatrix D_j must not be able to represent samples of class i; that is, S_i^j should have mostly zero coefficients, as expressed by the term $\sum_{j=1, j \neq i}^{C} ||D_j S_i^j||$.

- The Fisher criterion is used to make dictionary D more discriminative; that is, the addend $tr\big(SC_W(S) - SC_B(S)\big)$ is used to increase the within scatter and to reduce the between scatter matrices, and the addition of the term $\gamma||S||_F^2$ is necessary to archive convexity.

Overall, since Eq. (3) is not convex, the objective function is iteratively implemented in two steps: fixing D to update S, and fixing S to update D.

As formalized in Eqs. (4) and (5), classification is performed by mapping the test sample into a high dimension, with sparse representation, and assigning it to the nearby class, i.e. the class producing the minimum residual:

$$argmin_{s_i}||x_t - D_i s_i||_2^2 + \gamma||s_i||_1, \tag{4}$$

$$residual_{class_i}||x_t - D_i s_i||_2. \tag{5}$$

3.4. Interclass sparsity-based discriminative least square regression (ICS_DLSR)

The ICS_DLSR algorithm was introduced in 2018 by Wen et al. [28]. ICS_DLSR improves the standard least square regression (StLR) technique by adding to the cost function an error matrix, E, and a discriminative addend with the aim of stepping away from the zero-one label matrix and imposing a similar sparsity structure to samples of the same class.

In more detail, Eqs. (6) and (8) show the cost functions of StLR and ICS_DLSR. Focusing on StLR first:

$$argmin_Q||L - QX||_F^2 + \alpha||Q||_F^2, \tag{6}$$

where $X \in \mathbb{R}^{N\times T}$ is the training matrix, L is the label matrix, and $Q \in \mathbb{R}^{C\times N}$ is the transformation matrix. Eq. (6) can be easily solved as $Q = LX^T(XX^T + \gamma I)^{-1}$, and the predicted label of sample $x \in \mathbb{R}^N$ is given by the position of the maximum coefficient of vector Qx. In formula form:

$$argmax_i(Qx)_i, \forall i = 1, ..., C, \tag{7}$$

where $(Qx)_i$ is the ith element of vector Qx.

Eq. (8) shows the cost function of ICS_DLSR:

$$argmin_{Q,E}||L + E - QX||_F^2 + \alpha||Q||_F^2 + \beta\sum_{i=1}^{C}||QX_i||_{2,1} + \gamma||E||_{2,1}, \tag{8}$$

where $E \in \mathbb{R}^{C\times T}$ is the error matrix added to have a target matrix, $L + E$, appropriate for classification. The third addend of the cost function, $\sum_{i=1}^{C}||QX_i||_{2,1}$, imposes the same sparsity structure to all samples belonging to the same class.

Finally, the k-nearest neighbor algorithm is used for classification.

The following section presents the experimental setup and Section 5 compares the performance of the three algorithms.

4. Experimental setup on FACES

We run two different experiments for mixed-age and within-age groups. In the first experiment, all faces belonging to one emotion and acted by young (Y), middle-aged (M), and old (O) actors are assigned to the same class, whereas in the second experiment we worked on FER separately within each age group.

In both cases, we worked with the original stimuli of group A of the FACES database for a total of 1.026; considering the neutral face with the six emotions, we have a six-class problem. We used the 10-fold cross-validation technique, and the given performance is the average over 10 trials; every run employs methodically separated train and test samples. To make a fair comparison among the three algorithms, every trial uses the same (train, test) division of samples. The author is willing to share the used code upon request.

During the preprocessing step, Intraface free software [36] was used to automatically locate FLs, which were necessary to align the original images and to extract blocks out of them.

In the mixed-age experiment, we worked with the entire face as well as with some parts of it; that is, while the first variation of the experiment uses the raw pixels of aligned and cut faces, the second one uses only the pixels of the blocks of the two eyes and the mouth. Those blocks have been claimed to be the most discriminative patches of the faces and therefore the most successful for FER [37, 38]. Figure 4 shows an original image, the cut face, and the two blocks of the eyes and the mouth.

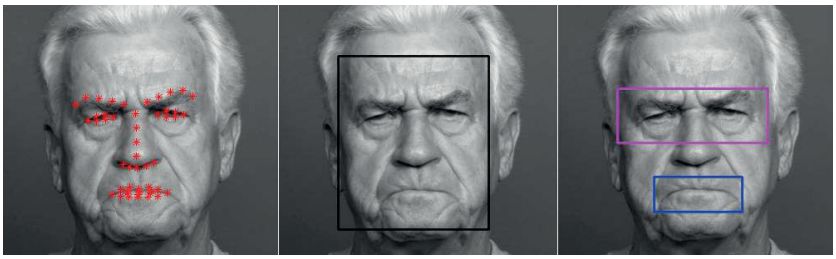

Figure 4. (Left) Original face with face landmarks; (center) the cut face of size 128 × 140 = 17920 pixels; (right) the block of the two eyes, and the one of the mouth; size (2 eyes) + size (mouth) = 56 × 136 + 40 × 80 = 7616 + 3200 = 10816 pixels.

5. Results and comparison with previous works

Table 1 shows the average performance and running time of the three classifiers in the mixed-age experiment: the 1st row stores results obtained when working with the cut face, while the 2nd row highlights the benefits of working in a block-based fashion. Results of Table 1 attest to the superior performance of the ICS_DLSR algorithm, considering both performance and running time; they also confirm our hypothesis that FER must be tackled in a block-based fashion. That is, the use of discriminative blocks of the face has the double advantage of increasing the performance as well as decreasing the running time of all algorithms.

All trials were run with an Intel CORE i7-3630QM CPU @ 2.40 GHz 2.40 GHz computer.

Table 1. Mixed-age experiment: average performance and running time of the three classifiers.

Perf (%)/time (min per trial)	SRC	FDDL	ICS_DLSR
Face	86/5	82/165	87/23
2 eyes + mouth	89/3	87/100	92/5

Table 2 gives the confusion matrix of ICS_DLSR for the mixed-age experiment with blocks; row labels, the true class and column the predicted one. Emotions are as follows: anger (A), disgust (D), fear (F), happiness (H), sadness (S), and neutral (N).

Table 2. Confusion matrix of ICS_DLSR using the two blocks of eyes and mouth (rows for ground truth, columns for predicted expression).

Perf (%)	A	D	F	H	S	N
A	92	3	0	0	2	3
D	4	90	0	0	6	0
F	0	0	98	0	0	2
H	0	0	0	99	0,5	0,5
S	5	5	0	0	84	6
N	2	0	2	0	5	91

Analyzing the results of Table 2, it is interesting to note that the happy face is the most successful one, with a recognition rate of 99%; also, fearful faces are easily acknowledged, reaching 98% . Finally, the recognition rates of angry, disgusted, sad, and neutral) expressions are respectively 92%, 90%, 84%, and 91%.

In the psychological field, Ebner et al. [11] ran a mixed-age emotion recognition experiment. Their performance results are reported in the 1st row of Table 3, while the 2nd row summarizes the results of the ICS_DLSR algorithm. A comparison between the accuracy of ICS_DLSR with that of [11] highlights that the artificial system has a higher recognition rate for all expressions.

Table 3. Mixed-age experiment: comparison of recognition rates.

Perf (%)	A	D	F	H	S	N	AvScore
Ebner et al. [11]	81	68	81	96	73	87	81
ICS_DLSR [28]	92	90	98	99	84	91	92

Finally, the performance of the ICS_DLSR algorithm in the block-based within-age experiment is (85, 89, 96)% when using blocks out of (old, middle-aged, young) actors, corresponding to an overall accuracy of 90%.

Since the two computational models previously built on FACES in [13] and [14] utilized manually labeled FLs, a fair comparison with them is not possible. However, all computational studies run the mixed-age and within-age experiments and they use the 10-fold cross-validation technique.

The author of [13] applied a three-scale and six-orientation Gabor filter bank to the FLs; classification was done with a support vector machine. When using the original stimuli, their mixed-age group performance was 64%, whereas the within-age group experiment had a hit rate of 98%. Analyzing those results, it is interesting to note that their artificial system is very sensitive to the mixture of faces with different ages. That is, the computational model proposed in [13] works quite well in the within-age experiment, where both train and test stimuli have similar patterns, but it fails in the mixed-age experiment, where faces of (young, middle-aged, old) actors are assigned to the same class. On the contrary, the sparse representation (SR)-based artificial system is less affected by the age disturbance element; that is, in the mixed-age experiment ICS_DLSR reaches the best

performance of 92%, which is 11 percentage points higher than [11] and 28 percentage points better than [13]. Nevertheless, we believe that there is still room for improvement, and future work aims to couple the SR-based classifier with features that can capture the essence of subtle expressions.

The author of [14] combined active shape model, active appearance model, and local binary pattern features with the support vector machine. The results proved that aging strongly affects the appearance of expressions since the presence of wrinkles together with the reduction of facial muscles changes the structure of the face, and young people show more exaggerated expressions than older individuals. Classification results are very high in both the mixed-age and within-age experiments. It would be interesting to check the same algorithms with automatically located FLs and this is part of our future work.

6. Performance of sparse classifiers on FER

This section compares the performance of the sparse classifiers strictly on FER, without the aging disturbance element. That is, the three sparse classifiers are used to challenge the leave-one-subject-out (LOSO) experiment of the Extended Cohn Kanade (CK+) dataset [5].

CK+ is a collection of videos with expressive faces acted by 123 middle-aged subjects; each video starts with a neutral face and ends with a peak expression. Faces in the CK+ database mimic the 6 basic emotions plus contempt; in more detail, the acronyms and frequency of every expression are as follows: anger (A, 45), contempt (C, 18), disgust (D, 59), fear (F, 25), happiness (H, 69), sadness (Sa, 28), and surprise (Su, 83). The CK+ database provides FLs associated with every picture. Having 123 subjects, the LOSO experiment results in 123 trials; at every run all expressive faces of one subject are used for testing and all remaining pictures for training; it is an 'open system' experimental setup where images of the same subject are not in the training and testing groups at the same time.

Figure 5 shows a face of CK+ acting the surprise emotion; the red dots are the given facial landmarks. The face has been already aligned, and FLs will be used again to extract the discriminative blocks of eyes and mouth, as shown in Figure 6.

Figure 5. Original face of CK+ acting the surprise emotion; the red dots are the facial landmarks.

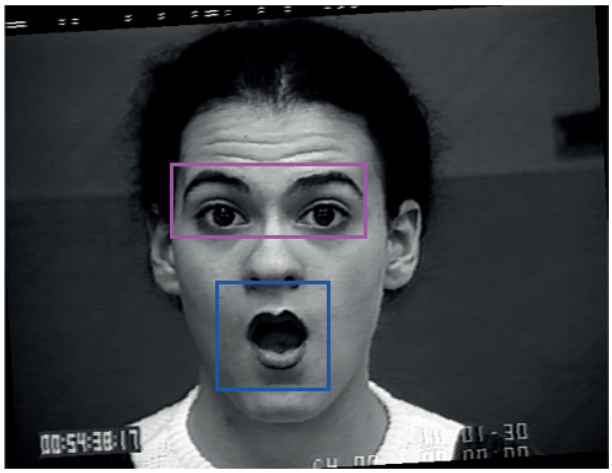

Figure 6. The block of the eyes of size 40 × 90 = 3600 pixels, and the block of the mouth of size 40 × 75 = 3000 pixels; the raw pixels of those blocks are the input to all classifiers.

Table 4 shows the average performance and running time of the three classifiers in the LOSO experiment; as in Section 4, the inputs to all experiments are raw pixels belonging to the blocks of the two eyes and the mouth.

Table 4. LOSO experiment: average performance and running time of the three classifiers.

Perf (%)/time (min per trial)	SRC	FDDL	ICS_DLSR
2 eyes + mouth	90/0.01	89/4	91/1

Results of Table 4 confirm the superior performance of the ICS_DLSR algorithm, considering both accuracy and running time.

In 2017, Lopes et al. [39] introduced a new technique for data augmentation based on a combination of translations, rotations, and skewing for increasing the dataset; for each original image of CK+, 70 additional synthetic pictures were generated. The authors of [39] used the FLs given with the CK+ database to determine the location of the center of the eyes of every image; they ran a variation of the LOSO experiment, and they did not consider the contempt expression. For all those reasons, performance comparison is possible but not totally correct.

In 2018, Yang et al. [40] tested a new geometric-based FER system on the LOSO experiment of CK+. The authors identified geometric features in two shape-deformation settings, one characterizing a face's shape changes in a neutral-to-peak sequence (NP method) and the other quantifying the within-category and between-category differences in facial shape (PE method). Best results were obtained using features extracted from both NP and PE methods. Since both methods require very accurate FLs, Yang et al. manually traced them on each image of every sequence.

Table 5 compares the performance of ICS_DLSR against the best accuracy obtained by Lopes et al. [39] and Yang et al. [40].

Table 5. LOSO experiment: comparison of recognition rates.

Perf (%)	A	C	D	F	H	Sa	Su	AvScore
ICS_DLSR [28]	78	89	91	76	97	79	98	90
Lopes et al. [39]	79	–	94	73	99	73	95	90
Yang et al. [40]	93	72	98	92	100	93	98	95

Analyzing the results of Table 5, it is possible to infer that ICS_DLSR performs better than CNN with augmented data proposed by [39], especially considering that Lopes et al. did not work with the contempt expression and ran a 6-class problem instead of the 7-class one. On the other hand, the accuracy reached by the geometric-based FER [40] is higher than ICS_DLSR but the use of manually labeled FLs done by Yang et al. make this comparison not logically correct.

Finally, looking at the recognition rate of every single expression, it is interesting to note that both methods, from [40] and [28], reached the top accuracy with happy and surprised faces; it is also interesting to note that the geometric-based FER system performs very well with all expressions except contempt.

7. Conclusions and future work

An automatic FER must be able to detect the correct expression of every person, regardless of the age of the actors. Previous psychological studies showed that elderly people express emotions with less intensity, in a subtle way; moreover, from a computational point of view, the presence of wrinkles in older faces may mislead the feature extraction stage, and the reduction of muscles decreases the intensity of action units. As a result, aging is a challenging disruption element in the construction of an automatic FER system. This study creates a fully automatic FER system with automatically located FLs and it compares both accuracy and running time of newly released DL algorithms. Using the FACES database it attests to the superior performance of ICS_DLSR, being robust against the age disturbance element. When working with the most discriminative blocks of the faces, the hit-rate of the proposed system is 92% in the mixed-age experiment, and (85, 89, 96)%, respectively, for (old, middle-aged, young) actors in the within-age experiment. Moreover, our experimental results confirmed the hypothesis that FER must be tackled in a block-based fashion, and, finally, the within-age experiment showed that the main problem of the proposed computational model is with the faces of older actors. A second set of experiments on the CK+ database compared the performance of DL algorithms strictly on FER, without the aging disturbance element. Future work includes: (1) making ICS_DLSR more sensitive to subtle expressions and increasing its capability to distinguish between wrinkles and action units; (2) comparing the performance of DL algorithms against deep learning methods; (3) conducting a deeper study on psychology and behavioral science so as to be able to discuss the effect of shape, texture, and appearance-based methods; (4) cross-validating happy and neutral subjects from the CAL/PAL and FACES datasets with the aim of double-checking the robustness of DL algorithms on FER with age; and (5) repeating the experiments of [14] using automatically located FLs.

References

[1] Puce A, Allison T, Gore JC, McCarthy G. Face-sensitive regions in human extrastriate cortex studied by functional MRI. Journal of Neurophysiology 1995; 74 (3): 1192-1199.

[2] Rybarczyk BD, Hart SP, Harkins SW. Age and forgetting rate with pictorial stimuli. Psychology and Aging 1987; 2 (4): 404-406.

[3] Grady CL, McIntosh AR, Horwitz B, Maisog JM, Ungerleider LG et al. Age-related reductions in human recognition memory due to impaired encoding. Science 1995; 269: 218-221.

[4] Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z et al. Comprehensive database for facial expression analysis. In: Fourth IEEE International Conference on Automatic Face and Gesture Recognition; Grenoble, France; 2000. pp. 46-53.

[5] Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z et al. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops; San Francisco, CA, USA; 2010. pp. 94-101.

[6] Lyons M, Akamatsu S, Kamachi M, Gyoba J. Coding facial expressions with Gabor wavelets. In: Third IEEE International Conference on Automatic Face and Gesture Recognition; Nara, Japan; 2018. pp. 200-205.

[7] Pantic M, Valstar M, Rademaker R, Maat L. Web-based database for facial expression analysis. In: IEEE International Conference on Multimedia and Expo; Amsterdam, the Netherlands; 2005. p. 5.

[8] Dhall A, Goecke R, Joshi J, Hoey J, Gedeon T. EmotiW 2016: Video and group-level emotion recognition challenges. In: 18th ACM International Conference on Multimodal Interaction; Tokyo, Japan; 2016. pp. 427-432.

[9] Minear M, Park DC. A lifespan database of adult facial stimuli. Behavior Research Methods, Instruments, & Computers 2004; 36 (4): 630-633.

[10] Ebner NC. Age of face matters: age-group differences in ratings of young and old faces. Behavior Research Methods 2008; 4: 130-136.

[11] Ebner NC, Riediger M, Lindenberger U. FACES—A database of facial expressions in young, middle-aged, and older women and men: development and validation. Behavior Research Methods 2010; 42: 351-362.

[12] Ebner NC, Johnson MK. Age-group differences in interference from young and older emotional faces. Cognition and Emotion 2010; 24 (7): 1095-1116.

[13] Guo G, Guo R, Li X. Facial expression recognition influenced by human aging. IEEE Transactions on Affective Computing 2013; 4 (3): 291-298.

[14] Algaraawi N, Morris T. Study on aging effect on facial expression recognition. In: World Congress on Engineering; London, UK; 2016.

[15] Johnston B, Chazal P. A review of image-based automatic facial landmark identification techniques. EURASIP Journal on Image and Video Processing 2018; 1: 86.

[16] Xu Y, Li Z, Yang J, Zhang D. A survey of dictionary learning algorithms for face recognition. IEEE Access 2017; 5: 8502-8514.

[17] Chellappa R. The changing fortunes of pattern recognition and computer vision. Image and Vision Computing 2016; 55 (1): 3-5.

[18] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image Processing 2006; 15 (12): 3736-3745.

[19] Fu Y, Lam A, Sato I, Sato Y. Adaptive spatial-spectral dictionary learning for hyperspectral image restoration. International Journal of Computer Vision 2017; 122 (2): 228-245.

[20] Chen YC, Sastry CS, Patel VM, Phillips PJ, Chellappa R. In-plane rotation and scale invariant clustering using dictionaries. IEEE Transactions on Image Processing 2013; 22 (6): 2166-2180.

[21] Xiang S, Meng G, Wang Y, Panm C, Zhang C. Image deblurring with coupled dictionary learning. International Journal of Computer Vision 2015; 114 (2): 248-271.

[22] Happy SL, Routray A. Robust facial expression classification using shape and appearance features. In: 8th International Conference on Advances in Pattern Recognition; Kolkata, India; 2015. pp. 1-5.

[23] Burkert P, Trier F, Afzal MZ, Dengel A, Liwicki M. DeXpression: Deep convolutional neural network for expression recognition. CoRR 2015; abs/1509.05371.

[24] Ouyang Y, Sang N, Huang R. Accurate and robust facial expressions recognition by fusing multiple sparse representation based classifiers. Neurocomputing 2015; 149: 71-78.

[25] Battini Sönmez E, Albayrak S. A facial component-based system for emotion classification. Turkish Journal of Electrical Engineering and Computer Sciences 2016; 28 (3): 1663-1673.

[26] Feng Q, Yuan C, Pan JS, Yang JF, Chou YT et al. Superimposed sparse parameter classifiers for face recognition. IEEE Transactions on Cybernetics 2017; 47 (2): 378-390.

[27] Moeini A, Faez K, Moeini H, Safai AM. Facial expression recognition using dual dictionary learning. Journal of Visual Communication and Image Representation 2017; 45 (C): 20-33.

[28] Wen J, Xu Y, Li Z, Ma Z, Xu Y. Inter-class sparsity based discriminative least square regression. Neural Networks 2018; 102: 36-47.

[29] Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 2009; 31 (2): 210-227.

[30] Yang M, Zhang L, Feng X, Zhang D. Sparse representation based Fisher discrimination dictionary learning for image classification. International Journal of Computer Vision 2014; 109 (3): 209-232.

[31] Baraniuk RG. Compressive sensing [lecture sotes]. IEEE Signal Processing Magazine 2007; 24 (4): 118-121.

[32] Candes EJ, Wakin MB. An introduction to compressive sampling. IEEE Signal Processing Magazine 2008; 25 (2): 21-30.

[33] Donoho DL. Compressed sensing. IEEE Transactions on Information Theory 2006; 52 (4): 1289-1306.

[34] Battini Sönmez E. Robust Classification Based on Sparsity. 1st ed. Saarbrucken, Germany: LAP Lambert Academic Publishing, 2013.

[35] Battini Sönmez E, Albayrak S. Critical parameters of the sparse representation-based classifier. IET Computer Vision Journal 2013; 7 (6): 500-507.

[36] De La Torre F, Chu W, Xiong X, Vicente F, Cohn JF. Intraface. In: IEEE International Conference on Face and Gesture Recognition; Slovenia; 2015. pp. 1-8.

[37] Shan C, Gong S, McOwan PW. Facial expression recognition based on Local Binary Patterns: a comprehensive study. Image and Vision Computing 2009; 27 (6): 803-816.

[38] Battini Sönmez E, Cangelosi A. Convolutional neural networks with balanced batches for facial expressions recognition. In: SPIE 10341, Ninth International Conference on Machine Vision (ICMV); Nice, France; 2016. doi: 10.1117/12.2268412

[39] Lopes AT, Aguiar Ed, Souza AF, Oliveira-Santos T. Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. Pattern Recognition 2017; 61: 610-628.

[40] Yang P, Yang H, Wei Y, Tang X. Geometry-based facial expression recognition via large deformation: diffeomorphic metric curve mapping. In: 25th IEEE International Conference on Image Processing; Athens, Greece; 2018. pp. 1937-1941.