Image and Vision Computing 144 (2024) 104952

Contents lists available at ScienceDirect

Image and Vision Computing

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

POSER: POsed vs Spontaneous Emotion Recognition using fractal encoding

Carmen Bisogni *, Lucia Cascone, Michele Nappi, Chiara Pero

University of Salerno, Via Giovanni Paolo II 132, Salerno 84084, Fisciano, Italy

ARTICLE INFO

Keywords:
Emotion recognition
Face recognition
Partitioned iterated function systems
Fractal encoding
Machine learning
Spontaneous emotions

ABSTRACT

Emotion recognition from facial expressions is a fundamental human ability that can be harnessed and transferred to machines. The ability to differentiate between spontaneous and posed emotions holds significant importance in various domains, including behavioral biometrics, forensics, and security. This paper introduces a novel method, called POsed vs Spontaneous Emotion Recognition (POSER), which leverages a modified version of the Partitioned Iterated Functions System (PIFS) to obtain a Fractal Encoding. This encoding is used for the first time as facial features to train a machine learning approach for the classification of emotions as either spontaneous or posed. Furthermore, by adapting the original architecture, we demonstrate the effectiveness of these features in distinguishing seven different emotions in controlled as well as wild environments, within a framework referred to as POSER-EMO. Experimental results are presented on the SPOS and DISFA+ datasets for the first classification problem, where POSER outperforms the state of the art, and on the CK+ and SFEW datasets for the second classification problem.

1. Introduction

Emotion recognition based on facial cues is a fundamental cognitive ability of the human species, comparable to face recognition. Along with the capacity to replicate and simulate these emotional states, the ability to decipher human intentions through facial expressions has always fascinated the scientific community [1]. In the fields of behavioral biometrics and forensics, this previously mentioned concept is particularly applicable. However, it also presents a significant challenge due to the scarcity and imbalance of accessible data pertaining to spontaneous emotions. Motivated by these factors, our decision was to explore spontaneous emotion recognition from a new perspective by integrating facial features with machine learning methods. First of all, we observed that in human behavior, the specific facial parts involved in the emotion classification can drastically change the overall classification result [2]. Taking this into consideration, we made the decision to extract pertinent features from the face by employing Partitioned Iterated Function Systems (PIFS), which generate a Fractal Encoding. The fractal encoding has been used in the biometric field in three main domains [3]. Its application to face recognition has been proposed by Tang et al. [4] and Bisogni et al. [5], in both cases to speed up the search process in face templates. The same recognition approach is also effective in case of depth images [6]. It has been also applied to iris recognition by Bourkhriss et al. [7] and Al-Saidi et al. [8] to detect auto-similarities in normalized iris and to perform template protection, respectively. In the case of fingerprint templates it has been applied by Abdullahi et al. [9] and Bajahzar et al. [10] to privacy protection and template reconstruction, respectively. PIFS has been also applied to head pose estimation, very recently by the two works of Bisogni et al. [11,12].

Notably, in this study, we introduce a modified version of PIFS for the generation of a Fractal Encoding array, a novel approach specifically tailored for emotion recognition, marking its inaugural use in this context. This array is subsequently subjected to classification using a machine learning technique.

To showcase the capabilities of our proposed technique, known as POSER, we aim not only to differentiate between spontaneous and posed expressions but also to distinguish between different types of expressions. To achieve this, we combine classifiers in both cascade and parallel configurations, defining POSER-EMO, a method capable of determining the expressed emotion from the facial data. The architectures of POSER and POSER-EMO can be summarized in Figs. 1 and 2 that will be analyzed in detail in Section 4. In particular, the main contributions of this work are:

- The novel utilization of Partitioned Iterated Function Systems (PIFS) as descriptors for emotions, leading to Fractal Encoding. This pioneering approach marks the first-time application of PIFS in the context of emotion recognition.

* Corresponding author.
E-mail address: cbisogni@unisa.it (C. Bisogni).

https://doi.org/10.1016/j.imavis.2024.104952
Received 15 February 2023; Received in revised form 17 October 2023; Accepted 16 February 2024
Available online 19 February 2024

- POSER: This architecture efficiently extracts facial information from an image, employing a modified version of PIFS. It further utilizes PIFS to extract a feature array, enabling the subsequent classification of emotions as either posed or spontaneous.
- POSER-EMO: Going beyond the scope of POSER, this architecture takes the facial extraction process a step further. It applies the modified version of PIFS to extract features and subsequently classifies emotions as either negative or positive. Additionally, it employs a cascade classification approach to identify specific emotions, utilizing either a binary or a multi-class approach.
- A comprehensive range of experiments has been conducted to address both the SPOS versus POSED problem and the challenge of emotion classification. These experiments specifically focus on exploring the structural aspects of the datasets, thereby offering a comprehensive evaluation of the proposed architectures.

The following sections are organized as follows. Section 2 presents and discusses recent related work on both problems. Section 3 provides background information on the Fractal Encoding algorithm and its modified version. In Section 4, we introduce and analyze POSER and POSER-EMO, delving into the transition from the image to emotion prediction. The results obtained and the datasets used are presented in Section 5, accompanied by extensive discussions. Finally, in Section 6, we draw conclusions and propose potential future directions for further exploration in this field.

2. Related works

Emotion recognition from face is a very wide biometrics field. The volume of data and the variety of methods related to this topic continue to grow. For this reason, here we will report only the most recent works on this topic that are relevant in the field. First of all, it is necessary to operate on a distinction between spontaneous emotion recognition and emotion classification. Emotion classification has been a longstanding area of research, but recently, there has been a surge of interest in determining the authenticity of expressed emotions, making it a prominent aspect of behavioral biometrics. Consequently, we will divide the study of facial emotions into these two sub-fields.

2.1. Spontaneous vs posed classification methods

Recognizing if a user is genuinely happy, sad, angry, etc. is undoubtedly an advantage in the study of behavioral biometrics. Different from emotion recognition from the face, in which humans are particularly able to detect the emotion, recognizing a spontaneous or posed expression from the face is a more difficult task for humans. It has been proven that an expression can be created not only by posing but also by acting in different ways. Humans can pose expressions by Mimic or by the Stanislavski method, well known to actors [13]. To detect posed expression, the kinematic of the face movements can be studied, as in [14]. Here, in particular, the authors find a connection between the speed of movement of a set of facial landmarks and the genuineness of the expression. Studies like this are also interested in another field strictly related, the micro-expression recognition [15]. Some facial expressions are more easily recognized; for this reason, some authors focus on an ad hoc study of the latter. The smile has gained particular attention in automatic posed expression recognition. In [16], Saito et al. used geometric features, such as temporal features and reflection intensity distribution, to train a Support Vector Machine able to distinguish between a posed and spontaneous smile with 94.6% of accuracy. In [17] Park et al. solved the same problem by using action units and by studying three-dimensional facial landmarks. Here they observed that spontaneous smiles have higher intensities for the upper face than the lower face, and the opposite happens for a posed smile. On the other hand, considering only an expression is quite limiting in a real-world context. For this reason, lately, some methods are focusing on spontaneous vs posed recognition for a large set of emotions. One of the problems that authors on this topic have to face is the limited amount of data, which makes it difficult to use deep learning techniques. In [18] Racoviteanu et al. use convolutional filters to solve this problem and further highlight the unbalance of state-of-the-art datasets that leads to a preference for the F1 score as a metric instead of classical accuracy. Wang et al., in [19] used several Latent Regression Bayesian Networks (LRBNs) to analyze the patterns of facial landmark points. In their study, the LRBN capture both the latent variables and the probabilistic dependencies among landmarks to model the spatial pattern that characterizes spontaneous expressions. Spatial and temporal patterns together are recently used to recognize a genuine expression, as in [20]. Here, Wang et al. used an interval-temporal-restricted Boltzmann machine (IT-RBM) to study the facial behavior during the expression. Facial expression is seen as a multifarious activity composed of sequential or overlapping primitive facial events. Comparing those methods to POSER, we do not use temporal features but a single frame, and we do not use landmarks but the whole facial image. Also in our study, we took care to present the F1 score to avoid distorted accuracy due to the unbalance of the available datasets. Table 1 outlines the classification methods described to distinguish Spontaneous vs. Posed.

2.2. Emotion classification methods

Compared to spontaneous vs posed expression recognition, which has gained more attention in recent times, emotion classification from the face is a very widely explored topic. The number of datasets and how they are acquired are very heterogeneous, leading to a burgeoning literature on the subject. In the already mentioned [20], the IT-RBM that uses both spatial and temporal variables is also able to classify the performed emotion. In [21], Tanfous et al. use the temporal information to track the facial landmarks and the geometric variations of the latter due to the acquisition of 3D landmarks by using RGB-D sensors. Also in [22], Kacem et al. analyze the temporal variation of landmarks. Here, in the Riemannian manifold of positive matrices of fixed rank, as parametrized trajectories. Then, we can find a set of works focused on deep learning to build an emotion classifier. Jung et al. [23] use two deep networks to extract temporal appearance features and temporal

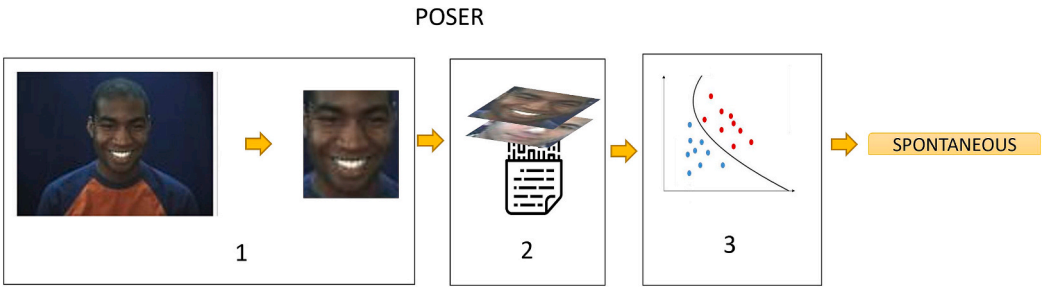

Fig. 1. POSER, SPOS vs POSED architecture: 1 extract the face, 2 obtain the fractal encoding with neutral reference, 3 use the binary classifier to obtain spontaneous or posed.

geometric features, respectively. The two models are then compared using an integration technique. We can also find models that do not use temporal variables, such as the one proposed by Tang et al. [24]. Here, a network called FreNet processes the image in the frequency domain, and the features are extracted at different levels thanks to discrete cosine transforms and weight-shared multiplication kernels. The idea to extract features from different levels using deep learning was also developed by Wang et al. [25]. The network they propose is called OAENet and it is composed of convolutional blocks that extract high-level features and an attention branch to highlight local information. W. Xie et al. presented two methods to solve emotion recognition with deep features. In the first one [26], they used feature sparseness to build a deep model with fewer parameters. In the latest [27] they improve the deep sparseness strategy to automatically adapt the hyper-parameter of the proposed network to different environments proposed by the state-of-the-art datasets. On the other hand, we can find works that are more focused on specific aspects or problems related to emotion recognition from faces. In [28], Umer et al. investigate the impact of data augmentation since, as previously mentioned, the amount of data needed to perform deep learning on this task is very low. S. Xie et al. [29] have focused their efforts on the research of salient regions of the face and on other biometric traits (age, gender, and ethnicity) that lead to different conclusions. In [30] Fu et al. are focused on the semantic perturbation in emotion recognition. Their architecture assumes the form of an asymmetrical autoencoder that takes into account the semantics of neighborhoods. Kar et al. [31] present an approach called KELM (Kernel Extreme Learning Machine). To decompose expression images, they adopted Variational Mode Decomposition (VMD). Then, they used a Whale Optimization (WO) algorithm to fine-tune RBF KELM parameters. To learn feature representation, Tong et al. [32] propose AWOBNet (Adaptive Weight Based on Overlapping Blocks Network). They use feature map blocking to extract local features of the network to improve performance. Gan et al. [33] focusing on eliminating the impact of redundant information from emotional-unrelated regions on facial expression recognition. To reach this aim, they used densely connected CNN and propose to combine the latter with a hierarchical spatial attention method. Poux et al. [34] focused on recognizing partial occluded faces with an optical flow reconstruction. They use denoising auto-encoders that have been trained to reconstruct corrupted optical flow. The loss function used compares the reconstruction with the ground-truth optical flow calculated on images without occlusions. Karnati et al. [35] proposed a network called FLEPNet. FLEPNet's building components are multi-scale convolutional and multi-scale residual block-based DCNN. They explore modified homomorphic filtering to successfully normalize the illumination, thereby minimizing the intraclass difference. Finally, there are also methods that are intended to solve the problem in the wild. In this category, we can find works focused on the occlusions, naturally present in this kind of environment (Wang et al. [36]), or on the challenges offered by variations in pose or illumination (Li et al. [37], Zhang et al. [38]). In some cases, emotion recognition is solved together with other tasks such as Face Synthesis and Face Alignment as proposed by Zhang et al. [39]. Table 2 summarizes the key concepts of the emotion classification methods.

Table 1
Overview of Spontaneous vs. Posed classification methods.

Authors	Method
Saito et al. [16]	Geometric features + SVM
Park et al. [17]	3D facial landmarks + ANOVA
Racoviteanu et al. [18]	CNN features vectors + SVM, RF, MLP
Wang et al. [19]	Facial landmark points + LRBNs
Wang et al. [20]	Interval temporal restricted Boltzmann machine (IT-RBM)
POSER	Fractal encoding algorithm + GNB

Table 2
Overview of emotion classification methods.

Authors	Method
Tanfous et al. [21]	2D/3D facial landmarks + Sparse Coding, Dictionary Learning
Kacem et al. [22]	Facial landmarks (Riemannian geometry) + ppfSVM
Jung et al. [23]	Deep temporal appearance network (DTAN) + Deep temporal geometry network (DTGN)
Tang et al. [24]	Frequency neural network (FreNet)
Wang et al. [25]	Oriented attention pseudo-siamese network (OAENet)
Xie et al. [26,27]	Feature sparseness-based regularization (deep features)
Umer et al. [28]	Deep learning features + augmentation techinques
Xie et al. [29]	Salient expression region descriptor + Encoder/Decoder
Fu et al. [30]	semantic perturbation
Kar et al. [31]	Variational Mode Decomposition (VMD) and Whale Optimization (WO) + KELM.
Tong et al. [32]	Adaptive Weight Based on Overlapping Blocks Network (AWOBNet)
Gan et al. [33]	Densely connected module + Spatial attention module
Poux et al. [34]	auto-encoder with skip connections (optical flow domain)
Karnati et al. [35]	Texture-based feature-level ensemble parallel network (FLEPNet)
Wang et al. [36]	Region Attention Network (RAN)
Li et al. [37]	Deep Locality-Preserving Convolutional Neural Network (DLP-CNN)
Zhang et al. [38]	Generative Adversarial Network (GAN)
Zhang et al. [39]	End-to-end deep learning framework
POSER	Fractal encoding algorithm + Bagging

3. Fractal encoding: An introductory overview

Fractal Encoding is mainly used in image processing to perform fractal compression, based on Partitioned Iterated Function System (PIFS). The first approach using PIFS to solve a biometrics problem was proposed in 1997 in [40]. Here, the Fractal Encoding is used to obtain a representation of the face to solve the face recognition problem. From this work, many other algorithms were developed based on PIFS, that

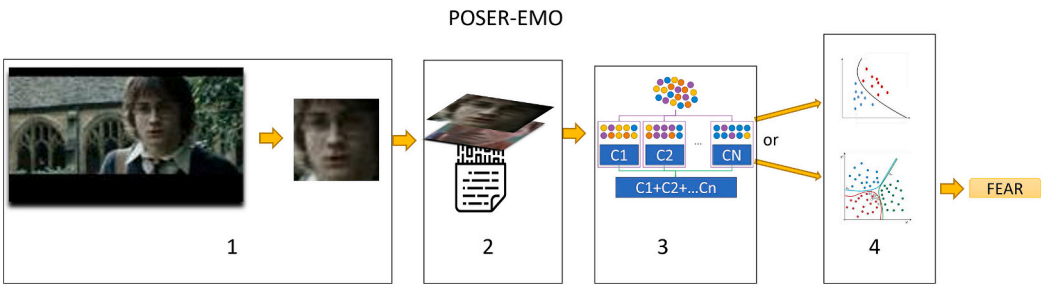

Fig. 2. POSER-EMO, Emotion classification architecture. 1 extract the face, 2 obtain the fractal encoding with neutral reference, 3 use of an ensable of classifier to obtain positive vs negative emotions, 4 based on the results of 3 use a binary or multiclass classifier.

also report the theoretical details of this technique, as in [12]. From a mathematical point of view, PIFS was developed by Arnaud Jacquin [41]. The idea is to codify a small part of the image in a downsampled version of another part of the image by using a set of affine transformations.

Before defining the transformations that can be used for this purpose, we must define the concept of a metric. In particular, for three generic points of the image x,y,z, we can define d as a metric if the following are satisfied:

$$\begin{aligned} d(x,y) = 0 \Leftrightarrow x = y \\ d(x,y) = d(y,x) \\ d(x,z) \leq d(x,y) + d(y,z). \end{aligned} \tag{1}$$

To obtain the reduction of the image, we are interested in functions that reduce the distances between points. For this reason, if d is the metric used to evaluate the distance between the points of the image, we are interested in functions that satisfy:

$$d(f(x),f(y)) \leq kd(x,y) \tag{2}$$

where f is called a *contraction mapping*, and the inequality is true for all the x and y points of the image and $0 \leq k < 1$. Thanks to the fixed point theorem, we are sure that, not only do contraction mappings reduce the distances between points, but also that if we iteratively apply these functions to a point in the image, the resulting point converges to the initial point:

$$\lim_{n\to\infty} f^n(x) = \lim_{n\to\infty} \underbrace{f(f(f(\ldots(x))))}_{n\ times} = x_i \tag{3}$$

where $f(*)$ is the contractive function and x_i is one point of the image.

At this point, we have only to define how a smaller block of pixels of the image can be made similar to a bigger block of pixels of the same image. To this aim, we define the affine transformations as:

$$W(X) = AX + B \tag{4}$$

where A is the transformation matrix, X is the array of the image composed by (x,y) coordinates and z gray level, and B is an offset vector. By affine transformation, an image can be translated, rotated, scaled or modified in contrast and brightness.

Now, putting together all the concepts above introduced, we can define an Iterated Function System as a set F of contractive affine functions $f_1,\ldots f_N$, where F is a contractive function itself of fixed point X called the *attractor*, such that

$$F(X) = \bigcup_{i=1}^{N} f_i(X) = X. \tag{5}$$

4. Method

In this section, we introduce our novel adaptation of the PIFS algorithm, which has been adjusted to address the specific task of Emotion Classification.

4.1. Modified PIFS

Defined as D a collection of sub-domains D_i and F a collection of contractive maps f_i, the step of our PIFS are summarized in the following pseudo-code.

Algorithm 1. Modified PIFS

Result: The set of selected domain blocks $D_{j,i}$. Split the image into R_i non-overlapping blocks
Split the reference into D_i non-overlapping blocks with $dim(D_i) = k \times dim(R_i)$

```
for each feature f_j do
    D_{j,i} ← f_j(D_i)
for each R_i block do
    for each D_{j,i} block near R_i do
        error ← MSE(D_{j,i}, R_i)
        if error < minError then
            minError ← error
            bestBlock ← D_{j,i}
        else
            continue
return The set of selected domain blocks D_{j,i}
```

The result of this algorithm is the list of the selected $D_{j,i}$ transformed blocks with their transformations in terms of translation, rotation, scaling, contrast, and brightness. The original version of the PIFS algorithm can be found in [42], Fig. 4. In particular, we can find in our version two main differences:

1. We use a reference image instead of the image itself to build the Domain blocks. In particular, instead of re-creating the domain blocks for each image to be computed, we use a neutral expression of a random subject as a reference image. On one side, we drastically reduce the necessary computational time; on the other side, we will use the same neutral reference for all the images.
2. The Domain block to be selected is searched only near the range block from a distance point of view. This means that we force the algorithm to encode near blocks in near blocks. By this operation, we further reduce the computational time and also ensure that similar parts of the image are compared. This step is not necessary for other kinds of analysis, on the other hand, it is very useful in emotion recognition.

The size of Domain blocks and range Blocks are chosen to be square, and the differences between them indicate the level of compression. Here, to further speed up the search for the optimal block, we used the optimization of the Fractal Encoding introduced in [43].

In Fig. 3 we can appreciate the modified PIFS extractor on the facial images. The set of blocks obtained is flattened to represent the feature array describing the emotion. Then, as described in the following sections, the feature arrays can be classified by machine learning algorithms.

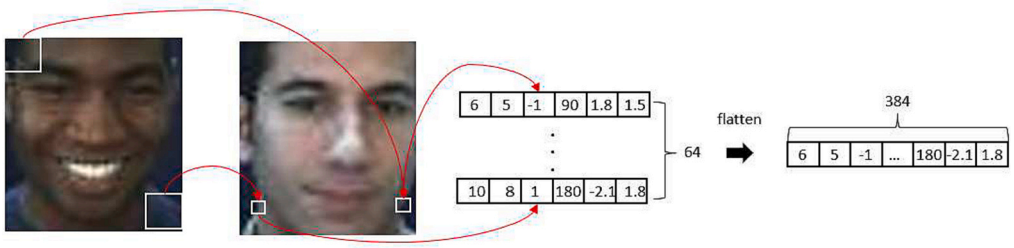

Fig. 3. The core of POSER and POSER-EMO: the modified PIFS to extract the emotion features array. The blocks identifier are composed by 6 elements each, for a total of 64 blocks. After the flatten, the final array has 384 entries.

4.2. Modified PIFS in POSER implementation

In this sub-section, we describe how the Modified PIFS algorithm is implemented in POSER. In particular, we discuss two distinct architectures. We mainly propose a Spontaneous vs. Posed classifier; however, to demonstrate the effectiveness of the Fractal Encoding to also detect the kind of emotion, we also propose an architecture able to distinguish between 7 different emotions, as above introduced. The steps in common between the two architectures concern the manipulation of the images, but the classification steps are different. The shared steps are:

- Step 1, *Face detection*. The face is detected by using Max-Margin Object Detection (MMOD) [44] adapted to faces. This detector uses a Convolutional Neural Network (CNN). The method is robust and able to detect faces from varying view angles, occlusions, and illumination variations.
- Step 2, *Fractal Encoding*. Here, we encode the detected face. By following the steps described in Section 3, we obtain a matrix representing the encoding of the face. Each row of the matrix represents an encoded block, and each column represents the correspondent transformation. To obtain a more easy to compare representation, we transform the matrix into an array by simply making the rows consecutive.

POSER classification

- Step 3, Spos vs Posed Classification. The classifier we used in this step is a Gaussian Naive Bayes classifier. The aim of this classifier is to maximize the probability of belonging to a class by a maximum a posteriori rule:

$$\bar{y} = \underset{k\in\{0,1\}}{arg\ maxp}(C_k)\prod_{i=1}^{n} p(x_i|C_k) \tag{6}$$

where k is the class Spontaneous or Posed, x_i represent the $i-th$ entry of the Fractal Encoding, $p(C_k)$ is the probability of the label C_k and $p(x_i|C_k)$ is the conditional probability that the feature x_i is present in the C_k class. The result of the classifier is a model that takes as input the face encoding and returns 0 if the emotion is Posed and 1 if it is Spontaneous.

POSER-EMO classification

- Step 3, *Positive* vs *Negative classification*. We built a binary classifier that distinguishes between Positive emotions (Happy, Surprise) and Negative emotions (Angry, Contempt, Sad, Fear, Disgust) + Neutral. The classifier used in this step is a Bagging Classifier. This classifier fits basic classifiers trained on a random subset of the original training set and then aggregates their predictions. The base classifiers are trained in parallel, belonging to the ensemble methods. This step takes as input the encoded face and returns 0 if the emotion is positive, 1 if it is negative.
- Step 4, *Emotion Classification*. Once the emotion is classified as positive or negative, in this step we use two different classifiers, one for the positives and one for the negatives. The *positive* classifier is a binary Gaussian NB, that return the positive emotion as *happy* or *surprise*. The *negative* classifier is a multi-class Gaussian NB that contemplates six classes.

In this last architecture, the final accuracy must take into account both the errors committed in Step 3 and the errors of the single classifier in Step 4. For this reason the final accuracy will be:

$$Acc = \frac{N - \left(err_b + err_p + err_n\right)^*}{N}\,100 \tag{7}$$

where N is the total number of tested images, and err_b is the number of errors in the first classifier. The images that are wrongly classified by the first classifier will automatically be considered errors since they will be passed to the wrong emotion classifier. err_p is the number of errors committed by the *positive* classifier, and err_n is the number of errors committed by the *negative* classifier. err_p and err_n consider only the error to classify emotion since the error between *positive* and *negative* has already been taken into account by err_b.

In Fig. 1 we can appreciate the workflow of POSER for SPOS vs POSED problem with the steps above mentioned. Here, starting from an image from the SPOS dataset we will introduce in the next section, we see the Fractal Encoding step with a neutral reference and the use of the binary Gaussian NB classifier. In Fig. 2 we show the Emotion Classification architecture of POSER-EMO, where starting from a SFEW image fractal encoded, we firstly classify the *positive* or *negative* emotion by a Bagging classifier and then we use a binary or multi-class Gaussian NB classifier to obtain the final prediction.

The total time to perform the pre-processing phase on an image, including face detection and localization, is 0.98 s. To compute the fractal encoding, the algorithm requires 0.46 s. The time taken by GaussianNB to predict a single sample was approximately 2.7×10^{-4} seconds. All the times were obtained by performing the experiments on a MacBook Pro 2.6 GHz Intel Core i7 6 core 16 GB 2667 MHz DDR4 Intel UHD Graphics 6,301,536 MB, with Python 3.7.6. We can notice that the majority of time is required for face detection and localization. In fact, other than this, the algorithm requires less than half a second to provide the final estimation. We can thus observe that a faster detector can drastically improve the computational time required, and an optimization technique could also be implemented on the fractal encoding search strategy (already speeded up in the modified PIFS we propose compared to the traditional one).

5. Experiments

This section illustrates the datasets, presents the comparison results with state-of-the-art methods, provides insightful discussions, and includes additional studies to further enhance our understanding.

5.1. Dataset descriptions

In this section we present the Dataset used for both Spontaneity Recognition and Emotion Recognition.

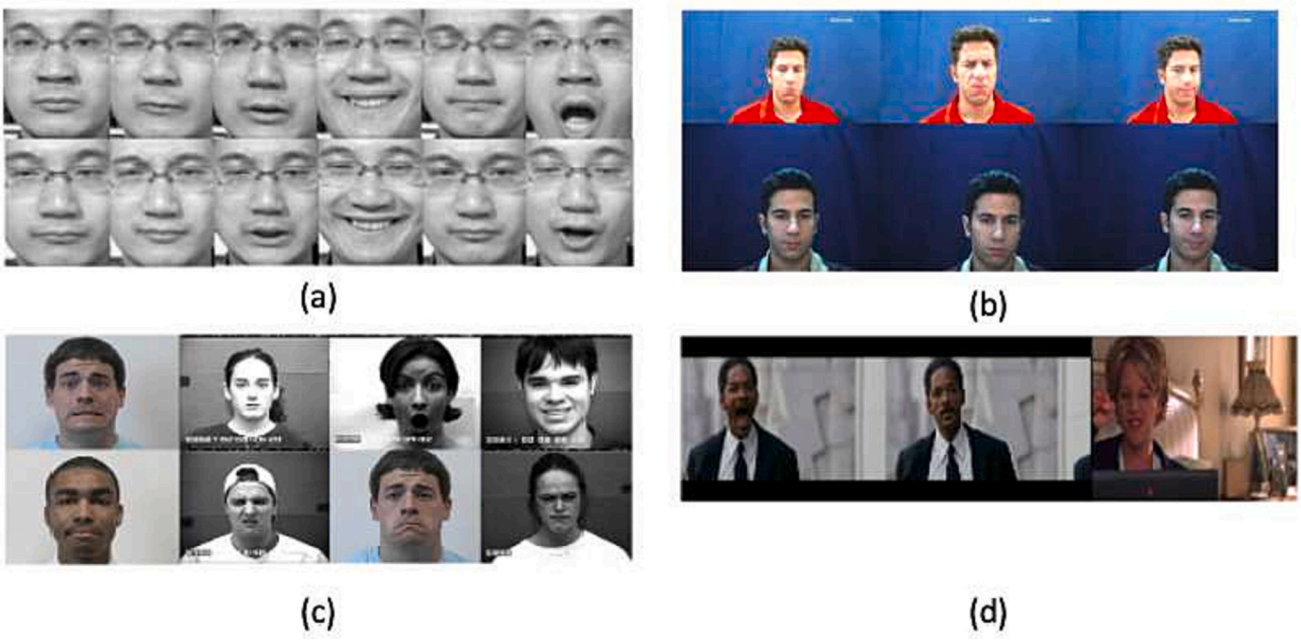

Fig. 4. The used Databases. (a) SPOS (b) DISFA and DISFA+, (c) CK+ (d) SFEW. On the top row the posed anger, disgust, fear, happy, sad, surprise. On the bottom row the spontaneous ones.

5.1.1. SPOS

The Spontaneous vs. Posed Facial Expression Database (SPOS) [45] includes both Spontaneous and Posed Expressions. In particular, 7 subjects were recorded twice. In the first session, they watched a movie clip to induce spontaneous expression, and in the second session, they repeated all the previous. There are 343 posed frames and 1583 spontaneous frames. We noticed that this dataset is highly unbalanced; for this reason, in experiments we also evaluated F1-score, other than accuracy. Examples of the SPOS database can be seen in Fig. 4 (a). On the top row the posed anger, disgust, fear, happy, sad, surprise. On the bottom row, the spontaneous ones.

5.1.2. DISFA and DISFA+

The Denver Intensity of Spontaneous Facial Action Database (DISFA) and the Extended Denver Intensity of Spontaneous Facial Action Database (DISFA+) are spontaneous and posed databases, respectively. DISFA [46] contains video sequences of 27 adults. During the video sequences, the spontaneous expressions are stimulated by showing different videos to the subjects. DISFA+ [47] has been released as an extended version of DISFA, in particular, to add to the previous one the respective posed expression, as seen in Fig. 4 (b).

5.1.3. CK+

The Extended Cohn-Kanade (CK+) Dataset [48] is composed of 123 different subjects recorded for a total of 593 video sequences in posed expressions. The frames that we extracted to perform the emotion recognition experiments are the same well known in the literature (as in [20]). In particular, 45 for Anger, 18 for Contempt, 59 for Disgust, 25 for Fear, 69 for Happy, 28 for Sadness, 83 for Surprise, 67 for Neutral. In Fig. Fig. 4 (c) some samples from the dataset can be appreciated.

5.1.4. SFEW

The Static Facial Expressions in the Wild (SFEW) dataset [49] is a dataset created by the frames of the previous AFEW Dataset. The dataset already comes split into Train, Validation, and Test with 958, 436, and 372 samples, respectively. The images are from film sequences, and each frame is labeled with one of the seven expressions: anger, disgust, fear, neutral, happiness, sadness, and surprise. Sample images of SFEW can be found in Fig. 4 (d).

We used SPOS and DISFA+ to perform the experiments of POSER on Spontaneous vs Posed recognition, CK+ to evaluate POSER on emotion recognition on a posed database, and SFEW to evaluate POSER on emotion recognition methods on a dataset in the wild.

5.2. Evaluation protocols

Spontaneous vs Posed detection. For the Spontaneous vs. Posed detection experiments, we conducted two types of experiments: an intra-dataset experiment and a cross-dataset experiment. In the intra-dataset experiment, we were aware of the potential impact that the biometric component could have on the performance, so we performed a k-fold cross-validation. Specifically, we used five-fold subject-independent cross-validation on the SPOS database and nine-fold subject-independent cross-validation on the DISFA+ database, following the methodology described in [20]. In the cross-dataset experiment, we alternated between using the two datasets as training and testing sets, respectively. These experimental setups allow us to assess the performance of the Spontaneous vs Posed detection model on both datasets and evaluate its generalization capabilities to unseen subjects and real-world scenarios.

Emotion Classification. For the emotion classification task, we conducted two experiments. For CK+, we performed a subject-independent experiment where 80% of the subjects were used for training. This ensures that the model is trained on a diverse set of subjects. The remaining 20% of subjects were used for testing to evaluate the model's performance on unseen individuals. On the other hand, for the SFEW dataset, the data is already split into training and testing sets. Therefore, we utilized the provided split, with the training set used for model training and the testing set used for evaluating the model's performance on unseen samples from the dataset.

5.3. Implementation details

Spontaneous vs Posed detection. For binary classification to distinguish between posed and spontaneous emotions, a binary Gaussian Naive Bayes classifier was employed. This classifier assumes a Gaussian distribution and applies the Naive Bayes assumption of conditional independence of features given the class label. To optimize its performance, a grid search methodology was used to search for the optimal

parameter 'var_smoothing'. This involved systematically evaluating the classifier's performance across a predefined set of 'var_smoothing' values to find the one that yielded the best results. During the training phase, a k-fold approach with $k = 5$ was used to mitigate overfitting. This involved dividing the training data into five equal-sized subsets or folds. The classifier was trained and evaluated five times, with each fold serving as the validation set once while the remaining folds were used for training. By repeating this process and averaging the performance across all folds, the impact of overfitting was minimized.

Emotion Classification. For the classification of emotion as negative or positive, a bagging classifier was used instead. The bagging classifier creates an ensemble of base classifiers by training each classifier on different subsets of the training data. This is done through bootstrapping, where random subsets of the original data are sampled with replacement to generate new training sets for each base classifier. By training the base classifiers on different subsets, the bagging classifier captures diverse patterns and reduces overfitting. Similarly, a k-fold strategy and a grid search were applied for the bagging classifier. The grid search explored different combinations of hyperparameters, such as the number of baseline estimators (n_estimators), the maximum number of samples used for each baseline estimator (max_samples), and the maximum number of features considered for each baseline estimator (max_features). The goal was to find the configuration that produced the best results. Finally, for the cascade classification of specific emotions, a Gaussian Naive Bayes classifier was used, following the same approach as described earlier.

5.4. POSER results

The first dataset we take under consideration for Spontaneous vs Posed detection is SPOS. The face is extracted and Fractal Encoding is performed as discussed in Section 4. Those operations have been conducted on both training than testing sets. The classifier we used is GaussianNB. In particular, the variance smoothing is set to 1e-07. We considered various configurations of the Fractal Encodings with different range and domain values. When the values of range and domain pixels increase, the resulting encoding array decrease. For this reason, we also tested the configuration that gives us the longest array (4 pixel range and 8 pixel domain) with PCA analysis to reduce the dimension. The experimental results on SPOS are reported in Table 3. As we can notice, due to an unbalanced dataset, the most balanced situation between classes is not the one with the highest accuracy. In particular, we noticed that when we reduce almost 1/4 of the original dimension of the array through PCA (from 384 to 100), the Spontaneous and Posed expressions have similar accuracy. In a real use case in-the-wild scenario, since the Spontaneous expressions are more frequent, it could be better to not use PCA, obtaining an accuracy 0.12 points higher. However, if we imagine using the algorithm in a controlled scenario (e.g. a lie detector test), we can take more advantage of a method able to detect posed expressions with higher accuracy. For this reason, from the results of SPOS, we can suggest that the PCA technique can be more or less desirable depending on the application scenario.

Then we conducted the same experiments on the DISFA+ Dataset (Table 4), using the same classifier, GaussianNB, and a variance smoothing of 1e-09. Here we can notice that there is no advantage in the use of PCA for the balance of the accuracy of the classes, the best result is in fact obtained in (4, 8) configuration without PCA. We can impute this difference to SPOS Dataset, to the fact that DISFA+ is balanced in posed and spontaneous expressions.

Table 3
SPOS vs POSED test on SPOS Dataset.

SPOS Dataset				
Encoding	PCA	Acc POSED	Acc SPOS	Mean Acc
(4,8)	**None**	**0.46**	**0.93**	**0.85**
(8,16)	None	0.47	0.68	0.65
(16,32)	None	0.60	0.22	0.29
(4,8)	50	0.49	0.72	0.67
(4,8)	**100**	**0.72**	**0.73**	**0.73**
(4,8)	150	0.68	0.76	0.75

Table 4
SPOS vs POSED test on DISFA+ Dataset.

DISFA+ Dataset				
Encoding	PCA	Acc POSED	Acc SPOS	Mean Acc
(4,8)	**None**	**0.9988**	**0.9988**	**0.9977**
(8,16)	None	0.9966	0.9955	0.9944
(16,32)	None	0.99	0.84	0.91
(4,8)	50	0.99	0.89	0.94
(4,8)	100	0.99	0.88	0.94
(4,8)	150	0.99	0.87	0.93

Table 5
Results in classification of Spontaneous vs Posed Expressions.

Dataset	SPOS		DISFA+	
Method	Accuracy	F1-score	Accuracy	F1-score
Wang et al. [20]	0.8291	0.8051	0.9490	0.9404
Racoviteanu et al. [18]	0.684	0.73	/	/
Wang et al. [19]	0.7607	0.6364	0.9053	0.9362
Yang et al. [16]	**0.8547**	0.8632	0.9296	0.9470
POSER	0.8483	**0.91**	**0.9977**	**0.9977**

In Table 5 we reported the comparisons on both datasets with the state of the art.

As can be observed by Table 5, POSER outperforms the state of the art in both SPOS than DISFA+. In particular, also on SPOS, where the accuracy of the method by [16] is higher than ours, our F1 score is significantly higher. Since the SPOS dataset is highly unbalanced, the F1-score is a more reliable index compared to accuracy. This means that POSER, even if with an accuracy of 0.0064 lower than the one of [16], has a better distribution of the error between the classes, with a difference of 0.0468 in the F1-score. On DISFA+, POSER is significantly better both in terms of Accuracy than F1-score. In this case, we can also underline that, differently from SPOS, the DISFA+ dataset is balanced in terms of spontaneous and posed, so in this case, the accuracy is more reliable with respect to the case of SPOS. Those results clearly demonstrate how Fractal Encoding is suitable to distinguish posed from spontaneous emotions. This can be explained by the property of the method to recognize differences in the same expressions.

To analyze the robustness of POSER we also performed across tests using SPOS as training set and DISFA+ as a test set and vice-versa. The results are shown in Table 6. Here we can appreciate that POSER is robust even if the datasets used are very different. In this experiment, the behavior connected to the dataset distribution can be further observed, since even if with a similar mean accuracy, the distribution over the Posed and Spontaneous classes are the opposite.

5.5. POSER-EMO results

In our study on emotion classification, we used the CK+ dataset as the main dataset. However, CK+ posed a challenge because of the limited number of total frames available. As a result, we only had a small number of samples per emotion, despite each sample having a high-

Table 6
Mixed experiments.

Config	Train	Test	POS Acc	SPOS Acc	Mean Acc
PCA	SPOS	DISFA+	0.94	0.53	0.75
PCA	DISFA+	SPOS	0.42	0.92	0.83

dimensional coding array of 576 elements. We included all emotions in the dataset, including neutrals and the often-avoided emotion of contempt, which had a very small number of instances (18) and shared similarities with disgust. We solved the emotion recognition task by using network in cascade. Firstly we divided the emotions into negative and positive. Angry, Contempt, Disgust, Fear, Neutral, Sadness are labeled as Negative. Happy and Surprise are labeled as Positive. A Bagging Classifier is firstly trained to distinguish between negative and positive emotions. The parameters of the classifier are set as: max features 0.6, max samples 0.8, estimators 50, random state 42. We will refer to this network as PosNeg. Then, two GaussianNB Classifiers with 1e-09 of variance are trained to distinguish positive the emotions inside positives and negatives, PosC and NegC respectively. Since the classifiers need to be executed in cascade, the overall accuracy is determined as follows: The images in the test set first pass through PosNeg. Based on the classification by PosNeg, the images are then sent to the corresponding network predicted by PosNeg. If PosNeg misclassifies an image, there is no opportunity for the image to be correctly classified since it will be directed to the wrong classifier. However, if PosNeg accurately classifies the image, the correctness of the prediction depends solely on PosC or NegC. Hence, achieving high accuracy in PosNeg is crucial to prevent error propagation throughout the network and ensure accurate classification.

Based on the data presented in Table 7, it is evident that classifying positive emotions is comparatively easier. Conversely, when examining negative emotions, we noticed a significant level of variability in terms of relative accuracy. This variability is clearly depicted in Table 8, which displays the confusion matrix for the NegC.

The observed phenomenon cannot be attributed solely to the fact that PosC is used for a binary problem and NegC for a multi-class problem. Even when separate binary networks are constructed for each negative expression, a significant portion of the expressions (approximately 15% for each network) are misclassified. This implies that multiple binary negative networks claim the emotion to belong to their respective classes with high accuracy, leading to an undecidable problem. Table 9 provides a detailed report on this phenomenon.

In Table 10 we compare POSER-EMO on CK+ with the state of the art. Since no one in the state of the art considers both neutral than contempt in the final accuracy reported, we also use VGGFace2 [50] as a feature extractor, by substituting the dense layers to be adaptable to CK+, as is customary in the use of large network architectures on a small dataset.

From the results, it is clear that POSER is comparable with the ones in [20]. In fact, POSER-EMO, VGGFace2 and the method in [20] are those who do not use DataAugmentation. Since CK+ is very small, this step makes a huge difference in terms of accuracy and it is not practicable where the data are numeric (like the features we extracted).

We then tested POSER-EMO on a dataset in the wild, SFEW. Also in this case we used the mouth and the split in binary. In particular, the result obtained on positive and negative expressions are in Table 11.

We can observe that, once an emotion is classified as positive, it is simple for POSER-EMO to correctly detect what positive it is (happy, surprise). On the other hand, classify the negative emotions is challenging. In particular, the final accuracy given by the cascade is highly impacted by the low accuracy obtained in distinguish negative from positive images, that generates an error in the first step of the method that does not allow the following networks to receive the correct emotions.

In particular, for the negative emotions, the confusion matrix of the first network is in Table 12.

It is clear that the positive spontaneous emotions are not well classified. If on one hand it means that only easy positive emotions pass to the following network (PosC), on the other hand, the remaining images are wrong regardless of the prediction of the following network.

The confusion matrix of the negative spontaneous images is shown in Table 13.

Table 7
The accuracies obtained on CK+.

Classifier	Accuracy
PosNeg	0.9665
PosC	1
NegC	0.77
Cascade	**0.8323**

Table 8
The confusion matrix of NegC on CK+.

Emotions	Angry	Cont.	Disg.	Fear	Neut.	Sad
Angry	**0.5806**	0.0645	0.0967	0.0967	0.1612	0
Cont.	0	**0.5555**	0	0.2222	0.2222	0
Disg.	0.0869	0	**0.7826**	0	0.0869	0.0434
Fear	0	0	0.0769	**0.6923**	0.2307	0
Neut.	0.0102	0.0408	0.0102	0	**0.8877**	0.0510
Sad	0.01428	0	0	0	0.3571	**0.50**

Table 9
Accuracy of the binary classifiers.

Emotion	Accuracy	Classifier
Fear	0.9491	DecisionTree
Sad	0.8034	SVC
Disgust	0.8728	SGDClassifier
Angry	0.8119	SVC
Contempt	0.8601	GaussianNB
Neutral	0.8432	AdaBoostClassifier

Table 10
Comparisons with the state of the art on CK+. C indicates that 'Contempt' has not been considered, N indicate that 'Neutral' has not been considered.

Method	Accuracy
Wang et al. [20]-N	0.8716
Tanfous et al. [21]-N	0.9573
Kacem et al. [22]-N	0.9687
Jung et al. [23]-N	0.9725
Fu et al. [30]-N,C	0.9858
Tang et al. [24]-N	0.9841
Wang et al.[25]-C	0.9813
W. Xie et al. [26]-C	0.9783
W. Xie et al. [27]-C	0.9759
S. Xie et al. [29]-N,C	0.9588
Umer et al. [28]-N	0.9769
Kar et al. [31]	0.9881
Gan et al. [33]-N	0.9571
Poux et al. [34]-N	0.928
Karnati et al. [35]-C	**0.9894**
VGGFace2	0.777
POSER-EMO	0.8323

Table 11
The accuracies obtained on Sfew.

Classifier	Accuracy
PosNeg	0.552
PosC	0.9411
NegC	0.4035
Cascade	**0.2911**

Table 12
The accuracies obtained on Sfew.

Emotions	Negative	Positive
Negative	0.6333	0.3666
Positive	0.6136	0.3863

Table 13
The confusion matrix of NegC on Sfew.

Emotions	Angry	Disg.	Fear	Neut.	Sad
Angry	**0.60**	0.066	0	0.20	0.133
Disg.	0	**0**	0	0.5	0.5
Fear	0.44	0.11	**0.11**	0	0.33
Neut.	0.2352	0.0588	0.05886	**0.3529**	0.2941
Sad	0.3333	0	0.0833	0	**0.5833**

Table 14
Comparisons with the state of the art on SFEW.

Method	Accuracy
Li et al. [37]	0.5105
Zhang et al. [38]	0.2724
Zhang et al. [39]	0.2910
Li et al. [51]	0.5229
Wang et al. [36]	0.564
Tong et al. [32]	**0.6241**
POSER-EMO	0.2911

As we can observe, the emotion with the best accuracy is angry, followed by Sad. The lower is Disgust since the method can not detect any emotion of this kind.

By comparisons with the state of the art, we observe that SFEW is a very challenging dataset. In Table 14 we reported those comparisons. In particular, we can observe that the accuracy of POSER-EMO is comparable with methods that do not perform pre-training on other datasets. As a consequence, even if the overall accuracy of all the methods is small, the best accuracy coincides with those deep learning methods that operate on SFEW by using transfer learning. In this way, they can compensate for the small amount of data available in SFEW.

6. Conclusions

The analysis of emotion recognition can be approached from two distinct perspectives. Firstly, it involves distinguishing between spontaneous and posed expressions, while secondly, it involves classifying different types of expressions. This paper addresses these challenges by first tackling the posed versus spontaneous recognition problem. We achieve this by utilizing a modified version of PIFS as facial features, which are subsequently classified using machine learning techniques. Our findings demonstrate that the proposed technique, known as POSER, outperforms the current state-of-the-art methods in this domain. Furthermore, we extend our investigation to encompass the detection of user expressions in both controlled and uncontrolled environments, by combining the proposed features with a cascade architecture comprising various classifiers. This integrated approach, named POSER-EMO, yields positive results. The utilization of POSER emphasizes the significance of Fractal Encoding, originally developed for image compression, and subsequently applied in face recognition and head pose estimation. Our study demonstrates its efficacy in resolving the challenges associated with emotion recognition. In future research, building upon the success of POSER in distinguishing between spontaneous and posed emotions, which predominantly involve micro-expressions, we plan to explore its application in this particular domain. Additionally, considering that several methods have achieved notable outcomes by incorporating temporal variables, we propose incorporating temporal features into our Fractal Encoding approach, considering consecutive frames. This integration of PIFS-based Fractal Encoding into the architectures of POSER and POSER-EMO not only enhances the accuracy and robustness of emotion classification but also opens up new possibilities for investigating the relationship between emotions and fractal geometry. In previous works on fractal encodings that involved face detection, it has been used a combination of Viola Jones detection algorithm refined by dlib. In the present work we used a mediapipe extension based on Max-Margin Object Detection (MMOD). This has been demonstrated to be superior in terms of performance compared to Viola-Jones [52]. In particular, the percentage of false positives is drastically reduced. In future, the face detector used can be easily substituted with the new implementations proposed by mediapipe (as BlazeFace Sparse) or other distributions.

CRediT authorship contribution statement

Carmen Bisogni: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation. **Lucia Cascone:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation. **Michele Nappi:** Conceptualization, Visualization, Supervision, Project administration. **Chiara Pero:** Conceptualization, Methodology, Software, Validation, Data curation, Investigation.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

This work was partially supported by the project Information Disorder Awareness (IDA) included in the Spoke 2 - Misinformation and Fakes of the Research and Innovation Program PE00000014, “SEcurity and RIghts in the CyberSpace (SERICS)”, under the National Recovery and Resilience Plan, Mission 4 “Education and Research” - Component 2 “From Research to Enterprise” - Investment 1.3, funded by the European Union - NextGenerationEU.

References

[1] J. Stanley, F. Blanchard-Fields, Challenges older adults face in detecting deceit: the role of emotion recognition, Psychol. Aging 23 (2008) 24–32, https://doi.org/10.1037/0882-7974.23.1.24.
[2] M. Wegrzyn, M. Vogt, B. Kireclioglu, J. Schneider, J. Kissler, Mapping the emotional face. How individual face parts contribute to successful emotion recognition, PLoS One 12 (2017) 1–15, https://doi.org/10.1371/journal.pone.0177239.
[3] C. Bisogni, A. Castiglione, F. Narducci, C. Pero, Fractal encoding uses and variations in contemporary biometrics applications, in: 2023 24th International Conference on Control Systems and Computer Science (CSCS), 2023, pp. 281–287, https://doi.org/10.1109/CSCS59211.2023.00051.
[4] Z. Tang, X. Wu, X. Leng, W. Chen, A fast face recognition method based on fractal coding, SIViP 11 (10) (2017), https://doi.org/10.1007/s11760-017-1078-7.
[5] C. Bisogni, M. Nappi, C. Pero, S. Ricciardi, Pifs scheme for head pose estimation aimed at faster face recognition, IEEE Trans. Biometrics Behav. Identity Sci. 4 (2) (2022) 173–184, https://doi.org/10.1109/TBIOM.2021.3122307.
[6] U. Bilotti, C. Bisogni, M. Nappi, C. Pero, Depth camera face recognition by normalized fractal encodings, in: G.L. Foresti, A. Fusiello, E. Hancock (Eds.), Image Analysis and Processing – ICIAP 2023, Springer Nature Switzerland, Cham, 2023, pp. 196–208.
[7] H. Boukhriss, A.Y. Jammoussi, I.K. Kallel, N. Derbel, Fractal analysis for Iris multimodal biometry, springer international publishing, Cham (2021) 41–62, https://doi.org/10.1007/978-3-030-81982-8_3.
[8] N.M.G. Al-Saidi, A.J. Mohammed, R.J. Al-Azawi, A.H. Ali, Iris features via fractal functions for authentication protocols, Int. J. Innov. Comput. Inform. Control 15 (4) (2019), https://doi.org/10.24507/ijicic.15.04.1441.
[9] S.M. Abdullahi, H. Wang, T. Li, Fractal coding-based robust and alignment-free fingerprint image hashing, IEEE Trans. Inf. Foren. Secur. 15 (2020) 2587–2601, https://doi.org/10.1109/TIFS.2020.2971142.
[10] A. Bajahzar, H. Guedri, Reconstruction of fingerprint shape using fractal interpolation, Int. J. Adv. Comput. Sci. Appl. 10 (2019) 103–114, https://doi.org/10.14569/IJACSA.2019.0100514.
[11] C. Bisogni, M. Nappi, C. Pero, S. Ricciardi, Hp2ifs: Head pose estimation exploiting partitioned iterated function systems, in: 2020 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 1725–1730, https://doi.org/10.1109/ICPR48806.2021.9413227.

[12] C. Bisogni, M. Nappi, C. Pero, S. Ricciardi, Fashe: a fractal based strategy for head pose estimation, IEEE Trans. Image Process. 30 (2021) 3192–3203, https://doi.org/10.1109/TIP.2021.3059409.
[13] M. Conson, M. Ponari, E. Monteforte, G. Ricciato, M. Sarà, D. Grossi, L. Trojano, Explicit recognition of emotional facial expressions is shaped by expertise: evidence from professional actors, Front. Psychol. 4 (2013) 382, https://doi.org/10.3389/fpsyg.2013.00382.
[14] S. Sowden, B. Schuster, C. Keating, D. Fraser, J. Cook, The role of movement kinematics in facial emotion expression production and recognition, Emotion (2021), https://doi.org/10.1037/emo0000835.
[15] B. Sun, S. Cao, D. Li, J. He, L. Yu, Dynamic micro-expression recognition using knowledge distillation, IEEE Trans. Affect. Comput. (2020) 1, https://doi.org/10.1109/TAFFC.2020.2986962.
[16] J. Yang, S. Wang, Capturing spatial and temporal patterns for distinguishing between posed and spontaneous expressions, in: Proceedings of the 25th ACM International Conference on Multimedia, 2017, pp. 469–477.
[17] S. Park, K. Lee, J.-A. Lim, H. Ko, T. Kim, J.-I. Lee, H. Kim, S.-J. Han, J.-S. Kim, S. Park, J.-Y. Lee, E.C. Lee, Differences in facial expressions between spontaneous and posed smiles: automated method by action units and three-dimensional facial landmarks, Sensors 20 (4) (2020).
[18] A. Racoviteanu, C. Florea, M. Badea, C. Vertan, Spontaneous emotion detection by combined learned and fixed descriptors, in: 2019 International Symposium on Signals, Circuits and Systems (ISSCS), 2019, pp. 1–4, https://doi.org/10.1109/ISSCS.2019.8801755.
[19] S. Wang, L. Hao, Q. Ji, Posed and spontaneous expression distinction using latent regression bayesian networks, ACM Trans. Multimed. Comput. Commun. Appl. 16 (3) (2020), https://doi.org/10.1145/3391290.
[20] S. Wang, Z. Zheng, S. Yin, J. Yang, Q. Ji, A novel dynamic model capturing spatial and temporal patterns for facial expression analysis, IEEE Trans. Pattern Anal. Mach. Intell. 42 (9) (2020) 2082–2095, https://doi.org/10.1109/TPAMI.2019.2911937.
[21] A.B. Tanfous, H. Drira, B.B. Amor, Sparse coding of shape trajectories for facial expression and action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 42 (10) (2020) 2594–2607, https://doi.org/10.1109/TPAMI.2019.2932979.
[22] A. Kacem, M. Daoudi, B.B. Amor, J.C. Alvarez-Paiva, A novel space-time representation on the positive semidefinite cone for facial expression recognition, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 3199–3208, https://doi.org/10.1109/ICCV.2017.345.
[23] H. Jung, S. Lee, J. Yim, S. Park, J. Kim, Joint fine-tuning in deep neural networks for facial expression recognition, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 2983–2991, https://doi.org/10.1109/ICCV.2015.341.
[24] Y. Tang, X. Zhang, X. Hu, S. Wang, H. Wang, Facial expression recognition using frequency neural network, IEEE Trans. Image Process. 30 (2021) 444–457, https://doi.org/10.1109/TIP.2020.3037467.
[25] Z. Wang, F. Zeng, S. Liu, B. Zeng, Oaenet: oriented attention ensemble for accurate facial expression recognition, Pattern Recogn. 112 (2021) 107694, https://doi.org/10.1016/j.patcog.2020.107694.
[26] W. Xie, W. Chen, L. Shen, J. Duan, M. Yang, Surrogate network-based sparseness hyper-parameter optimization for deep expression recognition, Pattern Recogn. 111 (2021) 107701, https://doi.org/10.1016/j.patcog.2020.107701.
[27] W. Xie, X. Jia, L. Shen, M. Yang, Sparse deep feature learning for facial expression recognition, Pattern Recogn. 96 (2019) 106966, https://doi.org/10.1016/j.patcog.2019.106966.
[28] S. Umer, R. Rout, C. Pero, M. Nappi, Facial expression recognition with trade-offs between data augmentation and deep learning features, J. Ambient. Intell. Humaniz. Comput. (2021), https://doi.org/10.1007/s12652-020-02845-8.
[29] S. Xie, H. Hu, Y. Wu, Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition, Pattern Recogn. 92 (2019) 177–191, https://doi.org/10.1016/j.patcog.2019.03.019.
[30] Y. Fu, X. Wu, X. Li, Z. Pan, D. Luo, Semantic neighborhood-aware deep facial expression recognition, IEEE Trans. Image Process. 29 (2020) 6535–6548, https://doi.org/10.1109/TIP.2020.2991510.
[31] N. Kar, K. Babu, S. Bakshi, Facial expression recognition system based on variational mode decomposition and whale optimized kelm, Image Vis. Comput. 123 (2022) 104445, https://doi.org/10.1016/j.imavis.2022.104445.
[32] X. Tong, S. Sun, M. Fu, Adaptive weight based on overlapping blocks network for facial expression recognition, Image Vis. Comput. 120 (2022) 104399, https://doi.org/10.1016/j.imavis.2022.104399.
[33] C. Gan, J. Xiao, Z. Wang, Z. Zhang, Q. Zhu, Facial expression recognition using densely connected convolutional neural network and hierarchical spatial attention, Image Vis. Comput. 117 (C) (2022), https://doi.org/10.1016/j.imavis.2021.104342.
[34] D. Poux, B. Allaert, N. Ihaddadene, I.M. Bilasco, C. Djeraba, M. Bennamoun, Dynamic facial expression recognition under partial occlusion with optical flow reconstruction, IEEE Trans. Image Process. 31 (2022) 446–457, https://doi.org/10.1109/TIP.2021.3129120.
[35] M. Karnati, A. Seal, A. Yazidi, O. Krejcar, Flepnet: feature level ensemble parallel network for facial expression recognition, IEEE Trans. Affect. Comput. 13 (4) (2022) 2058–2070, https://doi.org/10.1109/TAFFC.2022.3208309.
[36] K. Wang, X. Peng, J. Yang, D. Meng, Y. Qiao, Region attention networks for pose and occlusion robust facial expression recognition, IEEE Trans. Image Process. 29 (2020) 4057–4069, https://doi.org/10.1109/TIP.2019.2956143.
[37] S. Li, W. Deng, Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition, IEEE Trans. Image Process. 28 (1) (2019) 356–370, https://doi.org/10.1109/TIP.2018.2868382.
[38] F. Zhang, T. Zhang, Q. Mao, C. Xu, Geometry guided pose-invariant facial expression recognition, IEEE Trans. Image Process. 29 (2020) 4445–4460, https://doi.org/10.1109/TIP.2020.2972114.
[39] F. Zhang, T. Zhang, Q. Mao, C. Xu, A unified deep model for joint facial expression recognition, face synthesis, and face alignment, IEEE Trans. Image Process. 29 (2020) 6574–6589, https://doi.org/10.1109/TIP.2020.2991549.
[40] A. Kouzani, F. He, K. Sammut, Fractal face representation and recognition, in: 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation vol. 2, 1997, pp. 1609–1613, https://doi.org/10.1109/ICSMC.1997.638231.
[41] A. Jacquin, A Fractal of Iterated Markov Operators with Applications to Digital Image Coding, 1989.
[42] A. Ozturk, C. Gungor, A fast fractal image compression algorithm based on a simple similarity measure, Inform. Syst. Sci. 36 (2011) 159–178.
[43] R. Distasi, M. Nappi, D. Riccio, A range/domain approximation error-based approach for fractal image compression, 2005.
[44] S. Maji, J. Malik, Object detection using a max-margin hough transform, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1038–1045, https://doi.org/10.1109/CVPR.2009.5206693.
[45] T. Pfister, X. Li, G. Zhao, M. Pietikäinen, Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework, in: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011, pp. 868–875, https://doi.org/10.1109/ICCVW.2011.6130343.
[46] S.M. Mavadati, M.H. Mahoor, K. Bartlett, P. Trinh, J.F. Cohn, Disfa: a spontaneous facial action intensity database, IEEE Trans. Affect. Comput. 4 (2) (2013) 151–160, https://doi.org/10.1109/T-AFFC.2013.4.
[47] M. Mavadati, P. Sanger, M.H. Mahoor, Extended disfa dataset: investigating posed and spontaneous facial expressions, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016, pp. 1452–1459, https://doi.org/10.1109/CVPRW.2016.182.
[48] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2010, pp. 94–101, https://doi.org/10.1109/CVPRW.2010.5543262.
[49] A. Dhall, R. Goecke, S. Lucey, T. Gedeon, Static facial expression analysis in tough conditions: data, evaluation protocol and benchmark, in: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011, pp. 2106–2112, https://doi.org/10.1109/ICCVW.2011.6130508.
[50] Q. Cao, L. Shen, W. Xie, O. Parkhi, A. Zisserman, Vggface2: A dataset for recognising faces across pose and age, 2018, pp. 67–74, https://doi.org/10.1109/FG.2018.00020.
[51] Y. Li, J. Zeng, S. Shan, X. Chen, Occlusion aware facial expression recognition using cnn with attention mechanism, IEEE Trans. Image Process. 28 (5) (2019) 2439–2450, https://doi.org/10.1109/TIP.2018.2886767.
[52] D.E. King, Max-margin object detection, ArXiv abs/1502.00046, 2015. URL, https://api.semanticscholar.org/CorpusID:15906014.