Title: A survey on facial emotion recognition techniques: A state-of-the-art literature review

Outline

Figures (13)

Tables (3)

Information Sciences
Volume 582, January 2022, Pages 593-617
A survey on facial emotion recognition techniques: A state-of-the-art literature review
Author links open overlay panelFelipe Zago Canal a, Tobias Rossi Müller a, Jhennifer Cristine Matias a, Gustavo Gino Scotton a, Antonio Reis de Sa Junior c, Eliane Pozzebon a b, Antonio Carlos Sobieranski a b
Show more

https://doi.org/10.1016/j.ins.2021.10.005Get rights and content
Abstract

In this survey, a systematic literature review of the state-of-the-art on emotion expression recognition from facial images is presented. The paper has as main objective arise the most commonly used strategies employed to interpret and recognize facial emotion expressions, published over the past few years. For this purpose, a total of 51 papers were analyzed over the literature totaling 94 distinct methods, collected from well-established scientific databases (ACM Digital Library, IEEE Xplore, Science Direct and Scopus), whose works were categorized according to its main construction concept. From the analyzed works, it was possible to categorize them into two main trends: classical and those approaches specifically designed by the use of neural networks. The obtained statistical analysis demonstrated a marginally better recognition precision for the classical approaches when faced to neural networks counterpart, but with a reduced capacity of generalization. Additionally, the present study verified the most popular datasets for facial expression and emotion recognition showing the pros and cons each and, thereby, demonstrating a real demand for reliable data-sources regarding artificial and natural experimental environments.
* Previous article in issue
* Next article in issue
Keywords
Emotion Recognition
Facial emotion recognition
Pattern recognition
Systematic literature review
1\. Introduction
Facial Emotion Recognition performed computationally is a very interesting and challenging task to be explored. Besides interpreting facial emotion expression being a task naturally performed by humans, finding computational mechanisms to reproduce it in the same or similar way is still an unsolved problem [8]. Designing and developing algorithmic solutions able to interpret facial emotions from human faces opens a new window of possibilities for the human?computer interaction context, such as in robotics, gaming, digital marketing, intelligent tutor systems, among many others [59]. The sooner we are able to design such recognizers, the better we can help to understand natural areas of psychology, neuroscience, human cognition and learning [76]. Nonetheless, how human expression behavior is invariant among distinct individuals, and how biological and social aspects may interferes in human communication over time, are interesting questions to be studied and modeled computationally, as well as its correlation.
In order to provide analytical models for human expression recognition from facial images, engineers, mathematicians and computer scientists are exploring distinct ways to reproduce approaches able to effectively implement efficient algorithms. There are strong correlations among image processing, computer vision, pattern recognition and artificial intelligence fields exploring this topic, and interesting approaches can be verified over the literature. In general, specifically due to the recent achievements in computational processing power and new architectures for high-performance computing, applications for the aforementioned fields, previously restricted by computational limitations, have had their solution achieved [12]. With the advent of the computing graphic processing units (GPU?s), many methods were adapted for their parallel version, having execution time close to real-time [102]. Specifically for the emotion recognition problem from facial images, those computational improvements have also provided a new set of tools, paving the way for the development of several approaches based on what we named here

as classical methods1.
Another important achievement was obtained over the past few years with the advent of variants of neural networks ? the Convolutional Neural Networks (CNN?s). CNN?s have being applied as generic problem-solvers, where some input signal is decomposed (de-convolved) into a set of invariant-features, providing robust mechanisms to extract relevant features (such as texture, corners and key-points). After a training step, the classifier is ready to interpret image with zero-bias in a very effective way. Consequently, CNN is ascending lately in many areas, mainly the ones that involve image processing and pattern recognition, such as those used in medical image classification [65], [84], [7].
Over the literature it is possible to find many works fully dedicated to solve the emotion recognition problem from facial images, using classical methods or Neural Network based (NNB) approaches [27], [43], [50], [109]. The graph of the Fig. 1 shows a comparison of published approaches for emotion recognition from facial expressions, obtained from our literature review. For the selected papers, two main trends can be verified: (i) classical approaches, being composed of traditional image processing, pattern recognition and miscellaneous classifiers, more specifically provided by a hand-crafted feature design process, and (ii), NNB approaches, where classical neural networks and its convolutional counterpart were considered, whose features are learned from data using generic feature extractions, more specifically for the convolutional case. One can see a considerable acceptance of NNB approaches over the past five years, more specifically justified by the emergence of CNN?s approaches.

1. Download: Download high-res image (80KB)
2. Download: Download full-size image

Fig. 1. Number of methods related to facial emotion recognition since 2005 in our database. Approaches were classified into traditional image processing or computer vision approaches (Classic) and neural networks based (NNB) techniques. Only methods that had mentioned the classification algorithm, dataset and results clearly were counted.
Given the large number of classical methods, which still can be combined among them producing hybrid designs, and the emerging CNN?s approaches, a new literature review for the area of expression and emotion recognition from facial images is needed. In order to achieve this goal, in this paper, a systematic literature review was conducted to select the published works whose inclusion and exclusion criterion matched, and a total of 51 were obtained, resulting in 94 distinct approaches. Based on the selected papers, the objective of this work was to provide answers to the following questions: (i) what is the state-of-the-art in emotion recognition, how many methods, techniques or classes of methods can be identified over the literature? (ii) what is the importance of the CNN?s as a general trend for the facial emotion recognition problem? (iii) is there a general computational workflow for facial emotion recognition that can be identified? (iv) how many databases for facial emotion recognition can be identified, containing the number of individuals and expressions? how is their acceptance in the scientific community? (v) is it possible to predict further trends to solve the problem of emotion recognition from facial images, and how far from a generic effective solution are we?.
Given the importance of the aforementioned research questions, the remainder of this work is organized as follows: Section 2 provides the systematic literature review methodology used in this survey, describing in details the inclusion and exclusion criterion employed. Section 3 presents a general

computational flow we are able to identify as commonly used approach to solve the problem of emotion recognition from facial images or videos. Section 4 presents the most frequently used preliminary step found in the verified works, being defined by a preprocessing step used to locate the region of interest, as well as to perform image improvements. The process of extracting relevant information from the images, as known as feature extraction, is briefly discussed in Section 5. Datasets publicly available to the scientific community used to train the computational models are described in Section 6. The most used classification algorithms are presented in Section 7. The computational approaches and techniques used for emotion recognition from facial images or videos are described in Section 8. Finally, conclusions, discussions and further directions identified in this study are presented in Section 9.

2. Methodology for literature review

Over the past few years it is possible to find interesting literature review papers summarizing the most prominent works of the current state-of-the-art in the FER area. In general, journal papers such as [56], [63], [39], [79] present a very detailed description of many facial expression or emotion recognition systems, highlighting aspects of image or video acquisition, post-processing, feature extraction and recognition, as well as available datasets, essential sources to properly train a FER recognition system. Not less important, there are other papers majorly published in conferences, whose FER technologies are also presented [80], [103], [81], [67], focusing their discussion to traditional methods, or to the recently emerged convolutional neural network counterpart. Also, it is possible to find interesting literature reviews describing the combined use of FER with secondary information, such as the use of compound facial expression emotions [103], the combined use of augmented reality [79], gender and emotion classification using facial expression detection [32], and a comparison of methods specifically designed for genuine versus posed facial expressions of emotions [44].

In the literature review presented in this paper we intend to analyze the state-of-the-art in the area of emotion expression recognition captured from facial images or videos, based on the methodology described by Kitchenham [52], a well-established methodology used to provide systematic literature reviews in a verifiable and reliable manner. It is intended to provide a general panorama of the results achieved by computational facial emotion recognition, employing the most different approaches, yet mainly focuses on analyzing classical and NNB works, since those methods seems to be arising over the past few years. Classical approaches while effective, may be very dependent of the execution parameters and environment conditions to achieve efficient results. On the other hand, NNB approaches can be generalized to solve a series of problems, and have been used in the past few years to clarify the issue of facial emotion recognition as well. We will analyze the existing methods regarding in which emotion types they are applied; whether they are adaptive regarding environment changes; and whether they are able to distinguish possible faults or changes in terms of emotions; datasets and the capability to distinguish between different kind of expressions.

The systematic literature review for emotion expression recognition aims to postulate research questions as presented below:

**Question** : What is the state-of-the-art in computerized approaches for emotion recognition, obtained from facial images or videos, published over the past years?

**Population** : Related works describing a computational approach for

expression emotion recognition from facial images or videos, obtained from raw data or well-established datasets publicly available.
**Intervention** : Analyze the state-of-the-art in computational emotion recognition.
**Results** : Comparison among computational expression emotion recognition approaches and datasets.
**Context** : Papers found in digital libraries such as: _Science Direct_ , _IEEEXplore_ , _Scopus_ e _ACM Digital Library_.
The databases for literature review were chosen by taking into account their relevance to the areas of image processing, computer vision, pattern recognition and artificial intelligence, more specifically those that employed the problem domain previously mentioned. Therefore, the following databases were used: ScienceDirect, IEEE Xplore Digital Library, Scopus and ACM Digital Library: (i) ScienceDirect is a platform with a considerable amount of scientific papers, including high impact factor journals, along with the area of computer vision and pattern recognition. According to the website, it hosts around 2,500 journals and more than 39,000 books until the present time; (ii) IEEE Xplore digital library is a database of scientific and technical content, including journals, conferences and books. It?s supported by the IEEE (Institute of Electrical and Electronics Engineers) and it?s publishing partners. Accordingly to the maintainer, it is composed by more than four-million full-text documents, including 195 + journals and 1,800 + conference proceedings; (iii) Scopus is an abstract and citation database that contains peer-reviewer literature such as journals, books and conferences proceedings. It contains researches from the fields of science, technology, medicine, social sciences, and arts and humanities; (iv) ACM Digital Library is a database of full-text papers with focus in the computational field, including journals, conferences and books, maintained by the Association for Computing Machinery.
The access to the databases was performed through the CAPES Portal2. To avoid ambiguity and redundancy among the literature, the databases used were limited exclusively to the aforementioned sources, and other sources are out of the scope of this work.
2.1. Search definition
Search definitions used to support our manuscript are presented according to the Table 1. Firstly, the four databases used to initially select relevant papers is shown at the first row, as well as the search terms corresponding to facial expression recognition and facial emotion recognition. Inclusion and exclusion criterion are presented in the second and third row, respectively. These criterion are required to provide a second filter to the literature review methodology, describing particular desired aspects to be presented in the articles.
Table 1. Facial emotion - Search definition for our Literature Review.
Search Bases	**Search Terms**
\- ACM Digital Library;
\- IEEE Xplore;
\- Science Direct;
\- Scopus.| _* facial expression recognition;_
** facial emotion recognition;
Inclusion criteria
\- Papers written in English language;
\- Papers published between 2006 and 2019;
\- Papers that present a facial expression/emotion recognition approach;

\- Works with face images or sequence of images approaches to detect emotion;
Exclusion Criteria
\- Short papers, such as abstracts or expanded abstracts;
2.2. Search execution
After the initial selection of articles, a total of 117 manuscripts were filtered. These articles were elected according to their title, abstracts, keywords and images, as described by the literature review methodology. The second filter applied over the articles was based on analyzing their entire structure, which means, the documents were separated according to their contents. Articles with no significant information or lack of important data were discarded at this stage. Besides that, while reading carefully all the articles, we were able to classify them in terms of employed technologies, year of publication, dataset properties, considered expressions and result?s accuracy. After this final stage, 51 articles remained. Since some of these 51 studies propose not a single method, but multiple solutions, a total of 94 methods where analyzed (Table 3).
3\. General computational flow for facial emotion recognition
In our work, the literature review has revealed interesting aspects indicating for a general computational flow, which can be applied to solve the challenge of recognizing human facial emotion from images or videos. This flow can be illustrated in general terms according to the Fig. 1 as follows: (i) image or data acquisition step: input images or video sequences are used from well-established datasets or from proper image acquisition systems and environmental sets, using bi-dimensional signals or combined with tree-dimensional acquisition devices, such as kinect or sterescopic cameras; (ii) image preparation (preprocessing) step: performed in order to improve image quality, reduce noise, reduce dimensionality in terms of spatial resolution, or by the simplification of color information into gray scale one; (iii) feature selection/extraction step: aims to extract relevant features from the input signal in order to discretize the input signal and make them unique for each individual ? this step behaves like a signature or fingerprint for each specific facial emotion; (iv) classification step: the previously obtained features are used in some classification model, able to provide a similarity or dissimilarity score for facial emotions, ranked according to the number of recognizable expressions in the dataset; (v) results and validation procedure: this step is often required since several models are based on a training procedure, and a validation procedure is carried in order to compute a general precision overall for the proposed approach. This last step is usually performed against a well-known ground-truth information, provided by the dataset and used to train and validate the computational recognition model proposed.
The aforementioned steps illustrated by the Fig. 2 can be identified partially or even totally for a large number of approaches. For each step block, the algorithms used may vary according to the particularity of the problem approached, as well as characteristics of the dataset used, and they can be replaced in a tailored manner by another block method independently. In the next sections the steps cited will be explained in details, providing a general overview of the approaches proposed over the literature and its main constructive algorithmic steps.
1. Download: Download high-res image (19KB)
2. Download: Download full-size image
Fig. 2. General Computational Flow revealed from the literature review, often applied as a general problem-solver for the emotion recognition problem using images or videos as input signals.

4\. Preprocessing
Preprocessing step can be ignored for some approaches according to the dataset quality and whether previous preparation was done or not. For some cases, specially when a dataset containing raw images is used, preprocessing is a very important operation [49], [68], which may be responsible for the success or failure of a recognition system. Preprocessing has the objective to slightly suppress some imperfections such as noise, reduce distortions and highlight only the most important features for the data being analyzed.
4.1. Grayscale conversion
Besides gray scale conversion being a simple task, it has been demonstrated that for some areas such as face recognition and emotion recognition color information can be suppressed into a single dimension with no significant loss of precision [76]. Additionally, processing color inputs still provides an increment of computation power in the order of 3 times when compared to gray scale images at least until (iii)-feature extraction step, where information is usually invariant from color. On the other hand, nowadays almost every single digital camera captures images and store them into proprietary format or Bayer-8 (which in practice is uni-dimensional), and them convert it to color information afterwards (usually RGB) [100].
Over the literature it can be observed that reducing the input signal from color to gray scale is very common and effective, and at the same time preserving the relevant features and mitigate performance issues for the emotion recognition problem. The most common operation is the simply summation and normalization of the resulting unidimensional signal. Other approaches use the normalized grayscale conversion formula defined by the weights 0.2989,0.5870,0.1140, for red, green and blue channels, respectively. There are other approaches where a specific channel is taken into account as the dominant information thus generating a single channel image from it.
4.2. Face detection
The most relevant stage in recognizing facial expressions and emotions is the initial selection of the region of interest (ROI) in the input image [28].
There are some effective algorithmic approaches able to find the region of interest (the face region) despite the problematic perturbation that may affects their performance. Normally, these kind of algorithm has to get around some variations in the images, such as pose and illumination [5]. The variations between different faces (subjects), in general, are less impacting than environment ones, so some of these problems are usually solved with use of preprocessing techniques.
Generally, raw images provided by most datasets, have a lot of background information that has no use for expression and emotion classification. Therefore, the most common preprocessing technique is to detect the subject?s face and crop the area around it, discarding useless information and even increasing its accuracy. In the context proposed by this work, the most important preprocessing step is to detect a face inside the image that is being analyzed. In this section, we describe some detection algorithms and how they work with images to achieve what they are designed for.
4.2.1. Viola-Jones
The algorithm developed by Paul Viola and Michael J. Jones in 2003 had a great impact in the face detection area, because besides its great precision, it has low computational cost and also allows to visualize the results. Although it is still widely used, it needs a training step that can be slow, but after that training, face detection is fast.
According to the authors [46], this technique is based on the characteristics of Haar-like (used for pedestrian detection), which is given by the summation

of pixels that are within the area of white rectangles and then subtracted from the area of the gray rectangles. Although there are better filters, the advantage offered by Haar over the offered efficiency turns out to be advantageous according to [105], as shown in the Fig. 3-(a).

1. Download: Download high-res image (43KB)
2. Download: Download full-size image

Fig. 3. The most common approaches for facial localization. Viola Jones uses rectangular Haar-like features as show in (a). Haar classifier rectangular blocks are illustrated in (b).

The Viola-Jones detector performs face localization and has its constructive algorithmic based in four steps:

* 1.
Selection of rectangular Haar-Like Features as demonstrated in the Fig. 3-(a);
* 2.
Creation of the integral image that allows quick calculations of the previous mentioned Haar characteristics;
* 3.
Training of an AdaBoost-based classifier capable of selecting relevant characteristics;
* 4.
Using cascade classifiers that rule out regions where a face is unlikely to exist, and focus on regions likely to contain a face.

4.2.2. Haar cascade

The Haar Cascade Classifier is an object detection method that inserts Haar resources into a series of classifiers to identify objects in an image. Taking this into account, Haar?s features are based on Haar waves, rather than the usual image intensities. According to [92], this method was adopted and developed by Viola and Jones [105], despite having been proposed by [74]. The set of alternative resources was born due to the computational cost of calculating resources in an image with RBG pixels.

A Haar-like feature considers adjacent rectangular regions at a specific location in the detection window, sums up the pixel intensities in each region and calculates the difference between these sums. The difference calculated can be used then to identify subsections in the image. Haar classifiers must be trained by giving lots of images (positive and negative in therms of containing the object to be detected). With that done, features can be extracted using sliding windows of blocks, as shown in Fig. 3-(b). These blocks are almost like convolution matrices that are applied in different portions of the picture to find partial matches.

In sequence, even being able to fast calculate the sum intensities using integral images, there are lots of features extracted from the previous steps that do not add any value to the result. If a feature is detected in a region that is not of interest, it has no real value to the classifier. For that, a boosting technique is applied in Haar Cascade classifier, called AdaBoost [25]. These technique reduces the number of classifiers significantly, causing no harm to the end result.

The Haar Cascade Classifier is composed by the following stages:

* 1.
A section of the image is selected as working region;
* 2.
The first stage evaluates the presence of a feature as previously detailed;
* 3.
If the previous step?s return has a positive signal, the classifier jumps to the next stage and that goes until the last stage. On the other hand, if the

previous step?s return was a negative signal, another section of the image is selected and the process restarts;
 * 4.
Finally, if the image was entirely evaluated, the classifiers ends its execution.
The cascade process aforementioned is illustrated by the Fig. 4. Besides this approach is also used for several methods for object recognition and detection, it has been shown very effective for the problem of facial localization with a very good response in terms of execution time.
 1. Download: Download high-res image (40KB)
 2. Download: Download full-size image
Fig. 4. Haar classifier stages used for facial localization.
4.2.3. Convolutional neural network
A Convolutional Neural Networks (CNN) are nowadays one of the main categories when mentioning general problem-solvers for data mapping, image recognition and classification, object detection and, clearly, face localization/recognition. The deriving classifications provided by CNN [1] occur when an image is taken and classifying it the following way: the method sees the input image as a pixel matrix; this matrix will go through a series of convolution filters, pooling and functions for example softmax, that classifies an object with probabilistic values between 0 and 1; this way, with a trained network, it is possible to utilize those functions to classify images never ?seen? by the network before (zero bias) with a considerably well accuracy. Section 7.2 will provide further details regarding the functioning of a convolutional neural network. One CNN architecture that became very popular due to high performance in face detection was the Multi-Task Cascaded Convolutional Neural Network (MTCNN) [118]. Experiments performed on benchmarks datasets of facial detection in 2016 demonstrated a performance superior to all other algorithms tested so far. Several implementations of this architecture are open for use through python libraries, further popularizing the algorithm.
4.2.4. Miscellaneous
Beyond the aforementioned methods, there are still plenty of other approaches that could be potentially used to detect faces. Some of them are based on classic computer vision approaches (such as image segmentation, among others), but what is remarkable is that these technologies, besides obsolete nowadays, were essential to the development of the current technologies. Some of these miscellaneous technologies are describe below: At [116], three methods were used simultaneously: joint cascade detection and alignment (JDA) that has satisfactory performance with frontal detection, whereas the Deep Convolution Neural Network (DCNN) has good performances with non-frontal faces and, finally, a Mixture of Trees (MoT) took place in case the other two detectors fail. Another procedure is adopted by [43] where two combined methods for face detection are applied, HOG and a linear classifier. To run the recognizer in real time, the face location is initiated by the face detector with the first frame and uses the correlation tracker to track the face through the other frames.
4.3. Dimensionality reduction
Attempting to reduce the amount of data to work on, most of the researchers opt to apply some technique to reduce images dimensions and, therefore, reduce time consumption and lower requirement for processing power. The literature review clearly demonstrate that one of the most common and simple approach, is the down sampling itself. Down sampling is, in fact, a downscale of images with geometric transformation that is done without losing significant image

features for the recognition procedure. Scripts with this purpose are also very popular and enables researchers to process their images in machines with memory limitations.

Linear Discriminant Analysis (LDA) is a technique very cited on the articles reviewed by this survey. The main idea of LDA is to project some set of axis on a new and smaller set of axis in such away that the separation between the labels of the data is more evident. To do that, LDA uses two parameters, the mean of each label on the new set of axis, and the variation of each label on it. The algorithm maximizes the distance between the means and minimizes the variation for each label.

Another widely known and accepted by the scientific community approach is the use of Principal Component Analysis (PCA). PCA is a simple, non-parametric method that intends to isolate relevant information from a large set of input variables, using linear algebra?s analytical solutions to isolate the irrelevant feature vectors and eliminate them since there is no significance to keep them into account for the recognition procedure. With low effort, PCA is capable of provide a roadmap for reducing a complex data set into simplified structures that contain the important piece of information needed [87].

Performing a brief analysis of the methods presented in the articles it can be observed that among the ones that cite dimensionality reduction algorithms, 44% used PCA, 25% LDA and 25% Down Sampling. Only 27 methods out of the 94 analyzed mentioned explicitly if they used any dimensionality reduction algorithm. An interesting fact found was that 77.78% of the uses of down sampling were for methods that used neural networks as classifiers and 87.5% of the methods that used the PCA algorithm had some classic algorithm as a classifier.

5\. Feature Extraction

Feature extraction is a very important stage in recognition methods, being mainly applied when each sample has a large amount of data. It has as main goal to extract only the most important and descriptive piece of information from the samples, getting rid of what is not relevant for the given problem. This is usually necessary because of the high computational complexity of training in classification models and due to the fact that the more data there is in each sample, the more computing is needed to achieve good results on these models. In general, it is very important that the extracted characteristics are simple and independent from each other, that is, there is not such a high correlation index between the sets of characteristics. Some classical algorithms for extracting features from images are Active Shape Model (ASM), Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG). There are many variations of these algorithms being used on the Facial Emotion Recognition task to achieve distinct results [91]. In the last few years there can be observed some works that used Convolutional Neural Networks to extract features as well [4], [106].

The ASM algorithm was proposed in 1995 [14] with the intent of generating points that fit the contour of the object. Both LBP and HOG work splitting the image in small pieces called cells and generating histograms. LBP was proposed in 1994 by [72] and functions making comparisons between each cell?s pixels and their eight neighbors to build a binary number. Afterwards, the algorithm frames a histogram of this numbers for each cell and concatenates the histograms. More information about LBP can be found in [9]. The HOG algorithm, on the other hand, instead of making comparison between pixels, makes a histogram from the gradients previously calculated from them [17].

There are also feature extraction attempts based on the Facial Action Coding

System (FACS) [26]. FACS is a coding system created in 1976 by Paul Ekman and Wallace Friesen that measures the contraction and relaxation of facial muscles with degrees of intensity. It has undergone some changes over time making it more robust [24]. In general, works that try to find facial emotion using FACS use specific dataset to train neural networks that have facial images cataloged by FACS experts, so the feature extraction process is not computational but human.

The FACS analyzes the emotions in the face, which are classified through the movement of the facial musculature, which is coded by the system through AU - Action Units, AD - Action Descriptors and M - Movements. Each AU, AD or M corresponds to a code used to describe the specific movement of one or more muscles of the face, eye direction or head movement. Fig. 5 shows the AU classification.3

1. Download: Download high-res image (362KB)
2. Download: Download full-size image

Fig. 5. Facial Action Coding System (FACS).

6\. Datasets for emotion recognition from images or videos

Datasets containing facial images or videos specifically developed for emotion recognition are not a novelty. The first studies are related to [82], and since then the need to develop computational approaches to inspect emotion or behavior in face images or videos has been significantly growing. As a consequence, many facial expression datasets have emerged, varying in acquisition environment, number of recognizable expressions, regions, among other features [73]. On the other hand, due to the way some of these datasets were created, many other points were also included over time such as gender, ethnicity, age, image quality (size, color, older datasets performed a digitization from physical photos), and number of participants. All these variables tend to influence the quality of the chosen dataset, as well as impact the results that computational approaches for emotion recognition output.

Since datasets play a fundamental role for data-driven learning recognition based on emotion recognition [86], this section aims to present a review of the most important datasets for emotion recognition in images or videos, recognized and widely used by the academical society. Additionally, here we pointed relevant implementation details used for the construction of those datasets, as well as its particularities an possible drawbacks. Fig. 6 shows the relationship between the performance of the reviewed methods and the datasets in which they were tested.

1. Download: Download high-res image (38KB)
2. Download: Download full-size image

Fig. 6. This plot was generated using only classes with more than 3 members. The x axis represents the datasets used for testing the methods and the y axis represents the accuracy of the methods. Each colored dot represents one benchmarked method. It is important to emphasize that some of the approaches considered for this study may have applied the same method to multiple datasets.

6.1. JAFFE

The Japanese Female Facial Expression (JAFFE) [62] is a freely available dataset, published in 1998 by a group of Japanese researchers from the ATR Human Information Processing Research Laboratory and the Psychology Department of Kyushu University. The dataset features images of 10 Japanese women, expressing the following emotions: happiness, sadness, surprise, anger, disgust, fear and neutral. Each participant took around 3 to 4 pictures of herself for each emotion while looking through a semi-reflective plastic sheet towards the camera, totaling 219 images. The camera was also surrounded by a black box to prevent and mitigate light reflections. The hair was removed from the front of the face in a manner that the expressions were more evident. Because the photographic process was performed analogously, the photos were digitized afterward. The final digital resolution for each image is 256x256 pixels. Images were rated for each emotion by a group of 92 Japanese female undergraduate on a 5-point scale. Each image has a 5 or 6 component vector representing the average ratings for each emotion. Besides not recent, this dataset has been used for several emotion recognition approaches, such as [64], [3], [8].

6.2. Cohn-Kanade

Cohn-Kanade AU-Coded Expression Database (CK) [48] is a well-established dataset, being widely used in the literature since it?s first version [112], [94], [10], [57], [41]. This dataset was originally created with the name CMU-Pittsburgh AU-Coded Facial Expression Image Database, and is being managed ever since, by the (AAG) Affect Analysis Group of the Research Lab at the University of Pittsburgh. Back in the day, the first version of CK dataset was developed to fill the gap of large enough facial images sets to be able to support studies of facial feature tracking and analysis methods. Nowadays, the database counts with 2 versions: the first one was released in 2000 and includes 486 sequences about 97 different posers. From the subjects, there were 69% female, 31% male, 81% Euro-American, 13% Afro-American, and 6% other groups.

The CK is composed by sequences that varies from neutral to a peak expression, being the peak expression image fully FACS as presented in the previous sections, coded and given an emotion label. However, this emotion label refers to the expression that was requested, but not necessarily the one actually expressed by the participant. In the first version of the database, images were taken in a controlled room in terms of lights, background and camera position. Besides that, all expressions were collected by asking subjects to perform them, therefore, they may differ from spontaneously taken images. The CK dataset provides 2D images with either 640x490 or 640x480 resolution and pixel array with 8-bit gray-scale or 24-bit color values.

The second and the current latest version of the database was named The Extended Cohn-Kanade Dataset (CK+) [60] and released in 2010. For this version, there were included 107 sequences of emotion transitions and 26 subjects. Besides that, non-posed images were taken this time and each image received a nominal emotion label based on the subject?s impression of each of the 7 basic emotion categories, that is, Anger, Contempt, Disgust, Fear, Happy, Sadness and Surprise. The authors (The Affect Analysis Group) is still working in what should be the third version of this dataset, whose intention is to mainly add synchronized 30-degree images from the original frontal video, enabling a series of analysis such as 3D facial and emotion recognition.

6.3. MMI

MMI is a dataset created from 25 people of different ethnicities (European, Asian and South American), being 44% women and 56% men, with their age ranged from 19 to 62 years old. The expressions were recorded/photographed in a natural way, which means participants were induced to perform expressions based on videos used to stimulate (comedy videos and disgusting content) a most reliable natural condition. Papers [66], [16], [36] worked with this dataset, and one of the points analyzed by one of authors was about the use of glasses and the difficulty that ended up being high for the recognition of emotions.

Opposite to other databases that existed at the time when only 6 basic expressions were parsed, MMI has many non-basic expressions and not defined expressions, since they were performed during an expression exchange. The images in the database are all real colors but after a digitization procedure its resolution was downgraded to 720x576 pixels, and the videos were recorded at 24 frames per second (sequences range from 40 to 520 frames), ranging from neutral, expression and neutral again. In total, the dataset has one hour and thirty-two minutes of content. The backgrounds on which the photos and videos were captured is not the same for every sample. The dataset is accessible online and easily searchable, and is freely available to the scientific community.

6.4. FER2013

The FER2013 database was originally published in the International Conference on Machine Learning (ICML in 2013) [31]. This set of pictures is publicly open and was created for a project by Pierre-Luc Carrier and Aaron Courville. It was publicly shared to be used in a facial expression recognition competition in Kaggle, shortly named as ICML.

This dataset consists of 35.887 pictures of faces with 48x48 pixels in grayscale, all images are labeled in seven facial expressions and distributed as follows: 4953 angry images, 547 disgust images, 5121 fear images, 8989 happy images, 6077 sad images, 4002 surprise images and 6198 neutral images. The dataset was developed using the Google image search API in which was searched for face images that are within a set of 184 keywords related to emotion, such as ?happy?, ?sad?, and others.

During the Kaggle?s competition, 28709 images from this dataset were shared between the participants to train their neural network and 3589 images were used to the test set and validation to figure out the winning recognition algorithm on the competition. After the competition?s ending, the dataset were made accessible to the general public. Besides the large number of images, this dataset has a very reduced spatial resolution since its purpose is to be used as input to train and test computational classifiers.

6.5. BU-3DFE

Binghamton University 3D Facial Expression (BU-3DFE) [114] was designed to meet the need for a dataset with emotion-classified 3D facial images, thus allowing its use for the facial expression recognition area. In 2006, the publication date of the article ?The 3D Facial Expression Database for Facial Behavior Research? [114] did not detect any other dataset that fulfilled this role. This dataset is free for academic use. Images are linked to the emotional states anger, disgusting, fear, happiness, sadness, surprise and neutral. In addition, there is a degree of spontaneity associated with each image. A system with 6 cameras positioned at different angles was used to capture the images. The stereo photogrammetry technique was used to construct the three-dimensional face. A total of 100 participants had their expressions scanned, being 60 percent woman and 40 percent man, totalling 2500 3D models. In addition, all participants were part of the psychology, arts and engineering departments. Participants were asked to demonstrate the 7 emotions with 4 degrees of intensity, low, middle, high and highest. There are subsequent studies to BU-3DFE, such as BU-4DFE in which the time dimension was added, BP4D-Spontaneous and BP4D +.

6.6. Miscellaneous

Besides the databases presented above, there are some datasets that are worth mentioning, such as CASIA Webface [113], FEEDB [95], RaFD [54], National Institute of Mental Health Child Emotional Faces Picture Set (NIMH-ChEFS) [21], Taiwanese Facial Expression Image Database [11], Acted Facial

Expressions In The Wild [20].
As a public large dataset composed from images collected from the web, [113] contains almost five hundred thousand images from more than ten thousand subjects. [95] is a dataset of video sequences, recorded using Microsoft Kinect sensor and it was developed to serve as a dataset to be applied in facial recognition and facial expression/emotions studies. [54] is a initiative from the Behavioural Science Institute of the Radboud University Nijmegen. It contains 8 different emotions from 67 models and it is freely available for non-commercial scientific research by researchers who work for an officially accredited university. The [21] is a little bit different from the datasets presented above because it is composed by images of children only. It contains 482 high quality colored images about five emotions and of direct and averted gaze. [11] was developed by Brain Mapping Laboratory (National Yang-Ming University) and Integrated Brain Research Unit (Taipei Veterans General Hospital) and has 8 different kind so emotions from 40 models. The dataset is opened for scientific researches only. The [20] dataset proposes a face image extraction from movies, that is, a out of the scope of lab controlled datasets. The authors compare their dataset with [60], [62].
Over the internet, there are a lot of datasets with distinct characteristics available, most of the time for free. Even though there is such a great number of datasets, sometimes they are not exactly what researches are looking for. Therefore, likewise all datasets cited, a great quantity of projects on facial expression/emotion recognition prefer to use their own dataset, not being possible to the public to use it.

7\. Classification algorithms

Besides all operations being of extreme importance to accomplish good results when classifying emotions from face expressions, usually the most valued stage in the whole process is the classification. This step is also the one with most variations between works over the literature. As mentioned on Section 1, there have been some arising methods based on neural network architectures over the past few years, mainly because of the ascension of computer capabilities. Taken the 51 works selected for this scope, we divided them into two distinct groups: Classic and NNB approaches.
The first group refers to classical techniques used to categorize facial emotion expressions, that is, traditional image processing, computer vision, pattern recognition and misc classifiers, while the second one is composed exclusively by neural network based approaches, including the CNN arising methods.
Both groups will be presented with more details in the next subsections, such as the algorithms included in each one of them.

7.1. Classical approaches

For many years, the classical approaches for image processing in general, were the best attempt to solve some problems with high processing cost. Over time, with the development of computer capability and the advent of new architectures, new methods took place for multiple reasons, such as implementation complexity and achievement of great results. However, some methods (that we call classical in this paper) have achieved great results and are still widely used nowadays. From the chosen papers for this literature review, the classical approaches usage is shown in the Fig. 7. In the graph it is possible to see that Support Vector Machine (SVM) is the flagship when talking about classical classification methods, corresponding to 40% of the approaches. The second mostly applied method is the Dynamic Bayesian Network with only 10% of the appliances. The 13,3% corresponding to ?Others? is composed by algorithms that were applied only once over the 51 selected articles. These methods are: extreme learning machine [3], extreme sparse learning [88], support vector regression [120], sparse representation classification [124], random forest [123], random tree [123], many graph embedding [45] and a single modified Viola-Jones [51].

1. Download: Download high-res image (54KB)
2. Download: Download full-size image

Fig. 7. Scientific share for classical approaches verified for this literature review.

7.1.1. Support Vector Machines (SVM)

As it is possible to visualize in Fig. 7 that support vector machines (SVMs) are by far the most applied classical method to classify emotions from face images. SVMs are supervised learning methods, that is, algorithms capable of generating input?output mapping functions from a specific set of labeled data with one or more feature vectors. SVMs classify elements by determining boundaries, known as hyperplane, that separate the classes from each other. This hyperplane can be oriented in different ways and still fulfill it?s purpose, however, SVM goal is to orient it in such a way that it is as far as possible from the closest vector (called support vector) from each class [38]. Initially, these machine learning techniques were developed in 1963 to perform classification in linearly separable sets of data [104], however, with the assistance of nonlinear kernel functions, it is possible to transform input data to a high-dimensional feature space in which the input data become linearly separable and, therefore, classifiable by the SVM algorithm [108]. Because of it?s simplicity in therms of processing power needed (in comparison with NNBs), SVM is a great tool for machine learning and, with the help of nonlinear kernels, is capable of competing with the newest and most complex methods in therms of results. Besides the face/emotion recognition, SVMs have been used to solve classification of all sources over the past few years, including the medical area and many others [122].

7.1.2. Dynamic Bayesian Network

The Dynamic Bayesian Networks (DBNs) are Bayesian networks (BN) for processes with changing environments. Therefore, to understand what a DBN is, it is necessary to understand a BN. Thus, according to [89], Bayesian networks (BNs) are directed acyclic graphs, which represent the probabilistic relationships between a large number of variables. It is important to note that they work with incomplete knowledge and allow machines to make inferences, predictions and make decisions based on the probability of variables that assume some states, solving the problem.

A DBN or temporal Bayesian network, as well as BNs, must be acyclic and targeted. The difference is that DBN is a way of extending the BN to distribute the probabilities across infinite collections of random variables. Fig. 8 shows a DBN.

1. Download: Download high-res image (66KB)
2. Download: Download full-size image

Fig. 8. Dynamic Bayesian Network.

7.1.3. Fuzzy Logic

The fuzzy logic concept was born in 1965 with the proposal of the fuzzy set theory by Lofti Zadeh [117] with the objective of natural language processing. The goal of fuzzy logic in general, is to change the very basis of boolean computer processing, being able to produce results based on degrees of truth instead of 1 or 0. In other words, it is based on the idea that people make decisions over non-numerical information and, therefore, fuzzy is able to imitate these kind of decision making that sometimes is not absolute. For face emotion recognition, the fundamental operation is the selection of the most important features to be analyzed (usually made by an specialist or based on FACS) and the attribution of weights for each of these features. Besides that, the human involved has to build the fuzzy rules base, which is responsible for the evaluation of each sample and, therefore, each feature value may be classified into the best matching set. The overall process of a fuzzy system is presented in Fig. 9. Firstly, the raw input need to be fuzzyfied in order to have truth based values instead of boolean ones, as discussed before. The fuzzification process is done for every feature that is wanted to have some weight in the decision making. In the inference step, each feature is analyzed based on a previously determined set of rules with antecedents and consequences. The rules set is the very intelligence in a fuzzy set thus the determinative factor of a fuzzy system?s accuracy. Finally, the fuzzy output needs to be defuzzified in order to serve as an applicable action for the computer. Fuzzy logic have been used since it?s beginning mainly for control systems for the most different areas [101].

1. Download: Download high-res image (33KB)
2. Download: Download full-size image

Fig. 9. Schematic representation of a Fuzzy system to solve the recognition problem.

7.1.4. Miscellaneous

As showed in Fig. 7, some methods different from the ones presented above, were applied over the articles selected for these review. Since classical approaches, for this paper, means every single method that is not based on artificial neural network, there are a lot of particular methods that were applied by less than four works and will not be detailed in this document. The methods with unique occurrences were already mentioned in Section 7.1 and besides them, some methods with three article usages were plotted in the graph, they were: Adaboost, Naive Bayes (NB), K-Nearest Neighbors (KNN), Hidden Markov Model, Euclidean Distance and Decision Tree. Adaboost, as mentioned in Section 4.2.2, is in fact a booster and not a classification technique and therefore, was used in three articles in combination with other methods such as Support Vector Machines and Dynamic Bayesian Network. The Naive Bayes is a probabilistic technique and special case of Bayesian theorem that bases itself in particular features to classify an unknown sample. KNN is a method that projects the sample in a feature space and decides the class of the being-analyzed sampled based on the it?s k-nearest neighbours in the space, for that, it uses Euclidean distance for measurements. The definition of Hidden Markov Model states that it is a doubly stochastic process with a hidden stochastic process that can only be visualized with another set of processes [77]. The Euclidean Distance method basically consist of comparing the euclidean distance of the being-analyzed sample with the training set. In the Decision Tree method, the leafs of the tree represent each class to which the samples must be classified. The other nodes of the tree, correspond to the splitting rules. As a sample move throw the tree, starting from the root, it will be classified to the best fitting class.

7.2. Neural Network Based Approaches (NNB)

The use of convolutional neural networks to solve computer vision problems has been successful in various areas, including emotion processing in real time [43]. It is important to remember that what is considered in this paper as Neural Network Based Approaches include not only Convolutional Neural Networks but also, as presented by the authors, Multi-layer Perceptrons and Neural Networks. However, as it is visible in Fig. 10, that the predominant method used in the set of articles selected for this work is the CCN approach, which achieved over 65% of the NNBs applications. As we are working with NNBs, the

?Neural Network? that composes more than 21% in Fig. 10 are in fact, as reported by the authors as Neural Network in general. Therefore, the first step is to understand what is in fact a neural network.

1. Download: Download high-res image (27KB)
2. Download: Download full-size image

Fig. 10. Usage of NNB approaches on the chosen set of papers. The yellow portion in the figure represents the works that did not specify the architecture of the Artificial Neural Network used and, therefore, were considered as a generic Neural Networks.

Artificial neural networks (ANN) are an attempt to simulate (simplified, of course, but still) what happens in human brain where electrical signals are forwarded through nerve cells that are only capable of communicating with each other, the neurons [33]. In other words, an artificial neural network is a set of processing units called nodes, that simulates the neurons of an actual brain and are interconnected through a set of values called weights [15]. Weights are the simulation of synaptic connections that happen between neurons in the nervous system. This elements generally have variable parameters that affect the signal that pass throw them and, together with nodes, composes the basis of a ANN. The configuration of the network may vary depending on which problem it will be designated for and different configurations result in numerous types of ANNs [90]. Therefore, MLP and CNN are specific configurations of ANNs and yet have different results on classifying facial emotion expressions.

7.2.1. Multi Layer Perceptron (MLP)

MLP is an artificial neural network architecture that uses a structure known as hidden layers to increase the degree of accuracy of its classifications. Hidden layers are layers of neurons that receive the outputs of the previous layers multiplied by their respective connection weights, add them up, and multiply by an activation function. The sigmoid function is one of the most used activation functions and has its definition as:F(x)=11-e-xThe use of hidden layers was only possible thanks to the development of training algorithms for MLP, such as the classic Backpropagation algorithm. Backpropagation is a way to solve the problem of finding how much the weights of the network must be changed in order to the result to be closer to the correct one. It uses the observation that data at any point in the neural network is generated by a composite function and the derivative chain rule to calculate the downward gradient and then understand how the weights should be changed. The only problem with this technique is getting stuck in local minimums and failing to converge to a global minimum. To try to avoid the previous problem some dynamic learning rate heuristics are used. Genetic algorithms are also possible for training an MLP.

7.2.2. Convolutional Neural Netwoks (CNNs)

Writing about CNNs in a Section on classification algorithms can be a little inaccurate. The reason for this is that although CNNs are classifiers they do not only have this role within the pipeline established at the beginning of this article, but the combination of the feature extraction step with the classification step. Convolutional neural networks have emerged as an alternative to extracting characteristics through human knowledge on a given topic. The main idea of the algorithm is to use the convolution operation to extract characteristics from the image, making the position information of the pixel with its neighborhood more atomic for the classification process.

The first operation performed in the classification step by a CNN is the convolution between matrices called filters and a sub space of the original image, generating several new images of smaller size than the original one.

There is one output image for each filter applied to the image. The data generated by this process is called feature maps. The second step is to apply nonlinearity to feature maps, where one of the common nonlinear functions is Relu. The non-linear function is applied to each pixel of the feature maps, creating rectified feature maps. The new set of images obtained could be passed through the same two processes explained previously and this could be repeated many times. An optional step at this point would be to add a pooling layer to decrease the amount of data to be passed on to the next step. One of the possible common pooling functions in the CNN architectures used by the reviewed articles is max-pooling, which is basically characterized by dividing the images of the feature maps into sub spaces and taking the maximum value of the sub space as a representative of the entire subset. In the pipeline defined earlier by this article, this step would be within the dimensionality reduction box. After these preprocessing, a conventional MLP is normally used to classify the data. The architectural pattern for CNNs can be seen in Fig. 11.

1. Download: Download high-res image (93KB)
2. Download: Download full-size image

Fig. 11. Common CNN Architecture and its feature extraction convolutional pool.

8\. Emotion Recognition Methods from Facial Images or Videos

Due to the increasing use of NNB methods for the emotion classification (Fig. 1), these kind of methods?s performance was expected to be superior than the classical methods, however, when analysing the accuracy of each method proposed by the selected papers, it was possible to observe that, on average, the applications that applied classical methods achieved better results (Fig. 12).

1. Download: Download high-res image (105KB)
2. Download: Download full-size image

Fig. 12. The plot was generated filtering our data with two rules. The first one was take the combination of datasets for training that had at least two members to represents the class. The second was to delete the classes of the plot that has members only for neural based or classical algorithms.

However, looking closer to this information in the general flow of computational facial emotion recognition presented on Section 3, it is possible to identify that the comparison may not be trustworthy. When observing accuracy achieved from both method, we are ignoring all the other steps of the general flow, such as preprocessing, feature extraction and database selection. Taking the method all alone, despite the other steps of the flow, the accuracy may be better in classical applications, however, every step should be analyzed to obtain a fair comparison.

Fig. 13 presents a complement on the information by Fig. 12. The datasets usage is unbalanced in terms of classical and NNB approaches. The CK + dataset (Section 6.2, which is the most used by the NNB methods, is a much wider dataset than JAFFE. That may partially explain the accuracy difference between classical and NNB approaches, giving motive to the expansion of NNB usage, since they achieved great results while working with large amounts of data.

1. Download: Download high-res image (82KB)
2. Download: Download full-size image

Fig. 13. Usage of distinct datasets by both NNB and Classical approaches reflects directly in the final accuracy of the method.

Since each part of the general flow may influence the final result and a comparison between approaches is incomplete when looking into accuracy exclusively, this section intends to present some methods applied by 10 selected works, describing the complete process, from the dataset choice to the classification technique. For this purpose, a rank of both classical and NNB best accuracy papers for each principal dataset was designed (Table 2). The complete data with the 94 methods analyzed is presented in Table 3.

Table 2. Top Accuracy(Acc) papers for each one of the most used datasets by the papers in the study.

Dataset	Classification	Average Acc (%)	Best Acc (%)	Best Acc Paper(s)	Year
Jaffe	Classical	92,91	100	[3]	2015
	NNB	86,00	100	[3], [76]	2015/ 2018
CK	Classical	92,40	100	[91]	2018
	NNB	87,37	99,75	[3]	2015
CK+	Classical	83,88	94,09	[34]	2015
	NNB	96,76	99,68	[58]	2017
FER 2013	NNB	64,24	64,24	[27]	2018
Bu-3dfe	NNB	91,89	91,89	[58]	2017
MMI	Classical	87,73	100	[66]	2014
	NNB	78,40	78,40	[36]	2019

Table 3. This table contains a total of 94 methods that were extracted from the 51 studies selected for this literature review. Some cells are left blank because of the lack of that specific information at the source papers.

Acc (%)	Classification	Test dataset	Training dataset	Ref	Classic (C) vs NNB (N)	Facial detection	Other preprocessing algorithms	Feature extraction	Dimensionality reduction
78.64	src	oulu-casia nir&vis	oulu-casia nir&vis	[124]	C			lbp;asm	adaboost
72.09	svm	oulu-casia nir&vis	oulu-casia nir&vis	[124]	C			lbp;asm	adaboost
87.50	svm	jaffe	jaffe	[107]	C			luxand facesdk	lda
93.75	naive bayes	jaffe + tfeid + radboud	jaffe + tfeid + radboud	[2]	C	viola-jones		hog;lbp;nn	pca
52.00	bn	proprietary	proprietary	[123]	C			standard deviation skewness kurtosis correlation coeficient mean of psd standard deviation of psd	pca
70.00	decision tree	proprietary	proprietary	[123]	C			standard deviation skewness kurtosis correlation coeficient mean of psd standard deviation of psd	pca
49.00	random forest	proprietary	proprietary	[123]	C			standard deviation skewness kurtosis correlation coeficient mean of psd standard deviation of psd	pca
42.00	random tree	proprietary	proprietary	[123]	C			standard deviation skewness kurtosis correlation coeficient mean of psd standard deviation of psd	pca
36.00	svm	proprietary	proprietary	[123]	C			standard deviation skewness kurtosis correlation coeficient mean of psd standard deviation of psd	pca
93.57	euclidian distance	jaffe	jaffe	[64]	C		log-gabor filter		pca
90.90	hmm	ck	ck	[112]	C	nn			pca
99.75	elm-rbf	ck	ck	[3]	N		radon transform		pca;lda
100	elm-rbf	jaffe	jaffe	[3]	N		radon transform		pca;lda
99.61	knn	ck	ck	[3]	C		radon transform		pca;lda
100	knn	jaffe	jaffe	[3]	C		radon transform		pca;lda
99.71	svm	ck	ck	[3]	C		radon transform		pca;lda
100	svm	jaffe	jaffe	[3]	C		radon transform		pca;lda
94.09	svm	ck+	ck+	[34]	C	viola-jones	lowpass filter	haar-cascade;lbp	pca;lda;resize
92.00	svm	jaffe	jaffe	[34]	C	viola-jones	lowpass filter	haar-cascade;lbp	pca;lda;resize
93.43	cnn	rafd	rafd	[115]	N		cropping thresholding		resize
61.29	cnn	fer2013	sfew	[116]	N	jda cnn mixture of trees	gray-scale histogram equalization		resize

91.89| cnn| bu-3dfe| bu-3dfe| [58]| N| | rotation correction intensity normalization opping| | resize
Acc (%)| **Classification**| **Test dataset**| **Training dataset**|
Ref| **Classic(C) vs NNB (N)**| **Facial detection**| **Other
preprocessing algorithms**| **Feature extraction**| **Dimensionality
reduction**
96.76| cnn| ck+| ck+| [58]| N| | rotation correction intensity normalization opping| | resize
86.74| cnn| jaffe| jaffe| [58]| N| | rotation correction intensity normalization opping| | resize
68.00| cnn| rafd| fer2013| [83]| N| haar cascade| | | resize
50.50| cnn| fer2013| imagenet| [70]| N| viola-jones| | | resize
84.55| svm| ck| ck| [94]| C| | | aam|
95.00| svm| ck| ck| [10]| C| viola-jones| | asm|
99.00| svm| ck| ck| [8]| C| viola-jones| | contourlet transform|
98.00| svm| jaffe| jaffe| [8]| C| viola-jones| | contourlet transform|
93.96| fuzzy| rafd| rafd| [40]| C| | | edge detection skin color detection|
76.60| svm| otcbvs| otcbvs| [35]| C| viola-jones| | evolutionary computation|
87.43| adaboost + dbn| ck+| ck+| [57]| C| | | gabor;asm|
82.40| adaboost + dbn| mmi| ck+| [57]| C| | | gabor;asm|
47.00| euclidian distance| feedb| feedb| [96]| C| | | heuristic|
85.05| cnn| ck+;kdef| ck+;kdef| [49]| N| viola-jones| gray-scale| hog|
56.07| mlp| ck+;kdef| ck+;kdef| [49]| N| viola-jones| gray-scale| hog|
63.55| svm| ck+;kdef| ck+;kdef| [49]| C| viola-jones| gray-scale| hog|
82.87| svm| ava| ava| [106]| C| | | hog;sift;cnn|
94.36| svm| cuhk| cuhk| [106]| C| | | hog;sift;cnn|
93.77| svm| live-iq| live-iq| [106]| C| | | hog;sift;cnn|
86.63| svm| pne| pne| [106]| C| | | hog;sift;cnn|
96.42| fuzzy| jaffe| jaffe| [28]| C| | | integral projection|
94.28| svm| proprietary| proprietary| [42]| C| kinect| | kinect|
54.56| cnn| casia webface| casia webface| [55]| N| | | lbp|
100| knn| ck| ck| [91]| C| | | lbp|
97.14| knn| jaffe| jaffe| [91]| C| | | lbp|
77.67| svm| bu-3dfe| bu-3dfe| [68]| C| viola-jones| cropping;segmentation|
lbp;gmm|
91.60| svm| ck| ck| [68]| C| viola-jones| cropping;segmentation| lbp;gmm|
88.10| svm| jaffe| jaffe| [68]| C| viola-jones| cropping;segmentation|
lbp;gmm|
80.00| svm| multi-pie| multi-pie| [68]| C| viola-jones| cropping;segmentation|
lbp;gmm|
78.80| fuzzy| jaffe| jaffe| [71]| C| | thresholding| matlab|
78.80| fuzzy| ebner?s| ebner?s| [23]| C| matlab| | matlab|
75.00| nn| ck| proprietary| [119]| N| | | nn|
95.00| svm| ck| ck| [61]| C| | | rcsr;lbp;hog|
96.40| svm| jaffe| jaffe| [61]| C| | | rcsr;lbp;hog|
87.00| decision tree| jaffe| jaffe| [75]| C| | | template matching|
93.24| cnn| ck+| ck+| [41]| N| | gn| |
95.23| cnn| jaffe| jaffe| [41]| N| | gn| |
68.79| cnn + svm;| fer2013| fer2013| [109]| CN| | rotation correction histogram equalization| |
87.72| adaboost| mmi| mmi| [66]| C| viola-jones| segmentation;thresholding| |
76.60| naive bayes| mmi| mmi| [66]| C| viola-jones| segmentation;thresholding| |
84.31| svm| mmi| mmi| [66]| C| viola-jones| segmentation;thresholding| |
90.00| dbn| dtu-aam ioid-facedatabase| dtu-aam ioid-facedatabase| [53]| C| aam| | |
80.90| cnn| emotiw| emotiw| [97]| N| cnn| | |
87.52| nn| jaffe| jaffe| [111]| N| haar cascade| | |
Acc (%)| **Classification**| **Test dataset**| **Training dataset**|
Ref| **Classic (C) vs NNB (N)**| **Facial detection**| **Other

preprocessing algorithms**| **Feature extraction**| **Dimensionality reduction**
69.77| cnn| kaggle dataset| kaggle dataset| [43]| N| hog;linear classifier| | |
100| cnn| jaffe| jaffe| [76]| N| viola-jones| | |
89.58| cnn| kdef| kdef| [76]| N| viola-jones| | |
71.9| cnn| kdef + jaffe + sfew| kdef + jaffe + sfew| [76]| N| viola-jones| | |
90.9| svm| ck+| ck+| [16]| C| viola-jones| | |
86.4| svm| mmi| mmi| [16]| C| viola-jones| | |
90.38| fcm| ck+| ck+| [120]| C| viola-jones modified| | |
74| viola-jones modified| proprietary| proprietary| [51]| C| viola-jones modified| | |
89| adaboost| ck| ck| [22]| C| | | |
76| adaboost| ck+| ck+| [22]| C| | | |
51.2| cnn| afew| afew + proprietary| [36]| N| | | |
93.9| cnn| ck+| afew + proprietary + ck+| [36]| N| | | |
73.6| cnn| ck+| affectnet| [99]| N| | | |
78| cnn| ck+| ck+| [13]| N| | | |
64.46| cnn| fer2013| fer2013| [27]| N| | | |
62| cnn| ised| affectnet| [99]| N| | | |
46.5| cnn| jaffe| affectnet| [99]| N| | | |
78.4| cnn| mmi| afew + proprietary + mmi| [36]| N| | | |
94.48| cnn| multi-pie| multi-pie| [50]| N| | | |
41.03| cnn| tfd;google dataset| afew| [47]| N| | | |
96| knn| proprietary| proprietary| [98]| C| | | |
95.24| mge| jaffe;mug;ck+| jaffe;mug;ck+| [45]| C| | | |
90| mlp| proprietary| proprietary| [98]| N| | | |
77.8| naive bayes| ck| ck| [22]| C| | | |
72.9| naive bayes| ck+| ck+| [22]| C| | | |
87| svm| ck| ck| [22]| C| | | |
75.5| svm| ck+| ck+| [22]| C| | | |
71| svm| proprietary| proprietary| [6]| C| | | |
8.1. Jaffe best accuracy classical approach
A classical approach was demonstrated in [3], where the Jaffe (Section 6.1) and CK (Section 6.2) datasets were used. For preprocessing, a image crop was applied on the datasets based on a particular model [85]. As feature extraction techniques, the authors have chosen a quite complex processes. Firstly, a radon transform is applied to turn lines through images into points in the radon domain [30]. Besides that, to overcome the subtle changes, complexity and variability of the non-linear features, a empirical mode decomposition (EMD) based on the work by [37] was applied to minimize the effects of noise. As described by the authors, these method is capable of decomposing any complicated signal into oscillating component called intrinsic mode functions (IMF). These IMFs on the other hand, reflect to signal?s local characteristic features. With the signal extracted from the images, the authors applied 3 independently algorithms of dimensionality reduction: PCA \+ Linear Discriminant Analysis (LDA) [18], PCA \+ Local Fisher Discriminant Analysis (LFDA) [93], [110], [121], [78] and Kernel LFDA (KLFDA) [124]. After the previous process, the obtained features were still subjected to a statistical analysis named ANOVA test, with the goal of selecting only the significant features and then proceed to the classification step, in which the authors have decided to apply KNN previously mentioned in Section 7.1.4 and Gaussian SVM (Section 7.1.1) classifiers. Evaluating the model with the JAFFE dataset, both classification algorithms were able to achieve 100% accuracy. On the other hand, while using CK + dataset for evaluation, the KNN classifier was able to achieve 99.31% accuracy and the SVM got the classification write

on 99,43% of the time.
8.2. Jaffe best accuracy NNB approaches
With the highest achieved accuracy between NNB methods, a CNN approach is presented capable off reaching 100% precision [76]. Three datasets were used for training in this work: Jaffe (previously presented on Section 6.1) with 213 images of 10 subjects, Karolinska Directed Emotional Faces (KDEF) Database [29] with 4900 images of 70 subjects and Static Facial Expressions In The Wild (SFEW) Database [19] with 700 images extracted from movies. As most of the methods, this project applied grayscale conversion and downscaling with geometric transformations algorithms (Sections 4.1 Grayscale conversion, 4.3 Dimensionality reduction). The authors have opted to use Viola-Jones (Section 4.2.1) to detect faces in the datasets images, and than crop them to reduce even more the data inputted to the model, making them 256x256 pixels.
In order to achieve good results with CNN models, the training set must be very large, much more than the sum of the images of each dataset chosen. Therefore, a Synthetic Image Generation, as named by the authors, took place with the objective of augmenting the set for training. This process was responsible for generating new images by rotating the images already in the datasets and, therefore, feeding the model with much larger set of data. This technique is capable of augmenting the dataset and helping the CNN not to get over fitted, but in the other hand, the number of subjects does not change with this application and, therefore, it is still difficult for the model to get god results when introduced to an ?unknown face?. With the CNN trained, the authors claimed to achieve 100% precision when evaluating the model with the JAFFE dataset and over 89% when using KDEF for evaluation.
8.3. Jaffe/Cohn-Kanade best accuracy NNB approaches
As for Jaffe dataset applied in a NNB method, [3] makes use of an extreme learning machine (ELM) with radial basis function (RBF) kernel. As mentioned by the authors, ELM with RBF kernel (ELM-RBF) is an extension of ELM from single hidden layer feedforward neural network (SLFN) with additive neurons case to SLFNs with RBF kernel. With the same method mentioned in the previous paragraph, this approach was able to achieve 100% precision with the Jaffe dataset and over 95% accuracy in CK + images.
8.4. Cohn-Kanade best accuracy classical approach
The work from [91] is a quite different from the ones mentioned preciously. This paper has the main goal of reviewing 22 Local Binary Pattern (LBP)-like descriptors and provide a comparison between then in the field of facial expression recognition. The methods proposed in this context was a simple parameter-free Nearest Neighbor (NN) classifier, since the real goal was to evaluate the LBP variants in terms of their performance on extracting features from the images. For the CK dataset, the authors claimed to achieve 100% score with eight descriptors (AELTP, BGC3, CSALTP, dLNP ?, LDN, nLNPd, STS, and WLD). Besides the great results achieved, the paper does not describe any pre processing step or extremely complex process, just the LBP feature extractors associated with the NN classifier did the job quite well.
8.5. Extended Cohn-Kanade best accuracy classical approach
The technique used by [34] was elaborated on top of classic techniques and was tested in the CK + (Section 6.2) and JAFFE (Section 6.1) datasets. For the authors attempted to understand the contribution of different areas in the face for the expression recognition and, therefore, applied the following method:
 * 1.
Pre-processing - A low pass filter with 3x3 Gaussian mask to remove noise from the source images was used;

* 2.
Face detection - Viola-Jones with Adaboost learning for face detection is applied to locate the face inside the picture;
* 3.
Detection of eyes and nose - The noise was located using the geometric position of the face and Haar cascade. The eyes had a specific Haar classifier trained for each one of them;
* 4.
Eyebrows and mouth detection - The position of the mouth was obtained from the lower region of the nose, where a Sobel horizontal edge detector was applied to detect the upper lip position. For the detection of the eyebrows, the authors applied the same method as for the lips, however, with the subtle change that is the performance of an adaptive threshold operation before applying the operator.
* 5.
Extraction of active facial patches - Fixed regions of the face (based on the eyes, nose?) were extracted since they might contain relevant information.
* 6.
Feature extraction - For this action, LBP was chose because of its robustness in therms of illumination invariant feature description.
After extracting the important features from the pictures, a SVM algorithm was used to classify the image into the best matching category (anger, disgust, fear, happiness, sadness and surprise) and was capable of achieving a average accuracy of 94,09%.
8.6. Extended Cohn-Kanade/BU-3DFE best accuracy NNB approach
The article with the best performance in the CK + dataset using CNN was written by [58], in this article data sets that have a controlled environment were tested, these being CK+ (6.2), JAFFE(6.1), BU-3DFE(6.5). The authors opted to use some preprocessing techniques such as rotation correction, image cropping, down sampling and intensity normalization (4) in order to have only the meaningful data for the next steps and yet maintain a general standard for the images. For the classification step, a CNN was used and for its input, the images were configured in a 32x32 matrix of pixels. After the input, the the network was composed by two convolution and subsampling sets and a fully connected layer. The algorithm was developed to identify anger, disgust, fear, happiness, sadness and surprise. The proposed method was capable of achieving a average of 96,76 for CK + and 91,89 for BU-3DFE.
8.7. FER 2013 best accuracy NNB method
The best accuracy paper for the FER 2013 dataset is based on the usage of the already known CNN architectures in the facial expression recognition problem, these being GoogleNet, ResNet, VGGNet and AlexNet. Pre-processing is not mentioned by the authors, instead focusing on the classification methods themselves and the process of applying them. The authors decided to train the models for different numbers of epochs: 6, 8, 12, 20 and 25. The top average accuracy was 64,24 achieved by the AlexNet model trained for 20 epochs. Besides being a very small number of epochs in comparison of other methods, the authors claim that the model could be suffering from overfitting, since the accuracy was better when training the model for 20 epochs than at the 25 epochs attempt.
8.8. MMI best accuracy classic method
As used by a great partition of the papers analyzed on this work, [66] uses the Viola-Jones algorithm to detect faces in the images and, therefore, crop them appropriately. Additionally, a skin detector based on texture is applied to identify the non-face parts of the pictures and, finally, segment them by

this characteristic. Done with preprocessing, a facial graph is built containing some interest point of the face. For the classification step, the authors used three different classifiers, all of them based on FACS (5): Adaboost, Naive Bayes and SVM. The algorithm was developed to classify the following emotions: Surprise, Happiness, Disgust, Fear, Anger, Sadness and Neutral. The best accuracy achieved was 100% with the SVM algorithm for the neutral emotion, however, the best average accuracy across all the emotions (87,73%) was achieved by the Adaboost classifier.

8.9. MMI best accuracy NNB method

The last paper of the Table 2 presents a very interesting and dissimilar approach from the other works. [36] decided to analyze data from videos instead of images as usual. For that, facial landmarks are extracted, facial detection, alignment and input normalization are applied in the frames of the videos and attributed to a Long Short Term Memory (LSTM) neural network. The videos inputted to the LSTM are composed by 16 frames of 244 x 244 sized face images and were achieved from image sources, that being AFEW, CK + and MMI. The average accuracy for the MMI dataset was 78.40%.In addition to the accuracy in the MMI dataset, the algorithm also achieves good results in the AFEW and CK + datasets, with 51.2% and 93.9% respectively.

9. Conclusion and discussions

Computational Facial emotion recognition as performed by humans can be considered a very challenging task when using input images or videos solely. Since face-based human emotion recognition is performed in a natural way as a kind of non-verbal communication, it is also very dependent of other inputs to be effective, such as a context of application itself, and sometimes it is hard to be recognized by the available analytical ways. Besides its complexity, finding computational mechanisms that are able to effectively recognize emotion from facial images, is very desired in the most varied fields with significant implications.

This survey intended to present a systematic literature review of the state-of-the-art on emotion expression recognition from facial images, considering the works published over the past few years. We categorized a total of 51 manuscripts, analyzed according to its main constructive concept, totaling 94 distinct approaches, where two main trends were able to be identified: classical approaches and those based on neural networks approaches, including the emerging convolutional counterpart. For the approaches based on Digital Image Processing, Pattern Recognition and Computer Vision techniques, we named them classical approaches. In essence, classical approaches consider those based on a face detection procedure followed by some recognizer able to interpret feature vector and made some decision according to a fixed number of output classes. These approaches usually explore some _m_ -dimensional space problem partition, looking for some discrimination among input vectors and output clusters, such as the well-know Support Vector Machines (SVM), Naive Bayes, k-Nearest Neighbors, Fuzzy, among others. On the other spectrum of problem-solvers, we have the neural network based approaches (NNB approaches), where the input layer containing a discretization of the facial image or more commonly feature vectors extracted using some feature descriptor method is used. NNB approaches include the emerging convolutional neural network counterpart, and since its first use as computer vision solver-problems, it was also used to the facial emotion recognition as well.

The papers analyzed here used a huge variety of techniques for the construction of the emotion recognition algorithm, during the development of this survey. This survey highlights several aspects provided by the published papers over the literature. One of the most important, that arises some interest of the general public, is to answer the question if a general effective solution has been achieved to solve the facial emotion recognition problem, which can be translated in terms of the general precision. In this survey we can point to some interesting solutions, but the response is that the problem can be considered solved for some very specific situations. The most evident problem in terms of the published approaches is that some variability can always be introduced to decrease the performance of the proposed works. The use of controlled datasets or a very limited subset of those images makes in fact the proposed approaches effective, but very restricted to this specific domain. When applied in real-world problems, the precision overall tends to decrease significantly. This aspect can be mainly observed for the classical approaches when compared to the NNB ones, where a marginally better precision can be observed mainly justified due to the nature of the domain used. In general, what can be observed for classical approaches is a very high specificity and low generality to solve the emotion recognition problem. Here, we can point the approaches proposed by [91], [3], [8], where a high-precision was achieved using classical approaches. For the classical approaches, the best ranked approach was [91] achieving a precision of 100 percent using the 1 nearest neighbor for classification and local binary patterns for feature extraction, and we computed a mean overall precision of 83.89 percent for the classical approaches group.

On the other hand, using NNB approaches, what can be observed is a discrete decrease of precision when compared to classical approaches, more specifically due the generality increase and the abstraction capability. However, the cost is to consider a large input dataset to train the network properly. In terms of precision, the best NNB approache have achieved a accuracy of 100 percent, and we computed a general precision of 77,52 percent. As conclusion, when comparing classical approaches vs NNB ones, despite the fact that NNB approaches have a lower performance for the selected articles in this survey, NNB methods provide the possibility of abstraction, applying facial emotion recognition with larger datasets. This fact allows it to be applied in less restricted scenarios than the classical methods, which requires greater restrictions for their properly operation.

Another aspect verified in this survey is the importance of the dataset used to train the classical and the NNB approaches. Traditional datasets such as JAFFE are very common and explored over the past few years, besides its low resolution, restrictions in size, and low variability in terms of nationality of the participants and gender. We can observe the lack of reliable datasets with good resolution, number of images, and including the most different cases and variability, made with several countries and ethnic diversity. For instance, a very small group is observed in the MMI dataset in which only 25 people were used. In addition, some datasets are obsolete, and the aforementioned drawbacks makes the tested algorithms often restricted and they end up showing good results only for a limited domain. Nonetheless, there is a remarkable difference for the datasets constructed taking into account artificial and natural expressions, where its hard to extract from a participant the real meaning for a facial expression, specially those dependent from a context. This survey suggest that datasets including natural and artificial expressions must be developed in order to increase the overall precision of the presented approaches.

Besides the great and interesting results achieved by some of the analyzed approaches, this study showed that the research still face difficulties with some aspects of this type of problem, especially for reliable datasets, controlled environments, due to the variation of luminosity, occlusion, among others. NNB approaches have achieved a very interesting capacity of modelling the nature of the problem, and what can be observed is a trend for the further works explore the capabilities of NNB approaches.

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.

Acknowledgments

This work was carried out with the support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Recommended articles

References

1. [1]
Albawi, S., Abed Mohammed, T., ALZAWI, S., 2017. Understanding of a convolutional neural network. DOI: 10.1109/ICEngTechnol.2017.8308186.
Google Scholar
2. [2]
G. Ali, M.A. Iqbal, T.S. Choi
Boosted nne collections for multicultural facial expression recognition
Pattern Recognition, 55 (2016), pp. 14-27
View in ScopusGoogle Scholar
3. [3]
H. Ali, M. Hariharan, S. Yaacob, A.H. Adom
Facial emotion recognition using empirical mode decomposition
Expert Systems with Applications, 42 (2015), pp. 1261-1277
View in ScopusGoogle Scholar
4. [4]
V.M. Álvarez, R. Velázquez, S. Gutiérrez, J. Enriquez-Zarate
A method for facial emotion recognition based on interest points
2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), IEEE (2018), pp. 1-4
Google Scholar
5. [5]
Anila, S., Devarajan, N., 2012. Preprocessing technique for face recognition applications under varying illumination conditions. Global Journal of Computer Science and Technology.
Google Scholar
6. [6]
J.N. Bailenson, E.D. Pontikakis, I.B. Mauss, J.J. Gross, M.E. Jabon, C.A. Hutcherson, C. Nass, O. John
Real-time classification of evoked emotions using facial feature tracking and physiological responses
International Journal of Human-computer Studies, 66 (2008), pp. 303-317
View in ScopusGoogle Scholar
7. [7]
Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, H. Greenspan
Chest pathology detection using deep learning with non-medical training
2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), IEEE (2015), pp. 294-297
CrossrefView in ScopusGoogle Scholar
8. [8]
S. Biswas, J. Sil
An efficient expression recognition method using contourlet transform
Proceedings of the 2nd International Conference on Perception and Machine

Intelligence, ACM (2015), pp. 167-174
CrossrefView in ScopusGoogle Scholar
9. [9]
T. Bouwmans, C. Silva, C. Marghes, M.S. Zitouni, H. Bhaskar, C. Frelicot
On the role and the importance of features for background modeling and foreground detection
Computer Science Review, 28 (2018), pp. 26-91
View in ScopusGoogle Scholar
10. [10]
J. Chen, D. Chen, Y. Gong, M. Yu, K. Zhang, L. Wang
Facial expression recognition using geometric and appearance features
Proceedings of the 4th international conference on internet multimedia computing and service (2012), pp. 29-33
CrossrefView in ScopusGoogle Scholar
11. [11]
Chen, L.F., Yen, Y.S., 2007. Taiwanese facial expression image database. Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan.
Google Scholar
12. [12]
J. Cheng, Y. Deng, H. Meng, Z. Wang
A facial expression based continuous emotional state monitoring system with gpu acceleration
2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG), IEEE (2013), pp. 1-6
CrossrefView in ScopusGoogle Scholar
13. [13]
D.Y. Choi, D.H. Kim, B.C. Song
Recognizing fine facial micro-expressions using two-dimensional landmark feature
2018 25th IEEE International Conference on Image Processing (ICIP), IEEE (2018), pp. 1962-1966
CrossrefView in ScopusGoogle Scholar
14. [14]
T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham
Active shape models-their training and application
Computer Vision and Image Understanding, 61 (1995), pp. 38-59
View in ScopusGoogle Scholar
15. [15]
S.S. Cross, R.F. Harrison, R.L. Kennedy
Introduction to neural networks
The Lancet, 346 (1995), pp. 1075-1079
View in ScopusGoogle Scholar
16. [16]
A.C. Cruz, B. Bhanu, N.S. Thakoor
One shot emotion scores for facial emotion recognition
2014 IEEE International Conference on Image Processing (ICIP), IEEE (2014), pp. 1376-1380
CrossrefView in ScopusGoogle Scholar
17. [17]
Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection.
Google Scholar
18. [18]

H.B. Deng, L.W. Jin, L.X. Zhen, J.C. Huang, _et al._
A new facial expression recognition method based on local gabor filter bank and pca plus lda
International Journal of Information Technology, 11 (2005), pp. 86-96
CrossrefGoogle Scholar
19. [19]
A. Dhall, R. Goecke, S. Lucey, T. Gedeon
Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark
2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), IEEE (2011), pp. 2106-2112
CrossrefView in ScopusGoogle Scholar
20. [20]
A. Dhall, R. Goecke, S. Lucey, T. Gedeon, _et al._
Collecting large, richly annotated facial-expression databases from movies
IEEE Multimedia, 19 (2012), pp. 34-41
View in ScopusGoogle Scholar
21. [21]
H.L. Egger, D.S. Pine, E. Nelson, E. Leibenluft, M. Ernst, K.E. Towbin, A. Angold
The nimh child emotional faces picture set (nimh-chefs): a new set of children?s facial emotion stimuli
International Journal of Methods in Psychiatric Research, 20 (2011), pp. 145-156
CrossrefView in ScopusGoogle Scholar
22. [22]
M.T. Eskil, K.S. Benli
Facial expression recognition based on anatomy
Computer Vision and Image Understanding, 119 (2014), pp. 1-14
Google Scholar
23. [23]
F.S. Farahani, M. Sheikhan, A. Farrokhi
A fuzzy approach for facial emotion recognition
2013 13th Iranian Conference on Fuzzy Systems (IFSC), IEEE (2013), pp. 1-4
Google Scholar
24. [24]
Freitas-Magalhaes, A., 2018. Facial Action Coding System 3.0-Manual de Codificacao Cientifica da Face Humana-FM BabyFACS. Leya.
Google Scholar
25. [25]
Y. Freund, R.E. Schapire
A decision-theoretic generalization of on-line learning and an application to boosting
Journal of Computer and System Sciences, 55 (1997), pp. 119-139
View in ScopusGoogle Scholar
26. [26]
Friesen, E., Ekman, P., 1978. Facial action coding system: a technique for the measurement of facial movement. Palo Alto 3.
Google Scholar
27. [27]
Y. Gan
Facial expression recognition using convolutional neural network
Proceedings of the 2nd International Conference on Vision, Image and Signal Processing, ACM (2018), p. 29

View in ScopusGoogle Scholar
28. [28]
R. Ghasemi, M. Ahmady
Facial expression recognition using facial effective areas and fuzzy logic
2014 Iranian Conference on Intelligent Systems (ICIS), IEEE (2014), pp. 1-4
CrossrefGoogle Scholar
29. [29]
E. Goeleven, R. De Raedt, L. Leyman, B. Verschuere
The karolinska directed emotional faces: a validation study
Cognition and emotion, 22 (2008), pp. 1094-1118
CrossrefView in ScopusGoogle Scholar
30. [30]
Gonzales, R.C., Woods, R.E., 2002. Digital image processing.
Google Scholar
31. [31]
Goodfellow, I., Erhan, D., Carrier, P.L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., Ionescu, R., Popescu, M., Grozea, C., Bergstra, J., Xie, J., Romaszko, L., Xu, B., Chuang, Z., Bengio, Y., 2013. Challenges in representation learning: A report on three machine learning contests http://arxiv.org/abs/1307.0414.
Google Scholar
32. [32]
Gupta, A., Sharma, D., Sharma, S., Agarwal, A., 2020. Survey paper on gender and emotion classification using facial expression detection. Proceedings of the International Conference on Innovative Computing & Communications (ICICC) 2020 10.2139/ssrn.3565808.
Google Scholar
33. [33]
K. Gurney
An introduction to neural networks
CRC Press (2014)
Google Scholar
34. [34]
S. Happy, A. Routray
Automatic facial expression recognition using features of salient facial patches
IEEE Transactions on Affective Computing, 6 (2014), pp. 1-12
Google Scholar
35. [35]
B. Hernández, G. Olague, R. Hammoud, L. Trujillo, E. Romero
Visual learning of texture descriptors for facial expression recognition in thermal imagery
Computer Vision and Image Understanding, 106 (2007), pp. 258-269
View in ScopusGoogle Scholar
36. [36]
M. Hu, H. Wang, X. Wang, J. Yang, R. Wang
Video facial emotion recognition based on local enhanced motion history image and cnn-ctslstm networks
Journal of Visual Communication and Image Representation, 59 (2019), pp. 176-185
View in ScopusGoogle Scholar
37. [37]

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences 454, 903?995.
Google Scholar
38. [38]
S. Huang, N. Cai, P.P. Pacheco, S. Narrandes, Y. Wang, W. Xu
Applications of support vector machine (svm) learning in cancer genomics
Cancer Genomics-Proteomics, 15 (2018), pp. 41-51
CrossrefGoogle Scholar
39. [39]
Y. Huang, F. Chen, S. Lv, X. Wang
Facial expression recognition: A survey. Symmetry, 11 (2019), p. 1189, 10.3390/sym11101189
View in ScopusGoogle Scholar
40. [40]
M. Ilbeygi, H. Shah-Hosseini
A novel fuzzy facial expression recognition system based on facial feature extraction from color face images
Engineering Applications of Artificial Intelligence, 25 (2012), pp. 130-146
View in ScopusGoogle Scholar
41. [41]
D.K. Jain, P. Shamsolmoali, P. Sehdev
Extended deep neural network for facial emotion recognition
Pattern Recognition Letters, 120 (2019), pp. 69-74
View in ScopusGoogle Scholar
42. [42]
Jazouli, M., Majda, A., Zarghili, A., 2017. A p recognizer for automatic facial emotion recognition using kinect sensor, in: 2017 Intelligent Systems and Computer Vision (ISCV), IEEE. pp. 1?5.
Google Scholar
43. [43]
J. Jeon, J.C. Park, Y. Jo, C. Nam, K.H. Bae, Y. Hwang, D.S. Kim
A real-time facial expression recognizer using deep neural network
Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication, ACM (2016), p. 94
View in ScopusGoogle Scholar
44. [44]
Jia, S., Wang, S., Hu, C., Webster, P.J., Li, X., 2021. Detection of genuine and posed facial expressions of emotion: Databases and methods. Frontiers in Psychology 11, 3818. https://www.frontiersin.org/article/10.3389/fpsyg.2020.580287, DOI: 10.3389/fpsyg.2020.580287.
Google Scholar
45. [45]
R. Jiang, A.T. Ho, I. Cheheb, N. Al-Maadeed, S. Al-Maadeed, A. Bouridane
Emotion recognition from scrambled facial images via many graph embedding
Pattern Recognition, 67 (2017), pp. 245-251
View in ScopusGoogle Scholar
46. [46]
M. Jones, P. Viola
Fast multi-view face detection
Mitsubishi Electric Research Lab TR-20003-96, 3 (2003), p. 2
View in ScopusGoogle Scholar

 47. [47]
S.E. Kahou, C. Pal, X. Bouthillier, P. Froumenty, Ç. Gülçehre, R. Memisevic, P. Vincent, A. Courville, Y. Bengio, R.C. Ferrari, _et al._
Combining modality specific deep neural networks for emotion recognition in video
Proceedings of the 15th ACM on International conference on multimodal interaction (2013), pp. 543-550
CrossrefView in ScopusGoogle Scholar
 48. [48]
Kanade, T., Cohn, J.F., Tian, Y., 2000. Comprehensive database for facial expression analysis, in: Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), IEEE. pp. 46?53.
Google Scholar
 49. [49]
A. Kartali, M. Rogli?, M. Barjaktarovi?, M. Duri?-Jovi?i?, M.M. Jankovi?
Real-time algorithms for facial emotion recognition: A comparison of different approaches
2018 14th Symposium on Neural Networks and Applications (NEUREL), IEEE (2018), pp. 1-4
CrossrefGoogle Scholar
 50. [50]
V. Khryashchev, L. Ivanovsky, A. Priorov
Deep learning for real-time robust facial expression analysis
Proceedings of the International Conference on Machine Vision and Applications, ACM (2018), pp. 66-70
CrossrefView in ScopusGoogle Scholar
 51. [51]
K.C. Kirana, S. Wibawanto, H.W. Herwanto
Facial emotion recognition based on viola-jones algorithm in the learning environment
2018 International Seminar on Application for Technology of Information and Communication, IEEE (2018), pp. 406-410
Google Scholar
 52. [52]
B. Kitchenham
Procedures for Performing Systematic Reviews. Keele University. Technical Report TR/SE-0401. Department of Computer Science
Keele University, UK (2004)
Google Scholar
 53. [53]
K.E. Ko, K.B. Sim
Development of a facial emotion recognition method based on combining aam with dbn
2010 International Conference on Cyberworlds, IEEE (2010), pp. 87-91
View in ScopusGoogle Scholar
 54. [54]
O. Langner, R. Dotsch, G. Bijlstra, D.H. Wigboldus, S.T. Hawk, A. Van Knippenberg
Presentation and validation of the radboud faces database
Cognition and emotion, 24 (2010), pp. 1377-1388
CrossrefView in ScopusGoogle Scholar
 55. [55]
G. Levi, T. Hassner
Emotion recognition in the wild via convolutional neural networks and mapped

binary patterns
Proceedings of the 2015 ACM on international conference on multimodal interaction (2015), pp. 503-510
CrossrefView in ScopusGoogle Scholar
56. [56]
Li, S., Deng, W., 2018. Deep facial expression recognition: A survey. CoRR abs/1804.08348. http://arxiv.org/abs/1804.08348.
Google Scholar
57. [57]
Y. Li, S. Wang, Y. Zhao, Q. Ji
Simultaneous facial feature tracking and facial expression recognition
IEEE Transactions on Image Processing, 22 (2013), pp. 2559-2573
View in ScopusGoogle Scholar
58. [58]
A.T. Lopes, E. de Aguiar, A.F. De Souza, T. Oliveira-Santos
Facial expression recognition with convolutional neural networks: coping with few data and the training sample order
Pattern Recognition, 61 (2017), pp. 610-628
View in ScopusGoogle Scholar
59. [59]
A.T. Lopes, E. De Aguiar, T. Oliveira-Santos
A facial expression recognition system using convolutional networks
2015 28th SIBGRAPI Conference on Graphics, Patterns and Images, IEEE (2015), pp. 273-280
CrossrefView in ScopusGoogle Scholar
60. [60]
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
2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, IEEE (2010), pp. 94-101
View in ScopusGoogle Scholar
61. [61]
Y. Luo, L. Zhang, Y. Chen, W. Jiang
Facial expression recognition algorithm based on reverse co-salient regions (rcsr) features
2017 4th International Conference on Information Science and Control Engineering (ICISCE), IEEE (2017), pp. 326-329
View in ScopusGoogle Scholar
62. [62]
M. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba
Coding facial expressions with gabor wavelets
Proceedings Third IEEE international conference on automatic face and gesture recognition, IEEE. (1998), pp. 200-205
View in ScopusGoogle Scholar
63. [63]
Mehta, Dhwani, M.F.S., Javaid, A.Y., 2018. Facial emotion recognition: A survey and real-world user experiences in mixed reality. Sensors 18, 416.
Google Scholar
64. [64]
N. Mehta, S. Jadhav
Facial emotion recognition using log gabor filter and pca
2016 International Conference on Computing Communication Control and automation (ICCUBEA, IEEE (2016), pp. 1-5

Google Scholar
65. [65]
B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, _et al._
The multimodal brain tumor image segmentation benchmark (brats)
IEEE transactions on medical imaging, 34 (2014), pp. 1993-2024
Google Scholar
66. [66]
S. Mohseni, N. Zarei, S. Ramazani
Facial expression recognition using anatomy based facial graph in: 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE (2014), pp. 3715-3719
View in ScopusGoogle Scholar
67. [67]
M. Moolchandani, S. Dwivedi, S. Nigam, K. Gupta
A survey on: Facial emotion recognition and classification
2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (2021), pp. 1677-1686, 10.1109/ICCMC51019.2021.9418349
View in ScopusGoogle Scholar
68. [68]
D. Mushfieldt, M. Ghaziasgar, J. Connan
Robust facial expression recognition in the presence of rotation and partial occlusion
Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference (2013), pp. 186-193
CrossrefView in ScopusGoogle Scholar
69. [69]
Nanni, L., Ghidoni, S., Brahnam, S., 2017. Handcrafted vs. non-handcrafted features for computer vision classification. Pattern Recognition 71, 158?172. https://www.sciencedirect.com/science/article/pii/ S0031320317302224, doi: 10.1016/j.patcog.2017.05.025.
Google Scholar
70. [70]
H.W. Ng, V.D. Nguyen, V. Vonikakis, S. Winkler
Deep learning for emotion recognition on small datasets using transfer learning
Proceedings of the 2015 ACM on international conference on multimodal interaction (2015), pp. 443-449
CrossrefView in ScopusGoogle Scholar
71. [71]
Nicolai, A., Choi, A., 2015. Facial emotion recognition using fuzzy systems, in: 2015 IEEE international conference on systems, man, and cybernetics, IEEE. pp. 2216?2221.
Google Scholar
72. [72]
T. Ojala, M. Pietikäinen, D. Harwood
A comparative study of texture measures with classification based on featured distributions
Pattern recognition, 29 (1996), pp. 51-59
Google Scholar
73. [73]
M. Pantic, M. Valstar, R. Rademaker, L. Maat
Web-based database for facial expression analysis
2005 IEEE international conference on multimedia and Expo, IEEE (2005), p. 5

Google Scholar
74. [74]
Papageorgiou, C.P., Oren, M., Poggio, T., 1998. A general framework for object detection, in: Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271), IEEE. pp. 555?562.
Google Scholar
75. [75]
N. Perveen, S. Gupta, K. Verma
Facial expression recognition using facial characteristic points and gini index
2012 Students Conference on Engineering and Systems, IEEE (2012), pp. 1-6
CrossrefGoogle Scholar
76. [76]
R.V. Puthanidam, T.S. Moh
A hybrid approach for facial expression recognition
Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication, ACM (2018), p. 60
View in ScopusGoogle Scholar
77. [77]
L. Rabiner, B. Juang
An introduction to hidden markov models
ieee assp magazine, 3 (1986), pp. 4-16
View in ScopusGoogle Scholar
78. [78]
Y. Rahulamathavan, R.C.W. Phan, J.A. Chambers, D.J. Parish
Facial expression recognition in the encrypted domain based on local fisher discriminant analysis
IEEE Transactions on Affective Computing, 4 (2012), pp. 83-92
Google Scholar
79. [79]
S. Rajan, P. Chenniappan, S. Devaraj, N. Madian
Facial expression recognition techniques: a comprehensive survey
IET Image Processing, 13 (2019), pp. 1031-1040
CrossrefView in ScopusGoogle Scholar
80. [80]
raval, D., Sakle, M., 2015. A literature review on emotion recognition system using various facial expression. International Journal of Advance Research and Innovative Ideas in Education 1, 326?329.
Google Scholar
81. [81]
Revina, I., Emmanuel, W.S., 2021. A survey on human face expression recognition techniques. Journal of King Saud University - Computer and Information Sciences 33, 619?628.
https://www.sciencedirect.com/science/article/pii/ S1319157818303379, doi: 10.1016/j.jksuci.2018.09.002.
Google Scholar
82. [82]
T. Sakai, M. Nagao, S. Fujibayashi
Line extraction and pattern detection in a photograph
Pattern recognition, 1 (1969), pp. 233-248
View in ScopusGoogle Scholar
83. [83]
V.V. Salunke, C. Patil
A new approach for automatic face emotion recognition and classification based

on deep networks
2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), IEEE (2017), pp. 1-5
CrossrefGoogle Scholar
84. [84]
W. Shen, M. Zhou, F. Yang, C. Yang, J. Tian
Multi-scale convolutional neural networks for lung nodule classification
International Conference on Information Processing in Medical Imaging, Springer (2015), pp. 588-599
CrossrefView in ScopusGoogle Scholar
85. [85]
F.Y. Shih, C.F. Chuang
Automatic extraction of head and face boundaries and facial features
Information Sciences, 158 (2004), pp. 117-130
View in ScopusGoogle Scholar
86. [86]
H.C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers
Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning
IEEE transactions on medical imaging, 35 (2016), pp. 1285-1298
View in ScopusGoogle Scholar
87. [87]
J. Shlens
A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100 (2014)
Google Scholar
88. [88]
S. Shojaeilangari, W.Y. Yau, K. Nandakumar, J. Li, E.K. Teoh
Robust representation and recognition of facial emotions using extreme sparse learning
IEEE Transactions on Image Processing, 24 (2015), pp. 2140-2152
View in ScopusGoogle Scholar
89. [89]
Silander, T., Myllymaki, P., 2012. A simple approach for finding the globally optimal bayesian network structure. arXiv preprint arXiv:1206.6875.
Google Scholar
90. [90]
Singh, H.P., Bailer-Jones, C.A., Gupta, R., 2001. Introduction to artificial neural networks.
Google Scholar
91. [91]
K. Slimani, M. Kas, Y. El Merabet, R. Messoussi, Y. Ruichek
Facial emotion recognition: A comparative analysis using 22 lbp variants
Proceedings of the 2nd Mediterranean Conference on Pattern Recognition and Artificial Intelligence, ACM (2018), pp. 88-94
CrossrefView in ScopusGoogle Scholar
92. [92]
S. Soo
Object detection using haar-cascade classifier
University of Tartu, Institute of Computer Science (2014), pp. 1-12
View in ScopusGoogle Scholar
93. [93]
Sugiyama, M., 2006. Local fisher discriminant analysis for supervised

dimensionality reduction, in: Proceedings of the 23rd international conference on Machine learning, pp. 905?912.
Google Scholar
94. [94]
J.M. Sun, X.S. Pei, S.S. Zhou
Facial emotion recognition in modern distant education system using svm
2008 International Conference on Machine Learning and Cybernetics, IEEE (2008), pp. 3545-3548
View in ScopusGoogle Scholar
95. [95]
M. Szwoch
Feedb: a multimodal database of facial expressions and emotions
2013 6th International Conference on Human System Interactions (HSI), IEEE (2013), pp. 524-531
CrossrefView in ScopusGoogle Scholar
96. [96]
M. Szwoch, P. Pienia?ek
Facial emotion recognition using depth data
2015 8th International Conference on Human System Interaction (HSI), IEEE (2015), pp. 271-277
CrossrefView in ScopusGoogle Scholar
97. [97]
L. Tan, K. Zhang, K. Wang, X. Zeng, X. Peng, Y. Qiao
Group emotion recognition with individual facial emotion cnns and global image based cnns
Proceedings of the 19th ACM International Conference on Multimodal Interaction (2017), pp. 549-552
CrossrefView in ScopusGoogle Scholar
98. [98]
Tarnowski, P., Kolodziej, M., Majkowski, A., Rak, R.J., 2017. Emotion recognition using facial expressions., in: ICCS, pp. 1175?1184.
Google Scholar
99. [99]
I. Tautkute, T. Trzcinski, A. Bielski
I know how you feel: Emotion recognition with facial landmarks
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2018), pp. 1878-1880
Google Scholar
100. [100]
G.A. Taylor
Initial steps in image preparation
American Journal of Roentgenology, 179 (2002), pp. 1411-1413
CrossrefView in ScopusGoogle Scholar
101. [101]
P.R. Thrift
Fuzzy logic synthesis with genetic algorithms., in: ICGA (1991), pp. 509-513
Google Scholar
102. [102]
S. Turabzadeh, H. Meng, R.M. Swash, M. Pleva, J. Juhar
Real-time emotional state detection from facial expression on embedded devices
2017 Seventh International Conference on Innovative Computing Technology (INTECH), IEEE (2017), pp. 46-51
CrossrefView in ScopusGoogle Scholar

103. [103]
Ullah, S., Tian, W., 2020. A systematic literature review of recognition of compound facial expression of emotions, in: ICVIP 2020: 2020 The 4th International Conference on Video and Image Processing, Association for Computing Machinery, New York, NY, USA. p. 116?121. doi: 10.1145/3447450.3447469.
Google Scholar
104. [104]
V. Vapnik
Pattern recognition using generalized portrait method
Automation and remote control, 24 (1963), pp. 774-780
Google Scholar
105. [105]
P. Viola, M. Jones, _et al._
Rapid object detection using a boosted cascade of simple features
CVPR, 1 (1) (2001), p. 3
Google Scholar
106. [106]
F. Wang, J. Lv, G. Ying, S. Chen, C. Zhang
Facial expression recognition from image based on hybrid features understanding
Journal of Visual Communication and Image Representation, 59 (2019), pp. 84-88
CrossrefView in ScopusGoogle Scholar
107. [107]
H. Wang, H. Huang, Y. Hu, M. Anderson, P. Rollins, F. Makedon
Emotion detection via discriminative kernel method
Proceedings of the 3rd international conference on pervasive technologies related to assistive environments (2010), pp. 1-7
Google Scholar
108. [108]
L. Wang
Support vector machines: theory and applications, volume 177, Springer Science & Business Media (2005)
Google Scholar
109. [109]
X. Wang, J. Huang, J. Zhu, M. Yang, F. Yang
Facial expression recognition with deep learning
Proceedings of the 10th International Conference on Internet Multimedia Computing and Service, ACM (2018), p. 10
CrossrefGoogle Scholar
110. [110]
Z. Wang, Q. Ruan
Facial expression recognition based orthogonal local fisher discriminant analysis
IEEE 10th International Conference on Signal Processing Proceedings, IEEE (2010), pp. 1358-1361
CrossrefView in ScopusGoogle Scholar
111. [111]
D. Yang, A. Alsadoon, P. Prasad, A. Singh, A. Elchouemi
An emotion recognition model based on facial recognition in virtual learning environment
Procedia Computer Science, 125 (2018), pp. 2-10
Google Scholar
112. [112]

M. Yeasin, B. Bullot, R. Sharma
Recognition of facial expressions and measurement of levels of interest from video
IEEE Transactions on Multimedia, 8 (2006), pp. 500-508
Google Scholar
113. [113]
Yi, D., Lei, Z., Liao, S., Li, S.Z., 2014. Learning face representation from scratch. arXiv preprint arXiv:1411.7923.
Google Scholar
114. [114]
Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J., 2006. A 3d facial expression database for facial behavior research, in: 7th international conference on automatic face and gesture recognition (FGR06), IEEE. pp. 211?216.
Google Scholar
115. [115]
G. Yolcu, I. Oztel, S. Kazan, C. Oz, K. Palaniappan, T.E. Lever, F. Bunyak
Deep learning-based facial expression recognition for monitoring neurological disorders
2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE (2017), pp. 1652-1657
View in ScopusGoogle Scholar
116. [116]
Z. Yu, C. Zhang
Image based static facial expression recognition with multiple deep network learning
Proceedings of the 2015 ACM on international conference on multimodal interaction (2015), pp. 435-442
CrossrefView in ScopusGoogle Scholar
117. [117]
L.A. Zadeh
Fuzzy sets. Information and control, 8 (1965), pp. 338-353
View in ScopusGoogle Scholar
118. [118]
K. Zhang, Z. Zhang, Z. Li, Y. Qiao
Joint face detection and alignment using multitask cascaded convolutional networks
IEEE Signal Processing Letters, 23 (2016), pp. 1499-1503
View in ScopusGoogle Scholar
119. [119]
L. Zhang, M. Jiang, D. Farid, M.A. Hossain
Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot
Expert Systems with Applications, 40 (2013), pp. 5160-5168
View in ScopusGoogle Scholar
120. [120]
L. Zhang, K. Mistry, M. Jiang, S.C. Neoh, M.A. Hossain
Adaptive facial point detection and emotion recognition for a humanoid robot
Computer Vision and Image Understanding, 140 (2015), pp. 93-114
View in ScopusGoogle Scholar
121. [121]
S. Zhang, X. Zhao, B. Lei
Facial expression recognition based on local binary patterns and local fisher discriminant analysis

WSEAS transactions on signal processing, 8 (2012), pp. 21-31
View in ScopusGoogle Scholar
122. [122]
Y.D. Zhang, Z.J. Yang, H.M. Lu, X.X. Zhou, P. Phillips, Q.M. Liu, S.H. Wang
Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation
IEEE Access, 4 (2016), pp. 8375-8385
View in ScopusGoogle Scholar
123. [123]
Z. Zhang, L. Cui, X. Liu, T. Zhu
Emotion detection using kinect 3d facial points
2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI), IEEE (2016), pp. 407-410
View in ScopusGoogle Scholar
124. [124]
G. Zhao, X. Huang, M. Taini, S.Z. Li, M. PietikäInen
Facial expression recognition from near-infrared videos
Image and Vision Computing, 29 (2011), pp. 607-619
View in ScopusGoogle Scholar
Cited by (162)
* ### Deep learning-based multimodal emotion recognition from audio, visual, and text modalities: A systematic review of recent advancements and future prospects
2024, Expert Systems with Applications
Show abstract
Emotion recognition has recently attracted extensive interest due to its significant applications to human?computer interaction. The expression of human emotion depends on various verbal and non-verbal languages like audio, visual, text, _etc_. Emotion recognition is thus well suited as a multimodal rather than single-modal learning problem. Owing to the powerful feature learning capability, extensive deep learning methods have been recently leveraged to capture high-level emotional feature representations for multimodal emotion recognition (MER). Therefore, this paper makes the first effort in comprehensively summarize recent advances in deep learning-based multimodal emotion recognition (DL-MER) involved in audio, visual, and text modalities. We focus on: (1) MER milestones are given to summarize the development tendency of MER, and conventional multimodal emotional datasets are provided; (2) The core principles of typical deep learning models and its recent advancements are overviewed; (3) A systematic survey and taxonomy is provided to cover the state-of-the-art methods related to two key steps in a MER system, including feature extraction and multimodal information fusion; (4) The research challenges and open issues in this field are discussed, and promising future directions are given.
* ### Emotion recognition from unimodal to multimodal analysis: A review
2023, Information Fusion
Show abstract
The omnipresence of numerous information sources in our daily life brings up new alternatives for emotion recognition in several domains including e-health, e-learning, robotics, and e-commerce. Due to the variety of data, the research area of multimodal machine learning poses special problems for computer scientists; how did the field of emotion recognition progress in each modality and what are the most common strategies for recognizing emotions? What part does deep learning play in this? What is multimodality? How did it progress? What are the methods of information fusion? What are the most used datasets in each modality and in multimodal recognition? We can understand and

compare the various methods by answering these questions.

* ### EEG-based cross-subject emotion recognition using Fourier-Bessel series expansion based empirical wavelet transform and NCA feature selection method

2022, Information Sciences

Citation Excerpt :

There are two ways, verbal and non-verbal, through which emotions can be expressed. Hence, it is required to develop an HCI system that can identify emotions by using people?s facial or vocal expressions [6,7]. But, systems that use the subject?s facial expressions are unreliable as subjects involved in the experiments can fake emotions to achieve satisfactory performance.

Show abstract

Automated emotion recognition using brain electroencephalogram (EEG) signals is predominantly used for the accurate assessment of human actions as compared to facial expression or speech signals. Various signal processing methods have been used for extracting representative features from EEG signals for emotion recognition. However, the EEG signals are non-stationary and vary across the subjects as well as in different sessions of the same subject; hence it exhibits poor generalizability and low classification accuracy for an emotion classification of cross subjects. In this paper, EEG signals-based automated cross-subject emotion recognition framework is proposed using the Fourier-Bessel series expansion-based empirical wavelet transform (FBSE-EWT) method. This method is used to decompose the EEG signals from each channel into four sub-band signals. Manually ten channels are selected from the frontal lobe, from which entropy and energy features are extracted from each sub-band signal. The subject variability is reduced using an average moving filter method on each channel to obtain the smoothened feature vector of size 80. The three feature selection techniques, such as neighborhood component analysis (NCA), relief-F, and mRMR, are used to obtain an optimal feature vector. The machine learning models, such as artificial neural network (ANN), _k_ -nearest neighborhood (_k_ -NN) with two (fine and weighted) functions, and ensemble bagged tree classifiers are trained by the obtained feature vectors. The experiments are performed on two publicly accessible databases, named SJTU emotion EEG dataset (SEED) and dataset for emotion analysis using physiological signals (DEAP). The training and testing of the models have been performed using 10-fold cross-validation and leave-one-subject-out-cross-validation (LOSOCV). The proposed framework based on FBSE-EWT and NCA feature selection approach shows superior results for classifying human emotions compared to other state-of-art emotion classification models.

* ### HistNet: Histogram-based convolutional neural network with Chi-squared deep metric learning for facial expression recognition

2022, Information Sciences

Citation Excerpt :

In [14] (our previous work), a histogram distance metric learning has been proposed for facial expression recognition using histogram-based hand-crafted features such as LBP; however, in this study, a new CNN is proposed for facial expression recognition which learns histogram feature extraction and classification steps simultaneously. In the past few years, several deep learning based methods have been proposed for facial expression recognition [3,15?33]. In this regard, Yang et al. [16] employed the encoder of a generative adversarial network to separate expression related features from identity-related ones in the input image followed by performing facial expression recognition on the expression-related features using a CNN-based classifier.

Show abstract

Facial expression recognition is a challenging problem in machine learning. There is much research has been conducted in this field; however, the accuracy of facial expression recognition, especially in uncontrolled conditions, needs much improvement. In this paper, a deep histogram metric learning in a Convolutional Neural Network (CNN) is presented for facial expression recognition. The proposed CNN utilizes a histogram calculation layer to provide statistical description of feature maps at the output of the convolutional layers. To train the proposed CNN in histogram space, a learnable matrix (equivalent to the fully connected layer) is introduced in chi-squared distance equation. Then, the modified equation is used in the loss function. The recognition rates of the proposed CNN for seven-class facial expression recognition on four well-known databases including CK+, MMI, SFEW, and RAF-DB are 98.47%, 83.41%, 61.01%, and 89.28%, respectively. The results show superiority of the proposed CNN compared to the state-of-the-art methods.

* ### Emotion Recognition Using Different Sensors, Emotion Models, Methods and Datasets: A Comprehensive Review
2023, Sensors
* ### Masked Face Emotion Recognition Based on Facial Landmarks and Deep Learning Approaches for Visually Impaired People
2023, Sensors

View all citing articles on Scopus

1
Approaches based on image processing, pattern recognition, or classical artificial intelligence methods, whose task of feature selection is designed beforehand by human experts aiming to extract a given set of chosen features to train and build a recognizer (also said hand-crafted features), in contrast to CNN?s whose features are learned from data using generic feature extractors [69].

2
A worldwide portal for scientific knowledge access, managed by the Brazilian Ministry of Education provided for the authorized institutions, including universities and research institutes, government agencies and private companies.

3
Source: https://infograph.venngage.com/p/8769/microexpressions

View Abstract

Recommended articles

* ### A survey of emotion recognition methods with emphasis on E-Learning environments
Journal of Network and Computer Applications, Volume 147, 2019, Article 102423
Maryam Imani, Gholam Ali Montazer
View PDF
* ### Cross-dataset emotion recognition from facial expressions through convolutional neural networks
Journal of Visual Communication and Image Representation, Volume 82, 2022, Article 103395
William Dias, ?, Anderson Rocha
View PDF
* ### Facial expression recognition via ResNet-50
International Journal of Cognitive Computing in Engineering, Volume 2, 2021, pp. 57-64
Bin Li, Dimas Lima
View PDF
* ### Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods

Informatics in Medicine Unlocked, Volume 20, 2020, Article 100372
Aya Hassouneh, ?, M. Murugappan
View PDF
 * ### Boosted NNE collections for multicultural facial expression recognition
Pattern Recognition, Volume 55, 2016, pp. 14-27
Ghulam Ali, ?, Tae-Sun Choi
View PDF
 * ### Facial emotion recognition in major depressive disorder: A meta-analytic review
Journal of Affective Disorders, Volume 293, 2021, pp. 320-328
Fernando C. Krause, ?, Michael T. Moore
View PDF
Show 3 more articles

Functional Cookies
These cookies enable the website to provide enhanced functionality and personalisation. They may be set by us or by third party providers whose services we have added to our pages. If you do not allow these cookies then some or all of these services may not function properly.
Cookie Details List?
Performance Cookies
Performance Cookies
These cookies allow us to count visits and traffic sources so we can measure and improve the performance of our site. They help us to know which pages are the most and least popular and see how visitors move around the site.
Cookie Details List?
Targeting Cookies
Targeting Cookies
These cookies may be set through our site by our advertising partners. They may be used by those companies to build a profile of your interests and show you relevant adverts on other sites. If you do not allow these cookies, you will experience less targeted advertising.
Cookie Details List?
Back Button
Cookie List
Search Icon
Filter Icon
Clear
checkbox label label
Apply Cancel
Consent Leg.Interest
checkbox label label
checkbox label label
checkbox label label
Confirm my choices