3D Approaches and Challenges in Facial Expression Recognition Algorithms—A Literature Review

applied sciences

MDPI

Review

3D Approaches and Challenges in Facial Expression Recognition Algorithms—A Literature Review

Francesca Nonis *, Nicole Dagnes, Federica Marcolin and Enrico Vezzetti

Department of Management and Production Engineering, Politecnico di Torino, 10129 Torino, Italy; nicole.dagnes@polito.it (N.D.); federica.marcolin@polito.it (F.M.); enrico.vezzetti@polito.it (E.V.)

* Correspondence: francesca.nonis@polito.it

Received: 8 August 2019; Accepted: 10 September 2019; Published: 18 September 2019

Abstract: In recent years, facial expression analysis and recognition (FER) have emerged as an active research topic with applications in several different areas, including the human-computer interaction domain. Solutions based on 2D models are not entirely satisfactory for real-world applications, as they present some problems of pose variations and illumination related to the nature of the data. Thanks to technological development, 3D facial data, both still images and video sequences, have become increasingly used to improve the accuracy of FER systems. Despite the advance in 3D algorithms, these solutions still have some drawbacks that make pure three-dimensional techniques convenient only for a set of specific applications; a viable solution to overcome such limitations is adopting a multimodal 2D+3D analysis. In this paper, we analyze the limits and strengths of traditional and deep-learning FER techniques, intending to provide the research community an overview of the results obtained looking to the next future. Furthermore, we describe in detail the most used databases to address the problem of facial expressions and emotions, highlighting the results obtained by the various authors. The different techniques used are compared, and some conclusions are drawn concerning the best recognition rates achieved.

Keywords: facial expression recognition; 3D face analysis; deep learning-based FER; 2D/3D comparison; facial action coding system; action units

1. Introduction to Facial Expression Recognition (FER)

Facial Expression Recognition is a computer-based technology that uses mathematical algorithms to analyze faces in images or video. The facial analysis is developed in three primary phases: face detection, facial landmark detection, and facial expression and emotion classification. The last step analyzes the movement of facial features and classifies them into emotion or attitude categories, also taking the name of Facial Emotion Recognition, a topic of emotion recognition that involves the analysis of human facial expressions in multimodal forms. More generally, emotion recognition is the automatic processing of human emotions, most typically from facial expressions as well as from verbal expressions, but also body movement and gestures. The acronym FER, in literature, often refers to both facial expression recognition and facial emotion recognition [1]. In this paper, it stands for Facial Expression Recognition, the recognition of emotional states based on facial expressions.

Facial expressions play an important role in expressing internal emotions and intentions and are one of the most significant non-verbal ways in daily emotional communication. Nowadays, there is a considerable demand for improving performance in facial expression recognition, due to the broad set of its potential applications, such as surveillance, security, and communication, even in real-time.

The majority of face recognition research and commercial face recognition systems typically use intensity images of the face (i.e., 2D texture data), but other approaches, which are becoming more and

more common, consist of using 3D models of the face (i.e., 3D shapes) or both 2D and 3D face data (multi-modal FER) [2].

Although significant progress has been made, most existing algorithms that use 2D features fail to solve the challenging problems of illumination and pose variations, naturally overcome by 3D approaches. For this reason, and thanks to the rapid technological development of 3D scanning, 3D FER has attracted more and more attention. A promising research direction to meet the requirements of real applications is the multi-modal 2D+3D FER, as it exists a considerable complementarity among different modalities. Data modality is not the only possible classification. In general, it is possible to distinguish two other perspectives to classify existing FER methods: expression granularity and temporal dynamics. From the first one, they are divided into recognition of prototypical facial expression (basic emotions) and detection of facial Action Units; from the second one, they are classified into still images and image sequences.

In contrast to traditional approaches, in recent years, deep-learning-based algorithms have been used for feature extraction, classification, and recognition tasks. Deep learning architectures, such as the Convolutional Neural Networks (CNN) and the Recurrent Neural Networks (RNN), have been applied to the field of computer vision, yielding state-of-the-art results in many studies with the availability of big data, including object recognition, face recognition, and FER. One of the main advantages of CNN is to enable "end-to-end" learning directly from input images, removing altogether or highly reducing the dependence on pre-processing techniques [3].

The study of facial expressions and emotions is a topic that has interested in the research community for several years. In the absence of a uniquely comprehensive and widely accepted theory, multiple approaches for facial expression recognition have emerged, and extensive and detailed surveys of FER focusing on traditional approaches research are given in [4–7]; the reader is referred to [8,9] for 3D and [10,11] for 2D earlier approaches. Recently, deep learning-based FER approaches emerged, and have been surveyed in [1,12–14].

This study aims to highlight the advantages and drawbacks of 3D methods over 2D methods, commonly thought to have the potential for greater recognition accuracy and robustness. Compared to previous works focused on this topic, our survey offers a newcomer the very current scenario and the best outcomes in the whole facial expression recognition field. We present and analyze both traditional and deep-learning available methods for static and dynamic 3D FER, particularly focus on feature-based algorithms. Here, we mainly focus on facial landmarks and feature extraction and the main differences between 2D and 3D approaches.

After this brief overview of the possible types of classification of existing 3D FER algorithms, the remainder of the paper is structured as follows. Section 2 focuses on basic emotions and action units, and some models of emotion classification are presented. Section 3 gives an overview of the conventional and deep learning-based methods for 2D, 3D, and multimodal FER, keeping the distinction between feature-based and model-based approaches. Moreover, two exhaustive tables are used to outline and organize the articles surveyed. Contributions have been chosen among the years 2010 and 2019, to provide the most up-to-date view of the latest researches. Section 4 deals with facial animation, discussing some critical references. Section 5 presents a summary of available facial databases mainly used in this research field. Sections 6 and 7 investigated, respectively, the importance of facial landmarks and the role of time in FER algorithms. In Section 8, the results of the main works of the last decade on the three-dimensional recognition of facial expressions are reported, divided between basic emotions and action units. Finally, Section 9 presents a comparison between the advantages and the drawbacks of the 2D and 3D algorithms. It also mentions how some research groups have positively addressed these disadvantages and what challenges are still open for the future. Section 10 concludes the work and summarizes the obtained results to identify the best characteristics of an excellent automatic facial expression recognition system.

2. Basic Emotions and Action Units

Ekman [15], in 1971, defined a set of six emotions that are accepted as universal: anger, disgust, fear, happiness, sadness, and surprise. Ekman and Friesen named this group the basic emotions, which are universally recognized regardless of language and culture and cannot be decomposed into smaller semantic labels. The seven characteristics of emotions identified by Ekman et al. [16] are: "presence in other primates, distinctive physiology, universal commonalities in antecedent events, quick onset, brief duration, automatic appraisal, and unbidden occurrence". Most research studies on FER have been limited to these six "cardinal" categories of emotions. However, humans make use of a much fuller range of facial expressions for everyday communication than these six, some are even combinations of these basic ones [17]. Martinez et al. state that "there are approximately 7000 different expressions that people frequently use in everyday life" [18]. Furthermore, some of the expressions can have multiple interpretations depending on the context in which they are shown [19].

Another possibility for studying facial expressions is the use of action units. An action unit (AU) is the action of muscles typically seen when an individual produces facial expressions. Defined by Ekman and Friesen in 1978 [20], the Facial Action Coding System (FACS) is given by a set of AUs to classify the movements of a distinct muscle or a muscle group activation of facial expression. Facial features are often divided into two groups, the upper face and the lower face, since facial actions in the upper part have only small interactions with facial motion in the lower one, and vice versa. The parameters associated with the upper face features describe the motion and the shape of the eyes, the brows, and the check. On the other hand, the parameters associated with the lower face features describe the motion and the shape of the lips, and the furrows in the nasolabial and nasal root regions [21]. At the beginning, they divided the muscular activity in 46 AUs; although the number is relatively small, more than 7000 combinations of action units have been observed [22], allowing to describe the details of facial expression. For example, AU 12 (lip corner puller) defines the contraction of the zygomatic major muscle, typically observed in feelings of happiness, together with AU 25 (lips part). The anger expression, on the other hand, is characterized by AU 4, AU 7 and AU 24, which cause the lowering of the eyebrows, the tension of the eyelids and the pressure of the lips respectively. Table 1 summarizes the action units for Ekman's six universal emotions.

Table 1. Prototypical action units (AUs) observed in each basic emotion category [1,17].

Emotion [1]	Action Units [2]	Description [3]
Anger	4, 7, 24	Brow lowerer, Lid tightener, Lip pressor
Disgust	9, 10, 17	Nose wrinkle, Upper lip raiser, Chir raiser
Fear	1, 4, 20, 25	Inner brow raiser, Brow lowerer, Lip stretcher, Lips part
Happiness	12, 25	Lip corner puller, Lips part
Sadness	4, 15	Brow lowerer, Lip corner depressor
Surprise	1, 2, 25, 26	Inner brow raiser, Outer brow raiser, Lips part, Jaw drop

[1] Ekman's six basic emotions. [2] Number of corresponding action units, and [3] related description.

The emotion categories can be classified into two groups: basic emotions, described above, and compound emotions, constructed as a combination of two basic emotion categories. Some examples of compound emotions most typically expressed by people are: happily surprised, sadly fearful, fearfully angry, and disgustedly surprised. It is important to note that often in the categories of composed emotions, there is a conflict between the action units because it is impossible to perform some muscular movements simultaneously, such as lip pressor (AU 24) and lips part (AU25) in happily surprised. In total, the researchers mapped twenty-one different facial emotions from Ekman's six basic emotions, plus the neutral expression [17].

However, there are other emotions, such as guilt, jealousy, pride, and shame that people may experience in everyday lives, but do not tend to show clear and distinct expressions. Ekman investigated

them at the beginning of his studies and proposed an expanded list of basic emotions, including a range of positive and negative emotions that are not all encoded in facial muscles.

Basic emotion classification is, still today, a more popular research topic than automatic AU recognition, even though the latter is independent of interpretation and more suitable for describing spontaneous facial behaviors [11]. At first, this disparity in favor of basic emotion classification [23–30] at the expense of AU recognition [31–33] in both 2D and 3D domains, partly due to lack of FACS-codes databases. The Bosphorus is the first, and the only, 3D publicly available database that contains AU annotations and provides facial action coding. Sun et al. [31] proposed an AU recognition system manually labeling eight AUs in the BU-4DFE database, but their annotations are private.

Returning on the basic emotion method, in 2014 Jack et al. [34] from the University of Glasgow reported new research, published in the journal *Current Biology*, intimating that there are only four basic emotions. They reached this conclusion by studying the facial muscles involved in every emotion, as well as the activation over time of each of them. The results obtained stated that while happiness and sadness are distinct over time, fear and surprise share some common signals, like wide-open eyes. Similarly, anger and disgust share the wrinkled nose.

Other scientists have developed the theme of basic emotions. For example, Plutchik [35,36] developed a new model, called the "wheel of emotions", where basic emotions can be expressed at different intensities and can mix to form several emotions, as shown in Figure 1. Plutchik's eight basic emotions here called primary are joy, trust, fear, surprise, sadness, anticipation, anger, and disgust. Additional emotions exist, like aggressiveness, optimism, love and submission: these are all seen as a combination of primary emotions.

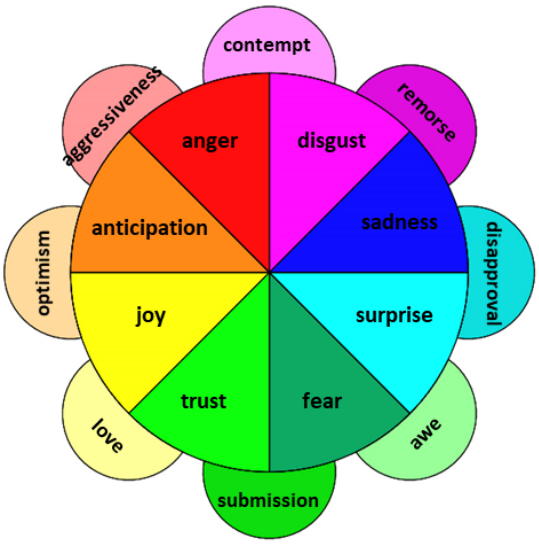

Figure 1. Wheel of emotions.

Russell [37,38] developed the Circumplex Model of Affect, presented in Figure 2. It is a circular model divided into quadrants, to show the level of valence and arousal of emotional states. The x-axis represents the continuum between pleasant and unpleasant emotions, the y-axis represents the continuum between high and low arousal emotions, while the center of the circle represents a neutral valence and a medium level of arousal.

Another theory that has gained notoriety is Parrott's approach [39], in which he identified over 100 emotions and conceptualized them as a tree-structured list, as seen from Figure 3. The first layer is composed of six primary emotions (love, joy, surprise, anger, sadness, and fear) that can be branched out into different forms of feeling, and the secondary emotions are the derivation of the primary ones instead of being a combination of them. Some of these emotions were not categorized by human expression, but rather, emotional states.

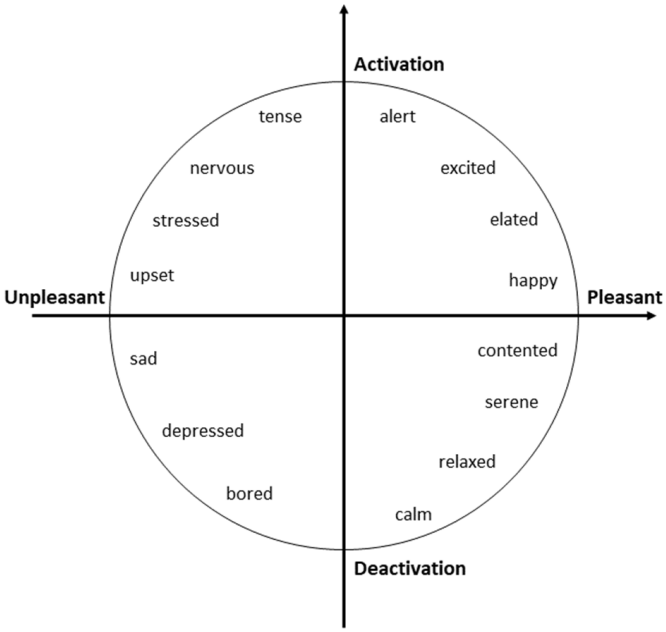

Figure 2. Circumplex Model of Affect.

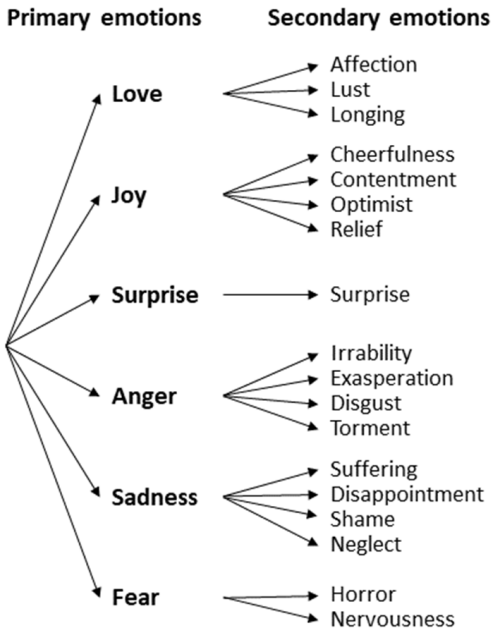

Figure 3. Parrots' classification of emotions.

In recent years, some research groups have tried to enlarge the number of emotions and facial expressions considered, testing their algorithm not only on basic emotions. For example, Fabiano et al. [40] consider in their study also embarrassment, nervousness, and pain. Zhang et al. [41,42] add to these expressions also startle, whereas Wei and Jia [43] include yawn. Others, on the contrary, have chosen to focus on a reduced set of these basic facial expressions, like Le et al. [44] who test their algorithm only on happiness, sadness, and surprise, or Sandbach et al. [45,46] whose algorithm works on anger, happiness and surprise emotions. An algorithm that works only on a small group of facial expressions has limited capabilities but can be useful for particular application fields. An in-depth analysis of the main works that use basic emotions or action units, with their respective recognition rates, is presented in Section 7, both for traditional and deep learning approaches.

3. Conventional and Deep Learning-Based Approaches

In this section, the main traditional methods for 3D facial expression recognition are presented, comparing feature-based, model-based, and multi-modal algorithms. Next, the leading deep learning techniques applied to FER for feature extraction and classification are described, distinguishing between 2D and 3D.

3.1. Feature-Based VS Model-Based Algorithms

Existing conventional approaches for FER based on 3D static data can be categorized into feature-based algorithms and model-based ones. The first category focuses on the extraction, directly from the input data, of facial surface geometric information such as curvature, spatial relations between pairs of interest landmarks (Euclidean and Geodetic distances), gradient and local shape. Features are usually calculated on the region surrounding principal facial landmarks or on the mouth and eyes that inherently contain essential information for emotion recognition. These key features that are considered closely related to expression categories, in order to perform FER are fed to various classifiers, as well as Support-Vector Machines (SVM) [47–51], Adaboost, k-Nearest Neighbors (k-NN), Linear Discriminant Analysis (LDA), Modified Principal Component Analysis (PCA), Hidden Markov Model (HMM) [44–46], Random Forest [52] or Neural Networks [51,53,54].

Feature-based methods are straightforward, but present two main drawbacks. First of all, most of these kinds of approach need a set of correctly located landmarks for feature extraction, an additional step of the process. If until a few years ago it could be considered a difficult task in real-world applications because it required the manual localization of landmarks for feature localization or surface registration, to date this step has been automated, as recently reported by Zeng [55] and Li [56]. Moreover, the performance of the model is directly related to the discriminative power of facial features adopted instead of shapes. Most of the works found in literature make use of the landmarks manually labeled on the BU-3DFE or Bosphorus dataset. Experiments have proven their excellent performance in recognizing universal expressions, but most of the features used have not yet shown enough discriminative power for distinguishing subtle facial AUs.

Thanks to the technological development in imaging and scanning, nowadays it is possible to capture 3D scans and extract geometric characteristics from certain regions around facial landmarks. Examples of some popular expression features are the distances between 3D facial landmarks [24,57–59], 3D facial curves [60], distances between locally extracted surface patches [56], facial geometry images and normal maps [61]. Another possible technique exploits 3D face descriptors, which derive from depth maps by using mathematical operators: first principal curvature, shape index, mean curvature, curvedness, etc. Descriptors are presented and fully explained in [62,63].

Table 2 gives a comparison of selected 3D FER algorithms that use facial features to perform expressions recognition. The works are listed chronologically by year of publication, and alphabetically by author.

The model-based methods, on the other hand, make use of a generic face model created using the neutral expression to determine emotions by measuring the feature vector formed by the coefficient of shape deformation. This approach needs to bring into correspondence the tracking model to 3D face scans by means of a registration step.

Zhao et al. [30] presented an automatic 3D FER approach to perform expression prediction by combining a Bayesian Belief Net (BBN) and Statistical Facial Feature Models (SFAM). They tested the developed automatic method on the BU-3DFE database, applying the SFAM for the automatic recognition of the six universal expressions reaching an average recognition rate of over 82%. In [64], Chen et al. proposed a real-time 3D model-based emotion recognition method. The 3D model of the face, which gives essential clues for improving sturdiness and allows managing large head rotations, rapid head movements, and partial facial occlusions, is restored by 2D images. More recently, Fabiano et al. [40], in 2018 developed a novel method for 3D facial expression recognition based on a statistical shape model with global and local constraints, showing that the combination of the

global shape of the face, along with local shape index-based information can be used to recognize a range of expressions, including the six basic emotions. The proposed method outperformed the state-of-the-art results achieving a classification accuracy of 99.99% on non-spontaneous data and 99.69% on spontaneous data (BP4D database). Zhen et al. [65,66] presented in their papers a novel approach to study FER problem based on the Muscular Movement Model (MMM) by combining the advantages of both feature-based and model-based methods. They formed without any manual landmark 11 muscle regions, each of which described by a certain number of geometric features (e.g., coordinate, normal and shape index) to capture shape characteristics. A genetic algorithm learns the weights of the several sections, and SVM and HMM classifiers are used for expression prediction in 3D and 4D FER, respectively.

Table 3 gives a comparison of selected 3D FER algorithms that use face model technique to perform expressions recognition. The works are listed chronologically by year of publication, and alphabetically by author.

3.2. Multi-Modal Algorithms Using 2D and 3D Data

Algorithms that combine results from 2D and 3D data did not appear until about the early 2000s. Up to date, in this area, the most straightforward approaches use combining the features obtained independently from the bi-dimensional or three-dimensional methods, such as texture information, landmarks location, facial shapes, and curvature, for recognizing expressions and measuring their intensity [12]. Experiments show that merging features obtained in different modalities helps to catch the general characteristics of facial deformation and to enhance recognition accuracy, even though the texture information is affected by the illumination and pose variations.

Table 2. Feature-based algorithms.

Author	Database [1]	2D [2]	Dynamic [3]	FLs [4]	Method [5]	Expressions [6]	Highlights
Berretti et al. [67]	BU-3DFE	No	No	20 man.	SIFT descriptors + SVM	6 BE	The neutral scans are not used
Maalej et al. [60]	BU-3DFE	No	No	68 man.	Curves and geodesic distance + SVM	6 BE	Curves used to capture the deformation on different faces under different expressions
Savran et al. [33]	Bosphorus	Yes	No	Not used	ROC curves	25 AUs	3D performs better for lower face AUs and low-intensity AUs
Soyel and Demirel [27]	BU-3DFE	No	No	11 man.	Fisher criterion + Neural Network	6 BE	Pose-invariant algorithm
Tsalakanid and Malassiotis [68]	Private	Yes	Yes	81 aut.	Geometry and curvature features	5-11 AUs	2D+3D images recorded in real-time
Venkatesh et al. [69]	BU-3DFE	No	No	Not used	Spectral flow matrices as features	6 BE	No manual intervention or Features selection step
Berretti et al. [70]	BU-3DFE	No	No	Aut.	SIFT feature + SVM	6 BE	Completely automatic approach
Le et al. [44]	BU-4DFE	No	Yes	Not used	Facial Level Curves + Hidden Markov Model	3 (happiness, sadness and surprise)	Parallelizable algorithm
Li et al. [61]	BU-3DFE	No	No	60 man.	HoG and HoS + SVM	6 BE	Local shape descriptors are used
Sandbach et al. [45]	BU-4DFE	No	Yes	Not used	3D facial geometry + Hidden Markov Model	3 (anger, happiness and surprise)	Fully automatic system
Vretos et al. [71]	BU-3DFE and Bosphorus	Yes	No	Not used	Zernike moments assures	6 BE	Real-time
Drira et al. [52]	BU-4DFE	Yes	Yes	Not used	DVF + Random Forest	6 BE	Temporal analysis of facial expressions
Li et al. [72]	BU-3DFE	No	No	Not used	Multi-Scale Local Normal Patterns + SimpleMKL	6 BE	Fully automatic approach
Rabiu et al. [73]	BU-3DFE and UPM-3DFE	No	No	32 man.	Minimum redundancy – maximum relevance + SVM	6 BE	Person and gender independent
Sandbach et al. [46]	BU-4DFE	No	Yes	-	Motion-based + Gentle Boost classifier and Hidden Markov Model	6 BE	Temporal analysis
Sandbach et al. [74]	Bosphorus and D3DFACS	No	Yes	6 man.	Depth map + APDI + SVM	22 AUs	Temporal analysis
Lemaire et al. [75]	BU-3DFE	No	No	Not used	Differential Mean Curvature Maps + Multiclass-SVM	6 BE	Face normalization step
Zeng et al. [55]	BU-3DFE	Yes	No	3 aut.	CFI and MCI + 2D texture descriptors	6 BE	Fully automatic

Table 2. *Cont.*

Author	Database [1]	2D [2]	Dynamic [3]	FLs [4]	Method [5]	Expressions [6]	Highlights
Zhang et al. [41]	BP4D	No	Yes	Not used	HMM	6 BE	Spontaneous expression data
Hayat and Bennamoun [76]	BU4DFE	Yes	Yes	Not used	Grassmannian manifold + SVM	6 BE	No computationally expensive pre-processing step and any user interventions
Li et al. [77]	FRGC v2.0, Bosphorus, BU-3DFE and 3D-TE	No	No	Not used	MSMC-LNP	6 BE	Different expression intensity levels
Srivastava and Roy [47]	BU-3DFE	No	No	83 man.	3D flow of facial points + SVM	6 BE	Neutral 3D facial model for each subject
Xue et al. [78]	BU-3DFE	No	No	5 aut. + 35 heuristic points	Depth features + SVM	6 BE	Fully automatic
Yurtkan and Demirel [79]	BU-3DFE	No	No	83 man.	3D geometrical facial feature point positions + SVM	6 BE	Entropy is used to extract the most discriminative features
Zhang et al. [42]	BP4D	Yes	Yes	83 man.	3D dynamic facial expression descriptors	8 (anger, disgust, fear, happiness, sadness, startle, embarrassment and pain) + 27 AUs	Spontaneous expressions
Azazi et al. [48]	BU-3DFE and Bosphorus	Yes	No	20 aut.	SURF + SVM	6 BE	Facial muscular movements exploiting
Jan and Meng [49]	BU-3DFE and Bosphorus	Yes	No	83 + 22 man.	Geometric and textured domains + SVM	6 BE	2D+3D to increase the overall performance
Li et al. [56]	BU-3DFE and Bosphorus	Yes	No	49 aut.	HSOG texture and SIFT + Surface and curvature	6 BE	Multi-order gradient-based local texture combined with shape descriptors
Li et al. [80]	BU-3DFE	Yes	No	Not used	Depth and texture information	6 BE	Fully automatic
Mao et al. [81]	UJSKED, FaceWarehouse, and real-time video sequences	Yes	No	Not used	Features of animation units (AUs) and feature point positions (FPPs)	6 BE	A real-time approach based on data captured by Kinect.
Yang et al. [82]	BU-3DFE	No	No	Not used	Shape index map	6 BE	Fully automatic 3D FER method
Huynh et al. [53]	BU-3DFE	Yes	No	Not used	Shape and texture descriptors + CNN	6 BE	Deep learning technique for 3D FER
Li et al. [83]	BU-3DFE	Yes	No	Not used	Different kinds of range images	6 BE	Range images do not lose primitive discriminative information for recognition

Table 2. *Cont.*

Author	Database [1]	2D [2]	Dynamic [3]	FLs [4]	Method [5]	Expressions [6]	Highlights
Derkach and Sukno [84]	BU-3DFE	No	No	68 man.	3D geometry and features decomposition in their special frequency components	6 BE + 17 AUs	3D geometry analysis extended from a curve-based representation into a spectral representation
Hussain et al. [50]	BU-4DFE	No	Yes	12 man.	Facial distances + SVM and Neural Network	2 (happiness and sadness) + 7 AUs	Facial distances are used to localized active muscles
Savran and Sankur [85]	BU-3DFE	No	No	Not used	Dynamic feature extraction mechanism	6 BE	Non-rigid registration incorporated in face-model-free analysis
Hariri et al. [86]	BU-3DFE	No	No	92 man.	Covariance matrices	6 BE	Covariance descriptors allow the combination of different types of features
Dong et al. [87]	Bosphorus	Yes	No	Not used	Global and local features + SFMs	16 AUs	A sign-based approach for FER
Binghua et al. [54]	Private	Yes	Yes	Not used	LBP features + CNN	6 BE	Testing on a private database, recorded by RealSense camera
Jan et al. [51]	BU-3DFE	Yes	No	49 aut.	2D texture + 3D depth map + CNN and SVM	6 BE	2D texture and 3D depth maps information showed a consistent RR improvement

[1] Denotes the database, or a portion of it, used for the study; [2] indicates the use of 2D data associated, and [3] the use of temporal information (3D image sequences). [4] The presence, number, and type of facial landmarks: manual or automatic. [5] Methods: scale-invariant feature transform (SIFT), support vector machines (SVM), receiver operating characteristic (ROC), neural network (NN), histogram of mesh gradient (HoG), histogram of shape index (HoS), deformation vector field (DVF), random forest (RF), azimuthal projection distance image (APDI), conformal factor image (CFI), mean curvature image (MCI), hidden Markov model (HMM), multi-scale and multi-component local normal patterns (MSMC_LNP), speed up robust features (SURF), feature point positions (FPPs), convolutional neural network (CNN), statistical feature model (SFM). [6] Expressions: basic emotions (BE) are anger, disgust, fear, happiness, sadness, and surprise.

Table 3. Model-based algorithms.

Author	Database [1]	2D [2]	Dynamic [3]	FLs [4]	Method [5]	Expressions [5]	Highlights [6]
Ramanathan et al. [88]	Private	No	No	Not used	Morphable Expression Model	3 (happiness, sadness and anger)	Morphable 3D model
Mpiperis et al. [26]	BU-3DFE	No	No	Man.	Elastically deformable model algorithm + bilinear models	6 BE	Face recognition invariant to facial expressions
Gong and Wang [28]	BU-3DFE	No	No	-	BFSC and ESC + SVM	6 BE	Single 3D face without any manual assistance
Zhao et al. [30]	BU-3DFE	No	No	19 aut.	SFAM + BBN	6 BE	Flexibility of the method, applicable to real-use cases
Zhao et al. [32]	Bosphorus	No	No	19 aut.	SFAM	16 AUs	Extended statistical facial feature model
Fang et al. [89]	BU-3DFE and BU-4DFE	No	Yes	12 aut.	AFM + PDM	6 BE	First approach that registers 4D data and bring them into dense correspondence without the need of texture information
Chen et al. [64]	AVEC 2012	Yes	Yes	42 inners	Random forest-based	6 BE	Real-time
Zhen et al. [66]	BU-3DFE and BU-4DFE	No	Yes	Not used	MMM + geometry features	6 BE	Introduction of a novel muscular model
Wei and Jia [43]	RGB-D and KinectFaceDB	Yes	No	Not used	FFPs and AUs + random forest	4 (smile, yawn, angry and sad) + 2 (smile and yawn)	Sensor Kinect for real-life applications
Fabiano et al. [40]	BU-4DFE	No	Yes	83 man.	Random forest	6 BE + embarrassment, nervousness and pain	Spontaneous and non-spontaneous data

[1] Denotes the database, or a portion of it, used for the study; [2] indicates the use of 2D data associated, and [3] the use of temporal information (3D image sequences). [4] The presence, number, and type of facial landmarks: manual or automatic. [5] Methods: basic facial shape component (BFSC), expression shape component (ESC), support vector machine (SVM), statistical facial feature model (SFAM), annotated face model (AFM), point distribution model (PDM), muscular movement model (MMM), facial feature points (FFPs), action units (AUs), random forest (RF). [6] Expressions: basic emotions (BE) are anger, disgust, fear, happiness, sadness, and surprise.

In 2014 Hayat and Bennamoun [76] used the 2D texture information incorporated along-with 3D features. After learning the SVM models from the 2D and 3D data separately, they performed the classification and undertook a fusion of results separately obtained for performance improvement. Hence, a feature dimensionality reduction technique is used to reduce the size of the feature vector while retaining its quality. Jan and Meng [49] in 2015, proposed to fuse the key features obtained from the geometric and textured domains, to investigate how the overall performance is affected. Also in this work, a feature dimensionality reduction method is used, before applying machine learning techniques; merging the many elements produced by the algorithms can result in a large feature vector which can slow down the system. Accordingly, in 2018, they continued to use a textured 3D face scan, since both detailed 3D geometric features and 2D texture information can provide key cues for FER. Experiments are conducted on the BU-3DFE database, demonstrating the effectiveness of combing texture and depth cues.

From a general point of view, multimodal techniques may deal with various data modalities, originating from other sources of information such as thermal image acquisitions, voice data, brain signals or cardiovascular activity, and context information. Other visual cues that could be combined with standard features for improving methods that detect facial expressions are eye gaze, head orientation, the motion of the head and body, mouth fidgeting, and FEs frequency or duration. Some researchers have already demonstrated that context can improve emotion recognition, as well as the body posture that becomes more important as the FE is more ambiguous. Meanwhile, also the voice can provide indications on emotions through acoustic properties such as pitch range, rhythm, and amplitude or duration changes.

Marginal studies have been conducted on the combination of 3D facial data and physiological or acoustic cues. Investigations on the possible integration of visual and non-visual modalities, like physiological data coming from wearable devices, could be a possible branch of research for the coming years.

3.3. Deep Learning Applied to FER

Deep learning [90,91] is a machine learning technique theorized for the first time in the 1980s [92] but only lately considered in practice because it needs a large amount of labeled data and considerable processing power. In recent years, deep learning techniques have been employed successfully in a wide range of tasks including the recognition of facial expressions, a difficult problem for machine learning, since people can show their feelings in very different ways. Deep Neural Networks (DNN) have been used to classify images of the human face into emotion categories in an end-to-end approach and overcome the difficulties of the traditional methods, reaching a level of recognition accuracy higher than that of man in some activities.

Automatic deep FER includes three different steps: pre-processing, deep feature learning, and deep feature classification [13]. Pre-processing is necessary before training the neural network to learn essential features for recognition; some examples are image cropping, rotation correction, data augmentation, and spatial normalization. Completed this preliminary phase, one of the deep learning techniques, particularly CNN or RNN, is applied for FER, and the given face is classified into one of the basic emotion categories. Using deep networks, feature extraction and their classification are performed in an end-to-end way [93], while in the traditional methods these two last phases are independent. Alternatively, the neural network can be used for feature extraction only, and then independent classifiers, such as support vector machine (SVM), can be applied to the extracted representations.

3.3.1. Convolutional Neural Networks (CNN)

Most of the deep learning methods use neural network architectures, and one of the well-known and used to recognize objects or faces is the Convolutional Neural Network (CNN). The three main factors that have contributed to the use of CNNs for deep learning are the elimination of manual

extraction of features, the cutting-edge recognition results and the possibility of retraining existing networks for other recognition activities [1].

A CNN is composed of an input layer, an output layer and up to 150 hidden layers in the middle (for this reason, it is called "deep") [94], while traditional neural networks have only 2–3. These layers are intended to find and learn features, repeating convolution, activation or Rectified Linear Unit (ReLU), and pooling operations; each hidden layer increases the complexity of the features of the image [95]. Below is a summary of the three main steps:

- Convolution is a mathematical operation that recalls the functioning of a scanner and consists of applying a smaller matrix called kernel to the input matrices, the images. Each of the convolutional filters is translated with a specific translation stride and activates certain features of the pictures.
- After a convolutional layer, in most ConvNets, there is an activation function. The main purpose of this layer is introducing a non-linearity in the system, using non-linear functions, such as tanh, sigmoid, and ReLU. The Rectified Linear Units returns 0 if it receives any negative input, but for any positive value x it returns that value back, so it can be written as f(x) = max (0, x): only activated features are transferred to the next layer.
- The pooling performs a non-linear subsampling, reducing the size of the output matrices and the number of parameters that must be learned by the network. The most common operations are MaxPooling and AveragePooling.

After the learning phase of the features, the classification phase begins. To provide the classification output, a fully connected layer, and a classification layer, such as SoftMax, are used. The penultimate layer generates a vector of dimensions equal to the number of classes, containing the probability for each one and computing the class score on the entire original image. The last layer of the CNN architecture assigns the decimal probabilities to each class, in a multi-class problem.

3.3.2. RNN-LSTM

A Recurrent Neural Network (RNN) is a type of advanced artificial neural network commonly used in speech recognition. While in a traditional neural network all inputs and outputs are independent of each other, RNNs use sequential information. This class of neural networks shows temporal dynamic behavior and uses its internal memory to calculate output depending on the earlier computations.

A variation of the recurrent net, the so-called Long Short-Term Memory (LSTM), was proposed in the mid-90s by the German researchers Hochreiter and Schmidhuber [96]. These networks use three gates to regulate and control the cell state and thus overcome the issue of the vanishing gradients and exploding problems that are common in training RNNs. LSTMs can keep the memory for a more extended period than RNNs, enabling them to model long-term dependencies in a sequence; common areas of application include speech recognition, language modeling, and video analysis (video-based expression recognition tasks).

3.3.3. State of the Art

In 1872 Darwin published the book "*The Expression of Emotion in Man and Animals*" [97], which first gave rise to the recognition and study of emotions. Subsequently, the researchers Izard and Ekman, inspired by Darwin, have conducted important studies on facial expressions, leading to significant advancements. The first study that presents an algorithm developed to perform Facial Emotion Recognition was conducted by Bartlett et al. in 2003. The paper [98] tried to make a method capable of automatically detecting frontal faces in a video stream and classifying facial expression as either anger, disgust, fear, happiness, sadness, surprise, and neutral. The real-time face-detection system, a development of Viola-Jones' work [99], was tested on the Cohn-Kanade dataset of posed facial expressions. Later, several types of conventional approaches for automatic FER were developed. These algorithms are able to accomplish the task, detecting the face region and extracting geometric features (such as facial landmarks, face shape and its components), appearance features (such as pixel

intensities and texture of the face), or a hybrid of geometric and appearance features [1,100]. The works presented in the literature can be divided according to the feature representations (static images or dynamic image sequences), the deep-learning-based algorithm (CNN or hybrid CNN-RNN), and the type of data (2D, 3D or 2D+3D). In this survey, the works of the last decades are presented, subdividing them into two-dimensional, three-dimensional, or multi-modal methods.

a. 2D FER

Matsugu et al. [101] developed the first facial expression recognition model with the property of subject independence as well as translation, rotation, and scale invariance. They employed a CNN, more efficient and compact than the Fasel's model [102] presented the year before, to "detect smiling or laughing faces based on differences in local features between a normal face and those not." In his paper, Fasel proposed two independent CNNs, one for facial expressions, and the other for face identity recognition, combined by a Multilayer Perceptron (MLP). FER systems are traditionally evaluated in a subject-independent way or with a cross-database approach, rarely in a subject-dependent way. This last technique, in which only a single person builds each classification model, is used in limited cases [103] where greater importance is given to recognition accuracy than generalization. The subject-independent manner, on the other hand, trains the classifier on a subset of images and evaluates it on the remaining part of the same database, while the cross-database method trains and evaluates the classifier on all of the pictures in different databases. Cross-database tasks are more complicated to satisfy as each database presents different settings of illumination, pose, resolution, etc. Deep neural networks are able to recognize subtle features, and for this reason, perform well in flexible learning tasks reaching high levels of accuracy in FER problems [104].

In 2014 there was one of the largest and the most challenging computer vision competition, ImageNet 2014 challenge. GoogLeNet, a new architecture of CNN inspired by LeNet [105] and implemented by Google, won the classification and object recognition challenges, achieving a top-5 error rate of 6.67%. In the following years, this same network, trained initially to distinguish between 1000 objects, was re-trained for different objectives, including facial recognition [106]. The performance of 2D facial expression recognition is degraded by pose variation and illumination problems, not present with a three-dimensional approach.

b. 3D FER

The technological progress in recent years has enabled us to achieve better performance in various areas, including facial expression recognition. The acquisition of high-quality 3D facial scans overcomes the problems of illumination and pose changes and contains more information about the movements of facial muscles induced by expressions.

In 2006 the first database including three-dimensional face models, the BU-3DFE, was made public, and many studies were performed. In addition to traditional approaches, categorized into model-based ones as well as feature-based ones, some researchers used deep networks in tasks of computer vision, often reaching higher accuracies.

The multimodal 2D+3D is a common approach for facial expression recognition, widely used in literature. Savran et al. [33] showed that in general 3D data perform better than 2D data, especially for lower face AUs, but with the fusion of two modalities higher detection rates are achieved (97.1 %). Depth maps (3D meshes) are fused with other 2D maps, such as texture maps [2,51,107], curvature maps [2,108] and normal maps [2,107].

Li et al. conducted two studies with Deep Learning-based methods. In the first [109], they generated the deep representation of the 3D face by putting the geometry map, normal maps, normalized curvature map, and texture map into a pre-trained CNN. Facial geometric attributes are then classified using SVM, achieving the facial expression prediction. Two years later, in their second work [2], they built a new deep CNN model for subject-independent multimodal 2D+3D facial expression recognition and trained it on six types of facial attribute maps. The single end-to-end training framework increases the

accuracy and achieves the best results. "This is the first work of introducing deep CNN to 3D FER and deep learning-based feature level fusion for multimodal 2D+3D FER" [2].

Oyedotun et al. [110] proposed a CNN model able to learn more discriminative features from both RGB and depth map latent representation for joint learning of facial expressions. Yang and Yin [108] presented a 3D facial expression recognition algorithm using CNNs and landmark clues, with the sole use of 3D geometrical facial models.

Chen et al. [111], in their paper, directly used 3D facial point-clouds for FER based on a fast and light manifold CNN model, FLM-CNN. Compared to the existing ones, the method proposed is better in speed and feature extraction, achieving state-of-art performance on BU-3DFE and showing a high tolerance to pose changes.

Some of the last year's works are [112–117]. Others are those of Jan et al. [51] and Zhu et al. [118]. Jan et al. [51] designed a novel system for 3D FER based on accurate facial parts extraction according to localized facial landmarks, and deep feature fusion of facial parts. With the use of facial parts, they achieved better performance than using the entire face. Finally, a multi-class SVM classifier is adopted for facial expression prediction. Zhu et al. [118] introduced a novel deep learning approach to 3D FER, namely Discriminative Attention-based Convolution Neural Network (DA-CNN), to capture more comprehensive expression related representations. They conducted several experiments to prove the effectiveness of their method, and state-of-art results are achieved for both 3D FER and multi-modal 3D+2D FER.

Lately, some researchers have started using 4D (3D dynamic) data [3,119,120]. Sandbach et al. [9,46] proposed a method that exploits 3D motion-based features for dynamic FER in the BU-4DFE database, performing a six-way classification, the six basic emotion. In the second paper they surveyed the progresses in 3D and 4D face acquisition and presented available databases for static and dynamic 3D FER.

4. Facial Animation

Facial animation is an area of computer graphics that consists of methods and techniques for generating and animating models of a human, an animal, or a fantasy character face. Parke made the first efforts to represent and animate three-dimensional faces using computers in 1972 [121]. Computer-based facial expression modeling and character animation is not a new endeavor but has been considerable growth of interest in recent years.

Following the success for describing movements of facial muscles of the FACS and Action Units developed by Ekman and Friesen in 1978 [20], Platt [122] and Brennan [123] in the early-1980s produced, respectively, the first physically based muscle-controlled face model, and techniques for facial caricatures. Their studies gave birth to the first animated human character able to express emotion through facial expressions and body movements.

Different techniques exist for the generation of facial animation data: marker-based motion capture [124,125], markerless motion capture, audio-driven, and keyframe animation. Marker-based techniques are widely used for real-time facial animation thanks to their robustness but are not useful for retrieving fine-scale dynamics and require specialized sensors. To simplify the motion capture process, techniques without requiring markers or specialized tracking hardware came out leveraging depth sensors and structured-light based devices [126]. The researchers demonstrated the ability to track detailed facial expressions from a 3D sensor in real-time, but the system required an extensive set of pre-processed facial expressions and consequently a lengthy training session. The year later, Li et al. [127] used the same system replacing the Principal Component Analysis (PCA) model with an optimized rig, in order to reduce training poses and enable retargeting.

Real-time 3D low-cost sensors such as Microsoft's Kinect favored the development of new methodologies that simplify the procedure [128,129]. In [128], a user-specific dynamic expression model is created in an offline preprocessing step. The novel face tracking algorithm combines 2D color image and 3D depth map, simultaneously captured by Kinect, in a systematic way with user-specific

blendshapes. This proposed method achieves high-quality performance-driven facial animation in real-time with a low-cost and markerless acquisition system, and more robust and accurate results than previous video-based methods. Li et al. [129] proposed a real-time facial animation system with adaptive tracking without any training or expression calibrations, achieving superior tracking fidelity than existing state-of-the-art techniques.

Other works in facial animation are [130–132]. Mousas and Anagnostopoulos [130] presented in their paper a novel mesh deformation method to automatically transfer facial blendshapes from a reference to a source face model. Parameters such as elasticity, mesh curvature descriptors were not considered, but the presented method achieves a lower error rate than the previous methodologies. Wei and Deng [131] studied speech animation in real-time based on live speech input, synthesizing but maintaining the realism of facial animation. Ouzounis et al. [132] presented a methodology that provides the ability to efficiently transfer facial animations to face models with different morphological variations.

Recently, Ma and Deng [133,134] presented a "complete pipeline to photo-realistically transform the facial expression for monocular video in real-time" [133] and a "real-time, automatic, geometry-based method for capturing fine-scale facial performance from monocular RGB video" [134]. Facial animation applications include communication, education, and scientific simulation area, even if the primary use remains animation films and computer games.

5. Databases

In the literature, there are numerous databases related to FER for comparative and extensive experiments; some of them are used in particular for conventional FER approaches with decision methods, others for FER systems based on deep learning with recognition algorithms. Traditionally, human facial emotions have been studied using 2D data, static or dynamic, with many difficulties attached. A 3D-based analysis of facial emotions will facilitate handling significant pose variations and subtle facial behaviors but has specific problems such as a high computational cost. The databases mainly used in this research field are described below and summarized in Table 4.

To date, various databases for 3D facial expression recognition exist and have been employed by the research community for the evaluations of the developed algorithms. Of these, only three publicly available have been designed specifically for emotion analysis, having datasets displaying the six basic emotions or different AUs of the FACS, and are BU-3DFE, BU-4DFE, and Bosphorus. The databases listed above are not the only ones to contain facial expressions; other public databases, for example, FRGC v2.0 and GavabDB, present a set of expressions variations, but incomplete or with an irregular distribution. The main databases with 3D images and video sequences used for the study of facial expressions are described below.

Table 4. A selection of face datasets employed in recent studies on expressions and AUs recognition.

Name	Year	Data [1]	Datasets [2]	Expressions	Occlusions [3]	Head Poses [4]	# of Landmarks [5]
JAFFE [135]	1999	2D images	213 (10)	6 basic emotions + neutral	No	No	-
FRGC ver2.0 [136]	2005	3D scans	4950 (577)	6 basic emotions	No	No	-
MMI [137]	2005	2D RGB images + 2D videos	740 + 2900 (75)	6 basic emotions + neutral + AUs	No	No	-
BU-3DFE [138]	2006	3D facial models	2500 (100)	6 basic emotions $*$ 4 levels of intensity + neutral	No	No	83
Bosphorus [139]	2008	2D RGB images + 3D data	4666 (105)	6 basic emotions + neutral + 28 AUs	Yes (4)	Yes (13)	24
BU-4DFE [140]	2008	3D video sequences	606 (101)	6 basic emotions + neutral	No	No	83
CK+ [141]	2010	2D video sequences	593 (123)	6 basic emotions + contempt and neutral + AUs	No	No	-
D3DFACS [142]	2011	2D+3D static + dynamic	519 (10)	19-97 AUs individually + combination	No	No	47
RGB-D [143]	2012	2D RGB images + depth maps	1581 (31)	Neutral, smile, sad, yawn and angry	No	Yes (17)	-
UPM-3DFE [144]	2012	3D images	350 (50)	6 basic emotions + neutral	No	No	32
BP4D [41,42]	2013	3D video sequences	328 (41)	8 facial expressions + 27 AUs	No	No	49
KinectFaceDB [145]	2014	2D RGB images + 2.5D depth maps + 3D point clouds + RGB video sequences	936 (52)	Neutral, smile and yawn	Yes (3)	Yes (2)	6
FERG-DB [146]	2017	2D images	55767 (6)	6 basic emotions + neutral	No	No	-

[1] Denotes the type of data, [2] indicates the number of samples and, in brackets, the number of subjects. Presence and number of [3] occlusions of eyes and mouth, [4] various poses of the head acquired for each subject in a systematic way, and [5] facial landmarks.

The most common for 3D FER systems is the Binghamton University 3D Facial Expression (BU-3DFE) database [138], the first one to become publicly available. It consists of raw 3D facial scans of 100 subjects, 56 females and 44 males, with different ethnic and racial ancestries, and ranging age from 18 to 70 years old. Each person performed the expressions corresponding to the six basic emotions, acquired at four levels of intensity (low, middle, high, and higher) using 3D sensors, along with a neutral expression. In total, therefore, the database contains 2500 3D facial expression models, 25 for each subject, plus the associated texture images captured at two views (about +45° and −45°). The manually annotated dense landmark set is provided with the release, contributing to its diffusion and use.

The BU-4DFE [140] is a 3D dynamic facial expression database, built to analyze the facial behavior in a dynamic 3D space. There are 58 female and 43 male subjects with a variety of ethnicities and an age range of 18–45, each of one performed gradually the six universal emotions starting and then finishing with the neutral expression. The sequences last approximately four seconds and are captured at a video rate of 25 frames per second; the database contains 606 3D facial expression sequences, giving a total of over 60thousand frame models.

The Bosphorus database [139] was designed for research on 2D and 3D FER tasks. There are 105 subjects, one-third of whom are professional actors and actresses, for a total of 4666 facial data acquired using a structured-light based 3D system. The Bosphorus is the only publicly available database to date that contains 3D face scans for AUs, including intensity and asymmetry codes for each AU. Moreover, there are scans in various poses, expressions and realistic occlusion conditions, such as glasses or hand around the mouth, or with mustache and beard. In this case, as in the previous ones, the data was collected in a controlled environment in which the subjects were instructed to perform specific emotions; therefore, the BU-3DFE, the BU-4DFE, and the Bosphorus are all non-spontaneous databases.

The Binghamton-Pittsburgh 3D Dynamic Spontaneous (BP4D), developed by Zhang et al. [41,42], actually is the only database that considers 3D dynamic spontaneous facial expressions [147]. The data was collected in the course of social interactions between the participants and an interviewer, with the use of 8 specific tasks that elicit different emotional expressions: happiness or amusement, sadness, surprise or startle, embarrassment, fear or nervous, physical pain, anger or upset, and disgust. Hence, the BP4D is suitable to design and test methods dealing with real-world scenarios.

Some works use the EUROKOM Kinect Face Dataset [145] or other private RGB-D database, explored extensively for various applications. The EUREKOM Kinect Face Dataset consists of face images of 52 people captured in nine states, including various facial expressions, with a Kinect RGB-D camera, a sensor characterized by low data acquisition time and low-quality depth information. The data are acquired in different modalities (2-D, 2.5-D, 3-D), and all the images are provided in the three sources of information: the RGB color image, the depth map, and 3D.

6. Landmarks

Facial landmarks are key points that are used to extract facial information, such as identity, expression, and emotion, in computer vision tasks. They are shared by all human faces and contain a geometric and biometric meaning.

Landmarks were initially introduced by Farkas [148] and extensively applied to different disciplines involving the human face [149]. Extensive handbooks and reference works have been written within the context of the anthropometry discipline, such as "*Anthropometry of the head and face*" [150], and "*Three-Dimensional Cephalometry: A Color Atlas and Manual*" [151], which report the truthful and medical meaning of each landmark. Up to 59 points can be identified in the human face, but the most famous and used are only 20.

The localization of the landmark points is an essential step in many facial expression recognition [152] and head pose estimation [153] algorithms. Most of the studies in literature used manual landmarks, usually exploiting the pool of landmark coordinates stored in the publicly available databases.

The Bosphorus database, for example, provides 2D and 3D coordinates of 24 labeled facial landmarks, which are: outer, middle, and inner left/right eyebrow, outer and inner left/right eye corner, nose saddle left/right, left/right nose peak, nose tip, left/right mouth corner, upper and lower lip outer/inner middle, chin middle, and left/right ear lobe. More recently, some researchers developed algorithms to automatically detect the locations of the landmarks points of the human face on images or videos without any assistance by the user [48,51,55,56,68,70].

However, facial key point detection is challenging due to significant variations in facial appearance under different facial expressions, head poses, environment and lighting conditions, and motion patterns, in particular for video sequences. Furthermore, facial occlusions can cause loss of landmarks information. The mouth region, like the lips, mainly presents this difficulty because it has high degrees of freedom in the movements. In more rigid parts, like eye corners, facial landmarks are easier identified automatically. For these reasons, Zeng et al in [55] proposed a general and fully automatic framework for 3D FER that only needs three main facial landmarks: nose tip, left and right inner eye corners. Similarly, Xue at al. [78] presented in 2014 a fully automatic method, including the detection of landmarks, for 3D FER. The algorithm, based on depth features, detect the four eye corners and nose tip in real-time, and then, from these five points, define another 25 heuristic points for facial expressions analysis.

In order to support the researchers' choices to use a set of landmarks instead of another, statistical analysis or assumptions are often used. For example, Derkach and Sukno [84] decided to discard 15 of the 83 manually labeled landmarks in the BU-3DFE database corresponding to the silhouette contour because they have arguably little validity in a 3D setting. Hence, only a subset of 68 landmarks laying within the face area was considered.

The motivation to use a small number of facial landmarks is a significant reduction in recognition time, a critical issue in real-time applications, where the expression must be determined almost immediately. Instead, more features will result in added tracking time and greater complexity in the classifier.

7. Role of Time

Facial expressions play an important role in the recognition of human emotions. At first, facial expressions analysis interested only psychologists, but later it had a wide diffusion in the field of scientific research.

In the past, many techniques, including neural networks, have been applied to recognize facial expressions in still images. The strategies developed had some limitations; one of these is the possibility to view only an instant of the expression, and usually, the still images capture the moment at which it is most marked, without considering that in their daily lives people show the apex of their facial emotions only for particular cases and for very brief periods of time. The subtle facial expressions that are used for most of the communication activities are not identifiable in still images but became visible in video sequences. For this reason, dynamic facial expressions became increasingly important in recent years. Temporal dynamics of facial expression provide additional relevant information that is not available in static 2D or 3D images. Indeed, an emotion lasts from 250 milliseconds to 5 seconds [154]; a dynamic method may be useful for evaluating the intensity level of muscle activities and for classifying emotions.

Referring to the work of Schmidt and Cohn [155], a majority of spontaneous smiles reach onset faster and show more action units than the posed smiles. In contrast to the 18 types of smile described by Ekman, during spontaneous smiles "the appearance of AU 12 was either simultaneous with or closely followed by one or more associated action units", such as AU 6 (cheek raiser), AU 15 (lip corner depressor) or AU 17 (chin raiser). The dynamic properties of spontaneous human facial expressions have significant implications for human-computer interaction to describe naturalistic interactive behavior.

Most of the works that choose video sequences as input use 2D FER algorithms; only a few attempts have been made to analyze facial expression using 4D data (i.e., 3D videos), partly due to the lack of public databases having three-dimensional dynamic sequences. Nowadays, thanks to the incredible development of capturing, reconstruction, alignment, and tracking techniques, dynamics 3D recordings are increasingly used in expression analysis research [9].

The public availability of the BU-4DFE database contributed to this innovation process. This database includes 3D sequences of the six universal emotions, but unfortunately, it does not provide AU annotations.

8. Results and Challenges for 3D FER

The most studied and tested facial expressions are the six basic facial expressions indicated by Paul Ekman as universal, but also other emotions have been considered, to develop sounder algorithms able to deal with any occlusions. Table 5 summarizes the performance results (recognition rates of the respective FER method) obtained by the different authors and reported in their papers through a confusion matrix. The overall accuracy is indicated, as well as the results obtained for the various types of expressions

Table 5. Recognition rates (RRs) in the presence of each type of facial expression and the overall average result obtained. For every kind of emotion, the best result is written in bold characters.

Author	Year	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Overall
Berretti et al. [67]	2010	81.7	73.6	63.6	86.9	64.6	**94.8**	**77.53**
Maalej et al. [60]	2010	96.50	**97.00**	94.50	94.67	96.00	97.83	**96.08**
Soyel and Demirel [156]	2010	91.7	93.9	90.0	94.1	90.8	**98.9**	**93.23**
Venkatesh et al. [69]	2010	87.17	84.45	73.47	**100**	74.14	94.18	**85.57**
Zhao et al. [30]	2010	79.20	87.60	79.20	93.3	90.8	**93.37**	**87.25**
Berretti et al. [70]	2011	81.7	73.6	63.6	86.9	64.6	**94.8**	**77.53**
Le et al. [44]	2011	/	/	/	**95**	91.67	90	**92.22**
Li et al. [61]	2011	76.8	78.1	73.2	91.4	75.5	**94.5**	**81.6**
Sandbach et al. [45]	2011	77.71	/	/	**89.37**	/	85.40	**83.03**
Vretos et al. [71]	2011	/	/	/	/	/	/	**/**
Drira et al. [52]	2012	93.11	92.46	91.24	**95.47**	92.46	94.53	**93.21**
Fang et al. [89]	2012	92.42	91.67	81.06	**98.48**	88.64	93.94	**91.04**
Li et al. [72]	2012	77.92	77.17	69.25	**93.17**	70.67	92.67	**80.14**
Rabiu et al. [73]	2012	86.8	95.9	89.7	97.6	85.5	**98.7**	**92.37**
Sandbach et al. [46]	2012	51.92	62.71	46.15	75.28	68.97	**82.56**	**64.60**
Lemaire et al. [75]	2013	74.1	74.9	64.6	89.8	74.5	**90.9**	**78.13**
Zeng et al. [55]	2013	59.58	63.67	51.00	84.58	59.58	**90.50**	**68.15**
Zhang et al. [41]	2013	/	/	/	/	/	/	**/**
Hayat and Bennamoun [76]	2014	92.71	93.46	90.09	98.00	93.12	**98.67**	**94.34**
Li et al. [77]	2014	/	/	/	/	/	/	**/**
Srivastava and Roy [47]	2018	80	95	90	95	90	**100**	**91.67**
Xue et al. [78]	2014	/	/	/	/	/	/	**/**
Yurtkan and Demirel [79]	2014	76.25	80.00	68.75	83.75	80.00	**91.22**	**80.00**
Zhang et al. [42]	2014	55.1	**83.4**	73.2	61.0	77.1	/	**69.96**
Azazi et al. [48]	2015	78.67	90.83	73.67	93.5	83.67	**94.50**	**85.81**
Chen et al. [64]	2015	/	/	/	/	/	/	**/**
Jan and Meng [49]	2015	88.24	90.60	86.56	91.34	85.32	**93.81**	**89.31**
Li et al. [56]	2015	82.33	82.83	72.33	**100**	89.00	81.83	**84.72**
Mao et al. [81]	2015	85.14	83.28	81.86	80.67	80.48	**98.49**	**84.99**
Yang et al. [82]	2015	83.16	83.27	78.67	**94.22**	77.18	92.31	**84.80**
Huynh et al. [53]	**2016**	**91.3**	**95.2**	**86.7**	**100**	**87.5**	**95.7**	**92.73**
Zhen et al. [66]	2016	79.5	85.7	63.3	94.6	79.2	**96.1**	**83.07**
Derkach et al. [84]	2017	85.58	75.31	65.12	89.5	77.2	**93.5**	**81.04**
Hariri et al. [86]	2017	86.25	90.00	86.25	**97.50**	79.75	90.50	**88.38**
Savran et al. [85]	2017	77	79	79	**94.5**	82	93	**84.03**
Yang and Yin [108]	2017	/	/	/	/	/	/	**/**
Wei and Jia [43]	2017	/	/	/	/	/	/	**/**
Fabiano and Canavan [40]	2018	/	/	/	/	/	/	**/**
Binghua et al. [54]	2018	/	/	/	/	/	/	**/**
Jan et al. [51]	2018	/	/	/	/	/	/	**/**

Only a few studies have dealt with data sets that explicitly incorporate AUs variation. Table 6 summarizes the performance results (recognition rates of the respective FER methods) obtained by the different research groups. The overall accuracy is indicated, as well as the results obtained for the single type of action units. The reported performances are greatly dependent on the inherent difficulty of the data.

Table 6. Recognition rates (RRs) in the presence of different action units and the overall average result obtained. For each type of AUs, the best result is written in bold characters.

Action Unit	Derkach and Sukno [84]	Dong et al. [87]	Hussain et al. [50]	Zhang et al. [42]	Zhao et al. [32]
AU 1	75	/	**100**	62.1	/
AU 2	78	**90**	/	68.2	**90.0**
AU 4	79	75	**94.67**	68.7	75
AU 5	80	/	/	/	/
AU 6	46	/	**96.89**	79.6	/
AU 7	73	**78.3**	/	69.6	**78.3**
AU 9	56	**81.7**	/	/	**81.7**
AU 10	67	**95**	/	79.1	**95**
AU 12	76	85	**90.22**	70.3	85
AU 14	/	/	/	68.2	**75**
AU 15	34	/	**86.67**	73.9	/
AU 16	52	/	/	/	/
AU 17	50	80	**92**	75.8	80
AU 18	/	/	/	/	**91.7**
AU 20	30	/	/	/	/
AU 22	/	90	/	/	**98.3**
AU 23	50	/	/	**70.4**	/
AU 24	61	/	/	**77.4**	76.7
AU 25	94	/	**94.67**	/	/
AU 26	88	**91.7**	/	/	**91.7**
AU 27	/	**91.7**	/	/	**91.7**
AU 28	/	/	/	/	**81.7**
AU 34	/	/	/	/	**88.3**
AU 43	/	/	/	/	**98.3**
Overall	**74**	**85.84**	**93.59**	**73.6**	**85.6**

9. Discussion

9.1. Computational Problems

3D faces offer more granular cues but also impose a higher dimensionality than 2D faces. Indeed, multiple scans from slightly different viewpoints are typically necessary to convert the raw data into a clean data set and reconstruct the 3D geometrical model. This problem becomes even more significant when the resolution and frame rates increase, raising the amount of 3D data acquired and consequently, the storage and computational costs. Moreover, after the data pre-processing step, post-processing which includes registration and merging of the scans, holes filling, smoothing, regularization, is needed, as well as a point detection step for head pose estimation and feature extraction.

For the reasons mentioned above and to reduce the dimensionality problem, 3D recognition tasks are often performed by mapping a 3D facial surface into the 2D plane. In other words, the depth image from the 3D mesh or the point cloud is computed, and the value of the z-coordinate of every point in the 3D space is assigned to the correspondent in the 2D plane.

In feature-based algorithms, also features selection step is crucial for the overall recognition method performance and accuracy. It aims to select the most relevant feature set from all the prospective features, and to eliminate the non-relevant features, consequently reducing the dimensionality. Hence, it needs to be well planned to deal with the high dimensionality problem inherent with the use of 3D face images.

Some research groups attempted to address this problem by proposing different strategies. For example, Azazi et al. [48] tackle the significant dimensionality problem through a twofold solution. First, the algorithm transforms the 3D faces into the 2D plane using conformal mapping. Then,

a Differential Evolution (DE) based optimization algorithm is undertaken to select the optimal facial feature set and the classifier parameters simultaneously. Only the features with the maximum discriminative power are chosen.

These steps may lead to an increase in time and computational complexity [7]. Hence, when a real-time response is desired, the 3D solution is not always right, although these strategies to reduce the response time of the algorithm have been developed.

9.2. Illumination and Limitation of Acquisition Technology

Changes in lighting due to skin reflectance properties and due to the internal camera control can significantly affect the performance of 2D FER systems. Illumination alterations cause troubles in AUs and facial expression recognition algorithms by the production of shadows and significant 2D texture variations [157]. To overcome these difficulties due to changes in the illumination conditions, image representations that are not very sensitive to these variations exist. Some examples are edge maps, image intensity derivatives, and images convolved with 2D Gabor-like filters, but Adini and Moses' results [158,159] showed that none of these representations is sufficient by itself. In general, it remains an open question of whether edge maps and the others provide an illumination-insensitive image for face recognition.

If on the one hand all or most of the 2D methods present illumination problems, on the other hand, the 3D systems are naturally robust to light variations. However, the technologies available for 3D data acquisition may be significantly affected by the difference of illumination conditions during the acquisition phase [160]. For example, artifacts, like holes or spikes, occur in oily facial regions or the proximity of eyebrows, mustache, or beard, even under ideal illumination conditions.

This problem is even more significant when tridimensional dynamic sequences are acquired. In these cases, to overpower the ambient light, the illumination provides by 3D sensors (i.e., flash) is not enough, and conspicuous lighting equipment is necessary for obtaining a constant illumination. Hence to develop a system robust to light variations for facial expression recognition remains an important goal. For this reason, in 2011, Stratou et al. [161] introduced a novel 3D dataset, called Relightable Facial Expression 3D database (ITC_3DRFE), which enables experimentations, having photometric information for studying the effect of illumination on FER. The author aimed to evaluate what kind of lighting can improve the algorithm's performances.

Despite recent advances in 3D technologies and devices, the acquisition of facial data can be accomplished only in controlled environments, representing a common drawback of 3D recognition systems. These solutions also require that the person stay still in front of the 3D scanning device for a time range from some seconds up to a few minutes. Besides, multiple scans from slightly different acquisition view-points are typically necessary to reconstruct parts of the face that can be affected by self-occlusions from a particular view.

9.3. Neutral Scan

The neutral scan that is the subjects' frontal face scan with the neutral expression is used to normalize the facial features. Based on the assumption of the availability of the subject's neutral scan during testing, it is possible to consider two different scenarios: person-dependent and person-independent. In a person-dependent approach, the feature distances of the neutral scan of a subject are subtracted from the features of his/her expressive scans.

The neutral scan of the input subject can be taken for granted in laboratory applications, but not in real-life ones. During case-studies, people gradually perform the six basic emotions from (and then go back to) the neutral in approximately four seconds. On the contrary, during real applications, the capture time is minimal, and facial expressions change very rapidly from one to another without necessarily passing through the neutral one, making it challenging to acquire it. Berretti et al. [67] in 2010 proposed an approach capable of achieving state of the art results, without using neutral scans. The same strategy was used by Fang et al. [8] in 2012. Three years later, Azazi et al. [48] proposed

a person-independent approach. Their algorithm can recognize the facial expression in any arbitrary 3D textured face image, without any prior information about the neutral face of the person. The strength of this choice lies in the fact that it does not need a neutral face, taking advantage of its use for real-time applications is required for real-life applications.

9.4. 2D VS 3D: Pros and Cons

The use of 3D facial data overcomes some problems encountered on solutions based on 2D models, but, at the same time, it has some drawbacks that make three-dimensional techniques convenient only for a set of specific applications. Table 7 lists the main advantages and disadvantages of 2D and 3D methods, next highlighting the common challenges.

Table 7. Pros and cons of 2D and 3D methods.

Comparisons	2D	3D
Illumination changes, head motions, aging, and facial make-up	2D images and videos suffer from these variations, which can affect performance	3D data is naturally robust to these variations, immune to illumination and to some extent to pose variations
Data acquisition	Trivial acquisition, possible with any device	Technology makes 3D acquisition easier and easier
Amount of data available	Large amount of data and public datasets	Only a few datasets are available but are meant to increase
Dimensional and computational costs	Very low costs	Greater dimensionality and, consequently, higher storage and computational costs
Facial surface measurements	Not enabled, it is a difficulty inherent in 2D modality	3D enables true facial surface measurements
Hidden acquisition cameras	Available	Not available
Performances for low-intensity AUs	Poor performance, achieving recognition rate lower than 3D	Good performance for lower face AUs and low-intensity AUs
Acquisition and recognition	Frontal view recognition	Ear-to-ear frontal face acquisition
Neutral scanning	No need for neutral scanning	Often need for neutral scanning, a disadvantage for real-time applications
Availability of databases with AUs and dynamic facial samples	High availability of databases, including public ones	Still low availability due to technical problems. Increase in recent years, thanks to the ease of 3D video capture
Real-time	Easy for a small amount of data	Not always good due to large time consumption

The main common challenges that both 2D and 3D methods still have to overcome are listed below:

- Current publicly available databases include posed facial expressions; lack of spontaneous datasets
- Many features or facial points are marked by hand
- Algorithms usually depend on an excessive number of feature points
- Few works are fully automatic; many of the systems still require manual intervention
- Face images collected in commonly used facial expression databases ignore the effect of time and age
- Due to the large-scale dissimilarity in the nature of the facial data, systems can accurately recognize or classify expressions only for the human faces for which it is trained
- The detection of subtle AUs or the combinations of several co-occurring AUS is still challenging
- More expressions are needed to be handled by future facial expression recognition methodologies as well as arbitrary and spontaneous ones, and micro expressions
- All the expressions have to be recognized with equal accuracy

9.5. Next Steps

Considering the different approaches analyzed in the paper, the highest recognition rates were obtained by working on three-dimensional data, or with multi-modal algorithms using 2D and 3D data. These results confirm the advantages of using 3D images or videos compared to the most common

2D methods. Based on the amount of data available, it is possible to decide whether to apply a deep learning technique to FER. Neural networks need a large amount of labeled data and considerable processing power, but allow one to perform feature extraction and their classification in an end-to-end way, reaching a state-of-the-art level of recognition accuracy. Alternatively, a conventional approach is recommended, and the most widespread is the feature-based algorithm, whereas fewer works focused on model-based or multimodal-based approaches. It is used both for the automatic AU recognition and for the basic emotions, obtaining better performances for the latter. The AUs are independent of the interpretation, and therefore more suitable to describe spontaneous facial behaviors. For this reason, thanks also to the technological development and the birth of the first databases that consider 3D dynamic and spontaneous facial expressions, in the future works we could expect a more in-depth study of the action units.

Technological advances have allowed the development of research in 3D facial expression analysis. A large number of works still use databases for static 3D FER, most with posed and exaggerated expressions of the six basic emotions, in contrast to real life. The analysis of facial expression dynamics (4D FER) is increasingly expected shortly, along with the other significant challenge of recording spontaneous behavior, "captured in a range of contexts having both high spatial and temporal resolution in real-time" [9]. The next step will be to get closer to the real world. Nowadays, with the acquisition and processing tools available, accessible to all at affordable prices, it is easier to work with three-dimensional images and videos. The transition from 2D to 3D, although with some drawbacks above all related to dimensional and computational costs, is almost complete. The next step will see the use of 4D with the introduction of the variable *time*, with the need to speed up the recognition algorithms and the creation of new databases.

Once the initial problems related to the amount of data needed and the required computing power have been overcome, a combination of deep-learning algorithms improved the performance of FER. In the future, it will be possible to integrate deep learning or other advanced algorithms with Internet-of-Things sensors, improving the current recognition rates and including spontaneous micro-expressions.

10. Conclusions

This survey explores and aims to group and organize all the works and the various methods that tried to face the problem of facial expression recognition through the analysis of human emotions, focusing on 3D solutions. Traditional and deep-learning approaches for facial expression analysis are detailed, with a further distinction between data modality (2D, 3D, and multi-modal 2D+3D), expression granularity (prototypical facial expression and facial action units), and temporal dynamics (still images and image sequences). With this study, we want to provide a guideline for newcomers who will address this topic, and take stock of neural networks, taking advantage of the golden age of AI. The most important works of recent years have been presented, highlighting the pros and cons and the best outcomes in the entire facial expression recognition field.

The overall recognition accuracy of the expressions is included in a range between 60% and 90%. Some expressions, like anger and fear, generally have the lowest recognition rates. Indeed, the motions of these expressions are moderate compared to happiness or surprise, and thus more challenging to recognize. Regarding the action units, the experiments reached recognition rates in a broader range, between 50% and 95%. The number of features for each AU to be detected will be increased to achieve more accurate results, and the neural networks will gradually be used more and more in the field of facial expression recognition.

Despite the higher dimensional and computational costs, and the greater difficulty of working in real-time, 3D methods have achieved better recognition rates than the more common 2D methods. For this reason, it is essential to use a dataset that contains 3D facial models or 3D video sequences, such as BU-3DFE, Bosphorus, BU-4DFE, D3DFACS, UPM-3DFE, BP4D, or a private one. Predicting the expression of the human face in real-time requires recognition as accurately and as quickly as possible, but it becomes quite complicated when compared to the static images because a video is

a collection of many frames, not just a single frame. Soon the recognition of emotions in real-time will be beneficial in the field of artificial intelligence research, with the need to recognize as many emotions of different people in one frame and detect mixed emotions. Many researchers developed algorithms that recognize the six basic expressions, but fewer contributions investigated other types of facial expressions or action units. For applications able to work in uncontrolled conditions, further improvements dealing with FER analysis must be done, enlarging the set of facial expressions and considering spontaneous emotions. Reducing the computational time and memory usage are the other two objectives. The primary motivation behind this is the urgent need to have robust and useful applications, which can be used in the real-world.

In our future work, we plan to include deep learning techniques, working on a private database containing three-dimensional videos and psychological validation of labeled emotions, to perform emotion recognition in the wild.

Author Contributions: Conceptualization, F.M.; methodology, F.N. and F.M.; software, N.D.; validation, F.N., F.M. and N.D.; formal analysis, F.N.; investigation, F.N and N.D.; resources, F.N. and N.D.; data curation, F.N. and N.D.; writing—original draft preparation, F.N. and N.D.; writing—review and editing, F.N. and F.M.; visualization, F.N. and F.M.; supervision, F.M. and E.V.; project administration, E.V.; funding acquisition, E.V.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ko, C.B. A brief review of facial emotion recognition based on visual information. *Sensors* **2018**, *18*, 401. [CrossRef] [PubMed]
2. Li, H.; Sun, J.; Xu, Z.; Chen, L. Multimodal 2D + 3D facial expression recognition with deep fusion convolutional neural network. *IEEE Trans. Multimed.* **2017**, *19*, 2816–2831. [CrossRef]
3. Bejaoui, H.; Ghazouani, H.; Barhoumi, W. Fully automated facial expression recognition using 3D morphable model and mesh-local binary pattern. In Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVIS), Antwerp, Belgium, 18–21 September 2017; Springer: Cham, Switzerland, 2017; pp. 39–50.
4. Mishra, B.; Fernandes, S.L.; Abhishek, K.; Alva, A.; Shetty, C.; Ajila, C.V.; Shetty, D.; Rao, H.; Shetty, P. Facial expression recognition using feature based techniques and model based techniques: A survey. In Proceedings of the 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 26–27 February 2015; pp. 589–594.
5. Danelakis, A.; Theoharis, T.; Pratikakis, I. A survey on facial expression recognition in 3D video sequences. *Multimed. Tools Appl.* **2015**, *74*, 5577–5615. [CrossRef]
6. Sariyanidi, E.; Gunes, H.; Cavallaro, A. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1113–1133. [CrossRef] [PubMed]
7. Deshmukh, S.; Patwardhan, M.; Mahajan, A. Survey on real-time facial expression recognition techniques. *IET Biom.* **2016**, *5*, 155–163. [CrossRef]
8. Fang, T.; Zhao, X.; Ocegueda, O.; Shah, S.K.; Kakadiaris, I.A. 3D facial expression recognition: A perspective on promises and challenges. In Proceedings of the Face and Gesture 2011, Santa Barbara, CA, USA, 21–23 March 2011; pp. 603–610.
9. Sandbach, G.; Zafeiriou, S.; Pantic, M.; Yin, L. Static and dynamic 3D facial expression recognition: A comprehensive survey. *Image Vis. Comput.* **2012**, *30*, 683–697. [CrossRef]
10. Pantic, M.; Rothkrantz, L.J.M. Automatic analysis of facial expressions: The state of the art. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1424–1445. [CrossRef]
11. Zeng, Z.; Pantic, M.; Roisman, G.I.; Huang, T.S. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 39–58. [CrossRef]
12. Corneanu, C.A.; Simón, M.O.; Cohn, J.F.; Guerrero, S.E. Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 1548–1568. [CrossRef]

13. Li, S.; Deng, W. Deep facial expression recognition: A survey. *arXiv* **2018**, arXiv:180408348. Available online: https://arxiv.org/abs/1804.08348 (accessed on 25 July 2019).
14. Revina, I.M.; Emmanuel, W.R.S. A survey on human face expression recognition techniques. *J. King Saud Univ. Comput. Inf. Sci.* **2018**. [CrossRef]
15. Ekman, P.; Friesen, W.V. Constants across cultures in the face and emotion. *J. Personal. Soc. Psychol.* **1971**, *17*, 124–129. [CrossRef]
16. Ekman, P.E.; Davidson, R.J. *The Nature of Emotion: Fundamental Questions*; Oxford University Press: New York, NY, USA, 1994; ISBN 978-0-19-508943-1.
17. Du, S.; Tao, Y.; Martinez, A.M. Compound facial expressions of emotion. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E1454–E1462. [CrossRef]
18. Martinez, B.; Valstar, M.F. Advances, challenges, and opportunities in automatic facial expression recognition. In *Advances in Face Detection and Facial Image Analysis*; Kawulok, M., Celebi, M.E., Smolka, B., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 63–100. ISBN 978-3-319-25958-1.
19. Kawulok, M.; Celebi, E.; Smolka, B. *Advances in Face Detection and Facial Image Analysis*; Springer: New York, NY, USA, 2016; ISBN 978-3-319-25958-1.
20. Ekman, P.; Friesen, W.V. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*; Consulting Psychologists Press: Palo Alto, CA, USA, 1978.
21. Tian, Y.; Kanade, T.; Cohn, J. Recognizing action units for facial expression analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 97–115. [CrossRef] [PubMed]
22. Scherer, K.R.; Ekman, P. (Eds.) *Handbook of Methods in Nonverbal Behavior Research*; Cambridge University Press: Cambridge, UK, 1982; ISBN 978-0-521-28072-3.
23. Yin, L.; Wei, X.; Longo, P.; Bhuvanesh, A. Analyzing facial expressions using intensity-variant 3D data for human computer interaction. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; Volume 1, pp. 1248–1251.
24. Soyel, H.; Demirel, H. Facial expression recognition using 3D facial feature distances. In Proceedings of the International Conference on Image Analysis and Recognition, Montreal, QC, Canada, 22–24 August 2007; Springer: Berlin, Germany, 2007; pp. 831–838.
25. Sun, Y.; Yin, L. Facial expression recognition based on 3D dynamic range model sequences. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2008; pp. 58–71.
26. Mpiperis, I.; Malassiotis, S.; Strintzis, M.G. Bilinear models for 3-D face and facial expression recognition. *IEEE Trans. Inf. Forensics Secur.* **2008**, *3*, 498–511. [CrossRef]
27. Soyel, H.; Demirel, H. Optimal feature selection for 3D facial expression recognition with geometrically localized facial features. In Proceedings of the 2009 Fifth International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control, Famagusta, Cyprus, 2–4 September 2009; pp. 1–4.
28. Gong, B.; Wang, Y. Automatic facial expression recognition on a single 3D face by exploring shape deformation. In Proceedings of the 17th ACM International Conference on Multimedia, Beijing, China, 19–24 October 2009; pp. 569–572.
29. Venkatesh, Y.V.; Kassim, A.A.; Ramana Murthy, O.V. A novel approach to classification of facial expressions from 3D-mesh datasets using modified PCA. *Pattern Recognit. Lett.* **2009**, *30*, 1128–1137. [CrossRef]
30. Zhao, X.; Huang, D.; Dellandrea, E.; Chen, L. Automatic 3D facial expression recognition based on a bayesian belief net and a statistical facial feature model. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 3724–3727.
31. Sun, Y.; Reale, M.; Yin, L. Recognizing partial facial action units based on 3D dynamic range data for facial expression recognition. In Proceedings of the 2008 8th IEEE International Conference on Automatic Face Gesture Recognition, Amsterdam, The Netherlands, 17–19 September 2008; pp. 1–8.
32. Zhao, X.; Dellandréa, E.; Chen, L.; Samaras, D. AU recognition on 3D faces based on an extended statistical facial feature model. In Proceedings of the 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), Washington, DC, USA, 27–29 September 2010; pp. 1–6.
33. Savran, A.; Sankur, B.; Bilge, M.T. Facial action unit detection: 3D versus 2D modality. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition—Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 71–78.

34. Jack, R.E.; Garrod, O.G.B.; Schyns, P.G. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Curr. Biol.* **2014**, *24*, 187–192. [CrossRef] [PubMed]
35. Plutchik, R. Emotions: A general psychoevolutionary theory. In *Approaches to Emotion*; Psychology Press: Road Hove, UK, 1984.
36. Plutchik, R. The Nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *Am. Sci.* **2001**, *89*, 344–350. [CrossRef]
37. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161–1178. [CrossRef]
38. Russell, J.A.; Fehr, B. Relativity in the perception of emotion in facial expressions. *J. Exp. Psychol. Gen.* **1987**, *116*, 223–237. [CrossRef]
39. Parrott, W.G. *Emotions in Social Psychology: Essential Readings*; Psychology Press: Road Hove, UK, 2001; ISBN 978-0-86377-682-3.
40. Fabiano, D.; Canavan, S. Spontaneous and Non-spontaneous 3D facial expression recognition using a statistical model with global and local constraints. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 3089–3093.
41. Zhang, X.; Yin, L.; Cohn, J.; Canavan, S.; Reale, M.; Horowitz, A.; Liu, P. A High-resolution spontaneous 3D dynamic facial expression database. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 22–26 April 2013.
42. Zhang, X.; Yin, L.; Cohn, J.F.; Canavan, S.; Reale, M.; Horowitz, A.; Liu, P.; Girard, J.M. BP4D-Spontaneous: A high-resolution spontaneous 3D dynamic facial expression database. *Image Vis. Comput.* **2014**, *32*, 692–706. [CrossRef]
43. Wei, W.; Jia, Q. 3D facial expression recognition based on Kinect. *Int. J. Innov. Comput. Inf. Control* **2017**, *13*, 1843–1854.
44. Le, V.; Tang, H.; Huang, T.S. Expression recognition from 3D dynamic faces using robust spatio-temporal shape features. In Proceedings of the Face and Gesture 2011, Santa Barbara, CA, USA, 21–25 March 2011; pp. 414–421.
45. Sandbach, G.; Zafeiriou, S.; Pantic, M.; Rueckert, D. A dynamic approach to the recognition of 3D facial expressions and their temporal models. In Proceedings of the Face and Gesture 2011, Santa Barbara, CA, USA, 21–25 March 2011; pp. 406–413.
46. Sandbach, G.; Zafeiriou, S.; Pantic, M.; Rueckert, D. Recognition of 3D facial expression dynamics. *Image Vis. Comput.* **2012**, *30*, 762–773. [CrossRef]
47. Srivastava, R.; Roy, S. Utilizing 3D flow of points for facial expression recognition. *Multimed. Tools Appl.* **2014**, *71*, 1953–1974. [CrossRef]
48. Azazi, A.; Lutfi, S.L.; Venkat, I.; Fernández-Martínez, F. Towards a robust affect recognition: Automatic facial expression recognition in 3D faces. *Expert Syst. Appl.* **2015**, *42*, 3056–3066. [CrossRef]
49. Jan, A. Hongying Meng Automatic 3D facial expression recognition using geometric and textured feature fusion. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 5, pp. 1–6.
50. Hussain, N.; Ujir, H.; Hipiny, I.; Minoi, J.-L. 3D Facial action units recognition for emotional expression. *arXiv* **2017**, arXiv:171200195. Available online: https://arxiv.org/abs/1712.00195 (accessed on 20 April 2019).
51. Jan, A.; Ding, H.; Meng, H.; Chen, L.; Li, H. Accurate facial parts localization and deep learning for 3D facial expression recognition. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 466–472.
52. Drira, H.; Amor, B.B.; Daoudi, M.; Srivastava, A.; Berretti, S. 3D dynamic expression recognition based on a novel Deformation Vector Field and Random Forest. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 1104–1107.
53. Huynh, X.-P.; Tran, T.-D.; Kim, Y.-G. Convolutional neural network models for facial expression recognition using BU-3DFE database. In *Information Science and Applications (ICISA)*; Springer: Singapore, 2016; pp. 441–450.
54. Binghua, H.E.; Zengzhao, C.; Gaoyang, L.I.; Lang, J.; Zhao, Z.; Chunlin, D. An expression recognition algorithm based on convolution neural network and RGB-D Images. *MATEC Web Conf.* **2018**, *173*, 03066. [CrossRef]
55. Zeng, W.; Li, H.; Chen, L.; Morvan, J.; Gu, X.D. An automatic 3D expression recognition framework based on sparse representation of conformal images. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 22–26 April 2013; pp. 1–8.

56. Li, H.; Ding, H.; Huang, D.; Wang, Y.; Zhao, X.; Morvan, J.-M.; Chen, L. An efficient multimodal 2D + 3D feature-based approach to automatic facial expression recognition. *Comput. Vis. Image Underst.* **2015**, *140*, 83–92. [CrossRef]
57. Tang, H.; Huang, T.S. 3D facial expression recognition based on automatically selected features. In Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2008, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
58. Tang, H.; Fu, Y.; Tu, J.; Huang, T.S.; Hasegawa-Johnson, M. EAVA: A 3D emotive audio-visual avatar. In Proceedings of the 2008 IEEE Workshop on Applications of Computer Vision, Copper Mountain, CO, USA, 7–9 January 2008; pp. 1–6.
59. Soyel, H.; Demirel, H. 3D facial expression recognition with geometrically localized facial features. In Proceedings of the 2008 23rd International Symposium on Computer and Information Sciences, Istanbul, Turkey, 27–29 October 2008; pp. 1–4.
60. Maalej, A.; Amor, B.B.; Daoudi, M.; Srivastava, A.; Berretti, S. Local 3D shape analysis for facial expression recognition. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 4129–4132.
61. Li, H.; Morvan, J.-M.; Chen, L. 3D facial expression recognition based on histograms of surface differential quantities. International Conference on Advanced Concepts for Intelligent Vision Systems, Ghent, Belgium, 22–25 August 2011; Springer: Berlin, Germany, 2011; pp. 483–494.
62. Vezzetti, E.; Tornincasa, S.; Moos, S.; Marcolin, F.; Violante, M.G.; Speranza, D.; Buisan, D.; Padula, F. 3D Human Face Analysis: Automatic expression recognition. *Biomed. Eng.* **2016**. [CrossRef]
63. Vezzetti, E.; Marcolin, F.; Tornincasa, S.; Baldassarre, F.; Vicente, D.B. 3D face expression recognition via geometry A preparatory path. *Int. J. Imag. Robot.* **2018**, *18*, 1–40.
64. Chen, H.; Li, J.; Zhang, F.; Li, Y.; Wang, H. 3D model-based continuous emotion recognition. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015; pp. 1836–1845.
65. Zhen, Q.; Huang, D.; Wang, Y.; Chen, L. Muscular movement model based automatic 3D facial expression recognition. In *International Conference on MultiMedia Modeling*; Springer: Cham, Switzerland, 2015; pp. 522–533.
66. Zhen, Q.; Huang, D.; Wang, Y.; Chen, L. Muscular movement model-based automatic 3D/4D facial expression recognition. *IEEE Trans. Multimed.* **2016**, *18*, 1438–1450. [CrossRef]
67. Berretti, S.; Bimbo, A.D.; Pala, P.; Amor, B.B.; Daoudi, M. A Set of Selected SIFT Features for 3D Facial Expression Recognition. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 4125–4128.
68. Tsalakanidou, F.; Malassiotis, S. Real-time 2D+3D facial action and expression recognition. *Pattern Recognit.* **2010**, *43*, 1763–1775. [CrossRef]
69. Venkatesh, Y.V.; Kassim, A.K.; Murthy, O.V.R. Resampling Approach to Facial Expression Recognition Using 3D Meshes. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 3772–3775.
70. Berretti, S.; Ben Amor, B.; Daoudi, M.; Del Bimbo, A. 3D facial expression recognition using SIFT descriptors of automatically detected keypoints. *Vis. Comput.* **2011**, *27*, 1021–1036. [CrossRef]
71. Vretos, N.; Nikolaidis, N.; Pitas, I. 3D facial expression recognition using Zernike moments on depth images. In Proceedings of the 2011 18th IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; pp. 773–776.
72. Li, H.; Chen, L.; Huang, D.; Wang, Y.; Morvan, J. 3D facial expression recognition via multiple kernel learning of Multi-Scale Local Normal Patterns. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 2577–2580.
73. Rabiu, H.; Saripan, M.I.; Mashohor, S.; Marhaban, M.H. 3D facial expression recognition using maximum relevance minimum redundancy geometrical features. *EURASIP J. Adv. Signal Process.* **2012**, *2012*, 213. [CrossRef]
74. Rajamanoharan, G.; Zafeiriou, S.; Pantic, M. Binary pattern analysis for 3D facial action unit detection. In Proceedings of the British Machine Vision Conference (BMVC), Guildford, UK, 3–7 September 2012.

75. Lemaire, P.; Ardabilian, M.; Chen, L.; Daoudi, M. Fully automatic 3D facial expression recognition using differential mean curvature maps and histograms of oriented gradients. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 22–26 April 2013; pp. 1–7.
76. Hayat, M.; Bennamoun, M. An automatic framework for textured 3D video-based facial expression recognition. *IEEE Trans. Affect. Comput.* **2014**, *5*, 301–313. [CrossRef]
77. Li, H.; Huang, D.; Morvan, J.M.; Chen, L.; Wang, Y. Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns. *Neurocomputing* **2014**, *133*, 179–193. [CrossRef]
78. Xue, M.; Mian, A.; Liu, W.; Li, L. Fully automatic 3D facial expression recognition using local depth features. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs, CO, USA, 24–26 March 2014; pp. 1096–1103.
79. Yurtkan, K.; Demirel, H. Feature selection for improved 3D facial expression recognition. *Pattern Recognit. Lett.* **2014**, *38*, 26–33. [CrossRef]
80. Li, X.; Ruan, Q.; Jin, Y.; An, G.; Zhao, R. Fully automatic 3D facial expression recognition using polytypic multi-block local binary patterns. *Signal Process.* **2015**, *108*, 297–308. [CrossRef]
81. Mao, Q.; Pan, X.; Zhan, Y.; Shen, X. Using Kinect for real-time emotion recognition via facial expressions. *Front. Inf. Technol. Electron. Eng.* **2015**, *16*, 272–282. [CrossRef]
82. Yang, X.; Huang, D.; Wang, Y.; Chen, L. Automatic 3D facial expression recognition using geometric scattering representation. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 1, pp. 1–6.
83. Li, X.; Ruan, Q.; An, G. Analysis of range images used in 3D facial expression recognition. In Proceedings of the 2013 IEEE International Conference of IEEE Region 10 (TENCON 2013), Xi'an, China, 22–25 October 2013; pp. 1–4.
84. Derkach, D.; Sukno, F.M. Local shape spectrum analysis for 3D facial expression recognition. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017.
85. Savran, A.; Sankur, B. Non-rigid registration based model-free 3D facial expression recognition. *Comput. Vis. Image Underst.* **2017**, *162*, 146–165. [CrossRef]
86. Hariri, W.; Tabia, H.; Farah, N.; Benouareth, A.; Declercq, D. 3D facial expression recognition using kernel methods on Riemannian manifold. *Eng. Appl. Artif. Intell.* **2017**, *64*, 25–32. [CrossRef]
87. Dong, Z.; Jia, X.; Gao, W.; Wang, K. Training on statistical feature models of action units for 3D facial expression recognition. In Proceedings of the Information Science and Cloud Computing (ISCC 2017), Guangzhou, China, 16–17 December 2017; p. 021.
88. Ramanathan, S.; Kassim, A.; Venkatesh, Y.V.; Wah, W.S. Human Facial Expression Recognition using a 3D Morphable Model. In Proceedings of the 2006 International Conference on Image Processing, Atlanta, GA, USA, 8–11 October 2006; pp. 661–664.
89. Fang, T.; Zhao, X.; Ocegueda, O.; Shah, S.K.; Kakadiaris, I.A. 3D/4D facial expression analysis: An advanced annotated face model approach. *Image Vis. Comput.* **2012**, *30*, 738–749. [CrossRef]
90. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [CrossRef] [PubMed]
91. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; ISBN 978-0-262-33737-3.
92. Deep Learning: 3 Cose da Sapere. Available online: https://it.mathworks.com/discovery/deep-learning.html (accessed on 31 July 2019).
93. Walecki, R.; Rudovic, O.; Pavlovic, V.; Schuller, B.; Pantic, M. Deep structured learning for facial action unit intensity estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21-26 July 2017; pp. 3405–3414.
94. Rete Neurale Convoluzionale. Available online: https://it.mathworks.com/solutions/deep-learning/convolutional-neural-network.html (accessed on 31 July 2019).
95. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
96. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

97. Darwin, C.; Prodger, P. *The Expression of the Emotions in Man and Animals*; Oxford University Press: Oxford, UK, 1998; ISBN 978-0-19-515806-9.
98. Bartlett, M.S.; Littlewort, G.; Fasel, I.; Movellan, J.R. Real time face detection and facial expression recognition: Development and applications to human computer interaction. In Proceedings of the 2003 Conference on Computer Vision and Pattern Recognition Workshop, Madison, WI, USA, 16–22 June 2003; Volume 5, p. 53.
99. Viola, P.; Jones, M. Robust real-time object detection. *Int J. Comput. Vis.* **2001**, *4*, 4.
100. Mayya, V.; Pai, R.M.; Manohara Pai, M.M. Automatic facial expression recognition using DCNN. *Procedia Comput. Sci.* **2016**, *93*, 453–461. [CrossRef]
101. Matsugu, M.; Mori, K.; Mitari, Y.; Kaneda, Y. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Netw.* **2003**, *16*, 555–559. [CrossRef]
102. Fasel, B. Robust face analysis using convolutional neural networks. In Proceedings of the Object Recognition Supported by User Interaction for Service Robots, Quebec City, QC, Canada, 11–15 August 2002; Volume 2, pp. 40–43.
103. Valstar, M.; Jiang, B.; Mehu, M.; Pantic, M.; Scherer, K. The first facial expression recognition and analysis challenge. In Proceedings of the Face and Gesture 2011, Santa Barbara, CA, USA, 21–25 March 2011; Volume 42, pp. 921–926.
104. Mollahosseini, A.; Chan, D.; Mahoor, M.H. Going deeper in facial expression recognition using deep neural networks. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; pp. 1–10.
105. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
106. Koehrsen, W. Facial Recognition Using Google's Convolutional Neural Network. Available online: https://medium.com/@williamkoehrsen/facial-recognition-using-googles-convolutional-neural-network-5aa752b4240e (accessed on 26 July 2019).
107. Wei, X.; Li, H.; Sun, J.; Chen, L. Unsupervised domain adaptation with regularized optimal transport for multimodal 2D+3D facial expression recognition. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 31–37.
108. Yang, H.; Yin, L. CNN based 3D facial expression recognition using masking and landmark features. In Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, USA, 23–26 October 2017; pp. 556–560.
109. Li, H.; Sun, J.; Wang, D.; Xu, Z.; Chen, L. Deep Representation of Facial Geometric and Photometric Attributes for Automatic 3D Facial Expression Recognition. *arXiv* **2015**, arXiv:1511.03015. Available online: https://arxiv.org/abs/1511.03015 (accessed on 2 August 2019).
110. Oyedotun, O.K.; Demisse, G.; Shabayek, A.E.R.; Aouada, D.; Ottersten, B. Facial expression recognition via joint deep learning of RGB-depth map latent representations. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 3161–3168.
111. Chen, Z.; Huang, D.; Wang, Y.; Chen, L. Fast and light manifold CNN based 3D facial expression recognition across pose variations. In Proceedings of the 2018 ACM Multimedia Conference on Multimedia Conference, Seoul, Korea, 22–26 October 2018; pp. 229–238.
112. Agrawal, A.; Mittal, N. Using CNN for facial expression recognition: A study of the effects of kernel size and number of filters on accuracy. *Vis. Comput.* **2019**, 1–8. [CrossRef]
113. Christou, N.; Kanojiya, N. Human facial expression recognition with convolution neural networks: ICICT 2018, London. In *Advances in Intelligent Systems and Computing*; Springer: Berlin\Heidelberg, Germany, 2019; pp. 539–545, ISBN 9789811311642.
114. Chavan, U.; Kulkarni, D. Optimizing deep convolutional neural network for facial expression recognitions. In *Data Management, Analytics and Innovation*; Springer: Singapore, 2019; pp. 185–196.
115. Jain, D.K.; Shamsolmoali, P.; Sehdev, P. Extended deep neural network for facial emotion recognition. *Pattern Recognit. Lett.* **2019**, *120*, 69–74. [CrossRef]
116. Wang, F.; Lv, J.; Ying, G.; Chen, S.; Zhang, C. Facial expression recognition from image based on hybrid features understanding. *J. Vis. Commun. Image Represent.* **2019**, *59*, 84–88. [CrossRef]

117. Minaee, S.; Abdolrashidi, A. Deep-emotion: Facial expression recognition using attentional convolutional network. *arXiv* **2019**, arXiv:1902.01019. Available online: https://arxiv.org/pdf/1902.01019.pdf (accessed on 14 June 2019).
118. Zhu, K.; Du, Z.; Li, W.; Huang, D.; Wang, Y.; Chen, L. Discriminative attention-based convolutional neural network for 3D facial expression recognition. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019), Lille, France, 14–18 May 2019; pp. 1–8.
119. Zhen, Q.; Huang, D.; Drira, H.; Ben Amor, B.; Wang, Y.; Daoudi, M. Magnifying subtle facial motions for effective 4D expression recognition. *IEEE Trans. Affect. Comput.* **2017**. [CrossRef]
120. Li, W.; Huang, D.; Li, H.; Wang, Y. Automatic 4D Facial expression recognition using dynamic geometrical image network. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 24–30.
121. Parke, F.I. Computer generated animation of faces. In Proceedings of the ACM Annual Conference—Volume 1, Boston, MA, USA, 1 August 1972; pp. 451–457.
122. Platt, S.M.; Badler, N.I. Animating facial expressions. In Proceedings of the 8th Annual Conference on Computer Graphics and Interactive Techniques, Dallas, TX, USA, 3–7 August 1981; pp. 245–252.
123. Brennan, S.E. Caricature Generator: The Dynamic Exaggeration of Faces by Computer. *Leonardo* **1985**, *18*, 170–178. [CrossRef]
124. Dagnes, N.; Ben-Mansour, K.; Marcolin, F.; Marin, F.; Sarhan, F.R.; Dakpé, S.; Vezzetti, E. What is the best set of markers for facial movements recognition? *Ann. Phys. Rehabil. Med.* **2018**, *61*, e455–e456. [CrossRef]
125. Dagnes, N.; Marcolin, F.; Vezzetti, E.; Sarhan, F.-R.; Dakpé, S.; Marin, F.; Nonis, F.; Ben Mansour, K. Optimal marker set assessment for motion capture of 3D mimic facial movements. *J. Biomech.* **2019**, *93*, 86–93. [CrossRef]
126. Weise, T.; Li, H.; Van Gool, L.; Pauly, M. Face/Off: Live facial puppetry. In Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation—SCA '09, New Orleans, LA, USA, 1–2 August 2009; 2009; p. 7.
127. Li, H.; Weise, T.; Pauly, M. Example-based facial rigging. In Proceedings of the ACM SIGGRAPH 2010 Papers, Los Angeles, CA, USA, 26–30 July 2010; pp. 32:1–32:6.
128. Weise, T.; Bouaziz, S.; Li, H.; Pauly, M. Realtime performance-based facial animation. In Proceedings of the ACM SIGGRAPH 2011 Papers, Vancouver, BC, Canada, 7–11 August 2011; pp. 77:1–77:10.
129. Li, H.; Yu, J.; Ye, Y.; Bregler, C. Realtime facial animation with on-the-fly correctives. *ACM Trans. Graph.* **2013**, *32*, 42. [CrossRef]
130. Mousas, C.; Anagnostopoulos, C.-N. Structure-aware transfer of facial blendshapes. In Proceedings of the 31st Spring Conference on Computer Graphics - SCCG '15, Smolenice, Slovakia, 22–24 April 2015; pp. 55–62.
131. Wei, L.; Deng, Z. A Practical model for live speech-driven lip-sync. *IEEE Comput. Graph. Appl.* **2015**, *35*, 70–78. [CrossRef]
132. Ouzounis, C.; Kilias, A.; Mousas, C. Kernel projection of latent structures regression for facial animation retargeting. *arXiv* **2017**, arXiv:1707.09629. Available online: https://arxiv.org/abs/1707.09629 (accessed on 2 September 2019).
133. Ma, L.; Deng, Z. Real-time facial expression transformation for monocular RGB video. *Comput. Graph. Forum* **2019**, *38*, 470–481. [CrossRef]
134. Ma, L.; Deng, Z. Real-time hierarchical facial performance capture. In Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games—I3D '19, Montreal, QC, Canada, 21–23 May 2019; pp. 1–10.
135. Lyons, M.J.; Budynek, J.; Akamatsu, S. Automatic classification of single facial images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1999**, *21*, 1357–1362. [CrossRef]
136. Phillips, P.J.; Flynn, P.J.; Scruggs, T.; Bowyer, K.W.; Chang, J.; Hoffman, K.; Marques, J.; Min, J.; Worek, W. Overview of the face recognition grand challenge. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 947–954.
137. Pantic, M.; Valstar, M.; Rademaker, R.; Maat, L. Web-based database for facial expression analysis. In Proceedings of the 2005 IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands, 6–8 July 2005; p. 5.

138. Yin, L.; Wei, X.; Sun, Y.; Wang, J.; Rosato, M.J. A 3D facial expression database for facial behavior research. In Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06), Southampton, UK, 10–12 April 2006; pp. 211–216.
139. Savran, A.; Alyüz, N.; Dibeklioğlu, H.; Çeliktutan, O.; Gökberk, B.; Sankur, B.; Akarun, L. Bosphorus database for 3D face analysis. In Proceedings of the European Workshop on Biometrics and Identity Management (BioID), Roskilde, Denmark, 7–9 May 2008; Springer: Berlin, Germany, 2008.
140. Yin, L.; Chen, X.; Sun, Y.; Worm, T.; Reale, M. A high-resolution 3D dynamic facial expression database. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG 2008), Amsterdam, The Netherlands, 17–19 September 2008; pp. 1–6.
141. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition—Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 94–101.
142. Cosker, D.; Krumhuber, E.; Hilton, A. A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2296–2303.
143. Hg, R.I.; Jasek, P.; Rofidal, C.; Nasrollahi, K.; Moeslund, T.B.; Tranchet, G. An RGB-D database using microsoft's kinect for windows for face detection. In Proceedings of the 2012 Eighth International Conference on Signal Image Technology and Internet Based Systems, Naples, Italy, 25–29 November 2012; pp. 42–46.
144. Habibu, R.; Syamsiah, M.; Hamiruce, M.M.; Iqbal, S.M. UPM-3D Facial expression recognition database (UPM-3DFE). In Proceedings of the Pacific Rim International Conference on Artificial Intelligence (PRICAI 2012), Kuching, Malaysia, 3–7 September 2012; Springer: Berlin, Germany, 2012; pp. 470–479.
145. Min, R.; Kose, N.; Dugelay, J. KinectFaceDB: A kinect database for face recognition. *IEEE Trans. Syst. Man Cybern. Syst.* **2014**, *44*, 1534–1548. [CrossRef]
146. Aneja, D.; Colburn, A.; Faigin, G.; Shapiro, L.; Mones, B. Modeling stylized character expressions via deep learning. In Proceedings of the Asian Conference on Computer Vision (ACCV), Taipei, Taiwan, 20–24 November 2016; Springer: Cham, Switzerland, 2016; pp. 136–153.
147. Krumhuber, E.; Skora, L.; Küster, D.; Fou, L. A review of dynamic datasets for facial expression research. *Emot. Rev.* **2016**, *9*, 175407391667002. [CrossRef]
148. Farkas, L.G. *Anthropometry of the Head and Face in Medicine*; Elsevier: Amsterdam, The Netherlands, 1981; ISBN 978-0-444-00557-1.
149. Marcolin, F. Miscellaneous expertise of 3D facial landmarks in recent literature. *Int. J. Biom.* **2017**, *9*, 279–304.
150. Farkas, L.G. *Anthropometry of the Head and Face*; Raven Press: Norris, MT, USA, 1994; ISBN 978-0-7817-0159-4.
151. Swennen, G.R.J.; Schutyser, F.A.C.; Hausamen, J.-E. *Three-Dimensional Cephalometry: A Color Atlas and Manual*; Springer Science & Business Media: Berlin, Germany, 2005; ISBN 978-3-540-29011-7.
152. Pantic, M.; Rothkrantz, L.J.M. Expert system for automatic analysis of facial expressions. *Image Vis. Comput.* **2000**, *18*, 881–905. [CrossRef]
153. Murphy-Chutorian, E.; Trivedi, M.M. Head pose estimation in computer vision: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 607–626. [CrossRef]
154. ScholarlyBrief. *Biometrics—Advances in Research and Application*; ScholarlyBrief: Atlanta, GA, USA, 2013; ISBN 978-1-4816-8576-4.
155. Schmidt, K.L.; Cohn, J.F. Dynamics of facial expression: Normative characteristics and individual differences. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Tokyo, Japan, 22–25 August 2001; pp. 547–550.
156. Soyel, H.; Demirel, H. Optimal feature selection for 3D facial expression recognition using coarse-to-fine classification. *Turk. J. Electr. Eng. Comput. Sci.* **2010**, *18*, 1031–1040.
157. Braje, W.L.; Kersten, D.; Tarr, M.J.; Troje, N.F. Illumination effects in face recognition. *Psychobiology* **1998**, *26*, 371–380.
158. Moses, Y.; Adini, Y.; Ullman, S. Face recognition: The problem of compensating for changes in illumination direction. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 1994; pp. 286–296.
159. Adini, Y.; Moses, Y.; Ullman, S. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 721–732. [CrossRef]

160. Bowyer, K.W.; Chang, K.; Flynn, P. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Comput. Vis. Image Underst.* **2006**, *101*, 1–15. [CrossRef]
161. Stratou, G.; Ghosh, A.; Debevec, P.; Morency, L. Effect of illumination on automatic expression recognition: A novel 3D relightable facial database. In Proceedings of the Face and Gesture 2011, Santa Barbara, CA, USA, 21–25 March 2011; pp. 611–618.