Pattern Recognition 100 (2020) 107108

Contents lists available at ScienceDirect

Pattern Recognition

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

Systematic review of 3D facial expression recognition methods

Gilderlane Ribeiro Alexandre*, José Marques Soares, George André Pereira Thé

Departamento de Engenharia de Teleinformática, Universidade Federal do Ceará, Campus do Pici s/n, Bloco 725, 60455-970, Fortaleza, CE, Brazil

ARTICLE INFO

Article history:
Received 25 February 2019
Revised 3 October 2019
Accepted 10 November 2019
Available online 11 November 2019

Keywords:
3D facial expression recognition
Systematic literature review
Preprocessing techniques
Face representation
Deep learning
Classification setup

ABSTRACT

The three-dimensional representation of the human face has emerged as a viable and effective way to characterize the facial surface for expression classification purposes. The rapid progress in the area continually demands its up-to-date characterization to guide and support research decisions, specially for newcomer researchers. This systematic literature review focus on investigating three major aspects of 3D facial expression recognition methods: face representation, preprocessing and classification experiments. The investigation of 49 specialized studies revealed the preferential types of data and regions of interest for face representation in recent years, as well as a trend towards keypoint-independent methods. In addition, it brings to light current weaknesses regarding the report of preprocessing techniques and identifies challenges concerning the current possibility of fair comparison among multiple methods. The presented findings outline essential research decisions whose the regardful report is of great value to this research community.

1. Introduction

Human emotional reactions are of great interest to science, what leads emotion-related investigations to be present in several fields of study. The recognition, interpreting, processing and simulation of emotions based on computer-aided approaches have attracted the interest of areas such as security, entertainment, education and health [1,2]. Facial expressions are only one of many resources that can be explored to apprehend the emotional state of individuals. However, besides it is an intrinsically natural alternative for communicating emotions, it also constitutes one of the least invasive means to collect emotional data from subjects, dispensing, for example, the use of wearable devices.

In the context of emotion recognition from the face, Paul Ekman gave important contributions, such as the identification of a set of universal expressions [3], which are manifested through the same facial muscular movements, regardless the ethnicity or the cultural environment in which the individuals are inserted. The emotions referred to as basic by Ekman are six: Anger, Disgust, Fear, Happiness, Sadness and Surprise. The Facial Action Coding System (FACS) [4], another of his important contributions, proposes the representation of facial movements in terms of muscular action units and allows the precise association between facial muscular movements and emotional expressions.

The majority of the applications that acquire emotional clues from faces resort to 2D static or dynamic images and require popular and low-cost devices, such as digital cameras. The popularization of 3D sensors has allowed the use of tridimensional images for biometry and emotional expression recognition, among other applications. This type of image makes it possible to overcome some limitations imposed by 2D representations, such as sensibility to light conditions, pose and use of makeup. Specifically, some muscular facial action units are difficult to be distinguished by means of 2D images, while their 3D representations are capable of overcoming that limitation [5].

Allied to the fact that the study of facial expressions has been a hot topic for decades, the conveniences offered by 3D representations might be the reason why the research activity in the 3D facial expression recognition (3D FER) field is increasing. Indeed, data collected from the scientific indexing service Scopus revealed that 3D FER has demonstrated an increasing publication activity over the past ten years (see Fig. 1).

In this context, whereas investigations and propositions of original methods, or primary studies, become available, a demand for secondary studies[1] is spontaneously stimulated to propose taxonomies considering technical and methodological characteristics of primary studies, in addition to identify scientific gaps to help other researchers.

* Corresponding author.
E-mail addresses: gilderlane@alu.ufc.br (G.R. Alexandre), marques@ufc.br (J.M. Soares), george.the@ufc.br (G.A. Pereira Thé).

[1] Secondary studies aggregate, summarize and establish conclusions about a collection of primary studies.

https://doi.org/10.1016/j.patcog.2019.107108
0031-3203/

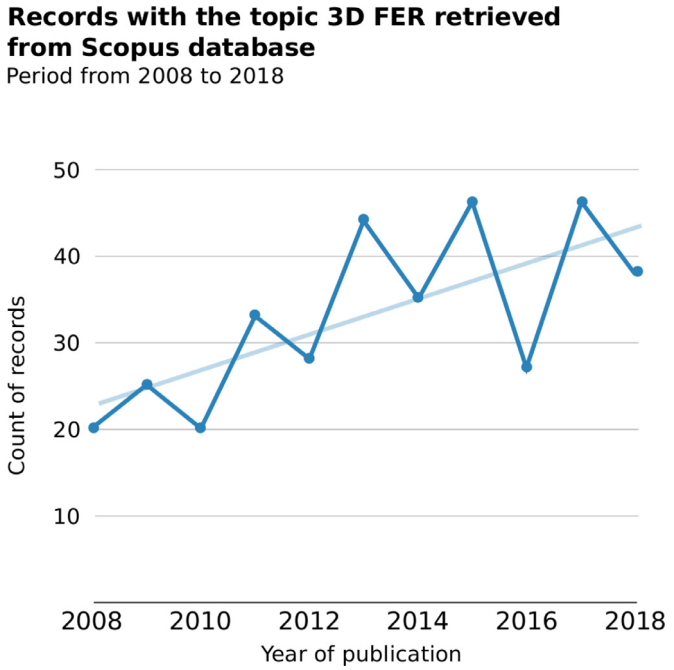

Fig. 1. Count of published works in 3D FER indexed by Scopus from 2008 to 2018. Search string: "3D" AND "facial" AND "expression recognition".

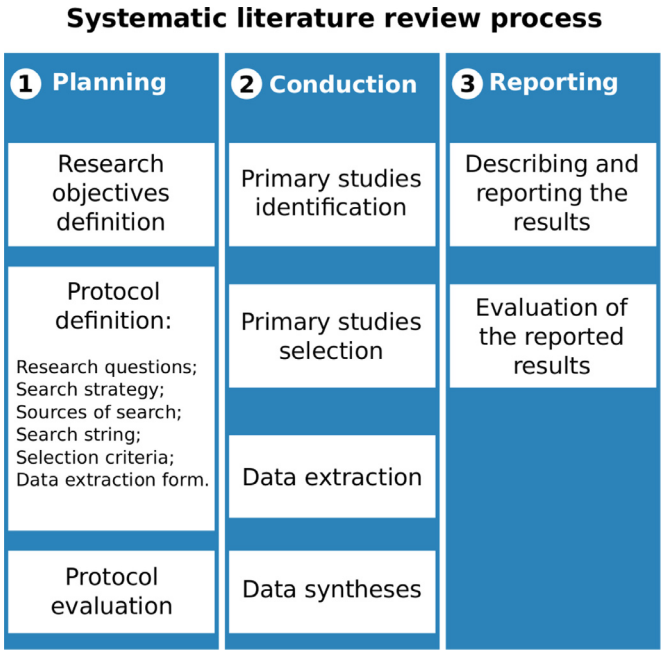

Fig. 2. SLR phases and their activities.

For example, in [5], Sandbach et al. present a comprehensive survey about 3D FER methods and bring the characterization of several aspects including acquisition technologies, 3D face databases and feature extraction techniques. Corneanu et al. [6] present a detailed literature review about multimodal methods for automatic FER, which includes not only 3D techniques but also RGB and thermal ones. This broad approach contributes to the understanding of the history and trends in the area and helps position 3D techniques within a scenario of a variety of other imaging techniques. In turn, in [7], Soltanpour et al. present a survey about local feature methods focused on 3D facial recognition (3D FR), bringing an important contribution in the description and categorization of face representation techniques. In [8], studies about 3D FER are surveyed in dynamic scenarios constituted of sequences of 3D images, also referred to as 3D videos. The deep learning based methods for FER are surveyed in [9], which presents a comprehensive description of those methods applied to static and dynamic, 2D and 3D FER.

The present study brings an up-to-date review of 3D FER works and aids the understanding of the current scenario of publications in the area, focusing on aspects that have been barely addressed in secondary studies, including statistics in this domain. Our major contributions are: (1) characterization of the most frequent preprocessing techniques in recent 3D FER studies; (2) description of the techniques adopted for face representation in this domain; and (3) summary of classification experiments employed in the surveyed studies.

The detailed description of those aspects aims to serve as a guide to newcomer researchers in the 3D FER field, and also to bring to light weaknesses to be overcome by the community in future works. This study was conducted as a Systematic Literature Review (SLR), since we understand it as a scientifically robust approach that minimizes the chances of bias and provides the means for being audited and replicated.

2. Systematic literature review process

This SLR is based on the guidelines proposed in [10] and [11] with adaptations. The systematic method provides the means for the standardized collection and summary of scientific evidences available in primary studies. A SLR is, therefore, a secondary study that differs from others by the systematization of the whole process of review, which minimizes the researchers' bias and leads to more reliable conclusions. Fig. 2 depicts the activities performed in each phase of the systematic review process we followed, which includes *planning, conduction* and *reporting*. Following in this section we present a detailed description of the outcomes of the activities performed in planning and conduction phases.

2.1. Research objective

The aim of the SLR reported in this paper is to investigate the techniques that have been developed from 2013 to 2018 for emotion recognition via 3D facial expression, emphasizing three major aspects and their relations: preprocessing, face representation, and classification experiments, to evidence not only technical but also methodological aspects. Those aspects are linked to the research questions guiding the review as discussed in the following.

2.2. Research questions

Complementary to the research objective, the following research questions are established to guide this review:

RQ1 What are the most frequently adopted face representation methods in 3D FER domain?
RQ2 Which preprocessing techniques have been required for 3D FER?
RQ3 Which degree of key points dependence is found in recent 3D FER studies?
RQ4 What procedures have been followed for classification experiments in 3D FER?

2.3. Search process

We establish as search strategies an automated search in four bibliographic databases, in addition to a manual search in the list of references of recent secondary studies in the area. We following detail the search procedure regarding both strategies.

Table 1
Inclusion and exclusion criteria.

Inclusion criteria	
IC01	Study presents a technique for automatic emotional expression recognition using 3D face images.
IC02	Study proposes or describes an application or system that utilizes an automatic emotional expression recognition using 3D face images.
Exclusion criteria	
EC01	Study is not available in full-text.
EC02	Study is not available in English.
EC03	Study is a duplicate.
EC04	Study was not published between 2013 and 2018.
EC05	Study presents technique exclusively dependent on texture images.
EC06	Study does not mention the dataset utilized to validate its method OR Study does not validate its method with a 3D dataset.
EC07	Study only approaches face recognition under expression variation, not emotional expression recognition.
EC08	Study does not consider the six basic emotions: Anger, Disgust, Fear, Happiness, Sadness, Surprise.
EC09	Study does not present the recognition results per emotion.
EC10	Study does not present either IC01 neither IC02.
EC11	Study is a course description or lecture note or patent or editorial or tutorial or survey or review.

2.3.1. Automated search

The IEEE Xplore,[2] the ACM Digital Library,[3] Scopus[4] and the Web of Science[5] (former ISI Web of Knowledge) were selected for being well reputed databases that index important journals and conference proceedings in the fields of technology and sciences as well as on the basis of the analysis of results from preliminary investigations performed by the authors. The basic difference between them is that Scopus and the Web of Science (WoS) index studies of many databases under certain criteria. Therefore, it is expected that, while there are identical documents being retrieved from more than one database, each database also contributes with a number of unique documents, so that diversity is guaranteed. The search was performed in February 2018 and restricted to studies published in a time range of 2013 to 2018, which also included documents available as "early access" in the databases. The language was limited to English. From the keywords identified in the research questions and their synonyms, we elaborated the following generic search string to be further adapted to each tool of search: **TITLE:** (3D OR " Three dimension" OR "Three dimensional") AND (face OR facial) AND (recognition OR recognizer OR detection OR detector OR detecting OR identification OR identifier OR identifying) AND (expression OR affect OR affection OR emotion OR emotional OR sentiment OR feeling) AND **TITLE OR ABSTRACT OR KEYWORDS:** ("key point extraction" OR "key point extraction" OR "key points extraction" OR "key points extraction" OR "feature selection" OR "features selection" OR "attribute selection" OR "attributes selection" OR descriptor OR signature OR classification OR classifying OR classifier OR application OR system).

2.3.2. Manual search

The manual search for primary studies was performed by consulting the lists of references of recent secondary studies on the "facial expression recognition" topic. Since a significant volume of primary studies is expected to be retrieved from bibliographic databases in automatic searches, the manual search is not intended to be exhaustive, but complementary, making it possible to consider potentially relevant studies that could not be reached by means of a specific search string. With that aim, we selected the surveys [9,12] as sources of manual search, that are made available in 2018 and include 3D FER in their discussion.

[2] https://ieeexplore.ieee.org/Xplore/home.jsp.
[3] https://dl.acm.org/.
[4] https://www.scopus.com/search/form.uri?display=basic.
[5] https://www.webofknowledge.com/.

2.4. Selection criteria

After the automated and manual searches, a manual selection of relevant documents is performed in two phases: a preliminary selection and a final selection. The aim of the preliminary selection is to reduce the number of documents to be read in full, in such way that first, only certain parts of all the documents are evaluated and the accepted documents are further appraised by the reading of the full-text. In order to minimize the bias in this process, we define two sets of selection criteria: the inclusion and exclusion criteria, listed in Table 1.

Notice that there are only two inclusion criteria which basically group two kinds of original works: we accept works that propose a technique for the emotional recognition via 3D facial expression, as well as works that describe a system or application that uses such method in practice. The exclusion criteria specify some aspects that make a document out of scope in this review, including whether a document does not meet any of the inclusion criteria (EC10). Since multiple selection criteria can be assigned to a document under evaluation, additional exclusion criteria can be used to detail the reasons for an exclusion. Some of these criteria are noteworthy: for example, according to EC05, documents that deal uniquely with texture images and therefore do not explore 3D techniques are neglected; and according to EC07, studies that deal with the facial recognition problem only (for which the expression itself is disregarded as a class) are neglected as well. In EC08, in turn, we restrict our scope to documents that include in their investigation the six basic emotions: Anger, Disgust, Fear, Happiness, Sadness and Surprise, since the universe of facial expressions is vast and not all of them are categorized as emotional expressions.

A document is only accepted if no exclusion criteria is met. On the other hand, an exclusion criteria is only assigned to a document if it is verifiable: in case of doubt, a work is always accepted and subject to deeper evaluation in the next selection phase.

2.5. Data extraction form

In order to collect evidences to answer the research questions presented in Section 2.2, we elaborate a data extraction form to be applied to each study included in the review after the selection processes. The research questions also guided the organization proposed to accommodate data collected from the studies. The form is structured in six sections: (1) Metadata, (2) Face Representation, (3) Preprocessing, (4) Key points dependence, (5) Classification experiments and (6) Results. A detailed description of the fields in the data extraction form is presented in Table 2.

Table 2
Data extraction form, filled for each selected study.

Extracted data	Description
1. Metadata	Title, authors and publication year of the study.
2. Face representation	
2.1. Regions of interest	The degree of granularity employed for the definition of ROI (e.g. surfaces, curves or points).
2.2. Descriptors	The descriptors computed from the regions of interest, that compose the feature vector (e.g. euclidean distance between feature points).
2.3. Type of data	The nature of data employed to represent the face and from which descriptors are computed (e.g. 3D, 2D+3D or 4D).
3. Preprocessing	The preprocessing steps reported in the study, e.g. registration, cropping, etc.
4. Key points dependence	
4.1. Number of key points	The number of key points required in the study (for any purpose).
4.2. Mode of detection	The means for the definition of those points (e.g. automatically by authors, manually by authors or provided in the dataset.)
4.3. Type of data	The nature of data from which key points are detected (e.g. 3D or texture).
4.4. Which specific tasks in the whole method are key points dependent?	e.g. preprocessing, feature extraction, etc.
5. Classification experiments	
5.1. 3D face database	The 3D face database used to validate the method proposed in the study.
5.2. Data sample	The subset actually employed (i.e. number of subjects, levels of intensity, whether neutral samples are included).
5.3. Training/test proportion	e.g. 90%/10%, 10-CV
5.4. Repetition	How many times the classifier ran over the training/test split.
5.5. Classifier	The algorithm for classification e.g. SVM, k-NN.
6. Results	The recognition accuracy reached per class: anger, disgust, fear, happiness, sadness, surprise, neutral.
Extracted data	Description

2.6. Protocol evaluation

The aspects described previously in this section are the result of an incremental process of evaluation and refinement. The evaluation of the protocol was performed with regard to two aspects: the definition of the search string to query each database and the definition of the selection criteria.

Each database interface specifies a set of rules to the queries supported, in such way that, for example, the allowed number of keywords and logical operators is variable. Also, some resources, such as the possibility of combining queries, work differently among databases. Those aspects contribute to the adapted search strings differ from the generic search string, which might impact the expected results. Likewise, the degree of assertiveness of the selection criteria might impact the final set of documents.

Therefore, experimental searches and selections were performed in order to evaluate whether the adaptations in the generic search string lead to acceptable results. In this process, the search string was refined and some selection criteria were improved for clarity and completeness before the conduction phase. To this end, from authors' *a priori* knowledge, the studies [13–17] were taken as controls. In other words, the adapted search strings were considered acceptable only if those five studies could be found in the total set of documents retrieved from the databases. Similarly, in the desirable scenario, the selection criteria should allow the inclusion of [13,14] and [15] as well as the exclusion of [16] and [17].

2.7. Search and selection results

The search string previously established in the protocol was conformably applied to each search tool. It is expected that duplicates were to be retrieved, since there is an intersection of documents indexed in each database as well as listed as references of the papers selected as sources of the manual search. A total of 536 unique documents remained after duplicates removal and compose the initial set of documents candidate to be evaluated in this review.

The preliminary document selection was conducted based on the reading of only title and abstract of the 536 unique documents in the initial set. Five researchers were involved in this task and evaluated the fitness of each document to the previously established inclusion and exclusion criteria. As result, 142 documents were accepted in this phase, i.e. considered for further analysis.

The documents accepted in the preliminary selection phase were further evaluated by the reading of their full-text. As result, a total of 49 documents, which corresponds to 34% of the evaluated documents in this phase and 9% of the initial set of documents retrieved, were preserved in this review. Fig. 3 summarizes search and selection procedures.

2.8. Strategy of data summarization

The 49 selected documents undergone data extraction procedure by using the data extraction form presented in Section 2.5. The collected evidences resulting from that activity are further summarized with respect to: preprocessing, face representation and classification experiment setup. Although in primary studies those aspects are commonly presented in the chronological order of execution (see Fig. 4), in this work the summary of face representation aspects precedes preprocessing, since some concepts related to the first are also used in further sections.

3. Face representation

Traditional 3D FER methods adopt particular face representation approaches that serve as input to an automatic classification system. In fact, face representation decisions are the most important aspect in FER tasks. Those decisions involve feature extraction and selection processes and can be described in terms of nature of data, type of descriptors and regions of interest (ROIs).

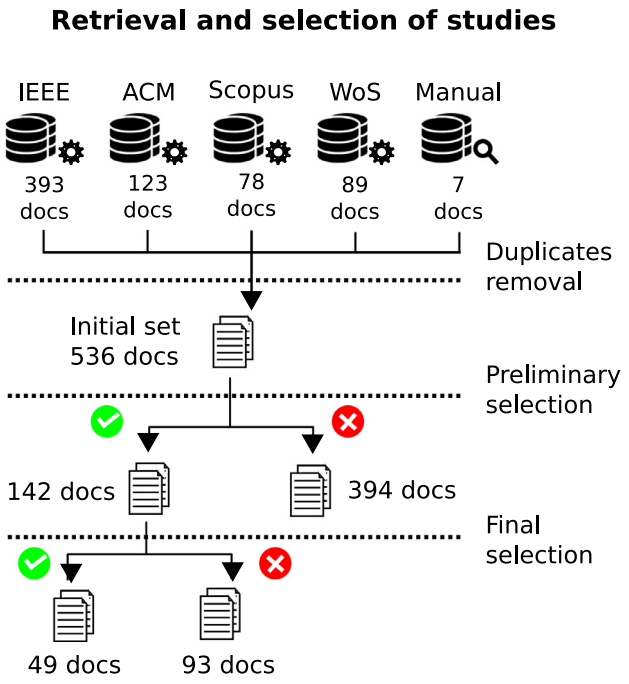

Fig. 3. Summary of the documents' search and selection procedures.

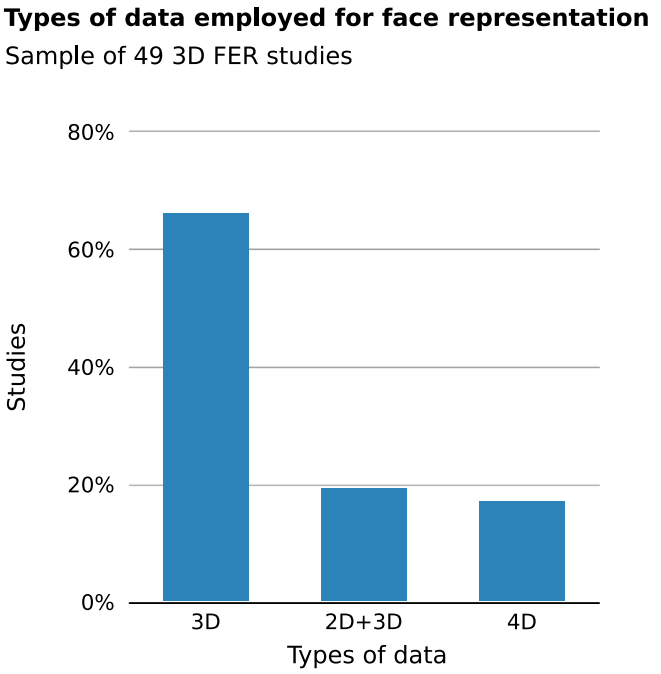

Fig. 5. Type of data employed in the 3D FER surveyed studies. Purely 3D representations are the most frequently adopted, while methods that aggregate texture (2D+3D) and variation over time (4D) are the minority. Among the studies categorized as 4D, the textured 3D videos are also included.

Next, we summarize the most relevant findings about traditional face representation approaches in the surveyed studies.

3.1. Types of data and descriptors

3D FER is approached in the literature through the use of purely 3D data or of its combination with texture and time variation information. The designation 2D+3D is given to the approaches that combine information of 2D and 3D images, captured simultaneously; the approach so called 4D employs information of the variation perceived in sequences of frames over time, and because of that, also called 3D videos. The majority of the surveyed studies propose solutions established upon solely geometric information of 3D static images, as opposed to multimodal solutions, as depicted in Fig. 5.

The extracted descriptors depend on the type of data employed to represent the face. Studies that make use of purely 3D data usually extract geometric descriptors such as coordinate values, distances (euclidean or geodesic) and curvatures. Studies that also deal with texture information extract, additionally, descriptors appropriate for this type of data, such as Local Binary Patterns (LBP) and descriptors relying on Scale-Invariant feature transform (SIFT). Moreover, even in studies that utilize solely 3D data, the geometric coordinates of the face are usually mapped into 2D representation, such as depth maps and their derivations: normal and shape index maps. That transformation enables the indirect extraction of descriptors originally designed for the 2D domain from 3D data.

3.2. Regions of interest

In the context of 3D FER, ROIs may adopt global, local and hybrid scopes, in which the latter combines aspects of the first two. That taxonomy have been employed by Soltanpour et al. [7] and Corneanu et al. [6]. In Table 3, the 49 selected studies subject to analysis in this work are categorized according to the type of scope of the ROIs of their face representation methods.

Methods that make use of global scope ROIs consider a unique region, usually comprising the whole face surface. The local scope

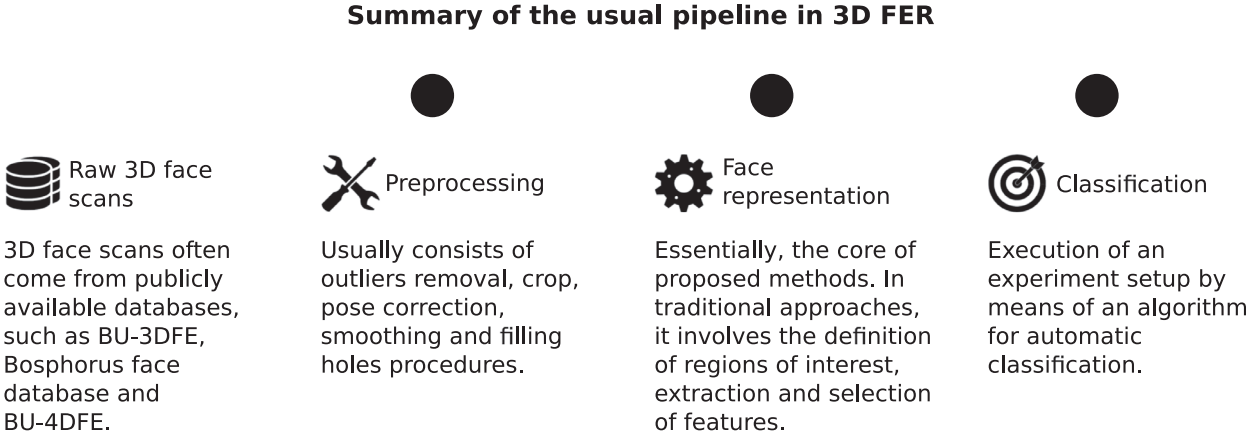

Fig. 4. The usual pipeline followed in traditional 3D FER studies.

Table 3
Surveyed studies according to the types of scopes of their ROIs. Notice that Sheng et al. [59] appears twice, since it combines two types of local ROIs: keypoints and surfaces.

Global	Local			Hybrid
	keypoints	Curves	Surfaces	
An and Ruan [18]	Aly et al. [19]	Amor et al. [20]	Berretti et al. [21]	Aly et al. [22]
Li et al. [23]	Azazi et al. [24]	Maalej et al. [25]	Derkach and Sukno [26]	Jan and Meng [27]
Fu et al. [28]	Azazi et al. [29]	Samad and Iftekharuddin [30]	Han and Ming [31]	
Li et al. [32]	Azazi et al. [13]	Zhen et al. [33]	Hariri et al. [15]	
Li et al. [34]	Elhoufi et al. [35]		Hayat et al. [36]	
Savran and Sankur [37]	Jazouli et al. [38]		Lemaire et al. [39]	
Trimech et al. [40]	Khashman and Conkbayir [41]		Li et al. [42]	
Zeng et al. [43]	Suja et al. [44]		Li et al. [45]	
	Tao and Matuszewski [46]		Moeni et al. [47]	
	Yurtkan and Demirel [14]		Ocegueda et al. [48]	
	Yurtkan and Demirel [49]		Reale et al. [50]	
	Yurtkan and Demirel [51]		Vieriu et al. [52]	
	Yurtkan et al. [53]		Xue et al. [54]	
	Zarbakhsh and Demirel [55]		Xue et al. [56]	
	Zhang et al. [57]		Yang et al. [58]	
	Sheng et al. [59]		Yao et al. [60]	
			Zhen et al. [61]	
			Sheng et al. [59]	

methods utilize one of three types of specific regions over the face: keypoints, curves and local surfaces; or a combination of them, as in [59], in which Sheng et al. consider both keypoints and local surfaces.

Local methods based on keypoints to represent the face use features directly extracted from those relevant points. For example, in [35], the spatial coordinates of 121 keypoints are directly utilized to form feature vectors to describe faces. In [55], Xue et al. select the most discriminating measures among some euclidean distances between 83 keypoints. In [44], in turn, feature vectors are formed from the concatenation of a vector of distances between pairs of keypoints and a vector of angles between those points. In [46], the feature vector is defined as the difference between the positions of keypoints in expressive faces and their corresponding points in a reference neutral face. Those works have in common the requirement of keypoints detection in their face representation method.

Local methods based on surfaces utilize strategies to determine ROIs based on regular and irregular grids or based on the area neighboring keypoints. As an example of the first case, in [60] some frames are selected from 3D textured videos. For face representation, features are extracted from local patches formed by a regular grid over each frame of each type of data, texture and 3D. Specifically, LBP are extracted from patches of texture frames and scattering coefficients are extracted from patches of 3D frames. In [15], in turn, keypoints are determined as references to delimit local surfaces of interest, in which 30 points are uniformly distributed over the faces and represent the center of 30 local surfaces with overlap.

Less frequent, local scope methods based on curves represent the face by a set of curves over its surface. Those curves are, generally, of the following types: radial, from a central point to a face extremity, or contour, closed around a central point.

The majority of the studies surveyed in this work adopt, as a form of face representation, local regions from which features are extracted, utilizing purely 3D data. In Fig. 6, we highlight the proportion of the choice for different scopes of ROIs for each type of data employed.

3.3. Discussion about face representation

The literature about 3D FER consistently points to the effectiveness of multimodal methods. For example, Li et al. [42] and Jan and Meng [27] demonstrate that, in equivalent experiments, methods combining 3D and texture descriptors present superior recognition rates, compared to methods that employ only 3D or only texture attributes. Likewise, Sandbach et al., in [62] demonstrate that the information extracted from sequential samples, which allows capturing the variation of expressions over time, also improve results in 3D FER. That claim has stimulated further researches in that direction. Despite of this, the majority of the surveyed studies propose solutions based on a purely 3D representation. The advantage of that representation over multimodal methods is that it is significantly simpler, requiring the handling of a smaller volume of data and the extraction of features of a unique type of data. Furthermore, those characteristics are desirable in applications and contribute to lower computational cost in the proposed methods.

From another perspective, in traditional methods, the preference for local scope ROIs, which consider keypoints, curves and

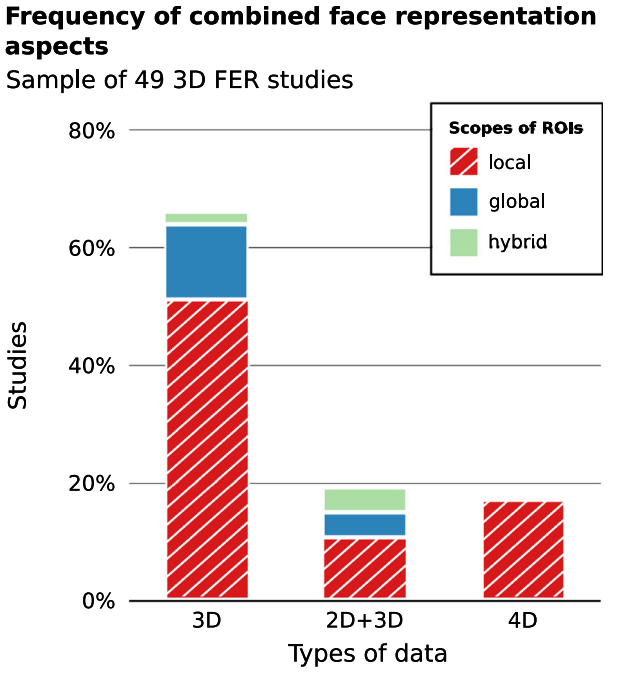

Fig. 6. Proportions of scopes of ROI employed per type of data. The biggest portion of documents combine local scope methods with solely 3D data.

Table 4
Surveyed studies reporting the adopted preprocessing steps. Columns are sorted, from left to right, in the decreasing order of their frequency of use.

Studies	Pose correction	Face crop	Smoothing	Holes filling	Other
Li et al. [45]	✔	✔	✔	✔	✔
Fu et al. [28]	✔	✔	✔	✔	✖
Amor et al. [20]	✔	✔	✖	✔	✖
Li et al. [34]	✔	✔	✖	✖	✔
Lemaire et al. [39]	✔	✔	✖	✖	✔
Zhen et al. [33]	✔	✔	✖	✖	✔
Hayat et al. [36]	✔	✔	✖	✖	✖
Maalej et al. [25]	✔	✔	✖	✖	✖
Zhen et al. [61]	✔	✖	✔	✔	✖
Sheng et al. [59]	✔	✖	✔	✔	✖
Xue et al. [56]	✔	✖	✖	✖	✔
Aly et al. [19]	✔	✖	✖	✖	✔
Zeng et al. [43]	✔	✖	✖	✖	✔
Reale et al. [50]	✔	✖	✖	✖	✔
Ocegueda et al. [48]	✔	✖	✖	✖	✔
Yao et al. [60]	✔	✖	✖	✖	✖
Savran and Sankur [37]	✔	✖	✖	✖	✖
Berretti et al. [21]	✖	✔	✔	✔	✔
Samad and Iftekharuddin [30]	✖	✔	✔	✔	✔
Hariri et al. [15]	✖	✔	✔	✔	✖
Li et al. [23]	✖	✔	✔	✖	✔
Xue et al. [54]	✖	✔	✔	✖	✔
Azazi et al. [13]	✖	✖	✔	✔	✔
Trimech et al. [40]	✖	✖	✔	✖	✔
Azazi et al. [29]	✖	✖	✖	✖	✔
Yang et al. [58]	✖	✖	✖	✖	✔
Derkach and Sukno [26]	✖	✖	✖	✖	✔
Vieriu et al. [52]	✖	✖	✖	✖	✔
Azazi et al. [24]	✖	✖	✖	✖	✖

surfaces, also suggests that the trade-off between simplicity and robustness of the reported techniques is often pursued. Those methods describe the face in terms of the most relevant regions instead of the use of the whole face surface, and can eliminate the influence of regions redundant amongst classes of emotions. As formalized in the FACS [4], the adoption of those face representation techniques is based on the perception that there are regions whose deformations caused by emotional expressions are more pronounced than others. It is no accident that the points considered relevant in the surveyed studies, generally, coincide with fiducial facial landmarks, which mark characteristic regions on face, such as eyes corners, mouth corners, eyebrows and nose tip. Likewise, works that consider local surfaces, in general, cover those same regions.

4. Preprocessing

Among the surveyed studies, the major preprocessing techniques applied to 3D faces are: ***(i)*** **pose correction**, ***(ii)*** **face crop**, ***(iii)*** **smoothing** and ***(iv)*** **holes filling**. Less frequently, other techniques such as re-sampling and 3D-2D mappings are reported as preprocessings. Multiple preprocessing techniques are commonly employed in the same study, as shown in Table 4, in which all the surveyed studies that mentioned their preprocessing steps are listed.

A total of 18 studies, which amounts to nearly 37% of the total of the surveyed studies in this work, do not mention any preprocessing applied to data. Those studies, however, do not make it clear whether the previous processing is actually dispensable. The exception is the study [58], in which the authors clearly claim that no preprocessing technique was employed, since the data available in the BU-3DFE database already had satisfactory quality for their purposes.

(i) **Pose correction:** The algorithm Iterative Closest Point (ICP) is the most mentioned for pose correction. It performs the rigid registration of 3D faces, in which the rotation matrix and translation vector that are the most adequate to the matching of two sets of points are computed. That algorithm was employed in [20,37,45,56,60,61] and [28]. The points utilized are generally keypoints, their neighbors or all the points of the face. ICP is dependent on the initial position of points, in addition to parameters for its adjustment, which affects the goodness of the registration. In spite of that, the reporting of the use of ICP, generally does not include the parameters used, which can make it harder to replicate results. Other techniques for pose correction are the MÃ¶bius transform for the normalization of conformal maps [43], pose correction based on Principal Component Analysis (PCA) [36] and Procrustes superimposition [19]. Moreover, it is common the description of pose correction procedures by the use of solely generic terms such as, “alignment”, “normalization” or “face registration”, with no further details being provided.

(ii) **Face crop:** The crop of facial area aims to the removal of peripheral points that are noisy and redundant for classification of 3D face models, such as hair, ears, neck and, in some cases, the bust. The spherical crop is the most frequently reported; it consists in the rejection of regions that are located beyond a spherical neighborhood of a certain reference point, usually the nose tip. Measures between 70 mm and 90 mm are usually employed as radius for crop. That preprocessing is a necessary and sufficient procedure for the definition of the ROIs in global scope methods. However, in local scope methods, it can correspond to a preliminary step, preceding the segmentation of other smaller ROIs. Similar to what happens with pose correction, authors usually neglect important information about crop procedure, employing only the generic denomination “crop” to refer to this preprocessing step as a whole, which is not enough to describe it. The studies that reported the crop of facial area as a preprocessing step are listed in Table 5.

(iii) **Smoothing:** Surface smoothing is applied to reduce the effect of noises inherent in the acquisition process. Similar to previous cases, smoothing is not frequently detailed in 3D FER studies. To describe that preprocessing, the surveyed studies only superfi-

Table 5
Characteristics of crop procedure reported in the surveyed studies. NA = Not applicable. Notice that, even though, some studies mention performing "crop", in some cases, details about that procedure are missing.

Studies	Spherical crop?	Centered in the nose tip?	Radius for crop
Hayat et al. [36]	Yes	Yes	85 mm
Berreti et al. [21]	Yes	Yes	90 mm
Li et al. [34]	–	–	–
Hariri et al. [15]	–	–	–
Amor et al. [20]	Yes	Yes	90 mm
Lemaire et al. [39]	Yes	Yes	80 mm
Zhen et al. [33]	–	–	–
Fu et al. [28]	Yes	Yes	90 mm
Li et al. [23]	Yes	Yes	–
Maalej et al. [25]	–	–	–
Xue et al. [54]	Yes	Yes	70 mm
Samad and Iftekharuddin [30]	No	No	NA
Li et al. [45]	Yes	Yes	distance between nose tip and chin.

cially name the filters employed, such as median filter [23,45,59], gaussian filter [13,21], average filter [30] and Laplacian operator [40] or, even more generically, only use the term "smoothing filters" [15,28,61]. Among the surveyed studies, the exception to that practice is found in [54], where authors provide relevant information to the replication of the smoothing procedure. In that study, a filter based on the average of points is applied, considering as outliers all of those beyond five standard deviations and rejecting them.

***(iv)* Holes filling:** During data acquisition, the position of the scanned faces relative to the 3D sensor may favor the occurrence of holes in the 3D face models. Surface gaps may also appear as a result of some facial expressions, such as the open mouth, characteristic of expressions of surprise and fear. Those openings in the face surface may disturb the face representation scheme and may be subject to preprocessing. The filling of holes is the fourth most frequently reported preprocessing in the surveyed studies. Cubic interpolation is the most mentioned technique to this purpose, being employed in [13,21,30,45,59,61]. None of the studies provided further information about that procedure.

4.1. Note on keypoints detection

Although keypoints detection is not always reported as a preprocessing technique, among the surveyed studies, it is required for supporting two major tasks: feature extraction and other preprocessing procedure. In the first case, that dependence mainly occurs in local scope face representation methods based on keypoints, in which features are directly derived from those coordinates (see Table 3). However, the local scope methods based on surface and curves may also depend on the identification of some relevant points to determine ROIs around of or from those points, as in [20,33]. Even in face representation methods that do not directly depend on keypoints detection, such procedure can still be necessary in other preprocessing procedures, such as the removal of peripheral regions, in which specially the nose tip is requested (see Table 5), and pose correction.

A total of 38 out of the 49 3D FER studies surveyed in this SLR (78%) present dependence on the detection of keypoints, varying from 1 to 121. Accordingly, approximately 20% of the surveyed studies are keypoints independent. In Fig. 7 we define 12 ranges for the count of required keypoints and, for each range, we present the percentage of studies in it. That arrangement reveals that more than 50% of the studies make efforts to detect more than 10 keypoints. The dependence on a large number of relevant points has been discussed by Soltanpour et al. in [7], where authors point to the high computational cost and need for a very precise detection of those points as the main weaknesses associated to the automatic detection techniques.

Beyond the keypoints dependence aspect from a quantitative perspective, the surveyed studies also revealed that current methods depend on the detection of keypoints whose positions carry some *a priori* known semantics, specifically, anatomical characteristics of the face. Except for Hariri et al. [15], all the keypoints required in the surveyed studies correspond to fiducial facial landmarks, which are those commonly used to identify facial features such as eye corners, mouth, eyebrows and nose tip.

Keypoints can be automatically detected or manually marked over 3D face models. Automatic detection is predominant among the surveyed studies, as shown in Fig. 8. Specifically for fiducial facial landmarks detection, there are well-consolidated algorithms available to solve that problem in the 2D domain. Those algorithms also succeed in that task when applied to 3D-2D mappings. In [29], Azazi et al. utilize a two-dimensional representation of 3D faces, by using conformal mappings and conveniently applying the algorithm Structured Output Support Vector Machine (SO-SVM) with Deformable Part Model (DPM) for landmarks detection. In [56], Xue et al. employ Haar-cascade algorithm based on AdaBoost over depth maps and their gradients for fiducial facial landmarks detection. In [13,27,42,54], Azazi and Co-authors perform automatic

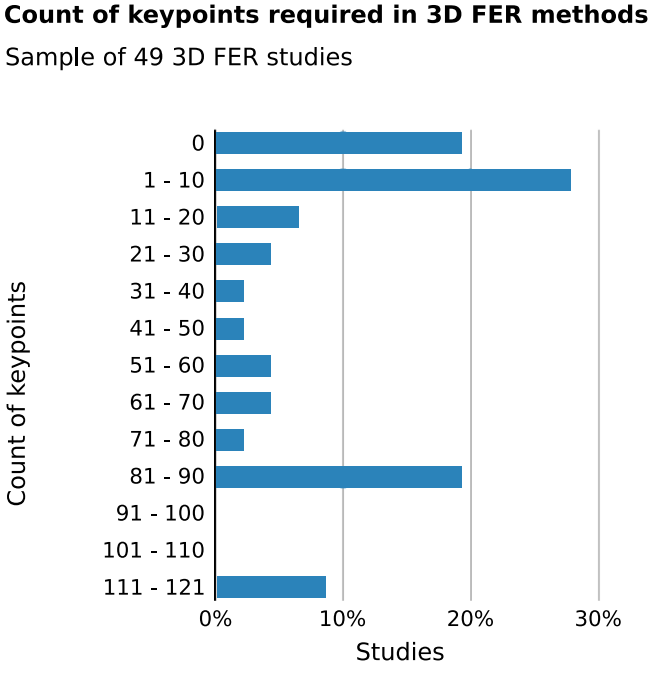

Fig. 7. Count of keypoints required in the 49 3D FER studies surveyed. Only 20% do not present any dependence on keypoints detection.

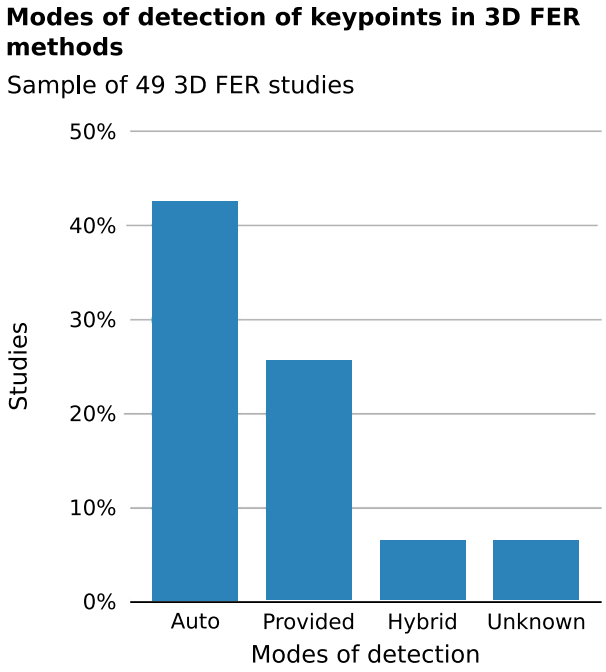

Fig. 8. Modes of detection of keypoints. Almost 20% of the surveyed studies do not require the detection of keypoints.

detection of fiducial facial points on texture images and then map those points to the correspondent 3D models.

Studies dependent on manual marking present solutions so called semi-automatic, that, to be applied in concrete problems, presume a previous and independent step for the detection of those points. None of the surveyed studies required exclusively manual marking of keypoints. On the other hand, a significant portion of the studies utilize markings provided in the 3D face datasets employed for validation. That kind of dependence also makes the method semi-automatic.

Hybrid methods of keypoints identification combine manual detection or manual marking with an automatic step. Methods with that characteristics are identified in only few studies. For example, in [61], Zhen et al. perform the manual marking of 71 keypoints that are used for the delimitation of 11 ROIs over a unique template face. The patches defined in the model are used to segment equivalent regions in all faces by using the Iterative Closest Normal Point (ICNP) algorithm. In the same study, for pose normalization, the nose tip is detected in each face as the point with the greatest magnitude of its z-coordinate.

In turn, in [23,28,33] only the nose tip is required as reference for either face representation or preprocessing, but the mode of detection is not reported by the authors and therefore are classified as unknown.

4.2. Discussion about preprocessing

Characteristics intrinsic of the data acquisition procedure can impose important limitations to the methods proposed in the studies. For that reason, preprocessing is an undeniably important aspect in machine learning tasks.

Specifically in 3D FER, raw data generally consists of 3D meshes or point clouds (textured or not), in such way that proposed methods hardly ever dispense previous treatment. Although, a wide range of procedures preceding the tasks of face representation may be considered preprocessing, frequent disturbances such as pose variations, presence of noise, presence of non-facial regions, as well as the presence of holes on face surface may impair the quality of representation and should be handled. Therefore pose correction, face crop, smoothing and holes filling are the procedures most commonly reported as such.

In spite of the relatively easy identification of those procedures in literature, not infrequently their description is insufficient, which makes the occasional attempt of replication and fair comparison a current challenge in this community. For example, several methods that depend on finding the nose tip for cropping or for ROI definition do so by taking the point with the greatest value in z. That procedure although simple is effective in well behaved scenarios since it assumes that face scans have had their pose corrected and that they do not have noise or other particular facial feature, such as a projected chin that could potentially mislead the detection of the actual nose tip. Thus, such preprocessing procedures become absolutely necessary for an appropriate face representation. A problem arises when, for example, the most popular algorithm for pose correction, ICP, is sensitive to parameters and, naturally, its outcome may be quite different for varied values, even when starting from the same exact database. Therefore, the starting point for face representation is susceptible, for example, to variations depending on ICP parameters. Similar consequences are applicable to the quality of smoothing and holes filling. That affects the ability of other researchers to replicate those methods, since there are no guarantees that the replicated method was conducted in the same conditions as the original and, therefore, it impairs the possibility of fair comparison between techniques, an important premise for the scientific progress.

Moreover, the application of preprocessing procedures is related to the nature of data employed for validation and reveals the degree of robustness of the proposed method to adverse conditions. Therefore, it is important that the need for that step or its absence are explicitly reported. In spite of that, the results achieved in this SLR show that it is usual that 3D FER studies do not mention or do not appropriately describe their preprocessing steps. That practice is observed in 43% of the conference papers and 33% of the journal articles surveyed in this work, reinforcing the evidence of a generalized omission of that aspect, which can be indispensable for the results achieved by some of the techniques presented.

Such issue presents itself both as a technical and methodological concern, since it arises from the need for preprocessing for adequate face representation and deepens with the neglect of relevant information to the community. In this direction, the development of robust methods to some adverse conditions and that, therefore, may dispense some preprocessing procedures, appear as a research opportunity. Advancements have already been presented with deep learning based methods for expression recognition robust to rotation [63] and evidence a potential for improvements.

Moreover, keypoints detection has been identified as a traditional requirement either for supporting feature extraction or other preprocessing steps. In this context, in which the dependence on keypoints implies either a computational cost as high as the count of required keypoints or severe restrictions in the applicability of the method, it is justifiable the interest in the proposition of methods completely independent of keypoints detection. Fig. 9 presents the proportion of keypoints-independent studies per year of publication identified in this review. Although only 20% of the analyzed studies are completely free of the need for those markings, we notice a tendency, although modest, for researchers in the area to abandon keypoints dependent techniques over the last years.

5. Classification experiments

The appropriate description of the experimental design is a crucial element of every scientific production, since it is intrinsically related to the achieved results, which are generally presented in the form of mean accuracies of automatic classification system's

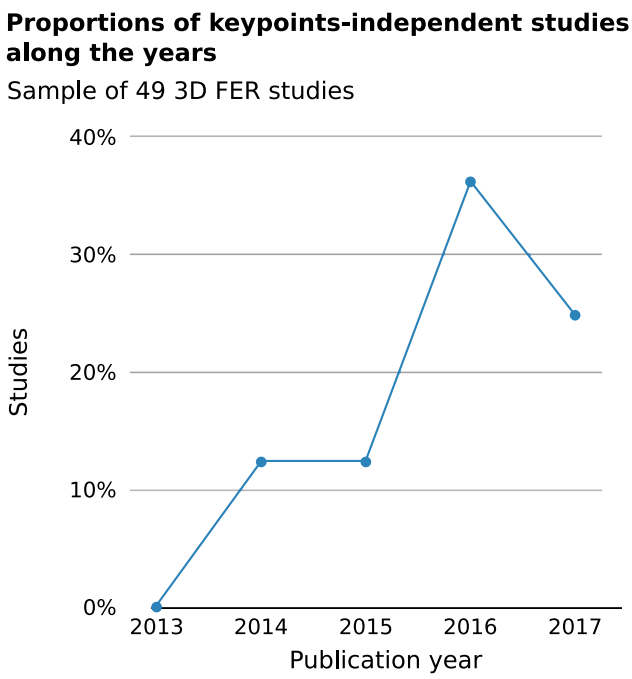

Fig. 9. Proportion of keypoints independent studies among the 49 surveyed studies per publication year.

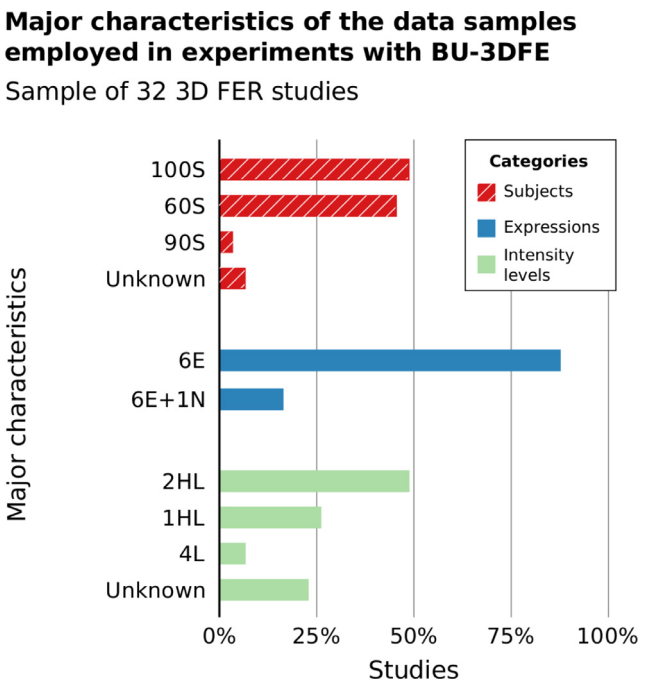

Fig. 10. The frequency of the major characteristics of data samples employed in experiments with BU-3DFE. Since the same study may report multiple classification experiments, each of them with different characteristics of data samples, it is registered in the graph the studies in which at least one experiment presents such characteristic.

performance. Such values are, therefore, estimates of the real performances of a certain system and are obtained by means of the presentation of samples of a population. Those results, in turn, are used to legitimate the contributions presented by the study and may indicate the state-of-the-art in the field and, therefore, their analysis for eventual comparisons should take into account a variety of aspects. Particularly in 3D FER studies, besides the decision for the classifier itself, other prominent aspects should be observed for a thorough description of the experimental design: the actual sample of data employed, the count of rounds of classification and the proportion between training and test data. In this section those aspects are grouped into data sample related and classification strategy related and thoroughly discussed in light of the surveyed studies.

5.1. Data sample

Among the 49 analyzed studies, the data employed in the experiments are restricted to four publicly available and two private datasets. Each of them presents inherent characteristics that may affect the experimental design, specially regarding the sample of data actually utilized. In Table 6, we list some relevant aspects of the composition of the datasets targeted in the surveyed studies. We following identify the major characteristics of the most frequent samples of data extracted from each of them.

5.1.1. BU-3DFE

Thirty-one out of the forty-nine studies surveyed in this review perform classification experiments with data from the BU-3DFE database. As summarized in Table 6, the BU-3DFE database has a total of 2.500 scans acquired from 100 subjects, each of them performing one neutral expression as well as the six basic emotions graduated in four levels of intensity (so as to capture the degree of happiness, sadness, etc). However the experiments using that database, in general, do not utilize the entire set of scans. Those experiments vary with respect to the counts of subjects, expressions and the levels of intensity of the expressive samples used for validation. As shown in Fig. 10, there is a preference for experiments with all the 100 subjects available (100S). A slightly inferior number of studies employ 60 subjects (60S) out of 100. With regard to the expressions considered in the experiments, the large

Table 6
The databases utilized for validation among the 3D FER surveyed studies. Letter S follows the count of subjects, while E and N follow the counts of prototypical expressions and neutral scans, respectively, available in the database.

Databases	Type of data	Composition	Total of samples
BU-3DFE [64]	texture + 3D	100S × (6E* + 1N)	2.500 scans
Bosphorus [65]	texture + 3D	105S × (6E + 4N)**	4.666 scans
BU-4DFE [66]	texture + 4D	101S × 6E	606 sequences (60.600 frames)
VT-KFER [67]	texture + 4D	32S × (6E + 1N)	32 sequences (61.374 frames)
Bertacchini et al. [35]	3D	10S × 6E	60 scans
Aly et al. [19]	3D	10S × 6E	17.000 scans

* The BU-3DFE database has four levels of intensity per prototypical expression, graduated from 1, the mildest, to 4, the most intense.

** The Bosphorus database has an unbalanced number of neutral and expressive samples per subject. The count of neutral samples vary from 1 to 4. Only 65 subjects have samples of all six prototypical expressions.

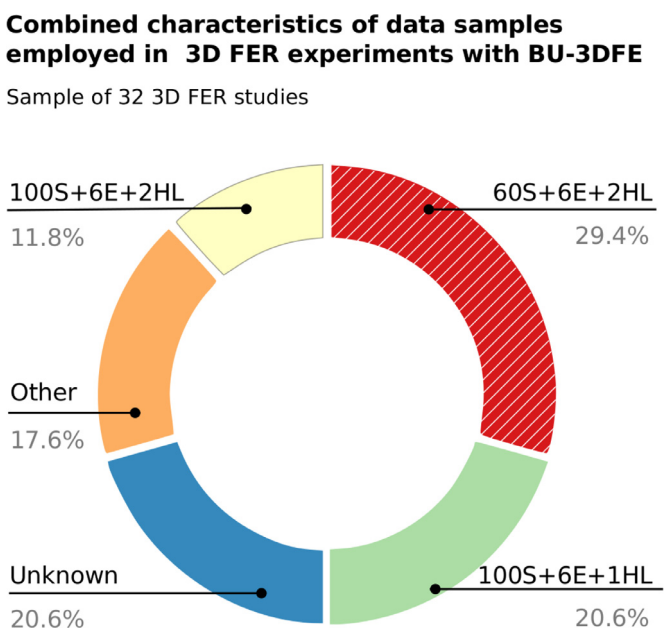

Fig. 11. Occurrences of characteristic portions of data in 3D FER experiments with BU-3DFE.

majority of the studies include scans of the six basic emotions (6E), ignoring neutral scans (N). Despite the availability of 4 intensity levels of the expressive scans, the 2 least intense levels are frequently disregarded. The scans of the 2 highest levels of intensity (2HL) are the most employed, followed by the scans of the most intense level (1HL).

Isolated, those characteristics can give us a hint of how is composed the prevalent portion of data used in studies with BU-3DFE. However, it is the combined occurrences of those characteristics that completely describe the sample of data. The count of those combinations points to the preference for scans of the 2 highest intensity levels of the 6 prototypical expressions of 60 subjects randomly selected out of 100, i.e. the portion referred to as 60S+ 6E+2HL. Yet, approximately 21% of the studies utilizing BU-3DFE lack information about at least one of those three aspects. In Fig. 11, the most frequent combinations of data samples characteristics are presented.

5.1.2. Bosphorus

The Bosphorus face database is the second most frequent for validation in 3D FER. Among the 49 surveyed studies, 10 report classification experiments with that database. Bosphorus is usually adopted as an extra database to demonstrate the generality of the proposed method, in such way that studies report experiments performed with data from two databases, usually BU-3DFE and Bosphorus. Examples of the strategy of generality verification are found in [15,27,34,42]. In this review, amongst those studies making use of Bosphorus, only [29,40,57] use solely that database for validation. Bosphorus do not provide indication of graduated intensity of facial expressions. Therefore, solely the counts of subjects and expressions are considered as sufficient elements to completely describe the samples of data employed in the studies.

Despite Bosphorus database includes scans of 105 subjects, only 65 of them have samples of all the prototypical expressions. That is the reason why 3D FER studies that employ data from that database usually consider samples of all 65 subjects or of a subset of them. Among the studies in this review that employ data from Bosphorus, the studies [15,40,56] utilize samples of 65 subjects, while [13,34,42] consider 60 subjects randomly selected out of 65. In [27,47,52], however, that information is not explicitly stated. In the surveyed studies, the use of neutral samples in addition to expressive samples (6E+1N) is as frequent as the use of solely expressive 3D models (6E).

5.1.3. BU-4DFE

The BU-4DFE is the database considered for validation in 9 studies surveyed in this review. As summarized in Table 6, the BU-4DFE is a database of 3D videos. Accordingly, it consists of sequences of 3D frames, capturing subjects while they perform facial expressions of emotions. The sample of data employed in experiments with that database is characterized by the count of subjects, expressions and the set of selected frames. Among those studies reporting experiments with BU-4DFE, only experiments selecting 60 and 100 subjects were identified (7 and 2 studies, respectively). Only in [21,61] Ben Amor and Co-authors execute experiments employing the full sequences provided in the database. The majority select key-frames, as in [25,33,60]; or sub-sequences as in [20,36,50,54]. The most frequent sample of data is formed by key-frames, selected from samples of the 6 prototypical expressions of 60 randomly selected subjects, the so called 60S+6E+KF portion.

5.1.4. Other databases

Other databases were identified in classification experiments of only 3 studies in this review. The publicly available 3D videos database, VT-KFER, is employed in [22], in which only expressions (6E) and their intensities (2HL) were reported. In [19,35], An and Co-authors acquire their own database to be subject to classification experiments, both with the composition 10S+6E.

5.2. Classifiers & classification strategy

The choice for Support Vector Machines (SVMs) is verified in the experiments of 28 surveyed studies. Among them, the use of SVMs is usually justified by its high performance in situations of limited data and high dimensional feature spaces [14,15,41]. Indeed, SVM has registered superior classification accuracies when compared to some other classifiers, as demonstrated in [18,24]. In the worst case, in [26,33], SVM reaches recognition rates sufficiently close to the ones achieved, in similar experiments, with Hidden Markov Models (HMMs) and Fisher's Linear Discriminant Analysis (FLDA), respectively. Some other algorithms for automatic classification are less frequently found in the surveyed studies, such as Random Forests [20,46,52], Nearest-neighbor [28,38,54], K-nearest neighbors [18,24,31] and Artificial Neural Networks [35,44].

Regarding the classification strategy, 43% of the surveyed studies report the use of 10-fold cross-validation (10-CV) to estimate the effectiveness of the predictive model, which makes that technique the most frequently adopted. In that scheme, the data sample is split into 10 groups and, for each unique group, the data is used as a hold out or test set, while the data belonging to the remaining groups compose the training set. In a unique execution of 10-CV, each turn, the classifier is guaranteed to be tested against new data. A simpler alternative to 10-CV, the 90%/10% split, is the second most employed procedure for validation. In that case, 90% of the data sample is selected as training set, while the remaining data is considered for test. Many other variations of split are less frequently found in the 3D FER studies surveyed. Fig. 12 shows all the validation strategies found regarding the training/test proportions.

Data partitioning, followed by automatic classification may be repeated for statistical analysis of the classifier, when the average of the recognition rates along multiple rounds is reported. The number of classification rounds adopted in experiments, in the surveyed studies, vary from a unique execution to 1000. The unique

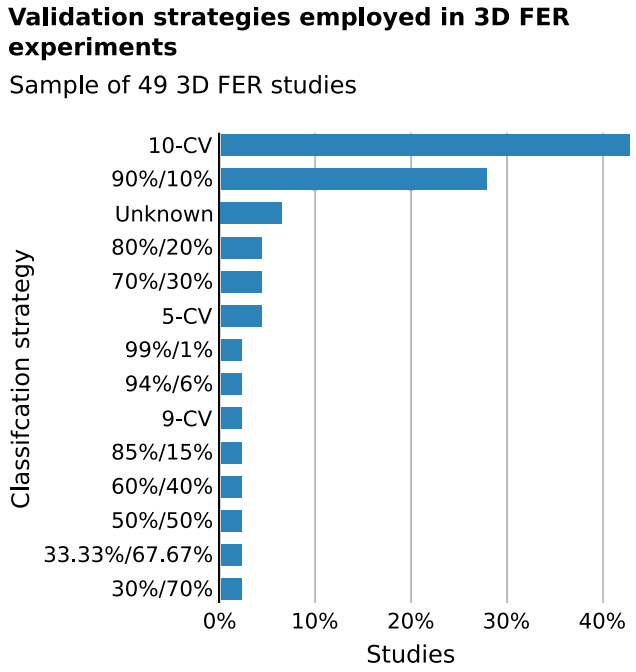

Fig. 12. The employed proportion between training and test data in the surveyed 3D FER studies.

execution of 10-fold cross-validation is the most frequent setup, present in 17% of the studies. Yet, 25% neglected that information.

The evaluation of the predictive model is usually presented to assess the effectiveness of the proposed method. The classification accuracy, which is the most basic metric for evaluation, is widely utilized to report classification performances and is usually referred to as recognition rate (RR) in studies in this domain. The median values of the multiple RRs per class of emotion reported in the surveyed studies evidence the expressions of Happiness and Surprise as the most easily distinguishable from the others, as well as the expressions of Fear and Sadness as the most challenging ones. From another perspective, the standard deviation of the RRs of diverse 3D FER methods reveals a greater consistency in the classification accuracies reached for Happiness and Surprise and higher fluctuations for the other expressions. Fig. 13 presents the median classification accuracies achieved in the surveyed studies discriminated by class of emotion.

The comparison between RRs must be taken as a careful task, since there might exist several particularities unrelated to the proposed method itself that can influence the final results, such as the data and the classification strategy. In Table 7 the five best performing experiments with regard to the achieved RRs are grouped by database and type of data.

5.3. Discussion about classification experiments

The appropriate definition of the classification experiment is a critical aspect for the comparison between 3D FER methods. In addition to quality metrics used for the quantitative evaluation of the effectiveness of a new method, it is important to be able to perform the comparison between studies that employ the same databases. To accomplish that, the experimental scenarios should keep maximum similarities, so the comparison between results of different methods is as fair as possible.

Among the databases considered in this review, the BU-3DFE stands out for being utilized in experiments of 66% of the surveyed studies. The high utilization of data from BU-3DFE is justified, in part, by that being the first 3D face database created for the study of facial expressions, as revealed by the chronology presented in [6]. Even with the advent of other databases suitable to the same purpose, the BU-3DFE database has been consolidated as the benchmark for experiments in 3D FER. Despite this preference, there is still no agreement about the composition of the sample of data that constitutes the data set effectively utilized in the classification experiments reported in the studies. On the contrary, we registered a significant variability in the choice of the count of subjects, expressions and degree of intensity considered for the expressions. The most frequent combination of those characteris-

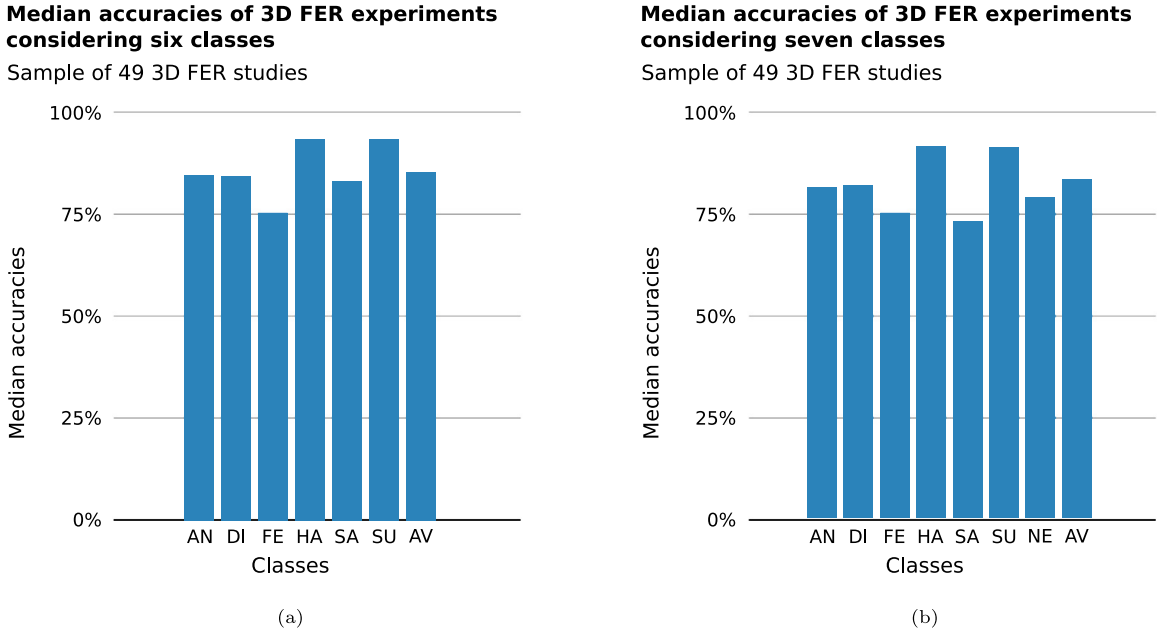

Fig. 13. Median accuracies reported in the 3D FER studies surveyed. AN = Anger, DI = Disgust, FE = Fear, HA = Happiness, SA = Sadness, SU = Surprise, NE = Neutral and AV = AVERAGE. Considering the impact that the number of classes might have over the classification effectiveness, the accuracies are grouped by (a) experiments that consider the six basic emotions as the classes and (b) experiments that also consider the neutral expressions as a 7th class.

Table 7
The best performing experiments regarding the RRs categorized by database, data type and count of classes.

	FER studies	Data sample	Classifier	Classification strategy	RR
6 basic emotions					
BU-4DFE					
4D					
	Zhen et al. [33]	60S+6E+KF	HMM	10-CV	95.13%
	Maalej et al. [25]	100S+6E+KF	HMM	10-CV	94.45%
	Amor et al. [20]	60S+6E+WIN	HMM	10-CV	93.83%
	Li et al. [68]	60S+6E+Full seq.	DGIN	10-CV	92.22%
	Hayat et al. [36]	60S+6E+WIN	Clustering on Grasmannian manifold	10 × 10-CV	90.97%
BU-3DFE					
2D+3D					
	Li et al. [45]	60S+6E+2L	RBF-SVM	100 × 90%/10%	94.56%
	Jan and Meng [27]	60S+6E+2L	SVM	100 × 10-CV	90.04%
	Li et al. [34]	100S+6E+2L	DF-CNN (svm)	100 × 10-CV	86.86%
	Li et al. [42]	60S+6E+2L	RBF-SVM	100 × 90%/10%	86.32%
	Azazi et al. [13]	60S+6E+NR	SVM with EPE	100 × 10-CV	85.81%
3D					
	Hariri et al. [15]	100S+6E+2L	SVM	10 × 90%/10%	92.62%
	Sheng et al. [59]	60S+6E+2L	SVM	20 × 94%/6%	92.1%
	Li et al. [23]	60S+6E+NR	SVM	1000 × 90%/10%	91.3%
	Yurtkan and Demirel [49]	100S+6E+1L	SVM	2 × 90%/10%	90.8%
	An and Ruan [18]	100S+6E+1L	SVM	90%/10%	90.17%
Bosphorus					
2D+3D					
	Li et al. [42]	60S+6E	RBF-SVM	100 × 90%/10%	84.33%
	Li et al. [34]	60S+6E	DF-CNN (svm)	100 × 10-CV	80.28%
	Moeni et al. [47]	6E*	SVM	10-CV	97.3%
3D					
	Zhang et al. [57]	56S+6E	CMTNN	5-CV	92.2%
	Trimech et al. [40]	65S+6E	SVM	80%/20%	73.07%
6 basic emotions + 1 neutral					
BU-3DFE					
2D+3D					
	Jan and Meng, [27]	60S+7E+2L	SVM	100 × 10-CV	88.32%
3D					
	Khashman and Conkbayir [41]	60S+7E+1L	MSVM	80%/20%	95.2%
	Savran and Sankur [37]	100S+7E+2L	Adaboost with NBC as "weak learner"	10-CV	83.2%
	Han and Ming [31]	100S+7E *	k-NN	50%/50%	77.3%
	Vieriu et al. [52]	100S+7E+2L	Random Forest	5-CV	73.71%
Bosphorus					
2D+3D					
	Moeni et al. [47]	7E**	SVM	10-CV	96%
	Azazi et al. [13]	60S+7E	SVM with EPE	10-CV	84%
	Jan and Meng [27]	7E**	SVM	100 × 10-CV	79.46%
3D					
	Azazi et al. [29]	65S+7E	RBF-SVM	10-CV	79%
	Hariri et al. [15]	65S+7E	SVM	10 × 90%/10%	86.17%

* Level of intensity of expressions used in the experiments is not reported.
** The count of subjects considered in the experiments is not reported.

tics form the portion 60S+6E+2HL, which represents only 29% of the studies with BU-3DFE. Moreover, approximately 21% of studies reporting experiments with data from that database do not sufficiently characterize the sample of data, which might impair credibility and difficult comparison between distinct methods.

The classification strategy is another important aspect for the fair comparison between methods. The literature in the 3D FER domain consistently adopts the 90%/10% partitioning into training and test sets, either in the 10-CV scheme or in the form of simple split. Those validation setups are frequently executed a unique time or repeated only a few times. However, the impact of the count of rounds of classification in 3D FER is acknowledged. It has been demonstrated that large fluctuating recognition accuracies result from experiments repeated only 10 and 20 times [69], and that higher counts of rounds of independent classification experiments guarantee stable results. Although a more methodological than technical issue, that practice arises as a current challenge in this community for purposes of comparison, as it can severely impact the reported results and must always be considered for fair comparison.

6. Deep learning based methods

Traditional methods adopt an approach based on the sequential execution of preprocessing, face representation decisions and automatic classification. It is assumed that the outcome of each task feeds the next. Usually, major contributions presented in traditional 3D FER methods are those related to face representation, which include aforementioned points, such as definition of ROIs, extraction and selection of attributes, etc. On the other hand, deep learning based methods, present in 6% of the studies in this review, have been recently more explored in this domain. Such approaches often propose an end-to-end training and prediction system, in which the operation of the deep network itself go without an *a priori* explicit definition of ROIs, feature extraction and selection, in such way that deep networks learn the appropriate features to discriminate classes. Typically, in those works, the focus is the particularization of a deep network.

For example, in [34], Li et al. present a powerful combination of multimodal data (2D+3D) with a Deep Fusion Convolutional Neural Network (DF-CNN). In that study, textured 3D face scans

are initially mapped into 2D facial attribute representations (geometry map, three normal component maps, normalized curvature map and texture map), which are fed into the feature fusion subnet. Further, the generated multi-channel feature maps are then fed into the feature fusion subnet, resulting in a highly concentrated facial representation followed by network training with the softmax-loss layer. In [68], Li et al. propose Dynamic Geometrical Image Network (DGIN) for end-to-end training and prediction and also have as input geometrical images derived from 3D videos.

Naturally, some degree of handcrafted face representation may precede the operation of a deep network. For example, in [70] Uddin et al. perform the extraction of Modified Local Direction Pattern (MLDP) features prior to deep learning and recognition. In [63], Chen et al. present a 3D deep learning manifold based method by means of the proposition of a Fast and Light Manifold Convolutional Neural Network (FLM-CNN). In that study, Sampling Patch Operators (SPO) are calculated to produce patch based features of each sampled point in face scans.

Moreover, although experiments with learned features [34] consistently demonstrate superior recognition rates compared to handcrafted ones, the employment of deep learning based methods involves having to deal with some important limitations. In practice, deep networks need to be trained with a large amount of data since they are highly prone to overfitting. Especially in 3D FER, that requirement is potentially critical, since there are a very limited number of 3D face databases available which have, in turn, a very limited number of 3D face scans (we discuss the employment of 3D face databases in Section 5.1). In virtue of that, 3D FER solutions based on deep learning can, for example, employ pre-trained deep models to initialize their feature extraction subnet [34] or resort to an strategy of data augmentation [68]. In addition, multiple geometrical representations of a same 3D face scan are often the input of those deep networks.

7. Conclusions

The outcomes of the systematic review process described in this work allows the description of three major aspects of 3D FER studies published between 2013 and 2018: face representation techniques, types of required preprocessing and, finally, the nature of classification experiments.

When considering traditional approaches of face representation, a preference towards purely 3D oriented methods stands out. Even though multimodal methods register better recognition accuracies, the lower complexity and, therefore, lower computational cost seems to justify the employment of simpler data, as opposed to the use of textured 3D images or 3D sequences. Complementary, in this context, techniques based on local scope ROIs are also preferred, as handcrafted features extracted from such regions allows to disregard irrelevant characteristics and makes it possible to ignore areas with coincident characteristics among classes of emotions that do not contribute to and may decrease the accuracy of the model. Moreover, despite the high dependence on keypoints extensively reported in the surveyed studies, it is identified an increasing trend in the proportion of keypoints-independent methods published from 2013 to 2017. Investigations of this nature are consistent with the pursuit of complexity reduction, since methods dependent on keypoints, in general, also depend on highly precise detection, which is not a trivial task and increases the computational cost.

Deep learning based approaches have been increasingly adopted in 3D FER, especially in the past two years as an alternative to handcrafted feature extraction. Although such methods require some special conditions to function properly as for example large amount of data for training, which itself is an important limitation in 3D FER, they favor the exploitation of viable solutions robust to adverse conditions, such as pose variation.

Additionally, a more result-oriented analysis revealed that the expressions of Happiness and Surprise stand out as the most consistently distinguishable ones, while Fear and Sadness are revealed to be the most challenging expressions, representing, therefore an opportunity for dedicated future works.

Furthermore, some methodological issues are evidenced as important challenges to be overcome by this community. The capability to replicate scientific experiments is an important premise for scientific progress. In spite of that, we identify in 3D FER studies the practice of neglecting relevant information about, mostly, preprocessing procedures, but also regarding the experimental design. It may lead to undesirable consequences in the area, such as, the adoption of unfair comparisons, producing inaccurate conclusions and, in the worst case, discouraging continuity, and therefore the eventual improvements of such works. Specially with regard to the classification experiments, the community is currently dealing with miscellaneous setups. That includes heterogeneous data samples, even among studies using the same 3D face database, and diverse classification strategies.

We understand that it is not possible to have an absolute agreement about the experimental procedures in 3D FER. Scientific investigations are not rigid, but naturally exploitative. However, it is important to note that some efforts to appropriately detail methodological aspects could greatly benefit this research community. In that direction, we used the findings from this systematic review to outline essential methodological decisions whose a more regardful report is, in our understanding, of great value to the future works.

Declaration of Competing Interest

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

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. We would like to thank the fellow researchers Maria Estela de O. Paiva, Artur R. Rocha Neto, Polycarpo Souza Neto and Izaías Emídio M. Junior for their valuable contributions in the selection of the studies included in this review.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patcog.2019.107108.

References

[1] B. Schuller, M. Wimmer, D. Arsic, T. Moosmayr, G. Rigoll, Detection of security related affect and behaviour in passenger transport, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, ISCA-INT SPEECH COMMUNICATION ASSOC, Brisbane, Australia, 2008, pp. 265–268.
[2] A. Luneski, E. Konstantinidis, P.D. Bamidis, Affective medicine a review of affective computing efforts in medical informatics, Methods Inf. Med. 49 (3) (2010) 207–218, doi:10.3414/me0617.
[3] P. Ekman, Universals and cultural differences in facial expressions of emotion, in: Nebraska Symposium on Motivation, 19, 1971, pp. 207–283.
[4] P. Ekman, W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, Palo Alto, 1978.
[5] G. Sandbach, S. Zafeiriou, M. Pantic, L. Yin, Static and dynamic 3D facial expression recognition: a comprehensive survey, Image Vis. Comput. 30 (10) (2012) 683–697, doi:10.1016/j.imavis.2012.06.005.

[6] C.A. Corneanu, M.O. Simón, J.F. Cohn, S.E. Guerrero, C. Adrian Corneanu, M. Oliu Simon, J.F. Cohn, S. Escalera Guerrero, Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications, IEEE Trans. Pattern Anal. Mach. Intell. 38 (8) (2016) 1548–1568, doi:10.1109/tpami.2016.2515606.

[7] S. Soltanpour, B. Boufama, Q.M. Jonathan Wu, A survey of local feature methods for 3D face recognition, Pattern Recognit. 72 (C) (2017) 391–406, doi:10.1016/j.patcog.2017.08.003.

[8] A. Danelakis, T. Theoharis, I. Pratikakis, A survey on facial expression recognition in 3D video sequences, Multimedia Tools Appl. 74 (15) (2015) 5577–5615, doi:10.1007/s11042-014-1869-6.

[9] S. Li, W. Deng, Deep Facial Expression Recognition: A Survey, 2018, pp. 1–25.

[10] B. Kitchenham, Procedures for performing systematic reviews, Keele University, Keele, UK, 2004. 10.1.1.122.3308.

[11] J.C. de Almeida Biolchini, P.G. Mian, A.C.C. Natali, T.U. Conte, G.H. Travassos, Scientific research ontology to support systematic review in software engineering, Adv. Eng. Inform. 21 (2) (2007) 133–151, doi:10.1016/j.aei.2006.11.006.

[12] T. Zhang, Facial expression recognition based on deep learning: a survey, in: Advances in Intelligent Systems and Computing, 686, 2018, pp. 345–352, doi:10.1007/978-3-319-69096-4_48.

[13] A. Azazi, S. Lebai Lutfi, I. Venkat, F. Fernández-Martínez, Towards a robust affect recognition: automatic facial expression recognition in 3D faces, Expert Syst. Appl. 42 (6) (2015) 3056–3066, doi:10.1016/j.eswa.2014.10.042.

[14] K. Yurtkan, H. Demirel, Feature selection for improved 3D facial expression recognition, Pattern Recognit. Lett. 38 (1) (2014) 26–33, doi:10.1016/j.patrec.2013.10.026.

[15] W. Hariri, H. Tabia, N. Farah, A. Benouareth, D. Declercq, 3D facial expression recognition using kernel methods on Riemannian manifold, Eng. Appl. Artif. Intell. 64 (C) (2017) 25–32, doi:10.1016/j.engappai.2017.05.009.

[16] M. Emambakhsh, A. Evans, Nasal patches and curves for expression-robust 3D face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 39 (5) (2017) 995–1007, doi:10.1109/TPAMI.2016.2565473.

[17] B. Hasani, M.H. Mahoor, Facial expression recognition using enhanced deep 3D convolutional neural networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, IEEE, 2017, pp. 2278–2288, doi:10.1109/CVPRW.2017.282.

[18] S. An, Q. Ruan, 3D facial expression recognition algorithm using local threshold binary pattern and histogram of oriented gradient, in: 2016 IEEE 13th International Conference on Signal Processing (ICSP), IEEE, 2016, pp. 265–270, doi:10.1109/ICSP.2016.7877838.

[19] S. Aly, A. Youssef, L. Abbott, Adaptive feature selection and data pruning for 3D facial expression recognition using the Kinect, in: 2014 IEEE International Conference on Image Processing (ICIP), IEEE, Paris, 2014, pp. 1361–1365, doi:10.1109/ICIP.2014.7025272.

[20] B. Ben Amor, H. Drira, S. Berretti, M. Daoudi, A. Srivastava, 4-D facial expression recognition by learning geometric deformations, IEEE Trans. Cybern. 44 (12) (2014) 2443–2457, doi:10.1109/TCYB.2014.2308091.

[21] S. Berretti, A. Del Bimbo, P. Pala, Automatic facial expression recognition in real-time from dynamic sequences of 3D face scans, Vis. Comput. 29 (12) (2013) 1333–1350, doi:10.1007/s00371-013-0869-2.

[22] S. Aly, A.L. Abbott, M. Torki, A multi-modal feature fusion framework for kinect-based facial expression recognition using Dual Kernel Discriminant Analysis (DKDA), in: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2016, pp. 1–10, doi:10.1109/WACV.2016.7477577.

[23] Gaoyun An, Qiuqi Ruan, Xiaoli Li, 3D facial expression recognition using delta faces, in: 5th IET International Conference on Wireless, Mobile and Multimedia Networks (ICWMMN 2013), vol. 2013, Institution of Engineering and Technology, 2013, pp. 4–10, doi:10.1049/cp.2013.2415.

[24] A. Azazi, S.L. Lutfi, I. Venkat, Analysis and evaluation of SURF descriptors for automatic 3D facial expression recognition using different classifiers, in: 2014 4th World Congress on Information and Communication Technologies, WICT 2014, IEEE, 2014, pp. 23–28, doi:10.1109/WICT.2014.7077296.

[25] A. Maalej, H. Tabia, H. Benhabiles, Dynamic 3D facial expression recognition using robust shape features, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7944 LNCS, 2013, pp. 309–318, doi:10.1007/978-3-642-38886-6_30.

[26] D. Derkach, F.M. Sukno, Local shape spectrum analysis for 3D facial expression recognition, in: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), IEEE, 2017, pp. 41–47, doi:10.1109/FG.2017.143.

[27] A. Jan, Hongying Meng, Automatic 3D facial expression recognition using geometric and textured feature fusion, in: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 05, IEEE, 2015, pp. 1–6, doi:10.1109/FG.2015.7284860.

[28] Y. Fu, Q. Ruan, G. An, Y. Jin, Fast nonnegative tensor factorization based on graph-preserving for 3D facial expression recognition, in: 2016 IEEE 13th International Conference on Signal Processing (ICSP), in: International Conference on Signal Processing, IEEE, 2016, pp. 292–297, doi:10.1109/ICSP.2016.7877843.

[29] A. Azazi, S.L. Lutfi, I. Venkat, Identifying universal facial emotion markers for automatic 3D facial expression recognition, in: 2014 International Conference on Computer and Information Sciences (ICCOINS), IEEE, 2014, pp. 1–6, doi:10.1109/ICCOINS.2014.6868369.

[30] M.D. Samad, K.M. Iftekharuddin, Frenet frame-based generalized space curve representation for pose-invariant classification and recognition of 3-D face, IEEE Trans. Hum.-Mach. Syst. 46 (4) (2016) 522–533, doi:10.1109/THMS.2016.2515602.

[31] D. Han, Y. Ming, Facial expression recognition with LBP and SLPP combined method, in: 2014 12th International Conference on Signal Processing (ICSP), 2015, IEEE, 2014, pp. 1418–1422, doi:10.1109/ICOSP.2014.7015233.

[32] Q. Li, G. An, Q. Ruan, 3D Facial expression recognition using orthogonal tensor marginal fisher analysis on geometric maps, in: 2017 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), vol. 35, IEEE, 2017, pp. 65–71, doi:10.1109/ICWAPR.2017.8076665.

[33] Q. Zhen, D. Huang, H. Drira, B.B. Amor, Y. Wang, M. Daoudi, Magnifying subtle facial motions for effective 4D expression recognition, IEEE Trans. Affect. Comput. PP (99) (2017) 1, doi:10.1109/TAFFC.2017.2747553.

[34] H. Li, J. Sun, Z. Xu, L. Chen, Multimodal 2D+3D facial expression recognition with deep fusion convolutional neural network, IEEE Trans. Multimed. 19 (12) (2017) 2816–2831, doi:10.1109/TMM.2017.2713408.

[35] S. Elhoufi, M. Jazouli, A. Majda, A. Zarghili, R. Aalouane, Automatic recognition of facial expressions using microsoft kinect with artificial neural network, in: 2016 International Conference on Engineering & MIS (ICEMIS), IEEE, 2016, pp. 1–5, doi:10.1109/ICEMIS.2016.7745376.

[36] M. Hayat, M. Bennamoun, A.A. El-Sallam, Clustering of video-patches on Grassmannian manifold for facial expression recognition from 3D videos, in: Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV), in: WACV '13, IEEE Computer Society, Washington, DC, USA, 2013, pp. 83–88, doi:10.1109/WACV.2013.6475003.

[37] A. Savran, B. Sankur, Non-rigid registration based model-free 3D facial expression recognition, Comput. Vis. Image Underst. 162 (C) (2017) 146–165, doi:10.1016/j.cviu.2017.07.005.

[38] M. Jazouli, A. Majda, A. Zarghili, A $P recognizer for automatic facial emotion recognition using Kinect sensor, in: 2017 Intelligent Systems and Computer Vision (ISCV), IEEE, 2017, pp. 1–5, doi:10.1109/ISACV.2017.8054955.

[39] P. Lemaire, M. Ardabilian, L. Chen, M. Daoudi, Fully automatic 3D facial expression recognition using differential mean curvature maps and histograms of oriented gradients, in: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), IEEE, 2013, pp. 1–7, doi:10.1109/FG.2013.6553821.

[40] I.H. Trimech, A. Maalej, N.E.B. Amara, 3D facial expression recognition using nonrigid CPD registration method, in: 2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), IEEE, 2016, pp. 478–481, doi:10.1109/SETIT.2016.7939917.

[41] A. Khashman, F.O. Conkbayir, Intelligent recognition of emotional expressions in 3D face images, in: 2013 21st Signal Processing and Communications Applications Conference (SIU), IEEE, 2013, pp. 1–4, doi:10.1109/SIU.2013.6531222.

[42] H. Li, H. Ding, D. Huang, Y. Wang, X. Zhao, J.-M. Morvan, L. Chen, An efficient multimodal 2D + 3D feature-based approach to automatic facial expression recognition, Comput. Vision Image Understand. 140 (C) (2015) 83–92, doi:10.1016/j.cviu.2015.07.005.

[43] W. Zeng, H. Li, L. Chen, J.-M. Morvan, X.D. Gu, An automatic 3D expression recognition framework based on sparse representation of conformal images, in: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), IEEE, 2013, pp. 1–8, doi:10.1109/FG.2013.6553749.

[44] P. Suja, D. Krishnasri, S. Tripathi, Pose invariant method for emotion recognition from 3D images, in: 12th IEEE International Conference Electronics, Energy, Environment, Communication, Computer, Control: (E3-C3), INDICON 2015, IEEE, 2016, pp. 1–5, doi:10.1109/INDICON.2015.7443288.

[45] X. Li, Q. Ruan, G. An, Y. Jin, R. Zhao, Multiple strategies to enhance automatic 3D facial expression recognition, Neurocomputing 161 (C) (2015) 89–98, doi:10.1016/j.neucom.2015.02.063.

[46] L. Tao, B.J. Matuszewski, Is 2D unlabeled data adequate for recognizing facial expressions? IEEE Intell. Syst. 31 (3) (2016) 19–29, doi:10.1109/MIS.2016.25.

[47] A. Moeini, K. Faez, H. Sadeghi, H. Moeini, 2D Facial expression recognition via 3D reconstruction and feature fusion, J. Vis. Commun. Image Represent. 35 (C) (2016) 1–14, doi:10.1016/j.jvcir.2015.11.006.

[48] O. Ocegueda, Tianhong Fang, S.K. Shah, I.A. Kakadiaris, 3D face discriminant analysis using Gauss-Markov posterior marginals, IEEE Trans. Pattern Anal. Mach. Intell. 35 (3) (2013) 728–739, doi:10.1109/TPAMI.2012.126.

[49] K. Yurtkan, H. Demirel, Entropy-based feature selection for improved 3D facial expression recognition, Signal Image Video Process. 8 (2) (2014) 267–277, doi:10.1007/s11760-013-0543-1.

[50] M. Reale, X. Zhang, L. Yin, Nebula feature: a space-time feature for posed and spontaneous 4D facial behavior analysis, in: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), IEEE, 2013, pp. 1–8, doi:10.1109/FG.2013.6553746.

[51] K. Yurtkan, H. Demirel, Person independent facial expression recognition using 3D facial feature positions, in: Computer and Information Sciences III - 27th International Symposium on Computer and Information Sciences, ISCIS 2012, 2013, pp. 321–329, doi:10.1007/978-1-4471-4594-3-33.

[52] R.-L. Vieriu, S. Tulyakov, S. Semeniuta, E. Sangineto, N. Sebe, Facial expression recognition under a wide range of head poses, in: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 1, IEEE, 2015, pp. 1–7, doi:10.1109/FG.2015.7163098.

[53] K. Yurtkan, H. Soyel, H. Demirel, Feature selection for enhanced 3D facial expression recognition based on varying feature point distances, in: E. Gelenbe, R. Lent (Eds.), Information Sciences and Systems 2013, Lecture Notes in Electrical Engineering, vol. 264, 2013, pp. 209–217, doi:10.1007/978-3-319-01604-7_21.

[54] M. Xue, A. Mian, W. Liu, L. Li, Automatic 4D facial expression recognition using DCT features, in: 2015 IEEE Winter Conference on Applications of Computer Vision, IEEE, 2015, pp. 199–206, doi:10.1109/WACV.2015.34.
[55] P. Zarbakhsh, H. Demirel, Fuzzy SVM for 3D facial expression classification using sequential forward feature selection, in: 2017 9th International Conference on Computational Intelligence and Communication Networks (CICN), IEEE, 2017, pp. 131–134, doi:10.1109/CICN.2017.8319371.
[56] M. Xue, A. Mian, W. Liu, Ling Li, Fully automatic 3D facial expression recognition using local depth features, in: IEEE Winter Conference on Applications of Computer Vision, IEEE, 2014, pp. 1096–1103, doi:10.1109/WACV.2014.6835736.
[57] Y. Zhang, L. Zhang, M. Hossain, Adaptive 3D facial action intensity estimation and emotion recognition, Expert Syst. Appl. 42 (3) (2015) 1446–1464, doi:10.1016/j.eswa.2014.08.042.
[58] Xudong Yang, Di Huang, Yunhong Wang, Liming Chen, Automatic 3D facial expression recognition using geometric scattering representation, in: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), IEEE, 2015, pp. 1–6, doi:10.1109/FG.2015.7163090.
[59] N. Sheng, Y. Cai, C. Zhan, Changyan Qiu, Yize Cui, Xurong Gao, 3D facial expression recognition using distance features and LBP features based on automatically detected keypoints, in: 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE, Datong, 2016, pp. 396–401, doi:10.1109/CISP-BMEI.2016.7852743.
[60] Y. Yao, D. Huang, X. Yang, Y. Wang, L. Chen, Texture and geometry scattering representation-based facial expression recognition in 2D+3D videos, ACM Trans. Multimed. Comput. Commun. Appl. 14 (1s) (2018) 1–23, doi:10.1145/3131345.
[61] Q. Zhen, D. Huang, Y. Wang, L. Chen, Muscular movement model-based automatic 3D/4D facial expression recognition, IEEE Trans. Multimed. 18 (7) (2016) 1438–1450, doi:10.1109/TMM.2016.2557063.
[62] G. Sandbach, S. Zafeiriou, M. Pantic, D. Rueckert, A dynamic approach to the recognition of 3D facial expressions and their temporal models, in: Face and Gesture 2011, IEEE, 2011, pp. 406–413, doi:10.1109/FG.2011.5771434.
[63] Z. Chen, D. Huang, Y. Wang, L. Chen, Fast and light manifold CNN based 3D facial expression recognition across pose variations, in: Proceedings of the 26th ACM international conference on Multimedia, 2018, pp. 229–238, doi:10.1145/3240508.3240568.
[64] L. Yin, X. Wei, Y. Sun, J. Wang, M.J. Rosato, A 3D facial expression database for facial behavior research, in: FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, 2006, 2006, pp. 211–216, doi:10.1109/FGR.2006.6.
[65] A. Savran, N. Alyüz, H. Dibekliolu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, Bosphorus database for 3D face analysis, in: European Workshop on Biometrics and Identity Management, 5372 LNCS, 2008, pp. 47–56, doi:10.1007/978-3-540-89991-4_6.
[66] L. Yin, X. Chen, Y. Sun, T. Worm, M. Reale, A high-resolution 3D dynamic facial expression database, Automatic Face Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on (1) (2008) 1–6, doi:10.1109/AFGR.2008.4813324.
[67] S. Aly, A. Trubanova, L. Abbott, S. White, A. Youssef, VT-KFER: a kinect-based RGBD+time dataset for spontaneous and non-spontaneous facial expression recognition, in: 2015 International Conference on Biometrics (ICB), 2015, pp. 90–97, doi:10.1109/ICB.2015.7139081.
[68] W. Li, D. Huang, H. Li, Y. Wang, Automatic 4D facial expression recognition using dynamic geometrical image network, in: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE, 2018, pp. 24–30, doi:10.1109/FG.2018.00014.
[69] J.L. B. Gong, Y. Wang, X. Tang, Automatic facial expression recognition on a single 3D face by exploring shape deformation, in: ACM International Conference on Multimedia, 2009, pp. 569–572.
[70] M.Z. Uddin, M.M. Hassan, A. Almogren, M. Zuair, G. Fortino, J. Torresen, A facial expression recognition system using robust face features from depth videos and deep learning, Comput. Electric. Eng. 63 (2017) 114–125, doi:10.1016/j.compeleceng.2017.04.019.

Gilderlane Ribeiro Alexandre received her bachelor's degree in computer engineering from Universidade Federal do Ceará, Brazil. Currently she pursues a master's degree in Teleinformatics Engineering at Universidade Federal do Ceará. Her research interests are in computer vision, pattern recognition and affective computing.

José Marques Soares received his Ph.D. degree from Institut National de Telecommunications, Evry, France. He is currently a professor with the Department of Teleinformatics Engineering at Universidade Federal do Ceará, Brazil. His research interests include distributed systems and computer vision.

George André Pereira Thé received his Ph.D. degree from Polytechnic of Turin, Turin, Italy. He is currently a professor with the Department of Teleinformatics Engineering at Universidade Federal do Ceará, Brazil. His research interests include robotics and computer vision.