J. Vis. Commun. Image R. 98 (2024) 104033

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

Full Length Article

CalD3r and MenD3s: Spontaneous 3D facial expression databases☆

Luca Ulrich [a], Federica Marcolin [a,*], Enrico Vezzetti [a], Francesca Nonis [a], Daniel C. Mograbi [b], Giulia Wally Scurati [d], Nicolò Dozio [c], Francesco Ferrise [c]

[a] Department of Management and Production Engineering, Politecnico di Torino, C.so Duca degli Abruzzi 24, Torino 10129, Italy
[b] Department of Psychology, Pontifical Catholic University of Rio de Janeiro, Rua Marquês de São Vicente 225, Gávea, Rio de Janeiro, Brazil
[c] Department of Mechanical Engineering, Politecnico di Milano, Via Privata Giuseppe La Masa 1, Milano 20156, Italy
[d] Department of Mechanical Engineering, Blekinge Tekniska Högskola, Valhallavägen 1, Karlskrona 371 41, Sweden

ARTICLE INFO

Keywords:
3D facial expression
Spontaneous expressions
Facial expression recognition
Ecological validity
Affective database
Human-computer interaction

ABSTRACT

In the last couple of decades, the research on 3D facial expression recognition has been fostered by the creation of tailored databases containing prototypical expressions of different individuals and by the advances in cost effective acquisition technologies. Though, most of the currently available databases consist of exaggerated facial expressions, due to the imitation principle which they rely on. This makes these databases only partially employable for real world applications such as human-computer interaction for smart products and environments, health, and industry 4.0, as algorithms learn on these 'inflated' data which do not respond to ecological validity requirements. In this work, we present two novel 2D + 3D spontaneous facial expression databases of young adults with different geographical origin, in which emotions have been evoked thanks to affective images of the acknowledged IAPS and GAPED databases, and verified with participants' self-reports. To the best of our knowledge, these are the first three-dimensional facial databases with emotions elicited by validated affective stimuli.

1. Introduction

In the era of affective computing and behavior analysis, developing methodologies for understanding users' emotions has become an essential practice to meet specific needs in the fields of human-computer interaction, smart products design, industry 4.0, education, psychology, communication and medicine [1]. Besides the psychological studies that are nurturing the field with different theories regarding generation, universality, and manifestation of the emotions [2], recently, the advances in artificial intelligence (AI) have paved the way for accurate computer-based facial expression recognition (FER) with machine/deep learning techniques, making of emotional interpretation an interdisciplinary task and a hot topic of research across The advent of low cost technologies for three-dimensional acquisition has encouraged the study of 3D facial data for analysis and recognition purposes [3,4], in particular after the Face Recognition Vendor Test in 2002 [5] and the subsequent Face Recognition Grand Challenge (FRGC) in 2005 [6]. The addition of the third dimension allows to overcome issues like lightning variations, pose and make-up [7–9]. Also, it reflects some aspects of the human visual and cognitive system regarding the mental representation of objects (and faces), as the viewer seems to have an actual three-dimensional perception and understanding of the facial surface, allowing the recognition from any viewpoint [10,11].

Some facial expression databases based on 3D data have been created and made available for the public with the aim to sup-different research fields port the research in the area, in particular to validate novel FER algorithms. Though, the current database landscape presents a limitation regarding the ecological validity of the expressions. *Ecological validity* has typically been taken to refer to whether or not one can generalize from observed behavior in the laboratory to natural behavior in the world [12]. In other words, most of these databases contain non-spontaneous expressions, meaning that the facial muscular behaviour corresponding to a specific emotion has not been generated with emotion elicitation but with other techniques such as imitation. Despite the general initial enthusiasm of dealing with apparently well-acted and prototypical expressions, the result is that these repositories contain over-exaggerated and unnatural expressions that people in real life seldom display, with the consequence that algorithms are spoiled with

☆ This paper has been recommended for acceptance by Caifeng Shan.
* Corresponding author.
E-mail address: federica.marcolin@polito.it (F. Marcolin).

https://doi.org/10.1016/j.jvcir.2023.104033
Received 5 September 2022; Received in revised form 25 October 2023; Accepted 20 December 2023
Available online 21 December 2023

Table 1
Comparison of databases including 3D data and facial expressions, in chronological order of release. S-a (self-assessment) refers to the adoption of self-reports to assess the 'ground truth' emotion; the last column refers to the method to induce the facial expressions, i.e. posed, unposed or elicited.

Database	Year	Subjects	Emotions	S-a	Method
ND-2006	2006	888	7	NA	Posed
BU-3DFE	2006	100	7*4 intensities	NA	Posed
CASIA	2007	123	5	NA	Posed
York	2008	350	Few	NA	Posed
BU-4DFE	2008	101	6	NA	Posed
Bosphorus	2008	105	7 + AUs	NA	Posed
Texas 3DFRD	2010	118	Few	NA	Posed
UMB-DB	2011	143	Few	NA	Posed
3D TEC	2011	214	2	NA	Posed
Photoface	2011	261	Few	NA	Unposed
FaceWarehouse	2013	150	20	NA	Posed
KinectFaceDB	2014	52	Few	NA	Posed
BP4D-Spont	2014	41	8	Yes	Elicited
DMCSv1	2015	35	3	NA	Posed
BP4D+	2016	140	10	Yes	Elicited
CalD3r + MenD3s	2023	104 + 92	7	Yes	Elicited

unrealistic data in the training sets and cannot 'generalize' to reasonable scenarios. The problem of ecological validity of facial databases has been already mentioned in previous literature by referring to missing authenticity of data [11,13–16] and became more prominent with the upcoming urgency to apply FER algorithms to real contexts. In fact, there are substantial differences between posed and spontaneous expressions regarding appearance, timing, and accompanying head movements [17].

In this work, we propose two novel 3D facial expression databases with spontaneous expressions, elicited with a selection of image stimuli from the International Affective Picture System (IAPS) [18] and the Geneva Affective Picture Database (GAPED) [19], to contribute to overcome the ecological validity issue. IAPS and GAPED gather images validated to arouse specific emotions. Despite the existence of other affective databases, such as those recently developed relying on Virtual Reality solutions [20,21], these databases have been chosen because they received a major acknowledgment from the scientific community [22], with the purpose of minimizing the bias of eliciting different emotions in different subjects with the same stimulus. The obtained expressions are assessed with the support of participants' self-reports. The first database, acquired in Italy and called CalD3r (pronounced "Calder"), involves 104 subjects aged 19–35, and is composed by 54 women and 50 men mostly of Southern European origin. The second database, acquired in Brazil and called MenD3s (pronounced "Mendes"), includes 92 participants aged between 18 and 55 and is composed by 46 women and 46 men, mostly Brazilians. They will be made publicly available for free upon publication of this paper. To our knowledge, there is no database that collects 3D facial expressions evoked with the aid of validated affective databases.

The remainder of the paper is structured as follows: Section 2 offers a framework and comparison of existing facial expression databases based on 3D data; Section 3 describes the data acquisition, in terms of participants' characteristics, image selection, experimental setup, and acquisition technology; Section 4 depicts the database content and organization; Section 5 proposes a validation of the present databases with a classifier based on neural networks; Section 6 presents the limitations of the present contribution, and draws some research streams for improvements and future work; in Section 7 conclusions are outlined.

2. Related work

Various databases involving three-dimensional facial expressions are available (Table 1). The most popular ones among the research community include the six basic emotions anger, disgust, fear, joy, sadness, surprise, firstly discretized by Darwin in the XIX century [23] and then

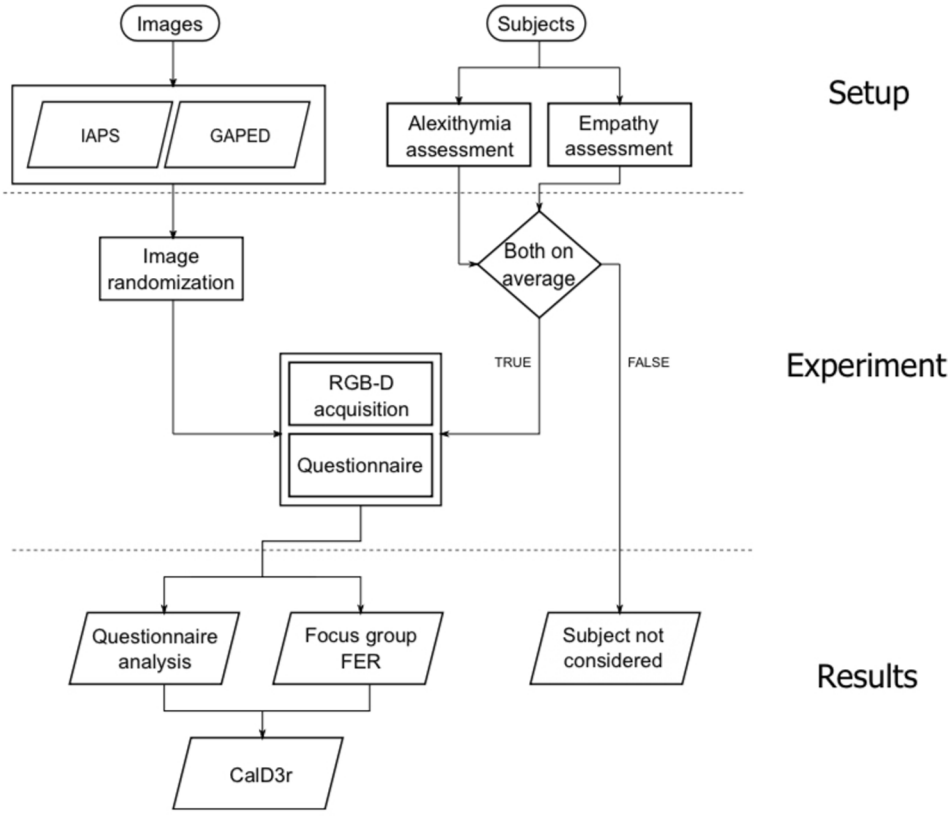

Fig. 1. Methodology flow-chart to produce CalD3r database. The methodology used to produce the MenD3s database was the same.

Table 2
Selected images from IAPS and GAPED. Column Emotion refers to the emotion that the image is supposed to elicit. Columns Description and Code refer to the content and the original IAPS or GAPED image code, respectively. In Valence and Arousal columns, the value in the parenthesis refers to the original numerical valence/arousal value proposed by the Database, reported in the last column, while the other value refers to the re-scaled (1–9) one.

Emotion	Description	Code	Valence	Arousal	Database
Anger	Animal	A008	2.12	5.89	GAPED
Anger	mistreatment	A014	(13.94)	(61.12)	GAPED
Anger	#1	A029	2.08	6.46	GAPED
Anger	Animal	H032	(13.45)	(68.21)	GAPED
Anger	mistreatment	H064	2.40	6.88	GAPED
	#2		(17.56)	(73.50)	
	Animal		1.15	7.23	
	mistreatment		(1.929)	(77.861)	
	#3		1.71	7.46	
	Human		(8.839)	(80.739)	
	mistreatment				
	#1				
	Human				
	mistreatment				
	#2				
Anger	Beaten woman	6315	2.31	6.38	IAPS
Anger	Soldiers	9163	2.10	6.53	IAPS
Anger	Soldier	9414	1.51	7.07	IAPS
Disgust	Mutilation #1	3010	1.79	7.26	IAPS
Disgust	Mutilation #2	3060	1.79	7.12	IAPS
Disgust	Mutilation #3	3068	1.80	6.77	IAPS
Disgust	Mutilation #4	3069	1.70	7.03	IAPS
Disgust	Mutilation #5	3080	1.48	7.22	IAPS
Disgust	Mutilation #6	3130	1.58	6.97	IAPS
Disgust	Baby with tumor	3170	1.46	7.21	IAPS
Disgust	Injury	3266	1.56	6.79	IAPS
Fear	Dog attack	1525	3.09	6.51	IAPS
Fear	Shark	1932	3.85	6.47	IAPS
Fear	Snake #1	1120	3.79	6.93	IAPS
Fear	Snake #2	Sn103	4.94	6.09	GAPED
Fear	Snake #3	Sn125	(49.25)	(63.656)	GAPED
			2.44	6.50	
			(17.954)	(68.729)	
Fear	Spider #1	Sp044	4.85	6.40	GAPED
			(48.162)	(67.493)	
Fear	Spider #2	Sp064	3.94	5.63	GAPED
			(36.696)	(57.856)	
Fear	Spider #3	Sp146	3.20	6.59	GAPED
			(27.499)	(69.825)	
Happiness	Baby #1	P017	8.07	3.38	GAPED
Happiness	Baby #2	P018	(88.431)	(29.739)	GAPED
Happiness	Baby #3	P024	8.03	2.86	GAPED
Happiness	Puppies #1	P074	(87.821)	(23.29)	GAPED
Happiness	Puppies #2	P114	8.21	2.72	GAPED
Happiness	Baby fox	P081	(90.101)	(21.514)	GAPED
			8.19	3.37	
			(89.89)	(29.608)	
			8.68	3.30	
			(95.97)	(28.742)	
			7.83	3.11	
			(85.388)	(26.408)	
Happiness	Kitten	P096	7.77	3.10	GAPED
			(84.617)	(26.211)	
Happiness	Kiss	2352.1	7.27	5.16	IAPS
Neutrality	Antenna	N041	5.40	2.97	GAPED
Neutrality	Bowl	7006	(54.996)	(24.654)	IAPS
Neutrality	Chairs	N089	4.88	2.33	GAPED
Neutrality	Lamp	7175	5.01	2.06	IAPS
Neutrality	Lamp and sofa	N108	(50.168)	(13.26)	GAPED
Neutrality	Mushroom #1	5500	4.87	1.72	IAPS
			5.84	2.10	
			(60.457)	(13.713)	
			5.42	3.00	
Neutrality	Mushroom #2	5531	5.15	3.69	IAPS
Neutrality	Spoon	7004	5.04	2.00	IAPS
Sadness	Animal in	A006	3.20	6.43	GAPED
Sadness	captivity #1	A007	(27.54)	(67.90)	GAPED
Sadness	Animal in	A062	1.80	7.48	GAPED
	captivity #2		(9.95)	(80.94)	

endorsed and deepened by Ekman in the Seventies [24], plus the neutral one. After the already mentioned FRGC [6], the ND-2006, made available in 2006 with 888 individuals, was one of the first with these characteristics [25]. The BU-3DFE, involving 100 subjects displaying four different intensities of emotions [11], was released in the same year, followed by its dynamic version, called BU4DFE [15]. In 2008, Savran et al. proposed the Bosphorus database, with 105 participants including some actors and also occluded data [16]. Despite the large adoption of them for their specificity on emotions, the acted rather than authentic expressions were seen as a major limitation by the authors themselves.

The need for spontaneity was met in 2014 by Zhang et al., who introduced the BP4D-Spontaneous database with 41 participants [26], which was later integrated with physiological data and re-proposed with 140 participants [13]. To elicit target emotional expressions, an experimenter with a professional actor/director career performed specific activities with participants such as interviews, video-clip viewing, improvisation, insults, smells. Self-reports and naive observers were adopted for emotion assessment. The Photoface database is another spontaneous database with 2D + 3D data, involving 261 subjects. The capture process was carried out in a realistic commercial setting, but without the specific intention or result to elicit certain emotions [27].

Other databases such as Gavadb [28], CASIA [29], York [30], Texas 3DFRD [31], UMB-DB [7], 3D TEC [32], FaceWarehouse [33], KinectFaceDB [8], DMCSv1 [34] present different facial expressions but do not involve all the six basic emotions. Also, considering that they have been mainly conceived for subject identification purposes rather than FER, none of them meets the ecological validity desiderata of the spontaneity of emotions.

3. Data acquisition

Differently from the vast majority of the available facial datasets regarding emotions, these databases aim to frame spontaneous facial expressions. For this purpose, an *ad hoc* experiment has been set up, selecting the visual stimuli as the source of emotion elicitation. The methodology behind the experiment setup is explained in the following subsections and schematized in Fig. 1.

3.1. Image selection

Participants have been asked to see a selected set of images retrieved from IAPS [18] and GAPED [19], two of the so-called 'affective databases'. These databases are composed of a wide range of pictures selected to elicit a specific emotional response and have been validated on the dimensions of affect valence and arousal. The first dimension quantifies whether the perceived emotion has a positive or negative prevalence, while the latter identifies the level of activation linked to the perceived emotional response. Nevertheless, some inconsistencies regarding valence and arousal values of images with similar semantic contents have been noticed; therefore, also considering the work of Bradley and Lang [35], and Bradley et al. [36], images have been selected considering cultural and generational aspects and categorized into six emotional entities, labelled as anger, disgust, fear, happiness, sadness, and neutrality. Surprise has been excluded as target as its

Table 2 (*continued*)

Emotion	Description	Code	Valence	Arousal	Database
Sadness	Animal in	A127	2.11	6.38	GAPED
Sadness	captivity #3	9908	(13.92)	(67.25)	IAPS
	Animal in		2.08	5.68	
	captivity #4		(13.46)	(58.44)	
	Car accident		2.34	6.63	
Sadness	Injured child	3301	1.80	5.21	IAPS
Sadness	Sad child	2800	1.78	5.49	IAPS
Sadness	Toddler	2095	1.79	5.25	IAPS

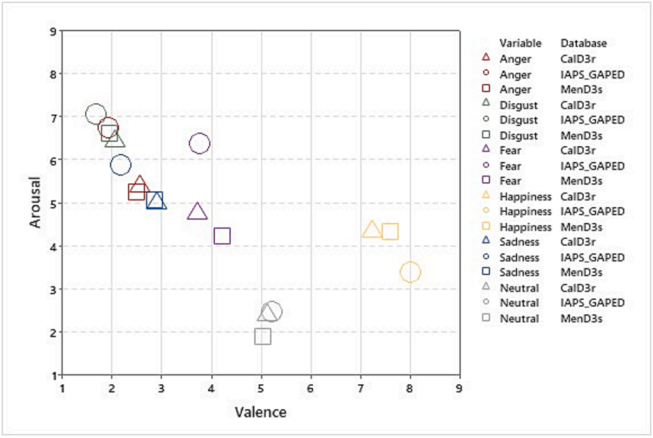

Fig. 2. Valence and arousal comparison between collected SAM responses and the original values provided by IAPS and GAPED databases. Matching colours between CalD3r (triangles)/MenD3s (squares) and IAPS-GAPED (circles) mean that the most rated emotion by participants was effectively the expected one (red for anger, green for disgust, purple for fear, yellow for happiness, blue for sad, grey for neutral). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

triggers are sudden and unexpected occurrences, such as accidents, hardly reproducible with static images, and for its controversial valence (could be pleasant or unpleasant) [37]. In Bradley and Lang's [35] and Bradley et al.'s [36] studies, the authors classified the picture content categories showing the images to a wide audience of men and women, providing a solid baseline to assess the expected emotion related to specific scenarios. Details of the emotional labelling process are reported in Nonis et al.'s work [38]. 60 images have been initially selected, but following a pilot test the number has been reduced not to cause a drop of attention in participants due to the experiment length, ending up with 8 pictures per category, for a total of 48 images (Table 2).

3.2. Participants

104 participants (54 women and 50 men), mostly of Southern European origin and aged between 19 and 35 ($M = 24.7$, $SD = 3.27$), have been involved in the creation of the CalD3r database. The MenD3s database includes 92 participants (46 women and 46 men), aged between 18 and 55 ($M = 23.2$, $SD = 6.70$), mostly Brazilian. Overall, the two databases counted 196 participants.

All the participants were asked to have their alexithymia and empathy assessed. Two different questionnaires were used to evaluate their ability to identify and to understand others' point of views, thoughts, intentions and beliefs [39] and their ability in recognizing, describing, and understanding their own emotions [40]. The Toronto Structured Interview for Alexithymia (TAS-20) has been adopted to evaluate participants' alexithymia [41]. For the empathy assessment of CalD3r participants, an Italian translated and validated version of the Balanced Emotional Empathy Scale (BEES) [42] has been adopted. Results proved that all the selected participants were suitable to attend the experiment and to provide consistent results.

3.3. Experimental setup

Before starting the experiment, participants were asked to read an informative document and to sign a consent form. All participants were informed about the presence of images that could have bothered their sensibility and that they were free to abandon the experiment at any moment. Then, a detailed presentation of the experimental procedure was provided by the experimenter, followed by a training phase to let participants familiarize themselves with the experimental task. No specific information regarding the image content was provided, and different images were used in the test phase in order to preserve the integrity of the experiment. The training phase and the following

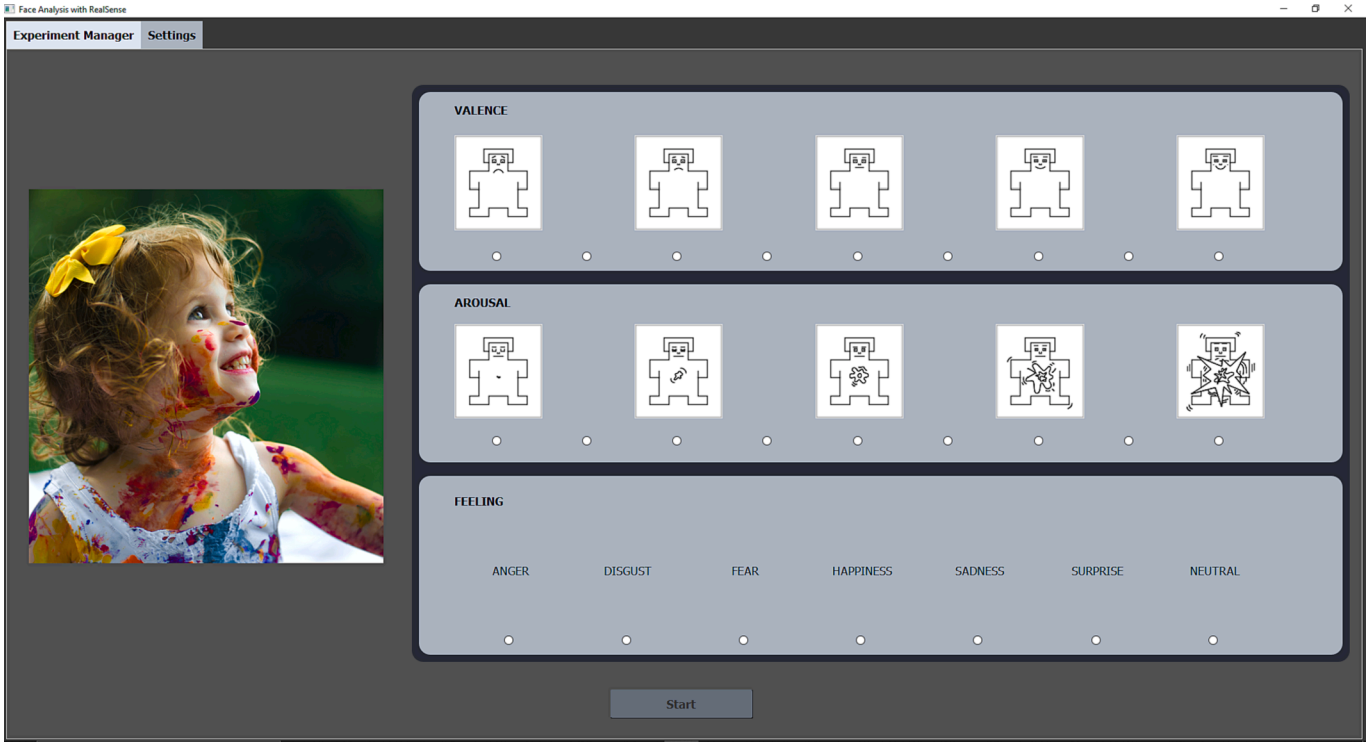

Fig. 3. Layout of the questionnaire proposed to the participants in order to assess the aroused emotion. On the left, the image from IAPS or GAPED is reprised in smaller size (the image here reported is a sample image, not from any of the two database, for reasons related to the requested non-diffusion). On the right, SAM icons support the users in understanding their level of valence (top line) and arousal (middle). Emotion labels are on the bottom line.

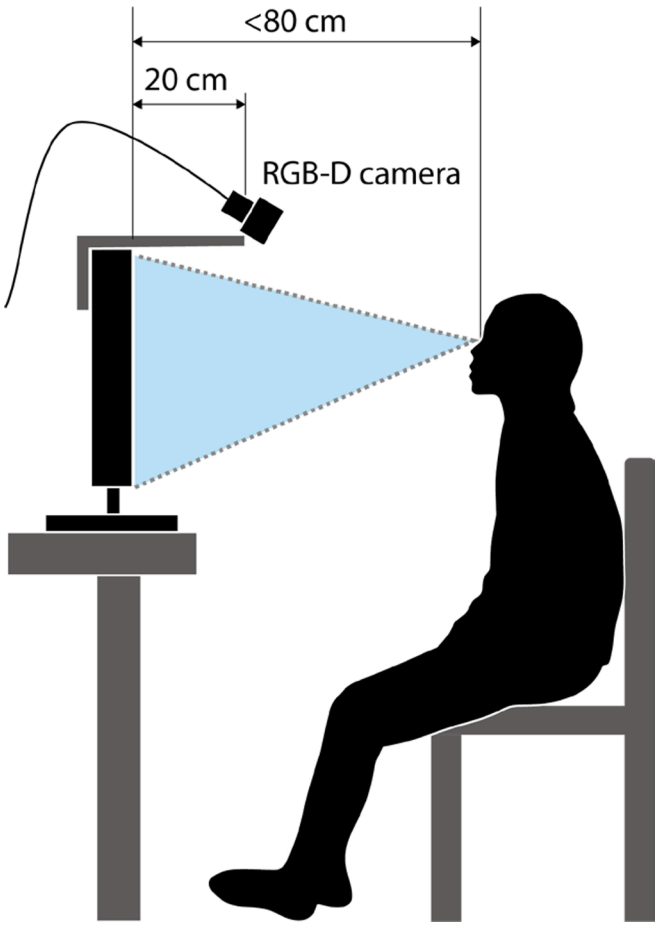

Fig. 4. Experimental setup.

experimental phase shared the same structure which, according to IAPS guidelines [18], was constituted by a 6 s image presentation, followed by a 15 s rating phase. Participants' responses were collected using a digital version of the Self-Assessment Manikin (SAM) questionnaire [43] consisting of two pictorial 9-point scales related to valence and arousal dimensions. SAM answers have been compared with valence and arousal values provided by the IAPS and GAPED databases; Fig. 2 displays the graphical result of this comparison, showing the barycenter of the expected and the obtained self-reported emotions during the two experiments. For each category, excluding the neutral state, the expected arousal was higher than the obtained values, while valence ratings were closer to those of the affective databases. Overall, the most selfreported emotion matched the expected one.

Furthermore, participants were asked to select the emotional label that, in their opinion, best represented the emotional response elicited by the observed image, choosing between 'anger', 'disgust', 'fear', 'happiness', 'sadness', 'surprise' and 'neutral' (Fig. 3). The label surprise has been included in the questionnaire not to limit participants' emotional choices, even if images arousing surprise have not been inserted in the final stimuli set. The testing phase was composed of 48 images randomized for every participant, and lasted about 20 minutes.

3.4. Acquisition technology

All the experiments were performed in uncontrolled lighting conditions, using a 27-inches monitor, and the participants seated within the maximum distance of 80 cm (Fig. 4).

A coded-light RGB-D camera, the Intel RealSense SR300, was used to acquire subjects' faces [44]. The employment of an RGB-D camera guarantees that both color and depth information (3D) are available. These cameras are low cost and can be easily integrated within personal devices such as smartphones and tablets [3]; hence, Intel RealSense SR300 has been chosen in order to acquire images in a naturalistic way both related to the spontaneity of facial expressions and from the technological point of view. With the expression 'coded-light', a subgroup of structured-light depth acquisition methodologies using spatial or temporal codes is identified. The Intel RealSense SR300 emitted light wavelength is in the infrared (IR) range, around 860 nm, which is safe and invisible to the human eye. A micro electro-mechanical system (MEMS) mirror is used to generate a set of predefined increasing spatial frequency coded IR vertical bar patterns to be projected on the scene. The IR camera resolution is 640 × 480 pixels and each pixel is 3.6 μm. Furthermore, the IR camera is equipped with an IR band filter. For the sake of completeness, it is worthy to mention that also the RGB camera is equipped with an IR band filter in order not to be affected by the projected patterns. The depth reconstruction is performed by an Application Specific Integrated Circuit and consists of two main steps: the codeword extraction and the depth generation phase. During the first step, the codeword referring to each pixel is extracted after that the scene has been illuminated with different IR patterns; during the second step, the depth map is generated considering all the patterns projected on the scene, through a pipeline which consists of codeword decoding and error correction, triangulation using codeworks and camera calibration data, and post processing to minimize the noise introduced by the temporal multiplexing, due to the motion sensitivity of this technique [45].

4. Database content

The CalD3r database consists of 4678 images showing different spontaneous facial expressions: anger, disgust, fear, happiness, sadness, surprise, and the neutral expression (Fig. 5).

In order to induce these emotions, all the participants underwent the same experiment. Nonetheless, considering the differences among the individuals in terms of emotional perception and manifestation, a non-homogeneous subdivision of facial expressions has been obtained, even if the stimuli, i.e., the images, were uniformly selected to elicit specific emotions. Therefore, the recorded facial expressions could differ from the expected ones, resulting in a final dataset of not equally distributed, but spontaneous, facial expressions. This approach aimed to prioritize the spontaneity of facial expressions makes the reliability of participants' faces assessment of paramount importance. Along the same lines of previous works on the evaluation of facial expressions related to specific stimuli [20,21], the assessment process was twofold: at first instance images have been categorized according to the arousal, valence and emotional label identified by the participants during the self-assessment phase; at second instance each image was coded by a focus group including a psychologist. The most frequent expression was the neutral ($N = 1949$), followed by happiness ($N = 766$), disgust ($N = 621$), sadness ($N = 541$), anger ($N = 492$), fear ($N = 209$) and surprise ($N = 100$) (Table 4). The paucity of facial expressions related to *surprise* is coherent with the absence of images that should have elicited this emotion, as discussed in Section 3.1.

Most of the faces are framed in frontal position, while some images display slightly rotated faces. The latest have not been discarded, since face movement can be considered a component of facial expression and, consequently, a sign of emotional activation [46]. Light occlusions can be present on some images, such as those caused by hair, beard, and glasses. Although a recommendation to avoid gestures that could cover the camera field of view was given, it was not explicitly forbidden to the participants. This choice was made to give the priority to experimental ecological validity, namely to allow participants to act spontaneously. Some frames have been discarded due to major occlusions or positioning problems with respect to the camera, such as partial face acquisition caused by the excessive proximity, for a total of 4678 pairs of 2D images and 3D depth maps of 104 subjects.

Original videos have been recorded with a 640 × 480 resolution both

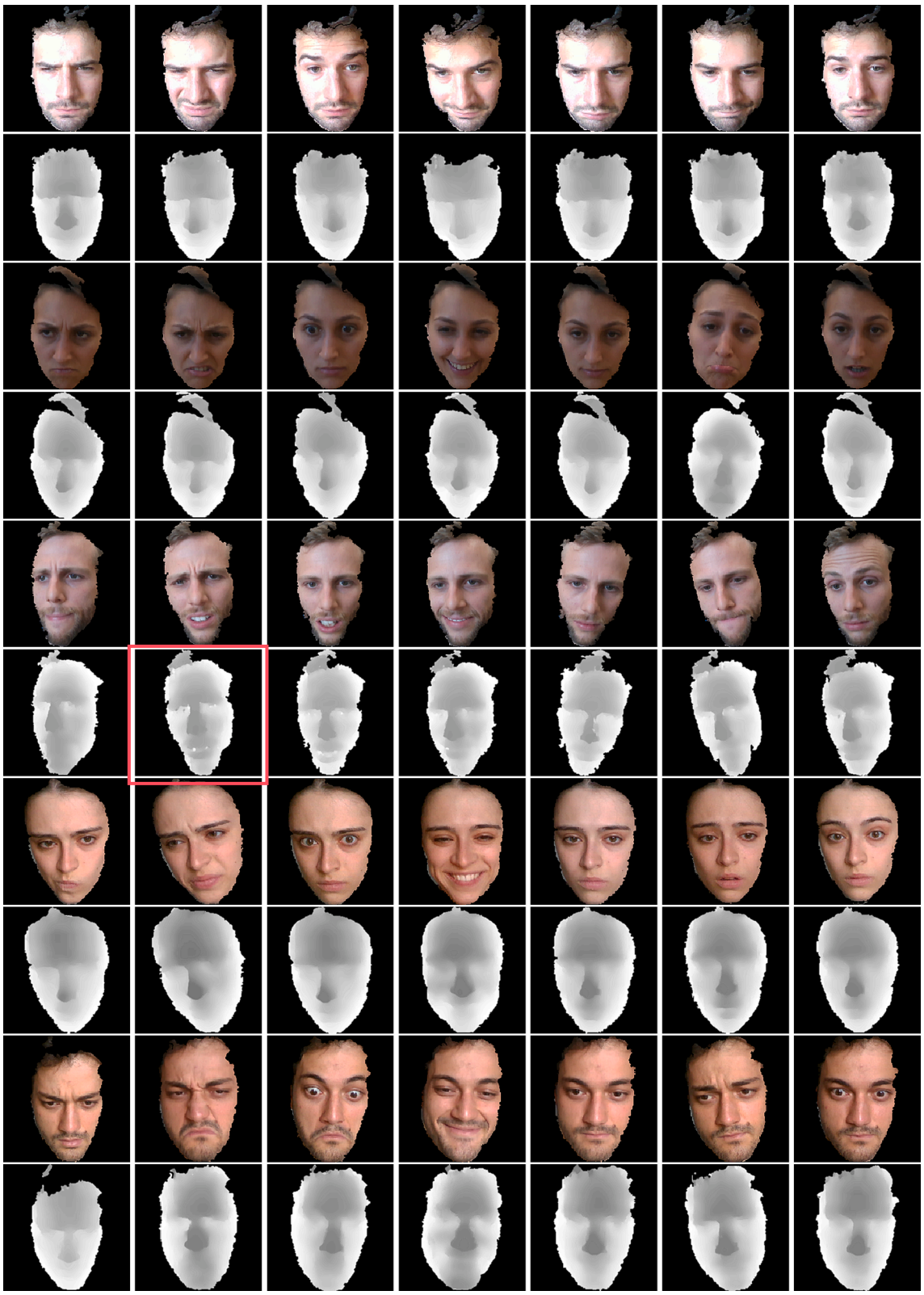

Fig. 5. From the left to the right: anger, disgust, fear, happiness, neutrality, sadness, and surprise of different subjects selected from the CalD3r. For each subject, images on the first line contain the RGB information, while on the second line there are the depth maps, i.e., the greyscale images providing the depth information.

for RGB and Depth streams. Subsequently, images provided by the streams have been aligned and a face detection algorithm has been applied to extract 224 × 224 images, a standard resolution for several deep learning approaches. In the end, the PNG images have a 224 × 224 resolution, while the RAW depth data are 640 × 480. Information related to video and image acquisition has been reported in Table 3.

The final database is subdivided by subject, identified with an alphanumeric string with the following information: gender, identified by a letter ('F' stands for 'female', while 'M' stands for 'male'); a numeric code to identify every subject; the eliciting IAPS or GAPED image code; a

Table 3
RGB and Depth videos and images details.

Description	Value
Original RGB video resolution	640 × 480
Original RGB video frame rate	30 FPS
Original Depth video resolution	640 × 480
Original Depth video frame rate	30 FPS
RGB image resolution	224 × 224
Depth image resolution	224 × 224

code to identify whether the image contains color or depth information. The alphanumeric string template has been reported below:

<Gender initial letter>_<subject number>_<affective image code>_<facial expression>_<source stream >. < file extension>

Each subject's folder contains seven subfolders with the cataloged facial expressions: anger, disgust, fear, happiness, neutral expression, sadness, and surprise. In each subfolder, for every facial expression, three files are provided: the PNG color image, the PNG depth image, and the RAW data about the depth. A script to obtain a 3D textured model by providing the RGB and the corresponding Depth image as inputs has been included.

Also, for each subject, self-assessment results (aroused emotion, valence and arousal values) are provided. To our knowledge, it is the first time that valence and arousal values felt and assessed by the subjects themselves have been integrated into a facial expression database.

The MenD3s database consists of 4038 images showing the same facial expressions of basic emotions (Fig. 6). The distribution of emotions is shown in Table 4. Similarly to the CalD3r database, the most frequent expression was neutral ($N = 1474$), while for the other expressions a different distribution was reported: sadness ($N = 645$), disgust ($N = 479$), happiness ($N = 430$), anger ($N = 390$), fear ($N = 345$), and surprise ($N = 275$).

An identical procedure was adopted for assessment and cataloging.

5. Validation

With the rapid development of deep learning, neural networks have achieved remarkable success in various visual recognition applications and currently dominate macro-FER [47] and micro-FER [48,49] automatic tasks. In this study, a multi-modal deep learning algorithm, exploiting both 2D and 3D data, has been adopted for spontaneous emotion classification to validate the CalD3r and MenD3s databases. For 2D data handling, ResNet [50], Vision Transformer (ViT) [51], and MobileNet [52] have been implemented using pre-trained models and then fine-tuned, while a custom architecture has been used to process the 3D point cloud data in its tensor form. The network performs FER over point clouds, purely 3D data obtained from depth maps, passing a $112 \times 112 \times 56 \times 3$ input through four 3D convolution layers, using the non-linear activation function Rectified Linear Unit (ReLU), followed by MaxPooling3D and BatchNormalization layers. Due to the low number of images labelled as 'surprise', since no stimuli were intended to elicit this emotion during the experiment, this class has been excluded from the neural network classification, trained to learn five facial expressions of emotion (i.e., anger, disgust, fear, happiness, and sadness) plus the neutral one. Six different configurations for the 2D + 3D FER were explored first on the MenD3s database to compare the multimodal architectures and select the best setup, and the validation accuracy is depicted in Fig. 7.

Among the available models implemented in Keras, ResNet50 and ResNet-101, introduced in 2006 [50] to solve the problem of dealing with deep architectures, MobileNetV3-Small and MobileNetV3-Large, presented by Google researchers in [52] for mobile and embedded vision applications with a priority on latency and network size, and ViT-B16 and ViT-L16, models for image classification that employ a transformer originally designed for text-based tasks over patches of the image [51], were selected. All networks have been trained for 80 epochs using Early Stopping setting the patience equal to 20 epochs, and ReduceLROnPlateau callback has been used in order to reduce the learning rate when the validation accuracy stopped improving. The initial learning rate was set to 0.0001 and reduced by a factor of 0.2, with the patience of 10 epochs, until a minimum fixed value of 1e-6. To decrease the training cost and since the multi-class nature of the task, the optimization algorithm Adam [53] and the categorical crossentropy, used and recommended when there are two or more classes [54], have been used respectively as optimizer and loss function hyperparameters to compile the model. Equation (1) shows the weighted categorical cross-entropy loss [55]:

$$CE = -\frac{1}{M}\sum_{k=1}^{K}\sum_{m=1}^{M} w_k \times y_m^k \times \log\left(h_\theta(x_m, k)\right) \tag{1}$$

where M is the number of training examples, K is the number of classes, w_k is the weight for class k, y_m^k is the target label for training example m for class k, x_m is the input for training example m, and h_θ represents the model with neural network weights θ.

The batch size varied according to the model size and the available computational power, choosing between 8 and 16. Keras, an open-source neural-network library written in Python, running on top of TensorFlow on Windows 10 Pro with NVIDIA Quadro RTX 6000, has been used. Most models act in similar ways, with no one standing above the others, and the best setup is MobileNetV3Small, the smallest model among all those tested. For this reason, it was selected for training and validation on the MenD3s + CalD3r (Fig. 8). The model that integrates MobileNetV3-Small for the 2D branch, trained on both datasets, achieved the highest validation accuracy (0.583) on six classes.

Regarding the testing, fear was the class with the lowest recognition rate. Fear was the second least aroused emotion during the experiment and has similarities with the expressions of anger and disgust, as they share some Action Units (AUs) [56], particularly in the upper part of the face. Sadness was mislabeled in the 31 % of cases as neutrality, confirming that a spontaneous sadness is arduous to elicit on a face [37,57]. Fig. 9 shows Grad-CAM maps [58] with four correctly classified (top two rows) and four misclassified (bottom two rows) RGB images to visually identify the parts of the face that most impacted the classification score. Considering the remaining expressions, the multi-modal architecture achieved a recognition rate of 75.4 % using both 2D and 3D data, demonstrating the advantage of employing three-dimensional data.

The results, comparable with others obtained in previous literature when spontaneous facial expression data are adopted, corroborate the finding that spontaneous expressions are more challenging to recognize and classify than posed ones due to their subtlety, complexity, and mixture [26,59,60]. Previous studies tried to recognize spontaneous expressions, most using BP4D and BP4D+, two spontaneous emotion databases. Zhang et al. [26] obtained an accuracy of 76.1 % working on the six prototypical facial expressions of 41 subjects of the BP4D, while the recognition rate for classifying six posed prototypic expressions of the BU-4DFE was 83 %. However, when the training was performed on selected data (happiness, disgust, and neutral) from BU-4DFE and then

Table 4
Distribution of the images for each emotion in the CALD3R (1st raw) and MEND3S (2nd raw) databases.

Database	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral	Total
CalD3r	492	621	209	766	541	100	1949	4678
MenD3s	390	479	345	430	645	275	1474	4038

Fig. 6. From the left to the right: anger, disgust, fear, happiness, neutrality, sadness, and surprise of different subjects selected from the MenD3s. For each subject, images on the first line contain the RGB information, while on the second line there are the depth maps, i.e., the greyscale images providing the depth information.

tested directly on the spontaneous data, the overall average accuracy was 71 %. From their experiment, authors find that happy-onset is often mistaken for disgust-onset, probably due to the fact that spontaneous smiles show activity around the eyes similar to some of the disgust expressions, while posed smiles do not always show this. Yang et al. [59] obtained an accuracy of 81.39 % working on only four expressions (neutral, happiness, pain, and surprise) of a subset of subjects selected from the BP4D + database. Focusing on the architecture proposed in this study, experiencing both widespread, such as ResNet, and recent, such as Vision Transformer, deep learning models, the recognition rate for

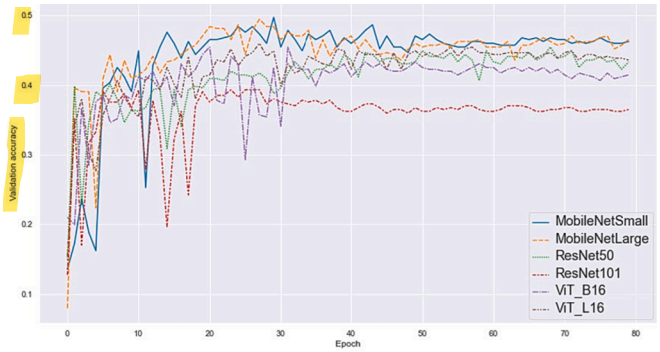

Fig. 7. Validation accuracy plot over the training epochs for 2D + 3D models, trained on the MenD3s database.

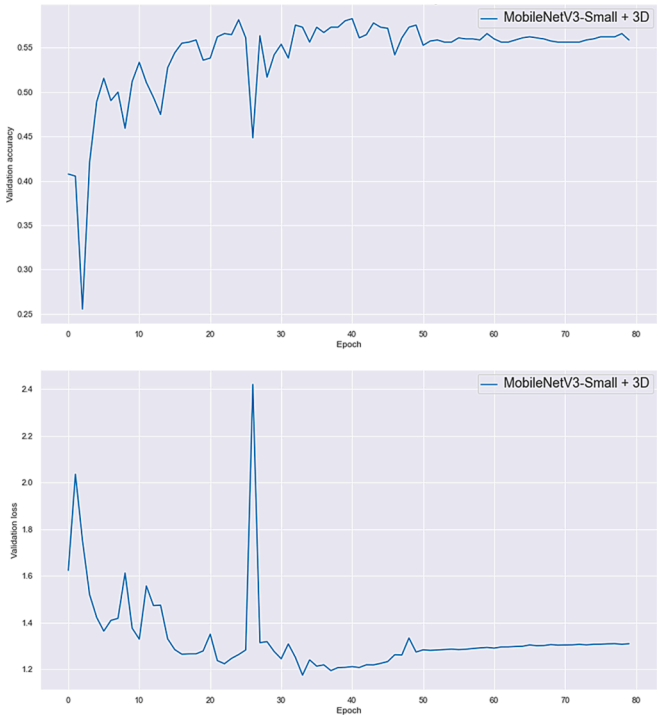

Fig. 8. Validation accuracy (up) and validation loss (down) over the training epochs for MobileNetV3-Small + 3D model, trained on the CalD3r and MenD3s databases.

classifying the six basic emotions of the posed (non-spontaneous) facial expression database BU-3DFE reached 87.5 % by using the multimodal configuration ViT + 3D.

FUSION?

6. Limitations and future work

There are some limitations in the current version of the CalD3r and MenD3s databases. Regarding the database content, the first limitation regards missing annotated reference points (landmarks) on the faces. The reason is that current FER methodologies based on neural networks do not adopt reference points, and several methods are available in the literature to obtain accurate automatic localization [61,62]. Facial inclinations rather than head poses or occlusions have been included. This is due to the experimental setup (and purpose), which was based on looking at images on a monitor. Thus, obtaining different head poses was not feasible and out of the scope of this study. For the same experimental and purpose reasons, similarly to all available databases with the exception of Bosphorus, separate facial Action Units [63] have

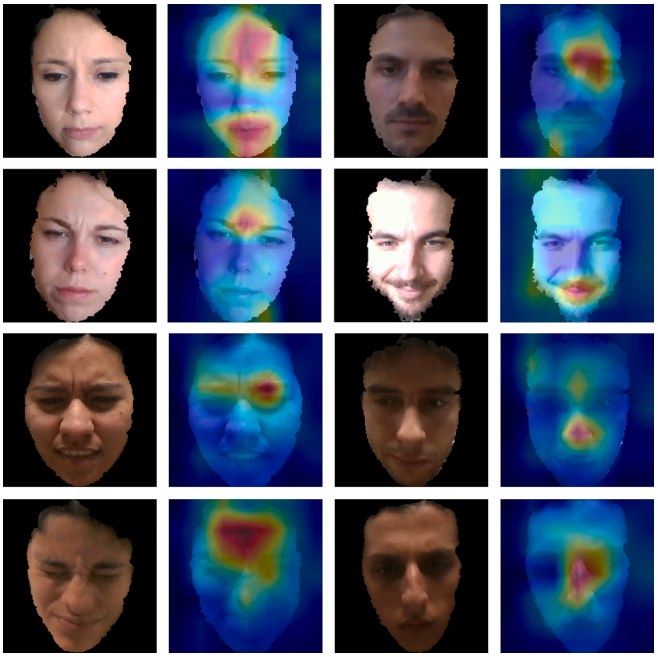

Fig. 9. Grad-CAM technique to identify the parts that most impact the classification score. The top two rows show some correctly classified expressions: from left to right, sadness, neutral, anger, and happiness. The bottom two rows show some misclassified expressions: two expressions of sadness classified as disgust and neutrality, respectively, and two expressions of fear classified as disgust and neutrality, respectively.

not been obtained nor catalogued in the obtained facial images and depths.

The last concern in terms of database content regards the geographical origin of the participants, which are just two; thus, it is only partially multi-racial and with a predominance of White faces. Facial recognition algorithms have been found particularly accurate for White males, and the least accurate for Black females [64] In particular, FER algorithms have shown to contain racial biases, for instance detecting negative emotions (e.g., anger) on Black people's faces [65]. This is also due to a consistent under-representation of darker-skinned people and females in training data sets. The presence of Brazilian acquisitions only partially bridges this gap. In this sense, the procedure to acquire other subjects makes this work easily 'scalable', as testified by the acquisition of the second dataset (MenD3s). The possibility to further expand these datasets with others including in turn other geographical provenance goes in the direction of providing data more and more generalizable, and consequently improving datasets efficacy during the training of automatic FER algorithms. As a consequence, a future development consists in the inclusion of additional data to ensure diversity in representation and prevent biases.

Regarding the stimuli, only visual static images were adopted in this experimentation, and the 'surprise' stimulus was excluded due to reasons that have been previously exposed (Section 3.1), even if faces displaying surprise have been obtained and included anyway.

The database can be improved over time by involving also affective audio stimuli, such as the International Affective Digitized Sounds (IADS) [66] accompanying the images, or dynamic contents. In this sense, Virtual Reality looks like a promising tool for evoking emotions [20,67,21,68] and could be adopted for conceiving an expanded version of this database containing also dynamic facial data. Physiological data such as electroencephalography (EEG) ones may also be included.

A major enhancement of this database concerns its possible and easy expandability with the collaboration of other worldwide research groups that could contribute to the database by enrolling multiracial participants. Our experimental setup can be easily replicated at low cost

and the acquisition process can be reproduced elsewhere to obtain a wider and more diverse attendance. Indeed, the Brazilian acquisition session is the first step towards this direction. A protocol is being made by our research group describing details of our setup to replicate it, so that this overall database will become the first open source spontaneous 3D facial expression database.

In the era of metaverse, the need to interact with other people in a shared virtual environment fosters the design of avatars capable of reproducing human facial expressions, in order to enhance the user experience. Reliable facial expressions can be achieved using cutting-edge deep learning methodologies, so that the metaverse itself could become a framework where specific context-related emotions could be elicited in order to further expand the available datasets for the training of AI algorithms. Generally speaking, human–computer interaction is currently transitioning to human-AI interaction [69]. Emotions play a pivotal role, as human beings are willing to assign human characteristics to computers/machines/robots, resulting in "humanizing AI". Therefore, emotions have become an essential actor in the ongoing conversation on psychology, ethics and morality of AI, pointing towards an 'artificial emotional intelligence'.

7. Conclusion

In this work, we introduce the CalD3r and MenD3s databases, including 2D and 3D facial data of seven emotions of 196 gender-balanced participants of Southern Europe and Brazilian origins, respectively. The emotions are elicited with a selection of images from IAPS and GAPED image databases, making of it the first three-dimensional facial expression database in which emotions are evoked with validated affective stimuli. The database is specifically thought for testing novel FER algorithms intended for 'ecologically valid' experimentation, i.e., naturalistic scenarios, but can be adopted for studies in face analysis, subject recognition and psychology.

The database is conceived to be expandable with other types of data such as physiological and dynamic, and its experimental setup can be easily replicated at low cost. Thus, it could be integrated with data coming from the scientific community, with an 'open source' philosophy. This approach will hopefully support the inclusion of data of additional participants from other racial groups, contributing to the development of more equitable face recognition technologies.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgments

The authors thank Francesca Giada Antonaci, Emanuele Cadeddu, Michela Putzu, Simona Cannizzaro, Gabriele Marsocci, Alba La Rosa, Fabio Tatti and Anna Di Lorenzo for contributing to the design and creation of the databases, and the subjects who voluntarily let their faces to be acquired.

This study was carried out within the Ministerial Decree no. 1062/2021 and received funding from the FSE REACT-EU - PON Ricerca e Innovazione 2014-2020. This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

References

[1] M. Egger, M. Ley, S. Hanke, Emotion recognition from physiological signal analysis: A review, Electron. Notes Theor. Comput. Sci. 343 (2019) 35–55.
[2] L.F. Barrett, R. Adolphs, S. Marsella, A.M. Martinez, S.D. Pollak, Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements, Psychol. Sci. Public Interest 20 (2019) 1–68.
[3] L. Ulrich, E. Vezzetti, S. Moos, F. Marcolin, Analysis of rgb-d camera technologies for supporting different facial usage scenarios, Multimed. Tools Appl. 79 (2020) 29375–29398.
[4] L. Ulrich, J.-L. Dugelay, E. Vezzetti, S. Moos, F. Marcolin, Perspective morphometric criteria for facial beauty and proportion assessment, Appl. Sci. 10 (2020) 8.
[5] P. J. Phillips, P. Grother, R. Micheals, D. M. Blackburn, E. Tabassi, M. Bone, Face recognition vendor test 2002, in: 2003 IEEE International SOI Conference. Proceedings (Cat. No. 03CH37443), IEEE, 2003, p. 44.
[6] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), volume 1, IEEE, 2005, pp. 947–954.
[7] A. Colombo, C. Cusano, R. Schettini, Umb-db: A database of partially occluded 3d faces, in: 2011 IEEE international conference on computer vision workshops (ICCV workshops), IEEE, 2011, pp. 2113–2119.
[8] R. Min, N. Kose, J.-L. Dugelay, Kinectfacedb: A kinect database for face recognition, IEEE Trans. Syst., Man, Cybernet.: Syste. 44 (2014) 1534–1548.
[9] E.C. Olivetti, J. Ferretti, G. Cirrincione, F. Nonis, S. Tornincasa, F. Marcolin, Deep cnn for 3d face recognition, in: In: International Conference of the Italian Association of Design Methods and Tools for Industrial Engineering, 2019, pp. 665–674.
[10] V. Bruce, A. Young, Understanding face recognition, Br. J. Psychol. 77 (1986) 305–327.
[11] L. Yin, X. Wei, Y. Sun, J. Wang, M.J. Rosato A,, in: 3d Facial Expression Database for Facial Behavior Research, in, 2006, pp. 211–216.
[12] M.A. Schmuckler, What is ecological validity? a dimensional analysis, Infancy 2 (2001) 419–436.
[13] Z. Zhang, J.M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, et al., Multimodal spontaneous emotion corpus for human behavior analysis, in: In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3438–3446.
[14] J.M. Chen, J.B. Norman, Y. Nam, Broadening the stimulus set: introducing the american multiracial faces database, Behav. Res. Methods 53 (2021) 371–389.
[15] L. Yin, X. Chen, Y. Sun, T. Worm, M. Reale, A high-resolution 3d dynamic facial expression database, in: 2008 8th IEEE International Conference on Automatic Face Gesture Recognition, 2008, pp. 1–6. doi:10.1109/AFGR.2008.4813324.
[16] A. Savran, N. Alyuz, H. Dibeklio¨glu, O. C¸eliktutan, B. G˘okberk¸, B. Sankur, L. Akarun, Bosphorus database for 3d face analysis, in: European workshop on biometrics and identity management, Springer, 2008, pp. 47–56.
[17] S. Wang, Z. Liu, Z. Wang, G. Wu, P. Shen, S. He, X. Wang, Analyses of a multimodal spontaneous facial expression database, IEEE Trans. Affect. Comput. 4 (2012) 34–46.
[18] P.J. Lang, M. M. Bradley, B. N. Cuthbert, et al., International affective picture system (iaps): Technical manual and affective ratings, NIMH Center for the Study of Emotion and Attention 1 (1997) 3.
[19] E.S. Dan-Glauser, K.R. Scherer, The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance, Behav. Res. Methods 43 (2011) 468–477.
[20] N. Dozio, F. Marcolin, G.W. Scurati, F. Nonis, L. Ulrich, E. Vezzetti, F. Ferrise, Development of an affective database made of interactive virtual environments, Sci. Rep. 11 (2021) 1–10.
[21] N. Dozio, F. Marcolin, G.W. Scurati, L. Ulrich, F. Nonis, E. Vezzetti, G. Marsocci, A. La Rosa, F. Ferrise, A design methodology for affective virtual reality, Int. J. Hum Comput Stud. 102791 (2022).
[22] C. Redies, M. Grebenkina, M. Mohseni, A. Kaduhm, C. Dobel, Global image properties predict ratings of affective pictures, Front. Psychol. 11 (2020) 953.
[23] C. Darwin, The expression of the emotions in man and animals, University of Chicago press, 2015.
[24] P. Ekman, Are there basic emotions? (1992).
[25] T.C. Faltemier, K.W. Bowyer, P.J. Flynn, Using a multi-instance enrollment representation to improve 3d face recognition, in: In: 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2007, pp. 1–6.
[26] X. Zhang, L. Yin, J.F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, J.M. Girard, Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database, Image Vis. Comput. 32 (2014) 692–706.
[27] S. Zafeiriou, M. Hansen, G. Atkinson, V. Argyriou, M. Petrou, M. Smith, L. Smith, The photoface database, in: CVPR 2011 WORKSHOPS, IEEE, 2011, pp. 132–139.
[28] A. Moreno Gavabdb: a, in: 3d Face Database, in: Proc, 2nd, 2004, 2004,, pp. 75–80.
[29] C. Zhong, Z. Sun, T. Tan, Robust 3d face recognition using learned visual codebook, in: In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–6.
[30] T. Heseltine, N. Pears, J. Austin, Three-dimensional face recognition using combinations of surface feature map subspace components, Image Vis. Comput. 26 (2008) 382–396.
[31] S. Gupta, K. R. Castleman, M. K. Markey, A. C. Bovik, Texas 3d face recognition database, in: 2010 IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI), IEEE, 2010, pp. 97–100.

[32] V. Vijayan, K. W. Bowyer, P. J. Flynn, D. Huang, L. Chen, M. Hansen, O. Ocegueda, S. K. Shah, I. A. Kakadiaris, Twins 3d face recognition challenge, in: 2011 international joint conference on biometrics (IJCB), IEEE, 2011, pp. 1–7.
[33] C. Cao, Y. Weng, S. Zhou, Y. Tong, K. Zhou, Facewarehouse: A 3d facial expression database for visual computing, IEEE Trans. Vis. Comput. Graph. 20 (2013) 413–425.
[34] W. Sankowski, P. S. Nowak, P. Krotewicz, Multimodal biometric database dmcsv1 of 3d face and hand scans, in: 2015 22nd International Conference Mixed Design of Integrated Circuits & Systems (MIXDES), IEEE, 2015, pp. 93–97.
[35] P. Lang, M.M. Bradley, The international affective picture system (iaps) in the study of emotion and attention, Handbook of Emotion Elicitation and Assessment 29 (2007) 70–73.
[36] M.M. Bradley, M. Codispoti, D. Sabatinelli, P.J. Lang, Emotion and motivation ii: sex differences in picture processing, Emotion 1 (2001) 300.
[37] P. Ekman, Emotions revealed, BMJ 328 (2004).
[38] F. Nonis, L. Ulrich, N. Dozio, F. G. Antonaci, E. Vezzetti, F. Ferrise, F. Marcolin, Building an ecologically valid facial expression database– behind the scenes, in: International Conference on Human-Computer Interaction, Springer, 2021, pp. 599–616.
[39] Y. Moriguchi, J. Decety, T. Ohnishi, M. Maeda, T. Mori, K. Nemoto, H. Matsuda, G. Komaki, Empathy and judging other's pain: an fmri study of alexithymia, Cereb. Cortex 17 (2007) 2223–2234.
[40] C.-L. Mul, S.D. Stagg, B. Herbelin, J.E. Aspell, The feeling of me feeling for you: Interoception, alexithymia and empathy in autism, Journal of Autism and Developmental Disorders 48 (2018) 2953–2967.
[41] R.M. Bagby, J.D. Parker, G.J. Taylor, The twenty-item toronto alexithymia scale—i. item selection and cross-validation of the factor structure, J. Psychosom. Res. 38 (1994) 23–32.
[42] A. Meneghini, R. Sartori, L. Cunico, Adattamento italiano della balanced emotional empathy scale (bees) di albert mehrabian [the italian adaptation of the balanced emotional empathy scale (bees) by albert mehrabian], Florence, Giunti Organizzazioni Speciali, Italy, 2012.
[43] M.M. Bradley, P.J. Lang, Measuring emotion: The self-assessment manikin and the semantic differential, J. Behav. Ther. Exp. Psychiatry 25 (1994) 49–59.
[44] G. Maculotti, L. Ulrich, E.C. Olivetti, G. Genta, F. Marcolin, E. Vezzetti, M. Galetto, A methodology for task-specific metrological characterization of low-cost 3d camera for face analysis, Measurement 200 (2022) 111643.
[45] A. Zabatani, V. Surazhsky, E. Sperling, S.B. Moshe, O. Menashe, D.H. Silver, Z. Karni, A.M. Bronstein, M.M. Bronstein, R. Kimmel, Intel® realsense™ sr300 coded light depth camera, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2019) 2333–2345.
[46] J.N. Bassili, Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face, J. Pers. Soc. Psychol. 37 (1979) 2049.
[47] F. Nonis, N. Dagnes, F. Marcolin, E. Vezzetti, 3d approaches and challenges in facial expression recognition algorithms—a literature review, Appl. Sci. 9 (2019) 3904.
[48] X. Ben, Y. Ren, J. Zhang, S.-J. Wang, K. Kpalma, W. Meng, Y.-J. Liu, Video-based facial micro-expression analysis: A survey of datasets, features and algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2021) 5826–5846.
[49] X. Ben, C. Gong, T. Huang, C. Li, R. Yan, Y. Li, Tackling microexpression data shortage via dataset alignment and active learning, IEEE Trans. Multimedia (2022).
[50] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[51] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint arXiv: 2010.11929 (2020).
[52] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861 (2017).
[53] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
[54] P.-T. De Boer, D.P. Kroese, S. Mannor, R.Y. Rubinstein, A tutorial on the cross-entropy method, Annals of Operations Research 134 (2005) 19–67.
[55] Y. Ho, S. Wookey, The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling, IEEE Access 8 (2019) 4806–4813.
[56] P. Ekman, W.V. Friesen, Facial action coding system, Environ. Psychol. Nonverbal Behavior (1978).
[57] S. Namba, S. Makihara, R.S. Kabir, M. Miyatani, T. Nakao, Spontaneous facial expressions are different from posed facial expressions: Morphological properties and dynamic sequences, Curr. Psychol. 36 (2017) 593–605.
[58] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.
[59] H. Yang, U. Ciftci, L. Yin, Facial expression recognition by de-expression residue learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2168–2177.
[60] S. Wan, J. Aggarwal, Spontaneous facial expression recognition: A robust metric learning approach, Pattern Recogn. 47 (2014) 1859–1868.
[61] J. Deng, A. Roussos, G. Chrysos, E. Ververas, I. Kotsia, J. Shen, S. Zafeiriou, The menpo benchmark for multi-pose 2d and 3d facial landmark localisation and tracking, International Journal of Computer Vision 127 (2019) 599–624.
[62] E. Vezzetti, F. Marcolin, S. Tornincasa, L. Ulrich, N. Dagnes, 3d geometry-based automatic landmark localization in presence of facial occlusions, Multimed. Tools Appl. 77 (2018) 14177–14205.
[63] E.L. Rosenberg, P. Ekman, What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS), Oxford University Press, 2020.
[64] J. Buolamwini, T. Gebru, Gender shades: Intersectional accuracy disparities in commercial gender classification, in: In: Conference on Fairness, Accountability and Transparency, 2018, pp. 77–91.
[65] L. Rhue, Racial influence on automated perceptions of emotions, Available at SSRN 3281765 (2018).
[66] M. M. Bradley, P. J. Lang, International affective digitized sounds (iads): Stimuli, instruction manual and affective ratings (tech. rep. no. b-2), Gainesville, FL: The Center for Research in Psychophysiology, University of Florida (1999).
[67] F. Marcolin, G.W. Scurati, L. Ulrich, F. Nonis, E. Vezzetti, N. Dozio, F. Ferrise, Affective virtual reality: How to design artificial experiences impacting human emotions, IEEE Comput. Graph. Appl. 41 (2021) 171–178.
[68] I. A. Castiblanco Jimenez, F. Marcolin, L. Ulrich, S. Moos, E. Vezzetti, S. Tornincasa, Interpreting emotions with eeg: an experimental study with chromatic variation in vr, in: International Joint Conference on Mechanics, Design Engineering & Advanced Manufacturing, Springer, 2022, pp. 318–329.
[69] J.A. Crowder, J.N. Carbone, S.A. Friess, J.A. Crowder, J.N. Carbone, S.A. Friess, The psychology of artificial intelligence, Artificial Cognition Archit. (2014) 17–26.