Chapter 3

Review on Emotion Recognition Databases

Rain Eric Haamer, Eka Rusadze, Iiris Lüsi,
Tauseef Ahmed, Sergio Escalera and
Gholamreza Anbarjafari

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72748

Abstract

Over the past few decades human-computer interaction has become more important in our daily lives and research has developed in many directions: memory research, depression detection, and behavioural deficiency detection, lie detection, (hidden) emotion recognition etc. Because of that, the number of generic emotion and face databases or those tailored to specific needs have grown immensely large. Thus, a comprehensive yet compact guide is needed to help researchers find the most suitable database and understand what types of databases already exist. In this paper, different elicitation methods are discussed and the databases are primarily organized into neat and informative tables based on the format.

Keywords: emotion, computer vision, databases

1. Introduction

With facial recognition and human-computer interaction becoming more prominent with each passing year, the amount of databases associated with both face detection and facial expressions has grown immensely [1, 2]. A key part in creating, training and even evaluating supervised emotion recognition models is a well-labelled database of visual and/or audio information fit for the desired application. For example, emotion recognition has many different applications ranging from simple human-robot computer interaction [3–5] to automated depression detection [6].

There are several papers, blogs and books [7–10] fully dedicated to just describing some of the more prominent databases for face recognition. However, the collection of emotion databases is disparate, as they are often tailored to a specific purpose, so there is no complete and thorough overview of the ones that currently exist.

Even though there already are a lot of collected databases out there that fit many specific criteria [11, 12], it is important to recognize that there are several different aspects that affect the content of the database. The selection of the participants, the method used to collect the data and what was in fact collected all have a great impact on the performance of the final model [13]. The cultural and social background of participants as well as their mood during recordings can sway the results of the database to be specific to a particular group of people. This can even happen with larger sample pools, like the case with the Bosphorus database [14], which suffers from a lack of ethnic diversity compared to databases with a similar or even smaller size [15–17].

Since most algorithms take an aligned and cropped face as an input, the most basic form of datasets is a collection of portrait images or already cropped faces, with uniform lighting and backgrounds. Among those is the NIST mugshot database [18], which has clear gray-scale mugshots and portraits of 1573 individuals on a uniform background. However, real-life scenarios are more complicated, requiring the authors to experiment with different lighting, head pose and occlusions [19]. For example in the M2VTS database [20], which contains the faces of 37 subjects in different rotated positions and lighting angles.

Some databases have focused on gathering samples from even less controlled environments with obstructed facial data like the SCface database [21], which contains surveillance data gathered from real world scenarios. Emotion recognition is not solely based on a person's facial expression, but can also be assisted by body language [22] or vocal context. Unfortunately, not many databases include body language, preferring to completely focus on the face, but there are some multi-modal video and audio databases that incorporate vocal context [11, 23].

2. Elicitation methods

An important choice to make in gathering data for emotion recognition databases is how to bring out different emotions in the participants. This is the reason why facial emotion databases are divided into three main categories [24]:

- posed
- induced
- spontaneous

Eliciting expressions can be done in several different ways and unfortunately, they yield wildly different results.

2.1. Posed

Emotions acted out based on conjecture or with the guidance from actors or professionals are called posed expressions [25]. Most facial emotion databases, especially the early ones i.e. Banse-Scherer [26], CK [27] and Chen-Huang [28], consist purely of posed facial expressions, as it is the easiest to gather. However, they also are the least representative of real world authentic emotions as forced emotions are often over-exaggerated or missing subtle details,

http://dx.doi.org/10.5772/intechopen.72748

like in **Figure 1**. Due to this, human expression analysis models created through the use of posed databases often have very poor results with real world data [13, 30]. To overcome the problems related to authenticity, professional theatre actors have been employed, e.g. for the GEMEP [31] database.

2.2. Induced

This method of elicitation displays more genuine emotions as the participants usually interact with other individuals or are subject to audiovisual media in order to invoke real emotions. Induced emotion databases have become more common in recent years due to the limitations of posed expressions. The performance of the models in real life is greatly improved, since they are not hindered by overemphasised and fake expressions, making them more natural, as seen in **Figure 2**. There are several databases that deal with audiovisual emotion elicitation like the

Figure 1. Posed expressions over different age groups from the FACES database [29].

Figure 2. Induced facial expressions from the SD database [32].

SD [32], UT DALLAS [33] and SMIC [34], and some that deal with human to human interaction like the ISL meeting corpus [35], AAI [36] and CSC corpus [37].

Databases produced by observing human-computer interaction on the other hand are a lot less common. The best representatives are the AIBO database [23], where children are trying to give commands to a Sony AIBO robot, and SAL [11], in which adults interact with an artificial chat-bot.

Even though induced databases are much better than the posed ones, they still have some problems with truthfulness. Since the emotions are often invoked in a lab setting with the supervision of authoritative figures, the subjects might subconsciously keep their expressions in check [25, 30].

2.3. Spontaneous

Spontaneous emotion datasets are considered to be the closest to actual real-life scenarios. However, since true emotion can only be observed, when the person is not aware of being recorded [30], they are difficult to collect and label. The acquisition of data is usually in conflict with privacy or ethics, whereas the labelling has to be done manually and the true emotion has to be guessed by the analyser [25]. This arduous task is both time-consuming and erroneous [13, 38], having a sharp contrast with posed and induced datasets, where labels are either predefined or can be derived from the elicitation content.

With that being said, there still exist a few databases out there that consist of data extracted from movies [39, 40], YouTube videos [41], or even television series [42], but these databases have inherently fewer samples in them than their posed and induced counterparts. Example images from these databases are in **Figures 3–5** respectively.

Figure 3. Images of movie clips taken from the AFEW database [39, 40].

Figure 4. Spanish YouTube video clips taken from the Spanish Multimodal Opinion database [41].

Figure 5. TV show stills taken from the VAM database [43].

3. Categories of emotion

The purpose of a database is defined by the emotions represented in it. Several databases like CK [27, 44], MMI [45], eNTERFACE [46], NVIE [47] all opt to capture the six basic emotion types: anger, disgust, fear, happiness, sadness and surprise as proposed by Ekman [48–50]. In the tables, they are denoted as primary 6. Often authors tend to add contempt to these, forming seven primary emotions and often neutral is included. However, they cover a very small subcategory of all possible emotions, so there have been attempts to combine them [51, 52].

Several databases try to just categorise the general positive and negative emotions or incorporate them along with others, e.g. the SMO [41], AAI [36], and ISL meeting corpus [35] databases. Some even try to rank deception and honesty like the CSC corpus database [37].

Apart from anger and disgust within the six primary emotions, scientists have tried to capture other negative expressions, such as boredom, disinterest, pain, embarrassment and depression. Unfortunately, these categories are harder to elicit than other types of emotions.

TUM AVIC [53] and AVDLC [12] databases are amongst those that try to label levels of interest and depression while GEMEP [31] and VAM [43] attempt to divide emotions into four quadrants and three dimensions, respectively. The main reason why most databases have a very small number of categories (mainly, neutral and smile/no-smile) is that the more emotions added, the more difficult they are to label and also more data is required to properly train a model.

Relatively newer databases have begun recording more subtle emotions hidden behind other forced or dominant emotions. Among these are the MAHNOB [51] database, which focuses on emotional laughter and different types of laughter, and others that try to record emotions hidden behind a neutral or straight face like SMIC [34], RML [54], Polikovsky's [55] databases.

One of the more recent databases, the iCV-MEFED [52, 56] database, takes on a different approach by posing varying combinations of emotions simultaneously, where one emotion takes the dominant role and the other is complimentary. Sample images can be seen in **Figure 6**.

3.1. Action units

The Facial Affect Sorting Technique (FAST) was developed to measure facial movement relative to emotion. They describe the six basic emotions through facial behaviour: happiness, surprise and disgust have three intensities and anger is reported as controlled and uncontrolled [57]. Darwin [58], Duchenne [59] and Hjortsjo [60], Ekman and Friesen [61] developed the Facial Action Coding System (FACS), a comprehensive system, which catalogues all possible visually distinguishable facial movements.

FACS describes facial expressions in terms of 44 anatomically based Action Units (AU). They are meant for facial punctuators in conversation, facial deficits indicative of brain lesions, emotion detection, etc. FACS only deals with visible changes, which are often induced by a combination of muscle contractions. Because of that, they are called action units [61]. A small

Figure 6. Combinations of emotions from the iCV-MEFED [52].

a)Neutral b)AU1 c)AU2 d)AU1+2 e)AU4 f)AU1+2+4

Figure 7. Induced facial action units from the DISFA database [62].

Database	Participants	Elicitation	Format	Action units	Additional information
CMU-Pittsburgh AU-Coded Face Expression Database [27] 2000	210	Posed	Videos	44	Varying ethnic backgrounds, FACS coding
MMI Facial Expression Database [63, 64] 2002	19	Posed and audiovisual media	Videos, images	79	Continuously updated, contains different parts
Face Video Database of the MPI [65, 66] 2003	1	Posed	Six viewpoint videos	55	Created using the MPI VideoLab
D3DFACS [67] 2011	10	posed	3D videos	19–97	Supervised by FACS specialists
DISFA [62] 2013	27	audiovisual media	Videos	12	

Table 1. Action unit databases.

sample of such expressions can be seen in **Figure 7**. A selection of databases based on AUs instead of regular facial expressions is listed in **Table 1**.

In 2002, the FACS system was revised and the number of facial contraction AUs was reduced to 33 and 25 head pose AUs were added [68–70]. In addition, there is a separate FACS version intended for children [71].

4. Database types

Emotion recognition databases may come in many different forms, depending on how the data was collected. We review existing databases for different types of emotion recognition. In order to better compare similar types of databases, we decided to split them into three broad categories based on format. The first two categories separated still images from video sequences, while the last category is comprised of databases with more unique capturing methods.

4.1. Static databases

Most early facial expression databases, like the CK [27], only consist of frontal portrait images taken with simple RGB cameras. Newer databases try to design collection methods that incorporate data, which is closer to real life scenarios by using different angles and occlusion (hats, glasses, etc.). Great examples are the MMI [45] and Multi-PIE [72] databases, which were some of the first well-known ones using multiple view angles. In order to increase the accuracy of the human expression analysis models, databases like the FABO [22] have expanded the frame from a portrait to the entire upper body.

Static databases are the oldest and most common type. Therefore, it's understandable that they were created with the most diverse of goals, varying from expression perception [29] to neuropsychological research [73], and have a wide range of data gathering styles, including self-photography through a semi-reflective mirror [74] and occlusion and light angle variation [75]. Static databases usually have the largest number of participants and a bigger sample size. While it is relatively easy to find a database suited for the task at hand, categories of emotions are quite limited, as static databases only focus on six primary emotions or smile/neutral detection. In the future, it would be convenient if there were databases with more emotions, especially spontaneous or induced, because, as you can see in **Table 2**, almost all static databases to date are posed.

4.2. Video databases

The most convenient format for capturing induced and spontaneous emotions is video. This is due to a lack of clear start and end points for non-posed emotions [93]. In the case of RGB Video, the subtle emotional changes known as microexpressions have also been recorded with the hope of detecting concealed emotions as in USF-HD [94], YorkDDT [95], SMIC [34], CASME [96] and Polikovsky's [55] databases, the newest and most extensive among those being CASME.

Posed video databases in **Table 3** suggest that they tend to be quite small in the number of participants, usually around 10, and often professional actors have been used. Unlike with still images, scientists have tried to benefit from voice, speech or any other type of utterances for emotion recognition. Many databases have also tried to gather micro-expressions, as they do not show up on still images or are harder to catch. The posed video databases have mainly focused on six primary emotions and a neutral expression.

Media induced databases, as in **Table 4**, have a larger number of participants and the emotions are usually induced by audiovisual media, like Superbowl ads [107]. Because the emotions in these databases are induced via external means, this format is great for gathering fake [108] or hidden [34] emotions.

Interaction induced video databases have more unique ways of gathering data, like child-robot interaction [23] or reviewing past memories [36]. This can be seen in **Table 5**. This type of databases takes significantly longer time to create [113], but this does not seem to affect the sample size. Almost all of the spontaneous databases are in video format from other media sources, purely because of how difficult they are to collect. Spontaneous databases are also some of the rarest, compared to other elicitation methods. This is reflected in **Table 6**, which has the lowest number of databases among the different elicitation methods.

4.3. Miscellaneous databases

Apart from the formats mentioned above, 3D scanned and even thermal databases of different emotions have also been constructed. The most well-known 3D datasets are the BU-3DFE [15],

Database	Participants	Primary 6	Neutral	Contempt	Embarrassment	Pain	Smile	Positive	Negative	Other	Additional information
JACFEE [76] 1988	4	X	X								Eight images of each emotion
POFA (or PFA) [73] 1993	14	X									Cross-cultural studies and neuropsychological research
AT-T Database for Faces (formerly ORL) [77, 78] 1994	40						X			X	Dark homogeneous background, frontal face
Yale [75] 1997	15		X								Frontal face, different light angles, occlusions
FERET [79] 1998	1199		X				X				Standard for face recognition algorithms
KDEF [80] 1998	70	X	X								Psychological and medical research (perception, attention, emotion, memory and backward masking)
The AR Face Database [81] 1998	126	✓[1]	X				X		X		Frontal face, different light angles, occlusions
The Japanese Female Facial Expression Database [74] 1998	10	X	X								Subjects photographed themselves through a semi-reflective mirror
MSFDE [82] 2000	12	X	X								FACS coding, ethnical diversity
CAFE Database [83] 2001	24	X	X								FACS coding, ethnical diversity
CMU PIE [84] 2002	68		X				X		X		Illumination variation, varying poses
Indian Face Database [85] 2002	40	✓					X				Indian participants from seven view angles
NimStim Face Stimulus Set [86] 2002	70	X					X			X	Facial expressions were supervised
KFDB [87] 2003	1920						X		X		Includes ground truth for facial landmarks
PAL Face Database [88] 2004	576	✓							X		Wide age range
UT DALLAS [33] 2005	284	✓					X				Head and face detection, emotions induced using audiovisual media
TFEID [89] 2007	40	X							X		Taiwanese actors, two simultaneous angles
CAS-PEAL [90] 2008	1040	X	X				X				Chinese face detection
Multi-PIE [72] 2008	337		X				X				Multiple view angles, illumination variation
PUT [91] 2008	100		X							X	High-resolution head-pose database
Radboud Faces Database [92] 2008	67	X	X	X							Supervised by FACS specialists
FACES database [29] 2010	154	X									Expression perception, wide age range, evaluated by participants
iCV-MEFED [52] 2017	115	X	X								Psychologists picked best from 5

[1]A selection of six primary emotions has been used in databases with this symbol.

Table 2. Posed static databases.

BU-4DFE [16], Bosphorus [14] and BP4D [17]. BU-3DFE and BU-4DFE both contain posed datasets with six expressions, the latter having higher resolution. Bosphorus tries to address the issue of having a wider selection of facial expressions and BP4D is the only one among the four using induced expressions instead of posed ones. A sample of models from a 3D database can be seen in **Figure 8**.

Database	Participants	Primary 6	Neutral	Contempt	Embarrassment	Pain	Smile	Positive	Negative	Other	Additional information
University of Maryland DB [97] 1997	40	X									1–3 expressions per clip
CK [27] 2000	97	X									One of the first FE databases made public
Chen-Huang [28] 2000	100	X									Facial expressions and speech
DaFEx [98] 2004	8	X	X								Italian actors mimicked emotions while uttering different sentences
Mind Reading [99] 2004	6		X				X				Teaching tool for children with behavioural disabilities
GEMEP [31] 2006	10	✓								X	Professional actors, supervised
AONE [100] 2007	75										Asian adults
FABO [22] 2007	4	✓								X	Face and upper-body
IEMOCAP [101] 2008	10	✓	X							X	Markers on face, head, hands
RML [54] 2008	8	X									Suppressed emotions
Polikovsky's database [55] 2009	10	X	X								Low intensity micro-expressions
SAVEE [102] 2009	4	X	X								Blue markers, three images per emotion
STOIC [103] 2009	10	X	X			X					Face recognition, discerning gender, contains still images
YorkDDT [95] 2009	9	X	X								Micro-expressions
ADFES [104] 2011	22	X		X	X				X		Frontal and turned facial expressions
USF-HD [94] 2011	16	✓								X	Micro-expressions, mimicked shown expressions
CASME [96] 2013	35	✓	X							X	Micro expressions, suppressed emotions

Table 3. Posed video databases.

With RGB-D databases, however, it is important to note that the data is unique to each sensor with outputs having varying density and error, so algorithms trained on databases like the IIIT-D RGB-D [115], VAP RGB-D [116] and KinectFaceDB [117] would be very susceptible to hardware changes. For comparison with the 3D databases, an RGB-D sample has been provided in **Figure 9**. One of the newer databases, the iCV SASE [118] database, is RGB-D dataset solely dedicated to headpose with free facial expressions.

Even though depth based databases, like the ones in **Table 7**, are relatively new compared to other types and there are very few of them, they still manage to cover a wide range of different emotions. With the release of commercial use depth cameras like the Microsoft Kinect [120], they will only continue to get more popular in the future.

http://dx.doi.org/10.5772/intechopen.72748

Database	Participants	Elicitation	Primary 6	Neutral	Contempt	Embarrassment	Pain	Smile	Positive	Negative	Other	Additional information
IAPS [105] 1997	497–1483	Visual media									X	Pleasure and arousal reaction images, subset for children
SD [32] 2004	28	AVM[1]	✓	X							X	One of the first international induced emotion data-sets
eNTERFACE'05 [46] 2006	42	Auditory media	X									Standard for face recognition algorithms
CK+ [44] 2010	220	Posed and AVM	X									Updated version of CK
SMIC [34] 2011	6	AVM	✓									Supressed emotions
Face Place [106] 2012	235	AVM	X	X							X	Different ethnicities
AM-FED [107] 2013	81–240	AVM		X				X				Reactions to Superbowl ads
MAHNOB [51] 2013	22	Posed and AVM	✓								X	Laughter recognition research
SASE-FE [108] 2017	54	AVM	✓		X							Fake emotions

[1]Audiovisual media.

Table 4. Media induced video databases.

As their applications are more specific, thermal facial expression datasets are very scarce. Some of the first and more known ones are IRIS [123] and Equinox [121, 122], which consist of RGB and thermal image pairs that are labelled with three emotions [124], as can be seen in **Figure 10**. Thermal databases are usually posed or induced by audiovisual media. The ones in **Table 8** mostly focus on positive, negative, neutral and six primary emotions. The average number of participants is quite high relative to other types of databases.

4.3.1. Audio databases

There are mainly two types of emotion databases that contain audio content: stand-alone audio databases and video databases that include spoken words or utterances. The information extracted from audio is called context and can be generally categorized into a multitude, wherein the three important context subdivisions for emotion recognition databases are the semantic, structural, and temporal ones.

Semantic context is where the emotion can be isolated through specific emotionally marked words, while structural context is dependent on the stress patterns and syntactic structure of longer phrases. **Temporal context** is the longer lasting variant of the structural context as it involves the change of emotion in speech over time, like emotional build-up [42].

Database	Participants	Elicitation	Primary 6	Neutral	Contempt	Embarrassment	Pain	Smile	Positive	Negative	Other	Additional information
ISL meeting corpus [35] 2002	90	Human-human interaction		X					X	X		Collected in a meeting fashion
AAI [36] 2004	60	Human-human interaction			X				X	X	X	Induced via past memories
AIBO database [23] 2004	30	Child-robot interaction	✓	X							X	Robot instructed by children
CSC corpus [37] 2005	32	Human-human interaction									X	Honesty research
RU-FACS [109] 2005	90	Human-human interaction	X	X								Subjects were all university students
SAL [11] 2005	24	human-computer interaction	✓	X								conversations held with a simulated "chat-bot" system
MMI [45] 2006	61/29	Posed/child-comedian interaction, adult-audiovisual media	X									Profile views along with portrait images
TUM AVIC [53] 2007	21	Human-human interaction									X	Commercial presentation
SEMAINE [110, 111] 2010/2012	150	Human-human interaction	X		X						X	Operator was thoroughly familiar with SAL script
AVDLC [12] 2013	292	Human-computer interaction									X	Mood disorder and unipolar depression research
RECOLA [112] 2013	46	Human-human interaction									X	Collaborative tasks. Audio-video, ECG and EDA were recorded

Table 5. Interaction induced video databases.

Database	Participants	Primary 6	Neutral	Contempt	Embarrassment	Pain	Smile	Positive	Negative	Other	Additional information
Belfast natural database [42] 2003	125	X	X	X						X	Video clips from television and interviews
Belfast Naturalistic Emotional Database [114] 2003	125	X								X	Studio recordings and television program clips
VAM [43] 2008	47									X	Video clips from a talk-show
AFEW [39, 40] 2011/2012	330	X	X								Video clips from movies
Spanish Multimodal Opinion [41] 2013	105							X	X		Spanish video clips from YouTube

Table 6. Spontaneous video databases.

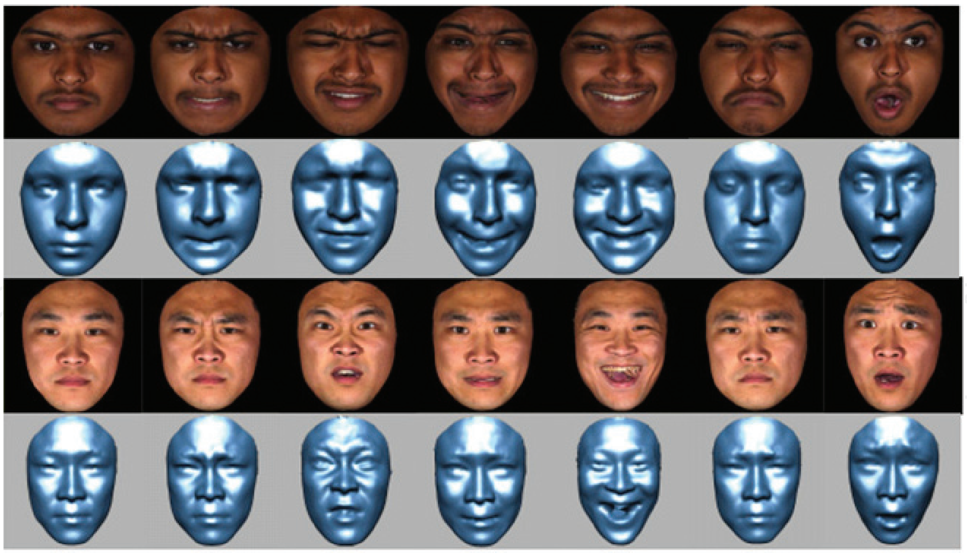

Figure 8. 3D facial expression samples from the BU-3DFE database [15].

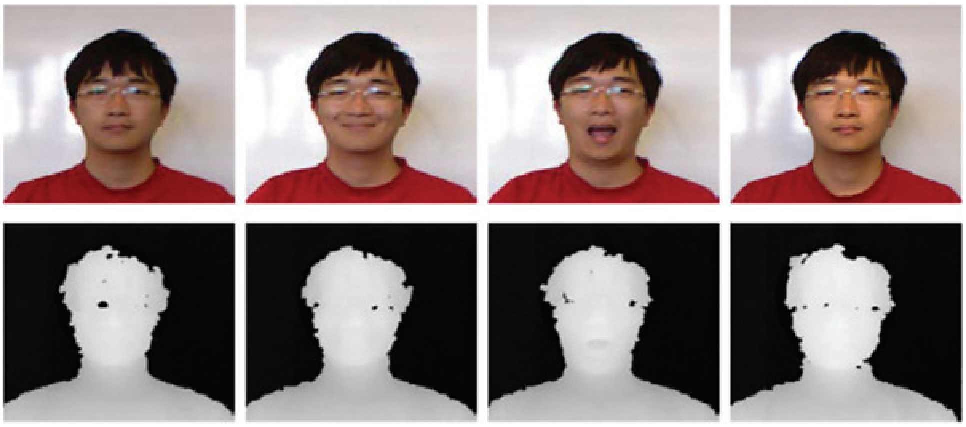

Figure 9. RGB-D facial expression samples from the KinectFaceDB database [117].

In case of multimodal data, the audio component can provide a semantic context, which can have a larger bearing on the emotion than the facial expressions themselves [11, 23]. However, in case of solely audio data, like the Bank and Stock Service [126] and ACC [127] databases, the context of the speech plays a quintessential role in emotion recognition [128, 129].

The audio databases in **Table 9** are very scarce and tailored to specific needs, like the Banse-Schrerer [26], which has only four participants and was gathered to see whether judges can deduce emotions from vocal cues. The easiest way to gather a larger amount of audio data is from call-centres, where the emotions are elicited either by another person or a computer program.

Even with all of the readily available databases out there, there is still a need for creating self-collected databases for emotion recognition, as the existing ones don't always fulfil all of the criteria [130–133].

Database	Participants	Format	Primary 6	Neutral	Contempt	Embarrassment	Pain	Smile	Positive	Negative	Other	Additional information
BU-3DFE [15] 2006	100	3D images	X									Ethnically diverse, two angled views
Bosphorus [14] 2008	105	3D images	X									Occlusions, less ethnic diversity than BU-3DF
BU-4DFE [16] 2008	101	3D videos										Newer version of BU-3DFE, has 3D videos
VAP RGB-D [116] 2012	31	RGB-D videos						X			X	17 different recorded states repeated 3 times for each person
PICS [119] 2013	—	Images, videos, 3D images										Includes several different datasets and is still ongoing
BP4D [17] 2014	41	3D videos	X			X	X					Human-human interaction
IIIT-D RGB-D [115] 2014	106	RGB-D images		X				X				Captured with Kinect
KinectFaceDB [117] 2014	52	RGB-D images, videos		X				X				Captured with Kinect, varying occlusions

Table 7. 3D and RGB-D databases.

Figure 10. Thermal images taken from the Equinox database [121, 122].

Database	Participants	Elicitation	Primary 6	Neutral	Contempt	Embarrassment	Pain	Smile	Positive	Negative	Other	Additional information
Equinox [121, 122] 2002	340	Posed		X					X	X		Captured in SWIR, MWIR and LWIR
IRIS [123] 2007	4228	Posed		X					X	X		Some of the first thermal FE data-sets
NVIE [47] 2010	215	Posed and AVM[1]	X									Spontaneous expressions are not present for every subject
KTFE [125] 2014	26	Posed and AVM	X	X								

[1]Audiovisual media.

Table 8. Thermal databases.

Database	Participants	Elicitation	Primary 6	Neutral	Contempt	Embarrassment	Pain	Smile	Positive	Negative	Other	Additional information
Banse-Scherer [26] 1996	4	Posed	X		X	X					X	Vocally expressed emotions
Bank and Stock Service [126] 2004	350	Human-human interaction	✓	X							X	Collected from a call center and Capital Bank Service Center
ACC [127] 2005	1187	Human-computer interaction		X						X		Collected from automated call center applications

Table 9. Audio databases.

5. Conclusion

With the rapid increase of computing power and size of data, it has become more and more feasible to distinguish emotions, identify people, and verify honesty based on video, audio or image input, taking a large step forward not only in human-computer interaction, but also in mental illness detection, medical research, security and so forth. In this paper an overview of existing face databases in varying categories has been given. They have been organised into tables to give the reader an easy way to find necessary data. This paper should be a good starting point for anyone who considers training a model for emotion recognition.

Acknowledgements

This work has been partially supported by Estonian Research Council Grant PUT638, The Scientific and Technological Research Council of Turkey 1001 Project (116E097), The Spanish project TIN2016-74946-P (MINECO/FEDER, UE), CERCA Programme/Generalitat de Catalunya, the COST Action IC1307 iV&L Net (European Network on Integrating Vision and Language) supported by COST (European Cooperation in Science and Technology), and the Estonian Centre of Excellence in IT (EXCITE) funded by the European Regional Development Fund. We also gratefully acknowledge the support of the NVIDIA Corporation with the donation of the Titan X Pascal GPU.

Author details

Rain Eric Haamer[1], Eka Rusadze[1], Iiris Lüsi[1], Tauseef Ahmed[1], Sergio Escalera[2] and Gholamreza Anbarjafari[1,3]*

*Address all correspondence to: shb@icv.tuit.ut.ee

1 iCV Research Group, Institute of Technology, University of Tartu, Tartu, Estonia

2 The Computer Vision Center and University of Barcelona, Barcelona, Spain

3 Department of Electrical and Electronic Engineering, Hasan Kalyoncu University, Gaizantep, Turkey

References

[1] Dix A. Human-computer interaction. In Encyclopedia of database systems. US: Springer. 2009:1327-1331

[2] Noroozi F, Marjanovic M. Njegus A, Escalera S, Anbarjafari G. Audio-visual emotion recognition in video clips. IEEE Transactions on Affective Computing; 2017

[3] Toumi T, Zidani A. From human-computer interaction to human-robot social interaction. arXiv preprint arXiv:1412.1251; 2014

[4] Daneshmand M, Abels A, Anbarjafari G. Real-time, automatic digi-tailor mannequin robot adjustment based on human body classification through supervised learning. International Journal of Advanced Robotic Systems. 2017;**14**(3):1729881417707169

[5] Bolotnikova A, Demirel H, Anbarjafari G. Real-time ensemble based face recognition system for NAO humanoids using local binary pattern. Analog Integrated Circuits and Signal Processing. 2017;**92**(3):467-475

[6] Valstar MF, Schuller BW, Smith K, Eyben F, Jiang B, Bilakhia S, Schnieder S, Cowie R, Pantic M. AVEC 2013: The continuous audio/visual emotion and depression recognition challenge. In: AVEC-ACM Multimedia, Barcelona, Spain; 2013

[7] Gross R, Baker S, Matthews I, Kanade T. Handbook of face recognition. In: Li SZ, Jain AK, editors. Handbook of Face Recognition. 2005:193-216

[8] Jain AK, Li SZ. Handbook of Face Recognition. Springer; 2011

[9] Face databases. http://web.mit.edu/emeyers/ www.face_databases.html [Accessed 31 March 2017]

[10] 60 facial recognition databases. https://www.kairos.com/blog/60-facial-recognition-databases [Accessed 31 March 2017]

[11] Athanaselis T, Bakamidis S, Dologlou I, Cowie R, Douglas-Cowie E, Cox C. ASR for emotional speech: Clarifying the issues and enhancing performance. Neural Networks. 2005;**18**(4):437-444

[12] Valstar M, Schuller B, Smith K, Eyben F, Jiang B, Bilakhia S, Schnieder S, Cowie R, Pantic M. AVEC 2013: The continuous audio/visual emotion and depression recognition challenge. In: Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge; ACM; 2013. pp. 3-10

[13] Jaimes A, Sebe N. Multimodal human–computer interaction: A survey. Computer Vision and Image Understanding. 2007;**108**(1):116-134

[14] Savran A, Alyüz N, Dibeklioğlu H, Çeliktutan O, Gökberk B, Sankur B, Akarun L. Bosphorus database for 3D face analysis. In: European Workshop on Biometrics and Identity Management; Springer; 2008. pp. 47-56

[15] Yin L, Wei X, Sun Y, Wang J, Rosato MJ. A 3D facial expression database for facial behavior research. In: Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on; IEEE; 2006. pp. 211-216

[16] Yin L, Chen X, Sun Y, Worm T, Reale M. A high-resolution 3D dynamic facial expression database. In: 8th IEEE International Conference on Automatic Face and Gesture Recognition, 2008. FG08. ; IEEE; 2008. pp. 1-6

[17] Zhang X, Yin L, Cohn JF, Canavan S, Reale M, Horowitz A, Liu P, Girard JM. Bp4d-spontaneous: A high-resolution spontaneous 3D dynamic facial expression database. Image and Vision Computing. 2014;**32**(10):692-706

[18] NIST. Special database 18: Mugshot Identification Database (MID)

[19] Bruce V, Young A. Understanding face recognition. British Journal of Psychology. 1986;**77**(3): 305-327

[20] Richard G, Mengay Y, Guis I, Suaudeau N, Boudy J, Lockwood P, Fernandez C, Fernández F, Kotropoulos C, Tefas A, et al. Multi modal verification for teleservices and security applications (M2VTS). IEEE International Conference on Multimedia Computing and Systems, 1999; IEEE. 1999;**2**:1061-1064

[21] Grgic M, Delac K, Grgic S. Scface–surveillance cameras face database. Multimedia Tools and Applications. 2011;**51**(3):863-879

[22] Gunes H, Piccardi M. Bi-modal emotion recognition from expressive face and body gestures. Journal of Network and Computer Applications. 2007;**30**(4):1334-1345

[23] Batliner A, Hacker C, Steidl S, Nöth E, D'Arcy S, Russell MJ, Wong M. "You stupid tin box"-children interacting with the AIBO robot: A cross-linguistic emotional speech corpus. In: LREC, Lisbon, Portugal; 2004

[24] Wu C-H, Lin J-C, Wei W-L. Survey on audiovisual emotion recognition: Databases, features, and data fusion strategies. APSIPA Transactions on Signal and Information Processing. 2014;**3**:e12

[25] Sebe N, Cohen I, Gevers T, Huang TS. Multimodal approaches for emotion recognition: A survey. In: Electronic Imaging 2005; International Society for Optics and Photonics; 2005. pp. 56-67

[26] Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. Journal of personality and social psychology. 1996;**70**(3):614

[27] Kanade T, Cohn JF, Tian Y. Comprehensive database for facial expression analysis. In: Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000; IEEE; 2000. pp. 46-53

[28] Lawrence Shao-Hsien Chen. Joint processing of audio-visual information for the recognition of emotional expressions in human-computer interaction [PhD thesis]. Citeseer; 2000

[29] Ebner NC, Riediger M, Lindenberger U. Faces—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. Behavior Research Methods. 2010;**42**(1):351-362

[30] Zeng Z, Pantic M, Roisman GI, Huang TS. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009;**31**(1):39-58

[31] Bänziger T, Pirker H, Scherer K. Gemep-geneva multimodal emotion portrayals: A corpus for the study of multimodal emotional expressions. Proceedings of LREC. 2006;**6**:15-19

[32] Sebe N, Lew MS, Sun Y, Cohen I, Gevers T, Huang TS. Authentic facial expression analysis. Image and Vision Computing. 2007;**25**(12):1856-1863

[33] O'Toole AJ, Harms J, Snow SL, Hurst DR, Pappas MR, Ayyad JH, Abdi H. A video database of moving faces and people. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005;**27**(5):812-816

[34] Pfister T, Li X, Zhao G, Pietikäinen M. Recognising spontaneous facial micro-expressions. In: IEEE International Conference on Computer Vision (ICCV), 2011; IEEE; 2011. pp. 1449-1456

[35] Burger S, MacLaren V, Yu H. The ISL meeting corpus: The impact of meeting type on speech style. In: INTERSPEECH, Denver, Colorado, USA; 2002

[36] Roisman GI, Tsai JL, Chiang K-HS. The emotional integration of childhood experience: Physiological, facial expressive, and self-reported emotional response during the adult attachment interview. Developmental Psychology. 2004;**40**(5):776

[37] Hirschberg J, Benus S, Brenier JM, Enos F, Friedman S, Gilman S, Girand C, Graciarena M, Kathol A, Michaelis L, et al. Distinguishing deceptive from non-deceptive speech. In: Interspeech; 2005. pp. 1833-1836

[38] Kirouac G, Dore FY. Accuracy of the judgment of facial expression of emotions as a function of sex and level of education. Journal of Nonverbal Behavior. 1985;**9**(1):3-7

[39] Dhall A, Goecke R, Lucey S, Gedeon T. Acted facial expressions in the wild database. Australian National University, Canberra. Technical Report TR-CS-11, 2; 2011

[40] Dhall A, Lucey S, Joshi J, Gedeon T. Collecting Large, Richly Annotated Facial-Expression Databases from Movies, IEEE MultiMedia, 2012;**19**(3):34-41

[41] Rosas VP, Mihalcea R, Morency L-P. Multimodal sentiment analysis of Spanish online videos. IEEE Intelligent Systems. 2013;**28**(3):38-45

[42] Douglas-Cowie E, Campbell N, Cowie R, Roach P. Emotional speech: Towards a new generation of databases. Speech Communication. 2003;**40**(1):33-60

[43] Grimm M, Kroschel K, Narayanan S. The Vera am Mittag German audio-visual emotional speech database. In: IEEE International Conference on Multimedia and Expo, 2008; IEEE; 2008. pp. 865-868

[44] Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010; IEEE; 2010. pp. 94-101

[45] Pantic M, Patras I. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). 2006;**36**(2):433-449

[46] Martin O, Kotsia I, Macq B, Pitas I. The enterface'05 audio-visual emotion database. In: . Proceedings of 22nd International Conference on Data Engineering Workshops, 2006; IEEE; 2006. p. 8

[47] Wang S, Liu Z, Lv S, Lv Y, Wu G, Peng P, Chen F, Wang X. A natural visible and infrared facial expression database for expression recognition and emotion inference. IEEE Transactions on Multimedia. 2010;**12**(7):682-691

[48] Ekman P, Friesen WV. Pictures of facial affect. Consulting Psychologists Press; 1975

[49] Ekman P. Facial expression and emotion. American Psychologist. 1993;**48**(4):384

[50] Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor JG. Emotion recognition in human-computer interaction. Signal Processing Magazine, IEEE. 2001;**18**(1):32-80

[51] Petridis S, Martinez B, Pantic M. The mahnob laughter database. Image and Vision Computing. 2013;**31**(2):186-202

[52] Gorbova J, Baró X, Escalera S, Demirel H, Allik J, Ozcinar C, Lüsi I, Jacques JCS, Anbarjafari G. Joint challenge on dominant and complementary emotion recognition using micro emotion features and head-pose estimation: Databases. IEEE; 2017

[53] Schuller B, Müeller R, Höernler B, Höethker A, Konosu H, Rigoll G. Audiovisual recognition of spontaneous interest within conversations. In: Proceedings of the 9th International Conference on Multimodal Interfaces; ACM; 2007. pp. 30-37

[54] Wang Y, Guan L. Recognizing human emotional state from audiovisual signals. IEEE Transactions on Multimedia. 2008;**10**(5):936-946

[55] Polikovsky S, Kameda Y, Ohta Y. Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor. In:, 3rd International Conference on Crime Detection and Prevention (ICDP 2009); IET; 2009. pp. 1-6

[56] Loob C, Rasti P, Lüsi I, Jacques JCS, Baró X, Escalera S, Sapinski T, Kaminska D, Anbarjafari G. Dominant and complementary multi-emotional facial expression recognition using c-support vector classification. In: 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017); IEEE; 2017. pp. 833-838

[57] Ekman P, Friesen WV, Tomkins SS. Facial affect scoring technique: A first validity study. Semiotica. 1971;**3**(1):37-58

[58] Darwin C. The Expression of the Emotions in Man and Animals. New York: Oxford University Press; 1998

[59] Guillaume-Benjamin Duchenne. Mécanisme de la physionomie humaine: où, Analyse électro-physiologique de l'expression des passions. J.-B. Baillière, 1876.

[60] Hjortsjö C-H. Man's Face and Mimic Language. Lund: Studentlitteratur; 1969

[61] Ekman P, Friesen WV, Hager J. The facial action coding system (FACS): A technique for the measurement of facial action. Palo Alto: Consulting Psychologists Press, Inc.; 1983. Ekman P, Levenson RW, Friesen WV. Auto-nomic nervous system activity distinguishes among emotions. Science. 1978;**221**:1208-1212

[62] Mavadati SM, Mahoor MH, Bartlett K, Trinh P, Cohn JF. DISFA: A spontaneous facial action intensity database. IEEE Transactions on Affective Computing. 2013;**4**(2):151-160

[63] Pantic M, Valstar M, Rademaker R, Maat L. Web-based database for facial expression analysis. In: IEEE International Conference on Multimedia and Expo, 2005. ICME 2005; IEEE; 2005. p. 5

[64] Valstar M, Pantic M. Induced disgust, happiness and surprise: An addition to the MMI facial expression database. In: Proceedings of the 3rd International Workshop on EMOTION (Satellite of LREC): Corpora for Research on Emotion and Affect; 2010. p. 65

[65] Kleiner M, Wallraven C, Bülthoff HH. The MPI VideoLab-a system for high quality synchronous recording of video and audio from multiple viewpoints. Tübingen: MPI; 2004. p. 123

[66] Kaulard K, Cunningham DW, Bülthoff HH, Wallraven C. The MPI facial expression database—A validated database of emotional and conversational facial expressions. PloS One. 2012;**7**(3):e32321

[67] Cosker D, Krumhuber E, Hilton A. A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling. In: Computer Vision (ICCV), 2011 IEEE International Conference on; IEEE; 2011. pp. 2296-2303

[68] Hager JC, Ekman P, Friesen WV. Facial action coding system. Salt Lake City: A Human Face. Technical Report. ISBN: 0-931835-01-1, 2002

[69] Cohn JF, Ambadar Z, Ekman P. Observer-based measurement of facial expression with the facial action coding system. In: The Handbook of Emotion Elicitation and Assessment; 2007. pp. 203-221

[70] Julle-Daniere E, Micheletta J, Whitehouse J, Joly M, Gass C, Burrows AM, Waller BM. Maqfacs (macaque facial action coding system) can be used to document facial movements in *Barbary macaques* (*Macaca sylvanus*). PeerJ. 2015;**3**:e1248

[71] Oster H. Baby FACS: Facial action coding system for infants andyoung children (Unpublished monograph and coding manual). New York: New York University; 2006

[72] Gross R, Matthews I, Cohn J, Kanade T, Baker S. Multi-PIE. Image and Vision Computing. 2010;**28**(5):807-813

[73] Ekman P, Freisen W. Pictures of Facial Affect. Palo Alto: Consulting Psychologists; 1976

[74] Michael Lyons, Shigeru Akamatsu, Miyuki Kamachi, and Jiro Gyoba. Coding facial expressions with Gabor wavelets. In: Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998; IEEE; 1998. pp. 200-205

[75] Belhumeur PN, Kriegman DJ. The Yale face database. http://cvc.yale.edu/projects/yalefaces/yalefaces.html. 1997;**1**(2):4

[76] Matsumoto D, Ekman P. Japanese and Caucasian Facial Expressions of Emotion (JACFEE) and Neutral Faces (JACNeuF). 1995

[77] Samaria FS, Harter AC. Parameterisation of a stochastic model for human face identification. In: Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on; IEEE; 1994. pp. 138-142

[78] Cambridge AL. The Olivetti Research Ltd. database of faces

[79] Phillips PJ, Wechsler H, Huang J, Rauss PJ. The FERET database and evaluation procedure for face-recognition algorithms. Image and Vision Computing. 1998;**16**(5):295-306

[80] Karolinska Directed Emotional Faces (KDEF). http://www.emotionlab.se/resources/kdef [Accessed: 31 March 2017]

[81] Martinez AM. The AR face database. CVC Technical Report, 24, 1998

[82] Beaupré M, Cheung N, Hess U. La reconnaissance des expressions émotionnelles faciales par des décodeurs africains, asiatiques, et caucasiens. In: Poster presented at the annual meeting of the Société Québécoise pour la Recherche en Psychologie, Hull, Quebec; 2000

[83] Dailey M, Cottrell GW, Reilly J. California Facial Expressions (Cafe). Unpublished digital images, University of California, San Diego, Computer Science and Engineering Department; 2001

[84] Sim T, Baker S, Bsat M. The CMU pose, illumination, and expression (PIE) database. In: Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002; IEEE; 2002. pp. 53-58

[85] Jain V, Mukherjee A. The Indian Face Database, 2002

[86] Nimstim Face Stimulus Set. http://www.macbrain.org/resources.htm [Accessed: 31 March 2017]

[87] Roh M-C, Lee S-W. Performance analysis of face recognition algorithms on Korean face database. International Journal of Pattern Recognition and Artificial Intelligence. 2007;**21**(06):1017-1033

[88] Minear M, Park DC. A lifespan database of adult facial stimuli. Behaviour Research Methods, Instruments, & Computers. 2004;**36**:630-633

[89] Chen L-F, Yen Y-S. Taiwanese Facial Expression Image Database. Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei, 2007

[90] Gao W, Cao B, Shan S, Chen X, Zhou D, Zhang X, Zhao D. The CAS-PEAL large-scale Chinese face database and baseline evaluations. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans. 2008;**38**(1):149-161

[91] Kasinski A, Florek A, Schmidt A. The PUT face database. Image Processing and Communications. 2008;**13**(3-4):59-64

[92] Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A. Presentation and validation of the Radboud Faces Database. Cognition and Emotion.2010;**24** (8):1377-1388

[93] Ekman P, Friesen WV. Nonverbal leakage and clues to deception. Psychiatry. 1969;**32** (1):88-106

[94] Shreve M, Godavarthy S, Goldgof D, Sarkar S. Macro-and micro-expression spotting in long videos using spatio-temporal strain. In: IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011); IEEE; 2011. pp. 51-56

[95] Warren G, Schertler E, Bull P. Detecting deception from emotional and unemotional cues. Journal of Nonverbal Behavior. 2009;**33**(1):59-69

[96] Yan W-J, Wu Q, Liu Y-J, Wang S-J, Fu X. CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces. In: 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013; IEEE; 2013. pp. 1-7

[97] Black MJ, Yacoob Y. Recognizing facial expressions in image sequences using local parameterized models of image motion. International Journal of Computer Vision. 1997;**25**(1):23-48

[98] Battocchi A, Pianesi F. Dafex: Un database di espressioni facciali dinamiche. In: Proceedings of the SLI-GSCP Workshop; 2004. pp. 311-324

[99] Baron-Cohen S, Golan O, Wheelwright S, Hill JJ. Mind Reading: The Interactive Guide to Emotions. London: Jessica Kingsley; 2004

[100] Jiang P, Ma J, Minamoto Y, Tsuchiya S, Sumitomo R, Ren F. Orient video database for facial expression analysis. Age. 2007;**20**:40

[101] Busso C, Bulut M, Lee C-C, Kazemzadeh A, Mower E, Kim S, Chang JN, Lee S, Narayanan SS. IEMOCAP: Interactive emotional dyadic motion capture database. Language Resources and Evaluation. 2008;**42**(4):335

[102] Haq S, Jackson PJB, Edge J. Speaker-dependent audio-visual emotion recognition. In: AVSP; 2009. pp. 53-58

[103] Roy S, Roy C, Fortin I, Ethier-Majcher C, Belin P, Gosselin F. A dynamic facial expression database. Journal of Vision. 2007;**7**(9):944-944

[104] Wingenbach TSH, Ashwin C, Brosnan M. Validation of the Amsterdam dynamic facial expression set–bath intensity variations (ADFES-BIV): A set of videos expressing low, intermediate, and high intensity emotions. PLoS One. 2016;**11**(1):e0147112

[105] Lang PJ, Bradley MM, Cuthbert BN. International affective picture system (IAPS): Technical manual and affective ratings. In: NIMH Center for the Study of Emotion and Attention; 1997. pp. 39-58

[106] Face Place. http://wiki.cnbc.cmu.edu/Face_Place [Accessed: 31 March 2017]

[107] McDuff D, Kaliouby RE, Senechal T, Amr M, Cohn JF, Picard R Affectiva-MIT facial expression dataset (AM-FED): Naturalistic and spontaneous facial expressions collected "In-the-Wild". In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2013. pp. 881-888

[108] Corneanu CA, Escalera S, Baro X, Hyniewska S, Allik J, Anbarjafari G, Ofodile I, Kulkarni K. Automatic recognition of deceptive facial expressions of emotion. arXiv preprint arXiv:1707.04061, 2017

[109] Bartlett MS, Littlewort G, Frank M, Lainscsek C, Fasel I, Movellan J. Recognizing facial expression: machine learning and application to spontaneous behavior. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 2; IEEE; 2005. pp. 568-573

[110] McKeown G, Valstar M, Cowie R, Pantic M, Schroder M. The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. IEEE Transactions on Affective Computing. 2012;**3**(1):5-17

[111] McKeown G, Valstar MF, Cowie R, Pantic M. The SEMAINE corpus of emotionally coloured character interactions. In: Multimedia and Expo (ICME), 2010 IEEE International Conference on; IEEE; 2010. pp. 1079-1084

[112] Ringeval F, Sonderegger A, Sauer J, Lalanne D. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In: Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on; IEEE; 2013. pp. 1-8

[113] Henry SG, Fetters MD. Video elicitation interviews: A qualitative research method for investigating physician-patient interactions. The Annals of Family Medicine. 2012;**10**(2):118-125

[114] Douglas-Cowie E, Cowie R, Schroeder M. The description of naturally occurring emotional speech. In: Proceedings of 15th International Congress of Phonetic Sciences, Barcelona; 2003

[115] Goswami G, Vatsa M, Singh R. RGB-D face recognition with texture and attribute features. IEEE Transactions on Information Forensics and Security. 2014;**9**(10):1629-1640

[116] Hg RI, Jasek P, Rofidal C, Nasrollahi K, Moeslund TB, Tranchet G. An RGB-D database using Microsoft's Kinect for windows for face detection. In: Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on; IEEE; 2012. pp. 42-46

[117] Min R, Kose N, Dugelay J-L. KinectFaceDB: A Kinect database for face recognition. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2014;**44**(11):1534-1548

[118] Lüsi I, Escarela S, Anbarjafari G. SASE: RGB-depth database for human head pose estimation. In: Computer Vision–ECCV 2016 Workshops; Springer; 2016. pp. 325-336

[119] Psychological image collection at Stirling (PICS). http://pics.psych.stir.ac.uk/ [Accessed: 31 March 2017]

[120] Microsoft, "Microsoft Kinect." http://www.xbox.com/en-US/xbox-one/accessories/kinect-for-xbox-one [Accessed: 28 March 2017]

[121] Wolff LB, Socolinsky DA, Eveland CK. Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery. In Proceedings of SPIE. 2002;**4820**:140-151

[122] Equinox Corporation. "Equinox face database". 2002

[123] Akhloufi M, Bendada A, Batsale J-C. State of the art in infrared face recognition. Quantitative InfraRed Thermography Journal. 2008;**5**(1):3-26

[124] Corneanu CA, Simón MO, Cohn JF, Guerrero SE. Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related

applications. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016;**38**(8): 1548-1568

[125] Nguyen H, Kotani K, Chen F, Le B. A thermal facial emotion database and its analysis. In: Pacific-Rim Symposium on Image and Video Technology; Springer; 2013. pp. 397-408

[126] Devillers L, Vasilescu I. Reliability of lexical and prosodic cues in two real-life spoken dialog corpora. In: LREC; 2004

[127] Lee CM, Narayanan SS. Toward detecting emotions in spoken dialogs. IEEE Transactions on Speech and Audio Processing. 2005;**13**(2):293-303

[128] Robert Ladd D, Scherer K, Silverman K. An integrated approach to studying intonation and attitude. Intonation in Discourse. London/Sidney: Crom Helm. 1986;**125**:138

[129] Cauldwell RT. Where did the anger go? The role of context in interpreting emotion in speech. In: ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion; 2000

[130] Song M, You M, Li N, Chen C. A robust multimodal approach for emotion recognition. Neurocomputing. 2008;**71**(10):1913-1920

[131] Zeng Z, Jilin T, Pianfetti BM, Huang TS. Audio–visual affective expression recognition through multistream fused HMM. IEEE Transactions on Multimedia. 2008;**10**(4):570-577

[132] Wan J, Escalera S, Anbarjafari G, Escalante HJ, Baró X, Guyon I, Madadi M, Allik J, Gorbova J, Chi L, Yiliang X. Results and analysis of ChaLearn LAP multi-modal isolated and continuous gesture recognition, and real versus fake expressed emotions challenges. In ChaLearn LaP, Action, Gesture, and Emotion Recognition Workshop and Competitions: Large Scale Multimodal Gesture Recognition and Real Versus Fake Expressed Emotions, ICCV; 2017;**4**(6)

[133] Lu K, Jia Y. Audio-visual emotion recognition with boosted coupled HMMM. In: 21st International Conference on Pattern Recognition (ICPR), 2012; IEEE; 2012. pp. 1148-1151