A Survey on Databases for Facial Expression Analysis

Raphaël Weber, Catherine Soladié and Renaud Séguier
FAST team, IETR lab, CentraleSupelec, Avenue de la Boulaie, 35576 Cesson-Sévigné, France

Keywords: Facial Expression, Database, Multimodal, Survey.

Abstract: Facial expression databases are essential to develop and test a system of facial expressions analysis. We propose in this paper a survey based on the review of 61 databases. To the best of our knowledge, there are no other surveys with so many databases. We identify 18 characteristics to describe the database and group them in 6 categories, (population, modalities, data acquisition hardware, experimental conditions, experimental protocol and annotations). These characteristics are useful to create or choose a database relevant to the targeted context application. We propose to classify the databases according to these characteristics so it can be helpful for researchers to choose the database suited to their context application. We bring to light the trends between posed, spontaneous and in-the-wild databases. We finish with future directions, including crowd sourcing and databases with groups of people.

1 INTRODUCTION

Since human face conveys information about the emotional state, automatic facial expression analysis has gained a growing interest in the past decades. A wide range of applications are covered, such as human-computer interaction or medical applications. Facial expression databases are essential to develop and test a system of facial expression analysis.

The first public facial expression databases appeared in the late 1990s (Lyons et al., 1998), (Lundqvist et al., 1998), (Kanade et al., 2000) as automatic facial expression analysis was taking off. Some of these databases are still used today as a test bed in order to compare to other methods. These first databases contain posed expressions acquired in the laboratory environment. Differences exist between posed expressions and spontaneous expressions, the latter being expressions that a person naturally displays in everyday life (Cohn and Schmidt, 2004), (Schmidt et al., 2006), (Valstar et al., 2007). These differences rely in the intensity and the dynamics of the expression. A system trained on posed expressions will be less performant when testing on spontaneous expressions. Therefore databases of spontaneous expressions began to appear in the 2000s (Schmidt and Cohn, 2001), (Toole et al., 2005). Since then, new databases are made available almost every year. More recently, automatic facial expression analysis in in-the-wild conditions have been identified as one of the main challenge to tackle (Martinez and Valstar, 2016). In-the-wild conditions refer to an unconstrained environment, as met in real life conditions. Databases meeting this criteria know a growing interest since the beginning of the 2010s.

A database is defined by many characteristics going from the number of subjects to the annotations describing the data. They have a direct impact on the use of the database. For instance, a database of frontal posed expressions is not suitable to train and test a system aimed at analyzing in-the-wild expressions. So, before creating or choosing a database, one must properly identify the targeted application context in order to define the desired database characteristics. We propose to group the different characteristics of a database in 6 categories: population, modalities, data acquisition hardware, experimental conditions, experimental protocol and annotations.

To the best of our knowledge, the existing surveys on databases of facial expressions only review about 15 databases (Anitha et al., 2010), 20 databases (Cowie et al., 2005) (Fu et al., 2012) or 30 databases (Kaulard et al., 2012). In this paper, we propose a survey based on the review of 61 databases, taking into account unimodal databases (only facial expressions) as well as multimodal databases (facial expressions combined with other modalities). In section 2, we review the databases according to their characteristics while bringing to light the different trends between posed, spontaneous and in-the-wild databases. In section 3, we indicate the future directions. We conclude the paper in section 4.

Weber, R., Soladié, C. and Séguier, R.
A Survey on Databases for Facial Expression Analysis.
DOI: 10.5220/0006553900730084
In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 5: VISAPP*, pages 73-84
ISBN: 978-989-758-290-5

2 SURVEY ACCORDING TO THE CHARACTERISTICS

In this section, we review the existing databases according to their characteristics, each subsection corresponding to one of the 6 categories as reported in table 1. We use a codification of the characteristics in order to refer to them easily in the following tables. When possible, we attempt to compare databases of posed, spontaneous and in-the-wild expressions. In this case, we use the following formatting to distinguish them in tables: no particular formatting for posed databases, bold for spontaneous databases and italic for in-the-wild databases. For the sake of clarity, we don't report the references to the databases in this section. The reader can find them in the appendix.

Table 1: Characteristics of a database. We propose a codification of these characteristics in order to refer to them easily in the following tables.

Category	Characteristic	Code
Population	# of subjects	P.1
	Women/Men %	P.2
	Age range	P.3
	Ethnic group(s)	P.4
Modalities	Available modalities	M.1
Data acquisition hardware	# of cameras	AE.1
	Resolution	AE.2
	FPS	AE.3
Experimental conditions	Background	EC.1
	Lightning	EC.2
	Occlusions	EC.3
	Head pose	EC.4
Experimental protocol	Method of acquisition	EP.1
	Available expressions	EP.2
Annotations	Facial features	A.1
	Action units (FACS)	A.2
	Emotional labels	A.3
	Emotional dimensions	A.4

2.1 Population

The characteristics of population are the number of subjects (P.1), the women/men distribution (P.2), the age range of the subjects (P.3) and the ethnic groups contained in the population (P.4). The choice of population is important because of interpersonal variability: shape and texture of the face varies with identity, gender, age and ethnic group. For example, the mean opening of the eyes differs between Asians and Caucasians. In order to develop a method that is robust to interpersonal variability, the database should contain the broadest range of ethnic groups and a good distribution of age and sex among the subjects, *i.e.* an interpersonal variability as great as possible.

Table 2 reports a classification of the databases according the number of subjets (P.1). A majority of the databases of posed expressions have less than 50 subjects. Comparatively, more databases of spontaneous and in-the-wild databases contain more than 90 subjects.

The women/men percentage (P.2) is most of the time between 40/60 and 60/40. However, there are exceptions such as JAFFE only containing women. Databases with mostly women ($\geq$ 70%) are CK, Belfast Naturalistic, UT-Dallas and CAS(ME)2. Databases with mostly men ($\geq$ 70%) are Multi-PIE, NVIE and ICT-3DRFE.

There are two main trends for age range (P.3): low (18-30 years old) and moderate (18-40 to 18-60 years old). Radboud Faces and AFEW are the exceptions since they contain children.

Most of the databases contain various ethnic groups (P.4) such as Caucasian, African Subsaharian or Asian, Caucasian group being a majority in this case. However, some databases propose only one or two specific ethnic groups: Caucasian and Asian (OULU-CASIA), Caucasian and Mediterranean (Radboud Faces, ADFES), Caucasian and South-American (BINED - Set 3), Asian only (JAFFE, NVIE, CAS(ME)2, CHEAVD), Turkish only (BAUM-1) and Caucasian only (GEMEP, D3DFACS, BINED - Set 1 and 2, DynEmo).

2.2 Modalities

Modalities refer to the nature of the acquired signals. We can distinguish databases according to the number of modalities: unimodal (only one modality) vs. multimodal (two or more modalities). Historically, the first databases are unimodal with 2D video (University Of Maryland, CK) or image (JAFFE, KDEF) of the face. 2D video is essential to study the dynamics of facial expression. In the 2000s, bimodal databases began to appear, due to interest in audio-visual emotion analysis (Zeng et al., 2009). The first database combining facial expression and audio is Belfast Naturalistic. In parallel, Smile Database is the first database to combine 2D video of face and physiological signals, in order to analyze smile. Later, two new modalities are investigated: body movement and 3D face model. The first databases combining facial expression and body movement are FABO and GEMEP, the latter adding also audio. The first database of static 3D model of facial expression is BU-3D FE. The same research team rapidly proposed BU-4D FE with dynamic 3D model.

Table 2: Classification of the databases according to the characteristic P.1 (number of subjects). The following formatting distinguishes databases: no particular formatting for posed databases, bold for spontaneous databases and italic for in-the-wild databases.

P.1	Databases
≤ 50	University Of Maryland, JAFFE, PICS - Pain Expressions, MMI, GEMEP, FABO, D3DFACS, ICT-3DRFE, ADFES, MPI, B3D(AC)2, DISFA+, **ENTERFACE**, **SAL**, **EmoTABOO**, **IEMOCAP**, **MMI+**, **MUG**, **SEMAINE**, **CAM3D**, **MAHNOB-HCI**, **DEAP**, **DISFA**, **RECOLA**, **CAS(ME)2**, **BP4D-Spontaneous**, **BAUM-1**, *EmoTV*
$\in (50,90)$	KDEF, OULU-CASIA, MUG, Radboud Faces, **BINED - Set 2 and 3**
$\in (90,130)$	BU-3D FE, BU-4D FE, Bosphorus, **Smile Database**, **RU-FACS**, **BINED - Set 1**, **UNBS-McMaster Shoulder Pain Expression Archive**, **PICS - Stirling ESRC 3D Face DB**, **BioVid Emo**, **GFT**, *Belfast Naturalistic*, *VAM*
$\in (180,250)$	CK, **CK+**, **NVIE**, *AM-FED*, *Vinereactor*, *CHEAVD*
≥ 280	Multi-PIE, **UT-Dallas**, **DynEmo**, **AVEC 2013 AViD-Corpus**, *AFEW*, *SFEW*, *HAPPEI*, *Aff-Wild*

Thus, the available modalities are facial expression (2D or 3D), audio, body movement and physiological signals. The exception is NVIE, combining facial expression both in the visible and infrared domain. Table 3 reports a classification of the databases according to these modalities (M.1). We can notice that spontaneous databases are often multimodal and cover every possible modality. MAHNOB-HCI and RECOLA are particularly interesting because they combine facial expression, audio and physiological signals, which makes them ideal databases for multimodal emotion analysis. In-the-wild databases are either unimodal (facial expression) or audio-visual. It is indeed very challenging to get a high interpersonal variability with the modalities of body movement, 3D model and physiological signals since they require a heavy hardware setup.

2.3 Data Acquisition Hardware

We focus here on the data acquisition hardware for image and video. We consider 3 characteristics: number of cameras (AE.1), camera resolution (AE.2) and frame per second (FPS, AE.3).

Regarding the number of cameras (AE.1), approximately half of the reviewed databases use only 1 camera facing the subject. There are 3 use cases when several cameras are used: 3D acquisition, multi-view acquisition and body movement acquisition. For example for 3D acquisition, BU-4D FE and BP4D-Spontaneous use 2 stereo cameras and 1 texture video camera. Multi-view acquisition refer to simultaneous image or video acquisition of the face from different views. The cameras are always installed on different angles of profile view (angle of yaw). The exceptions are Multi-PIE and Bosphorus in which additional cameras are mounted above the subjects, thus combining yaw and pitch angles. Body movement acquisition also require several cameras. Most of the time 1 camera is dedicated for facial expression acquisition and 1 or more camera is dedicated for body movement acquisition. The following databases are concerned (in brackets the number of cameras for body movement acquisition): FABO (1), EmoTABOO (1), GEMEP (2) and RU-FACS (3). In MAHNOB-HCI, 5 cameras are dedicated to multi-view acquisition and 1 camera to body movement acquisition. In IEMOCAP, a motion capture system is used to capture facial expression and hand movements, which makes it a singular database.

The choice of camera resolution (AE.2) and the FPS (AE.3) depend on the application context or the topic of study. For instance, in real-life conditions, it is likely that the camera resolution and/or FPS are low (e.g. with low-cost webcam). Contrarily, if one wants to study facial expression dynamics, it is advised to have a high FPS. The majority of the databases contain images/videos with a resolution (AE.2) of approximately 720x576 pixels and videos with a FPS (AE.3) between 24 and 30. These are the typical characteristics of consumer cameras.

Regarding camera resolution (AE.2), a few databases propose a low resolution of approximately 320x240 pixels: OULU-CASIA, VAM, AM-FED, Vinereactor. Apart from OULU-CASIA, these are in-the-wild databases, confirming the idea that the resolution is likely to be low in real-life conditions. Contrarily, there are much more databases, both posed and spontaneous, with a high resolution of approximately 1024x768: FABO, Multi-PIE, BU-3D FE, BU-4D FE, Bosphorus, D3DFACS, ICT-3DRFE, DISFA+, BINED - Set 3, DISFA, PICS - Stirling ESRC 3D Face Database, RECOLA, BP4D-Spontaneous, BioVid Emo.

There are few databases with a low or high FPS (AE.3). The following databases have a FPS smal-

Table 3: Classification of the databases according to the characteristic M.1 (modalities). The following formatting distinguishes databases: no particular formatting for posed databases, bold for spontaneous databases and italic for in-the-wild databases. “Physio. sig.” refers to physiological signals.

M.1	Databases
2D image	JAFFE, KDEF, PICS - Pain Expressions, Multi-PIE, Radboud Faces, *SFEW*, *HAPPEI*
2D video	University Of Maryland, CK, OULU-CASIA, DISFA+, **CK+**, **MUG**, **UNBC-McMaster Shoulder Pain Expression Archive**, **DISFA**, **CAS(ME)2**, **GFT**, *AM-FED*, *Vinereactor*
2D video + 2D image	MMI, ADFES, **UT-Dallas**, *Aff-Wild*
2D video + infrared video	**NVIE**
2D video + audio	**SAL**, **MMI+**, **SEMAINE**, **AVEC 2013 AViD-Corpus**, **BAUM-1**, *Belfast Naturalistic*, *EmoTV*, *AFEW*, *CHEAVD*
2D video + 2D image + audio	*VAM*
2D video + audio + 3D image	MPI
2D video + body movement	FABO, **RU-FACS**, **BINED**, **DynEmo**
2D video + body movement + audio	GEMEP, **EmoTABOO**
Motion capture + audio	**IEMOCAP**
2D video + physio. sig.	**Smile Database**, **ENTERFACE**, **DEAP**
2D video + audio + physio. sig.	**RECOLA**
2D video + body movement + physio. sig. + audio	**MAHNOB-HCI**
2D video + 3D video + physio. sig.	**BioVid Emo**
3D image	BU-3D FE, Bosphorus, ICT-3DRFE
2D video + 2D image + 3D image	**PICS - Stirling ESRC 3D Face Database**
3D video	BU-4D FE, D3DFACS, **BP4D-Spontaneous**
3D video + audio	B3D(AC)2
2D video + 3D video + audio	**CAM3D**

ler or equal to 20: FABO, DISFA+, MUG, DISFA, AM-FED. The following databases have a FPS greater than 50: D3DFACS, MPI, IEMOCAP, MAHNOB-HCI. IEMOCAP has the greater FPS available so far (equals to 120), which makes it an interesting database to study the dynamics of spontaneous expressions.

2.4 Experimental Conditions

Experimental conditions include the background (EC.1) and lightning condition (EC.2) as well as head pose variation (EC.3) and occlusions (EC.4). Background and lightning conditions are about environment variability, whereas head pose variation and occlusions are about intra-personal variability. These characteristics are important to take into account if one wants to test the robustness of a method in real life conditions.

Most databases are acquired in the laboratory with a plain background (EC.1) and uniform or ambiant lightning (EC.2). In this case, face detection and facial landmarks tracking are eased. The background may not be plain, then it remains the same. In-the-wild databases propose to tackle this problem since they consist in video or audio-visual corpus or crowd sourcing (see subsection 2.5.3), offering high variability in background and lightning condition.

Only a few databases of posed and spontaneous expressions propose several lightning conditions (EC.2). Three lightning conditions are available in OULU-CASIA (normal, weak and dark) and NVIE (front, left and right), whereas Multi-PIE proposes 19 lightning conditions. ICT-3DRFE goes further with a static 3D model of the face that is relightable thanks to a light stage with 156 LEDs.

Five kinds of occlusions (EC.3) are considered: when the subject wears glasses, hair on face, data acquisition hardware, hands in front of face and others. Occlusions with data acquisition hardware may affect databases with physiological signals (Smile Database and ENTERFACE) or audio (RECOLA). Table 4 reports the classification of the databases according to occlusions (EC.3). As expected, very few posed databases contain occlusions, whereas in-the-wild databases cover a large range of occlusions.

There are three ways to obtain head pose variation (EC.4). First, multi-view acquisition (see subsection 2.3) consists in acquiring the face simultaneously from several views. In the existing databases,

Table 4: Classification of the databases according to the characteristic EC.3 (occlusions). The following formatting distinguishes databases: no particular formatting for posed databases, bold for spontaneous databases and italic for in-the-wild databases. If the database contains several kinds of occlusions, it appears in each corresponding row.

EC.3	Databases
Glasses	MMI, Multi-PIE, Bosphorus, DISFA+, **Smile Database**, **EmoTABOO**, **MMI+**, **MUG**, **NVIE**, **CAM3D**, **DEAP**, **PICS - Stirling ESRC 3D Face Database**, **DynEmo**, **RECOLA**, *EmoTV*, *AFEW*, *SFEW*, *AM-FED*, *Aff-Wild*, *Vinereactor*
Hair on face	Bosphorus, **MUG**, **CAM3D**, **PICS - Stirling ESRC 3D Face Database**, **RECOLA**, *EmoTV*, *AFEW*, *SFEW*, *AM-FED*, *Aff-Wild*, *Vinereactor*
Data acquisition hardware	**Smile Database**, **ENTERFACE**, **RECOLA**
Hands	Bosphorus, **CAM3D**, **RECOLA**, **GFT**, *Aff-Wild*
Others	*AFEW*, *SFEW*, *HAPPEI*, *Aff-Wild*

we can find variation for the yaw angle, the pitch angle or the combination of yaw and pitch. Second, 3D databases allow to generate 2D face with any pose. Third, there are databases where the subject can freely move her head, hence a natural head pose variation. Table 5 reports the classification of the databases according to head pose variation (EC.4). Multi-view acquisition is split in variation of yaw, pitch and combination of both. Most of the posed databases contain yaw variation or 3D model, whereas most of the spontaneous and in-the-wild databases contain natural variation. 3D databases are ideal to investigate facial expression analysis robust to head pose variation. Yaw variation is a good alternative and easier to set up. Multi-PIE and Bosphorus are the only database containing pitch variation combined with yaw variation.

2.5 Experimental Protocol

Experimental protocol describes the expressive/emotional content of the database (available expressions, EP.2) and the way it is obtained from the subjects (method of acquisition, EP.1). As mentioned in the introduction, we can distinguish 3 kinds of databases: posed, spontaneous and in-the-wild. The experimental protocol varies from one kind to another, so we discuss each kind of database in a separate subsection.

Table 5: Classification of the databases according to the characteristic EC.4 (head pose variation). The following formatting distinguishes databases: no particular formatting for posed databases, bold for spontaneous databases and italic for in-the-wild databases. For yaw and pitch variations, the number of poses (including frontal) is in brackets.

EC.4	Databases
Yaw	KDEF (5), PICS - Pain Expressions (3), MMI (2), Multi-PIE (13), Bosphorus (7), Radboud Faces (5), ADFES (2, continuous), MPI (3), **UT-Dallas (9)**, **BioVid Emo (3)**, **BAUM-1 (2)**
Pitch	Bosphorus (4)
Yaw + pitch	Multi-PIE (2), Bosphorus (2)
3D	BU-3D FE, BU-4D FE, D3DFACS, ICT-3DRFE, B(3D)AC2, **BP4D-Spontaneous**
Natural	University Of Maryland, **Smile Database**, **RU-FACS**, **SAL**, **EmoTABOO**, **BINED**, **IEMOCAP**, **MMI+**, **SEMAINE**, **CAM3D**, **UNBC-McMaster Shoulder Pain Expression Archive**, **DISFA**, **PICS - Stirling ESRC 3D Face Database**, **DynEmo**, **RECOLA**, **GFT**, *Befast Naturalistic*, *EmoTV*, *VAM*, *AFEW*, *SFEW*, *AM-FED*, *HAPPEI*, *Aff-Wild*, *Vinereactor*, *CHEAVD*

2.5.1 Posed Expressions

Posed expressions are deliberately displayed by the subject by reproducing specific facial deformations. There are three methods of reproduction (EP.1): free reproduction, ordered reproduction and portrayal.

With free reproduction, the subject is just informed about the emotion to reproduce and must do it in an expressive manner with no other instruction. The following databases use this method: University Of Maryland, JAFFE, ICT-3DRFE, FABO, BU-3D FE. With ordered reproduction, either the subject is trained beforehand to reproduce the expressions, or she is in the presence of an expert who gives her an order during the acquisition. Compared to free reproduction, here the subject is guided. The following databases use this method: KDEF, CK, MMI, Multi-PIE, BU-4D FE, Bosphorus, OULU-CASIA, Radboud Faces, MUG, NVIE. With portrayal of the emotion, the subject must improvise on an emotionaly rich scenario. This is an interesting alternative to get more realistic posed expressions. Moreover these databases often include professional actors as subjects. The following databases use this method: GEMEP, IEMOCAP, MPI, B3D(AC)2, BAUM-1. Some spontaneous databases also contain posed expressions, that is why we have included them (IEMOCAP, MUG, NVIE, PICS - Stirling ESRC 3D Face Database, BAUM-1).

Regarding the available expressions (EP.2), the 6 prototypic expressions corresponding to the 6 basic emotions (Ekman and Friesen, 1971) (anger, disgust, fear, joy, sadness, surprise) are always included in the posed databases, possibly with other expressions. A few databases just include a subset of these 6 expressions along with secondary emotions or non emotional expressions: B3D(AC)2, IEMOCAP, Multi-PIE and BAUM-1. Secondary emotions, also known as mental states, correspond to non basic emotions such as frustration, shame, anxiety. The classification of the posed databases according to the available expressions (EP.2) is reported in table 6. We precise if the databases only contain the 6 basic emotions (if neutral is added, it gives 7 expressions), if they include contempt, secondary emotions, pain expressions, combinations of action units (AUs) or non emotional expressions. AUs are used in the FACS system (Ekman and Friesen, 1977) to describe local activation of facial muscles that produce facial expressions.

2.5.2 Spontaneous Expressions

Spontaneous expressions are expressions that occur naturally and that are not controlled by the subject, contrary to posed ones. Basically, there are two acquisition methods (EP.1) to obtain spontaneous expressions: emotion elicitation methods that are used to induce a specific emotional state and interaction between two protagonists in order to get emotionaly rich content. The setup of emotion elicitation methods is not without difficulty (Sneddon et al., 2012). It is impossible to know objectively what emotion is felt by the subject, how it is perceived by a third party and how much the facial expression reflects it. The more spontaneous the expressions are, the less easy they can be captured, the less information is available about the emotional state and the less the experimental protocol is reproductible. In contrast, the acquisititon of posed expressions allows to perfectly control the reproductibility of the experimental protocol, but it does not give any information on the genuine emotional state. The idea of emotion elicitation methods is to find a compromise by controlling the experimental protocol thanks to relatively standardized tasks that collect information about the emotional state while allowing the subject to react naturally to the task (Sneddon et al., 2012).

There are 2 emotion elicitation methods: passive tasks and active tasks. Passive tasks consist in watching videos or images that are intended to induce specific emotions. In the case of DEAP database, the subject watches musical clips that intend to span the 4 quadrants of the arousal/valence emotional space (Russell and Pratt, 1980) instead of specific emotions. Active tasks were popularized by BINED database. By definition, active tasks are designed to directly involve the subject and induce specific emotions. An example of active task inducing disgust is to ask the subject to put his non-dominant hand in a box containing cold, cooked and cut spaghetti in sauce, while the subject cannot see what is inside (Sneddon et al., 2012). In the case of AVEC 2013 AViD-Corpus, the active tasks are not meant to induce specific emotions. Some databases combine active and passive tasks.

There are 2 methods of interaction: human-human interaction and human-computer interaction. In human-human interaction, one of the subject may be aware of the protocol and thus seeks to manage the interaction in order to make it emotionaly rich (RU-FACS, EmoTABOO), or both subjects have to interact naturally in a precise context (RECOLA). GFT database extends the latter case to 3 subjects interacting together. In human-computer interaction, the subject interacts with a virtual agent remotely monitored by the experimenter; this is the wizard-of-Oz setup. The experimenter can choose several characters for the virtual agent and thus influence the emotional content of the interaction.

Table 7 reports the classification of the spontaneous databases according to the acquisition method (EP.1) and the available expressions (EP.2). The available expressions (EP.2) are often the same as with posed databases (basic and secondary emotions, see table 6). Most of the spontaneous databases use passive tasks as it is the easiest protocol to set up.

2.5.3 In-The-Wild Expressions

In-the-wild conditions refer to an unconstrained environment in terms of population (see subsection 2.1) and experimental conditions (see subsection 2.4), as we can meet in real life context. There are 3 methods of acquisition (EP.1) to obtain in-the-wild expressions: corpus of videos/images of posed expressions, corpus of videos/images of spontaneous expressions and crowd sourcing.

The first databases that begin to meet the criteria of in-the-wild conditions date back to the 2000s: Belfast Naturalistic, EmoTV and VAM. They consist in a corpus of video of spontaneous expressions with extracts from television programs. Thus the spontaneous expressions result from human-human inte-

Table 6: Classification of the databases of posed expressions according to the characteristic EP.2 (available expressions). Neutral face is included in the number of expressions.

# of expressions	Only 6 basic emotions	Incl. contempt	Incl. AUs	Incl. secondary emotions	Incl. pain expressions	Incl. non emotional expressions
5	-	-	IEMOCAP	-	-	-
6	University Of Maryland, BU-4D FE, OULU-CASIA, NVIE	-	-	-	-	Multi-PIE
7	JAFFE, KDEF, BU-3D FE, MUG, PICS - Stirling ESRC 3D Face Database	-	-	-	-	-
$\in (8,25)$	-	Radboud Faces, ADFES	GEMEP, FABO, ADFES	PICS - Pain Expressions	CK	ICT-3DRFE
≥ 34	-	MPI	B3D(AC)2, MPI, BAUM-1	-	MMI, Bosphorus, D3DFACS, DISFA+	-

raction. We consider them as in-the-wild databases since the emotional content is totally uncontrolled, as well as the experimental conditions, but they lack of variability in population. More recently, in Aff-Wild, videos from Youtube are extracted. The selected videos show a person who displays spontaneous expressions by watching a video, practicing an activity or reacting to a joke or surprise. This database also contains images from Google Image. The advantage of this database is to provide a wide variability in population (500+ for videos and 2000+ for images) and experimental conditions.

Corpus of videos or images of posed expressions appeared with AFEW. The selection of videos is made automatically among 54 movies by analyzing subtitles for deaf and hard of hearing, which contain, among others things, information about emotional context of actors. SFEW is a static version of AFEW containing images extracted from the latter. The advantage is that they provide a high variability in population (330 subjects ranging from 1 to 70 years old) and experimental conditions. However, the available expressions remain the 6 basic emotions and the expressions are posed. More recently, CHE-AVD combines a corpus of videos of posed expressions from movies and television series and a corpus of videos of spontaneous expressions from television programms. HAPPEI also combines a corpus of images of posed and spontaneous expressions. The images are selected on Flickr and contain a group of people (2 or more subjects) showing different levels of happiness (from neutral to thrilled). We consider that this database contains both posed and spontaneous expressions because it is likely that on pictures taken in social events, people may be posing.

In order to provide spontaneous expressions with a high variability in population and experimental conditions, crowd sourcing has been investigated. The principle is to recruit subjects through the Internet for a study and to film them directly at home via their webcam. To the best of our knowledge, only two databases use this method: AM-FED and Vinereactor. In both cases, the subject watches an inducing video and her reaction is recorded. In AM-FED, only the smile is induced. Similarly, in Vinereactor, the induced emotions appear to be related only to amusement since the subjects fill out a questionnaire to note how much the induction video amused her.

2.6 Annotations

The annotations are meta-data provided with the database that give low-level information (facial features A.1 or action units A.2) or high-level (emotional labels A.3 or emotional dimensions A.4). The choice of annotations depends on the problem the database is meant to tackle since they will be used as ground truth. Emotional labels are aimed at facial expression recognition, action units annotations are aimed at action units recognition and emotional dimensions are aimed at emotional dimension estimation (such as arousal or valence (Russell and Pratt, 1980)). Facial features (e.g. facial landmarks, LBP, ...) could make a database attractive since they may be used to design quickly a system without computing them. Table 8 reports the classification of the databases according the annotations (A.1, A.2, A.3, A.4).

Table 7: Classification of the databases of spontaneous expressions according to the characteristics EP.1 (acquisition method) and EP.2 (available expressions). The row 'Various' (available expressions) refer to databases where the acquisition methods do not intend to induce specific emotions. "HHI" and "HCI" refer to human-human interaction and human-computer interaction respectively.

EP.2	EP.1				
	Emotion elicitation methods			Interaction	
	Passive tasks	Active tasks	Passive + active tasks	HHI	HCI
Basic emotions	ENTERFACE, MMI+, MUG, NVIE, CAS(ME)2	-	-	-	-
Basic and secondary emotions	UT-Dallas, MAHNOB-HCI, BioVid Emo, BAUM-1	CAM3D	BINED, PICS - Stirling ESRC 3D Face Database, DynEmo, BP4D-Spontaneous	Emo-TABOO, IEMOCAP, CAM3D	-
Various	DEAP	AVEC 2013 AViD-Corpus	-	RU-FACS, RECOLA, GFT	SAL, SE-MAINE
Smile	Smile Database	-	-	CK+	-
AUs	DISFA	-	-	-	-
Pain expressions	-	UNBC-McMaster Shoulder Pain Expression Archive, BP4D-Spontaneous	-	-	-

3 FUTURE DIRECTIONS

We identify three future directions for facial expression databases that have already been addressed by only very few databases.

During the 5 last years, in-the-wild databases have known a growing interest as it has been made clear that facial expression analysis in an unconstrained environment is one of the main current challenge (Martinez and Valstar, 2016). We have seen in the subsubsection 2.5.3 that 3 methods of acquisition of in-the-wild expressions exist so far: corpus of videos/images of posed expressions, corpus of videos/images of spontaneous expressions and crowd sourcing. Crowd sourcing seems to be a promising method to acquire realistic data with a high variability in population and experimental conditions. To the best of our knowledge, until now only positive emotions have been acquired with this method in the databases AM-FED (McDuff et al., 2013) and Vinereactor (Kim and Vangala, 2016). Enhancing crowd sourcing by acquiring positive as well as negative emotions could be an interesting direction to explore in order to enrich the available in-the-wild expressions. Though this is challenging because of ethical concerns.

Another direction that has been explored recently is building a database with groups of people. To the best of our knowledge, only two databases propose such data. In HAPPEI (Dhall et al., 2015), the number of subjects varies but the database only contains different levels of happiness expressions. The purpose of the database is to study the happiness intensity of the group. In GFT (Girard et al., 2017), there are 3 subjects in the videos interacting naturally. The subjects are facing each other so the overall video does not contain 3 frontal faces. The purpose of this database is to study social interaction and the originality is to have 3 subjects instead of 2 as in RU-FACS (ruf, 2006) or RECOLA (Ringeval et al., 2013). So, we can consider two directions within databases of groups of people: images or videos with with a group of people facing the camera in order to estimate the overall emotion of the group or videos of a group interaction in order to study social interaction.

At last, we report very few databases with time lapse between acquisitions for each subject. To the best of our knowledge, this has been proposed only in 3 databases. In Smile Database (Schmidt and Cohn, 2001), only spontaneous smiles are acquired in two sessions recorded a year apart. In Multi-PIE (Gross

Table 8: Classification of the databases according to the annotation characteristics A.1 (facial features), A.2 (AUs), A.3 (emotional labels) and A.4 (emotional dimensions). The following formatting distinguishes databases: no particular formatting for posed databases, bold for spontaneous databases and italic for in-the-wild databases

Annotation	Databases
A.1 (facial features)	BU-3D FE, BU-4D FE, Bosphorus, B3D(AC)2, **Smile Database**, **IEMOCAP**, **CK+**, **MUG**, **NVIE**, **UNBC-McMaster Shoulder Pain Expression Archive**, **DISFA**, **AVEC 2013 AViD-Corpus**, **BP4D-Spontaneous**, **GFT**, *AFEW*, *SFEW*, *AM-FED*, *Vinereactor*
A.2 (AUs)	CK, MMI, D3DFACS, ICT-3DRFE, DISFA+, **Smile Database**, **RU-FACS**, **CK+**, **MMI+**, **UNBC-McMaster Shoulder Pain Expression Archive**, **DISFA**, **CAS(ME)2**, **BP4D-Spontaneous**, **GFT**, *AM-FED*, *Aff-Wild*, *Vinereactor*
A.3 (Emotional labels)	JAFFE, CK, FABO, Radboud Faces, **UT-Dallas**, **EmoTABOO**, **BINED**, **IEMOCAP**, **CK+**, **SEMAINE**, **NVIE**, **CAM3D**, **DynEmo**, **CAS(ME)2**, **BioVid Emo**, **BAUM-1**, *Belfast Naturalistic*, *EmoTV*, *VAM*, *AFEW*, *SFEW*, *AM-FED*, *HAPPEI*, *CHEAVD*
A.4 (Emotional dimensions)	GEMEP, Radboud Faces, **SAL**, **EmoTABOO**, **BINED**, **IEMOCAP**, **SEMAINE**, **NVIE**, **UNBC-McMaster Shoulder Pain Expression Archive**, **MAHNOB-HCI**, **DEAP**, **DynEmo**, **RECOLA**, **AVEC 2013 AViD-Corpus**, *Belfast Naturalistic*, *EmoTV*, *VAM*, *AM-FED*, *Aff-Wild*

et al., 2010), 4 acquisitions of 5 expressions were done over the course of 6 months. In AVEC 2013 AViD-Corpus (Valstar et al., 2013), there are between 1 and 4 acquisitions for each subject recorded two weeks apart. It could be interesting to go further in this direction in order to study the stability of facial expression reaction to a particular event or the variation of behavior over time. This kind of problems could be of great interest for human-computer interaction or medical application for monitoring an individual's emotional state.

4 CONCLUSIONS

In this paper, we presented a survey of facial expression databases. We identified 18 characteristics to describe a database and we grouped them in 6 categories. We reviewed each characteristic and brought to light the differences between posed, spontaneous and in-the-wild databases. We finished the paper with the future directions: enhancing crowd sourcing to build in-the-wild databases with a greater variety of expressions, building databases with groups of people and building databases with time lapse between acquisition for each subject.

REFERENCES

(2006). http://mplab.ucsd.edu/grants/project1/research/rufacs1-dataset.html.

(2009). http://www.cse.oulu.fi/CMV/Downloads/Oulu-CASIA.

(2013). http://pics.stir.ac.uk.

Abrilian, S., Devillers, L., Buisine, S., and Martin, J.-C. (2005). Emotv1: Annotation of real-life emotions for the specification of multimodal affective interfaces. In *HCI International*.

Aifanti, N., Papachristou, C., and Delopoulos, A. (2010). The mug facial expression database. In *Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 11th International Workshop on*, pages 1–4. IEEE.

Anitha, C., Venkatesha, M., and Adiga, B. S. (2010). A survey on facial expression databases. *International Journal of Engineering Science and Technology*, 2(10):5158–5174.

Bänziger, T., Pirker, H., and Scherer, K. (2006). Gemep-geneva multimodal emotion portrayals: A corpus for the study of multimodal emotional expressions. In *Proceedings of LREC*, volume 6, pages 15–019.

Black, M. J. and Yacoob, Y. (1997). Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Journal of Computer Vision*, 25(1):23–48.

Busso, C., Bulut, M., Lee, C.-C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J. N., Lee, S., and Narayanan, S. S. (2008). Iemocap: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42(4):335–359.

Cohn, J. F. and Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(02):121–132.

Cosker, D., Krumhuber, E., and Hilton, A. (2011). 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*, pages 2296–2303. IEEE.

Cowie, R., Douglas-Cowie, E., and Cox, C. (2005). Beyond emotion archetypes: Databases for emotion modelling using neural networks. *Neural networks*, 18(4):371–388.

Dhall, A., Goecke, R., and Gedeon, T. (2015). Automatic group happiness intensity analysis. *IEEE Transactions on Affective Computing*, 6(1):13–26.

Dhall, A., Goecke, R., Lucey, S., and Gedeon, T. (2011). Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 2106–2112. IEEE.

Dhall, A., Goecke, R., Lucey, S., and Gedeon, T. (2012). Collecting large, richly annotated facial-expression databases from movies.

Douglas-Cowie, E., Cowie, R., Cox, C., Amier, N., and Heylen, D. (2008). The sensitive artificial listner: an induction technique for generating emotionally coloured conversation.

Douglas-Cowie, E., Cowie, R., and Schröder, M. (2000). A new emotion database: considerations, sources and scope. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*.

Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124.

Ekman, P. and Friesen, W. V. (1977). Facial action coding system.

Fanelli, G., Gall, J., Romsdorfer, H., Weise, T., and Van Gool, L. (2010). A 3-d audio-visual corpus of affective communication. *Multimedia, IEEE Transactions on*, 12(6):591–598.

Fu, S., Yang, G., Kuai, X., and Zheng, R. (2012). *A Parametric Survey for Facial Expression Database*, pages 373–381. Springer Berlin Heidelberg, Berlin, Heidelberg.

Girard, J. M., Chu, W.-S., Jeni, L. A., Cohn, J. F., and De la Torre, F. (2017). Sayette group formation task (gft) spontaneous facial expression database.

Grimm, M., Kroschel, K., and Narayanan, S. (2008). The vera am mittag german audio-visual emotional speech database. In *Multimedia and Expo, 2008 IEEE International Conference on*, pages 865–868. IEEE.

Gross, R., Matthews, I., Cohn, J., Kanade, T., and Baker, S. (2010). Multi-pie. *Image and Vision Computing*, 28(5):807–813.

Gunes, H. and Piccardi, M. (2006). A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 1148–1153. IEEE.

Kanade, T., Cohn, J. F., and Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 46–53. IEEE.

Kaulard, K., Cunningham, D. W., Bülthoff, H. H., and Wallraven, C. (2012). The mpi facial expression database a validated database of emotional and conversational facial expressions. *PloS one*, 7(3):e32321.

Kim, E. and Vangala, S. (2016). Vinereactor: Crowdsourced spontaneous facial expression data. In *International Conference on Multimedia Retrieval (ICMR)*. IEEE.

Koelstra, S., Mühl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., and Patras, I. (2012). Deap: A database for emotion analysis; using physiological signals. *Affective Computing, IEEE Transactions on*, 3(1):18–31.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and emotion*, 24(8):1377–1388.

Li, Y., Tao, J., Chao, L., Bao, W., and Liu, Y. (2016). Cheavd: a chinese natural emotional audio–visual database. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–12.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 94–101. IEEE.

Lucey, P., Cohn, J. F., Prkachin, K. M., Solomon, P. E., and Matthews, I. (2011). Painful data: The unbc-mcmaster shoulder pain expression archive database. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 57–64. IEEE.

Lundqvist, D., Flykt, A., and hman, A. (1998). The karolinska directed emotional faces - kdef, cd rom from department of clinical neuroscience, psychology section, karolinska institutet.

Lyons, M., Akamatsu, S., Kamachi, M., and Gyoba, J. (1998). Coding facial expressions with gabor wavelets. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 200–205. IEEE.

Mahmoud, M., Baltrušaitis, T., Robinson, P., and Riek, L. D. (2011). 3d corpus of spontaneous complex mental states. In *International Conference on Affective Computing and Intelligent Interaction*, pages 205–214. Springer.

Martinez, B. and Valstar, M. F. (2016). Advances, challenges, and opportunities in automatic facial expression recognition. In *Advances in Face Detection and Facial Image Analysis*, pages 63–100. Springer.

Mavadati, M., Sanger, P., and Mahoor, M. H. (2016). Extended disfa dataset: Investigating posed and spontaneous facial expressions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8.

Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., and Cohn, J. F. (2013). Disfa: A spontaneous facial action intensity database. *Affective Computing, IEEE Transactions on*, 4(2):151–160.

McDuff, D., El Kaliouby, R., Senechal, T., Amr, M., Cohn, J. F., and Picard, R. (2013). Affectiva-mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions collected" in-the-wild". In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 881–888. IEEE.

McKeown, G., Valstar, M. F., Cowie, R., and Pantic, M. (2010). The semaine corpus of emotionally coloured character interactions. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pages 1079–1084. IEEE.

Pantic, M., Valstar, M., Rademaker, R., and Maat, L. (2005). Web-based database for facial expression analysis. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 5–pp. IEEE.

Qu, F., Wang, S.-J., Yan, W.-J., and Fu, X. (2016). Cas(me)2: A database of spontaneous macro-expressions and micro-expressions. In *International Conference on Human-Computer Interaction*, pages 48–59. Springer.

Ringeval, F., Sonderegger, A., Sauer, J., and Lalanne, D. (2013). Introducing the RECOLA Multimodal Corpus of Remote Collaborative and Affective Interactions. In *Proceedings of EmoSPACE 2013, held in conjunction with FG 2013*, Shanghai, China. IEEE.

Russell, J. A. and Pratt, G. (1980). A description of the affective quality attributed to environments. *Journal of personality and social psychology*, 38(2):311.

Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., and Akarun, L. (2008). Bosphorus database for 3d face analysis. In *European Workshop on Biometrics and Identity Management*, pages 47–56. Springer.

Savran, A., Ciftci, K., Chanel, G., Mota, J., Hong Viet, L., Sankur, B., Akarun, L., Caplier, A., and Rombaut, M. (2006). Emotion detection in the loop from brain signals and facial images.

Schmidt, K. L., Ambadar, Z., Cohn, J. F., and Reed, L. I. (2006). Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling. *Journal of Nonverbal Behavior*, 30(1):37–52.

Schmidt, K. L. and Cohn, J. F. (2001). Dynamics of facial expression: Normative characteristics and individual differences. In *ICME*. Citeseer.

Sneddon, I., McRorie, M., McKeown, G., and Hanratty, J. (2012). The belfast induced natural emotion database. *Affective Computing, IEEE Transactions on*, 3(1):32–41.

Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *Affective Computing, IEEE Transactions on*, 3(1):42–55.

Stratou, G., Ghosh, A., Debevec, P., and Morency, L.-P. (2011). Effect of illumination on automatic expression recognition: a novel 3d relightable facial database. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 611–618. IEEE.

Tcherkassof, A., Dupré, D., Meillon, B., Mandran, N., Dubois, M., and Adam, J.-M. (2013). Dynemo: A video database of natural facial expressions of emotions. *The International Journal of Multimedia & Its Applications*, 5(5):61–80.

Toole, A. J., Harms, J., Snow, S. L., Hurst, D. R., Pappas, M. R., Ayyad, J. H., and Abdi, H. (2005). A video database of moving faces and people. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(5):812–816.

Valstar, M. and Pantic, M. (2010). Induced disgust, happiness and surprise: an addition to the mmi facial expression database. In *Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect*, page 65.

Valstar, M., Schuller, B., Smith, K., Eyben, F., Jiang, B., Bilakhia, S., Schnieder, S., Cowie, R., and Pantic, M. (2013). Avec 2013: the continuous audio/visual emotion and depression recognition challenge. In *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, pages 3–10. ACM.

Valstar, M. F., Gunes, H., and Pantic, M. (2007). How to distinguish posed from spontaneous smiles using geometric features. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 38–45. ACM.

Van Der Schalk, J., Hawk, S. T., Fischer, A. H., and Doosje, B. (2011). Moving faces, looking places: validation of the amsterdam dynamic facial expression set (adfes). *Emotion*, 11(4):907.

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

Yin, L., Chen, X., Sun, Y., Worm, T., and Reale, M. (2008). A high-resolution 3d dynamic facial expression database. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference On*, pages 1–6. IEEE.

Yin, L., Wei, X., Sun, Y., Wang, J., and Rosato, M. J. (2006). A 3d facial expression database for facial behavior research. In *Automatic face and gesture recognition, 2006. FGR 2006. 7th international conference on*, pages 211–216. IEEE.

Zafeiriou, S., Papaioannou, A., Kotsia, I., Nicolaou, M. A., Zhao, G., Antonakos, E., Snape, P., Trigeorgis, G., and Zafeiriou, S. (2016). Facial affect in-the-wild: A survey and a new database. In *International Conference on Computer Vision*.

Zara, A., Maffiolo, V., Martin, J. C., and Devillers, L. (2007). Collection and annotation of a corpus of human-human multimodal interactions: Emotion and others anthropomorphic characteristics. In *Affective computing and intelligent interaction*, pages 464–475. Springer.

Zeng, Z., Pantic, M., Roisman, G., Huang, T. S., et al. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):39–58.

Zhalehpour, S., Onder, O., Akhtar, Z., and Erdem, C. E. (2016). Baum-1: A spontaneous audio-visual face database of affective and mental states. *IEEE Transactions on Affective Computing*.

Zhang, L., Walter, S., Ma, X., Werner, P., Al-Hamadi, A., Traue, H. C., and Gruss, S. (2016). biovid emo db: A multimodal database for emotion analyses validated by subjective ratings. In *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on*, pages 1–6. IEEE.

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., Liu, P., and Girard, J. M. (2014). Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. *Image and Vision Computing*, 32(10):692–706.

APPENDIX

For a purpose of clarity, the references are not included in section 2. We report here the corresponding references to all the databases we review.

Table 9: References of posed databases.

Database	Reference
ADFES	(Van Der Schalk et al., 2011)
B3D(AC)2	(Fanelli et al., 2010)
BAUM-1	(Zhalehpour et al., 2016)
Bosphorus	(Savran et al., 2008)
BU-3D FE	(Yin et al., 2006)
BU-4D FE	(Yin et al., 2008)
CK	(Kanade et al., 2000)
D3DFACS	(Cosker et al., 2011)
DISFA+	(Mavadati et al., 2016)
FABO	(Gunes and Piccardi, 2006)
GEMEP	(Bänziger et al., 2006)
ICT-3DRFE	(Stratou et al., 2011)
IEMOCAP	(Busso et al., 2008)
JAFFE	(Lyons et al., 1998)
KDEF	(Lundqvist et al., 1998)
MMI	(Pantic et al., 2005)
MPI	(Kaulard et al., 2012)
MUG	(Aifanti et al., 2010)
Multi-PIE	(Gross et al., 2010)
NVIE	(Wang et al., 2010)
OULU-CASIA	(oul, 2009)
PICS - Pain Expressions	(han, 2013)
Radboud Faces	(Langner et al., 2010)
University Of Maryland	(Black and Yacoob, 1997)

Table 10: References of spontaneous databases.

Database	Reference
AVEC 2013	(Valstar et al., 2013)
AViD-Corpus	
BAUM-1	(Zhalehpour et al., 2016)
BINED	(Sneddon et al., 2012)
BioVid Emo	(Zhang et al., 2016)
BP4D-Spontaneous	(Zhang et al., 2014)
CAM3D	(Mahmoud et al., 2011)
CAS(ME)2	(Qu et al., 2016)
CK+	(Lucey et al., 2010)
DEAP	(Koelstra et al., 2012)
DISFA	(Mavadati et al., 2013)
DynEmo	(Tcherkassof et al., 2013)
EmoTABOO	(Zara et al., 2007)
ENTERFACE	(Savran et al., 2006)
GFT	(Girard et al., 2017)
IEMOCAP	(Busso et al., 2008)
MAHNOB-HCI	(Soleymani et al., 2012)
MMI+	(Valstar and Pantic, 2010)
MUG	(Aifanti et al., 2010)
NVIE	(Wang et al., 2010)
PICS - Stirling ESRC 3D Face Database	(han, 2013)
RECOLA	(Ringeval et al., 2013)
RU-FACS	(ruf, 2006)
SAL	(Douglas-Cowie et al., 2008)
SEMAINE	(McKeown et al., 2010)
Smile Database	(Schmidt and Cohn, 2001)
UNBC-McMaster Shoulder Pain Expression Archive	(Lucey et al., 2011)
UT-Dallas	(Toole et al., 2005)

Table 11: References of in-the-wild databases.

Database	Reference
AFEW	(Dhall et al., 2012)
Aff-Wild	(Zafeiriou et al., 2016)
AM-FED	(McDuff et al., 2013)
Belfast Naturalistic	(Douglas-Cowie et al., 2000)
CHEAVD	(Li et al., 2016)
EmoTV	(Abrilian et al., 2005)
HAPPEI	(Dhall et al., 2015)
SFEW	(Dhall et al., 2011)
VAM	(Grimm et al., 2008)
Vinereactor	(Kim and Vangala, 2016)