A Survey on Databases for Facial Micro-Expression Analysis

Jingting Li, Catherine Soladie and Renaud Seguier
FAST Research Team, CENTRALESUPELEC/IETR, Rennes, France

Keywords: Micro-Expression, Database, Result Evaluation, Survey.

Abstract: Micro-expression (ME) is a brief local spontaneous facial expression and an important non-verbal clue to revealing genuine emotion. The study on automatic detection and recognition of ME has been emerging in the last decade. However, the research is restricted by the number of ME databases. In this paper, we propose a survey based on the 15 existing ME databases. Firstly, the databases are analyzed by 13 characteristics grouped into four categories (population, hardware, experimental protocol, and annotation). These characteristics provide a reference not only for choosing a database for special ME analysis purpose but also for future database construction. Concerning the ME analysis based on databases, we firstly present the emotion classification and metric frequency for ME recognition. The most frequently used databases for ME detection are then introduced. Finally, we discuss the future directions of micro-expression databases.

1 INTRODUCTION

Micro-expression (ME) is a brief local spontaneous facial expression (Ekman and Friesen, 1969), particularly appearing in the case of high pressure. The movement only lasts between 1/25s and 1/5s. ME is a very important non-verbal communication clue. Its involuntary nature can reveal the genuine emotion and the personal psychological states (Birdwhistell, 1968). Thus, ME analysis has many potential applications in national security (Ekman, 2009), medical care (Endres and Laidlaw, 2009), and etc.

MEs were discovered by Haggard and Isaacs (Haggard and Isaacs, 1966) and then named by Ekman and Friesen (Ekman and Friesen, 1969). Ekman developed a ME training tool: Micro Expressions Training Tools (METT) (Eckman, 2003). It has several visual samples which belong to the universal emotions and aims at training people to detect and interpret MEs. Yet, the overall recognition rate for the 6 basic emotions by naked eyes is lower than 50%, even by a trained expert (Frank et al., 2009).

The ME analysis includes recognition and detection / spotting (MEDR). ME detection is a broader term for identifying whether there is a ME in a video or not. In contrast, ME spotting means more specifically locating the frame index of ME in videos. In this paper, we use detection to represent both definitions. As illustrated in Figure 1, research on automatic ME analysis begins to emerge in recent decade. However, the paper amount is limited due to ME characteristics and ME databases. Methods should conduct experiments on databases to verify the performance. Besides, the results of different methods can be compared on the same chosen database. And the database features, e.g. population, image/video quality etc., would influence the result evaluation. Compared with macro-expression databases, there is still plenty of room to improve for ME database.

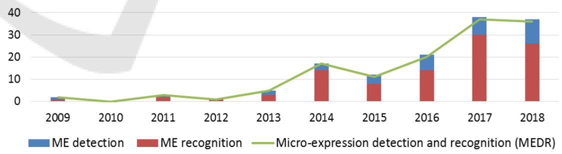

Figure 1: Micro-expression detection and recognition (MEDR) research trend. The histogram lists the small but increasing quantity of articles for MEDR research.

After 10 years of research, some surveys that aim at building guidelines for the further MEDR research have appeared. For instance, Oh et al. (Oh et al., 2018) summarized the databases, analysis methods and challenges. However, there is no systematic comparative analysis among the existing databases. In this paper, we propose a survey on ME databases. First of all, we have defined 4 categories: population, hardware, experimental protocol and annotation. It is inspired by Weber et al. (Weber et al., 2018), and more appropriate for ME databases description. Each ca-

Li, J., Soladie, C. and Seguier, R.
A Survey on Databases for Facial Micro-Expression Analysis.
DOI: 10.5220/0007309202410248
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 241-248
ISBN: 978-989-758-354-4

tegory contains several characteristics, and the databases are compared depending on them. This classification builds a reference for choosing databases for targeted research or for building a new ME database. Furthermore, the emotion classification and metric frequency are introduced for ME recognition based on databases. Meanwhile, the database frequency for ME detection is presented.

The article is organized as follows: section 2 presents an all-inclusive survey and comparison on ME databases. Section 3 reviews the applications of ME database and then discuss the future direction. Finally section 4 concludes the paper.

2 THE 13 CHARACTERISTICS OF MICRO-EXPRESSION DATABASES

To our knowledge, there are only 15 published ME databases. Since 2009, Canal9 (Vinciarelli et al., 2009), York-DDT (Warren et al., 2009), Polikovsky's database (PD) (Polikovsky, 2009) and USF-HD (Shreve et al., 2011) were published. Yet, these databases are not used nowadays. Canal9 and York-DTT do not dedicate to ME research. Meanwhile, the PD and USF-HD are posed ME databases. In the ensuing years, several spontaneous ME databases were created. Oulu University published SMIC-sub (Pfister et al., 2011) and SMIC (Li et al., 2013). Meantime, CASME I (Yan et al., 2013), CASME II (Yan et al., 2014) were created by Chinese Academy of Science. In 2015, Radlak et al. built a Silesian Deception Database (SDD) (Radlak et al., 2015), which provided video samples of deceivers and truth-tellers. Oulu University then published an extended version of SMIC: SMIC-E (Li et al., 2017) to provide video samples for ME detection. Afterwards, a spontaneous micro-facial movement dataset: SAMM (Davison et al., 2018b) is created. In 2017, a database which contains both macro and micro expression: CAS(ME)^2 (Qu et al., 2017) was published. Soon later, an in-the-wild database MEVIEW (Husák et al., 2017) was published. In addition, two databases: MobileDB and Grobova's database (GD) were mentioned in (He et al., 2017) and (Grobova et al., 2017) respectively but are not yet publicly available.

After a brief introduction of ME databases, we list 13 characteristics which can comprehensively represent the feature of ME databases. They are classified into 4 categories, as shown in Table 1. We give the trends and summarize information to facilitate the comparison among databases. Detailed information of databases can be found in Appendix. Databases in the following sections are named by their abbreviation, as cited in the above paragraph.

Table 1: Categories and characteristics of ME databases. Characteristics are coded to simplify further representation.

Category	Characteristic	Code
Population	# of subjects	P.1
	# of samples	P.2
	Gender (%)	P.3
	Age range	P.4
	Ethnic group(s)	P.5
Hardware	Modalities	H.1
	FPS(Frames per second)	H.2
	Resolution	H.3
Experimental protocol	Method of acquisition	EP.1
	Environment (Image/video quality)	EP.2
	Available expressions	EP.3
Annotations	Action Units	A.1
	Emotional labels	A.2

2.1 Population

This analysis of population focuses on the subject amount (P.1), the sample amount (P.2), the gender distribution (P.3), the age range (P.4) and the ethnic groups (P.5). MEs have a general variation pattern, but also differ for different subjects, due to the face shape, facial texture, their gender and the cultural influence. Thus, the population genericity is essential for improving the automatic ME analysis ability.

As shown in Table 2, most of ME databases contain less than 50 subjects (P.1). Moreover, the amount of ME samples (P.2) is not significant. Even the largest database CASME II does not exceed 255 samples, which make it difficult to train detection or recognition algorithms. This is because the ME samples are difficult to produce. It requires a strict recording environment and professional eliciting methods. Moreover, the annotation is time-consuming. Besides, even though MEs exist in our daily life, it is complicated to gather video samples and to identify the MEs precisely in the in-the-wild environment.

The women/man percentage (P.3) for ME databases is not well balanced. Canal9, CASME I, SMIC and MEVIEW contain much more male subjects than female, while the number of female subjects in York-DDT is almost two times the male subject amount. Yet, the percentage in the three most recent databases CASME II, SAMM and CAS(ME)^2 are well balanced between 40/60 and 60/40.

The age range (P.4) for most ME databases is quite low, since the majority of samples were produced by volunteers in university. The average age is around 25 years old and the standard deviation (std) is around

3. Yet, York-DDT has a moderate range (18-45), and the average age of SAMM is 33.24 with a large std (11.32). However, the age distribution is still far from the reality. A good database should also contain the samples gathered from children and elderly people.

For ME database, the ethnic groups (P.5) are not very diverse. China Academy of Science has built three databases, but there is only one Asian ethnic group. Meanwhile, SMIC has 3 ethnic groups: Caucasian, Asian and Africa, and PD have Caucasian, Asian and Indian groups. Furthermore, SAMM contains 13 ethnic groups, which makes it the most varied ME database in term of ethnic groups. A widely collected database is recommended for ME analysis in the real world. Yet, the construction of this kind of database may need the international cooperation.

Table 2: Classification of the databases according to the characteristic P.1, P.2, H.1, A.1 and A.2 (# of subjects, # of samples, modalities, action units and emotional labels). Databases are sorted by alphabetical order. The following formatting distinguishes databases: normal for posed databases, bold for spontaneous database, italic for in-the-wild databases, * means the database is not available online. 2D V: 2D video. SMIC and SMIC-E both have three sub-classes: NIR (N), VIS (V) and HS (H). Sub-class HS of SMIC / SMIC-E is separated from the other two because of the different number of ME video samples.

Databases	P.1	P.2	H.1	A.1	A.2
Canal9	$\in (200, 250)$	$\in (50, 100)$	2D V		
CASME I	≤ 50	$\in (100, 200)$	2D V	✓	✓
CASME II	≤ 50	$\in (200, 300)$	2D V	✓	✓
CAS(ME)^2	≤ 50	$\in (50, 100)$	2D V	✓	✓
GD*	≤ 50	$\in (50, 100)$	2D V		✓
MEVIEW	≤ 50	≤ 50	2D V	✓	✓
MDB*	≤ 50	$\in (200, 300)$	2D V		✓
PD	≤ 50	≤ 50	2D V	✓	
SAMM	≤ 50	$\in (100, 200)$	2D V	✓	✓
SDD	$\in (100, 200)$	$\in (100, 200)$	2D V		✓
SMIC-sub	≤ 50	$\in (50, 100)$	2D V		✓
SMIC-N, V	≤ 50	$\in (50, 100)$	2D V + IF		✓
SMIC-H	≤ 50	$\in (100, 200)$	2D V		✓
SMIC-E-N, V	≤ 50	$\in (50, 100)$	2D V + IF		✓
SMIC-E-H	≤ 50	$\in (100, 200)$	2D V		✓
USF-HD	≤ 50	$\in (100, 200)$	2D V		✓
York-DDT	$\in (50, 100)$	≤ 50	2D V		✓

2.2 Hardware

The first characteristic is modalities (H.1), i.e. the ME sample recorded format. Until now, as listed in Table 2, the modality for most ME databases is unified: an unimodal 2D video. However, SMIC and SMIC-E have three modalities: high speed (HS) video, normal visual (VIS) video and near-infrared (NIR) video. Multi-modalities (e.g. facial thermal variation from infrared images) can contribute to increasing the data-scale and enhance the reliability for ME analysis. Meanwhile, the synchronization should catch our attention. There is no audio, 3D model, or body movements. If the ME databases follow the same evolution as the macro-expression databases, we can imagine having more modalities in the future databases.

As the ME average duration is around 300ms (LI et al., 2018) and the ME usually appears on the local facial region, a high FPS (H.2) and a high resolution (H.3) will help to capture MEs. Most ME databases have at least 60 FPS with a facial resolution larger than 150×190. The FPS of PD, CASME II and SAMM reach to 200. The resolution of facial region in SAMM is 400×400. Samples in these databases were recorded by a high-speed camera in a strictly controlled laboratory environment to reduce the noise. Meanwhile, USF-HD, SMIC, CAS(ME)^2 and MEVIEW contain clips with low FPS, lower or equal to 30. These databases fit more the situation in real life. However, 30 FPS means that the video just contains 9 frames for ME (300ms), the data scale is small. Thus, this may make the ME analysis more complex and less reliable.

2.3 Experimental Protocol

The experimental protocol refers to the acquisition method (EP.1), experimental environment (image/video quality) (EP.2) and the available expressions (EP.3). As the protocols are quite different according to the type of database, we discuss them separately in the following paragraphs. Moreover, as image/video quality is a very important factor, it is specifically discussed in sub-section 2.3.1.

Posed Micro-Expressions. Posed ME means that the facial movement is expressed by a subject on purpose with simulated emotion. ME is challenging to produce because ME is a very brief and local facial movement. The ME sequences in PD, USF-HD and mobileDB are all reproduced by ordered reproduction (EP.1) (Weber et al., 2018). In the PD, volunteers were requested to perform 7 basic emotions slightly and quickly after trained by an expert. Subjects in USF-HD were demanded to mimic the MEs in sample video and the participants in mobileDB mimicked the expressions based on 6 basic MEs. Happiness, surprise, fear,sadness, disgust and anger, these 6 basic emotions (Ekman and Friesen, 1971) are regarded as available expression (EP.3) in these three databases. The PD has one more emotional content: contempt.

Spontaneous Micro-Expressions. Spontaneous ME is generated naturally by emotion affect. All the spontaneous ME database used passive task as the emotion elicitation method (EP.1). The most common method is the neutralization paradigm, i.e. asking participants to watch videos containing strong emotions and try to neutralize during the whole time or try to suppress facial expressions when they realized there is one. The samples in York-DDT and SDD were generated by lie generation activity. Moreover, there is another kind of ME, which is hidden behind other facial movements. It is called as Masked ME and it is more complicated than neutralized ME. We will discuss it in sub-section 3.2.

As already introduced for characteristic H.1, ME database modality is 2D video. The duration of video sequences is quite short: most videos are less than 10s. For ME recognition, most methods only use the frames between the onset and the offset. Yet, longer video, with sometimes several MEs, is better for ME spotting. SMIC-E provided a longer version of video samples in SMIC, The average length of raw videos in SAMM is 35.3s. In CAS(ME)^2, the longest video can reach to 148s.

Concerning available expressions (EP.3), there are two classification methods. One is respecting the 6 basic emotion classes, e.g. York-DDT, CASME I, CASME II and SAMM. The other one is classifying emotions into three or four classes: positive, negative, surprise and others, such as SMIC, SMIC-E and CAS(ME)^2. In addition, SDD, SAMM and CAS(ME)^2 consist of not only micro movements but also macro expressions.

In-The-Wild Micro-Expressions. In-the-wild ME means that the acquisition is not limited by population and experiment acquisition conditions (EP.1). There are only 2 in-the-wild ME databases: Canal9 and MEVIEW. They both consist of a corpus of videos of spontaneous expressions. Canal9 contains 70 political debates recorded by the Canal9 local station. ME can be found when the politicians try to conceal their real emotions. MEVIEW contains 31 video clips from poker games and TV interviews downloaded from the Internet. The poker game can help to trigger ME thanks to the stress and the need to hide emotions. The available expressions (EP.3) in these two databases are based on 6 basic emotions. It is a big challenge to analyze the MEs automatically since there area lot of other irrelevant movements.

2.3.1 Image/Video Quality

The experimental environment (image/video quality) (EP.2) contains the number of cameras, background, lighting condition and occlusions. This subsection dedicates to the discussion of this subject. It already exists various macro-expression databases which contain different image quality situations. Unfortunately, Video samples in the majority of published ME databases are recorded in a strictly controlled environment. The improvement of the latest published databases focuses more on population augmentation and video length rather than image quality.

For most ME databases, there is only one camera. Besides, to avoid unrelated movements, participants were required to stay still and face directly the camera. As a side note, the video samples in mobileDB were recorded by a mobile device, which could be used for daily emergency situations. As mentioned in sub-section 2.2, one exception is that SMIC has three cameras. The illumination condition is maintained to be stable. LED lights are commonly used, and in some cases, extra equipment is used to reduce the noise. E.g., light diffusers were placed around the lights to soften the light on the participant's faces in SAMM. CASME I has two different conditions: natural light and two LED lights. The background is normally white or gray. Concerning occlusions, almost all the databases contain subjects wearing glasses. However, other occlusions and the head pose variation are very rare. It is worth noting that image quality varies in MEVIEW because the camera is often zooming, as well as changing the angle and the scene. Furthermore, as most videos came from television programs, there are some body movements and head poses.

It is still challenging to accurately analyze ME in single viewing angle videos with few noises. Thus, the community has not paid sufficient attention to get various image quality situations. However, as it is an essential factor for macro-expression databases, we could expect its importance in future ME databases.

2.4 Annotations

Regarding of ME databases, low-level information: action units (A.1) and high-level information: emotional labels (A.2) are the two major annotations.

Facial Action Coding System (FACS) (Ekman and Friesen, 1978) is an essential tool for facial expression annotation. Indeed, the facial components of FACS, i.e. actions units (AUs), identify the local muscle movement, and the combination of AUs shows the emotional expression. Since ME is a local brief movement, identifying the AUs will help to facilitate the ME analysis. However, some databases were not labeled by AUs, e.g., USF-HD, SMIC and SMIC-E. Figure 2 shows a histogram for the sum of AU annotations in all the databases and lists the number of AUs annotation in ME databases. The highest AU amount represents the regions where has the most ME movements.

Davision et al. (Davison et al., 2018a) proposed an objective ME classification. The facial movements

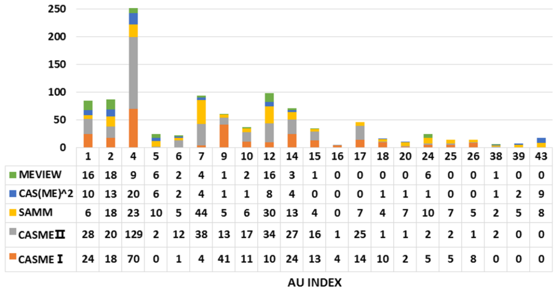

Figure 2: Histogram of action units (AUs) annotations for ME databases. The AU amount on the region of eyebrows (e.g. AU 1,2,4) and mouth (e.g. AU 12, 14) indicate that these two regions have the most frequent ME movement.

are labeled by AU combinations, and this would help to avoid the uncertainty cause by subjective annotation. In the Facial Micro-Expression Grand Challenge (MEGC) held by FG2018 (Yap et al., 2018), the ME classes for recognition are labeled by this objective classification.

Emotional labels (A.2) are used for ME recognition. As listed in Table 2, almost all the ME datasets have emotional labels except Canal9 and PD. As mentioned in sub-section 2.3, the emotion labels differ in different databases. However, there are two exceptions. One is the SDD, who contains micro-tension, eye closures and gaze aversion of subjects. Another one is Canal9, the samples are annotated by agree/disagree since the videos came from debate scenes. Until now, to our knowledge, there are no ME databases provide facial features or emotional dimensions. Table 3 and 4 show a quantitative summary for commonly used databases, and they list the emotion classes and the numbers of corresponding samples. SMIC-sub and CAS(ME)^2 are shown in both two tables. Because they have two types of emotional labels. In SMIC-sub, positive and happiness contain the same samples, negative is the ensemble of disgust and sad samples, fear and surprise samples are not included in 2 emotional classification. In addition, as the SAMM database is a micro facial movement dataset, there are also 26 video samples which are classified as emotion 'other'.

Table 3: Emotion classes and sample numbers for micro-expression databases - part 1.

Dataset	Positive	Surprise	Negative	Others
SMIC-sub	17	0	18	0
SMIC-HS	51	43	70	0
SMIC-VIS/NIR	28	20	23	0
SMIC-E-HS	51	42	71	0
CAS(ME)^2	8	9	21	19

Table 4: Emotion classes and sample numbers for micro-expression databases - part 2. H: Happiness, D: Disgust, Su: Surprise, R: Repression, T: Tense, F: Fear, C: Contempt, Sa: Sadness, He: helpless, Pa: pain, A: anger. S-Sub : SMIC-sub, CASIA: CASME I-A , CASIB: CASME I-B, CAS^2: CAS(ME)^2, MEV: MEVIEW, CASII: CASME II.

Dataset	H	D	Su	R	T	F	C	Sa	He	Pa	A
S-sub	17	10	20	0	0	16	0	8	0	0	0
CAS1A	4	4	7	30	48	1	2	0	0	0	0
CAS1B	5	42	14	10	23	1	0	6	0	0	0
CAS^2	15	16	10	0	0	4	0	1	1	2	0
MEV	6	1	9	0	0	3	7	0	0	0	2
SAMM	26	9	15	0	0	8	12	6	0	0	57
CAS2	33	60	25	27	102						

3 APPLICATIONS AND DISCUSSIONS OF MICRO-EXPRESSION DATABASES

3.1 Databases in ME Analysis

In the following sub-sections, we introduce the ME analysis based on databases. We firstly discuss the classes for ME recognition, and then introduce the frequency of recognition metrics for databases. Concerning ME detection, the most used databases are listed to provide a reference for the further research.

Classes for Micro-Expression Recognition. The emotion classes for recognition vary depending on different chosen databases. Furthermore, since the emotion samples distribute unevenly, some authors have defined their own emotion classes. They may combine emotion classes who have small proportions in the entire database into one class. For example, in (Xiaohua et al., 2017), the emotion classes for CASME I was set as positive, negative, surprise and others. Another solution is selecting useful samples for evaluating. For instance, in (Guo et al., 2015), only ME samples in CASME I correspond to happiness, surprise, repression and tense are used. For articles which performed their experiments on SMIC, the most common emotion classification is positive, negative and surprise (Li et al., 2017). Meanwhile, for articles using CASME II, the ME samples were usually classified as happiness, surprise, repression, disgust, and others (Wang et al., 2016). Table 5 lists all the emotion class numbers, the corresponding emotion types and article numbers. The emotion classes of database SMIC and CASMEII are used most frequently. Yet, there are 17 kinds of emotional classification, this makes it difficult to compare results between articles. Moreover, it's worth noticing that there are few articles recognizing videos samples who have no ME. Usually, the frame samples chosen for recognition are the frames from onset to offset.

Table 5: Summary of emotion classes for micro-expression recognition. The two most commonly used emotion classes are highlighted in bold. P: positive, N: negative, H: Happiness, D: Disgust, SU: Surprise, R: Repression, T: Tense, F: Fear, C: Contempt, SA: Sadness. The highest values are highlighted in bold.

# of emotions	Emotion types	Article numbers
2	P, N	2
3	**P, N, SU**	**28**
	P, N, Neutral	1
	H, SU, SA	1
4	P, N, SU, others	9
	SU, R, T, D	6
	SU, R, T, H	1
5	Attention, SU, D, R, T	1
	H, SU, D, R, T/others	**33**
	H, SU, SA, A, Neutral	1
6	H, SU, D, F, SA, A	2
	H, SU, D, F, SA, R	1
	H, SU, D, F, T, Neutral	1
7	H, SU, D, F, SA, A, Neutral	1
	H, SU, D, F, SA, A, C	2
8	H, SU, D, F, SA, A, C, Neutral	1
9	H, SU, D, F, SA, A, C, T, others	1

It is difficult to identify precisely the ME as one definite emotion without consideration of gesture and context. It occasionally exists some conflicts between emotional label manually annotated by the psychologists and the automatic recognition result. Hence, as introduced section 2.4, the objective classification has been encouraged. Recognizing MEs with AU combination would be more rational and reliable rather than defining the emotion type. Besides, objective classification can serve to unify the number of classes in different databases. It would facilitate the comparison between different methods.

Frequency of Recognition Metrics for Databases. Table 6 comprehensively listed the number of published articles which evaluated their results by these principal metrics and their corresponding databases. We can find that the accuracy is the most common metric, and CASMEII and SMIC are the two most used databases. Yet, few articles performed experiments on SAMM and CAS(ME)^2. Since these two databases contain more facial movements, to improve the recognition performance, we expect that more attentions could be paid on SAMM and CAS(ME)^2.

Micro-Expression Detection. There is only one article (Husák et al., 2017) which spotted ME in in-the-wild database MEVIEW. Thus, in this paragraph, we just review the spontaneous ME detection methods. Table 7 shows the number and frequency of databases used for detection. CASME II and SMIC-E are the two databases most frequently used. SAMM and CAS(ME)^2 contain longer video samples. There are more non-ME sample in these two databases than in previous ones, including neutral faces, eye blinks, subtle head movement, etc. As the ME detection applications in real life are usually performed on long videos, these two databases allow the methods adapting more easily to real situations.

Table 6: Summary of numbers of articles, with principal metrics and their corresponding databases. SMIC includes SMIC and SMIC-E. ACC: accuracy, CF: confusion matrix. CAS1: CASMEI, CAS2: CASMEII, CAS^2: CAS(ME)^2.

	CAS1	CAS2	SMIC	SAMM	CAS^2
ACC	22	61	37	5	2
CF	13	29	17	3	1
F1-score	5	22	13	6	-
Recall	3	13	8	2	-
Precision	2	11	7	-	-
Time	1	5	5	-	1
ROC	2	3	2	2	1

Table 7: Database numbers and frequency (%) for ME detection. The number of two most frequently used databases are highlighted in bold. F(%) means the used frequency of database for all the published articles. CAS1: CASMEI, CAS2: CASMEII, CAS^2: CAS(ME)^2, SMIC: SMIC-E.

	CAS2	SMIC	CAS1	SAMM	CAS^2	MEVIEW
#	**13**	**11**	7	2	2	1
F(%)	**62**	**52**	33	10	10	5

3.2 Discussion on Micro-Expression Databases

The posed ME is a reaction commanded by brain. The duration is longer than that of spontaneous ME. Yet, the short duration is one of the most important characteristics for ME. Hence, posed datasets are not used anymore. Nowadays, the majority of automatic ME analysis researches performed their experiments on spontaneous ME databases. Each database has its own advantages. CASME I, CASME II and SAMM have both emotional labels and AU labels. SMIC provides a possibility to analyze ME by multi-modalities. SAMM responds to the necessity of multi-ethnic groups. In addition, SAMM and CAS(ME)^2 have not only the ME but also other facial movements. Moreover, the video length of these two databases is longer than the others. Thus, even though CASME II and SMIC are two most commonly used databases, SAMM and CAS(ME)^2 are very promising for improving the ME analyzing performance in real world.

Nevertheless, there is still plenty of work to do. Firstly, the population genericity should be increased. 1). There are too few subjects and the most come from universities. The age range needs to be extended. E.g., facial wrinkles may affect the recognition result. Furthermore, the students do not have much experience of hiding emotions in high stake situations. To apply the ME analysis in the real world, we need more participants from society. 2). As it's a difficult task for children to hide their genuine emotions, the ME feature could be different from that of adults. Thus, ME samples collected from children should be considered. However, building a database containing children subjects would concern many legislative issues.

Secondly, more modalities, e.g. infrared video, could help improve the recognition ability by cross-modalities analysis. Thirdly, as the research on automatic MEDR just begun in recent decade, almost all the ME databases were built in a strictly controlled laboratory environment to facilitate the pre-processing. Along with the development of MEDR research, in-the-wild ME videos with more occlusions, such as pose variation, hair on the face, lightning change, etc., are expected by collecting from TV shows or by crowd-sourcing. Fourthly, concerning annotation, utilizing AU annotations could be a more objective way for ME classification. Meanwhile, the accuracy of annotation needs to be improved since there are still many non-labeled detected facial movements in the existing databases. Fifthly, the number of ME databases could be augmented by considering FACS Coded Databases like DISFA (Mavadati et al., 2013) and BP4D (Zhang et al., 2014). The spontaneous facial expressions in these databases are labelled with AUs intensities, expression sequences with short duration and low intensity could be used as ME samples.

In addition, due to the limited acquisition condition, we are looking forward to a comprehensively collected database by cooperation among worldwide research groups.

The last discussion is about definition of eye gaze change, subtle expression and masked expression. They have not attracted much attention. Nevertheless, it worth discussing them for the future ME database construction and ME analysis.

- The eye gaze shift also reveals the personal emotion, even without any action units that associate to it. It could be considered as a clue for identifying MEs. However, to officially use it as ME indicator, it still needs acknowledgment from psychologists and automatic MEDR research communities. Furthermore, samples in SDD could be used for analyzing the ME with eye gaze shift.
- The subtle expression is a small facial movement (spatial), but the duration could be longer than 500ms. The study on subtle expression would be a challenge due to the undefined duration.
- Regarding the masked expression, there might be some MEs masked in other facial movements. For example, the tense expression could be hidden during an eye blinking. Analyzing this kind of ME seems to be impossible based on currently proposed methods. We are looking forward to more studies on this problem.

4 CONCLUSIONS

By comprehensibly reviewing the existing databases, this paper gives some guidelines and suggestions for the further ME database construction. Regarding databases, 13 characteristics are presented in 4 categories. This classification could help other researchers to choose databases as needed. The future direction for databases is under discussion. The diversity of the population and the number of modalities should be increased. More in-the-wild databases are expected. Concerning the ME analysis based on databases, the objective classification is encouraged, and we are looking forward to having more experiments on recently published databases.

ACKNOWLEDGEMENTS

The authors would like to thank the financial support of China scholarship Council and ANR Reflet.

REFERENCES

Birdwhistell, R. L. (1968). Communication without words. *Ekistics*, pages 439–444.

Davison, A., Merghani, W., and Yap, M. (2018a). Objective classes for micro-facial expression recognition. *Journal of Imaging*, 4(10):119.

Davison, A. K., Lansley, C., Costen, N., Tan, K., and Yap, M. H. (2018b). Samm: A spontaneous micro-facial movement dataset. *IEEE Transactions on Affective Computing*, 9(1):116–129.

Eckman, P. (2003). Emotions revealed. *St. Martin's Griffin, New York*.

Ekman, P. (2009). Lie catching and microexpressions. *The philosophy of deception*, page 118–133.

Ekman, P. and Friesen, W. (1978). Facial action coding system: a technique for the measurement of facial movement. *Palo Alto: Consulting Psychologists*.

Ekman, P. and Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, 32(1):88–106.

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

Endres, J. and Laidlaw, A. (2009). Micro-expression recognition training in medical students: a pilot study. *BMC medical education*, 9(1):47.

Frank, M., Herbasz, M., Sinuk, K., Keller, A., and Nolan, C. (2009). I see how you feel: Training laypeople and professionals to recognize fleeting emotions. In *The Annual Meeting of the International Communication Association. Sheraton New York, New York City*.

Grobova, J., Colovic, M., Marjanovic, M., Njegus, A., Demire, H., and Anbarjafari, G. (2017). Automatic hidden sadness detection using micro-expressions. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*, pages 828–832. IEEE.

Guo, Y., Xue, C., Wang, Y., and Yu, M. (2015). Micro-expression recognition based on cbp-top feature with elm. *Optik-International Journal for Light and Electron Optics*, 126(23):4446–4451.

Haggard, E. A. and Isaacs, K. S. (1966). Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. In *Methods of research in psychotherapy*, page 154–165. Springer.

He, J., Hu, J.-F., Lu, X., and Zheng, W.-S. (2017). Multi-task mid-level feature learning for micro-expression recognition. *Pattern Recognition*, 66:44–52.

Husák, P., Čech, J., and Matas, J. (2017). Spotting facial micro-expressions "in the wild". In *Proceedings of the 22nd Computer Vision Winter Workshop*. Pattern Recognition and Image Processing Group (PRIP) and PRIP club.

LI, J., Soladié, C., and Séguier, R. (2018). Ltp-ml: Micro-expression detection by recognition of local temporal pattern of facial movements. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, pages 634–641. IEEE.

Li, X., Pfister, T., Huang, X., Zhao, G., and Pietikäinen, M. (2013). *A spontaneous micro-expression database: Inducement, collection and baseline*, page 1–6. IEEE.

Li, X., Xiaopeng, H., Moilanen, A., Huang, X., Pfister, T., Zhao, G., and Pietikainen, M. (2017). Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Transactions on Affective Computing*.

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

Oh, Y.-H., See, J., Le Ngo, A. C., Phan, R. C.-W., and Baskaran, V. M. (2018). A survey of automatic facial micro-expression analysis: Databases, methods and challenges. *Frontiers in Psychology*, 9:1128.

Pfister, T., Li, X., Zhao, G., and Pietikäinen, M. (2011). Recognising spontaneous facial micro-expressions. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1449–1456. IEEE.

Polikovsky, S. (2009). Facial micro-expressions recognition using high speed camera and 3d-gradients descriptor. In *Conference on Imaging for Crime Detection and Prevention, 2009*, volume 6.

Qu, F., Wang, S.-J., Yan, W.-J., Li, H., Wu, S., and Fu, X. (2017). Cas (me)ˆ 2: a database for spontaneous macro-expression and micro-expression spotting and recognition. *IEEE Transactions on Affective Computing*.

Radlak, K., Bozek, M., and Smolka, B. (2015). Silesian deception database: Presentation and analysis. In *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, pages 29–35. ACM.

Shreve, M., Godavarthy, S., Goldgof, D., and Sarkar, S. (2011). *Macro-and micro-expression spotting in long videos using spatio-temporal strain*, page 51–56. IEEE.

Vinciarelli, A., Dielmann, A., Favre, S., and Salamin, H. (2009). Canal9: A database of political debates for analysis of social interactions. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–4. IEEE.

Wang, S.-J., Yan, W.-J., Sun, T., Zhao, G., and Fu, X. (2016). Sparse tensor canonical correlation analysis for micro-expression recognition. *Neurocomputing*, 214:218–232.

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

Weber, R., Soladié, C., and Séguier, R. (2018). A survey on databases for facial expression analysis. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP,*, pages 73–84. INSTICC, SciTePress.

Xiaohua, H., Wang, S.-J., Liu, X., Zhao, G., Feng, X., and Pietikainen, M. (2017). Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition. *IEEE Transactions on Affective Computing*.

Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., and Fu, X. (2014). Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. *PloS one*, 9(1):e86041.

Yan, W.-J., Wu, Q., Liu, Y.-J., Wang, S.-J., and Fu, X. (2013). *CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces*, page 1–7. IEEE.

Yap, M. H., See, J., Hong, X., and Wang, S.-J. (2018). Facial micro-expressions grand challenge 2018 summary. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, pages 675–678. 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.