CAS(ME)^2: A Database for Spontaneous Macro-expression and Micro-expression Spotting and Recognition

Fangbing Qu, Su-Jing Wang, *Member, IEEE,* Wen-Jing Yan, He Li, Shuhang Wu, Xiaolan Fu, *Member, IEEE*

Abstract—Deception is a very common phenomenon and its detection can be beneficial to our daily lives. Compared with other deception cues, micro-expression has shown great potential as a promising cue for deception detection. The spotting and recognition of micro-expression from long videos may significantly aid both law enforcement officers and researchers. However, database that contains both micro-expression and macro-expression in long videos is still not publicly available. To facilitate development in this field, we present a new database, Chinese Academy of Sciences Macro-Expressions and Micro-Expressions (CAS(ME)^2), which provides both macro-expressions and micro-expressions in two parts (A and B). Part A contains 87 long videos that contain spontaneous macro-expressions and micro-expressions. Part B includes 300 cropped spontaneous macro-expression samples and 57 micro-expression samples. The emotion labels are based on a combination of action units (AUs), self-reported emotion for every facial movement, and the emotion types of emotion-evoking videos. Local Binary Pattern (LBP) was employed for the spotting and recognition of macro-expressions and micro-expressions and the results were reported as a baseline evaluation. The CAS(ME)^2 database offers both long videos and cropped expression samples, which may aid researchers in developing efficient algorithms for the spotting and recognition of macro-expressions and micro-expressions.

Index Terms—Macro-expression and micro-expression spotting, macro-expression and micro-expression recognition, micro-expression database, facial action coding system.

1 Introduction

INTERPERSONAL deception is a common aspect of human social interaction. Despite people's experiences with deceiving and being deceived by others, lies are notoriously difficult to detect, even for the skilled expert in law enforcement field. As a traditional system for lie detection, the polygraph is the extensively employed method as it can monitor uncontrolled changes in heart rate and electro-dermal responses, as a result of the subjects' arousal during telling a lie. However, the recording process of polygraph is intrusive and subjects may employ countermeasures to hide their true thoughts [1]. Recently, another possible index of lying, micro-expression, which is characterized by short duration, low intensity and typically local movements, has drawn the attention of affective computing researchers and psychologists. Micro-expressions are rapid and brief expressions that appear when individuals attempt to conceal their genuine emotions, especially in high-stake situations [2][3]. Ekman has even claimed that micro-expressions may be the most promising cues for lie detection [3].

Unlike a polygraph, the micro-expression can be captured with an unobtrusive camera during interaction or interviews, thus, subjects may not be aware that they are being monitored. Therefore, the automatic spotting of micro-expressions from video streams in interrogation interview contexts may significantly aid law enforcement officers in detecting suspects' usual or deception clues. Despite the potential of automatically spotting micro-expression from long videos, publicly available databases with micro-expression embedded in long videos remain rare.

In this paper, we first reviewed previous studies regarding the construction of facial expression databases (including conventional and micro-expression) and efforts in the automatic spotting and recognition of facial expressions by previous investigators. Then we revealed

This project was partially supported by the National Natural Science Foundation of China (61375009, 61379095, 31500875) and the Beijing Natural Science Foundation (4152055).
F.B Qu is with the State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, and with the University of the Chinese Academy of Sciences, Beijing, 100049, and also with the College of Preschool Education, Capital Normal University, Beijing, 100048. (e-mail: qufangbing@gmail.com).
S.J Wang is with the CAS Key Laboratory of Behavior Sciences, Institute of Psychology, Beijing, 100101. (e-mail: wangsujing@psych.ac.cn).
W.J Yan is with the College of Teacher Education, Wenzhou University, Wenzhou, 325035, China. (e-mail: yanwj@wzu.edu.cn).
H Li is with the State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, and also with the University of the Chinese Academy of Sciences, Beijing, 100049. (e-mail: qufangbing@gmail.com).
S.H Wu is with the College of Information Science and Engineering, Northeastern University, Shenyang, China. (e-mail: wushuhang818@126.com).
X.L Fu is with the State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, and also with the University of the Chinese Academy of Sciences, Beijing, 100049. (e-mail: fuxl@psych.ac.cn).
F.B Qu and S.J Wang contributed equally to this paper.

promising improvements according to previous research and presented our new database: the , Chinese Academy of Sciences Macro- and Micro- expression (CAS(ME)^2) database. In the third part of this paper, we conducted a database baseline evaluation and presented relevant results.

The paper was organized as follows. In Section 2, we review previous studies regarding the construction of facial expression databases, efforts in the automatic spotting and recognition of facial expression by previous investigators, analyses of previously employed emotion induction methods and the characteristics of CAS(ME)^2. In section 3, basic profile of this database, including the elicitation materials and procedure, coding and labelling process and user guide, are introduced. In Section 4, a database baseline evaluation is conducted and the relevant results are presented, while the paper ended in Section 5 with discussion and conclusions.

2 Related Work

2.1 Analyses of previous facial expression databases

To date, numerous methods and algorithms for automatically spotting and recognizing emotions from human faces have been developed [4][5][6]. Researchers have primarily focused on general facial expressions, which are usually referred to as macro-expressions, that can be easily noticed and typically last for more than 1/2 of a second, up to 4 seconds [7][1]. While the micro-expressions are those that occur over a brief period of time and can commonly go unnoticed. To be specific, Micro-expressions are rapid and brief expressions that appear when individuals attempt to conceal their genuine emotions, especially in high-stake situations [9][10]. Compared with the research on macro-expressions,the algorithms for micro-expression spotting and recognition , are partly constrained due to the limited number of micro-expression databases, especially databases that contain micro-expression in long videos that can be employed for automatic micro-expression spotting.

Substantial research has been conducted regarding macro-expression recognition. This progress would not have been possible without the construction of facial expression databases, which have significantly facilitated the development of facial expression recognition systems. Numerous facial expression databases, such as the Japanese Female Facial Expression Database (JAFFE) [11]; the CMU Pose, Illumination and Expression (PIE) database [12] and its new version the Multi-PIE database [13]; and the Genki-4K database [14], have been developed(see Table 1) [15][6]. However, these databases only contain still facial expression images that represent different emotional states.

Still facial expression images contain less facial motion information than dynamic facial expression image sequences, which are more discriminant for recognizing expression. Therefore, researchers are becoming attentive to dynamic information and have developed several databases that contain dynamic facial expression image sequences rather than still images, such as the RU-FACS database [16], the MMI facial expression database [17], Cohn-Kanade database (CK) [18] and the GEMEP database [19].

However, these databases only include posed expressions (i.e., the participants were asked to present or pose certain facial expressions, such as happy and sad expressions) rather than naturally expressed or spontaneous facial expressions. Previous studies have indicated that posed expressions may differ in appearance and timing from spontaneously occurring expressions [20] (For a detailed survey of posed and none-posed databases, refer to Sariyanidi *et al.* [6]). Therefore, a database that contains spontaneous facial expressions would have more ecological validity.

To address the issue of facial expression spontaneity, Lucey *et al.* [21] developed the Extended Cohn-Kanade Dataset (CK+) by collecting numerous spontaneous facial expressions. However, this dataset only included happy expressions that spontaneously occurred between the participants' facial expression posing tasks. Mavadati *et al.* [22] constructed the Denver Intensity of Spontaneous Facial Action (DISFA) Database by collecting participants' spontaneous facial expressions while they watched videos that were intended to elicit spontaneous emotion expression and coded presence, absence, and intensity of facial action units (AUs). Recently, Mcduff *et al.* [23] presented the Affectiva-MIT Facial Expression Dataset (AM-FED), which contains naturalistic and spontaneous facial expressions collected online from volunteers recruited on the Internet who agreed to be videotaped while watching amusing Super Bowl commercials. This database achieved improved validity via the collection of spontaneous expression samples recorded in natural settings. However, only facial expressions that were understood to be related to a single emotional state (amusement) were recorded; self-reports on the the experiences of the participants were not obtained. Wang *et al.* [24] published a new expression database, the natural visible and infrared facial expressions (NVIE), with both posed and spontaneous expression samples. They also collected self-reported data to analyze the effectiveness of emotion-eliciting videos and compared the differences between posed and spontaneous expressions by analyzing their thermal differences. This study provide an excellent example of expression database construction. Zhang *et al.* [25] recently published a new facial expression database of three-dimensional (3D) spontaneous elicited facial expressions, the Binghamton Pittsburgh Four-dimensional Spontaneous Expression Database (BP4D-Spontaneous), which is coded based on participants' self-reports, subjective ratings from naive observers and the Facial Action

1. Some researchers thought the duration of macro-expression to be 1/2 second to 2 seconds [8].

Coding System (FACS), which improved the validity of expression coding and labeling. However, only the general self-reported feelings of participants were not collected after each task, self-reports for every facial movement during the tasks were collected to exclude any unemotional movements that may contaminate the database, such as blowing of the nose, swallowing of saliva or rolling of the eyes.

2.2 Analyses of previous micro-expression databases

Compared with numerous macro-expression databases, however, databases that concern micro-expressions are rare. To our knowledge, only six micro-expression datasets were published, each with different advantages and disadvantages (see Table 1): USF-HD [8]; Polikovsky's database [26]; SMIC [27] and its successor, an extended version of SMIC [28]; and CASME [29] and its successor, CASME II [30].

These micro-databases have significantly facilitated the development of automatic micro-expression recognition. However, these databases only include cropped micro-expression samples, which are not suitable for automatic micro-expression spotting. The methods employed for emotion labeling were not consistent, emotions were usually labelled according to the FACS, the emotion type of the elicitation materials or both. These methods offered the possibility of including some meaningless facial movements in the database (such as blowing of the nose, swallowing of saliva or rolling of the eyes). In the CASME [29] and CASME II [30], the authors attempted to collect participants' self-reports after they watched each emotional videos, as supplementary information for emotion labeling. In this database, together with the FACS and emotion type of elicitation material, we collected participants' self-reports for each of their facial movements, which to the best extent guaranteed the purity of the micro-expression database.

Based on the micro-expression databases, several publications attempt to improve the development of automatic micro-expression recognition. Polikovsky *et al.* [26] employed a 3D-gradient descriptor for micro-expression recognition. Wang *et al.* [31] treated a gray-scale micro-expression video clip as a 3rd-order tensor and applied Discriminant Tensor Subspace Analysis (DTSA) and Extreme Learning Machine (ELM) approaches to recognize micro-expressions. However, subtle movements involved in micro-expressions may be lost in the process of DTSA. Pfister *et al.* [27] employed a temporal interpolation model (TIM) and local binary patterns on three orthogonal planes (LBP-TOP) [32] to extract the dynamic textures of micro-expressions. Wang *et al.* [33] used an independent color space to improve this work, and they [34] also applied Robust PCA [35] to extract subtle motion information regarding micro-expressions. Liu *et al.* [36] proposed a simple yet effective Main Directional Mean Optical-flow (MDMO) feature for micro-expression recognition. Xu *et al.* [37] proposed a Facial Dynamics Map (FDM) to describe the motion pattern of a micro-expression instance. Wang *et al.* [38] proposed Sparse Tensor Canonical Correlation Analysis (STCCA) for micro-expression recognition.

Compared with studies of micro-expression recognition, studies of micro-expression spotting are rare. Shreve *et al.* [39] primarily employed the facial strain and dense optical flow to detect and discriminate macro-expressions and micro-expressions. Polikovsky *et al.* [26][40] measured the duration of the three phases of micro-expressions by a 3D-gradient descriptor. Moilanen *et al.* [41] primarily employed LBP features to obtain both temporal and spatial locations for micro-expression spotting, but they used SMIC databases that included cropped micro-expression samples that include the frames of micro-expressions from onset to offset or the CASME database that include videos that are not very reasonable for micro-expression spotting - shortest video that only lasts 0.2 seconds.

Therefore, research on micro-expression spotting was primarily constrained by the development of micro-expression databases that contain micro-expressions in long videos. To our knowledge, there is no publicly available database that contains micro-facial expressions in long videos that can be used for micro-expression automatic spotting. The only related study on micro-expression spotting from long videos was conducted by Shreve [8][39], who applied a spatio-temporal strain method on a database with expression samples collected from different sources[2]. However, this database is not publicly available.

2.3 Analyses of previous emotion induction methods

In previous studies that construct facial expression databases, different types of emotion induction methods, such as emotional acting, other manipulation methods (olfactory stimulation, interview, and social challenge), and film watching, were employed (see Table 1); each method has advantages and disadvantages.

In the first method, emotional acting (JAFFE, PIE, multi-PIE,and GEMEP), participants or actors are asked to perform certain types of emotional facial expressions. This method is always criticized for the elicitation of posed expressions, which differ from spontaneous expressions in the utilized facial muscles and their dynamics [42][43]. For instance, many types of spontaneous smiles are smaller in amplitude, longer in total duration, and slower in onset and offset times than posed smiles [44][45].

In the second method, a series of tasks (such as the BP4D), such as olfactory stimulation to elicit disgust, interviews, the cold pressor test to elicit pain, and social challenges to elicit anger followed by reparation were

2. 56 videos from USF database, not public available; 6 video from Canal-9 political debates and 3 low quality video from the internet.

The final version of record is available at http://dx.doi.org/10.1109/TAFFC.2017.2654440

TABLE 1
The previous macro-expression and micro-expression databases

Databases	Numbers	Emotions include	Induction methods	Posed or Spontaneous	Tagging	Database includes
JAFFE	10 subjects	6(happiness/sadness/surprise /anger/disgust/fear)	Acting	Posed	Raters rating	Conventional/ Macro-expressions
PIE	337 subjects	4(neutral/smiling/ blinking/talking)	Acting	Posed	Pose and Expression-labeled	Conventional/ Macro-expressions and poses
Multi-PIE	337 subjects	6(neutral/smile/surprise /squint/disgust/scream)	Acting	Posed	Pose and Expression-labeled	Conventional/ Macro-expressions
Genki-4K	4000 images	Smiling or non-smiling	Download from the web	Not given	Pose- and Expression-labeled	Conventional expression images
RU-FACS	100 subjects	Not given	False opinion paradigm interviews	Posed and spontaneous	FACS	Conventional/ Macro-expressions
MMI	52 subjects	6(anger/happy/sad/ surprise/fear/disgust)	Acting	Mostly posed	FACS	Conventional/ Macro-expressions
CK	100 subjects	Smile	Acting and film watching	Posed and Spontaneous	FACS	Conventional/ Macro-expressions
CK+	593 videos	Smile	Acting and film watching	Posed and Spontaneous	FACS	Conventional/ Macro-expressions
GEMEP	7000 audiovisual emotion portrayals	18 emotions	Acting	Posed	FACS	Conventional/ Macro-expressions
DISFA	27 subjects	5(joy/surprise/ disgust/sad/fear)	Film watching	Spontaneous	FACS	Conventional/ Macro-expressions
AM-FED	242 videos	Smile	Film watching	Spontaneous	FACS	Conventional/ Macro-expressions
NVIE	215 subjects	7(happiness/sadness/surprise /fear/anger/disgust)	Acting and film watching	Posed and Spontaneous	Raters rating	Conventional/ Macro-expressions
BP4D	41subjects	8(happiness or amusement/ Sadness/Surprise or startle/ Embarrassment/Fear or nervous/ Pain/Anger or upset/Disgust)	Combined (interviews/planned activities/ film watching/cold pressor test/ social challenge/Olfactory stimulation)	Spontaneous	FACS	Conventional/ Macro-expressions
USD-HD	100 samples	6 (basic emotion)	Acting	Posed	Micro/non-Micro expressions	Micro/non-Micro expressions
Polikovsky's database	42 samples	6 (basic emotion)	Acting	Posed	FACS	Micro-expressions
SMIC	77 samples	3 (Positive/Negative/surprise)	Neutralization paradigm	Spontaneous	Self-reported emotion	Micro-expressions
SMICII HS	164 samples	3(Positive/Negative/surprise)	Neutralization paradigm	Spontaneous	Self-reported emotion	Micro-expressions
SMICII VIS	71 samples					
SMICII NIR	71 samples					
CASME	195	7 (basic emotion)	Neutralization paradigm	Spontaneous	Emotion/FACS	Micro-expressions
CASMEII	247	5 (basic emotion)	Neutralization paradigm	Spontaneous	Emotion/FACS	Micro-expressions

employed [25]. Although these methods have relatively high ecological validity in eliciting spontaneous expressions, they have limitations regarding the effectiveness of the elicitation of certain emotions or expressions. Expressions elicited with these methods may vary in their characteristics, such as duration and intensity, which renders them less comparable between each other.

The most extensively employed method for eliciting emotion or spontaneous facial expression is film watching (DISFA, AM-FED, SMIC, extended version of SMIC, CASME and CASMEII). Video episodes have a relatively high degree of ecological validity [46] and are usually better than pictures in term of emotional valence. Video episodes are lasting and dynamic emotional stimuli, making inhibition more difficult. Therefore, film watching was employed in the construction of this database.

2.4 Characteristics of current database: CAS(ME)^2

Considering the previously mentioned issues, we present the CAS(ME)^2) database, which includes both spontaneous macro-expressions and micro-expressions in long videos (part A) and cropped expression samples with frames from onset to offset (part B) for automatic macro-expression and micro-expression spotting and recognition training. The main contributions of this database are summarized as follows:

- This database is the first publicly available database that contains both macro-expressions and micro-expressions in long videos, which facilitates the development of algorithms for spotting micro-expressions from long video streams.
- All macro-expressions and micro-expressions samples were collected from the same participants in the same experimental conditions, which enables researchers to develop more efficient algorithms to extract features that are better able to discriminate macro-expressions and micro-expressions and compare the differences in their feature vectors.
- The difference in AUs between macro-expressions and micro-expressions can be acquired by the algorithms tested on this database.
- The database employed the combination of FACS AUs, the emotional type of the elicitation videos and participants' self-reported emotions for each expression sample. After the expression-inducing phase, the participants were asked to watch the videos of their recorded facial expressions and provide a self-report for every expression. This procedure enabled us to exclude almost all emotion-irrelevant facial movements and obtain relatively pure expression samples. The self-reported emotions for every expression sample were also provided, they can be employed for comparison with the AUs displayed by the participants.

This paper is an extended version of our paper presented at the International Conference of Human-Computer Interaction International (HCII 2016)[47]. The differences here are that we added the micro- and macro-expression embedded in long videos that can be used in automatic micro-expression spotting. We also conducted an automatic micro-expression spotting evaluation. In the following sections, we will firstly describe the database and its elicitation and coding procedures. We will also provide some basic evaluation results of the automatic spotting and recognition on this database as the baseline performance measure.

3 CAS(ME)^2 DATABASE PROFILE

The CAS(ME)^2 database contains two parts: Part A and Part B. Part A consists of 87 long videos that contain both macro-expressions and micro-expressions. Automatic macro-expression and micro-expression spotting were tested on this part using an appearance-based feature difference analysis method; the results were reported. Part B includes 357 cropped expression samples with 300 macro-expressions and 57 micro-expressions.

These expressions were filmed using a camera. The expression samples were selected from more than 600 elicited facial movements and were coded with the onset, apex, and offset frames[3], with AUs marked and emotions labeled [30].

To enhance the reliability of the emotion labeling, we obtained an additional emotion label by asking participants to review every recorded facial movement and give self-reported emotional experiences associated with each facial movements.

Macro-expressions with a duration of more than 500 ms and less than 4 s were selected for inclusion in this database [7]. Micro-expressions with a maximum duration of 500 ms were also selected.

The steps of the data acquisition and coding processes are presented in the following subsections. Table 2 presents descriptive statistics for the expression samples with different durations, which include 300 macro-expressions and 57 micro-expressions, defined in terms of total duration. Fig. 1 shows examples of a micro-expression (a) and a macro-expression (b). Considering that the intensity of these expressions is relatively low, we also provided the video clips that correspond to these two expressions to better illustrate their difference as supplementary materials. The distribution of expression samples according to both participants and emotion classes is also necessary when the expression databases are present [21] [48] [49]. Table 3 lists the numbers of samples according to each participant and different emotion classes in CAS(ME)^2.

3.1 Participants and Elicitation Materials

Twenty-two participants (13 females and 9 males), with a mean age of 22.59 years (standard deviation = 2.2), and

3. The onset frame was the first one which changes from the baseline (usually neutral facial expressions). The apex-1 frame is the first one that reached highest intensity of the facial expression and if it keeps for a certain time, the apex-2 frame is coded.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TAFFC.2017.2654440

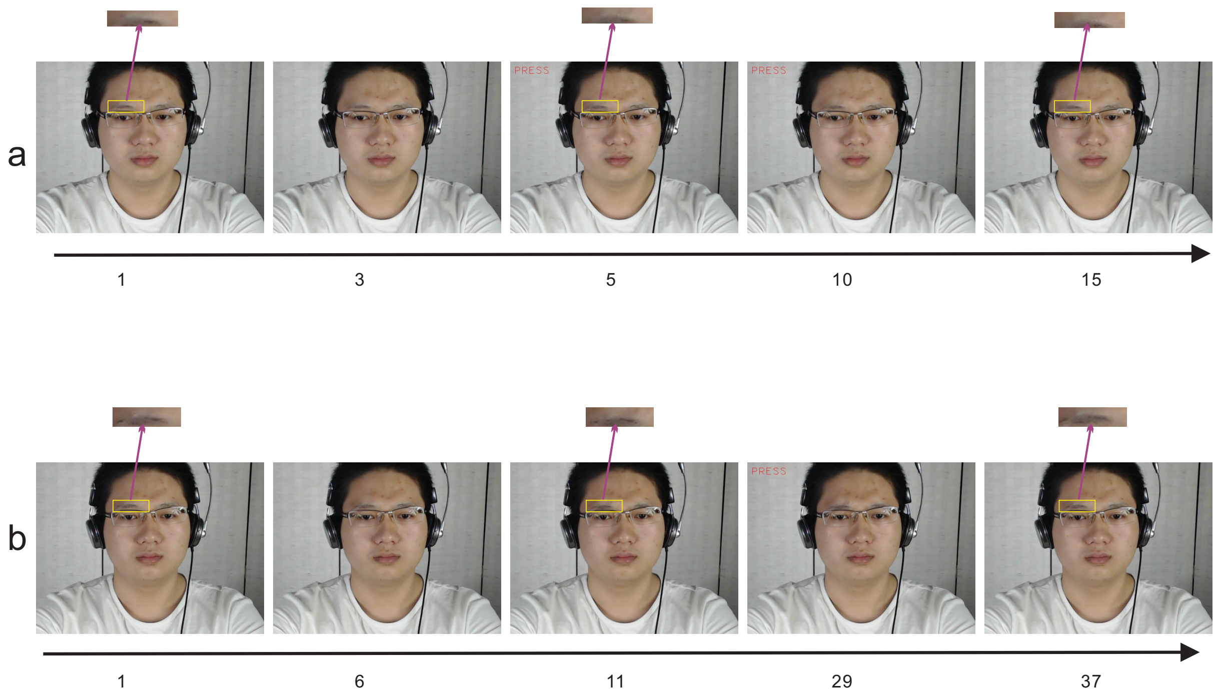


Fig. 1. Examples of micro-expression (a) and macro-expression (b). The apex frame presents at about frame 5 for the micro-expression and frame 11 for the macro-expression, which are all represents the negative emotion (anger). The AUs related on these two expressions are all AU4 (inner brow).

TABLE 2
Descriptive statistics for macro-expressions and micro-expressions

Expression type	Number	Total Duration	
		Mean (ms)	Standard Deviation
Macro-expression	300	1303	651
Micro-expression	57	419	66

an age range from 19-26 were recruited. All provided informed consent to the use of their video images for scientific research.

For the expression elicitation materials, nine emotional videos were selected from twenty videos that were rated by 20 other raters for their ability to induce expressions[4]. Based on the ability to elicit micro-expressions observed in previous studies, only three types of emotion-evoking emotion videos (videos that evoked disgust, anger, and happiness) were employed in this study. Their length ranged from 1 minute to approximately 2 minutes and 30 seconds (see Table 4). Two disgust-evoking emotion videos, two anger-evoking emotion videos, and five happiness-evoking emotion

4. Seventeen emotion videos used in previous studies [10] and 3 new emotion videos downloaded from the internet were rated with regard to their ability to elicit facial expressions. Twenty participants rated the main emotions associated with these emotion videos and assigned them arousal intensity scores from 1 to 5 (1 represents the weakest intensity, and 5 represents the strongest intensity).

TABLE 3
Numbers of samples and frames according to participants and emotions

Participant Number	Positive	Negative	Surprise	Others
1	3	10		4
2	13	26	3	12
3	2			3
4	3			
5		1		1
6	1	5		10
7	1	1		9
8	13	16	2	1
9	4	5		
10		2		4
11	14	23	1	5
12				1
13	12	1		6
14	1	3	14	14
15	21	26	4	3
16		1		3
17	1			2
18		1	1	
19	3			1
20	17	3		3
21	12			
22	3	2		

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TAFFC.2017.2654440

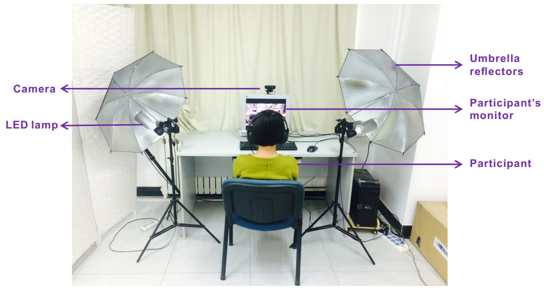


Fig. 2. The capturing environment.

videos were chosen[5]. Each emotion video predominantly elicited one type of emotion.

TABLE 4
Participants' ratings on the 9 emotion videos

Number	Duration	Emotion	Rate of Selection	Mean Score
1	1′07″	Disgust	0.86	4.14
2	1′35″	Disgust	0.92	4.33
3	1′57″	Anger	0.75	4.5
4	2′24″	Anger	0.77	3.92
5	1′18″	Happiness	0.92	4
6	1′32″	Happiness	0.86	3.07
7	1′16″	Happiness	0.86	3.28
8	1′48″	Happiness	0.73	3.64
9	1′09″	Happiness	0.71	3.17

3.2 Elicitation Procedure

During the elicitation procedure, each participant was seated in front of a Logitech Pro C920 camera (with 30 frames per second, and the resolution was set to 640×480 pixels), which was set on a tripod behind the monitor to record the full-frontal face of the participant. The participants were seated in a room with two light-emitting diode (LED) lights (see Fig. 2). The nine video episodes were presented by the experimenter in random order. The participant was told to closely watch the screen and maintain a neutral face.

Because the elicitation of micro-expressions requires strong motivation to conceal truly experienced emotions, motivation manipulation protocols were needed. Following the neutralization paradigm that were employed in previous studies [10], the participants were informed that the purpose of the experiment was to test their ability to control their emotional expressions, which was strongly related to their social success. The participants were also told that their payment would be directly related to their performance. To reduce noise and detection loss when applying automatic spotting algorithms, we informed the participants beforehand to avoid making large head movements. To ensure that the participants watched the emotion-eliciting films on the screen, we also asked participants to keep their eyes on the screen as much as possible.

After watching all nine emotion-eliciting videos, the participants were asked to review the recorded videos of their faces to identify any facial movements and provided a self-report of the inner feelings that they had experienced during every facial movement (when they were not sure about the emotional feelings associated with a certain facial movement, they could review the original emotion-eliciting video). These self-reports of the feelings associated with each expression were collected and employed as a separate emotion labeling system, as described in the following section. Only the facial movements that were reported with emotional meaning were included in this database.

3.3 Coding Process

Two well-trained FACS coders coded the videotaped facial expression videos for the presence and duration of emotional expressions in the upper and lower facial regions frame-by-frame. This coding required classifying the emotion exhibited in each facial region in each frame; recording the onset time, apex time and offset time of each expression; and arbitrating any disagreement that occurred between the coders. The coders coded 28 different AUs (the most frequently occurred AU is AU12 which occurred 129 times) and classified all emotions into four emotion labels: positive, negative, surprise and others. When the coders could not agree on the exact frame of the onset, apex or offset of an expression, the average of the values specified by both coders was employed. The two coders achieved a coding reliability (frame agreement) of 0.82 (from the onset frame to the offset frame). The coders also coded the AUs of each expression sample. The reliability between the two coders was 0.8, which was calculated as

$$R = \frac{2 \times AU(C_1C_2)}{All_{AU}} \quad (1)$$

where $AU(C_1C_2)$ is the number of AUs on which Coder 1 and Coder 2 agreed and All_{AU} is the total number of AUs in the facial expression scored by the two coders. The coders discussed and arbitrated the disagreements [30].

3.4 Emotion Labeling

When labeling the emotions that are associated with facial expressions, previous researchers have usually employed the emotion types that are associated with the corresponding emotion-evoking videos and employed the FACS AUs as the ground truth [28][30]. However, an

5. We chose five happiness-evoking videos because: Firstly, all the five happiness-evoking emotion videos were rated and selected for their relatively high ability to elicit happiness. Secondly, we chose five videos because the happy micro-expression was relatively difficult to elicit and happy expression maybe easier to suppress compared with other kind of facial expression such as disgust or fear. Therefore, we chose more than two elicitation videos (five happiness-evoking videos) to ensure the number of happy micro-expressions.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TAFFC.2017.2654440

emotion-evoking video may consist of multiple emotion-evoking events. Therefore, the emotion types that are estimated according to the FACS and the emotion types of the evoking videos are not representative, and many facial movements such as blowing of the nose, blinking, and swallowing of saliva may also be included among the expression samples.

In addition, micro-expressions may differ from macro-expressions because they may occur involuntarily, partially and in short durations; thus, the emotional labeling of micro-expressions that is only based on the FACS AUs and the emotion types of the evoking videos is incomplete. We must also consider the feelings that are reflected in the self-reported emotions of the participants when labeling micro-expressions.

In this database, a combination of AUs, emotion types of expression-elicitation videos and self-reported emotions for every facial movements was employed to enhance the validity of the emotion labels. Table 5 lists the emotion labeling criteria based on the FACS coding results, the emotion types of the emotion-evoking videos and the self-reported emotions.

In previous databases, facial expression samples were typically classified into basic emotions, such as happiness, sadness, surprise, disgust and anger [10] or more general terms such as positive, negative and surprise[28]. In this database, we classified the facial expression samples into four categories positive, negative, surprise and others according to the previously mentioned labeling method (see Table 5). "Positive" indicates the micro-expression of positive emotions such as happy, delightful, and amusement. "Negative" indicates a micro-expression associated with negative emotions such as fear, disgust and anger. Usually, the "surprise" expression is difficult to classify as either positive or negative; therefore, we classified the surprise facial expression into one independent class. "Others" indicates micro-expressions that have vague emotional meanings or micro-expressions that are difficult to categorize into six prototypical facial expressions.

3.5 CAS(ME)^2 User Guide

The CAS(ME)^2 database will be available online for download for research purposes. Part A includes all 87 raw facial expression video clips (in rawvideo.zip with .avi format) and image sequences (in rawpic.zip with .jpg format) without any preprocessing. These video clips and image sequences can be used for automatic macro-expression and micro-expression spotting. Both the onset and offset frame of macro- and micro-expression were shown in the file CAS(ME)^{2}code_final.txt. In this file, the second column represents the number of the expression and the name of the video clip. For example, in the first line, "anger1_1" is the first expression from the anger1 video clip, and 'anger1' is the name of the video clip.

Part B includes all 357 facial expressions with 300 macro-expressions and 57 micro-expressions (in selectedpic.zip with .jpg format). The current database also includes the cropped faces in the cropped.zip file and 68 feature points for each facial expression with the Active Shape Model (ASM) [50]. For details, please refer to Section 4.3.

The file CAS(ME)^{2}code_final.txt includes three sheet. The first one, "CAS(ME)^{2}code_final" includes nine columns. The first column contains the number of participants. The second column contains the number of the expression and the name of the video clip. For example, in the first line, "anger1_1" is the first expression from the anger1 video clip. The third column contains the first frame for the expression. The fourth column contains the apex frame of the expression. The fifth column contains the last frame for the expression. The sixth column contains the (AUs) (see FACS). The seventh column contains the estimated emotion. The eighth column contains the expression type (macro or micro expression). The ninth column contains the self-reported emotion. The second sheet introduces the naming rule of sub-files in the three files("rawvideo", "rawpic", and"selectedpic"). For example, "s15" means the first subject in the present database (third column). The third sheet includes the naming rules of every single file. For example, the file "15_0101disgustingteeth" means the facial recordings of subject 1 when watching the video number (disgust1, the first disgusting video).

4 Database Baseline Evaluation

In this section, the LBP method [51][41] was employed for macro-expression and micro-expression spotting evaluation. In Section 4.3, LBP-TOP [32] histograms were employed for the expression recognition evaluation.

4.1 LBP and LBP-TOP

LBPs [51] are used on gray images to extract texture features. Given the pixel c in the gray image, its LBP code is computed by comparing it with its P neighbors p. The neighbors lie on a circle with the center c and a radius equal to R.

$$
\begin{aligned}
LBP_{P,R} &= \sum_{p=0}^{P-1} s(g_p - g_c)2^p \\
s(u) &= \begin{cases} 1 & u \geq 0; \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
\tag{2}
$$

where g_c is the gray value of the given pixel c, and g_p is the value of its neighbor p. If the coordinates of c are (x_c, y_c), the coordinates of p are $(x_c + Rcos(2\pi p/P), y_c - Rsin(2\pi p/P))$. The coordinates of the neighbors that do not exactly fall on pixels are approximated by bilinear interpolation. The LBP encoding process is illustrated in Fig. 3.

LBP only extracts features of 2D objects. To extract features of 3D object, Zhao *et al.* proposed the dynamic LBP-TOP, which is a dynamic texture operator extended from LBP [32].

TABLE 5
Criteria for labeling the emotions and the frequency in the present database

Emotion Category	Criteria	Number	Macro-expression	Micro-expression
Positive	AUs needed for Happiness, At least AU6 or AU12 was present	124	116	8
Negative	AUs needed for Anger, Disgust, Sadness, Fear	126	105	21
Surprise	At least AU 1+2, AU 25, or AU 2 was present	25	16	9
Others	Other facial movements*	82	63	19

*Others include facial expressions that cannot be classified into basic emotions, such as tense and control, hurt, sympathy, confusing and helpless. Consistent with previous studies, emotion labeling are *partly* based on the AUs because micro-expressions are usually partial and in low intensity. In addition, the participants' self-reports on every single facial movements and the content of the video episodes were also considered.

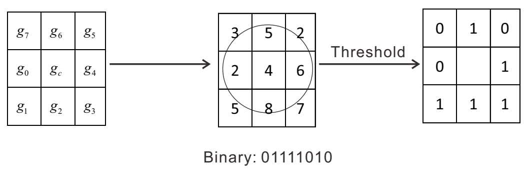


Fig. 3. Example of a basic LBP operator.

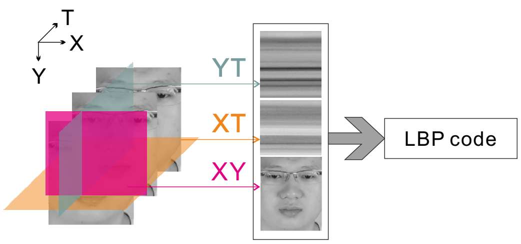


Fig. 4. Illustration of a spatiotemporal volume of a video, the XY plane (original frames) and the resulting temporal planes for LBP feature extraction.

Fig. 4 shows a micro-expression video clip and illustrates three orthogonal planes (XY, XT, and YT planes) from a single row and column of the clips. The LBP-TOP code is calculated by concatenating the LBP codes from the planes. The XT and YT planes encode the vertical motion patterns and the horizontal motion patterns, respectively. LBP-TOP will be employed for expression recognition evaluation in Section 4.3.

4.2 Expression Spotting Evaluation

A published LBP method [41] was employed to calculate the differences in the appearance-based features of the video frames within an alterable interval and automatically estimate the spotting of movements from videos, which can obtain both spatial and temporal locations. Due to relatively large head movements, 28 videos were not employed in the automatic spotting evaluation (see Table 6). The eyes were taken as the tracked points to perform nonreflective similarity transformation and accomplish facial alignment. After facial cropping based on the method [52], the face image was divided into 6 x 6 block structure (see Fig. 5). LBP histograms were calculated for each block.

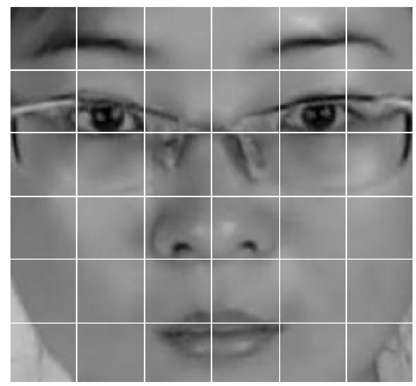


Fig. 5. An example of 6x6 block structure of a frame.

The average feature frame (AFF) represented the average of the features of the tail frame (TF) and the head frame (HF). HF denotes the k-th frame prior to the currently analyzed frame (CF), and TF denotes the k-th frame after the CF. The difference between AFF and CF was obtained by calculating the chi-squared distance of each pair of LBP histograms for every block; a total of 36 difference values were calculated due to the 6x6 block structure. CF valued all n frames of a video, with the exception of the first and last k frames of the video. The 36 pairs of difference values of each frame were arranged in descending order. F_i was defined as the average of the M greatest difference values of the i-th frame from all n frames, and M was set to 12 in this experiment. To avoid noise interference, we employed the contrasted difference vector C_i instead of F_i to represent the difference values with

$$C_i = F_i - \frac{1}{2}\left(F_{i+k} + F_{i-k}\right) \tag{3}$$

The threshold was used to obtain the peaks that represent the largest facial movement frames of the video. The threshold was the dashed line shown in Fig. 6a and Fig. 6b, which was measured by

$$T = C_{mean} + p \times (C_{max} - C_{mean}) \tag{4}$$

C_{mean} and C_{max} demote the average difference value and the maximum difference values, respectively , of the entire video, and p is a variable parameter in the range $[0, 1]$. We regarded spotted peak frames as valid

TABLE 6
Numbers of 28 videos not used in the automatic spotting evaluation

Number of video	Number of Subject	Number of emotion video	Number of expression sample
1	16	05	02
2	21	01	01
3	27	04	01
4	27	04	02
5	27	01	01
6	27	05	02
7	27	05	07
8	30	05	05
9	31	05	07
10	32	04	01
11	32	04	02
12	32	01	01
13	32	05	02
14	32	05	03
15	32	05	05
16	32	05	07
17	32	05	08
18	33	01	02
19	33	04	02
20	34	04	01
21	34	05	03
22	36	04	01
23	36	05	05
24	37	01	01
25	37	04	02
26	37	05	02
27	37	05	05
28	37	05	07

if they fell within the span of $k/2$ before or after the onset or offset of the provided truth frames. This method was used to spot both macro-expressions and micro-expressions. Due to serious head movement, 28 videos (i.e., 113 expressions) were eliminated. The 28 videos are listed in Table 6. A total of 59 videos (i.e., 244 expressions) were used for the automatic spotting test. For micro-expression spotting, k was set to 12 and p was set to 0.25. Fig. 6a shows the difference values on the vertical axis and the frame numbers on the horizontal axis for all n frames. As a result, 69.8% spotted peaks were valid when eye blinks and macro-expressions were also treated as true results in this experiment, and 47.3% micro-expressions were spotted. For spotting macro-expressions, k was set to 100 due to a longer expression duration, and the P value was set to 0.45, which was presented in the same manner as shown in Fig. 6b, in which three positive macro-expressions were spotted. Eventually, 70.1% spotted peaks were valid when eye blinks were also treated as true results in this experiment, and 75.7% macro-expressions were spotted.

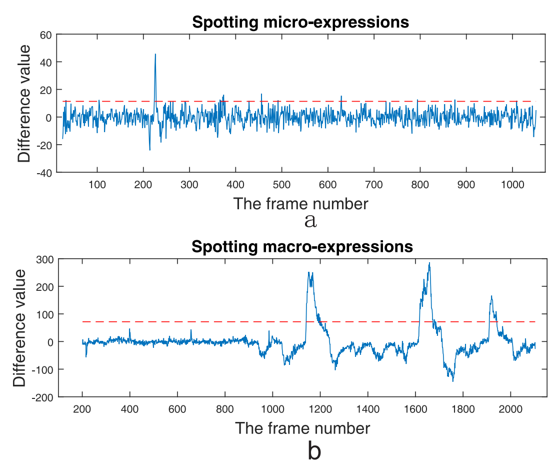


Fig. 6. An example of a contrasted difference vector of all n frames from a video for micro-expressions(a) and macro-expressions(b).

If we vary T from 0 to C_{max}, we obtain the ROC curve in Fig. 7. When the false positive rate is small, the slope of the curve is large. When false positive rate is large, the curve is parallel to the line whose angle with the abscissa is 45 degrees. This finding reveals that when the false positive rate is large, the predications

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TAFFC.2017.2654440

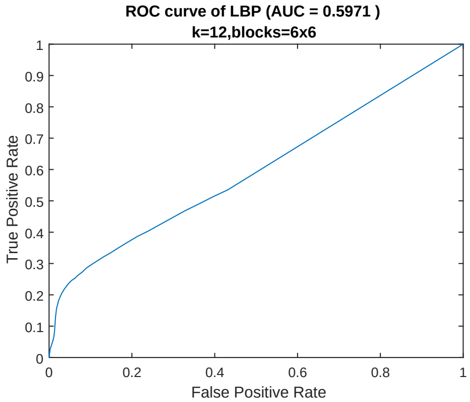


Fig. 7. ROC curve.

of LBP occur at the random level. For spotting facial movements, the random probability of predication is 0.5 (movement or no movement).

We employed the area under curve (AUC) to evaluate the performance of the spotting expression. The AUC is a common evaluation metric for binary classification problems. Consider a plot of the true positive rate vs the false positive rate as the threshold value for classifying an item as 0 or is increased from 0 to 1. If the classifier is very good, the true positive rate will rapidly increase and the AUC will be close to 1. If the classifier is not better than random guessing, the true positive rate will linearly increase with the false positive rate and the AUC will be approximately 0.5.

The numbers of the blocks are 5×5, 6×6, 7×7 and 8. The values of k are 6, 12, 18, 24, 48 and 96. The same experiments are performed, the AUCs are listed in Table 7. As the blocks and k value increase, the AUC increases.

TABLE 7
AUCs on various blocks and k

Blocks	5×5	6×6	7×7	8×8
$k=6$	0.5579	0.5608	0.5562	0.5626
$k=12$	0.5927	0.5971	0.5982	0.6011
$k=18$	0.6100	0.6157	0.6205	0.6208
$k=24$	0.6157	0.6251	0.6268	0.6300
$k=48$	0.6497	0.6559	0.6585	0.6639
$k=96$	0.6564	0.6597	0.6619	0.6606

4.3 Expression Recognition Evaluation

To evaluate the database, we employed LBP-TOP histograms [32] to extract dynamic textures and a support vector machine (SVM) approach to classify these dynamic textures.

To address the large variations in the spatial appearance of faces, all faces were normalized to a template face by registering 68 facial landmark points detected using the ASM [50]. First, we selected the frontal face image M with a neutral expression as the template, and the coordinates of 68 landmarks of the template face were detected by ASM as ψ_M. Second, for a sample micro-expression clip we detected the 68 facial landmarks on its first frame as ψ_{f1}, and estimated the 2D geometric transformation of the template face and the current given sample face as: $\psi_M = T\psi_{f1}$, where T is the transformation matrix. Third, we registered the sample face to the template by applying the transformation T to all frames of the micro-expression clip. Because the head movement in the video clip is not significant, the transformation T can be used for all frames in the same video clip. The sizes of each frame of samples are normalized to 163×134 pixels.

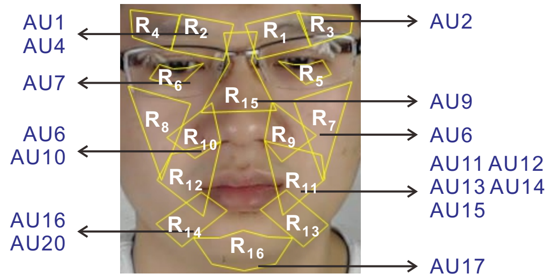


Fig. 8. The template face and 16 ROIs.

Because LBP-TOP is a local feature extraction method, facial images need to be divided into serval patches. We selected a frontal neutral facial image as the template face and divided the template face into 16 regions of interest (ROIs). Each ROI corresponds to one or more AUs. Fig. 8 shows the template face, in which the 16 ROIs and the AUs corresponds to the ROIs [53].

In Part B, we use 356 video clip samples (face in one video clip cannot be fitted by the ASM; video number "disgust2_1" from the third participant). Among the 356 samples from 22 subjects, four frames are included in the shortest sample, and the longest sample contains 117 frames. The frame numbers of all samples were normalized to 120 via linear interpolation. Here, we employed leave-one-subject-out (LOSO) cross-validation, i.e., in each fold, one participant was employed as the test set, and the other participants were employed as the training set. After the analysis of 22 folds, each participant had been used as the test set once, and the final recognition accuracy was calculated based on all results.

We extracted LBP-TOP to represent the dynamic texture features for each ROI and construct histograms. Then, the histograms concatenated a vector as an input for the classifier. A SVM classifier was selected, and the RBF kernel was employed as the kernel function. For LBP-TOP, the radii in axes X and Y (denoted as R_x and R_y) ranged from 1 to 4, and the radii in axes T (denoted as R_t) was assigned various values from 2 to 4. The number of neighboring points (denoted as P) in the XY, XT and YT planes were set to four and eight. The uniform pattern and the basic LBP were employed in LBP coding. Table 8 lists the results.

The lowest accuracy is 28.09% when $R_x = 4, R_y = 4$, and $R_t = 2$. The best accuracy is 40.95% when $R_x = 1, R_y = 1$, and $R_t = 4$. The recognition performance is sensitive to the parameters of LBP. In most cases, the performances of $P = 8$ are better than the performance of $P = 4$. The performance difference between $P = 8$ and $P = 4$ is approximately 8% when $R_x = 4, R_y = 4$, and $R_t = 2$. The larger is the number of neighboring points, the better is the performance of LBP. The $P = 8$ condition has more points than the $P = 4$ condition and a higher sampling rate, which generates a higher performance in the $P = 8$ condition.

TABLE 8
Performance on recognizing expressions

		Basic LBP		Uniform Pattern	
		$P=4$	$P=8$	$P=4$	$P=8$
$R_x=1, R_y=1$	$R_t=2$	37.24	37.44	36.40	33.13
	$R_t=3$	38.96	37.02	38.91	38.34
	$R_t=4$	36.68	40.95	36.76	40.83
$R_x=2, R_y=2$	$R_t=2$	31.36	31.56	30.63	32.14
	$R_t=3$	32.78	32.68	32.43	32.80
	$R_t=4$	31.45	33.01	31.70	32.50
$R_x=3, R_y=3$	$R_t=2$	33.33	34.12	33.02	34.88
	$R_t=3$	31.84	31.30	31.53	34.14
	$R_t=4$	32.92	31.55	33.28	35.26
$R_x=4, R_y=4$	$R_t=2$	28.50	36.05	28.09	32.66
	$R_t=3$	30.31	34.41	30.58	34.45
	$R_t=4$	32.43	37.69	31.46	36.91

5 Discussion and Conclusion

In this paper, we describe a new facial expression database, the CAS(ME)^2, which contains both expression samples in long video streams (Part A) that can be employed for macro-expression and micro-expression spotting from long videos and cropped expression samples with frames from onset to offset (Part B, 357 expression samples, comprising 300 macro-expression samples and 57 micro-expression samples) that can be employed for automatic expression recognition training. These expression samples were collected from the same individuals in the same experimental conditions. This database may enable researchers to develop more efficient algorithms to extract features that are capable of spotting and discriminating macro-expressions and micro-expressions from long videos.

Considering the unique features of micro-expressions, which occur rapidly, partially (on either the upper face or the lower face) and are low in intensity, the emotional labeling of these facial expressions is only based on the corresponding AUs, and the emotion types that are associated with the videos that evoked them may not be sufficiently precise. To improve the validity of emotional labeling when constructing the presented database, we also collected self-reports of the participants' emotions for each expression sample. The emotion labeling of each sample was based on a combination of the FACS AUs, the emotion type of the emotion-evoking videos and the self-reported emotions. This labeling method should considerably enhance the precision of the emotion category assignment. In addition, the three labels are independent and will be accessible when the database is published, to enable researchers to access specific expressions in the database.

Because the movement of micro-expression is very subtle, automatic detection of micro-expression is significantly influenced by factors such as illumination and head movements, which invoke a larger change than micro-expression-related facial changes. These factors make the automatic detection of micro-expression challenging. In addition, micro-expressions tend to occur with eye blinks, which increases the difficulty of correctly detecting micro-expressions. Therefore, the spotting rate of micro-expressions is very low (the highest result was approximately 40% in this evaluation).

Because research on the automatic detection of micro-expression is developing, we attempted to begin with a database in which data were collected in relatively rigid and controlled experimental settings (controlled illumination and head movements). We achieved a rather low detection rate with our detection methods. This result implies that future studies may have to elaborate and improve the current methods to better detect micro-expressions. Due to the inherent low and subtle intensity of micro-expressions, future studies may have to approach and develop new methods for addressing micro-expression detection problems in long videos.

Due to the difficulties encountered in micro-expression elicitation and the extremely time-consuming nature of manual coding, the size of the micro-expression sample pool in the current version of the database may not be completely sufficient. In the previous micro-expression databases, the sample size was usually small (e.g., USF-HD with 100 samples, SMIC with 77 samples). In the current database, we established stricter criteria for selecting the samples. We removed facial movements without emotion meaning according to the participants' self-reported emotions, such as nose blowing and eye blinking. This operation reduces the sample size but creates a ”cleaner” database. Regarding the issue of frame rate, a 30 fps video camera was employed and this is relatively low frame rate compared with those used in recently published databases. We intend to enrich the sample pool by eliciting additional micro-expression samples to provide researchers with sufficient testing and training data and to use cameras with higher frame rates in the future.The CAS(ME)^2 database is online public available now for testing (see http://fu.psych.ac.cn/CASME/cas(me)2-en.php for details).

REFERENCES

[1] N. Michael, M. Dilsizian, D. Metaxas, and J. K. Burgoon, "Motion profiles for deception detection using visual cues," in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 462–475.

[2] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," *Psychiatry*, vol. 32, no. 1, pp. 88–106, 1969.

[3] P. Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised Edition)*. WW Norton & Company, 2009.

[4] Y. Tong, J. Chen, and Q. Ji, "A unified probabilistic framework for spontaneous facial action modeling and understanding," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 2, pp. 258–273, 2010.

[5] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Transactions on affective computing*, vol. 1, no. 1, pp. 18–37, 2010.

[6] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 6, pp. 1113–1133, 2015.

[7] P. Ekman, *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. Macmillan, 2007.

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

[9] P. Ekman, "Darwin, deception, and facial expression," *Annals of the New York Academy of Sciences*, vol. 1000, no. 1, pp. 205–221, 2006.

[10] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, "How fast are the leaked facial expressions: The duration of micro-expressions," *Journal of Nonverbal Behavior*, pp. 1–14, 2013.

[11] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*. IEEE, 1998, pp. 200–205.

[12] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression database," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 12, pp. 1615–1618, 2003.

[13] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[14] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, "Toward practical smile detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 11, pp. 2106–2111, 2009.

[15] R. Cowie, E. Douglas-Cowie, and C. Cox, "Beyond emotion archetypes: Databases for emotion modelling using neural networks," *Neural networks*, vol. 18, no. 4, pp. 371–388, 2005.

[16] M. S. Bartlett, G. C. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan, "Automatic recognition of facial actions in spontaneous expressions," *Journal of multimedia*, vol. 1, no. 6, pp. 22–35, 2006.

[17] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, pp. 5–pp.

[18] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000, pp. 46–53.

[19] T. Bänziger and K. R. Scherer, "Introducing the geneva multimodal emotion portrayal (gemep) corpus," *Blueprint for affective computing: A sourcebook*, pp. 271–294, 2010.

[20] K. L. Schmidt and J. F. Cohn, "Human facial expressions as adaptations: Evolutionary questions in facial expression research," *American journal of physical anthropology*, vol. 116, no. S33, pp. 3–24, 2001.

[21] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "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*. IEEE, 2010, pp. 94–101.

[22] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, "Disfa: A spontaneous facial action intensity database," *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 151–160, 2013.

[23] D. McDuff, R. El Kaliouby, T. Senechal, M. Amr, J. F. Cohn, and R. Picard, "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*. IEEE, 2013, pp. 881–888.

[24] S. Wang, Z. Liu, Z. Wang, G. Wu, P. Shen, S. He, and X. Wang, "Analyses of a multimodal spontaneous facial expression database," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 34–46, 2013.

[25] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard, "Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database," *Image and Vision Computing*, vol. 32, no. 10, pp. 692–706, 2014.

[26] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor," in *3rd International Conference on Crime Detection and Prevention*. IET, 2009, pp. 1–6.

[27] T. Pfister, X. Li, G. Zhao, and M. Pietikainen, "Recognising spontaneous facial micro-expressions," in *12th IEEE International Conference on Computer Vision*. IEEE, 2011, pp. 1449–1456.

[28] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikainen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–6.

[29] W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, and X. FU, "CASME Database: A dataset of spontaneous micro-expressions collected from neutralized faces," in *10th IEEE Conference on Automatic Face and Gesture Recognition*, 2013, pp. 1–7.

[30] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," *PLoS ONE*, vol. 9, no. 1, p. e86041, 01 2014.

[31] S.-J. Wang, H.-L. Chen, W.-J. Yan, Y.-H. Chen, and X. Fu, "Face recognition and micro-expression based on discriminant tensor subspace analysis plus extreme learning machine," *Neural Processing Letters*, 2013.

[32] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[33] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, and X. Fu, "Micro-expression recognition using dynamic textures on tensor independent color space," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014, pp. 4678–4683.

[34] S.-J. Wang, W.-J. Yan, G. Zhao, X. Fu, and C.-G. Zhou, "Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features," in *Computer Vision - ECCV 2014 Workshops*, ser. Lecture Notes in Computer Science, L. Agapito, M. M. Bronstein, and C. Rother, Eds. Springer International Publishing, 2015, vol. 8925, pp. 325–338. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-16178-5_23

[35] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Advances in neural information processing systems*, 2009, pp. 2080–2088.

[36] Y. J. Liu, J. K. Zhang, W. J. Yan, S. J. Wang, G. Zhao, and X. Fu, "A main directional mean optical flow feature for spontaneous micro-expression recognition," *IEEE Transactions on Affective Computing*, vol. 7, no. 4, pp. 299–310, Oct 2016.

[37] F. Xu, J. Zhang, and J. Wang, "Microexpression identification and categorization using a facial dynamics map," *IEEE Transactions on Affective Computing*, vol. PP, no. 99, pp. 1–1, 2016.

[38] S.-J. Wang, W.-J. Yan, T. Sun, G. Zhao, and X. Fu, "Sparse tensor canonical correlation analysis for micro-expression recognition," *Neurocomputing*, vol. 214, pp. 218 – 232, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231216305501

[39] M. Shreve, S. Godavarthy, V. Manohar, D. Goldgof, and S. Sarkar, "Towards macro-and micro-expression spotting in video using strain patterns," in *Applications of Computer Vision (WACV), 2009 Workshop on*. IEEE, 2009, pp. 1–6.

[40] S. Polikovsky, Y. Kameda, and Y. Ohta, "Detection and measurement of facial micro-expression characteristics for psychological analysis," *Kameda's Publication*, vol. 110, pp. 57–64, 2010.

[41] A. Moilanen, G. Zhao, and M. Pietikainen, "Spotting rapid facial movements from videos using appearance-based feature differ-

ence analysis," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014, pp. 1722–1727.

[42] T. Bänziger and K. R. Scherer, *Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 476–487. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-74889-2_42

[43] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, Jan 2009.

[44] J. F. COHN and K. L. SCHMIDT, "The timing of facial motion in posed and spontaneous smiles," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 02, no. 02, pp. 121–132, 2004. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S021969130400041X

[45] M. F. Valstar, H. Gunes, and M. Pantic, "How to distinguish posed from spontaneous smiles using geometric features," in *Proceedings of the 9th International Conference on Multimodal Interfaces*, ser. ICMI '07. New York, NY, USA: ACM, 2007, pp. 38–45. [Online]. Available: http://doi.acm.org/10.1145/1322192.1322202

[46] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," *Cognition and Emotion*, vol. 9, no. 1, pp. 87–108, 1995. [Online]. Available: http://dx.doi.org/10.1080/02699939508408966

[47] F. Qu, S.-J. Wang, W.-J. Yan, and X. Fu, *CAS(ME)2: A Database of Spontaneous Macro-expressions and Micro-expressions*. Cham: Springer International Publishing, 2016, pp. 48–59. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-39513-5_5

[48] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[49] A. C. Le Ngo, R. C.-W. Phan, and J. See, "Spontaneous subtle expression recognition: Imbalanced databases and solutions," in *Computer Vision–ACCV 2014*. Springer, 2014, pp. 33–48.

[50] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

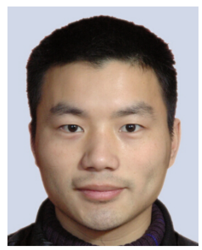
[51] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[52] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3444–3451.

[53] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, C.-G. Zhou, X. Fu, M. Yang, and J. Tao, "Micro-expression recognition using color spaces," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6034–6047, 2015.

Fangbing Qu is now a lecturer in College of Preschool Education, Capital Normal University. His interests include facial expression (especially micro-expression), deception detection and facial expression self-monitor. He is now trying interdisciplinary researches among behavioral science, affective computing and pattern recognition.

Su-Jing Wang received the Master's degree from the Software College of Jilin University, Changchun, China, in 2007. He received the Ph.D. degree from the College of Computer Science and Technology of Jilin University in 2012. He was a postdoctoral researcher in Institute of Psychology, Chinese Academy of Sciences from 2012 to 2015. He is now an Assistant Researcher in Institute of Psychology, Chinese Academy of Sciences. He has published more than 40 scientific papers. He is One of Ten Selectees of the Doctoral Consortium at International Joint Conference on Biometrics 2011. He was called as *Chinese Hawkin* by the Xinhua News Agency. His current research interests include pattern recognition, computer vision and machine learning. He serves as an associate editor of Neurocomputing (Elsevier). For more information, visit http://sujingwang.name.



Wen-Jing Yan received his Ph.D degree from Institute of Psychology, Chinese Academy of Sciences, Beijing, China, in 2014. He is now an Assistant Professor in Department of Psychology in Wenzhou University, China. His interests include facial expression and deception. His research interests include interdisciplinary research on facial expression and affective computing.

He Li is now a doctoral student in Institute of Psychology, Chinese Academy of Sciences. His main research interest is deception detection by the analysis of facial expression and body posture. His research methods include behavioral analysis, eye tracking and motion capture analysis.

Shuhang Wu is now pursuing master's degree in College of Information Science and Engineering, Northeastern University. She is interested in Computer Vision and Machine Learning. Her current research is trying to seek more accurate algorithm for micro-expression detection and recognition.

Xiaolan Fu received her Ph. D. degree in 1990 from Institute of Psychology, Chinese Academy of Sciences. Currently, she is a Senior Researcher at Cognitive Psychology. Her research interests include visual and computational cognition: (1) attention and perception, (2) learning and memory, and (3) affective computing. At present, she is the director of Institute of Psychology, Chinese Academy of Sciences and Vice Director, State Key Laboratory of Brain and Cognitive Science.