

•Review•

# Review of micro-expression spotting and recognition in video sequences

Hang PAN<sup>1</sup>, Lun XIE<sup>1\*</sup>, Zhiliang WANG<sup>1</sup>, Bin LIU<sup>2</sup>, Minghao YANG<sup>2</sup>, Jianhua TAO<sup>2</sup>

1. School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

2. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

\* Corresponding author, xielun@ustb.edu.cn

Received: 11 July 2020      Revised: 18 September 2020      Accepted: 15 October 2020

Supported by the National Key R&D Program of China (2017YFB1002804); National Natural Science Foundation of China (61672093); Beijing Municipal Natural Science Foundation (L192005).

**Citation:** Hang PAN, Lun XIE, Zhiliang WANG, Bin LIU, Minghao YANG, Jianhua TAO. Review of micro-expression spotting and recognition in video sequences. *Virtual Reality & Intelligent Hardware*, 2021, 3(1): 1—17  
DOI: 10.1016/j.vrih.2020.10.003

**Abstract** Facial micro-expressions are short and imperceptible expressions that involuntarily reveal the true emotions that a person may be attempting to suppress, hide, disguise, or conceal. Such expressions can reflect a person's real emotions and have a wide range of application in public safety and clinical diagnosis. The analysis of facial micro-expressions in video sequences through computer vision is still relatively recent. In this research, a comprehensive review on the topic of spotting and recognition used in micro-expression analysis databases and methods, is conducted, and advanced technologies in this area are summarized. In addition, we discuss challenges that remain unresolved alongside future work to be completed in the field of micro-expression analysis.

**Keywords** Facial expression; Micro-expression spotting; Micro-expression recognition; Database; Review

## 1 Introduction

With the advancements of new technologies such as artificial intelligence, machine learning, and computer vision, intelligent human-computer interaction has gradually become part of the daily lives of users, including smart speakers, unmanned driving, and accompanying robots. Future society will be intelligence based, and intelligent human-computer interaction will be applied to everyday human life. Emotionally intelligent human-computer interactions not only require the machine to complete tasks through different methods of interaction, they also require the machine to have emotional recognition, expression, and feedback capabilities similar to human interaction. Psychologists believe that 7% of human emotional expression is conveyed through language, 38% through speech, and the remaining 55% is conveyed through facial expressions<sup>[1]</sup>. Although facial expressions can reflect a person's mental state, people often disguise or deliberately express a certain facial expression in a specific situation. In this case, it is necessary to judge an individual's true emotional state based on facial micro-expressions.

Facial micro-expressions are short and imperceptible expressions that involuntarily reveal the true

emotions that an individual is attempting to suppress, hide, disguise, or conceal. They are difficult to modify by will and can reflect a person's actual emotional state. In 1966, Haggard and Isaacs first proposed the concept of micro-expressions when studying psychotherapy, and found the existence of short-lived, difficult-to-detect facial expressions<sup>[2]</sup>. In 1969, Ekman et al. observed a conversational video between a psychologist and a patient with depression, and found that the patient occasionally had a few very painful expressions when she tried to convince the doctor that she was no longer suicidal by smiling<sup>[3]</sup>. Researchers refer to the rapid, unconscious spontaneous facial movements that such people produce when they experience strong emotions as micro-expressions. Compared with facial expressions made consciously by humans, micro-expressions can better reflect real emotions. Because micro-expressions contain potentially true emotional information, they are greatly significant in high-risk scenarios<sup>[4–9]</sup>. The application scenario of micro-expression is shown in Figure 1.



**Figure 1 Application scenarios of micro-expressions: lie detection, criminal investigation, clinical medical treatment.**

Micro-expressions are short-duration, low-intensity facial expressions that usually appear when people deliberately or unconsciously attempts to hide their true emotions<sup>[10]</sup>. It is therefore not easy to identify such real emotional information. Among such difficulties, one of the main problems is that the duration of a micro-expression is extremely short, which is sustained with in 0.04s to 0.2s<sup>[11]</sup>, and some studies have shown that the duration of a micro-expression is less than 0.33s, and at most does not exceed 0.5s<sup>[12]</sup>. This rapid appearance and disappearance bring about significant challenges to the location and recognition of micro-expressions. Micro-expressions also have a low-intensity and are only related to the part of the facial expression action units<sup>[13]</sup>. Therefore, micro-expressions are easily overlooked by the human eye. Without professional training, most subjects cannot recognize micro-expressions quickly. To help humans learn to perceive and understand micro-expressions, Ekman et al. designed the Micro-Expression Training Tool (METT)<sup>[14]</sup> to allow subjects to identify micro-expressions. By using METT for extensive and repeated training, the average person can learn to recognize seven basic micro-expressions. However, Frank et al. found that the overall recognition rate of micro-expressions by people trained in METT only reaches approximately 40%<sup>[15]</sup>.

With the rapid development of visual sensors, computer vision and video processing technology has become a new direction for researchers to apply facial micro-expression recognition in clinical treatment or high-risk environments. Spontaneous facial expression recognition for different research backgrounds and application scenarios is a popular research topic<sup>[16–21]</sup>. Meanwhile, the Facial micro-expressions Grand Challenge (MEGC) promoted the development of micro-expressions<sup>[22–25]</sup>. However, micro-expression spotting and recognition using video is a new topic with additional challenges. First, it is difficult to collect micro-expression datasets through spontaneous induction, and current data samples are few in number. It is also difficult to accurately determine the time at which a micro-expression occurred from a video sequence. Because micro-expressions are subtle and fast, finding micro-expressions in longer video sequences is not easy. Finally, because changes in micro-expressions are only related to a few facial action units, and the intensity of the change is very low, this causes difficulties in the recognition of micro-expression.

In recent years, some developments have emerged in micro-expression analysis through computer

technology. In contrast to other micro-expression reviews, in this article, we divide micro-expression analysis into three parts: micro-expression datasets, micro-expression spotting, and video sequence recognition. This article is organized as follows. Section 2 introduces micro-expression datasets. Section 3 introduces the micro-expression spotting method in long video sequences. Section 4 introduces the micro-expression recognition method. Finally, Section 5 summarizes and discusses the challenges for micro-expression analysis and future areas worth studying.

## 2 Micro-expression datasets

A prerequisite of micro-expression analysis is sufficient data with affective labels. However, research on micro-expression spotting and recognition based on computer vision has just begun to attract attention, and few spontaneous micro-expression datasets have been published to date. Table 1 summarizes all available micro-expression datasets, which include two types, posed and spontaneous, used for spotting and the recognition of micro-expressions.

**Table 1 Summary of posed and spontaneous micro-expression datasets**

Dataset	Dataset-Sub	Subjects	Samples	FACS	FPS	Classes	Resolution	Frame annotations	Type
Polikovsky <sup>[26]</sup>		11	42	No	200	6	640 × 480	–	Posed
USF-HD <sup>[27]</sup>		–	100	No	29.7	4	1280 × 720	–	Posed
YorkDDT <sup>[28]</sup>		9	18	No	25	2	320 × 240	–	Posed
SMIC-sub <sup>[29]</sup>		6	77	No	100	3	640 × 480	–	Spontaneous
SMIC <sup>[30]</sup>	HS	16	164	No	100	3	640 × 480	–	Spontaneous
	VIS	8	71	No	25	3	640 × 480	–	Spontaneous
	NIR	8	71	No	25	3	640 × 480	–	Spontaneous
	E-HS	16	157	No	100	3	640 × 480	Onset, offset	Spontaneous
	E-VIS	8	71	No	25	3	640 × 480	–	Spontaneous
	E-NIR	8	71	No	25	3	640 × 480	–	Spontaneous
CASME <sup>[31]</sup>	Section A	7	96	Yes	60	7	1280 × 720	–	Spontaneous
	Section B	12	101	Yes	60	7	640 × 480	–	Spontaneous
CASME II <sup>[32]</sup>		35	247	Yes	200	5	640 × 480	Onset, offset, apex	Spontaneous
CAS(ME) <sup>2[33]</sup>		22	357	No	30	4	640 × 480	Onset, offset, apex	Spontaneous
SAMM <sup>[34]</sup>		32	159	Yes	200	7	2040 × 1088	Onset, offset, apex	Spontaneous
SAMM Long Videos <sup>[35]</sup>		32	502	Yes	200	2	2040 × 1088	Onset, offset, apex	Spontaneous

For a posed micro-expression, the subjects imitate the facial micro-expression which cannot counter the current emotional state of the subject. Therefore, it does not help to recognize genuine subtle emotions. Spontaneous micro-expressions are uncontrolled and consistent with real emotions<sup>[36]</sup>. During the collection of spontaneous micro-expression datasets, the subjects watched inductive video clips to stimulate a true emotional state. Simultaneously, they were required to suppress their true emotions or risk punishment. Ekman et al. suggested that micro-expressions are unconscious and involuntary. Therefore, the posed micro-expressions usually do not show spontaneous micro-expression characteristics<sup>[3,4]</sup>.

### 2.1 Posed micro-expression datasets

This section presents a review of earlier research, which relied on posed micro-expression datasets.

#### 2.1.1 Polikovsky dataset

In 2009, Polikovsky et al. created the first micro-expression dataset<sup>[26]</sup>. This dataset included 42 video

sequences of 11 college students captured through a camera with a pixel resolution of  $640 \times 480$ , at 200 fps in a laboratory environment. The ethnic distribution of the subjects was reasonable, including five Asians, four Caucasians, and one Indian. Participants were requested to generate six basic expressions with low-intensity facial muscle movement, and to quickly return to a neutral facial expression. Therefore, such images are not considered representative of real world situations. In addition, the dataset has not been made public for further study.

### 2.1.2 USD-HD dataset

The USD-HD<sup>[27]</sup> dataset is similar to the Polikovsky dataset in that it collects micro-expression data through the way the subjects play. During the process of data collection, the subjects were shown various micro-expression images and then asked to imitate each one smoothly. The dataset collected 100 micro-expression samples at 29.7 fps, including four categories: smile, surprise, anger, and sadness. However, the dataset has not been made public.

### 2.1.3 YorkDDT dataset

The York Deception Detection Test (YorkDDT) is used in psychological research. Warren et al. recorded 20 video sequences of 9 subjects using a camera with a pixel resolution of  $320 \times 240$  at 25 fps<sup>[28]</sup>. The participants used truthful or deceptive situations to describe two types of emotional or non-emotional movie clips. The emotional clips are stressful and unpleasant scenes. Non-emotional clips are neutral, pleasant scenes. The participants were required to observe the emotional clip and then describe them as non-emotional segments, and viewing non-emotional clips requires that they are described as emotional segments. Warren et al. believe that micro-expressions will occur in both cases, but they are not public. This dataset is mainly used for automatic deception recognition.

## 2.2 Spontaneous micro-expression datasets

Because micro-expressions are difficult to fake, when we collect spontaneous micro-expression video clips, subjects must hide their true emotional state. It is therefore difficult to collect spontaneous micro-expression datasets. Thus far, spontaneous micro-expression datasets include SMIC-sub<sup>[29]</sup>, SMIC<sup>[30]</sup>, SMIC-E CASME<sup>[31]</sup>, CASME II<sup>[32]</sup>, CAS(ME)<sup>[33]</sup>, SAMM<sup>[34]</sup>, and SAMM Long Videos<sup>[35]</sup>.

### 2.2.1 Spontaneous micro-expression corpus (SMIC) series datasets

Micro-expressions are important clues for the analysis of deception and detection. In 2011, Pfister et al. created the first spontaneous micro-expression dataset SMIC-sub to identify micro-expressions in deception detection<sup>[29]</sup>. The SMIC dataset induces spontaneous micro-expressions created by having subjects watch video clips with emotional stimulation, and requiring the subjects to hide their true feelings when watching the clips. If the subjects were unable to hide their true emotional state, they were required to fill out a long and boring questionnaire as a penalty. The authors believed that this setting would create a high-risk environment for lying and better induce micro-expressions. The SMIC-sub recorded 77 spontaneous micro-expression videos of 6 subjects (3 males and 3 females) using a 100 fps high-speed camera.

They released the full version of the SMIC dataset in 2013. The SMIC recorded 328 video sequences of 20 subjects and found 164 spontaneous micro-expression samples from 16 subjects. The subject was alone watching the stimulating emotional video, with the participant's facial video data remotely observed from another room. All video data had a pixel resolution of  $640 \times 480$ , captured at 100 fps using a high-speed (HS) camera. In the recording of the last 10 subjects, a 25-fps visual camera (VIS) and a near-infrared camera (NIR) were also used. Therefore, the SMIC database contains three sub-datasets, namely, SMIC-

HS, SMIC-VIS, and SMIC-NIR, which are used for micro-expression analysis through multi-source data; however, because the data of SMIC-VIS and SMIC-NIR are incomplete, they are not commonly used in micro-expression analysis. The SMIC database contains three emotion categories:

**Table 2 SMIC emotion category distribution**

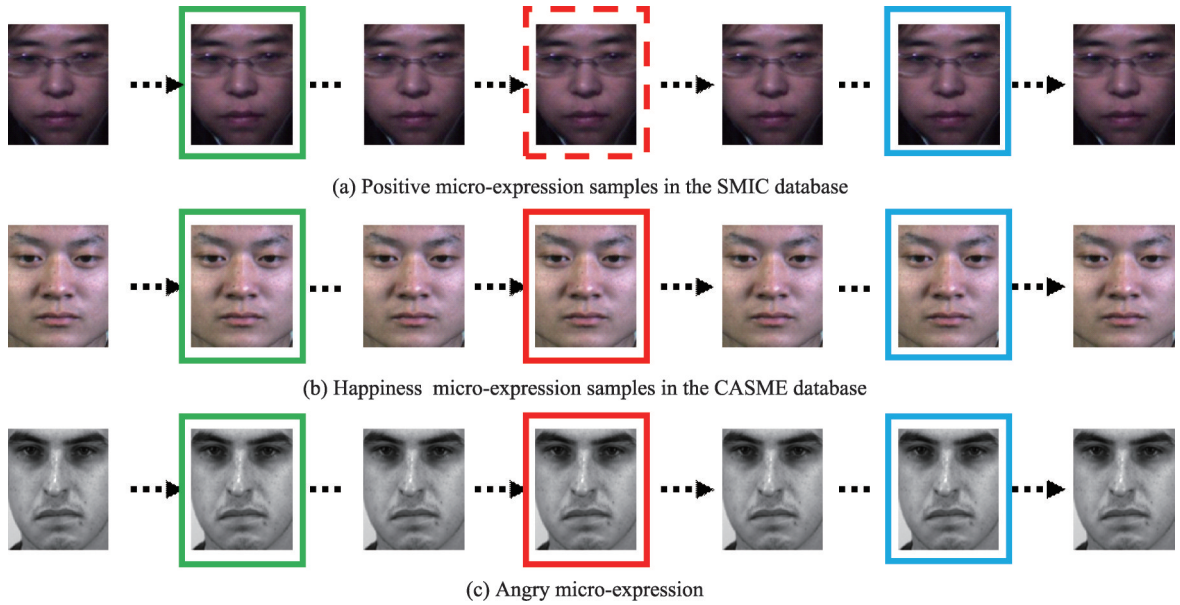
Dataset	Positive	Negative	Surprise	Total
HS	51	70	43	164
VIS	28	23	20	71
NIR	28	23	20	71

positive, negative, and surprised, sample category distributions are shown in Table 2.

The authors subsequently released an extended version of SMIC (SMIC-E), which includes a complete long video sequence containing non-micro-expression video frames before and after the start position and end position of the marked micro-expression, respectively. SMIC-E is similar to SMIC and has three sub-databases, namely, SMIC-E-HS, SMIC-E-VIS, and SMIC-E-NIR. The SMIC-E-VIS and SMIC-E-NIR databases contain 71 video clips, and the SMIC-E-HS dataset contains 157 video clips with an average length of 5.9s. Compared with the original SMIC, SMIC-E can be used for micro-expression positioning based on the start and end positions of the micro-expression.

### 2.2.2 Chinese academy of sciences micro-expressions (CASME) series datasets

In 2013, during the same period when the full version of the SMIC was released, Yan et al. created a more comprehensive dataset, CASME<sup>[31]</sup>. A total of 35 subjects (13 women and 22 men) participated in this experiment. The dataset contains 195 micro-expression sample video sequences captured using a 60-fps camera. A sample from this CASME dataset is shown in Figure 2. According to the environment settings and the camera, these clips are divided into two classes. Section A contains 96 samples. All samples were recorded at 60 fps using a BenQ M31 camera. Natural light was used for recording. Section B contains 101 samples. A GRAS-03K2C camera recorded such samples at 60 fps with a resolution of  $640 \times 480$ .



**Figure 2** Data displayed in the public micro-expression datasets, where green is the onset frame, red is the apex frame, and blue is the offset frame. The apex frame index is not provided in the SMIC dataset<sup>[30,32,34]</sup>.

Although CASME contains a relatively comprehensive sample of micro-expressions, some of the videos are extremely short, with duration of less than 0.2s, which makes it difficult for a micro-expression analysis. Therefore, Yan et al. improved the CASME dataset and released a new dataset CASME II<sup>[32]</sup>. This dataset collected 247 video sequences. These samples were collected using a 200-fps camera with a



resolution of  $640 \times 480$ . This dataset contains 5 classes of micro-expression: happiness (32 samples), surprise (25 samples), disgust (64 samples), repression (27 samples), and others (99 samples). To be more certain of the accuracy of the micro-expression emotion label, all samples have the action unit (AU) marked based on the FACS<sup>[37]</sup>, and were selected from almost 3000 facial movements. To promote the application of micro-expression spotting in deception detection, Qu et al. proposed a new CAS(ME)2 dataset for the spotting and recognition of micro-expressions in long video sequences<sup>[33]</sup>. The dataset collected 98 video sequences involving 22 subjects, with 30 fps at a resolution of  $640 \times 480$ , including 57 micro-expression samples and labeled the onset, apex, and offset time index of the micro-expression.

The sample category distribution of the CASME series datasets is shown in Table 3. It can be observed that due to the influence of factors such as the induction methods and personalization, the category distribution of the dataset presents an imbalance, and the accuracy of recognition will be affected when a micro-expression analysis is conducted using the data-driven learning algorithm. This is also an important challenge in micro-expression recognition.

**Table 3 Emotion category distribution of CASME series datasets**

Dataset	Positive	Negative						Surprise	Others	Total
	Happiness	Despise	Disgust	Fear	Repression	Sad	Tension			
CASME	9	1	44	2	38	6	69	20		189
CASME II	32	–	63	2	27	7	–	25	99	255
CAS(ME)2	8				21			9	19	57

### 2.2.3 Spontaneous actions and micro-movements (SAMM) series datasets

In 2016, Davison et al. released the first high-speed, 200-fps, high-resolution,  $2040 \times 1088$  SAMM dataset<sup>[34]</sup>. By stimulating the subjects to produce micro-expressions, the subjects were told to cover up their emotions to the greatest extent prior to starting. A total of 32 subjects elicited emotions, and 159 video sequences were collected. The emotion classes and FACS labels of these samples were determined by trained experts. The sample category distribution of the SAMM is shown in Table 4.

**Table 4 Emotion category distribution of SAMM dataset**

Dataset	Happiness	Surprise	Angry	Disgust	Sad	Sad	Despise	Others	Total
SAMM	26	15	57	9	6	8	12	26	159

During the data collection process of the SAMM database, self-reports of the emotions of the subject will not be collected after the end of the experiment. Before the start of the experiment, each subject needed to complete a questionnaire to tailor the different types of video stimuli to them, increasing the chance of emotional arousal, and a specific video was selected to show to the subjects to obtain the best inducing potential. To introduce a high-risk situation and increase the probability of inducing a micro-expression, if a micro-expression was shown, the subject was given a bonus of 50 British pounds. In 2019, they released an updated database version of the SAMM Long Videos<sup>[35]</sup>. In the SAMM Long Videos dataset, there are 32 subjects and 147 videos, including 343 macro-expressions and 159 micro-expressions. The dataset gives the micro-expression onset, apex, and offset time index labels.

In general, SMIC, CASME II, CAS (ME) 2, SAMM, and SAMM Long Videos are considered state-of-the-art datasets for micro-expression spotting and recognition that should be widely adopted for research purposes.

Figure 2 shows the sample data in the current three public micro-expression series datasets: SMIC, CAMSE, and SAMM. Among these, SMIC, CASME II, and SAMM are used for micro-expression recognition, and SMIC-E, CAS(ME)2, and SAMM Long Videos are used for micro-expression spotting.

### 3 Micro-expression spotting

Automatic micro-expression analysis usually includes two tasks: spotting and recognition. Micro-expression spotting is the time interval at which micro-expressions are detected from the video sequence. Micro-expression recognition is used to classify micro-expressions that occur in video sequences. Among them, the spotting of micro-expressions in video sequences is a prerequisite for advanced facial recognition. Automatic micro-expression spotting is used to detect the onset, apex, offset frame, and neutral phase of the micro-expression in a video sequence. Valstar et al. suggested that the onset frame phase is the moment when the facial muscle movements begin to increase, the apex frame phase is when a facial expression develops to its most obvious moment, and the offset frame phase is when the facial muscles return to a neutral appearance<sup>[38]</sup>. There are few published studies on micro-expression spotting<sup>[39]</sup>, and these methods can be classified as appearance-based, dynamic, and general methods.

#### 3.1 Micro-expression spotting methods

Facial micro-expression spotting is used to automatically detect the time point when the micro-expression in a video sequence occurs, which refers to spotting the movement or time interval of the micro-expression. Table 5 summarizes the existing techniques used for spotting a facial micro-expression.

**Table 5 Summary of studies in facial micro-expression spotting**

Work	Feature	Spotting method	Datasets
Polikovsky et al., 2009 <sup>[26]</sup>	3D-HOG	K means	Polikovsky
Polikovsky et al., 2013 <sup>[40]</sup>	3D-HOG	K means	Polikovsky
Moilanen et al., 2014 <sup>[41]</sup>	LBP	Threshold technique	CASME, SMIC
Davison et al., 2015 <sup>[42]</sup>	HOG	Threshold technique	SAMM
Patel et al., 2015 <sup>[43]</sup>	Optical flow	Threshold technique	SMIC
Xia et al., 2016 <sup>[44]</sup>	Geometrical motion	Random walk model	CASME, SMIC
Li et al., 2017 <sup>[45]</sup>	HOOF, LBP	Threshold technique	CASME II, SMIC
Wang et al., 2017 <sup>[46]</sup>	MDMD	Threshold technique	CAS(ME)2
Davison et al., 2018 <sup>[47]</sup>	3DHOG, LBP, OF	Threshold technique	SAMM, CASME II
Li et al., 2019 <sup>[48]</sup>	LBP- $\chi^2$	Threshold technique	SAMM, CASME II

In micro-expression spotting, the duration of the micro-expression can be located by sliding the time window, and the duration is identified from the onset frame to the offset frame. Moilanen et al. used the local binary pattern (LBP) to extract the feature difference (FD) between each frame of the video sequence to analyze the changes in facial motion, and calculated the feature Chi-square ( $\chi^2$ ) to generate the magnitude of differences in the features<sup>[41]</sup>. By calculating a feature difference vector, the index of the apex frame is identified from the video sequence, and is compared with the ground truth of the apex frame index. If both fall within half of the frame interval of the sliding frame before the start and after the end, they are considered to be true positives. This method was tested on CASME-A, CASME-B, and SMIC-VIS-E, and the true positive rates were 52%, 66%, and 71%, respectively.

Davison et al. spotted micro-expressions through a histogram of oriented gradients<sup>[42,47]</sup>. They expressed all sequences detected in less than 100 frames as true positives, including blinks and fixations. Motion sequences that are detected but not encoded are classified as false positives. The recall rate, accuracy, and F1-measure of this method using the SAMM database were 0.84, 0.70, and 0.76, respectively.

Patel et al. proposed a method to calculate the optical flow vector over a small local area and integrate it into the spatiotemporal area to identify the start and offset times<sup>[43]</sup>. Xia et al. applied a random walk model

to calculate the probability of frames containing a micro-expression by considering the geometric deformation correlation between frames in the time window<sup>[44]</sup>. Tran et al. constructed a multi-scale evaluation benchmark based on a sliding window to fairly and better evaluate the micro-expression spotting approaches<sup>[49]</sup>.

Li et al. proposed a micro-expression analysis system (MESR), which can spot and recognize micro-expressions in video sequences<sup>[45]</sup>. The results show that LBP is consistently better than the histogram of oriented optical flow (HOOF), with true positives in the four databases CASME II, SMIC-E-HS, SMIC-E-VIS, and SMIC-E-NIR exceeding 27.99%, 13.91%, 9.63%, and 7.37%, respectively. Wang et al. used the same method to spot micro-expressions in CAS(ME)2<sup>[46]</sup>. They also proposed a micro-expression spotting method based on the main directional optical flow (MDMD). The recall, precision, and F1-score on the CAS(ME)2 dataset were 0.32, 0.35, and 0.33, respectively.

In later studies, Davison et al. utilized 3D-HOG features to identify changes in facial muscles in local FACS regions<sup>[50]</sup>. They only focused on facial areas containing a specific AU. Then, 3D-HOG was used to extract features in three orthogonal planes to extract the changes in movement. Because this method ignores the influence of the overall facial emotions, it emphasizes local facial muscle changes, thereby reducing computational complexity and improving detection accuracy.

In the Second Facial micro-expressions Grand Challenge (MEGC2019)<sup>[24]</sup>, a micro-expression spotting challenge task in long video sequences was conducted for the first time in two CAS(ME)2 and SAMM databases. Li et al. used local temporal patterns (LTP-ML)<sup>[51]</sup> for spontaneous micro-expression spotting, which achieved better experimental results than the LBP- $\chi^2$ -distance (LBP- $\chi^2$ )<sup>[41]</sup> method<sup>[48]</sup>. These datasets and challenges established the foundation for micro-expression spotting.

### 3.2 Performance metrics

If the following conditions are met, the spotted interval  $W_{spotted}$  is regarded as a true positive (TP):

$$\frac{W_{spotted} \cap W_{groundTruth}}{W_{spotted} \cup W_{groundTruth}} \geq k \quad (1)$$

where  $k$  is set to 0.5, and  $W_{groundTruth}$  represents the ground truth of the micro-expression interval (onset-offset). If  $k > 0.5$ , the spotted interval is regarded as a false positive (FP).

Suppose there is an  $m$  ground truth interval in the video, and  $n$  intervals are spotted, where  $FP = n - a$  and  $FN = m - a$ . The spotting performance in one video can be evaluated using the following metrics:

$$recall = \frac{a}{m}, precision = \frac{a}{n}, \quad (2)$$

$$F1-score = \frac{2TP}{2TP + FP + FN} = \frac{2a}{m + n}. \quad (3)$$

## 4 Micro-expression recognition

Micro-expression recognition is a task used to classify a micro-expression video. Similar to facial expressions, micro-expressions contain human emotions and identifying these is the most common task. Recognizing the emotions expressed in a face sequence with known micro-expressions is called micro-expression recognition.

### 4.1 Micro-expression recognition methods

In previous studies, micro-expression analysis is mainly used to classify micro-expression samples and



conduct micro-expression recognition. Current mainstream micro-expression recognition is mainly divided into three aspects: micro-expression recognition through the local binary pattern-three orthogonal planes (LBP-TOP) operator and its improved texture features, the optical flow (OF) features, and direct recognition of micro-expression samples based on deep learning.

#### 4.1.1 LBP-TOP methods

LBP-TOP is an extension of local binary patterns<sup>[52]</sup> and uses binary codes to describe local changes in texture along a circular area, and then encodes them into a histogram. LBP-TOP has been widely used in many different studies. Pfister et al. proposed a micro-expression recognition framework and used a temporal interpolation model (TIM)<sup>[54]</sup> to align the lengths of short video samples<sup>[29]</sup>. The dynamic texture features were then extracted through LBP-TOP, and a support vector machine (SVM) was used to conduct the classification. They later extended the complete local binary patterns (CLBP) to three orthogonal planes (CLBP-TOP) to distinguish between spontaneous and posed facial micro-expressions<sup>[54]</sup>. Subsequently, some micro-expression recognition is based on this framework, and several variants of LBP-TOP have been proposed.

Huang et al. proposed spatiotemporal complementary local quantized patterns (STCLQP) based on a complete spatiotemporal completion, which uses information including the angle, amplitude, and direction components to achieve a more compact feature extraction, thereby solving the sparsity problem of LBP features<sup>[55]</sup>. In addition, Wang et al. proposed a local binary pattern-six interception point (LBP-SIP) to reduce the redundant information of LBP-TOP<sup>[56]</sup>. Wang et al. later constructed a more compact LBP-MOP based on the LBP features of three average images<sup>[57]</sup>. The performance of LBP-MOP is equivalent to that of LBP-SIP, although its calculation time is greatly reduced. Huang et al. proposed a spatio-temporal local binary pattern with integral projection (STLBP-IP) to enhance the features of LBP-TOP through an integrated projection<sup>[58]</sup>. Wang et al. also explored the influence of the color feature space on micro-expression recognition and proposed a tensor independent color space (TICS), in which LBP-TOP features are extracted for micro-expression recognition<sup>[59]</sup>. The results show that the performance in the TICS color space is better than that in the RGB color space. Le et al. utilized the sparsity-promoting dynamic mode decomposition (DMDSP) to eliminate the redundant features of LBP-TOP and used an SVM and linear discriminant analysis (LDA) for classification<sup>[60]</sup>. In addition, Huang et al. proposed a new variant of the binary pattern spatiotemporal local radon binary pattern (STRBP) to extract robust shape features<sup>[61]</sup>. Moreover, Ben et al. proposed the hot wheel patterns on three orthogonal planes (HWP-TOP) to encode the discriminative features of macro-expressions and micro-expression images<sup>[62]</sup>. Finally, Niu et al. proposed a novel local second-order gradient pattern (LTOGP) to describe the subtle changes in micro-expressions<sup>[63,64]</sup>.

Table 6 summarizes the micro-expression recognition methods based on the LBP-TOP series. LBP-TOP is the earliest attempt at micro-expression recognition and an adaptation of traditional facial expression recognition to micro-expression recognition. Many later developed tools were dedicated to the use of LBP-TOP to improve the recognition performance at the feature levels, such as sparseness and redundancy. Although changes in the spatiotemporal textures of the micro-expression are mined and a certain descriptive ability is achieved, the calculated performance is not ideal, and the recognition accuracy needs further promotion.

#### 4.1.2 OF methods

Thus far, many studies have found that the temporal dynamics of video sequences have a positive effect on the recognition of micro-expressions. Therefore, technology based on the OF<sup>[65]</sup> for micro-expressions has

**Table 6 Based on LBP-TOP series of micro-expression recognition**

LBP-TOP Series	Accuracy		F1-Score	
	SMIC	CASME II	SMIC	CASME II
LBP-TOP <sup>[29]</sup>	48.78	—	—	—
CLBP-TOP <sup>[54]</sup>	78.2	—	—	—
STCLQP <sup>[55]</sup>	64.02	58.39	0.6381	0.5836
LBP-SIP <sup>[56]</sup>	44.51	46.56	0.4492	0.4480
LBP-MOP <sup>[57]</sup>	50.61	44.13	—	—
STLBP-IP <sup>[58]</sup>	57.93	59.51	0.5800	0.5700
TICS <sup>[59]</sup>	—	61.47	—	—
DMDSP <sup>[60]</sup>	58.00	49.00	0.6000	0.5100
STRBP <sup>[61]</sup>	60.98	64.37	—	—
HWP-TOP <sup>[62]</sup>	—	64.80	—	—
LTOGP <sup>[64]</sup>	—	66.00	—	—

created controversy.

Xu et al. proposed a method of obtaining a facial dynamics map (FDM), and believed that extracting only the main direction mapping features of OF can eliminate an abnormal OF vector caused by noise or changes in illumination<sup>[66]</sup>. Aside from the work of FDM, which only uses the single dominant direction of OF in each facial area, Allaert et al. proposed a method for determining the multi-directional optical flow characteristics from a single facial area to construct an OF mapping in the same direction in adjacent facial areas' feature<sup>[67]</sup>.

Liong et al. were inspired by optical strain (OS) micro-expression positioning, which they used to recognize micro-expressions, and derived the OS by calculating the normal and tangent tensors of the OF, which can capture the subtleties of micro-expressions<sup>[68]</sup>. Variety. First, all OS images are temporarily merged into one OS map, and then the generated map is adjusted to a fixed resolution to represent the feature vector of the video. To emphasize the importance of the micro-expression activity area, they used the time-weighted OS feature map and the local LBP-TOP feature to conduct a weighted fusion<sup>[69]</sup>, making the feature vector of the activity area more representative, thereby increasing the distinction between the degree of emotion categories. Subsequently, Liong et al. proposed a bi-weighted oriented optical flow (BI-WOOF) feature descriptor, which uses two schemes to apply a weighted average of the global and local HOOF features<sup>[70]</sup>. In the local feature extraction, each ROI is weighted by using the amplitude component and then multiplied by the average light change amplitude of each ROI. The overall HOOF feature is then weighted to obtain the final histogram feature. They believe that pixel shifts or larger deformations can help generate more discriminative histogram features.

Zhang et al. proposed a method to generate local statistical features by traversing regions to extract the HOOF and LBP-TOP features<sup>[71]</sup>. They found that the local features merged in each region of interest were more detailed and more representative than the global feature information. Happy et al. proposed a fuzzy histogram of optical flow orientation (FHOFO) for micro-expression recognition, where the histogram is only a collection of directions, not a weighted optical flow size<sup>[72]</sup>. They assumed that the micro-expression was so subtle that the sensing amplitude was negligible. Simultaneously, they introduced a fuzzy membership function based on previous fuzzy membership functions to consider the influence of the orientation angle on its surroundings, creating a smooth histogram of the motion vector. Liu et al. proposed a main directional mean optical flow (MDMO) feature, which considers the average local statistics of the OF vector in each ROI and its spatial location<sup>[73]</sup>. As an advantage of this method, 72 features are extracted from 36 ROIs.

Table 7 summarizes the micro-expression recognition methods based on the optical flow series. The optical flow features can describe the micro-expression features from a motion perspective and provide good interpretability under the premise of ensuring recognition performance. However, the extraction of dense optical flow features is time-consuming. Although the improved optical flow features can reach recognition accuracy of 80%, a large amount of pre-processing is still required to align faces in micro-expression video sequences to eliminate the effects of head movements and rotations.

**Table 7 Based on optical flow series of micro-expression recognition**

Optical Flow Series	Accuracy		F1-Score	
	SMIC	CASME II	SMIC	CASME II
FDM <sup>[66]</sup>	54.88	45.93	0.5380	0.4053
OF Maps <sup>[67]</sup>	–	65.35	–	–
OS <sup>[68]</sup>	–	–	0.5300	0.5600
BI-WOOF <sup>[70]</sup>	50.61	44.13	–	–
HOOF <sup>[71]</sup>	–	62.50	–	–
FHOFO <sup>[72]</sup>	51.83	56.64	0.5243	0.5248
MDMO <sup>[73]</sup>	80.00	67.37	–	–

### 4.1.3 Deep Learning-based methods

Although a recognition method based on hand-crafted features can achieve good recognition results, hand-crafted features tend to ignore other information in the original image data. Convolutional neural networks (CNNs) have gradually emerged in recent years and have attracted significant attention. The use of such networks is an extremely efficient pattern classification approach, which was proposed by Hubel and Wiesel in the 1960s when studying the function of related neurons in the cerebral cortex of cats. This method is mainly used in the image processing field. CNNs can efficiently identify and classify images. Famous CNN network structures include LeNet<sup>[74]</sup>, AlexNet<sup>[75]</sup>, VGGNet<sup>[76]</sup>, GoogLeNet<sup>[77]</sup>, and ResNet<sup>[78]</sup>.

Kim et al. used the CNN structure to encode the spatial information of different onset, apex, and offset frames<sup>[79]</sup>. This work is one of the earliest to use CNNs in a micro-expression analysis. With this method, CNN features are input into long short-term memory (LSTM) to realize micro-expression recognition. Gan et al. introduced deep learning and proposed the optical flow features from the Apex frame (OFF-Apex) method<sup>[80]</sup>. The method uses the optical flow feature map of the micro-expression apex frame as the input of the CNN to enhance the features of optical flow. It should be noted that, unlike the above methods, these methods only utilize the apex and onset frame instead of the complete video sequence. Because a standard CNN is limited by a weakness in that the overall relationship is represented, Quang et al. use capsule networks (CapsuleNet) for micro-expression recognition. The experimental results show that the CapsuleNet method can achieve better results than a CNN model in micro-expression recognition<sup>[81]</sup>. Zhou et al. proposed a new network structure called a dual-inception network (DINet)<sup>[82]</sup>. The model learns high-dimensional feature representations from the horizontal optical flow and vertical optical flow of the onset frame and the intermediate frame for micro-expression recognition. Observing that an extremely deep CNN architecture cannot perform well under limited micro-expression data, Liong et al. proposed a shallow triple stream three-dimensional CNN (STSTNet) using three parallel stream feature maps input into the network to suppress underfitting problems<sup>[83]</sup>. Liu et al. proposed a part-based deep neural network (PB-DNN), which is enhanced by magnification and reduction of micro-expression samples<sup>[84]</sup>. Inspired by the domain confrontation network<sup>[85]</sup>, macro-expression samples were used in the CK + dataset and micro-expression samples in the SMIC, CASME II, and SAMM to minimize the combination loss function.

Table 8 summarizes the micro-expression recognition methods based on the deep learning series. Although deep learning has achieved surprising results in micro-expression recognition, some challenges still remain. For example, compared with CASME II, SMIC and SAMM are more challenging, which may be because the SMIC and SAMM databases are more widely distributed in terms of age and ethnicity, which has an impact on the recognition effect. At the same time, owing to the fast and low-intensity characteristics of micro-expressions, it is difficult for deep learning methods to capture subtle changes in micro-expressions. It is also necessary to consider the introduction of better methods to solve these problems in micro-expression recognition.

**Table 8 Based on optical flow series of micro-expression recognition**

Deep Learning	Unweighted F1-score (UF1)			Unweighted Average Recall (UAR)		
	SMIC	CASME II	SAMM	SMIC	CASME II	SAMM
OFF-Apex <sup>[80]</sup>	0.6817	0.8764	0.5409	0.6695	0.8681	0.5392
CapsuleNet <sup>[81]</sup>	0.5820	0.7068	0.6209	0.5877	0.7018	0.5989
DINet <sup>[82]</sup>	0.6645	0.8621	0.5868	0.6726	0.8560	0.5663
STSTNet <sup>[83]</sup>	0.6801	0.8382	0.6588	0.7013	0.8686	0.6810
PB-DNN <sup>[84]</sup>	0.7461	0.8293	0.7754	0.7530	0.8209	0.7152

## 4.2 Performance metrics

In the micro-expression classification based on the difference model, we used the LOSO cross-validation method to obtain the final recognition results. In evaluating micro-expression recognition, to deal with an imbalanced class distribution, we also used the accuracy and *F1\_score* for performance evaluation. Specifically, the *F1\_score* is expressed as follows:

$$F1\_score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

For

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

And

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

## 5 Conclusion

In this paper, we reviewed the datasets relevant to the spotting, and recognition methods of micro-expressions. First, we summarized the current micro-expression datasets of posed and spontaneous images and analyzed the advantages among them. Then, the methods and evaluation approaches for spotting micro-expressions in video sequences were summarized. Finally, we introduced the method of micro-expression recognition utilizing LBP-TOP, OF, and deep learning. However, some problems remain to be solved.

### 5.1 Pre-processing technology for micro-expressions

One advantage of using existing datasets is that new algorithms can be applied directly to the preprocessed images, reducing the pressure on the preprocessing stage. However, the pre-processing of micro-expression samples is an important step, and this goes beyond that of normal facial expression or other facial information recognition. During the recording of an existing dataset, the position of the subject is relatively

stable, and after further alignment processing, it is simpler to obtain high-quality images. However, this is difficult to achieve in practical applications, and thus refined preprocessing for micro-expressions is worth further study. Additionally, future research on preprocessing the micro-expression image sequence through methods such as face detection and alignment should be considered.

## 5.2 Unbalanced micro-expression sample distribution

However, the sample distribution of spontaneous micro-expression datasets may be unbalanced because of the collection equipment used, the experimental environment, and/or the individualization of the subjects. A sample imbalance is reflected in two aspects. The number of frames in the sample video clips and the sample category distribution in a dataset may be unbalanced. To reduce the influence of the number of sample frame imbalances in micro-expression recognition, researchers typically use a time interpolation model (TIM) to align the number of sample frames. For the imbalance of a category distribution in micro-expression recognition, it is also necessary to further consider the strategy balance method.

## 5.3 Identity information interference problem

Psychologists believe that facial muscle movements are not directly related to individual attributes such as gender, age, and ethnicity when micro-expressions occur. However, when micro-expression recognition is applied on a facial image acquired under various imaging conditions, the micro-expression image is a superposition of the personal identity attribute and facial muscle movements. Therefore, individual attributes will interfere with the micro-expression analysis.

Micro-expressions only occur in the local parts of the facial muscle movements, and these movements are very subtle in that micro-expression recognition pays more attention to local facial features. These characteristics make the identity information in the micro-expression image extremely significant. Therefore, reducing the interference of identity information is one of the challenges in micro-expression analysis.

## 5.4 Micro-expression fine-grained image classification problem

When facial expression images obtained by the imaging device are used for micro-expression spotting and recognition, micro-expression facial muscle movements only occur in the local area of the face and the problem of low intensity exists, resulting in a minor difference between a micro-expression facial image and a natural facial image. These problems pose significant challenges to micro-expression analysis.

A micro-expression is only related to the local part of the facial image and has low intensity, which results in small inter-class variations between the facial image and a natural facial image. This will lead to fine-grained image classification problems in micro-expression spotting and recognition.

### Declaration of competing interest

We declare that we have no conflict of interest.

### References

- 1 Mehrabian A. Communication without words. *Psychological Today*, 1968, 2(6): 53–55
- 2 Haggard E A, Isaacs K S. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. In: *Methods of Research in Psychotherapy*. Boston, MA, Springer US, 1966, 154–165  
DOI:10.1007/978-1-4684-6045-2\_14



- 3 Ekman P, Friesen W V. Nonverbal leakage and clues to deception. *Psychiatry* 1969, 32(1):88–106  
DOI:10.1080/00332747.1969.11023575
- 4 Ekman P. Lie catching and microexpressions. *The Philosophy of Deception*, 2009, 1(2): 5
- 5 Frank M, Maccario C, Govindaraju V. Protecting airline passengers in the age of terrorism. ABC-CLIO, Santa Barbara, 2009
- 6 O'Sullivan M, Frank M G, Hurley C M, Tiwana J. Police lie detection accuracy: the effect of lie scenario. *Law and Human Behavior*, 2009, 33(6): 530
- 7 Matsumoto D, Hwang H S. Evidence for training the ability to read microexpressions of emotion. *Motivation and Emotion*, 2011, 35(2): 181–191
- 8 Turner J H. The evolution of emotions: the nonverbal basis of human social organization. Lawrence Erlbaum Associates, Mahwah, New Jersey, 1997
- 9 Frank M, Herbasz M, Sinuk K, Keller A, Nolan C. 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, 2009, 1–35
- 10 Ekman P. *Telling lies: clues to deceit in the marketplace, politics, and marriage* (revised edition). WW Norton & Company, 2009
- 11 Ekman P, Friesen W V. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 1971, 17(2): 124–129  
DOI:10.1037/h0030377
- 12 Yan W J, Wu Q, Liang J, Chen Y H, Fu X L. How fast are the leaked facial expressions: the duration of micro-expressions. *Journal of Nonverbal Behavior*, 2013, 37(4): 217–230  
DOI:10.1007/s10919-013-0159-8
- 13 Porter S, ten Brinke L. Reading between the lies: identifying concealed and falsified emotions in universal facial expressions. *Psychological Science*, 2008, 19(5): 508–514  
DOI:10.1111/j.1467-9280.2008.02116.x
- 14 Ekman P. Microexpression training tool (METT). University of California, San Francisco, 2002
- 15 Frank M G, Maccario C J, Govindaraju V. Behavior and security. *Protecting airline passengers in the age of terrorism*, 2009, 86–106
- 16 Zhao G Y, Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(6): 915–928  
DOI:10.1109/tpami.2007.1110
- 17 Liu M, Shan S, Wang R, Chen X. Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, 1749–1756
- 18 Jung H, Lee S, Yim J, Park S, Kim J. Joint fine-tuning in deep neural networks for facial expression recognition. In: *Proceedings of the IEEE international conference on computer vision*. 2015, 2983–2991
- 19 Zeng N Y, Zhang H, Song B Y, Liu W B, Li Y R, Dobaie A M. Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing*, 2018, 273: 643–649  
DOI:10.1016/j.neucom.2017.08.043
- 20 Li S, Deng W H. Deep facial expression recognition: a survey. *IEEE Transactions on Affective Computing*, 2020: 1  
DOI:10.1109/taffc.2020.2981446
- 21 Tornincasa S, Vezzetti E, Moos S, Violante M, Marcolin F, Dagnes N, Ulrich L, Tregnaghi G. 3D facial action units and expression recognition using a crisp logic. *Computer-Aided Design and Applications*, 2018, 16(2): 256–268  
DOI:10.14733/cadaps.2019.256-268
- 22 Merghani W, Davison A, Yap M. Facial Micro-expressions Grand Challenge 2018: evaluating spatio-temporal features for classification of objective classes. In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, 662–666
- 23 Yap M H, See J, Hong X, Wang S J. Facial micro-expressions grand challenge 2018 summary. In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, 675–678
- 24 See J, Yap M H, Li J, Hong X, Wang S J. Megc 2019—the second facial micro-expressions grand challenge. In: *2019*

- 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019, 1–5
- 25 Li J, Wang S, Yap M H, See J, Hong X, Li X. MEGC2020-the third facial micro-expression grand challenge. In: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)(FG). 2020, 234–237
- 26 Polikovskiy S, Kameda Y, Ohta Y. Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. In: International Conference on Imaging for Crime Detection and Prevention. IET, 2009, 1–6
- 27 Shreve M, Godavarthy S, Goldgof D, Sarkar S. Macro-and micro-expression spotting in long videos using spatio-temporal strain. In: IEEE international conference and workshops on automatic face and gesture recognition. IEEE, 2011, 51–56
- 28 Warren G, Schertler E, Bull P. Detecting deception from emotional and unemotional cues. *Journal of Nonverbal Behavior*, 2009, 33(1): 59–69  
DOI:10.1007/s10919-008-0057-7
- 29 Pfister T, Li X, Zhao G, Pietikäinen M. Recognising spontaneous facial micro-expressions. In: IEEE International Conference on Computer Vision. IEEE, 2011, 1449–1456
- 30 Li X, Pfister T, Huang X, Zhao G, Pietikäinen M. A spontaneous micro-expression database: Inducement, collection and baseline. In: IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. IEEE, 2013, 1–6
- 31 Yan W J, Wu Q, Liu Y J, Wang S J, Fu X. CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces. In: IEEE International Conference And Workshops on Automatic Face and Gesture Recognition. IEEE, 2013, 1–7
- 32 Yan W J, Li X B, Wang S J, Zhao G Y, Liu Y J, Chen Y H, Fu X L. CASME II: an improved spontaneous micro-expression database and the baseline evaluation. *PLoS One*, 2014, 9(1): e86041  
DOI:10.1371/journal.pone.0086041
- 33 Qu F B, Wang S J, Yan W J, Li H, Wu S H, Fu X L. CAS(ME): a database for spontaneous macro-expression and micro-expression spotting and recognition. *IEEE Transactions on Affective Computing*, 2018, 9(4): 424–436  
DOI:10.1109/taffc.2017.2654440
- 34 Davison A K, Lansley C, Costen N, Tan K, Yap M H. SAMM: a spontaneous micro-facial movement dataset. *IEEE Transactions on Affective Computing*, 2018, 9(1): 116–129  
DOI:10.1109/taffc.2016.2573832
- 35 Yap C H, Kendrick C, Yap M H. Samm long videos: a spontaneous facial micro-and macro-expressions dataset. 2019
- 36 Hess U, Kleck R E. Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology*, 1990, 20(5): 369–385  
DOI:10.1002/ejsp.2420200502
- 37 Friesen E, Ekman P. Facial action coding system: a technique for the measurement of facial movement. Palo Alto, 1978
- 38 Valstar M F, Pantic M. Fully automatic recognition of the temporal phases of facial actions. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, 2012, 42(1): 28–43  
DOI:10.1109/tsmcb.2011.2163710
- 39 Hong X, Tran T K, Zhao G. Micro-expression spotting: a benchmark. 2017
- 40 Polikovskiy S, Kameda Y, Ohta Y. Facial micro-expression detection in hi-speed video based on facial action coding system (FACS). *IEICE transactions on information and systems*, 2013, 96(1): 81–92
- 41 Moilanen A, Zhao G, Pietikäinen M. Spotting rapid facial movements from videos using appearance-based feature difference analysis. In: 2014 22nd international conference on pattern recognition. IEEE, 2014, 1722–1727
- 42 Davison A K, Yap M H, Lansley C. Micro-facial movement detection using individualised baselines and histogram-based descriptors. In: 2015 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, 2015, 1864–1869
- 43 Patel D, Zhao G, Pietikäinen M. Spatiotemporal integration of optical flow vectors for micro-expression detection. In: International conference on advanced concepts for intelligent vision systems. Springer, 2015, 369–380
- 44 Xia Z, Feng X, Peng J, Peng X, Zhao G. Spontaneous micro-expression spotting via geometric deformation modeling. *Computer Vision and Image Understanding*. 2016, 147: 87–94
- 45 Li X, Hong X, Moilanen A, Huang X, Pfister T, Zhao G, Pietikäinen M. Towards reading hidden emotions: a comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Transactions on Affective*

- Computing, 2017, 9(4): 563–577
- 46 Wang S J, Wu S, Qian X, Li J, Fu X. A main directional maximal difference analysis for spotting facial movements from long-term videos. *Neurocomputing*, 2017, 230: 382–389
  - 47 Davison A, Merghani W, Lansley C, Ng C C, Yap M H. Objective micro-facial movement detection using faces-based regions and baseline evaluation. In: *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE, 2018, 642–649
  - 48 Li J, Soladie C, Seguier R, Wang S J, Yap M H. Spotting micro-expressions on long videos sequences. In: *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, 1–5
  - 49 Tran T K, Hong X, Zhao G. Sliding window based micro-expression spotting: a benchmark. In: *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2017, 542–553
  - 50 Davison A, Merghani W, Lansley C, Ng C C, Yap M H. Objective micro-facial movement detection using faces-based regions and baseline evaluation. In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, 642–649
  - 51 Li J, Soladie C, Seguier R. Ltp-ml: micro-expression detection by recognition of local temporal pattern of facial movements. In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, 634–641
  - 52 Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 1996, 29(1): 51–59
  - 53 Zhou Z, Zhao G, Pietikäinen M. Towards a practical lipreading system. In: *CVPR 2011*, IEEE, 2011, 137–144
  - 54 Pfister T, Li X, Zhao G, Pietikäinen M. Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework. In: *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. IEEE, 2011, 868–875
  - 55 Huang X, Zhao G, Hong X, Zheng W, Pietikäinen M. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neurocomputing*, 2016, 175: 564–578
  - 56 Wang Y, See J, Phan R C W, Oh Y H. Lbp with six intersection points: Reducing redundant information in lbp-top for micro-expression recognition. In: *Asian conference on computer vision*. Springer, 2014, 525–537
  - 57 Wang Y, See J, Phan R C W, Oh Y H. Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition. *PloS One*, 2015, 10(5): e0124674
  - 58 Huang X, Wang S J, Zhao G, Piteikainen M. Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection. In: *Proceedings of the IEEE international conference on computer vision workshops*. 2015, 1–9
  - 59 Wang S J, Yan W J, Li X, Zhao G, Fu X. Micro-expression recognition using dynamic textures on tensor independent color space. In: *2014 22nd international conference on pattern recognition*. IEEE, 2014, 4678–4683
  - 60 Le Ngo A C, Liong S T, See J, Phan R C W. Are subtle expressions too sparse to recognize? In: *2015 IEEE International Conference on Digital Signal Processing (DSP)*. IEEE, 2015, 1246–1250
  - 61 Huang X, Zhao G. Spontaneous facial micro-expression analysis using spatiotemporal local radon-based binary pattern. In: *2017 International Conference on the Frontiers and Advances in Data Science (FADS)*. IEEE, 2017, 159–164
  - 62 Ben X, Jia X, Yan R, Zhang X, Meng W. Learning effective binary descriptors for micro-expression recognition transferred by macro-information. *Pattern Recognition Letters*, 2018, 107: 50–58
  - 63 Niu M, Li Y, Tao J, Wang S J. Micro-expression recognition based on local two-order gradient pattern. In: *2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*, IEEE, 2018, 1–6
  - 64 Niu M, Tao J, Li Y, Huang J, Lian Z. Discriminative video representation with temporal order for micro-expression recognition. In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, 2112–2116
  - 65 Horn B K, Schunck B G. Determining optical flow. In: *Techniques and Applications of Image Understanding*. International Society for Optics and Photonics, 1981, 319–331
  - 66 Xu F, Zhang J, Wang J Z. Microexpression identification and categorization using a facial dynamics map. *IEEE Transactions on Affective Computing*, 2017, 8(2): 254–267
  - 67 Allaert B, Bilasco I M, Djeraba C. Consistent optical flow maps for full and micro facial expression recognition. In:

- International Conference on Computer Vision Theory and Applications (VISAPP). Springer, 2017
- 68 Liong S T, Phan R C W, See J, Oh Y H, Wong K. Optical strain based recognition of subtle emotions. In: 2014 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS). IEEE, 2014, 180–184
- 69 Liong S T, See J, Phan R C W, Le Ngo A C, Oh Y H, Wong K. Subtle expression recognition using optical strain weighted features. In: Asian Conference on Computer Vision. Springer, 2014, 644–657
- 70 Liong S T, See J, Wong K, Phan R C W. Less is more: micro-expression recognition from video using apex frame. *Signal Processing: Image Communication*, 2018, 62: 82–92
- 71 Zhang S, Feng B, Chen Z, Huang X. Micro-expression recognition by aggregating local spatio-temporal patterns. In: International Conference on Multimedia Modeling. Springer, 2017, 638–648
- 72 Happy S, Routray A. Recognizing subtle micro-facial expressions using fuzzy histogram of optical flow orientations and feature selection methods. *Computational Intelligence for Pattern Recognition*, 2018, 341–368
- 73 Liu Y J, Zhang J K, Yan W J, Wang S J, Zhao G, Fu X. A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Transactions on Affective Computing*, 2015, 7(4): 299–310
- 74 LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324
- 75 Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*. 2012, 1097–1105
- 76 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014
- 77 Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, 1–9
- 78 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, 770–778
- 79 Kim D H, Baddar W J, Ro Y M. Micro-expression recognition with expression-state constrained spatio-temporal feature representations. In: *Proceedings of the 24th ACM International Conference on Multimedia*. 2016, 382–386
- 80 Gan Y, Liong S T, Yau W C, Huang Y C, Tan L K. Off-apexnet on micro-expression recognition system. *Signal Processing: Image Communication*, 2019, 74:129–139
- 81 Van Quang N, Chun J, Tokuyama T. Capsulenet for micro-expression recognition. In: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019, 1–7
- 82 Zhou L, Mao Q, Xue L. Dual-inception network for cross-database micro-expression recognition. In: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019, 1–5
- 83 Liong S T, Gan Y, See J, Khor H Q, Huang Y C. Shallow triple stream three-dimensional cnn (ststnet) for micro-expression recognition. In: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019, 1–5
- 84 Liu Y, Du H, Zheng L, Gedeon T. A neural micro-expression recognizer. In: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019, 1–4
- 85 Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky V. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 2016, 17(1): 2096–2030