ECAI 2023
K. Gal et al. (Eds.)

doi:10.3233/FAIA230611

Micro-Expression Spotting Method Based on AU Prototype

Yuhong He[a], Guangyu Wang[a], Lin Ma[a] and Haifeng Li[a;*]

[a]Harbin Institute of Technology

Abstract. Micro-expressions (MEs) are brief, involuntary facial expressions that reveal genuine emotions, making their accurate detection crucial in various applications, such as security, psychology, and human-computer interaction. Due to its small intensity and short duration, how accurately capturing the subtle movements of micro-expression is a challenging problem. This paper presents a novel AU prototype-based method for micro-expression spotting, which offers high accuracy and robustness. Action Units (AUs) are basic facial actions, such as brow lower and lip corner puller, that are widely used for micro-expression analysis, and an expression can be encoded as a sequence of AUs. Our approach involves designing AU prototypes that record representative dynamic information of AUs. We then calculate the prototype matching index between AU prototypes and the image sequence to construct time-domain prototype matching curves for ME spotting. In the experimental section, AU prototypes derived from CASMEII dataset enable a more intuitive analysis of AU within micro-expressions. Results on the CAS(ME)^2 dataset demonstrate that our ME spotting method significantly outperforms existing approaches. This makes our method highly valuable for various application scenarios, potentially enhancing emotion recognition and analysis in real-world settings.

1 Introduction

Expression is an essential method for human emotional interaction [13] and can be divided into macro-expression (MaE) and micro-expression (ME) according to its duration and intensity [17]. Micro-expression, which is not controlled by subjective consciousness, often reflects real human emotions and has critical applications in public safety [1]. Micro-expression spotting is a key problem in ME research, aiming to locate micro-expression intervals in the video. Due to the small intensity and short duration [2], as shown in Figure 1, micro-expression features may be affected by many factors such as head-shaking, blinking, facial differences, et al., which makes the micro-expression spotting accuracy not high. This work aims to introduce a prototype-based method for micro-expression spotting with high accuracy and interpretability.

Many studies have been performed on hand-crafted feature extraction for capturing micro-expressions, mainly including optical flow features and texture features [14][10][12][5][4][21][6]. The optical flow method finds objects' movement between two frames according to the change of pixels in the time-domain. It offers excellent interpretability and intuitiveness, which enables its application in a wide range of high-reliability settings for micro-expression research. Researchers recognize the optical flow method as a key method in the micro-expression analysis [4][21][6]. In 2021, He [6] used the mean optical flow of the facial regions of interest to estimate local movements for micro-expression spotting, and the F1-score reached 0.3530 on the CAS(ME)^2 database [15]. However, the optical flow feature is susceptible to ME-independent movements and leads to a high false-positive rate.

Deep learning methods have recently achieved advanced results for ME spotting [23][18][20]. In 2021, Yu et al. [20] introduced the location suppression-based spotting network (LSSNet) for micro-expression spotting, achieving a F1-score of 0.327 on the CAS(ME)^2 database. In 2022, Leng et al. [9] proposed a ME Spotting framework based on Apex and Boundary Perception Network (ABPN), and the F1-score reached 0.2908 on the SAMM-LV database [19]. Nevertheless, the challenges in gathering micro-expression movements and labeling them have resulted in small-scale ME databases, which are insufficient for deep learning model requirements and may lead to issues such as overfitting. Moreover, deep learning-based ME spotting methods often lack a clear basis, and their results are not easily interpretable. In application contexts like abnormal behavior identification and emotion monitoring, ME spotting and recognition outcomes cannot be accepted without clear evidence.

(a) the onset image (b) the apex image (c) the offset image

Figure 1. Schematic diagram of a micro-expression onset, apex and offset image [16]. The apex image contains the most significant micro-expression movement information, and the red arrow points to the area where local movement occurs.

Ekman first introduced action units (AUs) in the face coding system (FACS) [3]. AUs are basic facial actions. Every expression can be encoded as a sequence of AU. AU classification is recognized as a powerful tool for analyzing micro-expressions and thus has become an important topic in micro-expression research [22][11]. In 2021, Li et al. [11] proposed the dual-view attentive similarity-preserving distillation method for robust micro-expression AU detection by leveraging massive facial expressions in the wild. Each AU is generated by corresponding facial muscle traction. Therefore, similar movement patterns show on the face when the same AU occurs [3]. However,

* Corresponding Author. Email: lihaifeng@hit.edu.cn

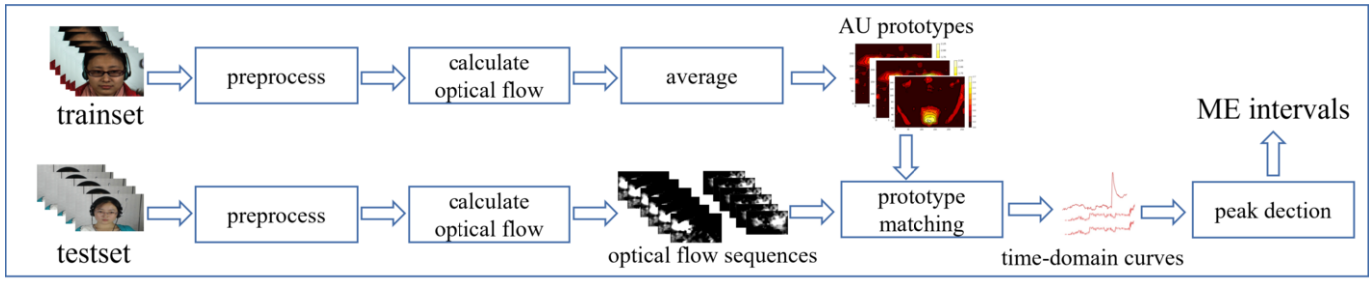


Figure 2. Overview of AU prototype based ME spotting method

due to the natural facial muscle structure differences, facial movements with the same AU are different between subjects. Subjects themselves can also vary in amplitudes of facial movements corresponding to different emotional intensities. Therefore, it is highly desirable to design methods that can obtain AU movement patterns and are robust to differences in facial movements with the same AU. This helps improve micro-expression spotting accuracy and is vital for interpretable micro-expression modeling.

In summary, this paper implements the optical flow method to capture the dynamic information of facial movement, analyzes complex micro-expressions based on AU prototypes, and then constructs a micro-expression spotting method with good reliability and robustness. The main contributions of this paper are summarized as follows: (1) AU prototypes for micro-expression analysis are proposed, which record the representative dynamic information of AUs. (2) A micro-expression spotting method based on the AU prototype is designed, dramatically improving ME spotting's accuracy.

The remainder of this paper is structured as follows: Section 2 details the optical flow technique, AU prototypes construction method and the prototype-based ME spotting method; Section 3 outlines the experiment and discusses the results of our method; Finally, the conclusion section summarizes the paper and explores potential avenues for future research.

2 Micro-expression Spotting Technology

The overview of our micro-expression spotting method is shown in Figure 2. During training, the videos of the trainset are first preprocessed, and optical flow fields are computed from them. On this basis, AU prototypes are learned. During testing, we calculate the prototype matching index between optical flow field sequences extracted from the video and all learned prototypes. And construct time-domain prototype matching curves corresponding to each AU. Next, we use the peak detection method to locate ME intervals from time-domain prototype matching curves.

2.1 Optical flow method

In this research, the optical flow method is employed to estimate facial movements. To effectively apply the optical flow approach, two prerequisites must be met: firstly, the luminosity of video frames remains consistent, and secondly, the spatial positioning of the object does not undergo abrupt alterations within the temporal domain.

As delineated in equations 1 and 2, polynomial approximations are carried out on images I_1 and I_2 to estimate grayscale intensity:

$$I_1(p) = p^T \mathbf{A}_1 p + \mathbf{b}_1^T p + c_1 \tag{1}$$

$$I_2(p) = p^T \mathbf{A}_2 p + \mathbf{b}_2^T p + c_2 \tag{2}$$

$\mathbf{A}_1$ and $\mathbf{A}_2$ are symmetric matrices, $\mathbf{b}_1^T$ and $\mathbf{b}_2^T$ are vectors, and c_1 and c_2 are scalar. All of them record parameters of polynomials. The pixel point p is represented by (x, y), where x denotes the horizontal coordinate, and y signifies the vertical coordinate. $I_1(p)$ refers to the grayscale intensity of image I_1 at point p.

Assuming a global displacement, denoted as $\mathbf{dis}$, exists between images I_1 and I_2, the polynomial approximation of the second image is derived:

$$\begin{aligned} I_2(p) =& I_1(p - \mathbf{dis}) \\ =& (p - \mathbf{dis})^T \mathbf{A}_1 (p - \mathbf{dis}) + \mathbf{b}_1^T (p - 1) + c_1 \\ =& p^T \mathbf{A}_1 p + (\mathbf{b}_1 - 2\mathbf{A}_1 \mathbf{dis})^T p \\ &+ \mathbf{dis}^T \mathbf{A}_1 \mathbf{dis} - \mathbf{b}_1^T \mathbf{dis} + c_1 \end{aligned} \tag{3}$$

Based on equations 2 and 3, the subsequent three equations are valid:

$$\mathbf{A}_2 = \mathbf{A}_1 \tag{4}$$

$$\mathbf{b}_2 = \mathbf{b}_1 - 2\mathbf{A}_1 \mathbf{dis} \tag{5}$$

$$c_2 = \mathbf{dis}^T \mathbf{A}_1 \mathbf{dis} - \mathbf{b}_1^T \mathbf{dis} + c_1 \tag{6}$$

If $\mathbf{A}_1$ is a non-singular matrix, the global displacement $\mathbf{dis}$ can be computed as:

$$\mathbf{dis} = -\frac{1}{2} \mathbf{A}_1^{-1} (\mathbf{b}_2 - \mathbf{b}_1) \tag{7}$$

The optical flow technique is adept at calculating the subtle movement information and is highly robust to variations in facial texture, making it particularly suitable for detecting minute movements like micro-expressions.

Suppose $I = \{I_1, I_2, ..., I_N\}$ is an image sequence of ME video, and the optical flow method is applied between the first image I_1 and each consecutive image of the sequence $\{I_2, I_3, ..., I_N\}$ to calculate the dense optical flow graph sequences $\{F_1, F_2, ..., F_{N-1}\}$. w and h are the weight and height of an image in sequence I.

$$F_i = \text{Flow}(I_1, I_i + 1), F_i \in \mathbb{R}^{w \times h \times 2} \tag{8}$$

F_i signifies the dense optical flow graph between I_{i+1} and the first image I_1. It comprises two matrices that represent the optical flow in horizontal and vertical directions.

2.2 Preprocessing

We normalize the subject's face before calculating AU prototypes. Considering the natural differences in facial organ distribution between subjects, we have identified the areas of the face that are most crucial in determining emotions, namely the eye region I^{EYE} and the mouth region I^{MTH}. To ensure the consistency and accuracy of our analysis across subjects, we have meticulously aligned these regions in each face using a standardized method, as depicted in Figure 3.

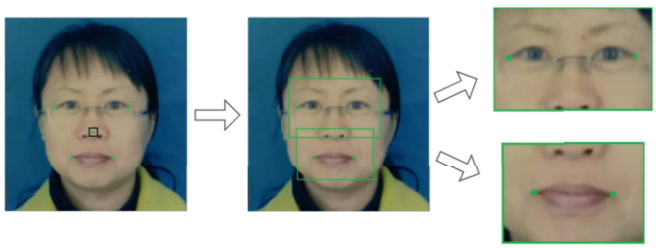


Figure 3. Schematic diagram of preprocessing step

Taking the calculation of the eye region I^{EYE} as an example, the scale of I^{EYE} is calculated as Equation 9 to Equation 12. $p_{leye} = (x_{leye}, y_{leye})$and $p_{reye} = (x_{reye}, y_{reye})$ are the left and right outer corner. $p_{ls}^{eye} = (x_{ls}^{eye}, y_{ls}^{eye})$ represents the upper left corner, and $p_{rx}^{eye} = (x_{rx}^{eye}, y_{rx}^{eye})$ is the lower right corner of I^{EYE}. We define the width and height of I^{EYE} are w_{eye} and h_{eye}.

$$x_{ls}^{eye} = x_{leye} - \gamma_1 \times (x_{reye} - x_{leye}) \tag{9}$$

$$y_{ls}^{eye} = \frac{(y_{reye} + y_{leye})}{2} - \gamma_2 \times (x_{reye} - x_{leye}) \tag{10}$$

$$x_{rx}^{eye} = x_{reye} + \gamma_3 \times (x_{reye} - x_{leye}) \tag{11}$$

$$y_{rx}^{eye} = \frac{(y_{reye} + y_{leye})}{2} + \gamma_4 \times (x_{reye} - x_{leye}) \tag{12}$$

2.3 AU prototypes construction

When AUs occur, $\vec{d}$ is the ME-related movement at point $p = (x, y)$ in the image. However, we may get an inaccurate estimate $\vec{u}$ using the optical flow method. We consider the optical flow vector $\vec{u}$ as the superposition of ME-related motion $\vec{d}$ and noise $\vec{n}$.

$$\vec{u} = \vec{d} + \vec{n} \tag{13}$$

Random noise may include head movements, blinks, and errors from the optical flow method. We consider the mean value of these random noises is approximately zero. Noise $\vec{n}$ may cover the micro-expression movement $\vec{d}$ due to the small intensity of MEs, resulting in a low signal-noise ratio of ME features, which brings difficulties to ME analysis. Therefore, based on the similar facial motion pattern of AU, this paper averages the optical flow field of micro-expressions containing the same AU to suppress random noise. The mathematic expectation of $\vec{u}$ is $E\{\vec{u}\} = E\{\vec{d}\}$. The average of M movements is shown in Equation 14.

$$\bar{\vec{u}} = \frac{1}{M} \sum_{g=0}^{M} \vec{u}_g \tag{14}$$

The new variance is $\frac{1}{M}$ of the original.

$$\sigma_{new}^2 = \frac{1}{M} \sigma_{\vec{u}}^2 \tag{15}$$

From the above analysis, we design the AU prototypes, and the calculation process is shown in Figure 4.

First, we calculate the optical flow fields of the eye region I^{EYE} and the mouth region I^{MTH} between the first and peak frame in the micro-expression video g, denoted as ${F_{apex}^{g}}^{eye} and {F_{apex}^{g}}^{mth}$.

$${F_{apex}^{g}}^{eye} = Flow({I_{first}^{g}}^{EYE}, {I_{apex}^{g}}^{EYE}) \tag{16}$$

$${F_{apex}^{g}}^{mth} = Flow({I_{first}^{g}}^{MTH}, {I_{apex}^{g}}^{MTH}) \tag{17}$$

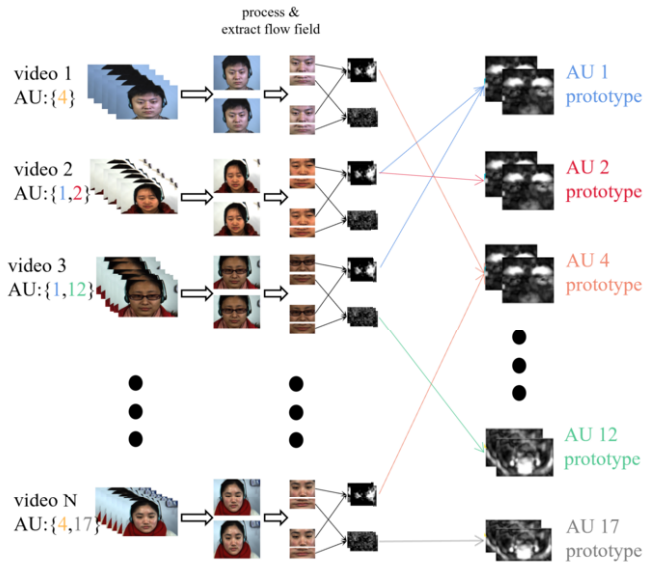


Figure 4. Process diagram of learning AU prototypes

Then, we average all ME optical flow fields with the same AU to obtain representative dynamic information and suppress random noise. AU prototypes are calculated as Equation 18.

$$M_j = \begin{cases} \dfrac{\sum_{g \in S_{AU_j}} {F_{apex}^{g}}^{eye}}{|S_{AU_j}|}, j \in H_{eye} \\ \dfrac{\sum_{g \in S_{AU_j}} {F_{apex}^{g}}^{mth}}{|S_{AU_j}|}, j \in H_{mth} \end{cases} \tag{18}$$

M_j is the prototype of AU_j. H_{eye} is a set of action units which happened in the eye region, and H_{mth} is a set of action units which happened in the month region. S_{AU_j} is a set of ME video indexes with AU_j.

We can obtain a sequence of AU prototype $mask = M_1, ..., M_V$, and V is the number of AU.

2.4 Micro-expression Spotting Method

For applied research, we design and implement a micro-expression spotting method. The input is a long video containing a front view of the face, while the output encompasses all spotted micro-expression

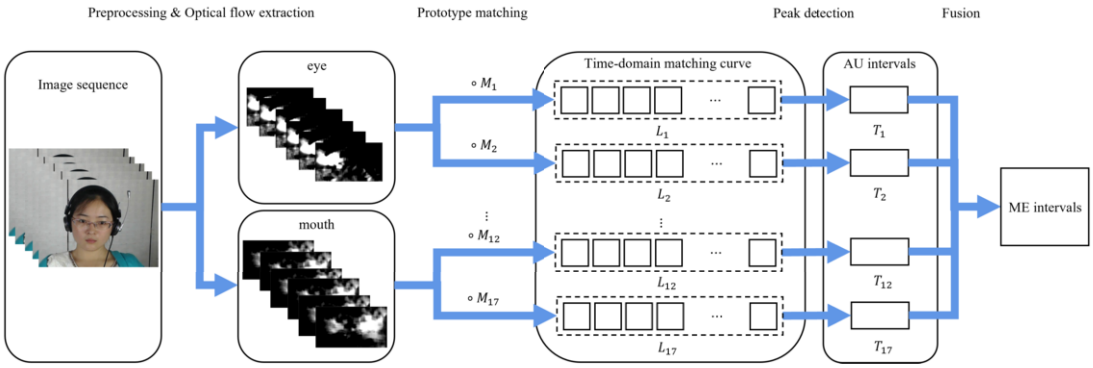


Figure 5. Process diagram of ME spotting step

intervals. We use a static sliding window encompassing N images to segment the video and locate MEs within the sliding window subsequently.

The process of spot micro-expressions based on AU prototypes in the image sequence involves four stages: facial region alignment, AU prototypes matching, AU detection and fusion. Consequently, the micro-expression intervals set $T = \{[start_1, end_1], [start_2, end_2], ..., [start_{last}, end_{last}]\}$ is acquired. Motivated by [6], the sliding window's step size (S) is adaptively modified based on the spotted expression's location.

If no micro-expression is identified within the current window:

$$S = N/2 \tag{19}$$

Conversely,

$$S = end_{last} + 1 \tag{20}$$

In the subsequent sections, we will elaborate on several techniques employed in our suggested micro-expression spotting method.

2.4.1 AU Prototypes Matching

The micro-expression spotting method based on AU prototype is shown in Figure 5. The first image I_1 is used as a reference for the image sequence $I = \{I_1, I_2, ..., I_N\}$ to perform preprocessing and optical flow feature extraction (as shown in section 2.1 and 2.2). After these operations, we get the optical flow field sequence $\{F_1^{eye}, F_2^{eye}, ..., F_{N-1}^{eye}\}$ of the eye region and the optical flow image sequence $\{F_1^{mth}, F_2^{mth}, ..., F_{N-1}^{mth}\}$ of the mouth region. F_i^{eye} is a dense optical flow field of the eye region between image I_1 and I_{i+1}, $F_i^{eye} \in \mathbb{R}^{h_{eye} \times w_{eye} \times 2}$.

AU prototype matching index p_i^j between the optical flow fields (F_i^{eye}, F_i^{mth}) and AU prototype M_j is calculated as equations 21 and 22.

$$D_i^j = \begin{cases} F_i^{eye} \circ M_j, j \in H_{eye} \\ F_i^{mth} \circ M_j, j \in H_{mth} \end{cases} \tag{21}$$

$$p_i^j = \begin{cases} \sum_{x=0}^{w_{eye}} \sum_{y=0}^{h_{eye}} D_i^j[x, y], j \in H_{eye} \\ \sum_{x=0}^{w_{mth}} \sum_{y=0}^{h_{mth}} D_i^j[x, y], j \in H_{mth} \end{cases} \tag{22}$$

The operation "$\circ$" is Hadamard product. If j belongs to H_{eye}, $D_i^j \in \mathbb{R}^{w_{eye} \times h_{eye}}$. Conversely, $D_i^j \in R^{w_{mth} \times h_{mth}}$. w_{mth} and h_{mth} are width and height of I^{MTH}.

The above calculation is performed on the **prototype** in $\{M_1, ..., M_V\}$ and the optical flow field in $\{F_1^{eye}, F_2^{eye}, ..., F_{N-1}^{eye}\}$ and $\{F_1^{mth}, F_2^{mth}, ..., F_{N-1}^{mth}\}$. Therefore, for an image sequence of length N, a $V \times (N-1)$ matrix W can be calculated to describe the occurrence of AUs.

$$\mathbf{W} = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_V \end{bmatrix} = \begin{bmatrix} p_1^1 & p_2^1 & \cdots & p_{N-1}^1 \\ p_1^2 & p_2^2 & \cdots & p_{N-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ p_1^V & p_2^V & \cdots & p_{N-1}^V \end{bmatrix} \tag{23}$$

Matrix $\mathbf{W}$ contains V time-domain prototype matching curves. When a micro-expression with AU i occurs, we can observe a peak appearing on the prototype matching curve L_i. Figure 6 shows the time-domain prototype matching curves of AU4 and AU12 when only a frowning action occurs, and the sample of ME is shown in Figure 7. There is an apparent peak in L_4 (curve of AU4), while L_{12} (curve of AU12) remains smooth. The highest point of the curve represents the peak image, which often contains the most ME information.

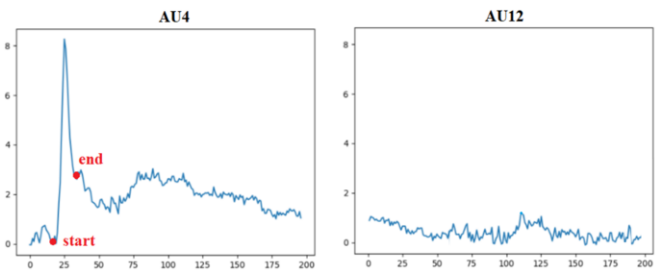
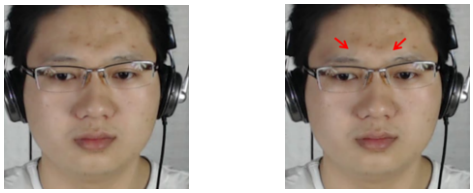


Figure 6. An example of time-domain prototype matching curves corresponding to AU4 and AU12 [The ordinate represents the AU prototype matching index, and the abscissa represents the image index]

Therefore, we can estimate AU intervals by detecting peaks of time-domain prototype matching curves.

2.4.2 Peak detection and fusion method

We detect peaks by calculating the difference between the curve value of point i and the minimum value of the surrounding areas.



(a) the onset image (b) the apex image

Figure 7. Schematic diagram of ME onset image and apex image [15]. This micro-expression only contains AU4 (frown action), and the arrow points to the area where AU4 occurs.

m is the radius of the search area.

$$\begin{cases} L_j[i] - min(L_j[i-1], ..., L_j[i-m]) > \theta_j \\ L_j[i] - min(L_j[i+1], ..., L_j[i+m]) > \theta_j \end{cases} \tag{24}$$

If the difference is large enough, we consider that point i belongs to a peak, as shown in equation 24. We calculate all points on the curve. θ_j is the peak detection threshold for AU j. After this operation, we obtain peak intervals of AU j $T_j = \{[start_1, end_1]_i, ..., [start_\tau, end_\tau]_j\}$.

We calculate the occurrence time of all action units. When multiple AUs appear at a similar time, we analyze whether two AUs ($[start_m$,$end_m]$ and $[start_n$,$end_n]$) belong to the same micro-expression by calculating the overlap index $\boldsymbol{\mu}$.

$$\boldsymbol{\mu} = \frac{max(min(end_n, end_m) - max(start_n, start_m), 0)}{min((end_m - start_m), (end_n - start_n))} \tag{25}$$

If $\boldsymbol{\mu}$ is larger than threshold δ, we consider these two AUs belong to the same micro-expression. And then, fuse two AU intervals to get a new interval $[start_{new}, end_{new}]$.

$$\begin{aligned} start_{new} &= min(start_n, start_m) \\ end_{new} &= max(end_n, end_m) \end{aligned} \tag{26}$$

After the fusion operation of each two AUs, we obtain the ME intervals $T_{ME} = \{[start_1, end_1], ..., [start_E, end_E]\}$.

3 EXPERIMENTS AND RESULTS

3.1 Databases and Metric

We evaluate our approach using the public dataset CASMEII [16] and CAS(ME)^2 [15]. CASMEII is the most popular micro-expression recognition database, which contains 256 micro-expression samples from 26 subjects. CASMEII samples are labeled as 5 categories: happiness, surprise, disgust, repression and others. CAS(ME)^2 is one of the most widely used databases for MaE and ME spotting, which is recorded with a frame rate of 30 fps and a resolution of 640 × 480. CAS(ME)^2 includes 57 micro-expressions and 300 macro-expressions. The total duration of the CAS(ME)^2 database is 138.34 minutes, and the time of the expression occurrence is 6.31 minutes. It is known that the occurrence of expressions in the CAS(ME)^2 database is sparse.

We discriminate whether a spotted interval $Q_{spotted}$ is a correct detection by calculating IoU between the spotted interval $Q_{spotted}$ and the ground-truth interval $Q_{groundTruth}$.

$$\frac{Q_{spotted} \cap Q_{groundTruth}}{Q_{spotted} \cup Q_{groundTruth}} \geq k \tag{27}$$

Table 1. Details of databases

	CAS(ME)^2	CASMEII
Video samples	87	256
MaEs	300	0
MEs	57	256
Resolution	640×480	640×480
FPS	30	200

The threshold k is set to 0.5. If the IoU is larger than threshold k, $Q_{spotted}$ is a true positive result (TP).

We calculate the F1-score of macro-expressions, micro-expressions and overall by Equation 30. It can be observed that the F1-score is calculated from the indicator Recall and Precision. This requires spotting as many micro-expressions as possible while avoiding too many false positive detections.

$$\text{Recall}_{\text{All}} = \frac{\text{A}_{\text{ME}} + \text{A}_{\text{MaE}}}{\text{M}_{\text{ME}} + \text{M}_{\text{MaE}}} \tag{28}$$

$$\text{Precision}_{\text{All}} = \frac{\text{A}_{\text{ME}} + \text{A}_{\text{MaE}}}{\text{N}_{\text{ME}} + \text{N}_{\text{MaE}}} \tag{29}$$

$$\text{F1-score} = \frac{2 \times (\text{Recall} \times \text{Precision})}{\text{Recall} + \text{Precision}} \tag{30}$$

A_{ME} and A_{MaE} are amounts of true positive results; M_{ME} and M_{MaE} are amounts of all ME and MaE intervals; N_{ME} and N_{MaE} are amounts of spotted ME and MaE intervals.

3.2 Experiment configuration

In this experiment, to test the transfer ability of the method and increase the challenge difficulty. We train the prototype on the CASMEII [16] dataset and conduct experiments on the CAS(ME)^2 [15] dataset for spotting micro- and macro-expressions, respectively. The specific experimental setup is shown below:

The length of the sliding window is set to the number of images contained in a seven-second video interval. We use the same method to spot macro- and micro-expressions. Expressions that last less than 0.5s are considered micro-expressions. We use the Dlib tool [8] to locate 68 landmarks of the face. When calculating the eye region, the parameters $\gamma 1$ is set to 0.2, $\gamma 2$ is set to 0.6, $\gamma 3$ is set to 0.2 and $\gamma 4$ is set to 0.6. The eye region I^{EYE} is normalized to 280 × 240 and the month region I^{MTH} is normalized to 260 × 160 after preprocessing.

Before the peak detection, the curves are passed through a low-pass filtering process, and any component with a value below zero is set to zero. In the peak detection method, the search radius m is set to the number of images contained in a one-second interval of the video. The threshold δ for interval fusion is 0.33 in the local movement fusion method.

3.3 Analysis of experimental results

3.3.1 AU Prototypes

In this paper, we construct AU prototypes based on the CASMEII database [16]. Eight AUs appearing more than ten times in the CASMEII are selected. The definition of selected AUs is shown in Table 1. Each AU in set {AU1, AU2, AU4, AU7} only occurs in the eye region, while the AU in set {AU12, AU14, AU15, AU17} occurs in the mouth region.

Table 2. The definition of selected AUs.

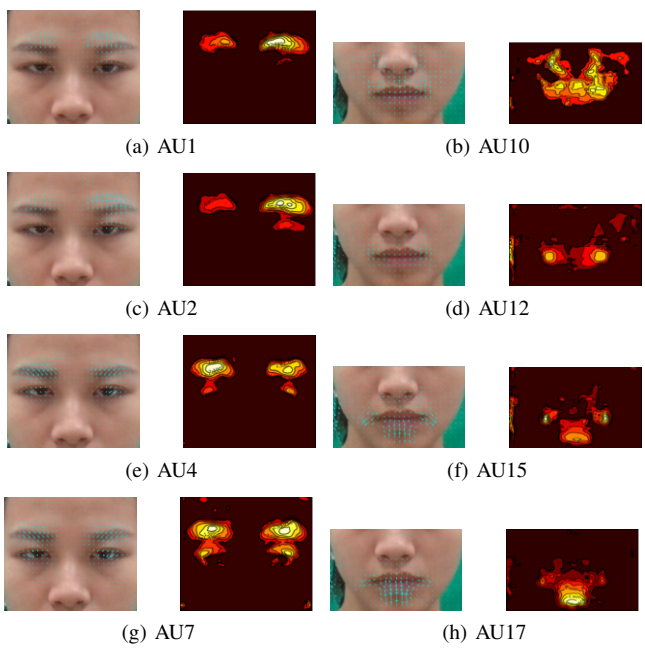
AU	definition	AU	definition
AU1	brow inner corner raise	AU12	Lip corner puller
AU2	brow outer corner raise	AU14	Lip corner tight
AU4	Brow lower	AU15	Lip corner down
AU7	Lids tight	AU17	Lower lip raise

An AU prototype comprises two matrixes representing the horizontal and vertical motion trends, $M_j = \{M_j^x, M_j^y\}$. To clearly observe the motion information represented by each AU prototype, we draw arrows on uniformly spaced pixels, as shown in Figure 8. We call them as the directional map. Specifically, for the arrow starting at pixel point $p = (\alpha, \beta)$ the pointing position p_{end} is shown in equation 31. φ is an amplification factor.

$$p_{end} = (\alpha + \varphi * M_j^x[\alpha, \beta], \beta + \varphi * M_j^y[\alpha, \beta]) \tag{31}$$

At the same time, in order to facilitate the observation of the movement amplitude information of the AU prototype, We calculate the magnitude matrix H_j from ${M_j}^x$ and ${M_j}^y$, j is the index of AU.

$$H_i = \sqrt{M_i^x \circ M_i^x + M_i^y \circ M_i^y} \tag{32}$$



(a) AU1 (b) AU10 (c) AU2 (d) AU12 (e) AU4 (f) AU15 (g) AU7 (h) AU17

Figure 8. Examples of heatmap and directional map related to AU prototypes

As shown in Figure 8, we draw heatmaps according to magnitude matrixes. We can see when the AU1 or AU2 occurs, the movement of the eyebrow area is significant. AU1 and AU2 often appear together in micro-expressions with surprising emotions, which have similar motion direction. However, we can observe from the heatmap that, compared to AU1, the area with the highest movement amplitude in the AU2 prototype is closer to the outer side of the eyebrow, which is consistent with the definition of AU2. From the Figure 8, Action Units (AUs) 1, 2 and 4 have similar motion locales, yet notable difference are discernible when examining the directional map. AU4 and AU7 are often accompanied by negative emotions and can occur either alone or together. It can be observed that the movement of AU7 is more pronounced at the lid compared to AU4, which is also in accordance with the definition in table 2. When AU12 and AU15 occur, heatmaps show that the movement of the lip corner area is significant. we can see the clear difference in the direction of their movement in the directional map. The definition of AU17 is raising the lower lip, and the heatmap shows that the movement of the middle part between the lower lip and the lower jaw is apparent.

In AU prototypes, the value of the AU-independent region approaches zero. Therefore, when analyzing facial movements, the AU prototype can effectively remove the interference of dynamic information irrelevant to this AU.

3.3.2 *Macro- and micro-expression spotting results*

We spot macro-expressions (MaEs) and micro-expressions (MEs) on CAS(ME)^2 dataset and the result is shown in Table 2. The F1-score is 0.4337 for macro-expressions, 0.2857 for micro-expressions, and 0.4162 for all expressions. Because of the small and fast movement of micro-expressions, the F1-score of ME is much lower than MaEs.

Table 3. Performance on Macro- and Micro-expression.

	Macro-expression	Micro-expression	overall result
Total number	300	57	357
TP	113	10	123
FP	94	3	101
FN	187	47	234
Precision	0.3766	0.1754	0.3445
Recall	0.5113	0.7692	0.5256
F1-score	0.4337	0.2857	0.4162

Thus far, micro-expression databases are manually annotated. Nevertheless, due to the subtle nature of micro-expression, delineating the boundaries of micro-expression occurrences is often challenging. Consequently, expressions location labels in the database are inherently susceptible to the presence of inaccuracies. In light of this, the present study aims to moderately lower the threshold k of the overlap index, in order to obtain spotting results that cater to different overlap degree requirements, as illustrated in Table 5. The results reveal that when the threshold requirement is diminished, the spotting result F1-score experiences a significant enhancement. This outcome is not readily apparent. As demonstrated in Section 3.1, the occurrence of expressions in the long videos of the CAS(ME)^2 database is sparse. This observation suggests that some expressions are indeed spotted, but with an IOU of less than 0.5 with the label.

Table 4. Spotting results of CAS(ME)^2 database with different threshold values.

k	Precision	Recall	F1-score
0.5	0.3445	0.5256	0.4162
0.4	0.3865	0.5897	0.4670
0.3	0.4341	0.6623	0.5245

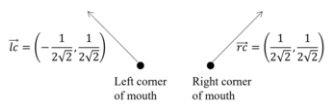
Table 5 compares different methods for spotting MaEs, MEs and overall in terms of F1-score on CAS(ME)^2. The result of our method is significantly higher than other studies. Among them, paper [20], [9] implemented deep learning methods to spot micro-expressions.

We think that in the case of smaller amounts of data in micro-expression databases, the performance of hand-crafted features may be more stable.

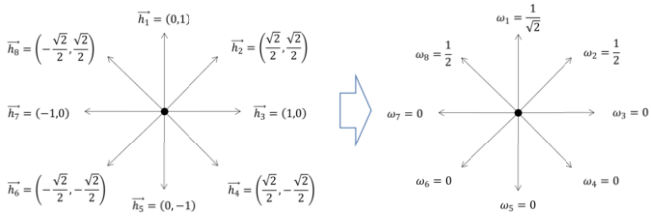
Table 5. Comparison with other methods.

	Macro-expression	Micro-expression	overall result
Gan [4]	0.1436	0.0098	0.0448
Zhang [21]	0.2131	0.0547	0.1403
Yu [20]	0.3800	0.0630	0.3270
Yang [18]	0.2599	0.0339	0.2118
He [7]	0.4169	0.1202	0.3530
Leng [9]	0.3357	0.1590	0.3117
Proposed	**0.4337**	**0.2857**	**0.4162**

While papers [21], [7] use optical flow methods to extract motion information. The paper [7] is the expanded version of work [6] which achieved first place in the 2021 MEGC competition. It estimates the magnitude of local facial motion by calculating the amplitude of the average optical flow in facial regions of interest and constructs a temporal waveform curve. Correspondingly, our study analyzes the presence of AUs corresponding to ME by matching with AU prototype. Since these two methods have similar peak detection techniques, we use [7] as a baseline. As can be observed from Table 5, the feature extraction based on AU prototype significantly improves the spotting performance. We believe that the reason for the better results is that our method has better robustness, effectively reducing the impact of expression-unrelated noise on micro-expression feature extraction and eliminating some false positive results. We use the matching process of head-shaking noise with the AU12 prototype as an example, as shown in Figure 9.



(a) Simplified prototype of AU12



(b) 8 head-shaking unit vectors and matching results

Figure 9. The matching process between the head-shaking noise and the AU12 simplified prototype.

For the convenience of representation, we simplify the AU12 prototype as vectors at the two points of the left and right mouth corners, denoted as $\vec{lc}$ and $\vec{rc}$, as shown in Figure 8(a). The actual movement at the corners of the mouth is noted as $\vec{h}_l$ and $\vec{h}_r$. The matching result is noted as ω.

$$\omega = \max(\vec{h}_l \circ \vec{lc} + \vec{h}_r \circ \vec{rc}, 0)$$

When AU12 appears at the corners of the mouth, the motions of the two mouth corners are set to be two unit vectors, and the matching result with the simplified AU12 prototype is noted as ω_{AU12}, with $\omega_{AU12} = 1$.

We then set eight unit vectors to represent eight head-shaking directions, noted as $\{\vec{h}_1, \vec{h}_2, \ldots, \vec{h}_8\}$, as shown in Figure 9(b); when only the head shakes, the actual movement at the corners of the mouth is $\vec{h}_l = \vec{h}_r = \vec{h}_v$. The results of matching the head-shaking vectors with the simplified AU12 prototype are calculated separately and noted as $\{\omega_1, \omega_2, \ldots, \omega_8\}$, and the results are shown in Figure 9(b). It can be observed that matching with the AU12 prototype has a certain attenuating effect on the head movement noise belonging to the $\{\vec{h}_1, \vec{h}_2, \vec{h}_8\}$ direction (the matching result is less than 1); for the head movement belonging to the $\{\vec{h}_3, \vec{h}_4, \ldots, \vec{h}_7\}$ direction, it completely will not match the AU12 prototype (the matching result is 0). In summary, our proposed method has good robustness to ME-unrelated noise such as head-shaking, and can effectively improve the reliability of spotting results.

3.4 Ethical Concern

It is essential to acknowledge potential biases of our method that may arise from collecting and labeling the training data. Current micro-expression data are collected in a lab-controlled environment, and labeled subjectively by human annotators. Our models may internalize these biases, leading to inaccuracies for certain demographic groups. Moreover, personal and sensitive information could be revealed by ME, so informed consent is needed for their ethical development and deployment. Safeguarding the privacy of both raw data and learned patterns is of utmost importance. Addressing these issues is essential to ensure the deployment and ethical development of ME analysis technologies.

4 Conclusions

In this study, we design an automatic micro-expression spotting method based on AU prototypes. It mainly includes AU prototype construction and matching process. First, we compute the optical flow field to capture the tiny facial movements. Secondly, we take advantage of the feature that AU has similar motion patterns to suppress expression-unrelated noise and construct pure AU prototypes. Subsequently, by calculating the prototype matching index, a comprehensive analysis of micro-expressions is achieved. This matching approach has good robustness and interpretability.

In experiment, we construct the prototype on the CASMEII dataset and spot micro- and macro-expressions on the CAS(ME)^2 dataset. Visualization results of AU prototype help us understand its motion patterns. The F1-score of micro- and macro-expressions spotting on the CAS(ME)^2 is 0.4162. Compared with other micro-expression spotting methods, the spotting results of this method have been greatly improved. In the future work, we will try to construct ME spotting and recognition system based on multi-feature fusion.

Acknowledgements

This work is supported in part by the National Key R&D Program of China (2022YFC3301800,2020YFC0833204), Provincial Key R&D Program of Heilongjiang (GY2021ZB0206), Shenzhen Foundational Research Funding (JCYJ20200109150814370), Funds for National Scientific and Technological Development (2021SZVUP087, 2021SZVUP088)

References

[1] P Ekman and W V Friesen, 'Nonverbal leakage and clues to deception', *Psychiatry Interpersonal Biological Processes*, **32**(1), 88–106, (1969).

[2] Paul Ekman, 'Lie catching and microexpressions', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**(1521), 118–133, (2009).

[3] Paul Ekman and Wallace V Friesen, 'Facial action coding system (facs): a technique for the measurement of facial actions', *Rivista Di Psichiatria*, (1978).

[4] YS Gan, Sai-Kit Liong, Dong Zheng, Shuang Li, and Chen Bin, 'Optical strain based macro-and micro-expression sequence spotting in long video', in *IEEE International Conference on Automatic Face and Gesture Recognition*, (2020).

[5] Ying He, Su-Jing Wang, Jingting Li, and Moi Hoon Yap, 'Spotting macro-and micro-expression intervals in long video sequences', 742–748, (2020).

[6] Yuhong He, 'Research on micro-expression spotting method based on optical flow features', in *Proceedings of the 29th ACM International Conference on Multimedia*, p. 4803–4807, New York, NY, USA, (2021). Association for Computing Machinery.

[7] Yuhong He, Zhongliang Xu, Lin Ma, and Haifeng Li, 'Micro-expression spotting based on optical flow features', *Pattern Recognition Letters*, **163**, 57–64, (2022).

[8] D. E. King, 'Dlib-ml: A machine learning toolkit', *Journal of Machine Learning Research*, **10**(Jul), 1755–1758, (2009).

[9] Wenhao Leng, Sirui Zhao, Yiming Zhang, Shiifeng Liu, Xinglong Mao, Hao Wang, Tong Xu, and Enhong Chen, 'Abpn: Apex and boundary perception network for micro- and macro-expression spotting', in *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, p. 7160–7164, New York, NY, USA, (2022). Association for Computing Machinery.

[10] Bo Li, Zhi Zhang, Rui Cao, et al., 'Automatic double region of interest selection for spotting micro-expression from long videos', in *2020 International Conference on Computer Engineering and Application (ICCEA)*. IEEE, (2020).

[11] Y. Li, W. Peng, and G. Zhao, 'Micro-expression action unit detection with dual-view attentive similarity-preserving knowledge distillation', in *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pp. 01–08. IEEE, (2021).

[12] Sai-Kit Liong, John See, Kok-Seng Wong, et al., 'Automatic apex frame spotting in micro-expression database', in *IAPR Asian Conference on Pattern Recognition*, pp. 665–669. IEEE, (2015).

[13] Albert Mehrabian, 'Communication without words', *Psychological Today*, **2**(4), 53–55, (1968).

[14] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää, 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**(7), 971–987, (2002).

[15] Fangbing Qu, Su-Jing Wang, Wen-Jing Yan, He Li, Shuhang Wu, and Xiaolan Fu, 'Cas(me)^2 : A database for spontaneous macro-expression and micro-expression spotting and recognition', *IEEE Transactions on Affective Computing*, **9**(4), 424–436, (2018).

[16] Wen-Jing Yan, Xiaobai Li, Su-Jing Wang, Guoying Zhao, Yong-Jin Liu, Yu-Hsin Chen, and Xiaolan Fu, 'Casme ii: An improved spontaneous micro-expression database and the baseline evaluation', *PLOS ONE*, **9**(1), 1–8, (01 2014).

[17] Wen-Jing Yan, Qi Wu, Jing Liang, Yu-Hsin Chen, and Xiaolan Fu, 'How fast are the leaked facial expressions: The duration of micro-expressions', *Journal of Nonverbal Behavior*, **37**(4), 217–230, (2013).

[18] Bo Yang, Jianming Wu, Zhiguang Zhou, et al., 'Facial action unit-based deep learning framework for spotting macro- and micro-expressions in long video sequences', in *Proceedings of the 29th ACM International Conference on Multimedia*, (2021).

[19] Chuin Hong Yap, Connah Kendrick, and Moi Hoon Yap, 'Samm long videos: A spontaneous facial micro- and macro-expressions dataset', in *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pp. 771–776, (2020).

[20] Wang-Wang Yu, Jingwen Jiang, and Yong-Jie Li, 'Lssnet: A two-stream convolutional neural network for spotting macro- and micro-expression in long videos', in *Proceedings of the 29th ACM International Conference on Multimedia*, (2021).

[21] Li-Wei Zhang, Jie Li, Sheng-Jun Wang, et al., 'Spatio-temporal fusion for macro-and micro-expression spotting in long video sequences', in *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) (FG)*, pp. 734–741. IEEE, (2020).

[22] Lijun Zhang, Ognjen Arandjelovic, and Xiaopeng Hong, 'Facial action unit detection with local key facial sub-region based multi-label classification for micro-expression analysis', in *Proceedings of the 1st Workshop on Facial Micro-Expression: Advanced Techniques for Facial Expressions Generation and Spotting (FME'21)*, pp. 11–18. Association for Computing Machinery, (2021).

[23] Zhi Zhang, Tianyou Chen, Hongying Meng, et al., 'SMEConvNet: A convolutional neural network for spotting spontaneous facial micro-expression from long videos', *IEEE Access*, **6**, 1–1, (2018).