Micro-expression detection using Integral Projections

Hua Lu Kidiyo Kpalma Joseph Ronsin

Université Bretagne Loire
UMR 6164,F-35708, Rennes, France

hua.lu@insa-rennes.fr kidiyo.kpalma@insa-rennes.fr Joseph.ronsin@insa-rennes.fr

ABSTRACT

Automatic detection of micro-expression from a video is the first step in the micro-expression analysis. In this paper, we present a method addressing the micro-expression detection problem based on the differences in the Integral Projection (IP) of sequential frames. The method can detect the temporal location of the micro-expression. It involves an observation of the Chi-squared distance of the IP to measure the difference between frames. The main advantage of using IP for micro-expression detection is its low computation cost, which brings an important merit in real-time application. To evaluate our method, experiments are completed on two micro-expression databases namely CASME and CASME II. Results on these two datasets show that the proposed method obtains positive promising results with much less computation time against state-of-the-art methods.

Keywords

Micro-expression detection, Integral Projection, Chi-squared distance

1 INTRODUCTION

As a non-verbal behavior, the facial expression plays an important role in our daily life. People can express their feelings by making facial expressions and can also communicate with others by reading facial expressions. During the past two decades, the macro-expression analysis has been paid huge attention in many fields. Psychologists studied the human psychology conveyed by the changing facial expression and computer scientists relied on the digital process to analyze the expression. With the progress of technology, automatic macro-expression detection and recognition can be achieved in real-time and has been successfully applied into business [Del15]. Most of the expressions which are called macro-expressions can be easily observed by naked eyes. However, researchers found some expressions appearing and fading quickly and often easily neglected by naked eyes, these are named as micro-expressions.

Approaches addressing the issues related to micro-expression have been considerably studied in the past 50 years [Hag66]. Recently, an increasing attention has been paid to micro-expression detection and recognition due to its potential application in medicine, business and many other fields, such as catching lies during diagnose, negotiation or interrogation [Ekm69]. Compared to macro-expression, a micro-expression lasts only 1/25s to 1/5s, and moreover, its subtle appearance in part of the face makes naked eyes-based detection and recognition difficult to achieve. Thus, computer vision analysis offers a potential solution.

Micro-expression approaches in computer vision area consist of detecting and classifying them from videos. This inspires a series of approaches on micro-expression analysis integrating computer-aided techniques. Most works of the micro-expression analysis concentrate on the classification step [Hua15], [Liu15], and few works have been devoted to the detection, which is the foundation of the analysis. So far, several methods have been developed for this detection, such as method based on 2D/3D histogram of oriented gradients, local binary patterns and optical flow which will be presented and then used for comparison with the proposed method.

Polikovsky et al. [Pol09] divided the face into different facial regions and used the 3D histogram of oriented gradients descriptor (3D HOG) for feature extraction. The recognition applied the k-means method to cluster the extracted features of each region. The results showed good performance rates (all over 80 %) in the regions of the forehead, between the eyes and lower nose. However, the experiments were conducted in a small dataset that only contained 13 posed micro-expressions instead of the spontaneous micro-

expressions. Davison et al. [Dav15] used 2D histogram of oriented gradients (HOG) to extract the features of each frame. The Chi-squared distance measure was applied to compute dissimilarity between the sequence frames. However, in paper [Dav15], all detected micro-movements up to 100 frames (200 fps) were classified as true positive including blinks and the eye gaze, without comparing the ground truth of the micro-expression.

Moilanen et al. [Moi14] adopted local binary patterns (LBP) to extract the features from the divided blocks of the face. The method relied on calculating the dissimilarity of features for each block by using Chi-squared distance. The detection experiments were conducted on the spontaneous facial micro-expression datasets in order to solve the problem in practice.

Shreve et al. [Shr14] developed a method for the segmentation of macro- and micro-expression frames by calculating the deformation of facial skin using optical flow (OF). The optical flow is a well-known motion estimation technology and can well spot the subtle movement, but its calculation costs expensively computation time.

Besides, some papers addressing the problem of the detection by training a model to determine if a sequence does or does not contain a micro-expression. Pfister et al. [Pfi11] extracted spatio-temporal local texture features from video sequences and used machine learning algorithms (SVM, MKL, RF) for classification. Xia et al. [Xia16] utilized an adaboost model to compute the initial probability for each frame and the correlation between frames in order to generate a random walk (RW) model. The random walk model was used to calculate the deformation correlation between frames and provide the probability of having micro-expressions in a sequence.

Instead of developing a training model, we propose a new micro-movement detection method by invoking the IP as a feature descriptor to characterize changes in the divided blocks of the face. The IP feature is extracted on each individual block. The Chi-squared distance is used to measure the IP feature dissimilarity between frames so as to observe for possible micro-expression in the frame sequence. The proposed method is evaluated on two widely used datasets through experimental comparison with some popular feature extractors such as the OF, LBP and HOG. The proposed method is an unsupervised model. One of the main advantages of our model is its computation complexity: our model can obtain better or comparable results than the existing models using the OF, LBP and HOG, but requiring much less computation time.

The rest of the paper is organized as follows: Section 2 outlines the background on micro-expression and the IP computation. Section 3 describes the procedure of micro-movement detection and experimental results are discussed in Section 5. Finally Section 6 concludes the paper with the discussions.

2 BACKGROUND

2.1 Micro-expression

A micro-expression is a brief, spontaneous expression, which reveals the true feeling that people try to hide and suppress. Regardless of whether they are macro or micro, facial expressions are dynamic temporal process that match the time and duration of facial deformations and are described with three important concepts: onset, apex and offset [Bet12], [Wei09]. The onset is the point at which the expression starts to show up, the apex is the instant when the deformation of the expression reaches the peak and the offset represents the instant when the expression fades away. Hence, the micro-expression detection is a temporal segmentation of videos, which includes locating the micro-expression appearance instant and providing the duration between the onset and offset. Fig. 1 present an image sequence of the micro-expression which is labeled by 'disgust' in CASME II. Due to the limit space, five frames are presented including 1th(onset), 15th, 29th (apex), 57th and 86th (offset).

Obviously, the duration of the micro-expression is variable. In general, the micro-expression lasts from 1/25s to 1/5s [Ekm69] but other papers show the lasting time of micro-expressions can extend to 1/2s [Mat00], [Mat11]. In our method, the duration from 1/25s to 1/2s will be considered as the guideline.

2.2 Integral Projection

Due to the difficulty for people to read micro-expressions, it is necessary to find appropriate methods for catching subtle and rapid changes of the face. The Integral Projection is presented in the following and it holds as a useful technique for the extraction of facial features. As IP can be extremely effective in determining the position of features, Brunelli et al. [Bru93] applied IP on the human face recognition. In a recent work [Hua15], a combinational method of the IP and LBP was chosen for micro-expression recognition thanks to its ability for providing the shape property of facial images.

IP is a simple and rapid feature extraction method which can reduce the 2D Image features to a simply 1D data. Let $\Omega \subset \mathbb{R}^2$ be the image domain and $I : \Omega \times D \to \mathbb{R}$ be a sequence of gray level images, where $D \subset \mathbb{R}$ is the time space. At each point $(x, y) \in \Omega$ and at time t, the intensity value is denoted by $I(x, y, t)$,

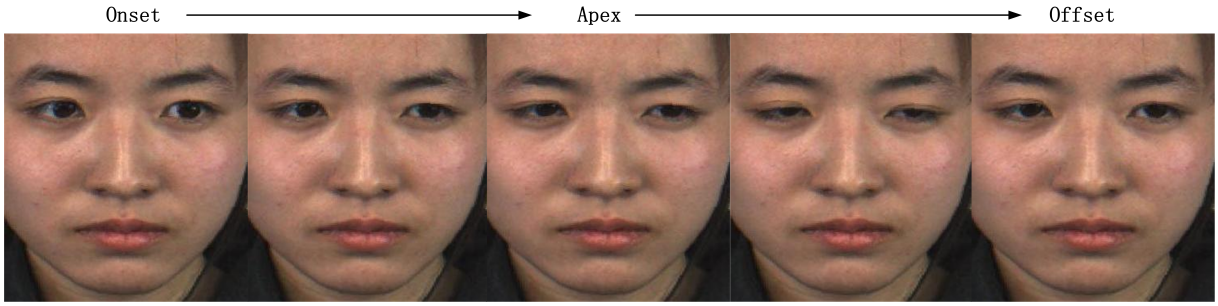


Figure 1: An example of a micro-expression sequence. Five frames are presented including 1th(onset), 15th, 29th (apex), 57th and 86th (offset).

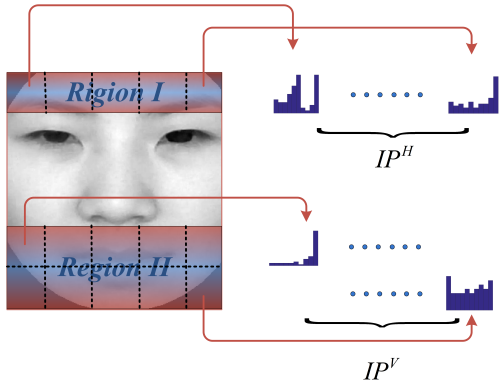


Figure 2: An example of the IP histograms. Plots in the first (resp. second) row correspond to the horizontal (resp. vertical) IP function from each block.

and the typical formula of the IP function can be expressed as:

$$IP_t^H(x) = \frac{1}{y_2 - y_1} \int_{y_1}^{y_2} I(x,y,t)dy, \quad (1)$$

$$IP_t^V(y) = \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} I(x,y,t)dx, \quad (2)$$

where IP_t^H and IP_t^V are the horizontal and vertical integral projection vectors in the rectangle $[x_1,x_2] \times [y_1,y_2]$ at time t, respectively. Fig. 2 shows examples of the IP histogram curves (horizontal and vertical).

3 PROPOSED METHOD

The flow chart of the proposed method is summarized in Fig. 3 and will be detailed in the following steps. Two main parts are presented in this flow chart: one part is the global pre-processing and featuring, and the other is the extraction of the micro-expression. The part I includes tracking, registering, cropping, masking, blocking the face, the IP features extraction, chi-squared distance analysis, thresholding, peak detection. The part II consists of the micro-expression extraction.

3.1 Face tracking and Registration

Determining the existence and locations of faces in each frame of the sequence is the first step for micro-expression detection. The crucial point of this step is the key points detection for cropping the face. In this section, we choose the Supervised Descent method [Xio13] for facial expression points tracking, from which we can obtain 49 facial key points to register and crop the face.

Since the algorithm depends on careful positioning of the face, the alignment step is necessary for the purpose of keeping eyes in horizontal line. By using the facial key points located on the inner eye corners to calculate the angle $\theta \in [0,\pi)$ between the line of the two eyes and a horizontal line, face alignment operation can be performed such that $\theta = 0$.

3.2 Crop, Mask and Divide face into blocks

The nasal spine point is considered as the fixed point for cropping the face. The regions containing inhomogeneous background, clothes, hairs, and eyes which may influence the micro-expression detection results will be removed by a face mask. Thus, one can focus on the regions which only contain useful information. During the process of calculating the IP over the whole face, some important spatial information may be missed due to global merging of observations and hence giving difficulties to identify subtle changes of face. Therefore, in order to obtain the accurate spatial information for the detection of micro-expression, two blocked regions of interest (ROIs) are defined for IP computing: Region I and Region II respectively involving N and $2N$ blocks, as shown in Fig. 2. The number N will be discussed in section 4.1.2. IP can be calculated in each block to locate small movement for micro-expression analysis.

3.3 Feature Extraction Using IP

Once obtained the cropped and blocked face regions, the IP histograms for each block is computed and then fused for the corresponding region. For the blocks in region I, horizontal IP will describe better the change of the facial skin such as the quick movement of the eyebrow. For the blocks in region II, the micro-movement of the mouth will be well featured by the vertical IP.

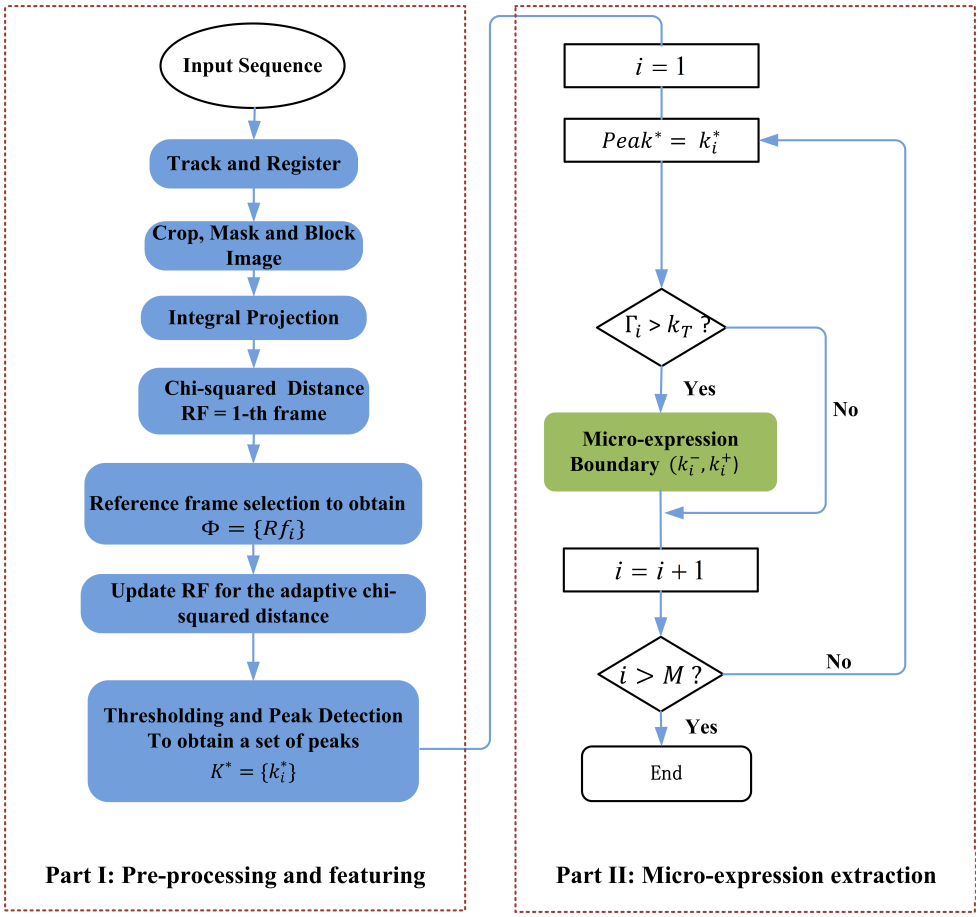


Figure 3: Flow diagram of the proposed algorithm.

Thus, the horizontal IP feature the region I, while the vertical IP feature the region II, as shown in Fig. 2.

3.4 Feature difference analysis

For feature difference analysis, scanning the sequence, subtraction will be performed between a reference frame (RF) and each successive frame denoted as current frame (CF). This reference frame must be a neutral face or onset frame of a temporal facial expression for highlighting differences along the sequence. Differences will be observed from integral projections. IP_t^H and IP_t^V features are extracted from each frame at each block of the two regions, followed by chi-squared distance computation [Moi14] to measure the dissimilarity between the ROIs of the CF and the ROIs of the RF. The chi-squared distance is an efficient method to compute the distance between the features. Given two IP features of P and Q, the chi-squared distance is defined by

$$\chi^2(P,Q) = \frac{1}{2}\sum_{i=1}^{n}\frac{(P(i)-Q(i))^2}{P(i)+Q(i)}, \qquad (3)$$

where n is the length of the IP features.

The regions I and II generate two chi-squared distance sequences which are denoted by S_1 and S_2, respectively. The computation of $S_j (j = 1,2)$ for the k-th frame can be expressed as

$$S_j(k) = \chi^2(P_0^j, P_k^j) \quad \forall k \in [1, L-1], \qquad (4)$$

where P_0^j and P_k^j are the IP features of the reference frame and the k-th frame at the regions I $(j = 1)$ and II $(j = 2)$. The chi-squared distance S used for micro-movement detection is computed by

$$S(k) = \frac{1}{2}(S_1(k) + S_2(k)), \qquad (5)$$

which involves the mean values of the normalization of the sequences S_1 and S_2 at the respective location. Normalize the sequences S_1 and S_2 respectively by the values of $\sqrt{\sum_k S_1^2(k)}$ and $\sqrt{\sum_k S_2^2(k)}$.

3.5 Reference frames selection

For very long videos segmentation, it is necessary to select different RFs since taking the first frame as the RF will lead to accumulating errors along the sequence. To solve this problem, a new reference frame selection method is proposed. Before RF selection, one needs to apply low-pass filtering in order to eliminate high frequency details that may influence the result. We give an

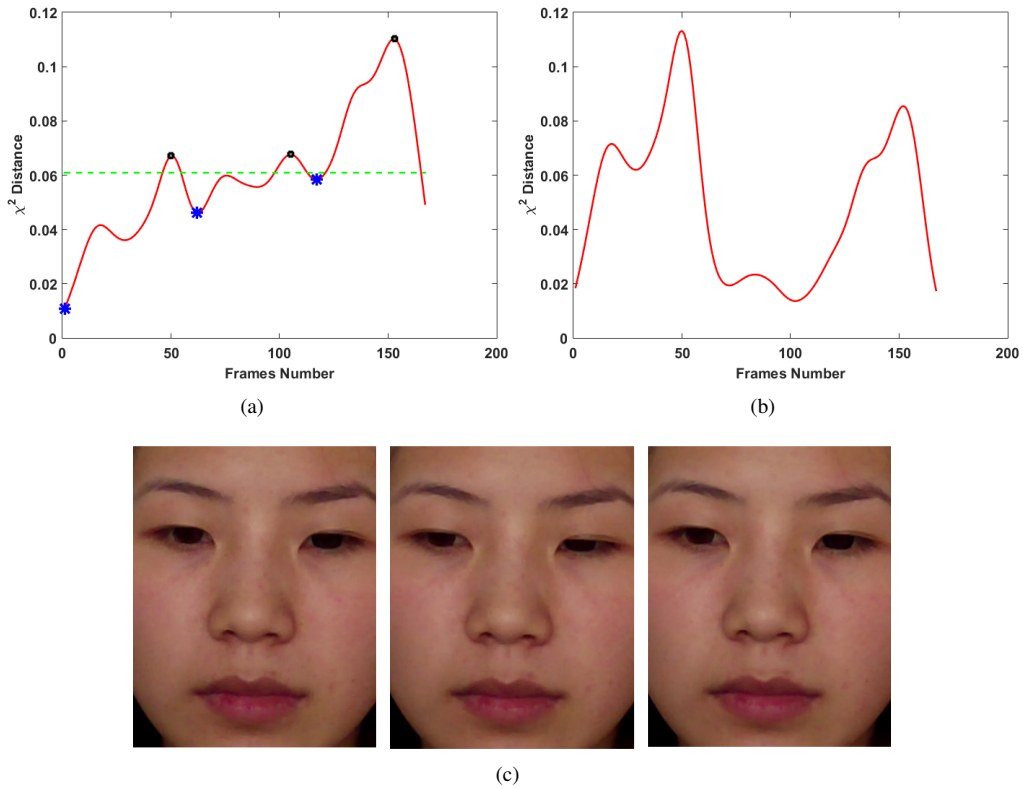


Figure 4: An example of the detection of Φ. **(a)** The red curve describes the chi-squared distance S. The mean value of S are denoted by a green line. The collection Ψ and Φ are described by black dots and blue star points, respectively. **(b)** illustrates the curve for the chi-squared distance S after updating the RF from the collection Φ. **(c)** shows the three RF of Φ at 1, 62, 117.

example in Fig. 4a, where we plot the curve (red solid curve) for the chi-squared distance S in Eq. (5) when the first frame is selected as the RF. We can see that local maximums of the values of S get larger along the sequence, which may introduce bias estimation to the threshold value used for the micro-expression detection in Section. 3.6.

To cope with this possible bias, we define Φ as a collection of the reference frame indexes which can be expressed as

$$\Phi = \{Rf_i\}_{1\leq i\leq m}, \quad m \in [1, L-1],$$

where m is the total number of the reference frames and Rf_i is the index of the i-th RF in the sequence. L is the total number of frames in the sequence. Let $Rf_1 = 1$ be the first frame of the sequence then the remaining elements of the collection Φ can be detected in the following two steps.

Firstly, apply the peak detection procedure to the chi-squared distance S in Eq. (5) to search for a collection Ψ of frame indexes

$$\Psi = \{\zeta_j\}_{1\leq j\leq \tau}, \quad \tau \in [1, L-1],$$

Each element $\zeta_j \in \Psi$ is a local maximizer of the chi-squared distance S. In other words, ζ_j indicates an admissible peak of S such that $S(\zeta)$ is a local maximum value which is larger than the mean of S. We further assume that the elements ζ_j of the collection Ψ admits that $\zeta_i < \zeta_j$, if $i < j$. Secondly, search for the nearest local minimum value from the maximum ζ_j along the positive direction. Each pair of adjacent elements ζ_i, $\zeta_{i+1} \in \Psi$ determines a subsequence of frames, among which a local *minimizer* Rf_i of the computed chi-squared distance S can be obtained. This minimizer is taken as the index of the i-th RF for its notation and it is called Rf_i and $Rf_i \in \Phi$. If there are more than one local minimizer in the subsequence between ζ_i and ζ_{i+1}, we choose the closest one to the frame ζ_{i+1} (in the sense of Euclidean distance of indexes) as the RF.

Starting from the reference frames collection Φ, a new distance sequence is generated by updating the RF. This is done by computing the chi-squared distance between Rf_i and Rf_{i+1} using Rf_i as RF. Fig. 4b is an example of the new chi-squared distance sequence. One can claim that after updating the RF, it is easier to obtain the location of the micro-expression, where the ground truth of the micro-expression given in this example is 39-59 frames. Fig. 4c is an illustration of the detected reference frames which are neutral faces or nearly ones.

In this section, we perform the reference frames selection and the Gaussian smoothing operation on the sequences S to obtain a new adaptive chi-squared distance sequence $S^{'}$ based on the RF collection Φ. The smoothing step aims to remove the noises from the sequences.

3.6 Thresholding and Peak Detection

Once computed the chi-squared distance, it is necessary to use a thresholding method to obtain the location of the micro-expression.

The following steps are applied for the process of the distance sequence $S^{'}$:

1. **Polynomial Fitting.** Apply a second order polynomial fitting operation to the sequence $S^{'}$ by the least square method [Shr11] and generate a fitting function ρ. In Fig. 5a, we demonstrate the plot curve of the function ρ.

2. **Micro-expression Appearance Computation.** Compute a sequence β with the same length of S by subtracting the fitting sequence ρ from the sequence $S^{'}$. All of the negative values of β are set to 0. The expression of β can be found in Eq. (6) and it is illustrated on Fig. 5b.

3. **Peaks Detection.** Apply a peak detection procedure to search for a collection of peaks of the sequence β as the indicators of the appearance of micro-expressions. This peaks detection procedure relies on two threshold values as described in the following.

The polynomial fitting step is able to suppress the cropping errors accumulated over the whole sequence. In step 2, the thresholded sequence β can be computed by thresholding $S^{'}$ with ρ as follows:

$$\beta(k) = \max\left\{S^{'}(k) - \rho(k), 0\right\}, \quad \forall k \in [1, L-1], \quad (6)$$

where ρ is the fitting function and L is the total number of frames in a sequence. The sequence of β involves the information of the existence and location information of the potential micro-expressions.

The peaks detection procedure is carried out dependently of a threshold value T that can be computed by

$$T = \beta_{\text{mean}} + p(\beta_{\max} - \beta_{\text{mean}}), \quad (7)$$

where β_{mean} and $\beta_{\max}$ are the corresponding mean and maximum values of the thresholded sequence β. The scalar value $p \in [0,1]$ is a tuning parameter [Moi14]. This procedure plays the crucial role in the entire course of the micro-expression detection. Thus we give a detailed introduction in the following.

We first detect a collection K^* of M admissible peaks points k_i^* from the sequence β in Eq. (6). Each peak point survives in a subsequence $\Gamma_i \subset [1, L-1]$, where L is the length of the processed frames including the reference frame. These subsequences Γ_i can be considered as the neighborhoods of the corresponding peak point. We supposed that each subsequence Γ_i has only one peak point and is disjoint to another, i.e.,

$$\Gamma_i \cap \Gamma_j = \emptyset, \quad \forall i \neq j.$$

The detection of the collection K^* and the subsequences Γ_i can be done in two sub-steps. First of all, a candidate peak point is a local maximizer of the sequence β within the subsequence Γ_i

$$\beta(k_i^*) \geq \beta(k), \quad \forall k \in \Gamma_i.$$

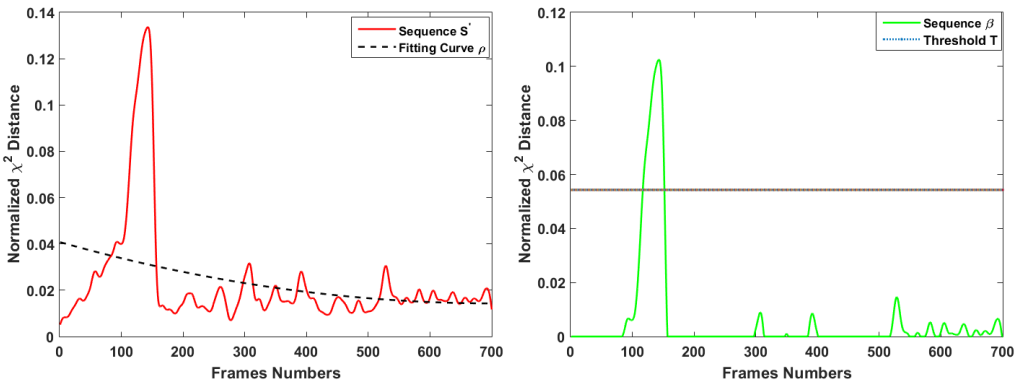
and has a value of β larger than the threshold T. Secondly, we detect the neighborhood Γ_i of this candidate peak point. A subsequence Γ_i can be characterized by the position k_i^* of the candidate peak point and two boundary points k_i^+ and k_i^- such that $\Gamma_i = [k_i^-, k_i^+]$. We search for the position k^+ from the candidate peak point β_i^* along the positive direction till we pass by a point k_* such that $\beta(k_*) < \beta(k_i^*)$, or $\beta(k_*)$ is a local minimum of β, i.e., $\beta(k_*) > \max(\beta(k_*-1), \beta(k_*+1))$, where $\alpha > 0$ is a constant value. Similarly, the position k_i^- is determined along the negative direction. In practice, the value of $\beta(k_*)$ is thought as a local minimum if $|\beta(k_*) - \beta(k_*+1)|$ is small enough. Based on the two sub-steps described above, a candidate peak point k_i^* is admissible if

$$|k_i^+ - k_i^-| > k_T,$$

where k_T is a given threshold value dependent of datasets. Note that the subsequence Γ_i is actually the i-th duration of a micro-expression. The value of $\beta(k_i^*)$ is the i-th value of the peak of the thresholded sequence β. In this step, the values of β at the boundary points k_i^+ and k_i^- are approximately equal to a fraction $\beta(k_i^*)$

$$\beta(k_i^+) \approx \beta(k_i^-) \approx \alpha\beta(k_i^*). \quad (8)$$

In this paper, the constant α determines the detected length of the micro-expression. Fig. 5a illustrates for a video of 700 frames the fitting curve ρ (black color) for $S^{'}$ (red color). The threshold T in Eq. (7) and the sequence β in Eq. (6) used for spotting the micro-expression are shown in Fig. 5b by a horizontal dash line and a green solid curve, respectively. In Fig. 5b, it can be seen from the green curve that a micro-expression is spotted around the frame 143. A duration of $128-150$ frames is detected with $\alpha = 0.8$ which will be kept inside the algorithm. Compared with the referenced ground truth of frames $131-160$ with the peak frame 142, one can see that the detected starting and ending frames are not exactly the same as those of the ground truth but are very close with long overlapping between both. Based on this observation, it is reasonable to claim that obtained results agree with the ground truth.



(a) (b)

Figure 5: An example of the process of the micro-expression detection. **(a)** illustrates the curve of S is fitted by the polynomial fitting and **(b)** provides the step of locating the micro-expression for thresholding the curve of β by T.

4 EXPERIMENTS

For the evaluation, experiments are conducted on two well-known datasets in micro-expression analysis namely CASME [Yan13] and CASME II [Yan14]. Micro-expressions in these two datasets are elicited spontaneously and labeled with reliable ground truth corresponding to the onset, apex and offset frames which can be used for comparisons in the experiments.

4.1 Datasets and experiment sets

4.1.1 Datasets

The dataset of CASME consists of 195 spontaneous micro-expressions which were selected from more than 1500 elicited facial movements and filmed with two cameras corresponding to two classes (class A and B) at a frame-rate of 60fps. As the lighting condition and resolution of pixels differ between two classes, experiments were carried separately. The videos were recorded in natural light in Class A and in a room with two LED lights in Class B. The resolutions of pixels in Class A and B are 1280×720 and 640×480 pixels, respectively. All samples in CASME database were used in the experiment.

CASME II contains 255 micro-expressions sequences selected among 3000 facial movements. Thirty five participants were recruited with a mean age of 22.03 years in the database. The resolution is 640×480 pixels with a frame rate of 200fps such that they could be recorded more detailed information on the facial muscle movements than CASME. All samples from CASME II dataset were used for the evaluation.

The micro-expression is featured by its rapid movement with respect to the short duration. Thus, the total number of frames of the micro-expression duration is limited. In CASME database, micro-expressions with the duration no more than 500ms or facial expressions lasted more than 500ms but their onset duration less than 250ms were selected. The total duration less than 500ms or onset duration less than 250ms were chosen as the final samples in CSAME II.

4.1.2 Parameters setup

In our experiments, a comparison with methods of the optical flow (OF), local binary patterns (LBP) and histogram of oriented gradients (HOG) is provided. Parameters are set up in the following.

For the OF, optical flow is computed using the MATLAB implementation of Black [Sun10], [Bla96].

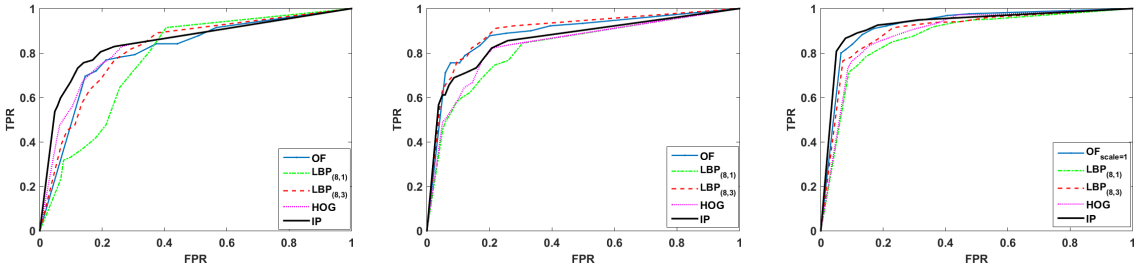
For the LBP, two uniform patterns [Oja96] of $((P,r) = (8,1),(P,r) = (8,3))$ are considered. P corresponds to the number of pixels on the local neighborhood of a circle defined by its radius r.

For the HOG [Dal05a], the histogram angle varies from 0 to π or from 0 to 2π, which corresponds to the 'unsigned' or 'signed' gradient. The number of orientation bins is a segmentation value of histogram angles. Here, 8 orientation bins on 2π angle corresponding to signed gradient are chosen as in [Dav15].

The variable parameter N defines the number of blocks mentioned in Section 3.2 and is set to 5 and α in Eq. (8) is set to 0.8. k_T mentioned in Section 3.6 is set to 2 in CASME dataset and set to 7 in CASME II dataset which corresponds to the minimum duration of the micro-expression.

We give a time window tolerance l to detect positively the appearance of the micro-expression peak. The locations of spotted peaks k_i^* are compared with the provided ground truth, and considered to be true positive if they fall within the frame span of $(onset+l, offset-l)$.

We set parameter $l = 5$ for CASME as discussed in [Moi14], and to $l = 16$ for CASME II same as in [Li15]. As eyes are masked in our experiment, spotted eye blinks are counted as false positives not true positives.



(a) (b) (c)

Figure 6: **a**-**c**: ROC curves for the datasets of CASME-A, CASME-B and Casme II, respectively.

5 RESULTS

Three indicators are adopted for assessing the performance of the algorithm: The receiver operating characteristic (ROC) curve, the area under the ROC curve (AUC) and processing time. The implementation was tested on an Intel Core i7 computer with 16GB of RAM which was equipped with Matlab 2015a.

The ROC curve is used for spotting performance comparison which is illustrated by plotting the true positive rate (TPR) in y-axis against the false positive rate (FPR) in x-axis. The TPR is defined as the number of frames of correctly spotted micro-expression divided by the total number of the ground truth micro-expression frames in the dataset. The FPR is computed as the number of incorrectly spotted frames divided by the total number of non-micro-expression frames in the database. The TPR is in the vertical axis, and the FPR is in the horizontal axis, and p in Eq. (7) is used as the varying threshold parameter with step size of 0.1.

Figs. 6a to 6c show the ROC curves obtained from CASME-A, CASME-B, CASME II for the 4 methods, respectively. Overall, we can observe that the proposed method achieves better performance than other methods (OF, LBP, HOG) in CASME-A and CASME II datasets. Some points with low FPR in ROC curves are meaningful. For example, our proposed method is able to detect 80% of the micro-expression with 4% FPR in CASME II dataset. $\text{LBP}_{(8,3)}$ outperforms $\text{LBP}_{(8,1)}$ on three database and provides the best results in CASME-B.

The area under the ROC curve (AUC) summarizes the spotting performance as shown in Table 1. The high values of the AUC means good performance of the method. The AUC results are positive overall and demonstrate that all 4 methods are efficient for spotting micro-expressions. Among two datasets, a better overall performance can be observed that in CASME II. Two reasons can explain this: one is that subjects in CASME dataset often move their head, and another one is that videos are recorded in a different lighting environment in CASME-B leading to the uneven distribution of the lighting in face. In contrast, CASME II contains short video clips at a frame rate of 200fps and no face moving rapidly leading to better detection results.

Dataset	CASME-A	CASME-B	CASME II
OF	0.8092	0.8888	0.9243
HOG	0.8268	0.8378	0.8939
$\text{LBP}_{(8,1)}$	0.7716	0.8244	0.8751
$\text{LBP}_{(8,3)}$	0.8177	**0.8987**	0.9014
Proposed IP	**0.8480**	0.8617	**0.9289**

Table 1: AUC performance for all datasets

Method	Time per frame(ms)	γ
OF	480	631.58
HOG	121.11	159.35
LBP	35.13	46.22
Proposed IP	**0.76**	**1**

Table 2: Computation time comparison (image size 320×260)

Among these methods, the proposed method can perform best except in CASME-B dataset because the IP is sensitive to illumination variance while the LBP is robust to illumination. However, the better performance of our method in CASME-A and CASME II shows that the IP is an efficient feature which can describe the temporal dynamic of the micro-expression. The processing time for different methods is presented in the Table 2. Here a ratio for computational time comparison is defined as:

$$\gamma = \frac{T_{method}}{T_{IP}}, \tag{9}$$

where T_{method} represents the processing time of the OF, HOG and LBP. T_{IP} indicates the processing time of the IP features.

As we can observe in Table 2, the algorithm of the optical flow is extremely slow taking 480ms for one image of 320×260 feature extracting. While the proposed method takes only 0.76ms and thus is promising for implementation in real-time process. It is also clear from Table 2 that the integral projections provide a huge reduction in computational time. Compared to the LBP

and HOG, our method still globally outperforms them with an advantage in the lower computational complexity.

6 CONCLUSION

The analysis on differences of the integral projection allows detecting micro-movements automatically with a low computation complexity. Experimental results are positive on the datasets CASME-A, CASME-B and CASME II, indicating that this method is capable of catching micro-expressions from videos. To the best knowledge of the authors, this is the fastest method for automatic micro-expression detection and it could be implemented for future real-time detection. During the experiment, it was noticed that large head motions and the illumination variation can cause the miss-detections. In the future, more robust algorithms will be studied for addressing these problems. The motion-based method is accurate at the cost of computation complexity, while appearance-based methods are a bit less efficient in spotting micro-movements but with a fast computation. Thus, a combination of motion- and appearance-based method could bring a solution for improving performance of micro-expression detection.

7 REFERENCES

[Hag66] Haggard, E.A., Isaacs, K.S.: Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. Methods of research in psychotherapy. pp.154-165, 1966.

[Ekm69] Ekman, P., Friesen, W.V. Nonverbal leakage and clues to deception. Psychiatry. pp.88-106, 1969.

[Hua15] Huang, X., Wang, S.-J., Zhao, G., Piteikainen, M. Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection. Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 1-9, 2015.

[Liu15] 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. pp. 299-310, 2015.

[Pol09] POLIKOVSKY, S., KAMEDA, Y., OHTA, Y. Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. Crime Detection and Prevention (ICDP 2009), pp. 1-6, 2009.

[Dav15] Davison, A.K., Yap, M.H., Lansley, C. Micro-Facial Movement Detection Using Individualised Baselines and Histogram-Based Descriptors. Systems, Man, and Cybernetics (SMC), pp. 1864-1869, 2015.

[Moi14] Moilanen, A., Zhao, G., Pietikainen, M. Spotting rapid facial movements from videos using appearance-based feature difference analysis. 2014 22nd International Conference on Pattern Recognition (ICPR), pp. 1722-1727, 2014.

[Shr14] Shreve, M., Brizzi, J., Fefilatyev, S., Luguev, T., Goldgof, D., Sarkar, S. Automatic expression spotting in videos. Image and Vision Computing. pp.476-486, 2014.

[Bet12] Bettadapura, V. Face expression recognition and analysis: the state of the art. arXiv preprint arXiv:1203.6722. 2012.

[Wei09] Weiss, J.: Ekman, P. Telling Lies : Clues to Deceit in the Marketplace, Politics, and Marriage. New York: Norton. American Journal of Clinical Hypnosis. pp.287-288, 2011.

[Mat00] Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., Yrizarry, N., Loewinger, S., Uchida, H., Yee, A. and Amo, L. A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). Journal of Nonverbal Behavior. pp. 179-209, 2000.

[Mat11] Matsumoto, D., Hwang, H.S. Evidence for training the ability to read microexpressions of emotion. Motivation and Emotion. pp. 181-191, 2011.

[Bru93] Brunelli, R., Poggio, T. Face recognition: Features versus templates. IEEE transactions on pattern analysis and machine intelligence. pp. 1042-1052, 1993.

[Xio13] Xiong, X., Torre, F. Supervised descent method and its applications to face alignment. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR 2013), pp. 532-539, 2013.

[Shr09] Shreve, M., Godavarthy, S., Goldgof, D., Sarkar, S.: Macro-and micro-expression spotting in long videos using spatio-temporal strain. 2011 IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG 2011), pp. 51-56, 2011.

[Yan13] 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. Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops, pp.1-7, 2013.

[Yan14] Yan, W.J., Li, X., Wang, S.J., Zhao, G., Liu, Y.J., Chen, Y.H. and Fu, X.. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. PloS one, 9(1), 2014.

[Sun10] Sun, D., Roth, S., Black, M.J. Secrets of optical flow estimation and their principles. Presented

at the Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference, pp. 2432-2439, 2010.

[Oja96] oja96 Ojala, T., Pietikäinen, M., Harwood, D. A comparative study of texture measures with classification based on featured distributions. Pattern recognition. pp. 51-59, 1996.

[Dal05] Dalal, N., Triggs, B. Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), pp. 886-893, 2005.

[Li15] Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., Pietikäinen, M. Reading hidden emotions: spontaneous micro-expression spotting and recognition. arXiv preprint arXiv:1511.00423. 2015.

[Bla96] Black, M.J., Anandan, P. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. Computer vision and image understanding. pp. 75-104, 1996.

[Del15] De la Torre, F., Chu, W. S., Xiong, X., Vicente, F., Ding, X., & Cohn, J. Intraface. In Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, pp. 1-8, 2015

[Xia16] Xia, Z., Feng, X., Peng, J., Peng, X., & Zhao, G. Spontaneous micro-expression spotting via geometric deformation modeling. Computer Vision and Image Understanding. pp. 87-94, 2016.

[Pfi11] Pfister, T., Li, X., Zhao, G., & Pietikäinen, M.. Recognising spontaneous facial micro-expressions. Computer Vision (ICCV), 2011 IEEE International Conference. pp. 1449-1456, 2011.