Signal Processing: Image Communication 78 (2019) 246–253

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

A weighted feature extraction method based on temporal accumulation of optical flow for micro-expression recognition ☆

Lei Wang [a,b], Hai Xiao [a], Sheng Luo [a], Jie Zhang [a], Xiyao Liu [a,*,1]

[a] Central South University, School of Computer Science and Engineering, Changsha, China
[b] Mobile Health - Ministry of Education-China Mobile Joint Laboratory, Changsha, China

ARTICLE INFO

Keywords:
Micro-expression recognition
Temporal accumulation
Weighted LBP-TOP feature
Optical flow

ABSTRACT

Spatiotemporal features are widely used in micro-expression (ME) recognition to represent facial appearance and action. The features extracted from different face regions are usually given different weights according to the motion intensities in the corresponding regions. The weighted features are reported to be more discriminative than the unweighted ones for ME recognition. However, MEs are so subtle that their motion intensities are usually as low as noises, therefore small image noises can cause similar weights with MEs and degenerate the effectiveness of these weights. To address this issue, a novel weighted feature extraction method is proposed in this paper, whereby the neighboring optical flows in a time interval are accumulated to compute motion intensities. In this manner, the displacements caused by image noises in optical flow are decreased because these displacements are random and direction-inconsistent. Meanwhile, the displacements caused by facial expressions are enhanced because the displacements caused by facial expressions are usually direction-consistent among neighboring frames. The weights computed from the accumulated optical flows are multiplied with the spatiotemporal features, then the weighted features are fed to SVM to classify MEs. The experimental results demonstrate that our method achieves comparable recognition performances with the state-of-the-art methods on SMIC-HS and outperforms the state-of-the-art methods on CASME II.

1. Introduction

Micro-expressions are involuntary and instant facial dynamics which occur when people attempt to conceal emotions [1]. Micro-expression recognition has diverse applications for various fields, such as criminal detection [2], and psychological counseling [3]. However, micro-expression recognition is still a significant challenge due to the following reasons. First, micro-expressions are rapid facial muscle movements, which only last between 170 ms to 500 ms [4]. Second, intensities of involved muscle movements are subtle [5]. Therefore, even for ordinary people, recognizing micro-expressions is a hard task [6].

To recognize micro-expressions, many methods use spatiotemporal features to represent the subtle movements of micro-expressions. Most of the spatiotemporal features are extracted by block-based feature extraction technique which divides facial images into $N \times N$ non-overlapping blocks and extracts features from each block. Then, the features of each block are generated to contain local information and finally concatenated into a global feature histogram. Yan et al. [7] used 5×5 and 8×8 blocks Local Binary Pattern on Three Orthogonal Planes (LBP-TOP) features for classification of micro-expression videos on CASME II database and chose the classification accuracy as the baseline of this database. Li et al. [8] proposed a micro-expression analysis system (MESR) to detect and recognize micro-expression. Wang et al. [9] extracted LBP-TOP from a tensor independent color space to recognize micro-expressions. They implemented motion magnification technique on block-based Histogram of Image Gradient Orientation (HIGO) feature. Huang et al. [10] designed a spatiotemporal local binary pattern with integral projection (LBP-IP) to preserve the shape properties of micro-expressions. Wang et al. [11] proposed Sparse Tensor Canonical Correlation Analysis (STCCA) for micro-expression recognition, which fused corresponding LBP and micro-expression data by correlation characteristics. Huang et al. [12] improved LBP-TOP and proposed a spatiotemporal Completed Local Quantization Patterns (STCLQP) to recognize micro-expressions. Liong et al. [13] extracts

☆ One or more of the authors of this paper have disclosed potential or pertinent conflicts of interest, which may include receipt of payment, either direct or indirect, institutional support, or association with an entity in the biomedical field which may be perceived to have potential conflict of interest with this work. For full disclosure statements refer to https://doi.org/10.1016/j.image.2019.07.011.

* Corresponding author.
E-mail address: lxyzoewx@csu.edu.cn (X. Liu).

[1] The author as a member of EURASIP.

https://doi.org/10.1016/j.image.2019.07.011
Received 28 December 2018; Received in revised form 13 July 2019; Accepted 15 July 2019
Available online 17 July 2019

spatiotemporal features only from eyes and mouth regions because these regions contain most facial movements of micro-expressions.

Except for the spatiotemporal features, optical flow as a classic feature to extract motion information is also used in some micro-expression detection and recognition methods. Liu et al. [14] proposed the Main Directional Mean Optical flow (MDMO) feature to encode motion information of micro-expressions and used MDMO to recognize micro-expressions. Wang et al. [15] proposed a Main Directional Maximal Difference (MDMD) analysis to detect micro-expression in long-term videos, which use optical flow to estimate facial motion of facial regions. Happy et al. [16] proposed a Fuzzy Histogram of Optical Flow Orientations (FHOFO) feature to represent motion information from 36 facial regions of interest for micro-expression recognition. Lu et al. [17] proposed a feature based on the Fusion of Motion Boundary Histograms (FMBH), which combines the horizontal and the vertical components of differential of optical flow. FMBH has great discriminative ability for subtle movements of micro-expressions.

All the methods mentioned above treat the local features from different facial regions equally. However, micro-expressions usually occur in some facial regions rather than the whole face area [5]. Therefore, the local features extracted from the regions with movements should have more contributions for micro-expressions classification.

To address this issue, weighted block-based feature methods were proposed [18–20]. In these methods, local features of each block are given different weights to emphasize their different contributions for ME recognition. The weights are calculated based on the estimated motion intensities of each block. Liong et al. [18] used optical strain, which is based on optical flow, to measure motion intensities of each block and generated weights for block-based LBP-TOP feature. He et al. [19] also used this optical strain based weighting method to enhance multi-task mid-level learning feature. Liong et al. [20] proposed a Bi-Weighed Oriented Optical Flow (Bi-WOOF) feature which extracts Histograms Of Optical Flow (HOOF) feature from each block and weights each block with optical strain weight.

Unfortunately, the weights based on motion intensities are vulnerable to image noises, especially in the micro-expression circumstances. Because micro-expressions are subtle, the movement intensities measured by optical flow are usually low. Therefore, even the unnoticeable image noises can cause some optical flow responses which degenerate the effectiveness of the weights.

We notice that the displacements in optical flows caused by facial movements are direction consistent among neighboring frames, whereas the displacements caused by image noises are random and direction inconsistent. Therefore, temporal accumulation of optical flows can enhance the directional consistent displacements caused by micro-expression movements and decrease the random displacements caused by image noises.

Based on the above observation, we propose a novel weighted feature extraction method based on temporal accumulation of optical flow to enhance the discrimination of the weighted features for ME recognition. First, the optical flows are calculated from micro-expression video clips and accumulated temporally to reduce the displacements caused by image noises. Second, the motion intensity in each block is calculated based on the magnitude of the accumulated optical flow. Third, the weight for each block is computed by normalizing the motion intensities. Finally, the local features of each block are multiplied with these weights to generate the overall features as the input of SVM for micro-expression recognition.

The rest of the paper is organized as follows: Section 2 describes the details of our method. Section 3 reports and discusses the experimental results. Section 4 gives conclusion and future work.

2. Proposed methods

Our proposed method consists of three main phases: weight generation phase, feature extraction phase, and micro-expression classification phase. The framework of our method is shown in Fig. 1.

2.1. Weight generation phase

2.1.1. Temporal accumulation of optical flows

As described in [8], all the faces in the input video clips are first aligned and cropped based on an overall neutral face. In this manner, the variations of spatial appearance among different video clips are normalized.

Then, the dense optical flows [21] are extracted from the aligned video clips to represent motion information of each pixel in micro-expression video clips. The optical flows $\boldsymbol{F_t(x, y)}$ are computed from each pair of neighboring frames. For an l-frames input video, $l-1$ optical flows can be obtained. The optical flows $\boldsymbol{F_t(x, y)}$ consist of horizontal displacements and vertical displacements. Therefore, the optical flows $\boldsymbol{F_t(x, y)}$ can be expressed as:

$$\boldsymbol{F_t(x, y)} = (U_h^t(x, y), U_v^t(x, y)) \tag{1}$$

where $U_h^t(x, y)$ and $U_v^t(x, y)$ indicate the horizontal and vertical displacements respectively.

The optical flows $\boldsymbol{F_t(x, y)}$ are further partitioned into S groups, and the optical flows in each group are temporally accumulated to form cumulative optical flows. S is the number of groups for accumulating optical flows. In each group, the group size is D which denotes the number of frames to be temporally accumulated. The relationship between the number of groups S and the group size D can be expressed as:

$$D = \lceil (l-1)/S \rceil \tag{2}$$

where $\lceil . \rceil$ is the ceiling function. Then, temporal cumulative optical flows are calculated as:

$$C_h^k(x, y) = \sum_{t=(k-1)\times D+1}^{k\times D} U_h^t(x, y), \quad C_v^k(x, y) = \sum_{t=(k-1)\times D+1}^{k\times D} U_v^t(x, y) \tag{3}$$

where $C_h^k(x, y)$ and $C_v^k(x, y)$ represent the horizontal and vertical components of kth group of cumulative optical flows. The range of k is $[1, S]$. By temporal accumulation, the displacements caused by image noises decreases since they are random and direction inconsistent. The displacements caused by micro-expression will be enhanced since they are direction consistent among consecutive frames.

The group size D has a considerable impact on the effectiveness of our proposed method. The smaller it is, the fewer optical flows are accumulated in a group. Then possibly there are not enough accumulations to remove the displacements caused by image noises. On the other hand, if D is large, many optical flows will be accumulated in a group. Therefore, the displacements of irrelevant expressions and the slightly head motions are mixed and enhanced, which will produce a weight matrix with high values in nearly all facial regions. The influence and choice of the group size is discussed in Section 3.3.

To further illustrate the effectiveness of temporal accumulation of optical flows, Fig. 2 shows an example of the optical flows before and after the accumulation. In Fig. 2, optical flows are visualized by using the color-coding method in [22]. The direction and magnitude of the displacement in optical flows are represented by different colors. The color-coding map is shown at the bottom-left of Fig. 2. As shown in Fig. 2, the optical flows are computed from the micro-expression video clip. There are many small colorful spots in optical flows, which represents the random displacements caused by image noises. After temporally accumulating, the cumulative optical flow image gets smoother and the green region at the top-right corner becomes more distinct, which indicates the random displacements caused by image noises have been effectively reduced and the displacements caused by facial movements have been strengthened.

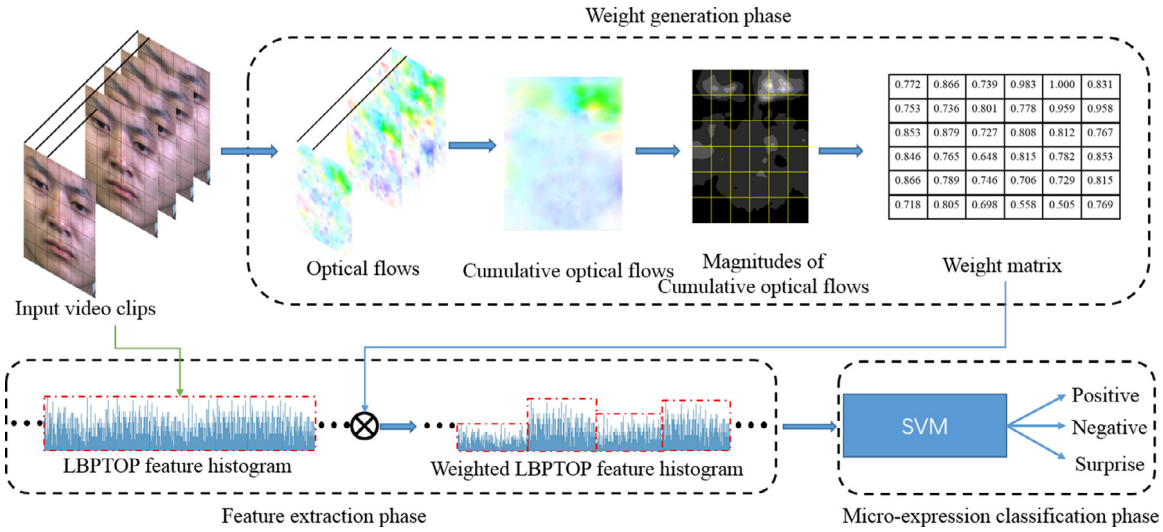

Fig. 1. Framework of our proposed micro-expression recognition system.

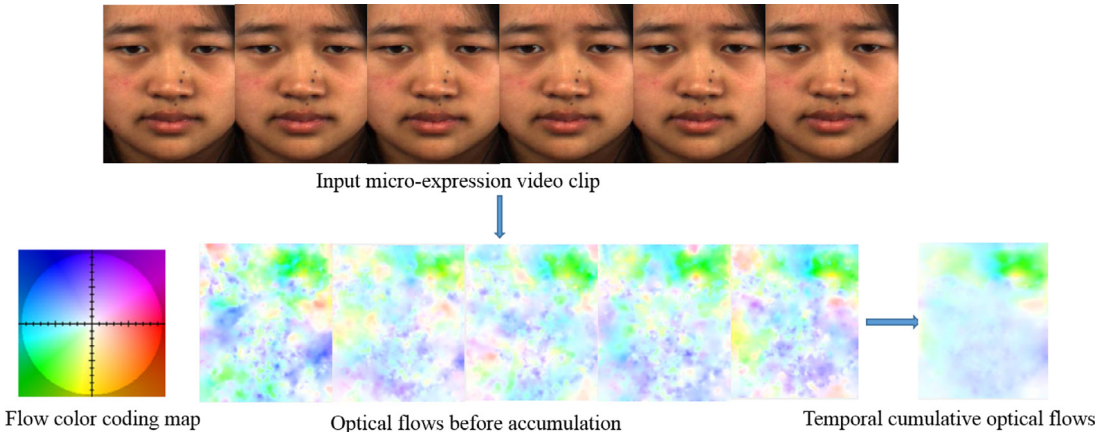

Fig. 2. An example to illustrate the noise removal by temporal accumulation of optical flows.

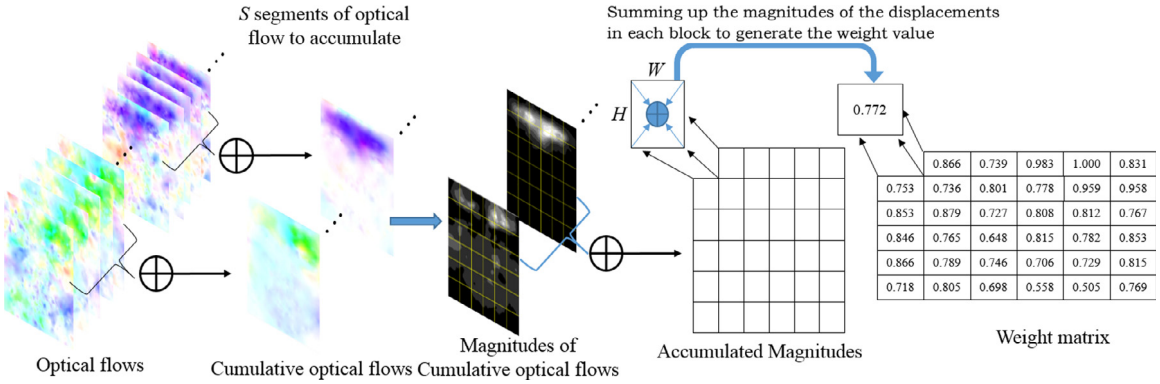

Fig. 3. Process of weight matrix computing.

2.1.2. Weight matrices computing

The weight matrix for block-based features is computed based on the magnitude of cumulative optical flows. Similar to the block-based features, the cumulative optical flows are divided into $N \times N$ non-overlapping blocks. The motion intensity $M_{i,j}$ in the block at the ith row and jth column is obtained by summing up the optical flows in this block and its neighboring blocks through time, as shown in (4). Fig. 3 illustrates the process of weight matrices computing.

$$M_{i,j} = \sum_{k=1}^{S} \sum_{x=(i-1)\times W+1}^{i\times W} \sum_{y=(j-1)\times H+1}^{j\times H} \sqrt{(C_h^k(x,y))^2 + (C_v^k(x,y))^2} \tag{4}$$

where i, j are the block indexes and $i, j = 1, 2, \ldots, N$. H and W are the height and width of each block. k is the index of cumulative optical flows. $k = 1, 2, \ldots, S$ and S is the number of groups for accumulating optical flows.

Finally, the weight matrix $E_{i,j}$ is obtained by normalizing the motion intensity $M_{i,j}$ into the range $[0, 1]$, expressed as:

$$E_{i,j} = M_{i,j} / MAX(\{M_{i,j} | i, j = 1, \ldots, N\}) \quad (5)$$

where the function $MAX()$ finds the maximal value in a set.

2.2. Feature extraction phase

To reduce the video length discrepancies of each micro-expression sample, all the video clips are interpolated into 10 frames with the temporal interpolation model [23]. Then the frames in the interpolated videos are normalized and cropped to address the variations of spatial appearance as described in [8].

The preprocessed video clips are divided into $N \times N \times T$ non-overlapping block volumes. The features are extracted from each block volume and concatenated into a vector for future classification. We use the LBP-TOP feature descriptors [24] to encode the spatiotemporal information of micro-expressions. LBP-TOP, a widely used feature in micro-expression recognition, consists of three LBP histograms extracted from the spatial plane XY and the spatial–temporal planes XT and YT respectively. Therefore, the feature histograms for a block volume can be expressed as:

$$H_{i,j,t} = [H^{XY}_{i,j,t}, H^{XT}_{i,j,t}, H^{YT}_{i,j,t}] \quad i, j = 1, \ldots, N; t = 1, \ldots, T \quad (6)$$

The extracted LBP-TOP histograms are multiplied with the corresponding weights to form the weighted features for micro-expression recognition, as shown in (7). And Fig. 4 shows the weighting process.

$$\tilde{H}_{i,j,t} = W_{i,j} \cdot H_{i,j,t} \quad (7)$$

Compared with the existing weighted feature-based method [18] which only weights the features in XY plane, our method weights all the three planes' features, as illustrated in Fig. 4. We will compare these two weighting strategies in Section 3.4.

2.3. Micro-expression classification phase

The weighted features are fed into a linear SVM-based [25] classifier to recognize micro-expressions. The one-against-one strategy is adopted to implement the multi-class recognition problem. For K types of micro-expressions, $K \times (K-1)/2$ binary SVMs are trained to distinguish the samples of one type from the samples of another type. A classification of an unknown expression is done according to the maximum voting, where each SVM votes for one type.

3. Experiments and analysis

In this section, performances of our proposed method are evaluated and analyzed. Section 3.1 introduces the experimental setting; Section 3.2 discusses the parameter selection for LBP-TOP feature; Section 3.3 discusses the influence of group size for optical flow accumulation; Section 3.4 compares two weighting strategies, Section 3.5 compares our proposed method with state-of-the-art micro-expression recognition methods and Section 3.6 evaluates the method with confusion matrices.

3.1. Experimental setting

To evaluate the performance of our proposed method, the experiments were carried out on two recent spontaneous micro-expression databases, namely SMIC-HS [26] and CASME II [7]. The SMIC-HS dataset includes 164 micro-expression video clips recorded from 16 subjects. These video clips are recorded with a resolution of 640 × 480 pixels and a frame rate of 100 fps. The resolution of cropped face images is around 140 × 180. For SMIC-HS, ME samples are classified into three categories of positive, negative and surprise. The CASME II dataset consists of 255 micro-expression video clips from 26 subjects. These video clips are recorded by a 200 fps high-speed camera with a resolution of 640 × 480. But the resolution of cropped face images is higher than SMIC-HS, which is around 250 × 300. Video clips in CASME II include seven classes of micro-expression: surprise, fear, happiness, repression, disgust, sadness and others. However, there are only 2 samples of fear and 7 samples of sadness, which are too few for training and evaluation. Therefore, these fear and sadness samples are not included in our experiments and the rest of 246 samples of 5 classes of micro-expressions are used in this paper.

Like many existing researches [7,8], leave-one-subject-out (LOSO) cross-validation is used to evaluate our proposed method. In this paper, all experiments are running on a PC of Windows 8.1 system with Intel(R) Core(TM) i5-4200H CPU @2.80 Hz and 8G RAM.

Table 1
The recognition accuracies for LBP-TOP features with varying R_{xy} and R_t in different divided block numbers N on SMIC-HS.

	$N = 4$		$N = 6$		$N = 8$	
	$R_t = 1$	$R_t = 2$	$R_t = 1$	$R_t = 2$	$R_t = 1$	$R_t = 2$
(a) SMIC-HS						
$R_{xy} = 1$	42.86	48.88	42.23	50.14	43.16	50.23
$R_{xy} = 2$	45.38	53.74	46.59	52.14	38.35	47.16
$R_{xy} = 3$	41.23	52.43	47.86	51.95	36.53	45.48
$R_{xy} = 4$	40.96	**53.75**	49.68	52.12	40.43	49.88
(b) CASME II						
$R_{xy} = 1$	50.18	51.48	50.23	52.34	49.68	**57.69**
$R_{xy} = 2$	53.59	56.04	51.64	54.26	51.48	54.92
$R_{xy} = 3$	52.88	54.27	51.16	53.38	50.59	56.28
$R_{xy} = 4$	53.09	53.88	51.74	54.36	53.12	50.23

3.2. Parameter selection for LBP-TOP

The aim of this first experiment is to explore how the parameters of LBP-TOP features affect the accuracy of ME recognition.

Parameters: We vary the radius of circular pattern and the number of divided blocks. Let the radius for the spatial plane (XY plane) be R_{xy} and the radius for the spatiotemporal planes (XT and YT planes) be R_t. The video clips are divided to $N \times N \times 2$ blocks. R_{xy} is varied from 1 to 4 and R_t is varied from 1 to 2. Three different values for N (4, 6, 8) are tested. The number of neighboring points p is set to $p = 8$ for all datasets [8].

Results: The ME recognition results are shown in Table 1. The best accuracy for the SMIC-HS dataset is achieved at $N = 4$, $R_{xy} = 4$ and $R_t = 2$. Whereas the best accuracy for the CASME II dataset is achieved at $N = 8$, $R_{xy} = 1$ and $R_t = 2$. The optimal number of divided blocks for SIMIC is smaller than the one for CASME II. The reason is that the size of face images in SMIC-HS is smaller (130 × 160 pixels on SMIC-HS and 250 × 300 pixels on CASME II), to keep the discrimination of LBP-TOP features, the size of divided blocks should not be too small. Smaller divided block number on SMIC-HS ensures that the size of each divided block is big enough. These optimal values of LBP-TOP parameters are used in the following experiments.

3.3. Effect of the group size for accumulating optical flows

In this experiment, we explore how the group size D for accumulating optical flows affects the accuracy of ME recognition.

Parameters: The group size is varied from 1 to 7 for CASME II. A smaller range (1 to 5) is used for SMIC-HS, because the frame rate of SMIC-HS is only half of CASME II. The largest group size is also tested, which means to accumulate all the optical flows of a clip into one single group.

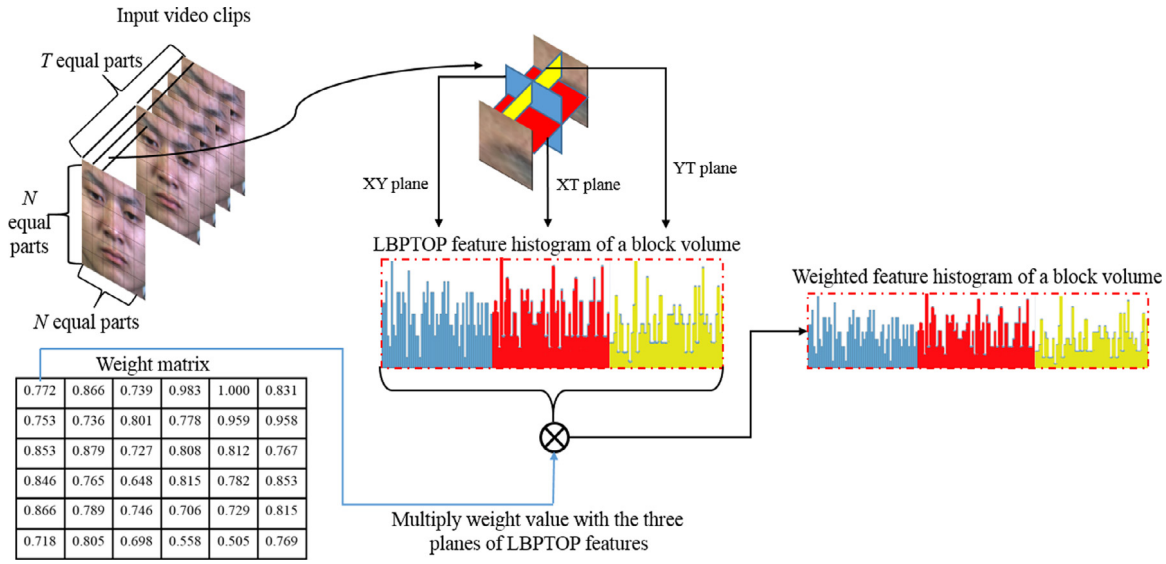

Fig. 4. The process of weighting the LBP-TOP features.

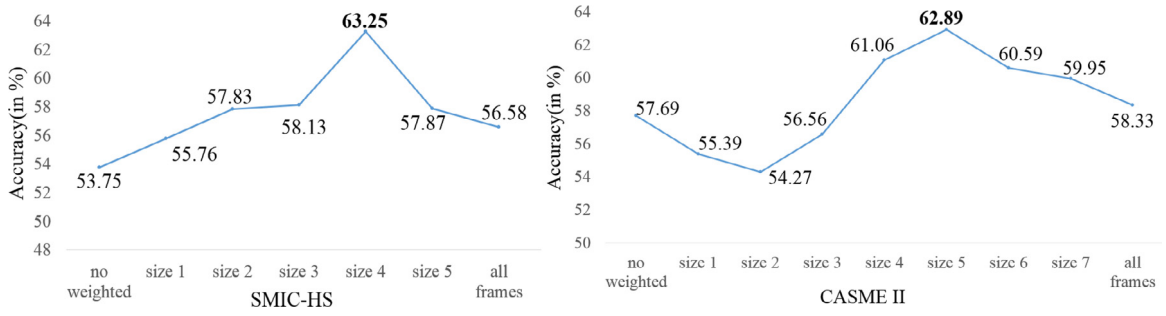

Fig. 5. Recognition accuracies of different accumulation group sizes D.

Results: The ME recognition results are shown in Fig. 5. The best group size is 4 for SMIC-HS and 5 for CASME II. We analyzed this observation from three aspects:

First, accumulation of optical flows does benefit ME recognition. When the group size is set to 1, which amounts to no accumulation is applied, the recognition accuracies are low for both datasets. When we increase the group size from 1 to 4, which means more and more neighboring optical flows are accumulated, the recognition accuracies increase for both datasets. By temporal accumulation, the displacements caused by image noises decrease since they are random and direction inconsistent. The displacements caused by micro-expressions will be enhanced since they are direction consistent among consecutive frames.

Second, the group size should be kept in a proper range. For SMIC-HS, the recognition accuracy decreases when the group size is larger than 4. For CASME II, the accuracy decreases when the group size is larger than 5. When the group size gets larger, more and more optical flows will be accumulated in a group. Then the displacements of irrelevant expressions and the slightly head motions are mixed and enhanced, which will produce weight matrices with large values in nearly all facial regions. Therefore, the weight matrices cannot emphasize the features in ME regions and the weighting method becomes less effective.

Third, the optimal group size should be chosen according to the frame rates of video clips. The higher the frame rate is, the larger the group size should be. This is because higher frame rate will produce more frames for micro-expression video clips with the same length. The facial movements with the same direction will last for more frames in video clips. For SMIC-HS, the optimal accumulation group size is 4; Whereas for CASME II whose videos have higher frame rates, the optimal group size increases to 5. In addition, when implementing our method in real-world application, the group size should be chosen based on the frame rate.

3.4. Comparison of two weighting strategies

In this experiment, we evaluate two weighting strategies. The first one is to weight the LBP-TOP features in just one plane (XY plane) [18]. The second one is to weight the features in all the three planes (XY, YT and XT planes). The ME recognition results are shown in Fig. 6. Weighting in three planes achieves better performance than just weighting in XY plane. Because LBP-TOP is spatiotemporal features, weighting in XY plane just emphasizes the spatial features in the regions with micro-expressions. However, weighting all the three planes emphasizes both spatial and temporal features in the regions with micro-expressions, which makes LBP-TOP more discriminative for ME recognition.

3.5. Comparison with the state-of-the-art methods

The best performance of our method is with the optimal group sizes for accumulating optical flows (4 for SMIC-HS, 5 for CASME II) and weighting LBP-TOP features in all the three planes (XY, YT and XT planes). We achieved 63.25% accuracy for ME recognition on the SIMC-HS dataset, and 62.89% on CASME II. The results were obtained without extra preprocessing like motion magnification [27] or more advanced features like HIGO [8]. We list state-of-the-art results for these datasets in Table 2 for comparison. We performed all experiments using the leave-one-subject-out validation protocol.

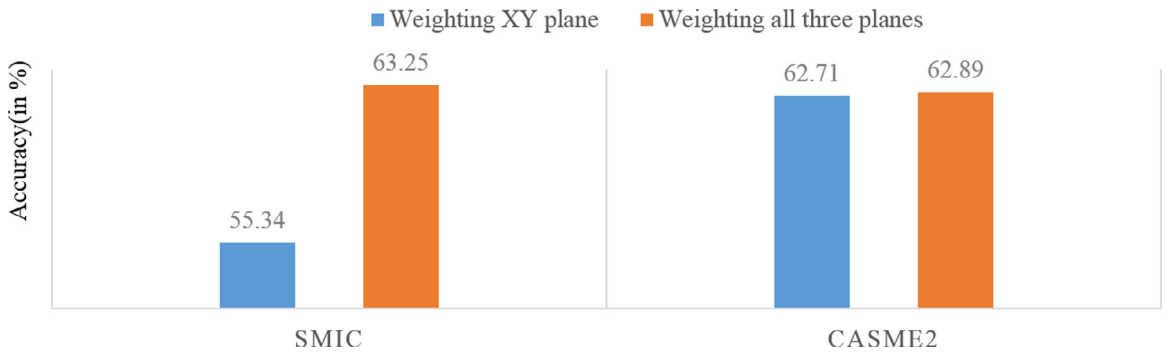

Fig. 6. The recognition accuracies of weighting on different planes of LBP-TOP features.

Table 2
The comparison results with the state-of-the-art methods in recognition accuracy (%).

	SMIC-HS	CASME II
OS [18]	53.56	–
MMFL [19]	63.15	59.81
Bi-WOOF [20]	58.85	62.20
HIGO [8]	65.24	58.39
LBP-IP [10]	57.90	59.51
STCLQP [12]	64.02	58.39
FHOFO [16]	51.83	56.64
MDMO[a] [14]	–	67.37
FMBH[b] [17]	71.95	69.11
Proposed method	63.25	62.89
Proposed method[a]	–	68.75
Proposed method[b]	71.70	69.56

[a]Means the method classifies samples into four classes.
[b]Means there is a manual mask step described in [17].

Table 3
The confusion matrix of micro-expression recognition on SMIC-HS dataset.

	Ground truth		
	Negative	Positive	Surprise
LBP-IP [10]			
Negative	0.57	0.31	0.35
Positive	0.29	0.63	0.12
Surprise	0.14	0.06	0.53
STCLQP [12]			
Negative	0.66	0.29	0.23
Positive	0.10	0.61	0.12
Surprise	0.24	0.10	0.65
Proposed method			
Negative	0.63	0.20	0.23
Positive	0.17	0.65	0.12
Surprise	0.20	0.15	0.65

Previous work includes weighted feature-based methods and non-weighted feature-based methods. The former contains include Optical Strain (OS) weighted method [18], Multi-task Mid-level Feature Learning method (MMFL) [19] and Bi-Weighed Oriented Optical Flow (Bi-WOOF) feature [20]. The latter contains Fuzzy Histogram of Image Gradient Orientation feature (HIGO) [8], spatiotemporal Local Binary Pattern with Integral Projection (LBP-IP) [10], Spatio-Temporal Completed Local Quantization Pattern (STCLQP) [12], Histogram of Optical Flow Orientations (FHOFO) [16], Main Directional Mean Optical flow feature (MDMO) [14] and Fusion of Motion Boundary Histograms (FMBH) [17].

Most of the compared methods classify MEs into five emotions on CASME II, however MDMO [14] classifies MEs into four emotions (*positive*, *negative*, *surprise* and *others*). To compare with MDMO fairly, we also test our method in the same way as MDMO. Because there is no result of MDMO on SMIC-HS in [14], therefore we only listed the its results on CASME II in Table 2.

Table 4
The confusion matrix of recognition accuracies for each micro-expression on CASME II dataset.

	Ground truth				
	Happy	Disgust	Surprise	Repression	Others
LBP-IP [10]					
Happy	0.34	0.02	0	0.19	0.02
Disgust	0.06	0.50	0	0.04	0.10
Surprise	0.03	0	0.64	0	0
Repression	0.03	0	0.04	0.22	0.05
Others	0.53	0.48	0.32	0.56	0.83
STCLQP [12]					
Happy	0.56	0.03	0.08	0.22	0.05
Disgust	0.06	0.38	0	0.11	0.19
Surprise	0.03	0	0.72	0	0
Repression	0.06	0.02	0.04	0.37	0.01
Others	0.28	0.58	0.12	0.30	0.75
Propose method					
Happy	0.66	0.05	0.24	0.26	0.08
Disgust	0	0.51	0.08	0.07	0.21
Surprise	0.03	0.16	0.60	0.07	0
Repression	0.07	0	0.04	0.37	0.01
Others	0.24	0.28	0.04	0.22	0.70

In FMBH [17], a manually created mask is used to remove the background in images. Although the manual mask does benefit MEs recognition, it introduces extra manipulations for different videos and increases the cost for implementing this method in real-world application. Therefore, it is unfair to compare the other automatic methods with FMBH. To compare with FMBH fairly, we also implement this manual mask and test the recognition accuracy by using our method with the manual mask.

As shown in Table 2, our proposed method achieves comparable recognition performances with the state-of-the-art methods on SMIC-HS and outperforms the state-of-the-art methods on CASME II. Compared with the weighted feature feature-based methods [18–20], the recognition accuracies of our method on both SMIC-HS and CASME II are higher than those of OS [18], MMFL [19] and Bi-WOOF [20]. Compared with non-weighted feature-based methods, the recognition accuracy of the proposed method is 11.42% higher than FHOFO [16] on SMIC-HS and 6.25% higher on CASME II. Compared with LBP-IP [10], the recognition accuracy of the proposed method is 5.35% higher on SMIC-HS and 3.38% higher on CASME II. Compared with STCLQP [12] and HIGO [8], which use more complicated and discriminative feature than FHOFO and LBP-IP, our method achieves comparable performances on SMIC-HS, and better performances on CASME II. Compared with MDMO [14] which achieves 67.37% in four emotions classification on CASME II, the proposed method achieves 68.75% in the same settings. FMBH [17] achieves superior accuracies of 71.95% on SMIC-HS and 69.11% on CASME II, which results from its discriminative feature and manually background removal. By augmented with the

Table 5
The confusion matrix of four emotions classification on CASME II.

	Ground truth			
	Positive	Negative	Surprise	Others
MDMO [14]				
Positive	0.45	0.03	0.05	0.08
Negative	0.06	0.54	0.05	0.11
Surprise	0.03	0.03	0.67	0.01
Others	0.45	0.40	0.24	0.80
Propose method				
Positive	0.53	0.03	0.04	0.09
Negative	0.06	0.56	0.04	0.10
Surprise	0.09	0.08	0.68	0.04
Others	0.31	0.33	0.24	0.78

Table 6
The confusion matrix of recognition with manual mask on SMIC-HS dataset.

	Ground truth		
	Negative	Positive	Surprise
FMBH [17]			
Negative	0.64	0.12	0.19
Positive	0.21	0.80	0.07
Surprise	0.15	0.08	0.74
Proposed method			
Negative	0.73	0.10	0.10
Positive	0.16	0.76	0.16
Surprise	0.11	0.14	0.74

same manual background removal method as FMBH, our proposed method achieves 71.70% on SMIC-HS and 69.56% on CASME II which are comparable with FMBH.

The reason of the above results is that the temporally accumulated optical flows in our method can reduce the displacements caused by image noises and enhance the displacement caused by facial movements to achieve better performances of micro-expression recognition.

3.6. Evaluate the performance for recognizing different micro-expressions

To further evaluate the performance for recognizing different emotions of our proposed method, we have tested the accuracies of recognition for different micro-expressions on SMIC-HS and CASME II and generated the confusion matrices at the best recognition accuracies by the LOSO cross-valid. Tables 3 and 4 show the confusion matrices of LBP-IP [10], STCLQP [12] and our proposed method on SMIC-HS and CASME II respectively. Table 5 lists the confusion matrix of MDMO [14] and our proposed method with four emotions classification on CASME II. Tables 6 and 7 show the confusion matrices of FMBH [17] and our proposed method with manual mask on SMIC-HS and CASME II respectively.

As shown in Table 3, compared with LBP-IP [10], our method achieves better performances than LBP-IP on all the three emotions on SMIC-HS. In addition, the average recognition accuracy for three emotions of our proposed method is 0.643 and is higher than that of LBP-IP, which is 0.576. These results demonstrate our method outperforms LBP-IP for recognizing different micro-expressions on SMIC-HS. Compared with STCLQP [12], our proposed method achieves comparable performances on three emotions on SMIC-HS. The average recognition accuracy of our method is comparable with that of STCLQP, which is 0.640. These results demonstrate that the performance of our proposed method for recognizing different micro-expressions on SMIC-HS is comparable with that of STCLQP.

As shown in Table 4, compared with LBP-IP [10], our proposed method achieves much better performances on *happy*, *repression*, and comparable performances on *disgust* and *surprise* on CASME II. Although the accuracy on *others* of our method is lower than that of LBP-IP, the average recognition accuracy for all emotions of our method is 0.568 and higher than that of LBP-IP, which is 0.506. Compared with STCLQP [12], our proposed method performs comparable on *repression* , and much better on *happy*, *disgust* on CASME II. Although the performance on *surprise* and *others* of our method is worse than that of STCLQP, the average recognition accuracy of our method is 0.568 and comparable with than that of STCLQP, which is 0.556. These results demonstrate the performance of our method for recognizing different micro-expressions on CASME II is better than that of LBP-IP and comparable with that of STCLQP.

Table 7
The confusion matrix of recognition with manual mask on CASME II dataset.

	Ground truth				
	Happy	Disgust	Surprise	Repression	Others
FMBH [17]					
Happy	0.63	0.16	0.08	0.07	0.07
Disgust	0.16	0.73	0.04	0.04	0.17
Surprise	0.03	0	0.80	0.04	0.01
Repression	0.06	0	0.04	0.52	0.04
Others	0.13	0.22	0.04	0.33	0.71
Propose method					
Happy	0.72	0.02	0.04	0.11	0.08
Disgust	0.06	0.57	0.08	0.07	0.06
Surprise	0.06	0	0.72	0.04	0.04
Repression	0.03	0.10	0.12	0.59	0.10
Others	0.13	0.32	0.04	0.19	0.72

As shown in Table 5, our proposed method achieves comparable performances on *negative*, *surprise* and *others*, and better performances on *positive* on CASME II. In addition, the average accuracy of all emotions of our method is 0.637. This value is higher than that of MDMO which is 0.615. These results demonstrate our proposed method outperform MDMO for recognizing different micro-expressions.

As shown in Table 6, on SMIC-HS, our proposed method achieves comparable performances on *positive* and *surprise*, and better performances on *negative* on SMIC-HS compared with FMBH [17]. The average accuracy of all emotions of our method on SMIC-HS is 0.743. This value is comparable with than that of FMBH which is 0.726. These results demonstrate that our proposed method have comparable performances for recognizing different micro-expressions with FMBH on SMIC-HS.

As shown in Table 7, for CASME II, our accuracies of *happy* and *repression* are higher than FMBH. Although our accuracies of *disgust* and *surprise* are lower, the average accuracy of our method is 0.664 and comparable with that of FMBH, which is 0.678. These results demonstrate that the performance of our proposed method for recognizing different micro-expressions on CASME II is comparable with that of FMBH.

4. Conclusion

In this paper, a novel weighted feature extraction method based on accumulation of optical flows is proposed. By accumulating optical flows, the displacements caused by image noises are reduced and the displacements caused by micro-expressions are enhanced. Therefore the motion intensities can be estimated more accurately. The experimental results on two popular ME datasets demonstrate the effectiveness of the proposed method. Our recognition performances are comparable with the state-of-the-art methods on SMIC-HS and outperforms the state-of-the-art methods on CASME II.

For future works, we plan to improve our methods by implementing an adaptive group size scheme which could determine the group size for different facial movements based on the action unit detection. In addition, how to design more effective optical flow based alignment method to correct small head movements and further reduce image noises will be our future research.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (61602527), Hunan Provincial Natural Science Foundation of China (2017JJ3416, 2018JJ2548), China Postdoctoral Science Foundation (2017M612585) and "Mobile Health" Ministry of Education-China Mobile Joint Laboratory.

References

[1] P. Ekman, Lie catching and microexpressions, Philos. Decept. (2009) 118–133.
[2] T.A. Russell, E. Chu, M.L. Phillips, A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool, Br. J. Clin. Psychol. 45 (4) (2006) 579–583.
[3] X. Ben, P. Zhang, R. Yan, M. Yang, G. Ge, Gait recognition and micro-expression recognition based on maximum margin projection with tensor representation, Neural Comput. Appl. 27 (8) (2016) 2629–2646.
[4] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, X. Fu, How fast are the leaked facial expressions: The duration of micro-expressions, J. Nonverbal Behav. 37 (4) (2013) 217–230.
[5] S. Porter, L. Ten Brinke, Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions, Psychol. Sci. 19 (5) (2008) 508–514.
[6] P. Ekman, Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage, revised ed., WW Norton & Company, 2009.
[7] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, X. Fu, CASME II: An improved spontaneous micro-expression database and the baseline evaluation, PLoS One 9 (1) (2014) e86041.
[8] X. Li, X. Hong, A. Moilanen, X. Huang, T. Pfister, G. Zhao, M. Pietikäinen, Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods, IEEE Trans. Affect. Comput. 9 (4) (2018) 563–577.
[9] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, C.-G. Zhou, X. Fu, M. Yang, J. Tao, Micro-expression recognition using color spaces, IEEE Trans. Image Process. 24 (12) (2015) 6034–6047.
[10] X. Huang, S.-J. Wang, G. Zhao, M. Piteikainen, Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, pp. 1–9.
[11] S.-J. Wang, W.-J. Yan, T. Sun, G. Zhao, X. Fu, Sparse tensor canonical correlation analysis for micro-expression recognition, Neurocomputing 214 (2016) 218–232.
[12] X. Huang, G. Zhao, X. Hong, W. Zheng, M. Pietikäinen, Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns, Neurocomputing 175 (2016) 564–578.
[13] S.-T. Liong, J. See, R.C.-W. Phan, K. Wong, S.-W. Tan, Hybrid facial regions extraction for micro-expression recognition system, J. Signal Process. Syst. 90 (4) (2018) 601–617.
[14] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, X. Fu, A main directional mean optical flow feature for spontaneous micro-expression recognition, IEEE Trans. Affect. Comput. 7 (4) (2016) 299–310.
[15] S.-J. Wang, S. Wu, X. Qian, J. Li, X. Fu, A main directional maximal difference analysis for spotting facial movements from long-term videos, Neurocomputing 230 (2017) 382–389.
[16] S. Happy, A. Routray, Fuzzy histogram of optical flow orientations for micro-expression recognition, IEEE Trans. Affect. Comput. (2017).
[17] H. Lu, K. Kpalma, J. Ronsin, Motion descriptors for micro-expression recognition, Signal Process., Image Commun. 67 (2018) 108–117.
[18] S.-T. Liong, J. See, R.C.-W. Phan, A.C. Le Ngo, Y.-H. Oh, K. Wong, Subtle expression recognition using optical strain weighted features, in: Asian Conference on Computer Vision, Springer, 2014, pp. 644–657.
[19] J. He, J.-F. Hu, X. Lu, W.-S. Zheng, Multi-task mid-level feature learning for micro-expression recognition, Pattern Recognit. 66 (2017) 44–52.
[20] S.-T. Liong, J. See, K. Wong, R.C.-W. Phan, Less is more: Micro-expression recognition from video using apex frame, Signal Process., Image Commun. 62 (2018) 82–92.
[21] M.J. Black, P. Anandan, The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields, Computer Vision and Image Understanding 63 (1) (1996) 75–104.
[22] S. Baker, D. Scharstein, J. Lewis, S. Roth, M.J. Black, R. Szeliski, A database and evaluation methodology for optical flow, Int. J. Comput. Vis. 92 (1) (2011) 1–31.
[23] Z. Zhou, G. Zhao, M. Pietikäinen, Towards a practical lipreading system, in: Computer Vision and Pattern Recognition, CVPR, 2011 IEEE Conference on, IEEE, 2011, pp. 137–144.
[24] G. Zhao, M. Pietikainen, Dynamic texture recognition using local binary patterns with an application to facial expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (6) (2007) 915–928.
[25] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (3) (2011) 27.
[26] X. Li, T. Pfister, X. Huang, G. Zhao, M. Pietikäinen, A spontaneous micro-expression database: Inducement, collection and baseline, in: Automatic Face and Gesture Recognition (Fg), 2013 10th IEEE International Conference and Workshops on, IEEE, 2013, pp. 1–6.
[27] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, W. Freeman, Eulerian video magnification for revealing subtle changes in the world, ACM Trans. Graph. 31 (4) (2012) 1–8.