2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)

Micro-Expression Recognition Based on Local Two-Order Gradient Pattern

Mingyue Niu
National Laboratory of Pattern Recognition, (NLPR)
Institute of Automation, CAS,
School of Artificial Intelligence
University of Chinese Academy of Sciences
Beijing, China
niumigyue2017@nlpr.ia.ac.cn

Ya Li
National Laboratory of Pattern Recognition, (NLPR)
Institute of Automation, CAS,
School of Artificial Intelligence
University of Chinese Academy of Sciences
Beijing, China
yli@nlpr.ia.ac.cn

Jianhua Tao
National Laboratory of Pattern Recognition, (NLPR)
Institute of Automation, CAS,
School of Artificial Intelligence
University of Chinese Academy of Sciences,
CAS Center for Excellence in Brain Science and Intelligence Technology,
Institute of Automation, CAS
Beijing, China
jhtao@nlpr.ia.ac.cn

Su-Jing Wang
Key Laboratory of Behavior Science,
Institute of Psychology, CAS
Beijing, China
wangsujing@psych.ac.cn

***Abstract*—Recently, micro-expression recognition has obtained more and more attention due to its spontaneity. Moreover, how to extract feature from these subtle image sequences is still a hard issue. Currently, a lot of local patterns have been applied to describe these micro-changes, but in essence they are seldom specialized to capture fine structure changes in video clips and often need tens of frames to generate a descriptor through interpolation between frames. To make up for these deficiencies, this paper proposes a novel local pattern i.e. Local Two-Order Gradient Pattern (LTOGP) to describe the subtle changes in micro-expression sequences. Using this pattern, we can not only capture subtle changes in videos, but also only make use of 5 frames to generate a descriptor of a micro-expression clip rather than tens or hundreds of frames in related literatures. Experiments are conducted on two databases of CASME and CASME2 and the results show that our proposed method gains promising performance for micro-expression recognition with proposed method.**

***Index Terms*—micro-expression, local pattern, local two-order gradient pattern**

I. Introduction

Micro-expression is a kind of spontaneous facial expression and it has two intrinsic characteristics: short duration that is less than 0.5s [1,2] and low intensity [3] that is even not realized by ourselves. In other words, micro-expression could reveal certain emotion state that people try to conceal. Due to these reasons, micro-expression gradually becomes a very useful pathway to analysis people's true emotion and could be applied in many aspects, such as interrogations [4], clinical diagnosis [5], national security [6]. On account of these applications, micro-expression recognition attracts more and more people's attention, and gradually becomes a research hotspot.

It is well known that feature extraction is a very critical stage for all recognition tasks. Nevertheless, as mentioned above, micro-expression is an image sequence where there are small changes between current frame and other frames. An example from Chinese Academy of Sciences Micro-expression Database (CASME) [14] could illustrate this situation in Fig. 1. Therefore, we can say that it is hard to extract proper feature that reflects the changing process in the sequence. As is reported in [7], even though highly trained individuals only achieve the accuracy of less than 50%. But the poor performance also prompts researchers to find a method to handle this hard issue.

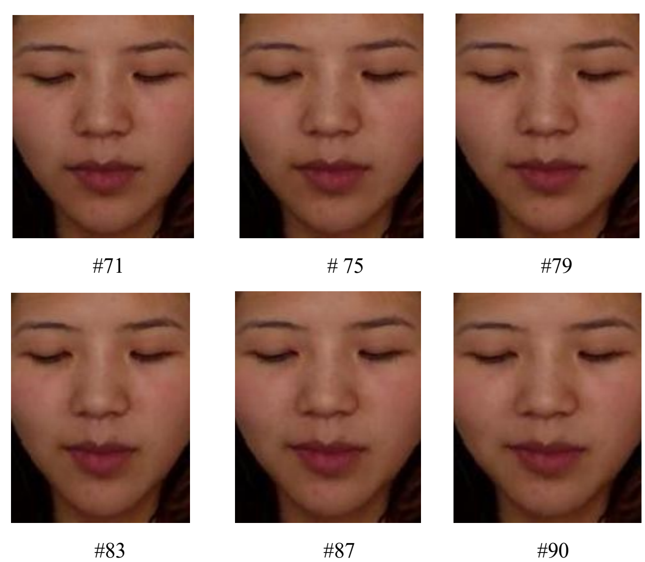


Fig.1. Six frames in a micro-expression sequence of 'contempt'

In face recognition field, a lot of local patterns have shown to be very effective to describe a face image[9][10]. This is mainly for two reasons: firstly, it reflect the overall information by integrating local pattern, which can capture local regional

This work is supported by the National Natural Science Foundation of China (NSFC) (No.61425017, No. 61773379), the National Key Research & Development Plan of China (No. 2017YFB1002804) and the Major Program for the National Social Science Fund of China (13&ZD189).

978-1-5386-5311-1/18/$31.00 ©2018 IEEE

structure in different positions in an image; secondly, these patterns could furthermore improve their robustness through some operations between pixels and thresholding. Intuitively, these methods in exploring local pattern also can be applied to micro-expression recognition directly, but it should be pointed out that face recognition technology describes one image, micro-expression recognition aims at an image sequence. Thus, some researchers proposed some corresponding methods to recognize micro-expression, such as [8, 12]. But these methods are difficulty in capturing micro-changes, because they mainly use the information of pixel intensity difference, while misses structure information. Besides, these methods usually need to interpolate between frames, which is time-consuming and introduces some fake images in videos and then probably reducing performance. Therefore, this paper proposes a novel local pattern i.e. Local Two-Order Gradient Pattern (LTOGP) to represent local image structure. LTOGP describes current local structure by neighbor pixels' two-order gradient and guarantee its robustness by thresholding to gradient value. These operations can capture micro structural changes in image sequences effectively. Meanwhile, this method can only use 5 frames to represent one micro-expression sequence rather than tens even hundreds of frames in some literatures.

The remainder of this paper is organized as follows. In Section 2, we introduce some related work about micro-expression recognition. Section 3 mainly present a detailed introduction about LTOGP. Experiment results and analysis will be demonstrated in Section 4. Finally, conclusion and future works are summarized in Section 5.

II. Related Work

Generally speaking, there are two kinds of feature in micro-expression recognition field, i.e. geometry-based and appearance-based features. However, it is very difficult to capture subtle facial movements (e.g., the eye wrinkles) to use geometry features, while appearance-based features can capture subtle changes due to it can extract the skin texture feature [8], we can also intuitively understand its superiority from the success of Local Binary Pattern (LBP) in face recognition, such as [9].

Amongst appearance-based methods, LBP is a common representation method to analysis face identity and expression. In [11], G.Zhao *et al*. proposed Local Binary Pattern from Three Orthogonal Planes(LBP-TOP) to analyze and recognition face expression, but LBP is one of relatively macro-appearance descriptors due to it is conducted by pixel difference and binarization. Similarly, Local Binary Pattern with Six Intersection Points(LBP-SIP) [12] not only calculate spatial texture but also temporal textual description, whereas the difference between frames is small, especially after quantization. Besides, Wang [13] *et al*. viewed image sequences as tensors and proposed Sparse Tensor Canonical Correlation Analysis (STCCA) for micro-expression recognition. Although the method captures the principle component in the sequences, the descriptors are produced by an overall way. Thus, it would miss some details and don't capture the subtle changing process in the image sequence. Besides, most methods normalize image sequences to tens even hundreds of frames before processing, this preconditioning would have two shortcomings, the first one is that it constructs some fictitious frames between exiting frames, which would influence recognition performance; the second one is that inserting frames is high time cost due to needing inserting a lot of pixel positions no matter which linear or nonlinear.

Given these problems, this paper proposes a novel local pattern named by Local Two-Order Gradient Pattern (LTOGP), this pattern describes a local structure by neighbor pixels' two-order gradient and directly binary gradient value. Moreover, this paper only makes use of five frames (i.e. apex frame and two frames before and after it) to construct the sequence descriptor and not only save time cost but storage. In addition, experimental results also show that few frames can represent a kind of micro-expression.

III. Local Two-Order Gradient Pattern

In this section, we will firstly review LBP from gradient view. From this novel perspective, we would understand its strengths and shortcomings. Moreover, this paper will also demonstrate our method motivated by LBP and exhibit how it to explore structure information in the images.

A. Relation between LBP and gradient

Local pattern has been widely used to face recognition[9], expression recognition[10] and so on and gains promising performance. However, there is rarely a local pattern aiming at micro-expression or micro-changes in sequence data. As we all know, high order gradient could reflect detail information better. To show this view, we use w1=[-1;1] and w2=[1;-2;1] to filter lena image to compare one-order and two-order gradient filter results as shown in Fig. 2. From the figure, we can observe that the curves in (c) are finer than that in (b), in other words, two-order gradient is prone to subtle description in the image and this is crucial in micro-expression recognition field, while one-order gradient captures coarse information. What's more, two-order gradient gains the change information of one-order gradient or two-order gradient captures structural information in images. Therefore, it is rational to introduce two-order gradient in local pattern to improve the performance because of its those strengths.

(a) Original image (b) w1 filter result (c) w2 filter result

Fig.2. Original image and different filter results

As mentioned in Ref [9], LBP is generated by the difference between central pixel and its eight neighbor pixels in 3x3 square and then binarization. Furthermore, if we unfold the masks applied in LBP as shown in Fig. 3, then we would observe that these eight masks indicate eight neighbor pixels' gradients at eight different direction. Based on this point of view, we extend it to two-order gradient to represent the variation in image sequences.

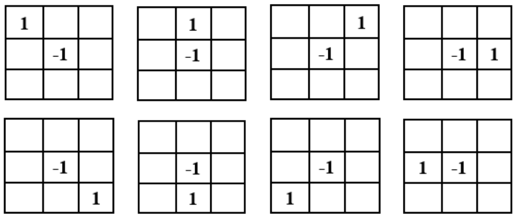


Fig.3. Eight masks in LBP

B. Local Two-Order Gradient Pattern

As mentioned above, for the sake of taking advantage of two-order gradient, this paper proposes LTOGP to more effectively capture subtle changes in micro-expression sequences by variation in structure. Naturally, in analogy with masks in LBP, this paper also proposes eight masks to capture structure information in neighbor, specifically as shown in Fig. 4. From them, we can see that these masks extract eight neighbor pixels' two order gradient at different direction and based on those analysis, we know that the proposed pattern is able to describe finer variation in local region. Moreover, the size of masks in LTOGP is larger than that in LBP, so the proposed method includes more information to represent local structure.

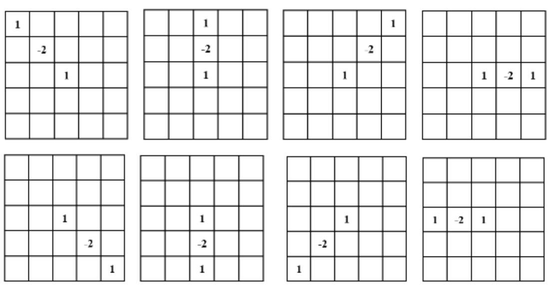


Fig.4. Eight masks in LTOGP

Similarly, as doing so in LBP, we also utilize threshold function as equation (1) to binary two-order gradient value. Thus, we can gain eight bits and then convert it into a decimal number with (2).

$$s(x)=\begin{cases}0 & ,x<0\\ 1 & ,x\geq 0\end{cases} \tag{1}$$

$$value(c)=\sum_{i=0}^{7} des_{i+1}\cdot 2^{i} \tag{2}$$

where $des_i, i=1,2,\cdots$ is the eight bits in descriptor and c is central pixel's position in local neighbor.

Also, these abilities to describe subtle changes in LTOGP, we use few frames to representation a kind of micro-expression in the image sequences. Specially, we choose five frames i.e. apex frame and two frames before and after it, the details will be shown in section IV. Then, we obtain final descriptor by concatenating these descriptors extracted from these five frames according to time sequence.

C. Descriptor generation

In two databases of CASME [14] and it's the second version (CASME2 [15]), there are two tables i.e. CASME-coded.xls and CASME2-coding-20140508.xlsx. From them, we can find ApexF1 and ApexF attributes in corresponding tables, which reflect apex in image sequences. In the process of descriptor generation, we firstly normalize each image into 128×128 and divide each image into 8×8 blocks, then calculate LTOGP value of each pixel in every block; secondly, we accumulate grayscale statistical histogram in each block and concatenate them, which forms a descriptor for the image; finally, video descriptor is generated by jointing descriptors from each frame. This procedure indicates that our descriptor is generated by using spatial information according to time sequence and this is also a kind of ways to represent time sequence. The overall process is shown in Fig. 5.

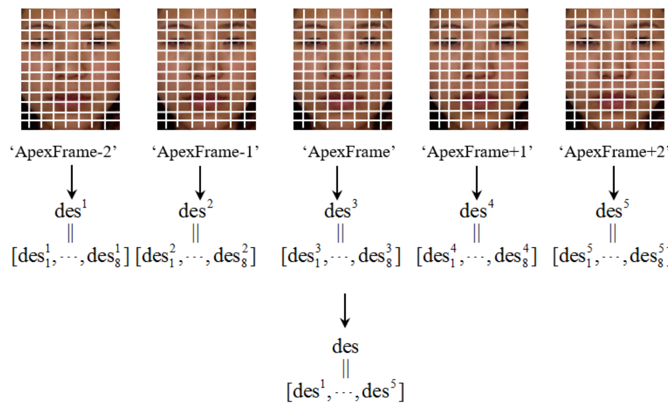


Fig.5. Overall process of descriptor generation
('ApexFrame' 'ApexFrame-1' 'ApexFrame-2' 'ApexFrame+1' 'ApexFrame+2' denote apex frame and before it and after it, respectively)

IV. EXPERIMENTS

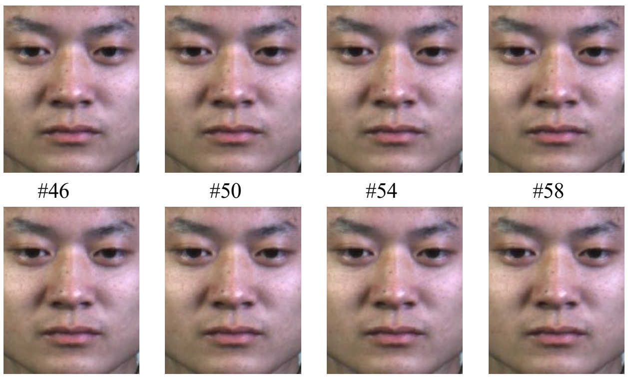
In this section, we will evaluate the proposed descriptor on CASME, CASME2 and mixture of both and show these experimental results.

A. Database description and protocol

In this subsection, we evaluate the performance of the proposed algorithm on Chinese Academy of Sciences Micro-expression Database (CASME) [14], CASME2[15] and mixture of CASME and CASME2. In Fig. 6 (a) and (b), we show two samples in these two database. In experiments, the size of each frame is normalized to 128×128 pixels. Moreover, we only utilize five frames i.e. ApexF1 or ApexF frame and two frames before and after it according to the table in CASME or CASME2. All the experiments were done on a PC with the same hardware i7-6700 CPU (3.40 GHZ), 12.0GB RAM, MATLAB 2017b and WIN 10.

58 #62 #66 #70

(a) Example of 'surprise' in CASME



(b) Example of 'happiness' in CASME2

Fig.6 Example of micro-expression in CASME and CASME2

B. *Evaluation*

In these experiments, we make three experiments on CASME and CASME2. What is more, we only use five classes data (happiness, surprise, disgust, repression and tense[1]) in these two database given the number of samples and the frame numbers of all samples are 5 (ApexF1 or ApexF and two frames after and before it in CASME and CASME2, respectively), while other methods utilize 70 frames and 150 frames in CASME and CASME2, respectively. The sizes of each frame are normalized to 128×128 pixels. To compare to other methods fairly, this paper also uses nearest neighbor (NN) as the classifier and randomly split the micro-expression samples so that n (n = 5,10,15,20) samples for each class as training samples and the remaining as test samples and this process is repeated 20 times. The mean recognition accuracies and standard deviations in CASME, CASME2 and mixture of CASME and CASME2 are listed in Table I, Table II, Table III. And their bar graphs are also shown in following.

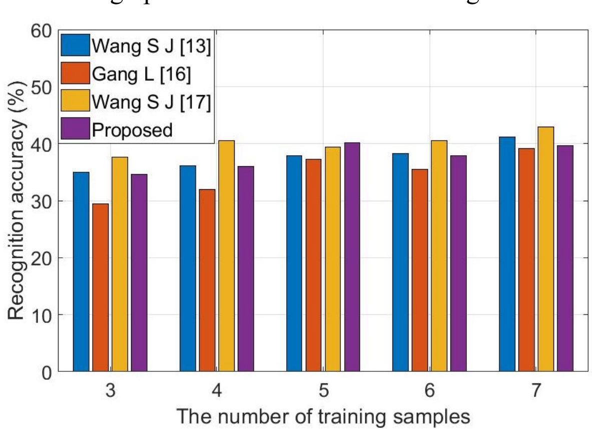


Fig.7. Bar graph of micro-expression mean recognition accuracies in CASME

[1]The term "others" in CASME2 is equivalent to "tense" in CASME[13].

TABLE I. MICRO-EXPRESSION MEAN RECOGNITION ACCURACIES(%) OF REF [13], [14], [15] AND PROPOSED IN CASME (MEAN ± STD)

n	Methods	Frames Used	Accuracy
3	Wang S J [13]	70	35.00 ± 6.27
	Gang L [16]	70	29.40 ± 7.16
	Wang S J [17]	70	**37.59 ± 7.23**
	Proposed	**5**	34.61 ± 8.26
4	Wang S J [13]	70	36.12 ± 7.06
	Gang L [16]	70	31.96 ± 6.05
	Wang S J [17]	70	**40.47 ± 5.51**
	Proposed	**5**	36.03 ± 6.56
5	Wang S J [13]	70	± 5.40
	Gang L [16]	70	37.21 ± 5.54
	Wang S J [17]	70	39.39 ± 6.01
	Proposed	**5**	**40.19 ± 5.17**
6	Wang S J [13]	70	38.25 ± 5.09
	Gang L [16]	70	35.50 ± 5.92
	Wang S J [17]	70	**40.53 ± 5.99**
	Proposed	**5**	37.93 ± 6.66
7	Wang S J [13]	70	41.20 ± 5.14
	Gang L [16]	70	39.14 ± 3.63
	Wang S J [17]	70	**42.95 ± 4.76**
	Proposed	**5**	39.65 ± 6.67

From Table I, the proposed method is comparable with others. Also, we can see that none is the best consistently, but it is worth mention that the proposed method only uses 5 frames while the remaining methods use 70 frames. Moreover, our method doesn't need to adjust any parameters, while others must adjust the dimension of projecting to achieve the best performance. Therefore, our method is better than others if we consider these reasons. Furthermore, we can intuitively see from Fig.7 that the proposed local structure pattern has power potential to describe those subtle changes in micro-expression image sequences with few frames.

TABLE II. MICRO-EXPRESSION MEAN RECOGNITION ACCURACIES(%) OF REF [13], [14], [15] AND PROPOSED IN CASME2 (MEAN ± STD)

n	Methods	Frames Used	Accuracy
5	Wang S J [13]	150	31.02 ± 4.85
	Gang L [16]	150	27.58 ± 14.64
	Wang S J [17]	150	31.56 ± 4.88
	Proposed	**5**	**31.79 ± 4.28**
10	Wang S J [13]	150	35.36 ± 3.98
	Gang L [16]	150	33.98 ± 4.03
	Wang S J [17]	150	34.11 ± 3.97
	Proposed	**5**	**38.00 ± 4.06**
15	Wang S J [13]	150	36.99 ± 3.53
	Gang L [16]	150	34.15 ± 4.22
	Wang S J [17]	150	35.56 ± 3.77
	Proposed	**5**	**40.18 ± 4.11**
20	Wang S J [13]	150	38.39 ± 4.98
	Gang L [16]	150	36.51 ± 4.23
	Wang S J [17]	150	36.82 ± 4.04
	Proposed	**5**	**44.00 ± 2.74**

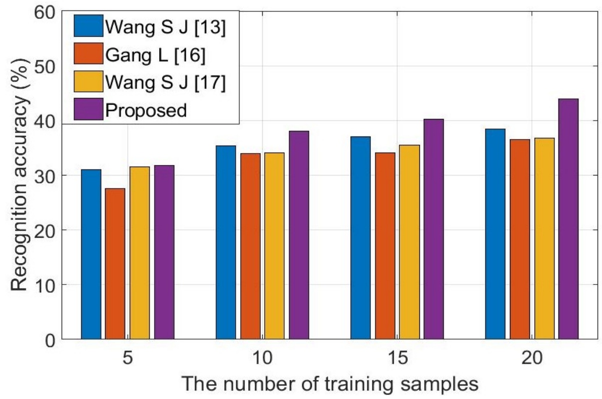


Fig.8. Bar graph of micro-expression mean recognition accuracies in CASME2

TABLE III. MICRO-EXPRESSION MEAN RECOGNITION ACCURACIES(%) OF REF [13], [14], [15] AND PROPOSED IN MIXTURE OF CASME AND CASME2 (MEAN ± STD)

n	Methods	Frames Used	Accuracy
15	Wang S J [13]	150	37.68 ± 3.94
	Gang L [16]	150	34.74 ± 3.58
	Wang S J [17]	150	36.29 ± 3.41
	Proposed	**5**	**38.41 ± 4.35**
20	Wang S J [13]	150	40.09 ± 2.61
	Gang L [16]	150	37.58 ± 4.44
	Wang S J [17]	150	39.20 ± 3.26
	Proposed	**5**	**41.84 ± 3.44**
25	Wang S J [13]	150	40.56 ± 3.13
	Gang L [16]	150	40.51 ± 4.26
	Wang S J [17]	150	38.20 ± 2.63
	Proposed	**5**	**42.12 ± 2.81**
30	Wang S J [13]	150	42.33 ± 2.79
	Gang L [16]	150	41.77 ± 3.87
	Wang S J [17]	150	39.31 ± 3.17
	Proposed	**5**	**43.83 ± 2.95**
35	Wang S J [13]	150	43.47 ± 2.76
	Gang L [16]	150	42.62 ± 2.71
	Wang S J [17]	150	41.61 ± 2.37
	Proposed	**5**	**44.55 ± 3.93**

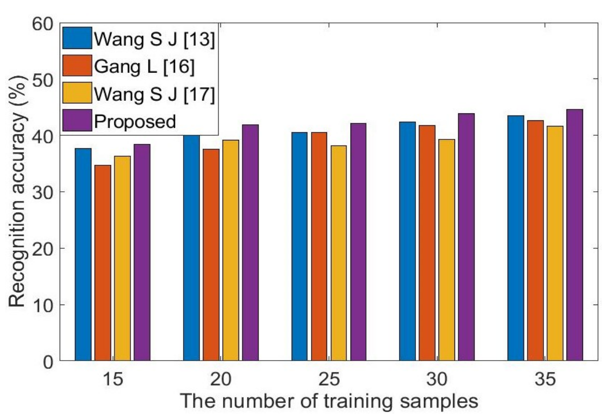


Fig.9. Bar graph of micro-expression mean recognition accuracies in mixture of CASME and CASME2

Besides, as shown in Table II and Table III, we can see that the proposed method obtains better performance than other methods consistently. We think that three reasons lead to this result. Firstly, there are more samples, fixed illumination and higher resolution (both temporal and spatial) in CASME2 than CASME [13]. Secondly, there are explicit and single ApexF frames in CASME2, but in CASME there are two kinds of apex i.e. ApexF1 and ApexF2, which will bring about uncertainty and fuzziness. Thirdly, the proposed method only uses five original frames, while the others use 150 frames by linear interpolation, which introduces some fake information and changes original subtle variation tendency. Furthermore, it is worth mentioning that from Fig.8 and Fig. 9 we can observe that our descriptor gains better results by means of extracting feature from each frame according to time order. This is equivalent to capture sequential variation through spatial concatenation.

V. CONCLUSIONS AND FUTURE WORK

This paper proposes a novel image pattern i.e. Local Two-Order Gradient Pattern and we illustrate the benefits of it that it can describe finer texture in images. In micro-expression recognition, our method only need five frames to represent a micro-expression video. Comparing to previous methods, our method obtains better performance in the experiments. However, in this paper we still need normalize the number of image frames, thus we would combine descriptor aggression with proposed method to get power image representation in future work.

REFERENCES

[1] 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. (2013) 1–14.

[2] D. Matsumoto, H. Hwang, Evidence for training the ability to read microexpressions of emotion, Motiv. Emot. 35 (2) (2011) 181–191.

[3] Porter S, Brinke L T. Reading between the Lies: Identifying Concealed and Falsified Emotions in Universal Facial Expressions[J]. Psychological Science, 2008, 19(5):508.

[4] M. Frank, C. Maccario, V. Govindaraju, Behavior and Security, Greenwood Publishing Group, Santa Barbara, California, 2009, pp. 86–106.

[5] M.G. Frank, M. Herbasz, A.K.K. Sinuk, C. Nolan, I see how you feel: traininglaypeople and professionals to recognize fleeting emotions, in: The AnnualMeeting of the International Communication Association, New York, 2009.

[6] M. OSullivan, M. Frank, C. Hurley, J. Tiwana, Police lie detection accuracy: the effect of lie scenario, Law Hum. Behav. 33 (6) (2009) 530–538.

[7] M. Frank, M. Herbasz, K. Sinuk, A. Keller, and C. Nolan, "I see howyou feel: Training laypeople and professionals to recognize fleetingemotions," in International Communication Association, 2009.

[8] Huang X, Wang S J, Liu X, et al. Discriminative Spatiotemporal Local Binary Pattern with Revisited Integral Projection for Spontaneous Facial Micro-Expression Recognition[J]. IEEE Transactions on Affective Computing, 1949, PP(99):1-1.

[9] Ahonen T, Hadid A, Pietikainen M. Face Description with Local Binary Patterns: Application to Face Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2006, 28(12):2037.

[10] Rivera, Adin Ramirez, Jorge Rojas Castillo, and Oksam Oksam Chae. "Local directional number pattern for face analysis: Face and expression

recognition." IEEE transactions on image processing 22.5 (2013): 1740-1752.

[11] G. Zhao and M. Pietikainen, “Dynamic texture recognition using local ¨binary pattern with an application to facial expressions,” IEEE Trans. On PAMI, vol. 29, no. 6, pp. 915–928, 2009.

[12] Wang Y, See J, Phan R C W, et al. LBP with Six Intersection Points: Reducing Redundant Information in LBP-TOP for Micro-expression Recognition[C]// ACCV. 2014:21–23.

[13] Wang S J, Yan W J, Sun T, et al. Sparse Tensor Canonical Correlation Analysis for Micro-expression Recognition[J]. Neurocomputing, 2016, 214(C):218-232.

[14] Yan, Wen-Jing, et al. "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 on. IEEE, 2013.

[15] Yan, Wen-Jing, et al. "CASME II: An improved spontaneous micro-expression database and the baseline evaluation." PloS one 9.1 (2014): e86041.

[16] Gang L, Yong Z, Liu Y L, et al. Three Dimensional Canonical Correlation Analysis and Its Application to Facial Expression Recognition[M]// Intelligent Computing and Information Science. Springer Berlin Heidelberg, 2011:56-61.

[17] Wang S J, Zhou C G, Zhang N, et al. Face recognition using second-order discriminant tensor subspace analysis[J]. Neurocomputing, 2011, 74(12):2142-2156.