MEGC2020 - The Third Facial Micro-Expression Grand Challenge

Jingting LI[1], Su-Jing Wang[1,2,*], Moi Hoon Yap[3], John See[4], Xiaopeng Hong[5], Xiaobai Li[6]

[1] Key Laboratory of Behavior Sciences, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[2] Department of Psychology, University of the Chinese Academy of Sciences, Beijing, China
[3] Faculty of Science and Engineering, Manchester Metropolitan University, UK
[4] Faculty of Computing and Informatics, Multimedia University, Malaysia
[5] Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, China
[6] Center for Machine Vision and Signal Analysis, University of Oulu, Finland

Abstract— **The recent emergence of automatic facial micro-expression analysis has attracted a lot of attention in the last five years. Compared to the advances made in micro-expression recognition, the task of micro-expression spotting from long videos is tremendously in need of more effective methods. This paper summarises the 3rd Facial Micro-Expression Grand Challenge (MEGC 2020) held in conjunction with the 15th IEEE Conference on Automatic Face and Gesture Recognition (FG) 2020. In this workshop, we propose a new challenge of spotting both macro- and micro-expressions from long videos, to spur the community to develop new techniques for micro-expression spotting and also to extend facial micro-expression analysis to more complex real-world scenarios where micro-expressions are likely to be intertwined among normal expressions. In this paper, we outline the evaluation protocols for the challenge task, and describe the datasets involved. Then, we summarize the methods from the accepted challenge papers, present the comparison and analysis of results, as well as future directions.**

I. INTRODUCTION

The facial expression is one of the most important external indicators to reveal the emotion and the psychological status of a person [2]. Micro-facial expressions [6] are involuntary local facial movements, that occur spontaneously when a person attempts to hide the genuine emotion, most likely in a high-stakes environment [7]. Meanwhile, the duration of micro-expressions is very short with a general duration of not more than 500 milliseconds (ms) [6], which is an important characteristic that distinguishes them from normal facial expressions. Computational analysis and automation of tasks on micro-expressions have emerged in the last decade of face research, with a strong interest appearing since 2014 [20], [14], [12], [5]. As such, the availability of a few spontaneously induced facial micro-expression datasets allows the research to expand further from the computational aspect. Among them, there are two facial macro- and micro- expression databases that contain long videos: the CAS(ME)^2 [16] and the SAMM Long Videos dataset [21].

The community organized the first and the second Micro-Expression Grand Challenge (MEGC2018 [22] and MEGC2019 [18]), which mainly focused on the task of

micro-expression recognition. For micro-expression spotting, while much research has been done on short videos, there has been not many attempts to spot micro-expressions on *long videos*. Hence, the third edition of this workshop aims to organize the challenge involving spotting both macro- and micro-expression on long videos using the two designated state-of-the-art benchmark datasets.

This workshop also aims to promote interactions between researchers and academics not only within the facial micro-expression research domain, but also by including those from broader and more general areas of research on expression and psychology. Therefore, we also solicit original works that address a variety of challenges of micro-expression research, including micro-expression research using self-supervised learning, micro-expression feature representation and computational analysis, unified micro-expression spot-and-recognize schemes, new micro-expression dataset and psychology studies of micro-expression.

II. CHALLENGE TASK

The goal of this challenge is to spot both macro- and micro-expression intervals (from onset to offset) in long video sequences. In this challenge, we focus on detecting facial expressions (macro and micro) from the CAS(ME)^2 and SAMM Long Videos datasets.

A. Databases

In MEGC2019, the spotting challenge used the CAS(ME)^2 [16] and SAMM [4] database. Yet, the previous task only spotted micro-expressions in these two databases. In MEGC2020, both macro and micro-expressions are involved in the spotting task, which presents a realistic real-world scenario. CAS(ME)^2 consists of 22 participants and 98 long videos, including 300 macro-expressions and 57 micro-expressions. Meanwhile, the authors of the SAMM dataset released their corresponding long videos, i.e. the SAMM Long Videos dataset [21], which consists of 147 long videos, including 343 macro-movements and 159 micro-movements in the long videos. Table I summarizes the two databases.

B. Performance Metrics

The 2nd MEGC [18] proposed a result evaluation method for micro-expression sequence spotting. The spotted intervals

Dataset	CAS(ME)^2	SAMM Long Videos
Video samples	98	147
Macro-expressions	300	343
Micro-expressions	57	159
Resolution	640×480	2040×1088
FPS	30	200

are measured in order to avoid potential inaccuracies arising from less precise annotations. Since in this 3rd MEGC, both macro- and micro-expressions are to be spotted, we adopt the metrics proposed by Li et al [11]. We consider that each ground-truth interval corresponds to one single spotted interval. Then, the True Positive (TP) for each video, and other related metrics are computed in similar fashion to the 2nd MEGC. The performance for the entire database is evaluated as follows.

The following representative notations are determined:

- There are V videos consisting of M_1 macro-expressions (MaEs) sequences and M_2 micro-expression (MEs) sequences;
- There are N_1 MaE intervals and N_2 ME intervals in total;
- There are A_1 TPs for MaE and A_2 TPs for ME in total.

For ease of representation, we can regard all videos in the dataset as "one long video", hence the metrics determine the overall performance across all videos. The results are firstly evaluated for the MaE spotting and ME spotting separately, before the overall result is determined for the entire dataset. The *recall* and *precision* for the entire dataset can be calculated by following formulas:

- For macro-expression:

$$Recall_{MaE\_D} = \frac{A_1}{M_1} \qquad (1)$$

$$Precision_{MaE\_D} = \frac{A_1}{N_1} \qquad (2)$$

- For micro-expression:

$$Recall_{ME\_D} = \frac{A_2}{M_2} \qquad (3)$$

$$Precision_{ME\_D} = \frac{A_2}{N_2} \qquad (4)$$

- For overall evaluation:

$$Recall_D = \frac{A_1 + A_2}{M_1 + M_2} \qquad (5)$$

$$Precision_D = \frac{A_1 + A_2}{N_1 + N_2} \qquad (6)$$

Correspondingly, the *F1-score* for all these three evaluations can be obtained by:

$$F1 - score = \frac{2 \times (Recall \times Precision)}{Recall + Precision} \qquad (7)$$

The first place submission to this challenge is the method with the best overall score for spotting micro- and macro-expressions.

III. METHODOLOGY

This section summarizes the methods proposed by five accepted submissions for macro- and micro-expression spotting on long video sequences in MEGC2020.

A. Baseline method by He et al. [9]

To establish the baseline of MEGC2020, He et al. implemented the method of Main Directional Maximal Difference Analysis (MDMD) which was proposed by Wang et al. [19]. The MDMD method uses the magnitude of maximal difference in the main direction of optical flow features to spot facial movements. The author first pre-processed the raw videos into cropped face region sequences. Then the MDMD method is applied to predict whether a frame belongs to specific facial movements. To obtain target intervals, the adjacent frames are used to predict if a macro- or micro-expressions is found based on its interval, while the intervals that are too long or too short are discarded. Parameters used were adjusted to specific expression types for specific datasets.

B. Local Bilinear Convolutional Neural Network by Pan [15]

In order to alleviate the problem of locating low intensity movements on the face when facial micro-expressions occur, Pan proposed a method based on Local Bilinear Convolutional Neural Network (LBCNN). Considering the low intensity of facial micro-expressions, and the occurrence of micro-expressions at only a local area of the face, the author turned the micro-expression spotting task in long videos sequences into a fine-grained image recognition problem. Since the Bilinear Convolutional Neural Network (BCNN) has proven to be effective in fine-grained image recognition, the author used the BCNN structure to extract global and local features of the face area from each frame in the long video sequence. The final classification result is obtained using a fusion of the extracted global and local features.

C. Optical Strain Based Spotting Method by Gan et al. [8]

Gan et al. introduced a novel spotting framework based on optical strain. The method is inspired by the apex spotting mechanism presented by Liong et al. [13]. The authors only considered features of salient regions in each image; only specific important regions (i.e., mouth and eyebrows) that contribute towards facial expressions are considered for feature extraction. Then, the optical flow components are computed between the first frame of the video and the consecutive frames. The frames with significant flow intensity are selected as the apex frames. Finally, certain frame numbers are added before and after the detected apex to form the final spotted macro- and micro-expression interval.

D. Spatio-Temporal Fusion by Zhang et al. [23]

Zhang et al. proposed an automatic spotting framework, so as to deal with the transformation caused by uncontrolled head movements. The author disentangled the local movement vector from the global optical flow field by estimating the mean optical flow in the nose region. The local optical flow field is obtained by obtaining the difference between superposition of optical field and standard global movement vector. Then, a spatio-temporal feature fusion matrix is used to describe the spatial and temporal information based on the row and column relationship in each region of interest (ROI). A specific pattern (SP-pattern) related to magnitude and angle is extracted from the matrix; this pattern contains all the information from a micro-expression interval. The onset, apex, and offset information can be obtained according to the SP-pattern. Finally, the authors employ a multi-scale filter to remove high frequency noise and preserve the crests of different intensities.

E. Expression spotting using AU detection by Yap et al. [21]

Yap et al. preprocessed the long videos using Open-Face [1], which includes face alignment and detection of facial action units (AUs). For the spotting task, the authors combine the intensity of all the AUs detected, and normalised it to a scale of 0 to 1. Subsequently, the Savitzky-Golay filter [17] with 21st order was used for noise reduction. Furthermore, by using a custom analysis algorithm modified from [10], the onset and offset frames of the processed data are obtained and compared against the ground truth. This algorithm uses Daubechies wavelet [3] with scaling function of 2 and a level 3 signal decomposition for signal smoothing. For peak detection, the peak intensity is defined as the point where it is higher than the 7 points that come immediately before and after it. Filters of a threshold value of 75th percentile are also implemented to filter out low peaks caused by noise. The spotted interval is finally obtained by subtracting the onset from the offset.

IV. CHALLENGE RESULT

In this section, we report and analyze the result of the challenge. We have received submissions from five participating teams. The results reported in this section are presented in the respective submitted papers, and verified again using the submitted output logs provided by the participants. The challenge also required the participants to provide their code implementations in Github to ensure reproducibility of the submitted methods and also to encourage more research in automatic micro-expression analysis. Table II summarizes the performance (F1-score) of four submissions against the baseline proposed by He at al. [9]. Yet, only three groups performed the experiments on both CAS(ME)^2 and SAMM Long Videos (SAMM-LV) databases. The other two groups reported their spotting results on either CAS(ME)^2 or SAMM-LV.

Overall, the method by Zhang et al. [23] emerged as the best method with a full comparison on both databases. Their proposed method outperformed the others on both

CAS(ME)^2 and SAMM-LV. The work of Yap et al. [21] is notably strong on the macro-expression spotting for SAMM-LV, obtaining the highest overall F1-score in this dataset, but no results were reported on the CAS(ME)^2.

From these results, we highlight a number of interesting observations:

1) On average, the spotting method performs better on macro-expressions than micro-expressions. This is expected since macro-expressions are normally regarded as a global facial movement, hence its movement intensity is much greater than that of micro-expressions. Also, representative features for macro-expression sequence are likely more well-formed than that of micro-expressions, which are minute and rapid in duration. Notably, since the OpenFace is trained by facial expressions, i.e. macro-expression, it did worked well for macro-expression spotting in [21].

2) All the micro-expression spotting performances surpass that of [11] (0.0055 for CAS(ME)^2 and 0.0229 for SAMM). Yet, we regard the overall performance as still considerably weak, i.e. the average F1-score is still less than 0.01. This is primarily due to the characteristics of micro-expressions: low in intensity, short duration and very localized movements. Moreover, the variations in illumination, head pose, eye blinks and other kind of unwanted influences are inevitable in a real-world scenario; as such, the current methods are still insufficiently strong to distinguish a micro-expression occurrence from all these noises (which in turn contributes towards a big number of false positives). Therefore, we see much room for improvement in the micro-expression spotting task.

3) There is only one method which utilized deep learning technique [15]. All other methods still rely on hand-crafted features, followed by locating the movement by feature difference computation. Nevertheless, the improvement brought about by the deep learning method is not very evident. We believe that data-driven machine learning techniques for micro-expression spotting could be restricted by small sample size problem.

4) It is difficult to tell which of CAS(ME)^2 and SAMM Long Videos dataset is more challenging for micro-expression spotting by way of Table II. The CAS(ME)^2 dataset was captured at a slower FPS and lower resolution whereas the SAMM Long Videos dataset contains quite a diverse range of age and ethnicity. Also, the long duration makes the videos contain many irrelevant movements. All these factors influence the final spotting performance of both datasets in varied amounts.

V. CONCLUSION AND FUTURE CHALLENGE

This paper summarizes the Third Facial Micro-Expressions Grand Challenge (MEGC2020) workshop which involves the new challenge of macro- and micro-expression spotting. We received a total of five submissions for the spotting task, which is much higher compared to

Database	CAS(ME)^2			SAMM Long Videos		
Method	Macro-expression	Micro-expression	Overall	Macro-expression	Micro-expression	Overall
Baseline [9]	0.1196	0.0082	0.0376	0.0629	0.0364	0.0445
Gan et al [8]	0.1436	0.0098	0.0448	-	-	-
Pan [15]	-	-	0.0595	-	-	0.0813
Zhang et al. [23]	0.0547	0.2131	**0.1403**	0.1331	0.0725	**0.0999**
Yap et al. [21]	-	-	-	0.4081	0.0508	**0.3299**

that in MEGC2019, which is promising. Although the task of spotting micro-expressions from long videos is very challenging, the current momentum in micro-expression research is encouraging.

Micro-expression spotting is an important and practical first-in-the-pipeline problem, in the quest towards fully automated analysis. As such, robust and precise tracking of local movements in micro-expression sequences from long videos or unconstrained in-the-wild scenarios may be beneficial in advancing this field. Therefore, we encourage the research community to pay more attention to new proposals and innovative ideas on micro-expression spotting. Further to this workshop, works are also in progress to further increase the amount of available data. To date, there are only around 832 samples (i.e. video clips) from a total of five public spontaneous databases. In addition, it is also impossible to unify the standardization of ME labeling by different annotators. To tackle this problem from a different direction, we expect that the recently emerged self-supervised learning technique can help provide the necessary means to improve the performance of micro-expression spotting and recognition.

REFERENCES

[1] T. Baltrušaitis, P. Robinson, and L.-P. Morency. Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10. IEEE, 2016.

[2] R. L. Birdwhistell. Communication without words. *Ekistics*, pages 439–444, 1968.

[3] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE transactions on information theory*, 36(5):961–1005, 1990.

[4] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap. SAMM: A spontaneous micro-facial movement dataset. *IEEE Transactions on Affective Computing*, 9(1):116–129, 2018.

[5] A. K. Davison, M. H. Yap, N. Costen, K. Tan, C. Lansley, and D. Leightley. Micro-facial movements: An investigation on spatio-temporal descriptors. In *European conference on computer vision*, pages 111–123. Springer, 2014.

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

[7] M. G. Frank and P. Ekman. The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology*, 72(6):1429, 1997.

[8] Y. S. Gan, S. T. Liong, D. Zheng, S. Li, and C. Bin. Optical strain based macro- and micro-expression sequence spotting in long video. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2020.

[9] Y. He, S.-J. Wang, J. Li, , and M. H. Yap. Spotting macro- and micro-expression intervals in long video sequences. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2020.

[10] B. M. Hewitt, M. H. Yap, E. F. Hodson-Tole, A. J. Kennerley, P. S. Sharp, and R. A. Grant. A novel automated rodent tracker (art), demonstrated in a mouse model of amyotrophic lateral sclerosis. *Journal of neuroscience methods*, 300:147–156, 2018.

[11] J. Li, C. Soladie, R. Sguier, S. Wang, and M. H. Yap. Spotting micro-expressions on long videos sequences. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 1–5, 2019.

[12] S.-T. Liong, J. See, R. C.-W. Phan, A. C. Le Ngo, Y.-H. Oh, and K. Wong. Subtle expression recognition using optical strain weighted features. In *Asian conference on computer vision*, pages 644–657. Springer, 2014.

[13] S.-T. Liong, J. See, K. Wong, and R. C.-W. Phan. Automatic micro-expression recognition from long video using a single spotted apex. In *Asian Conference on Computer Vision*, page 345–360. Springer, 2016.

[14] A. Moilanen, G. Zhao, and M. Pietikäinen. Spotting rapid facial movements from videos using appearance-based feature difference analysis. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, page 1722–1727. IEEE, 2014.

[15] H. Pan. Local bilinear convolutional neural network for spotting macro- and micro-expression intervals in long video sequences. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2020.

[16] F. Qu, S. J. Wang, W. J. Yan, H. Li, S. Wu, and X. Fu. CAS(ME)^2: A database for spontaneous macro-expression and micro-expression spotting and recognition. *IEEE Transactions on Affective Computing*, 9(4):424–436, 2017.

[17] A. Savitzky and M. J. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

[18] J. See, M. H. Yap, J. Li, X. Hong, and S.-J. Wang. MEGC 2019 - the second facial micro-expressions grand challenge. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 1–5, 2019.

[19] S.-J. Wang, S. Wu, X. Qian, J. Li, and X. Fu. A main directional maximal difference analysis for spotting facial movements from long-term videos. *Neurocomputing*, 230:382–389, 2017.

[20] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, and X. Fu. Micro-expression recognition using dynamic textures on tensor independent color space. In *2014 22nd International Conference on Pattern Recognition*, pages 4678–4683. IEEE, 2014.

[21] C. H. Yap, C. Kendrick, and M. H. Yap. Samm long videos: A spontaneous facial micro-and macro-expressions dataset. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2020.

[22] M. H. Yap, J. See, X. Hong, and S.-J. Wang. Facial micro-expressions grand challenge 2018 summary. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 675–678. IEEE, 2018.

[23] L.-w. Zhang, J. Li, S.-J. Wang, W.-j. Yan, X.-h. Duan, S.-c. Huang, and H. Xie. Spatio-temporal fusion for macro- and micro-expression spotting in long video sequences. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2020.