2nd Micro-Expression Grand Challenge (MEGC)

in conjunction with IEEE Automatic Face and Gesture Recognition (FG) 2019, in Lille, France

Cross-DB Challenge

The previous Cross-DB challenge in the 1st MEGC [1] used a combination of 2 datasets (CASME II and SAMM), with objective class labels as proposed in [2]. In this 2nd MEGC, the Cross-DB challenge increases its coverage to include the classic SMIC dataset, which is one of the earliest spontaneous micro-expression dataset to be created. To enable all three datasets to be used together, a common reduced set of emotion classes are used, with appropriate mappings from their original emotion classes.

The motivation behind this challenge is to mimic a more realistic scenario by:

- Increasing the number of subjects considered in the system, particularly with subjects captured from different environment and settings.
- Using a reduced set of general emotion classes to better accommodate contrasting types of emotions which have been elicited from different stimuli and environment setup. This also reduces the ambiguity in the elicited emotions caused by such differences.

As a side benefit, using more data samples through this consolidation also facilitates the use of more contemporary machine learning techniques that are data-driven in nature.

Guidelines

I. Download the data

To download the necessary data for this challenge, you need to obtain permission to use the **CASME II, SAMM** and **SMIC** datasets from their respective institutions that are hosting them. There are license agreements that are required to be submitted before the data is accessible.

The Cross-DB Challenge in MEGC 2019 uses three spontaneous facial micro-expression datasets:

- **SMIC dataset [3]**
- **CASME II dataset [4]**
- **SAMM dataset [5]**

This challenge involves a combination of all three datasets mentioned above. In order to facilitate classification based on common grouping of emotion, the original emotion classes are grouped based into three main classes (original classes in parenthesis):

- `negative` (i.e., 'Repression', 'Anger', 'Contempt', 'Disgust', 'Fear' and 'Sadness'),
- `positive` ('Happiness'), and
- `surprise` ('Surprise').

Videos containing other unrelated or undefined emotions are omitted.

The summary of the distribution of samples for all three datasets are given in the table below:

Emotion Class	SMIC	CASME II	SAMM	3DB-combined
Negative	70	88[†]	91[‡]	249

Positive	51	32	26	109
Surprise	43	25	15	83
TOTAL	164	145	132	441

† Negative class of CASME II consists of samples from its original emotion classes of Disgust and Repression.

‡ Negative class of SAMM consists of samples from its original emotion classes of Anger, Contempt, Disgust, Fear and Sadness.

The consolidated ground truth file is provided: combined_3class_gt.csv

Baseline results for the consolidated '3DB-combined' dataset and the individual datasets, on the new 3-emotion classes, will be released in mid-December 2018.

Submitted methods are expected to be able to exceed the baseline performance.

Challenge Task

In the previous MEGC, two protocols were established to evaluate in a cross-database setting: Holdout-Database Evaluation (HDE) and Composite Database Evaluation (CDE).

In this year's challenge, we will only adopt the CDE, with additional reporting of per-database performances. The HDE protocol is not adopted as it will be a lengthy process which may involve many possible permutations of train-test partitions from the three datasets.

Composite Database Evaluation (CDE):

All samples from the datasets (SMIC, CASME II and SAMM) are combined into a single composite database, based on the reduced emotion classes. Leave-One-Subject-Out (LOSO) cross-validation is used to determine the training-testing splits (i.e. each subject group is held out as the testing set while all remaining samples are used for training). There are altogether 68 subjects (16 from SMIC, 24 from CASME II, 28 from SAMM) after the databases are consolidated based on the new generic classes. This protocol mimics a realistic scenario where people from diverse backgrounds (ethnicity, gender emotional sensitivities) are enrolled separately in different environment and settings, into a single recognition system. The LOSO cross-validation also ensures subject-independent evaluation.

Performance Metric. The composite database is clearly imbalanced in terms of its class distribution, i.e. the distribution for `surprise:positive:negative` classes are in the ratio of 1 : 1.3 : 3 (e.g. Accuracy of the system is 0.565 simply by making a naive `negative` class prediction). To properly handle such class imbalances [6], the performance is to be reported with two balanced metrics:

- Unweighted F1-score (or F-measure), also commonly known as macro-averaged F1-score. This flavour of F1-score is a good choice in multi-class settings for providing equal emphasis on rare classes. To calculate this, firstly obtain all the True Positives (TP), False Positives (FP) and False Negatives (FN) over all k folds of LOSO[1] by each class c (of C classes), and proceed to compute their respective F1-scores. The final balanced F1-score is determined by averaging the per-class F1-scores:

$$TP_c := \sum_{i=1}^{k} TP_c^{(i)}$$

[1] See the paper by [7] for the most unbiased way of calculating F1-score in a k-fold cross-validation setting. It caters well for cases of strong class imbalance.

$$FP_c := \sum_{i=1}^{k} FP_c^{(i)}$$

$$FN_c := \sum_{i=1}^{k} FN_c^{(i)}$$

$$F1_c := \frac{2 \cdot TP_c}{2 \cdot TP_c + FP_c + FN_c}$$

$$UF1 = F1_c / C$$

- Unweighted Average Recall (UAR), or also known as "balanced accuracy" of the system. This is a more reasonable metric in place of the standard Accuracy (or Weighted Average Recall) metric which may be bias towards classifiers that predict the larger classes well. The per-class accuracy scores Acc_c are first calculated, before averaging by the number of classes:

$$UAR = \frac{1}{C} \sum Acc_c$$
$$\text{where } Acc_c = \frac{TP_c}{n_c}$$

Both these metrics provide a balanced judgement whether an approach can predict all classes equally well, hence reducing the possibility that an approach could be well-fitted to only work for certain classes.

Submission

For the purpose of result verification and to encourage reproducibility and transparency, all entries must submit the following:

- **An evaluation log file** (.txt, or .csv) indicating the fold, the ground truth class, and the predicted class. This is to ensure that all submissions are fairly and correctly evaluated for comparisons.
- **A paper** highlighting the contribution of the submission, but not limited to, the method, experimental results and analysis, prepared according to the format stipulated by IEEE FG 2019 (Refer to here for detailed instructions). All challenge entries should be accompanied by a paper submission.
- **GitHub repository URL** containing codes of your implemented method, and all other relevant files such as feature/parameter data. To help publicize our workshop and domain area, please do mention (or add relevant links on) MEGC Workshop 2019 and FG 2019. You may provide this URL in a simple text file while submitting.

For all files except for the paper, please submit in a single zip file and upload to the submission system as supplementary material.

Sample log file

For each LOSO fold, a header line indicates the database name and the subject name (or subject folder name), followed by one line for each video sample in the folder, indicating the video file name, ground truth label, and predicted label, in this exact order.

```
casme2 sub01
EP02_01f 1 1
EP19_05f 0 1
...

smic s01
s01_ne_01 0 0
s01_ne_02 0 1
...

samm 006
006_1_2 0 2
006_1_3 0 0
```

The submission portal is now open at Microsoft CMT: https://cmt3.research.microsoft.com/MEGC2019
Challenge Deadline: **27 January 2019, 2359 PST (UTC -8)**

Rules

The organizers reserve the right to disqualify submissions with on the basis of

- Incomplete submission
- Challenge results that do not tally with the run codes, or are likely to be suspicious, i.e. out-of-norm from the distribution of scores from submitters.
- Non-submission of accompanying paper.
- Submission of an accompanying paper that has a substantial overlap with any other paper already submitted or published, or to be submitted during the review period

For further enquiries, please contact:

John See johnsee@mmu.edu.my
Sze-Teng Liong stliong@fcu.edu.tw

References:

[1] Yap, M. H., See, J., Hong, X., & Wang, S. J. (2018). Facial Micro-Expressions Grand Challenge 2018 Summary. In Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on (pp. 675-678)

[2] Davison, A., Merghani, W., & Yap, M. (2018). Objective classes for micro-facial expression recognition. Journal of Imaging, 4(10), 119.

[3] Li, X., Pfister, T., Huang, X., Zhao, G., & Pietikainen, M. (2013). A Spontaneous Micro-expression Database: Inducement, collection and baseline. In IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition (FG) (pp. 22-26).

[4] Yan, W. J., Li, X., Wang, S. J., Zhao, G., Liu, Y. J., Chen, Y. H., & Fu, X. (2014). CASME II: An improved spontaneous micro-expression database and the baseline evaluation. PloS one, 9(1), e86041.

[5] Davison, A. K., Lansley, C., Costen, N., Tan, K., & Yap, M. H. (2016). SAMM: A spontaneous micro-facial movement dataset. IEEE Transactions on Affective Computing.

[6] Le Ngo, A. C., Phan, R. C. W., & See, J. (2014). Spontaneous subtle expression recognition: Imbalanced databases and solutions. In Asian Conference on Computer Vision (pp. 33-48).

[7] Forman, G., & Scholz, M. (2010). Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. ACM SIGKDD Explorations Newsletter, 12(1), 49-57.