Facial Micro-Expression Analysis – A Computer Vision Challenge

V. Challenges & Future Avenues

JOHN SEE  Multimedia University, Malaysia
ANH CAT LE NGO  TrustingSocial
SZE-TENG LIONG  Feng Chia University, Taiwan
So, we decided that we should meet up and have a “real-world” look at each other’s expressions...
5 “Objective classes” (grouped by Facial AU) instead of emotion classes

Cross-database protocols

- **Holdout Database Evaluation (HDE)**
  - Train on one dataset, Test on the other. Swap, repeat. (WAR, UAR)

- **Composite Database Evaluation (CDE)**
  - Combine both datasets, evaluate by LOSO (F1-score)
UAR results are very close
6 papers accepted (50%) – 3 challenge, 3 non-challenge
Enriched Long-term Recurrent Convolutional Network (ELRCN)

- 2 ways of enriching a CNN-LSTM pairing
  - **Spatially**: Gray, OF, OS images stacked along CNN channel, feats. passed to LSTM
  - **Temporally**: Separate CNN streams for Gray, OF and OS images, late fusion after FC, feats passed to LSTM

Enriched Long-term Recurrent Convolutional Network (ELRCN)

Transfer learning of macro-trained deep models

- Train deep models on macro-expression apex samples ➔ Transfer learning on micro-expression apex samples
  - **ResNet10** pre-trained on 4 macro-exp. datasets using apex frames
    - CK+ (852 images)
    - Oulu CASIA NIR & VIS (1200 images)
    - Jaffe (151 images)
    - MUGFE (8228 images)
    - TOTAL: 10,431 images ➔ oversample to 5,000 images/expression
  - Fine-tuning on micro-exp datasets using apex frames ➔ oversample to 200 images/expression
  - Assumption: That apex information is available!

Insights:

- Cross-database task is challenging
  - Leveraging macro-expression samples seem to work reasonably well
  - Lack of data -> LSTMs not suitable

- Efforts underway to create a new large-scale database

- We need more people to work on this area!
Subjectivity in humans

- Certain emotions (e.g. happiness) are easier to elicit compared to others (e.g. fear, sadness, anger)
- Some people are more “poker-faced” than others – they hide their emotions well!

Sample distribution

- Bias learning → Imbalanced distribution of samples per emotion, samples per subject

Creative strategies for inducement

- Complementary info from body region\(^1\), or heart rate from skin variations\(^2\)

---

\(^1\) Song et al. (2013). *Learning a sparse codebook of facial and body micro expressions for emotion recognition,* Proc of ACM Int. Conf. Multimodal Interaction

\(^2\) Gupta et al. (2018). *Exploring the feasibility of face video based instantaneous heart-rate for micro-expression spotting,* CVPR Workshops.
Subject diversity

- Most datasets contain a majority of subjects from one particular country or ethnicity

Environment and setting

- Real-world scenarios are much needed: Job interviews, criminal interrogation, patient assessment etc. (but many cannot pass ethic committees!)
- How about “two truths and a lie” game?
Challenges

SPOTTING

Landmark detection
- Room for improvement in existing methods. ME requires very stable detection / robust against noise to capture minute changes in facial muscles.

Threshold or classify?
- Most existing works employ rule-based strategies ➞ Not robust and adaptable!
- Per-frame classification of ME occurrence ➞ Rigid and noisy!

Onset and offset detection
- Current works do not consider detecting the start and end frames, which could be useful to trim ME sequences before classification
Block Selection

• Block-based methods of extracting features are quite popular

• Assignment of weights to blocks with key information ➔ New: Learn which blocks are discriminative

Eyes: To Keep or Not To Keep?

• Some\textsuperscript{4} works mask out eye regions to avoid eye blink motions, some\textsuperscript{5} think otherwise

\textsuperscript{3} Zong et al. (2018). Learning from hierarchical spatiotemporal descriptors for micro-expression recognition. IEEE T-MM


\textsuperscript{5} Duan et al. (2016). Recognizing spontaneous micro-expression from eye region,” Neurocomputing.
**Feature crafting / learning**
- Most crafted features circa 2014-2016 are still holding strong results – shape, motion
- DL getting popular 2016 onwards – pushing the limits

**Cross-DB recognition**
- Realistic setting (multi-environment)
- How to generalise across domains\(^6\)?
- MEGC\(^7\) leads this effort

---


Evaluation Protocol

- Use LOSO cross-validation\(^8\) instead of LOVO cross-validation (some works still do this! 😞)
  - LOVO exposes the training to samples belonging to the test sample subject

Performance Metrics

- Use F\(_1\)-score instead of Accuracy
  - Accuracy tends to be bias in imbalanced datasets or heavily skewed data
  - Use unweighted metrics that give equal emphasis to rare classes

Class Labels

- A few works consider fewer number of classes than it should be ➔ problem benchmarking!
- Emotion classes vs. Objective classes\(^9\)

---

\(^8\) Le Ngo et al. (2014). *Spontaneous subtle expression recognition: imbalanced databases and solutions*. ACCV

2nd MEGC @ IEEE FG 2019

14th IEEE International Conference on Automatic Face and Gesture Recognition

FG2019


2 Challenges:

- Cross-DB Recognition
- Spotting

Submission of papers: Challenge and non-challenge papers

Important Dates

- Submission Deadline: 27 January 2019
- Notification: 12 February 2019
- Camera-Ready: 15 February 2019
Many thanks to the our sponsors:

- Multimedia University, Malaysia
- Ministry of Higher Education, Fundamental Research Grant Scheme
- Belt and Road Initiative Research Scholar Exchange

our team throughout the years (2014-Current):

Anh Cat Le Ngo  Yandan Wang  Raphael C.W. Phan  Sze-Teng Liong  Yee-Hui Oh  Huai-Qian Khor

and collaborators:

Moi-Hoon Yap
Manchester Metro. Uni

Xiaopeng Hong
Univ of Oulu, Finland

Su-Jing Wang
Institute of Psychology, CAS
Thank you