Tutorial

ACCV 2020

Recent Advances and Challenges in Facial Micro-Expression Analysis

Datasets

John See Multimedia University, Malaysia

Moi-Hoon Yap Manchester Metropolitan University, UK

Su-Jing Wang Chinese Academy of Sciences, China

Jingting Li Chinese Academy of Sciences, China

Sze-Teng Liong Feng Chia University, Taiwan

Outline

How ME datasets are collected

- Inducement & Elicitation
- General Considerations

Datasets

- SMIC
- CASME / CASME II
- SAMM
- CAS(ME)^2
- MEVIEW

Inducement & Elicitation of MEs

- **Posed vs. Spontaneous**
  - Voluntary: MEs can either be acted out (upon instructions)
  - Involuntary: MEs can be induced spontaneously using appropriate stimuli that has high ecological validity[†] e.g. emotional video clips
  - **Spontaneity is crucial to mimicking real-world scenarios**
- **Expected issues during elicitation**
  - Not all participants will exhibit micro-expressions when induced (extremely poker-face!)
  - Not all emotions are necessarily induced from a participant (some people naturally do not look happy or able to show happiness easily)

[†] A measure of how test performances predicts behaviours in real-world settings

General Considerations when Collecting Data

- Careful selection of stimuli (the videos)
- Participants are asked to keep a “poker face” and not reveal their true emotions while watching the video clips
- While the participants watch the video with their faces recorded, a “moderator” observes the participants’ facial and body language to ensure the participant does not show his emotions clearly
  - A penalty (fill in long boring questionnaires) or incentive (some money) is used to get participants’ compliance

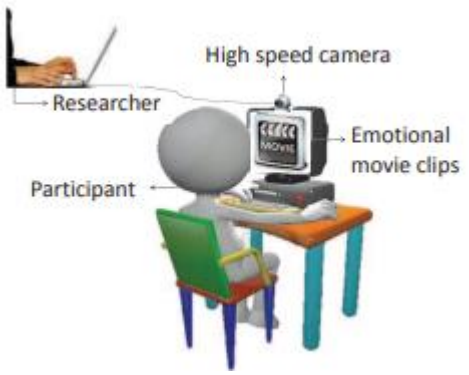
SMIC

Li et al. (2013)

- **Probably** the first spontaneous facial micro-expression dataset
- Collected 3 subsets from various modalities
  - HS: 100fps high speed camera
  - VIS: 25fps normal camera
  - NIR: 25fps near infrared camera
- Final dataset: 164 clips from 16 participants (mean age: 28.1 years, 6 females, 10 males, 8 Caucasians, 8 Asians), 640x480 pixels
- Emotions labelled following self-report from participants

Dataset	Experiment	Method	Accuracy
HS	Detection	LBP-TOP($5 \times 5 \times 1$) + TIM10	65.55 %
	Recognition	LBP-TOP($8 \times 8 \times 1$) + TIM10	48.78 %
VIS	Detection	LBP-TOP($5 \times 5 \times 1$)	62.68 %
	Recognition	LBP-TOP($5 \times 5 \times 1$) + TIM10	52.11 %
NIR	Detection	LBP-TOP($8 \times 8 \times 1$) + TIM20	59.15 %
	Recognition	LBP-TOP($8 \times 8 \times 1$) + TIM10	38.03 %

Data	Participants	Micro-expression clips			
		Po.	Ne.	Sur.	Total
HS	**16/20**	51	70	43	**164**
VIS	**8/10**	28	23	20	**71**
NIR	**8/10**	28	23	20	**71**



CASME

Yan et al. (2013)

- Collected from a 60 fps camera @ 640x480 (A) 1280x720 pixels (B)
- Final dataset: 195 clips from 35 participants (mean age: 22.03 years, 13 females, 22 males, All Asians)
- Action Units (AU) annotations
- Emotions labelled based on participant's self-report
- Penalty (5 CNY) imposed if participants fail to control their emotions

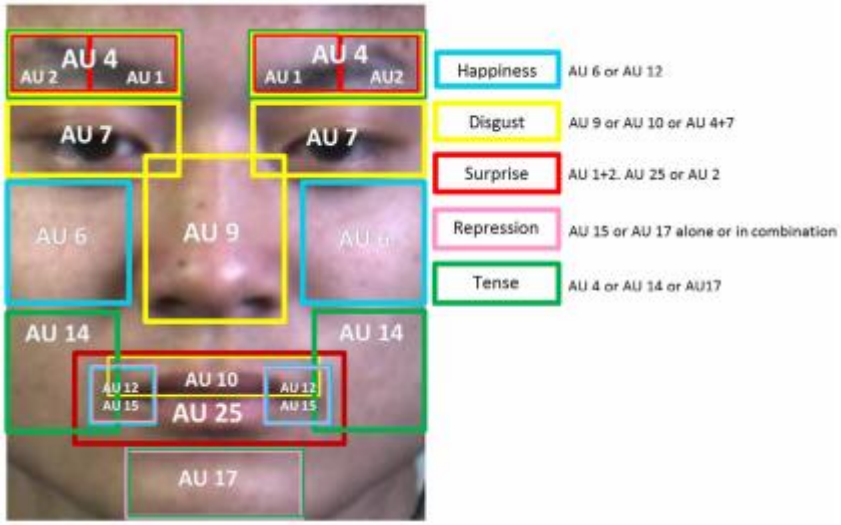
CRITERIA FOR LABELING THE EMOTIONS AND THE FREQUENCY IN THE DATABASE*.

Emotion	Criteria	N
Amusement	Either AU6 or AU12 must be present	5
Sadness	AU1 must be present	6
Disgust	At least one of AU9, AU10, AU4 must be present	88
Surprise	AU1+2, AU25 or AU2 must be present	20
Contempt	Either unilateral AU10 or unilateral AU12 be present	3
Fear	Either AU1+2+4 or AU20 must be present	2
Repression	AU14, AU15 or AU17 is presented alone or in combination	40
Tense	Other emotion-related facial movements	28

*The emotion labeling are just partly based on the AUs because micro-expressions are usually partial and in low intensity. Therefore, we also take account of participants' self-report and the content of the video episodes.

FACS Action Units (AU)

- **Facial Action Coding System (FACS)** – categorises human facial movements by their appearance and muscle components, known as **Action Units (AU)**



- FACS AUs (single or in combination) can be used to identify expressions, smiles, or other facial behavior

CASME II

Yan et al. (2014)

- Collected from a high-speed 200 fps camera, 640x480 res, face res of 280x340 pixels
- Final dataset: 247 clips labelled with Action Units (AU) and emotion classes
- Emotions are labelled based on **both** AUs and participant's self-report
- Baseline: LBP-TOP, 5x5 blocks

SAMM

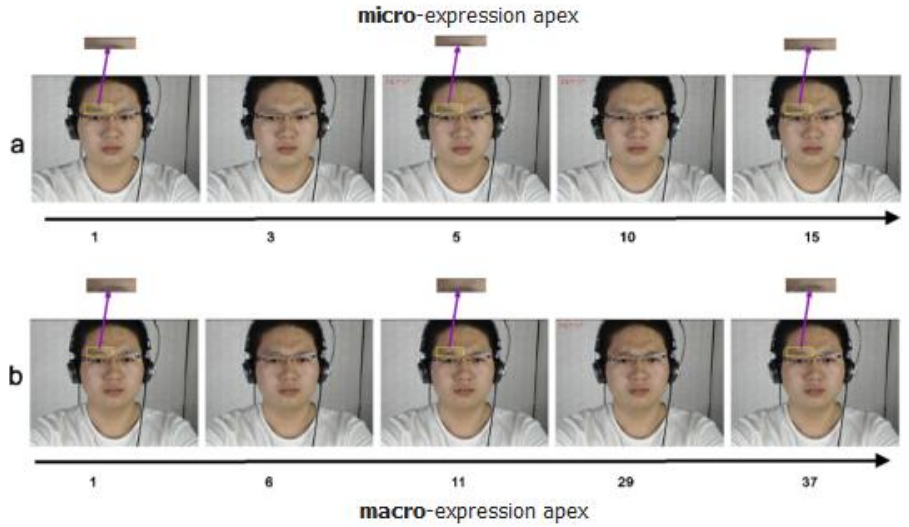
Davison et al. (2016)

- Collected from a high-speed 200 fps camera @ 2048x1088 pixel resolution
- Final dataset: 159 clips from 30 people (60% whites, 40% other ethnicities), 14 males, 16 females with mean age of 34.48 years old
- FACS coded by 3 coders
- Prize of 50 GBP was offered to participants who successfully suppress their emotions while watching the videos

CAS(ME)^2

Qu et al. (2017)

- First to contain both macro- and micro-expressions collected from same participants
- Samples:
  - 250 macro-expression samples (0.5s – 4s),
  - 53 micro-expression samples (<0.5s)
- Each sample is AU-annotated
- Samples collected from 30 fps webcam, with resolution of 640x480 pixels, from 22 participants (16 females, the rest male) with mean age 22.59 years.



MEVIEW

Husak et al. (2017)

- First to provide unconstrained "in the wild" situations and naturalistic high-stakes scenarios
- Consists of mostly YouTube-collected poker games and interviews
- An METT trained annotator labelled the onset and offset frames of the ME, with FACS coding and emotion types (some samples had no labels)
- Only 31 samples collected at 25 fps, from 16 individuals

Micro-Expression Examples

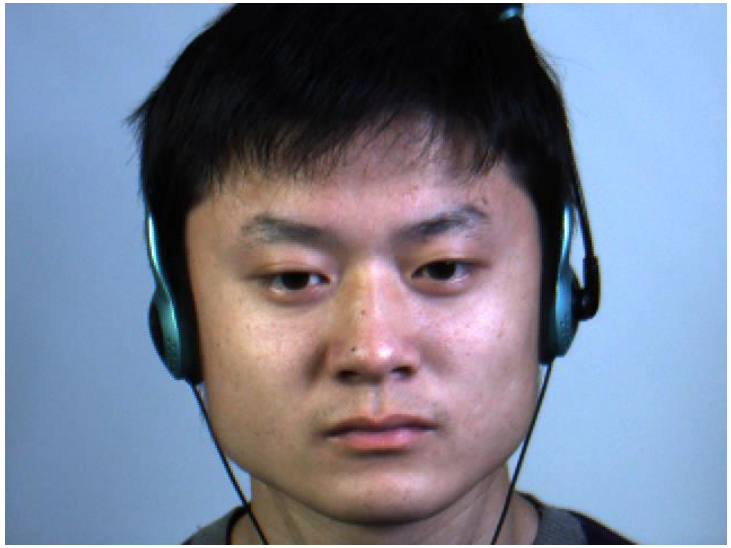
SMIC

"Positive"

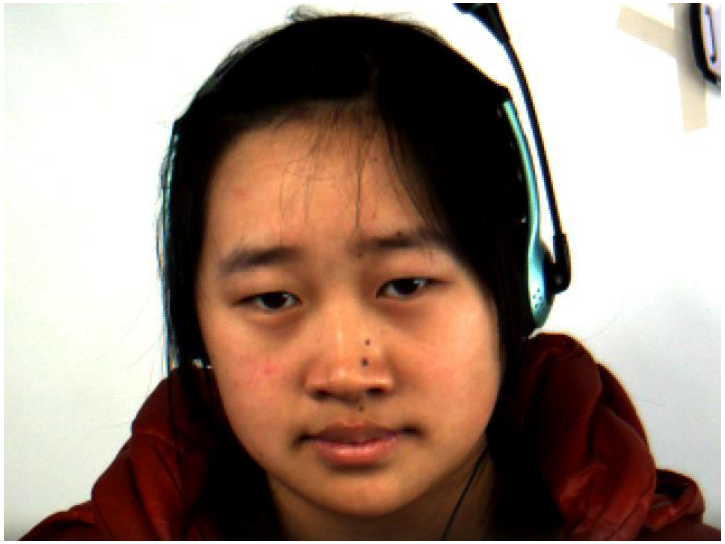
"Negative"

Micro-Expression Samples

CASME II



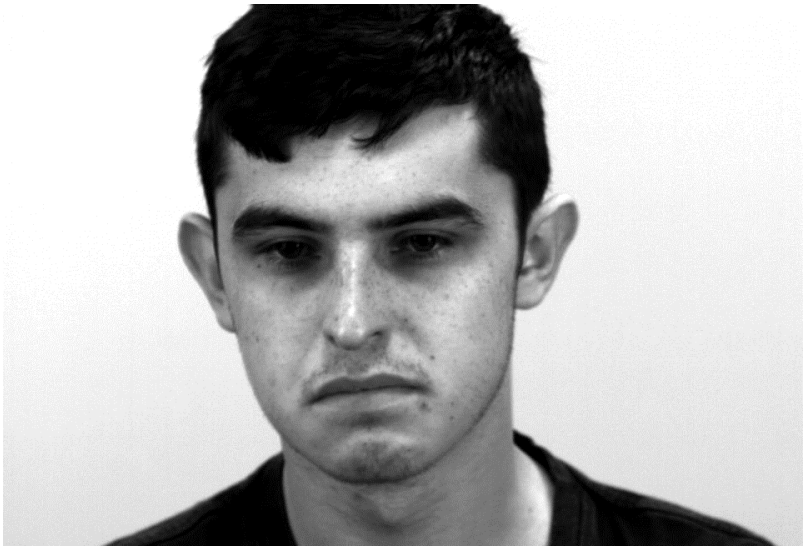
"Happiness"



"Repression"

Micro-Expression Samples

SAMM



"Anger"

"Surprise"

Micro-Expression Samples

MEVIEW

"Contempt"

"Surprise"

Less Is More: Micro-Expression Recognition from Video using Apex Frame

Signal Processing: Image Communication, 2018

Moi Hoon Yap, Adrian Davison, Chuin Hong Yap, Clive Lansley, Connah Kendrick

Manchester Metropolitan University

Less Is More: Micro-Expression Recognition from Video using Apex Frame

Signal Processing: Image Communication, 2018

Sze-Teng Liong, **John See**, KokSheik Wong, Raphael C.W. Phan

Manchester Metropolitan University

End of Part 2

Questions?