The Use of AI for Thermal Emotion Recognition

: A review of problems and limitations in standard design and data

AAAI 2020 Sep 22

안녕
10/28/2020

What you should know :-D

- for Facial Emotion Recognition (FER) :

Thermal imagery > RGB

Figure 9: The Tufts Face Database (Panetta et al., 2018)

What you should know :-D

Figure 1: RGB, near infrared and thermal images of a resting (up) and fatigued (down) face. In the thermal images, darker pixels corresponds to colder and lighter to hotter. (Lopez, del Blanco, and Garcia, 2017)

Contents

Ⅰ. Advantages of Thermal over Visible

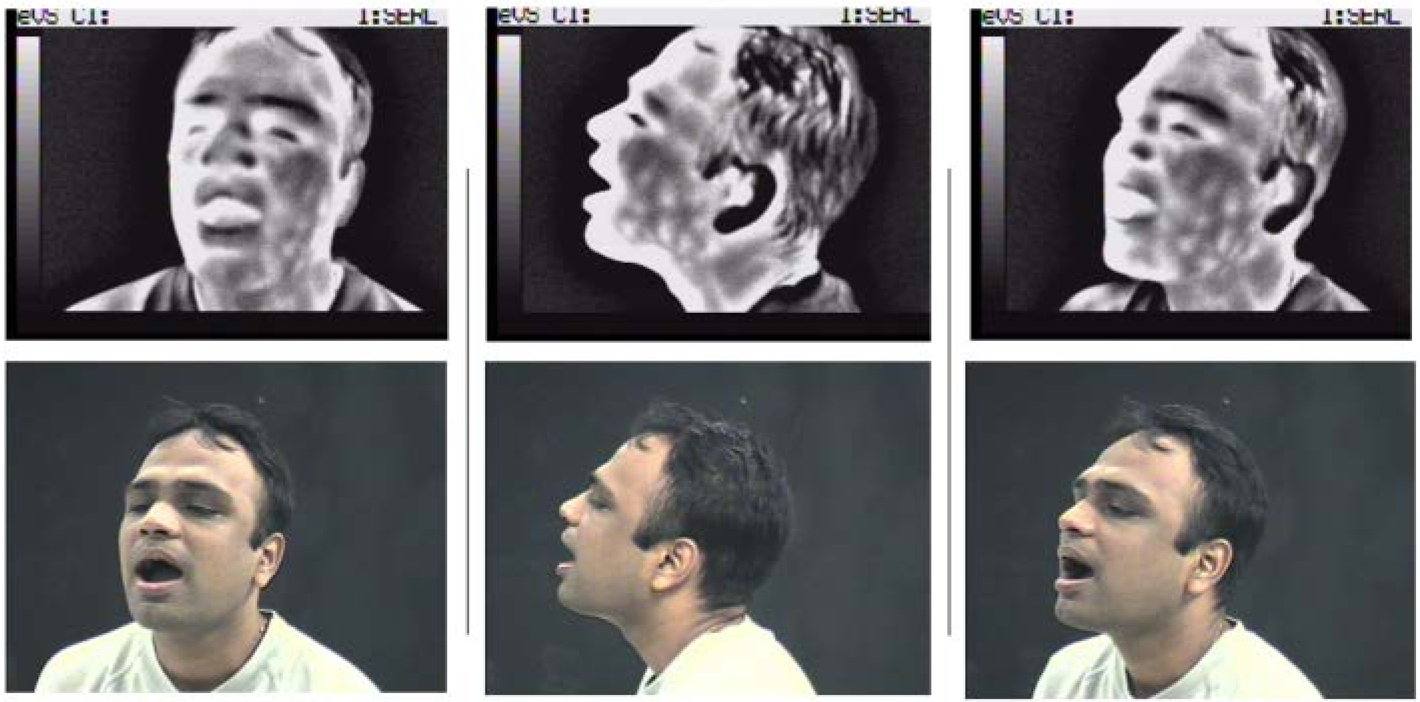


Figure 2: Example of data from the Iris dataset (Hammoud)

Ⅰ. Advantages of Thermal over Visible

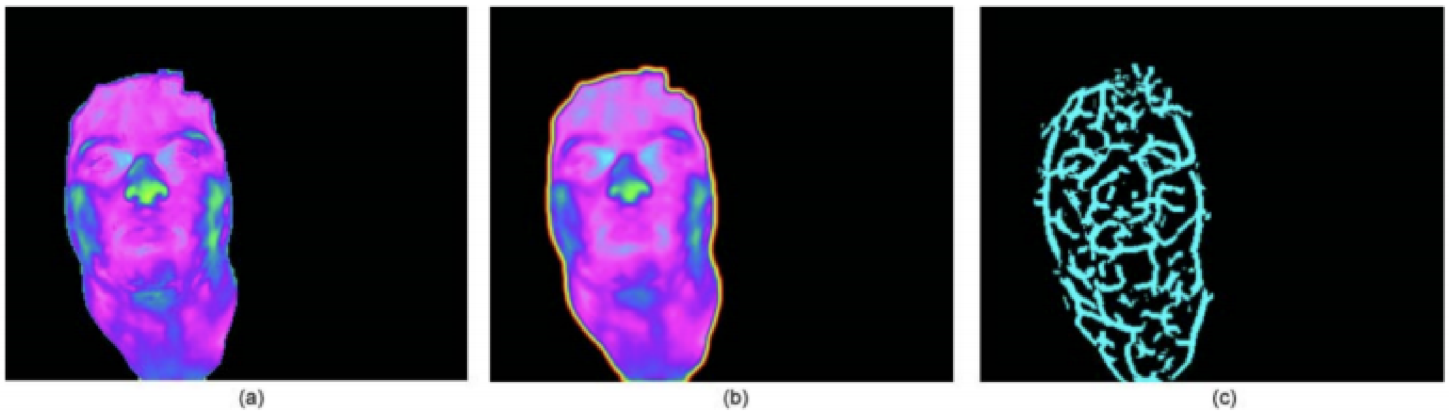


Figure 3: Vascular network extraction: (a) Original segmented image; (b) Anisotropically diffused image; (c) Blood vessels extracted using white top hat segmentation, per (Buddharaju et al., 2007)

Ⅰ. Advantages of Thermal over Visible

Figure 4: Digitization privacy in different scenes: digitization results in scenes with people, computers and buildings. The left column are the input 16 bit images and the right column is the simulated output. (Pittaluga, Zivkovic, and Koppal, 2016)

Ⅰ. Advantages of Thermal over Visible

1. Invariant to lighting conditions
2. Reliable and accurate correlation to standard physiological measures
3. Non-invasive
4. Resistant to intentional deceit
5. Able to reveal facial disguises (i.e. wigs, masks)

Ⅱ. Physiology and Thermal FER

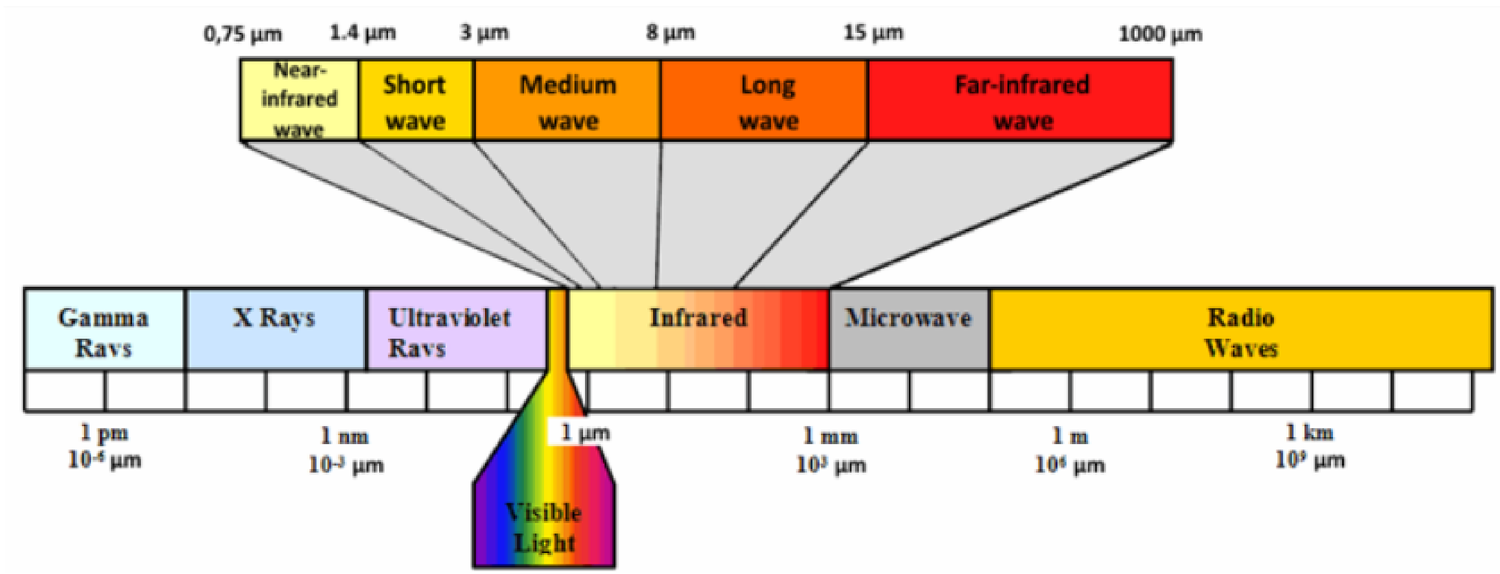


Figure 5: Long-Wave IR falls in the wavelength range of 8 µm to 15 µm

Ⅱ. Physiology and Thermal FER

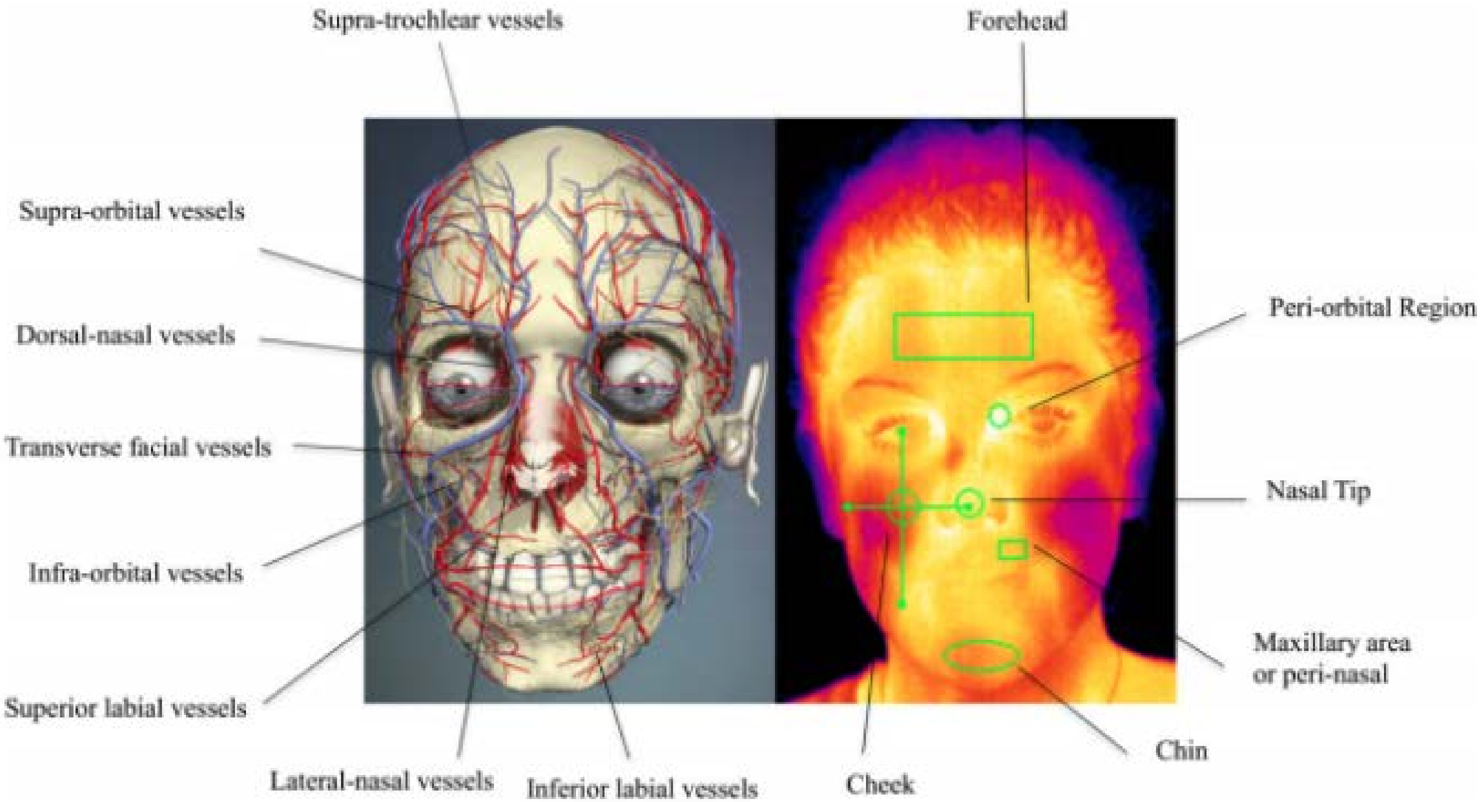


Figure 6: Thermal representation for extraction of ROIs by Ioannou

Ⅱ. Physiology and Thermal FER

	Emotions									
	Stress	Fear	Embarra-ssment	Startle	Sexual arousal	Anxiety	Joy	Pain	Guilt	Displea-sure (exercise)
Regions										
Nose	↓	↓			↑		↓		↓	
Cheeks				↓						↑
Periorbital				↑	↑	↑				
Supraorbital				↑		↑				
Forehead	↓↑	↓			↑	↑		↓		↑
Maxillary	↓	↓		↓				↓	↓	
Neck–carotid				↑						
Nose	↓									
Tail		↓						↓		
Fingers/palm		↓						↓		
Lips/mouth			↑		↑					

Figure 7: Skin thermal variations in the considered regions of interest across emotions

Ⅲ. AI and Thermal FER

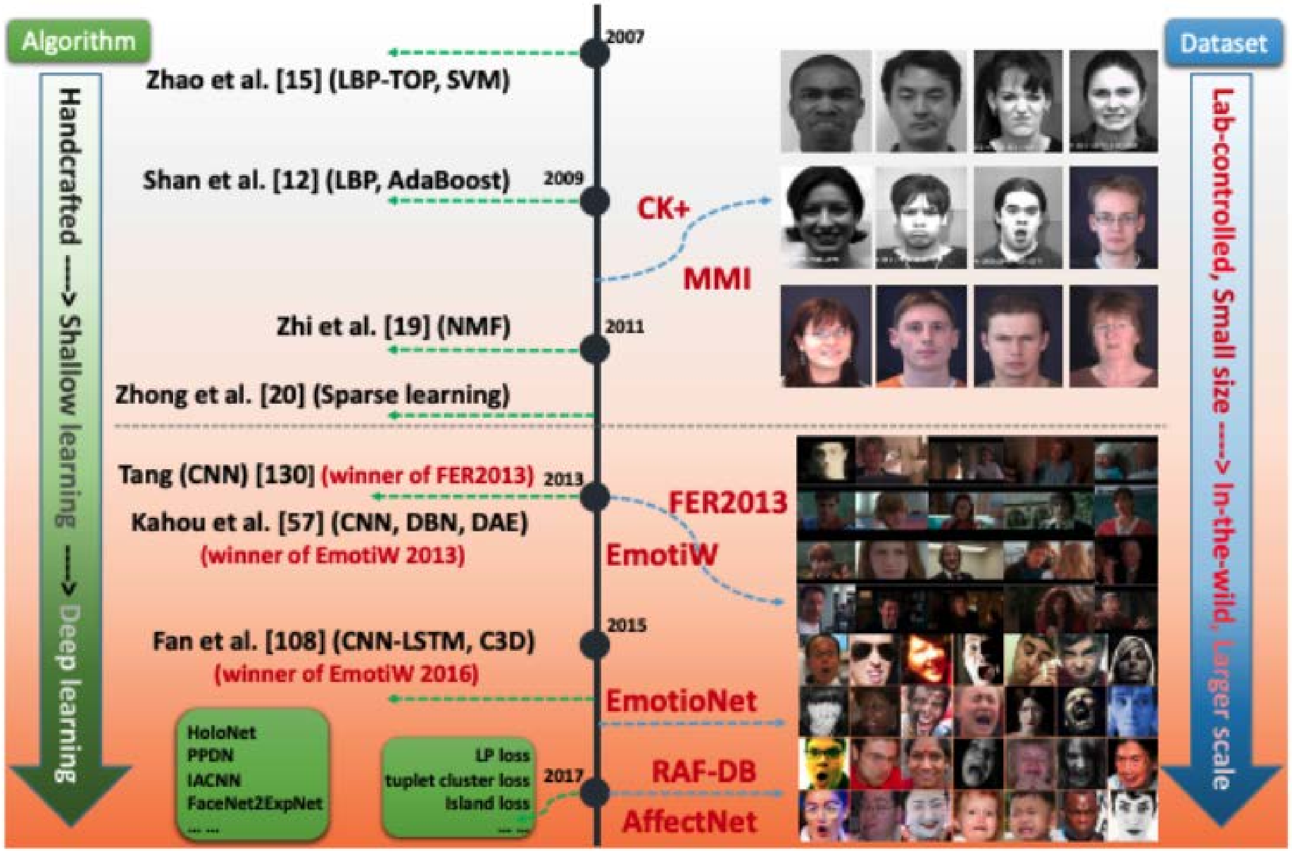


Figure 8: Growth of lab-controlled, small size data to "in-the-wild", larger scale data encouraged use of deep learning algorithms in visible FER (Li and Deng, 2018)

Ⅲ. AI and Thermal FER

Table 2: Selected Thermal Facial Emotion Recognition AI Papers

Author	Year	Affect	ROIs	Model	Dataset	Target	Acc	Data	Code	Params
Stemberger	2010	Cognitive Workload	7 ROIs	ANN	Custom dataset	Multiple Workload	81.0%	(-)	(-)	(+)
Wang	2014	Spont. Affect	Whole face	DBM	USTC-NVIE	Valence	62.9%	(+)	(-)	(+)
Wu	2016	Posed Affect	Whole face	CNN	RGB-D-T	Multiple Affects	99.40%	(-)	(-)	(-)
Simon	2016	Posed Affect	Whole face	CNN	RGB-D-T	Multiple Affects	UNK	(-)	(-)	(+)
Cho	2017	Stress	Nose	CNN	Custom dataset	Binary Stress	85.59%	(-)	(-)	(+)
Lopez	2017	Exercise Fatigue	Whole face, 3 ROIs	CNN, SVM	Custom dataset	Binary Fatigue	23.3% - 81.8%	(+)	(-)	(+)
Haque	2018	Pain	Whole face	CNN, LSTM	Custom dataset	5 Pain Levels	18.33% CNN	(+)	(-)	(+)
Ilyas	2018	Spont. Affect	Whole face	CNN, LSTM	Custom dataset	Multiple Affects	89.74%	(-)	(-)	(-)
Elbarawy	2019	Posed Affect	Whole face	CNN	Iris	Multiple Affects	96.7%	(+)	(-)	(+)
Ilikci	2019	Posed Affect	Whole face	CNN	Iris	Multiple Affects	92.72%	(+)	(-)	(+)
Shreyas Kamath	2019	Posed Affect	Whole face	CNN	Tufts Face Database	Multiple Affects	96.2%	(+)	(-)	(+)

Year - Publication year, Affect - Expression type (Posed and Spont. mean basic discrete emotions), ROIs - facial regions of interest, Model - Deep learning algorithm type, Dataset - name of database, Target - the predicted class (all papers identified were classification), Acc - Best classification accuracy across models reported. Data - link to database provided if custom or name of public database provided, Code - link to code provided, Params - model parameters disclosed in paper, Annotations of (-) indicate information not disclosed, and (+) means it was disclosed in the paper.

Ⅲ. AI and Thermal FER

Table 1: Thermal Facial Emotion Recognition Datasets.

Dataset	Year	Pose	Pairs	Affect	Subj	Access	Seq	Multi	THR	VIS
Univ. Notre Dame (UND)	2002	Spont.	Yes	UNK	241	R	UNK	Yes	LWIR	Yes
Equinox (Equinox; Heo et al., 2004)	2004	Posed	UNK	3	90	N/A	No	No	MW, LWIR	Yes
IIT Delhi (Kumar)	2007	Posed	UNK	UNK	108	R	No	UNK	NIR	No
Univ. Houston (Buddharaju et al., 2007)	2007	Both	Yes	0	138	UNK	UNK	No	MWIR	Yes
SC-Face (Grgic)	2009	None	Yes	0	130	R	No	No	NIR	Yes
USTC-NVIE (Wang et al., 2010)	2010	Both	UNK	6	100	N/A	Yes	No	LWIR	Yes
Zhang (Zhang et al., 2010)	2010	Posed	UNK	0	350	R	No	UNK	NIR	No
UCHThermalFace (Hermosilla et al., 2012)	2012	Posed	No	3	102	UNK	Yes	No	LWIR	UNK
KTFE Database (Nguyen et al., 2013)	2013	Spont.	Yes	7	26	UNK	Yes	No	LWIR	Yes
Iris (Hammoud)	2013	Posed	Yes	3	30	P	No	No	LWIR	Yes
RGB-D-T (Simón et al., 2016)	2016	Posed	Yes	5	51	UNK	UNK	UNK	LWIR	Yes
VIS-TH (Eurecom) (Mallat and Dugelay, 2018)	2018	Posed	Yes	4	50	R	Yes	Yes	LWIR	Yes
RWTH Aachen Univ. (Kopaczka, Kolk, and Merhof, 2018)	2018	Posed	No	8	90	R	Yes	UNK	LWIR	No
Tufts Face Database (Panetta et al., 2018)	2018	Posed	Yes	5	113	R	Yes	No	NIR, LWIR	Yes
UL-FMTV (Ghiass et al., 2014)	2018	Posed	Yes	UNK	238	R	Yes	Yes	N, MW, LWIR	No
ThermalWorld (Kniaz et al., 2018)	2019	Spont.	Yes	0	516	R	No	No	LWIR	Yes
RFLDDJ (Seo and Chung, 2019)	2019	UNK	Yes	UNK	UNK	P	UNK	No	LWIR	Yes

Dataset - Database name, Year - publication year, Pose - Posed, Spontaneous, or Both, Pairs - Visible and Thermal, Affect - Number of labeled expressions, Subj - Number of unique human subjects, Access - R (requires permission from authors), P (publicly downloadable), Seq - Yes or No for availability in dataset of video sequences, Multi- Yes or No for multi-session recording, THR - Thermal image modality, VIS - Yes or No for presence of visible images, UNK means information was not provided in the paper.

Ⅲ. AI and Thermal FER

Table 3: Examples of Thermal FER Experimental Design Parameters

Author	Year	Thermal Cam.	Dual Sensor	Thermal Res.	Dem.	Exclusion	Subjects	Temp.	Rest Time	Lighting	Stimulus
Nhan	2010	ThermaCAM	UNK	UNK	9F, 3M, mean 24 yo	UNK	12	UNK	20 min	UNK	Static images
Wang	2010	SAT-HY6850	UNK	320 x 240	58F, 157M, 17 - 31 yo	UNK	215	Means 23.29	UNK	Yes	Emotional videos
Hermosilla	2012	Flir 320 TAU	UNK	324 x 256	UNK	UNK	102	UNK	UNK	UNK	UNK
Nguyen	2013	NEC R300	Yes	UNK	UNK gender, 11 - 32 yo	UNK	26	24 - 26	2 hrs.	UNK	Emotional videos
Salazar-Lopez	2015	ThermoVision A320G	UNK	UNK	60F, 60M, 24 - 27 yo	Yes	120	18 - 25	10 - 15 min.	UNK	Static images
Lopez	2017	Therm-App	UNK	288 x 384	8F, 11M, 23 - 27yo	UNK	19	UNK	Until heart rate below 20 bpm	UNK	Exercise
Mallat	2018	Flir Duo R	Yes	160 x 120	No	UNK	50	25	No	Yes	UNK
Goulart	2019	Therm-App	UNK	384 x 288	8F, 9M, 8 - 12 yo	UNK	17	20 - 24	10 min.	Yes	Questionairre
Sonkusare	2019	Flir A615	UNK	640 x 480	11F, 9 M, 22 - 30 yo	Yes	20	22	No alcohol & caffeine 2 hrs. prior	Yes	Auditory stimulus
Panetta	2020	FLIR Vue Pro	UNK	UNK	UNK	UNK	113	UNK	UNK	Yes	UNK

Year - Publication year, Thermal Cam. - Type of LWIR camera, Dual Sensor - Yes or No, captures visible and thermal simultaneously, Thermal Res. - Reported thermal pixel resolution, Dem. - Demographics of subjects, Exclusion - Yes or No, exclusion or inclusion criteria documented, Subjects - Number of unique human subjects, Temp. - Room temperature for experiment reported in degrees Celsius, Rest Time - Time subjects reach relaxed state prior to image capture, Lighting - Yes or No, illumination design documented, Stimulus - Type of stimulus to provoke spontaneous response, if spontaneous, UNK means information was not found in the paper.

IV. Thermal FER Data challenges

Table 4: Summary of Thermal FER Data Challenge

Challenge	**Consequence**	**Mitigation**	**Opportunities**
Include video sequences	Static images fail to capture the complete temporal dynamics of emotional response.	Including labeled videos in thermal FER dataset.	Spatio-temporal labeling of thermal onset, delay, duration of physiological response.
Enable spontaneous response	Discrete posed expressions may not invoke realistic physiological response.	Add spontaneous elicitation where possible, in addition to discrete set.	Natural, “in the wild” expressions that offer accurate representations of emotion.
Provide social or personal context	Thermal data collected without social stimuli may not be useable for social use cases.	If appropriate, label social context or if controlling for, document how social response has been minimized.	Social interaction thermal FER expressions, with labeled context and scenarios.
Collect multimodal pairs	No opportunity to increase accuracy or learn from additional modality mappings if only one modality (thermal) is collected.	May require dual sensor, or experimental design for simultaneous capture using two cameras.	Multimodal pairs for various social, spontaneous elicited thermal FER domains.
Document experimental setup	Confounding through uncontrolled environmental variables can lead to misleading images.	Report at minimum, the parameters shown in in Table 3.	Standard thermal FER experimental protocol for design and demographic documentation.
Accounting for Sensor Differences	Untested margin of error for images collected using different thermal sensors.	No mitigation strategy. This is an open research question.	Assessment with optical engineers to determine margin of error across sensors for human thermal FER.

IV. Thermal FER Data challenges

2. Enable spontaneous response

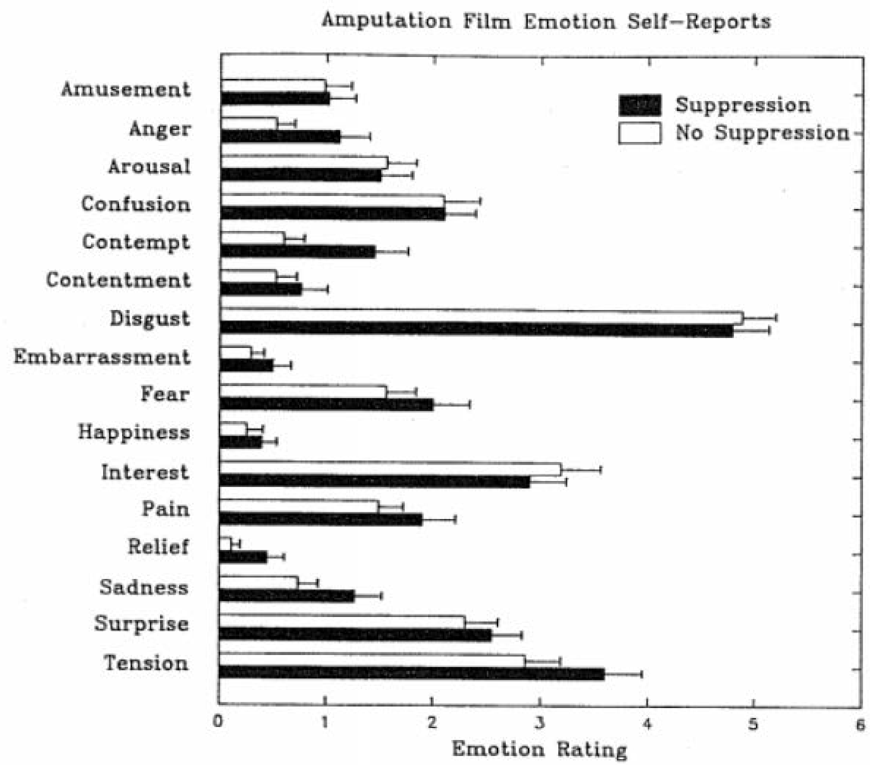


Figure 2. Emotion self-reports by condition for the amputation film, with standard errors of the mean.

Figure 10: Multiple feelings self-reported after exposure to high arousal video (Gross and Levenson, 1993)

IV. Thermal FER Data challenges

2. Enable spontaneous response

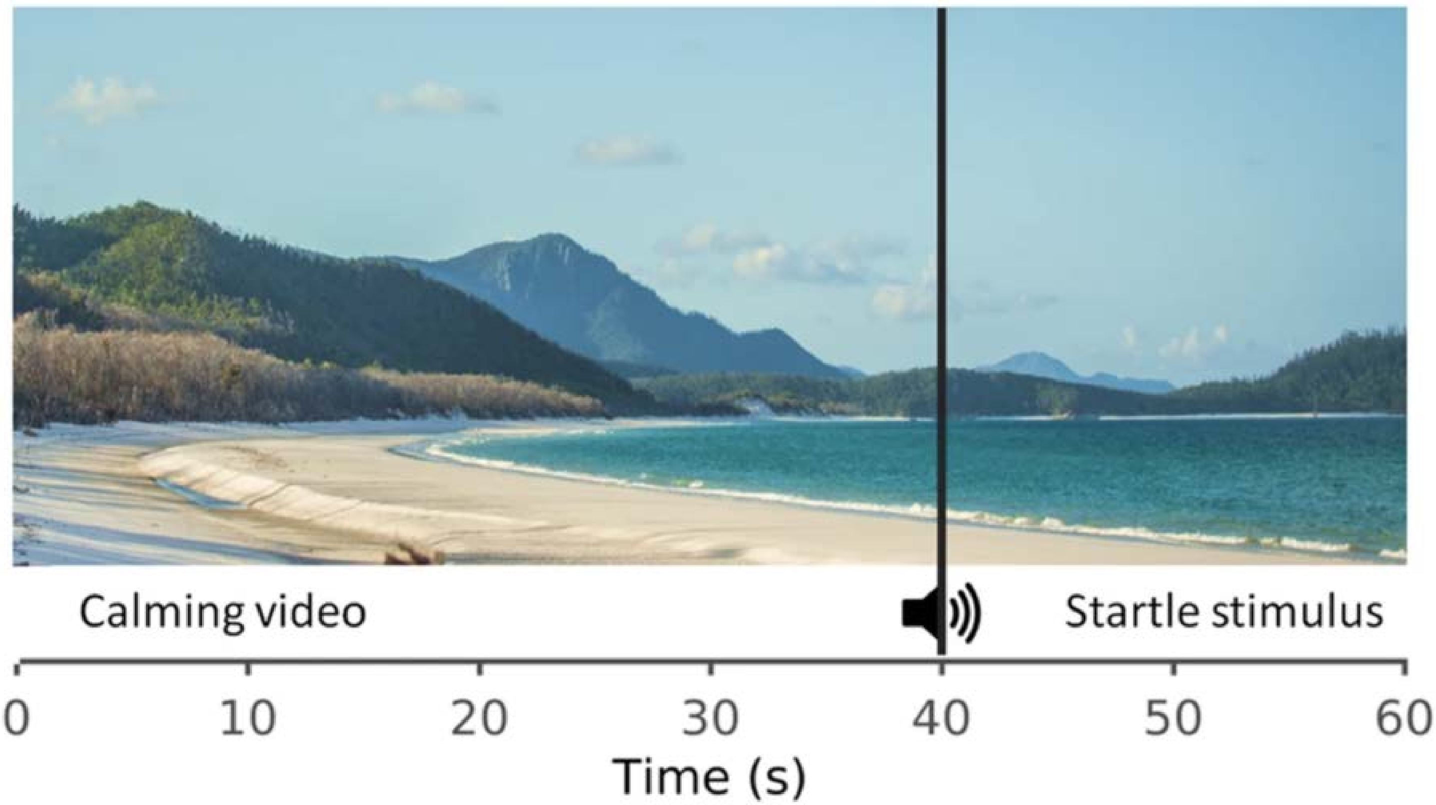


Figure 11: Example of an emotional stimulus by Sonkusare et al. to elicit a spontaneous response. A calming ocean video clip was played for 60 seconds. A loud gunshot sound (80dB) was played at 40seconds to mimic a startle response. (Sonkusare et al., 2019)

IV. Thermal FER Data challenges

3. Provide social or personal context

Figure 12: Experimental setup showing the child-robot interaction by Goulart et al. (2019) (a) Before showing the robot; (b) After presenting it.

IV. Thermal FER Data challenges

4. Collect multimodal pairs

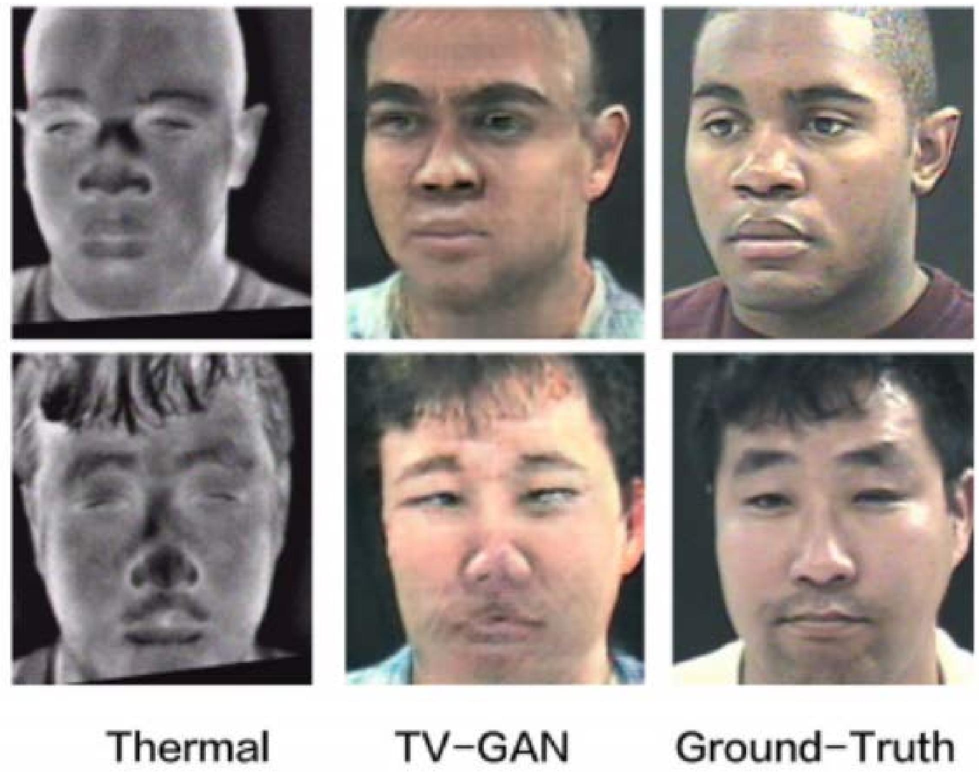
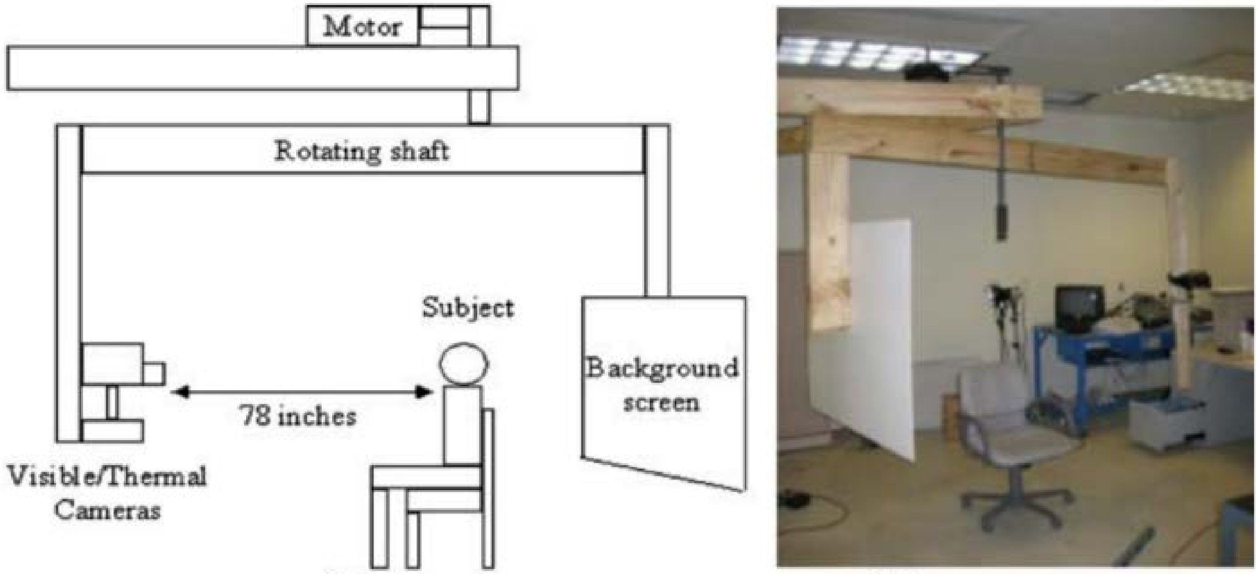


Figure 13: Example of TV-GAN trained on multimodal pairs for thermal-to-visible image translation (Zhang et al., 2018)

IV. Thermal FER Data challenges

5. Document experimental setup



(a) (b)

Figure 14: Experimental Setup for Iris dataset capture (Kong et al., 2007)

IV. Thermal FER Data challenges

5. Document experimental setup

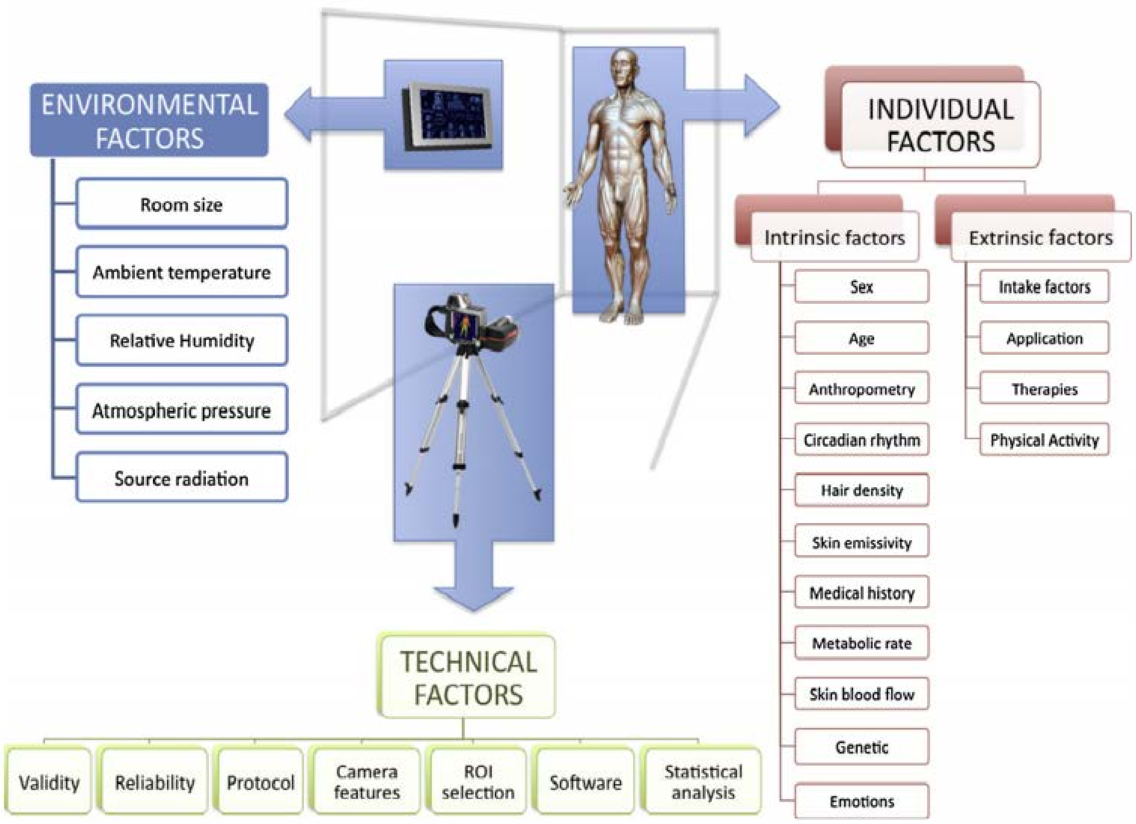


Figure 15: Factors influencing thermal imagery of humans (Fernández-Cuevas et al., 2015)

Recap ;-D

Why Thermal Imagery for FR?

- Psycho-physiological perspective
- Covid-19 pandemic
- Black lives matter
- Affective Computing

Recap ;-D

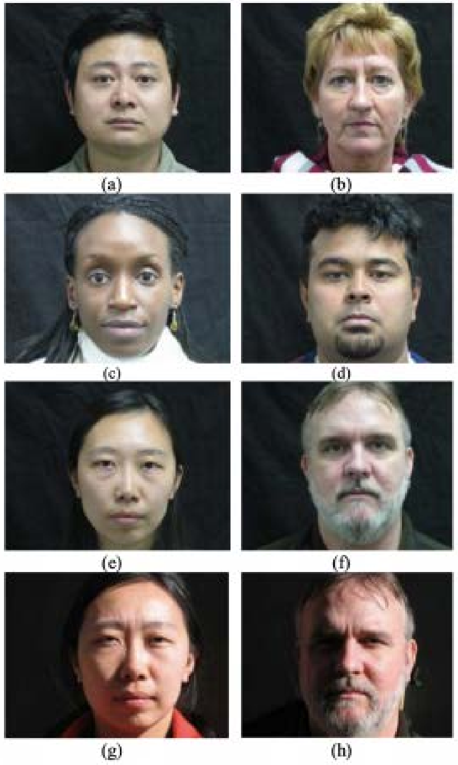


Figure 16: Participants from diverse multimodal dataset collected by the IRIS Lab in 2006 (Chang et al., 2006)

Recap ;-D

- Advantages of using thermal imagery over RGB for facial FER
  - Technincal benefits :
    Semi-anonymity
- Proposed challenges
- Insights on the consequences
- Mitigation
- Opportunities