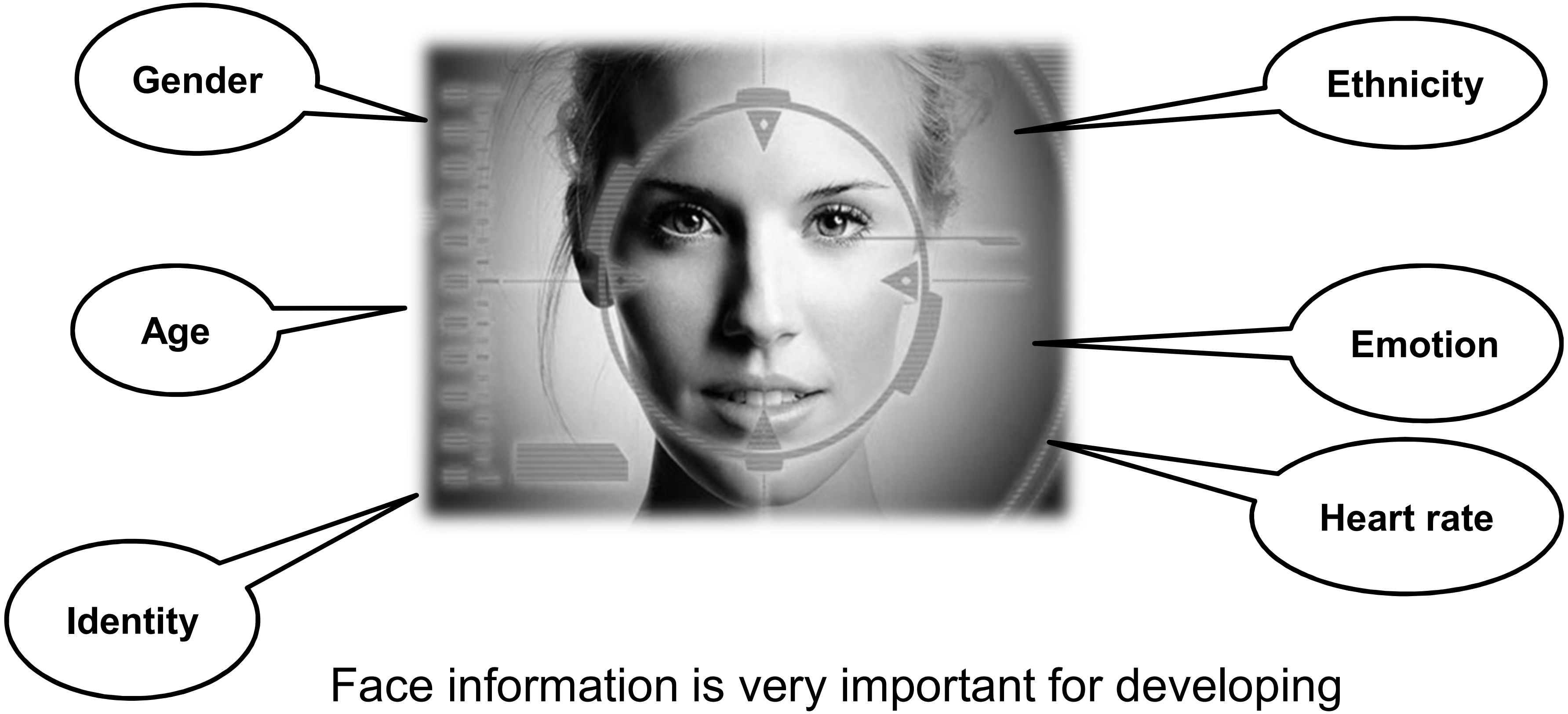
Emotional Interfaces with Face Analysis

Guoying Zhao
Academy Professor
University of Oulu, Finland
guoying.zhao@oulu.fi

The Face is special for us:



Face information is very important for developing **emotionally intelligent systems** to various applications

Possible applications

Emotional
Well beings

E-Teaching

Human-computer interaction

Security

User experience analysis

Job interview and
video conferences

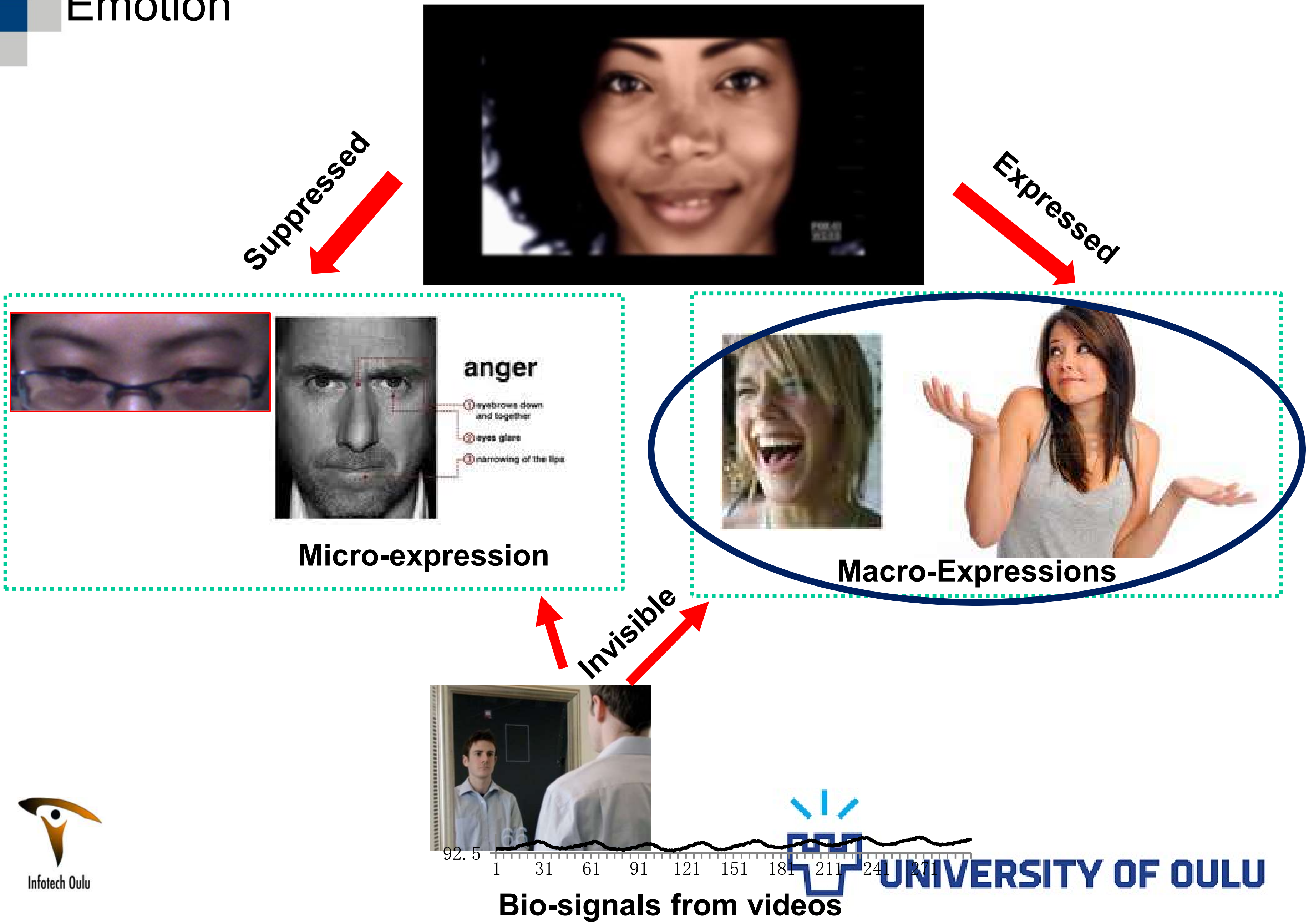
Safe driving

Health care

Infotech Oulu

UNIVERSITY OF OULU

Emotion



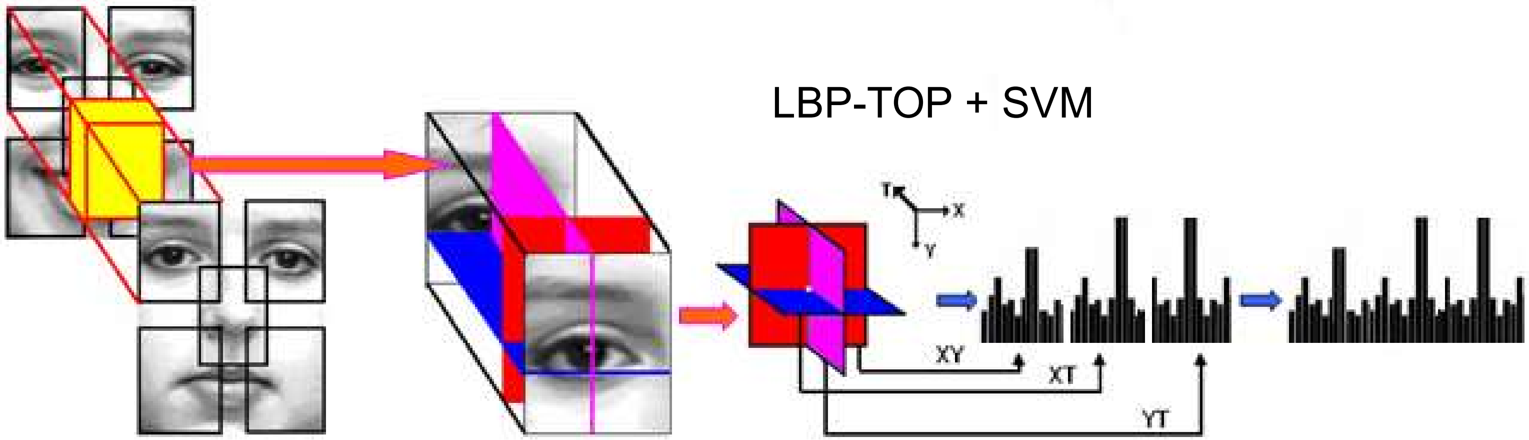
Infotech Oulu

UNIVERSITY OF OULU

Macro-expression: from static to dynamic

Input: **Images** **Videos**

LBP-TOP + SVM



G. Zhao & M. Pietikäinen (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(6):915-928.

Macro-expression: Deal with Challenges

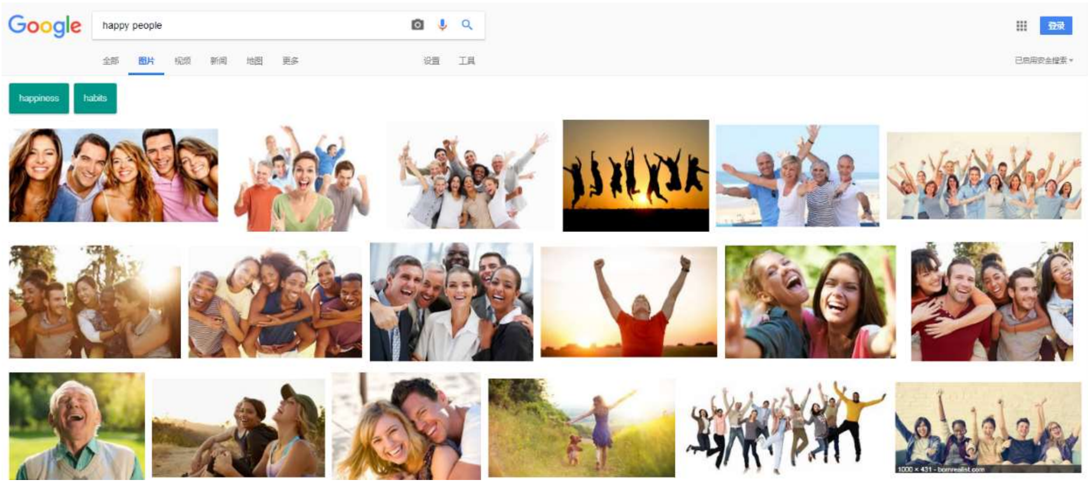
Illumination: NIR + LBP-TOP

Subtle-Expression:
Video magnification

View:
Multi-view canonical correlation

Occlusion: Sparse
representation

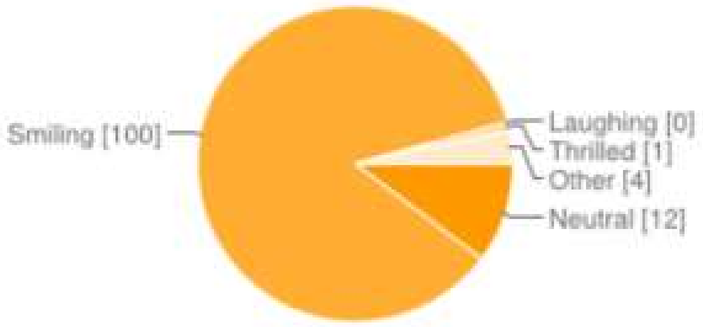
Macro-expression: from single people to a group of people



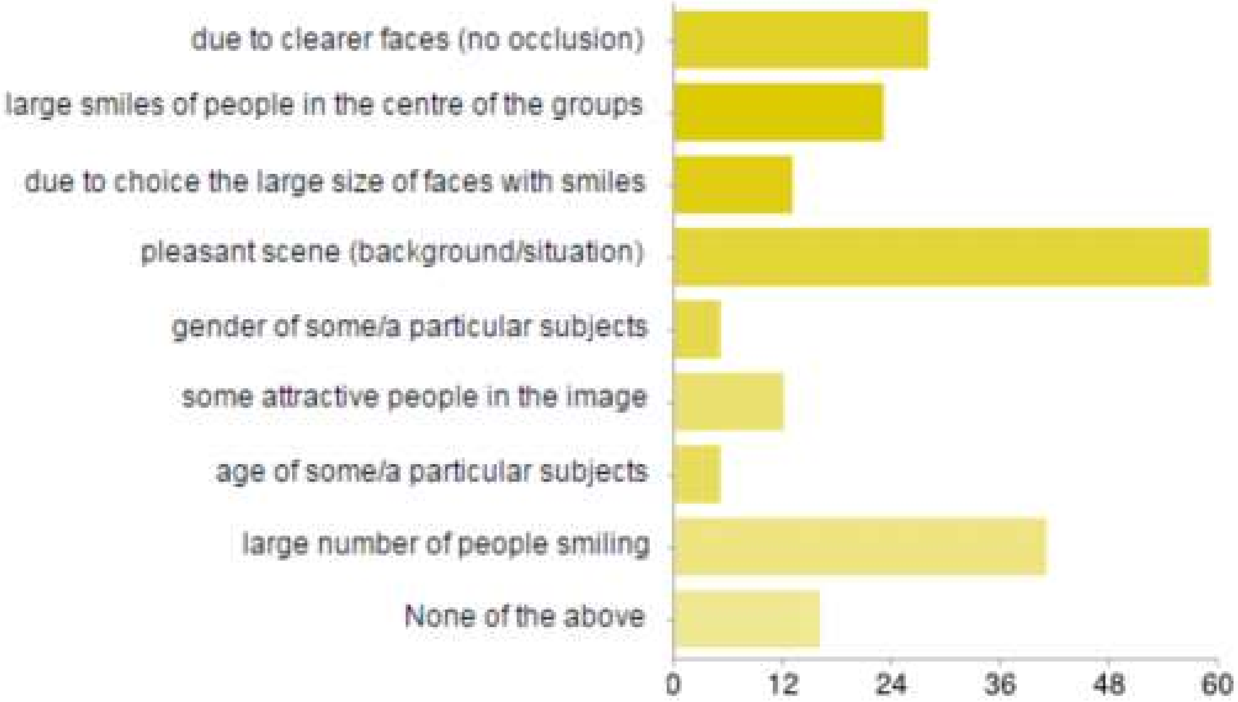
The presence of a large pool of data on multimedia enables us to explore images that contain more than one person.

The analysis of the emotion of multiple people in an image has various applications in multimedia such as image management and retrieval, photo album and event detection.

How will you describe the expression of the group?



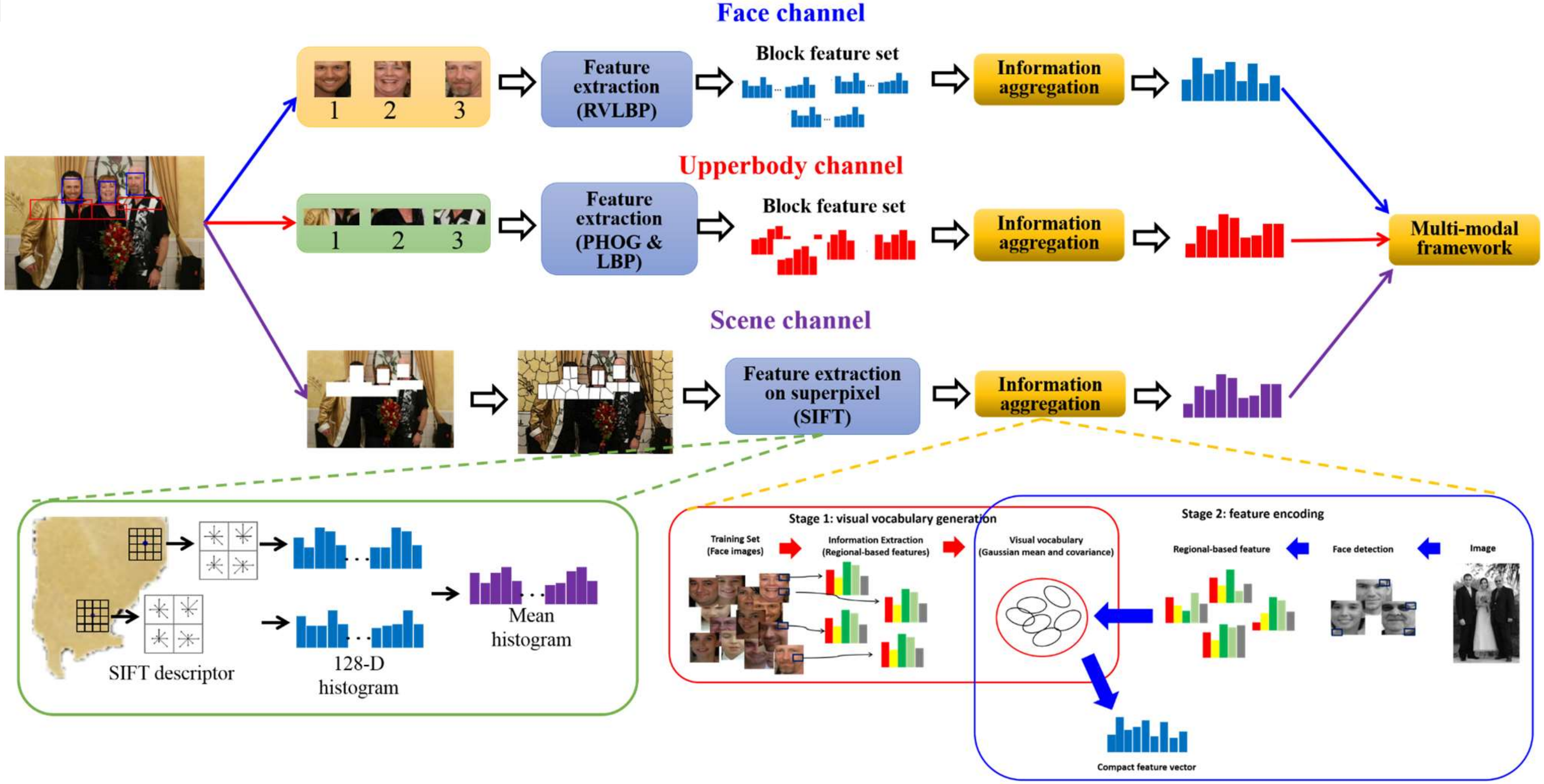
Was your choice motivated by: (multiple answers acceptable)n question 5 & 6



Any other reason for your choice?

Forced smiles | nothing | Smile (not necessarily in the center) | no | People tilted to come in the frame makes it more formal

Macro-expression: from single people to a group of people



X. Huang, A. Dhall, R. Goecke, M. Pietikäinen, G. Zhao. Multi-modal Framework for Analyzing the Affect of a Group of People. IEEE Transactions on Multimedia, 20(10): 2706-2721. 2018

Macro-expression: from 2D to 3D

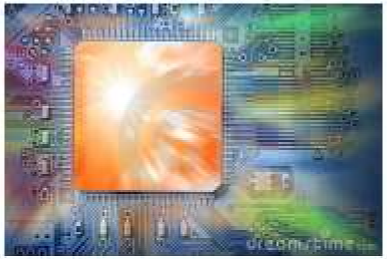
The majority of the works have focused primarily on the 2D data (images and videos)

- Broad range of use
- Being readily available
- Computational limitations

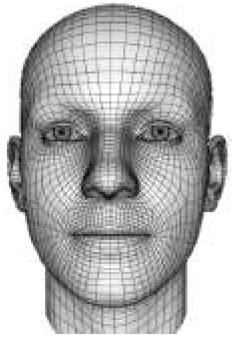
Accessible 3D Face

Capturing

Processing



Storing



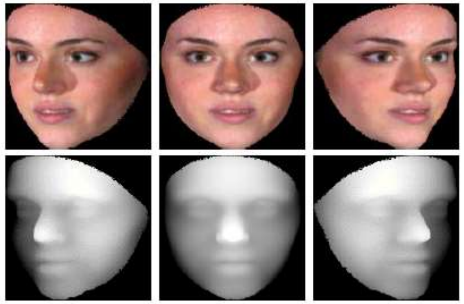
Infotech Oulu

UNIVERSITY OF OULU

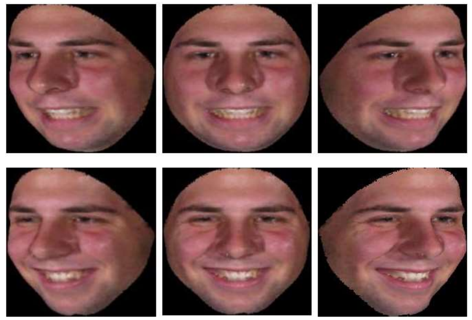
Macro-expression: from 2D to 3D

Benefit of multi-views for the FER

1. Provide more clues to recognize the low-intensity emotions.
   - Low-intensity Happy expression looks quite similar to the neutral or surprise expression.
   - The emotion is expressed more clearly in the side view.



2. Some similar expressions can be distinguished more easily with side views.
   - On the frontal view, Happy and Fear express similar movements on the face.
   - On side-view, the fear and happy expressions are quite different.



Infotech Oulu

UNIVERSITY OF OULU

Macro-expression: from 2D to 3D

Collaborative Network for 4D facial expression recognition

- Spatial information from multi-views is fed into a subnet for training.
- Information from different domains is represented in cross-domain images.
- Temporal information for each sample is represented as one dynamic image.
- The collaboration performed over multi-views is an added benefit of our proposed automatic 4D FER method.
- CCDN enjoys an expanded training size and lends itself a competing performance under general experimental settings.

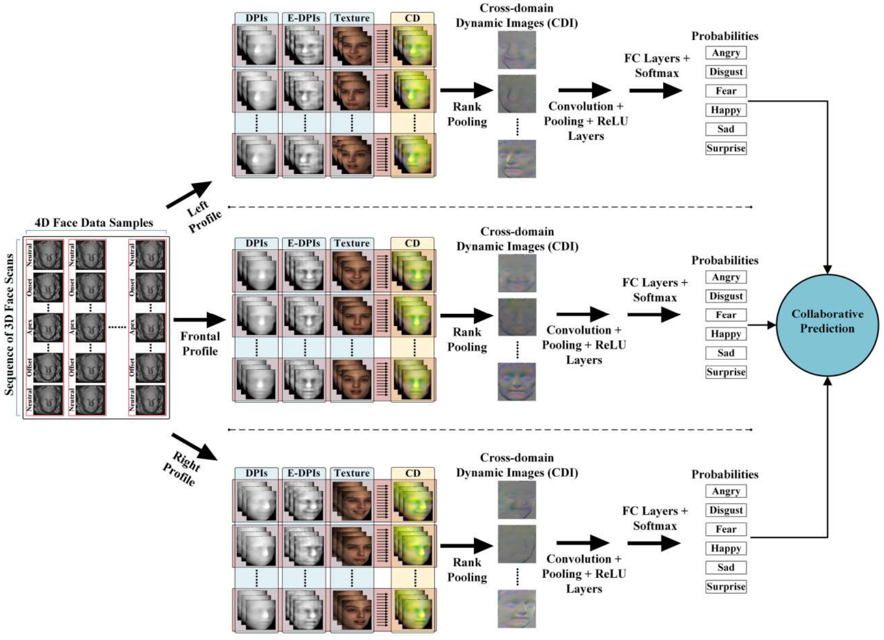
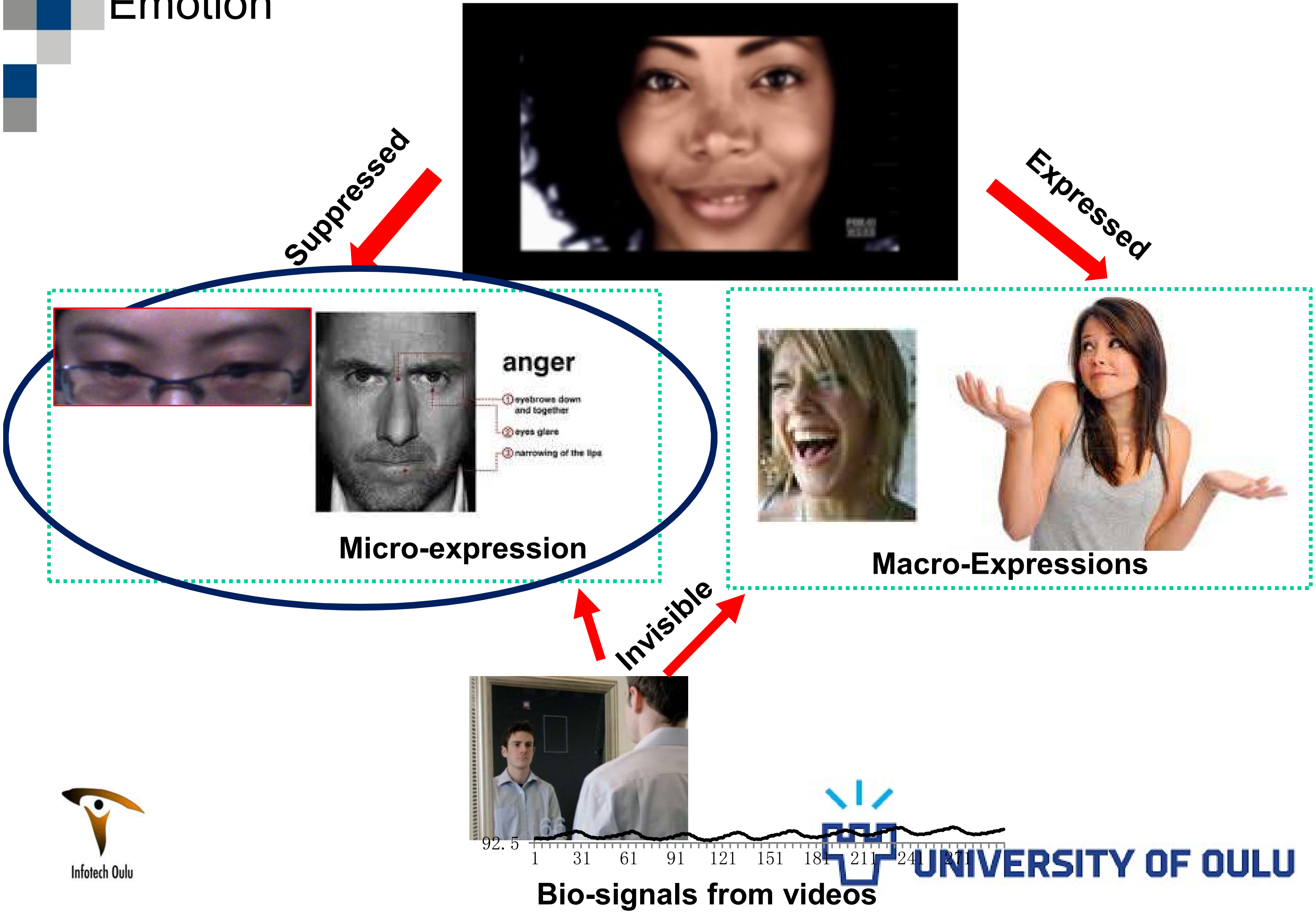


Fig. Flowchart of the proposed 4D FER network.

M. Behzad, N. Vo, X. Li, G. Zhao. Automatic 4D Facial Expression Recognition via Collaborative Cross-domain Dynamic Image Network. BMVC 2019.

Emotion



Infotech Oulu

UNIVERSITY OF OULU

Smiling

Micro-expression anguish

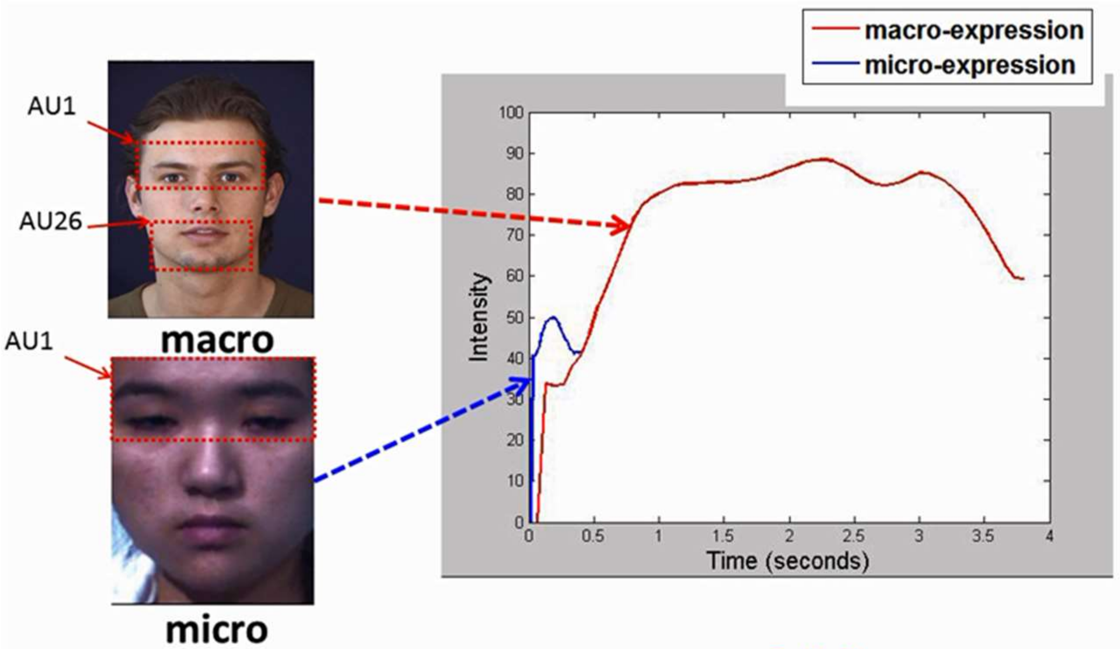
Micro-Expressions

- Rapid involuntary facial expressions
- Reveal suppressed affect

- Contradictions between facial expressions and the emotional state, enabling recognition of suppressed emotions.

T. Pfister, X. Li, G. Zhao & M. Pietikäinen (2011) Recognising spontaneous facial micro-expressions. ICCV 2011

Attribute	Micro-expressions	Macro-expressions
Duration	< 0,5 s	0,5 - 4 s
Facial movements	Subtle	Clearly visible
Emotion	Repressed	Expressed



W.J. Yan, Q. Wu, J. Liang, Y.H. Chen, X. Fu. How fast are the leaked facial expressions: The duration of micro-expressions. Journal of Nonverbal Behavior 37(4), 217-230 (2013)

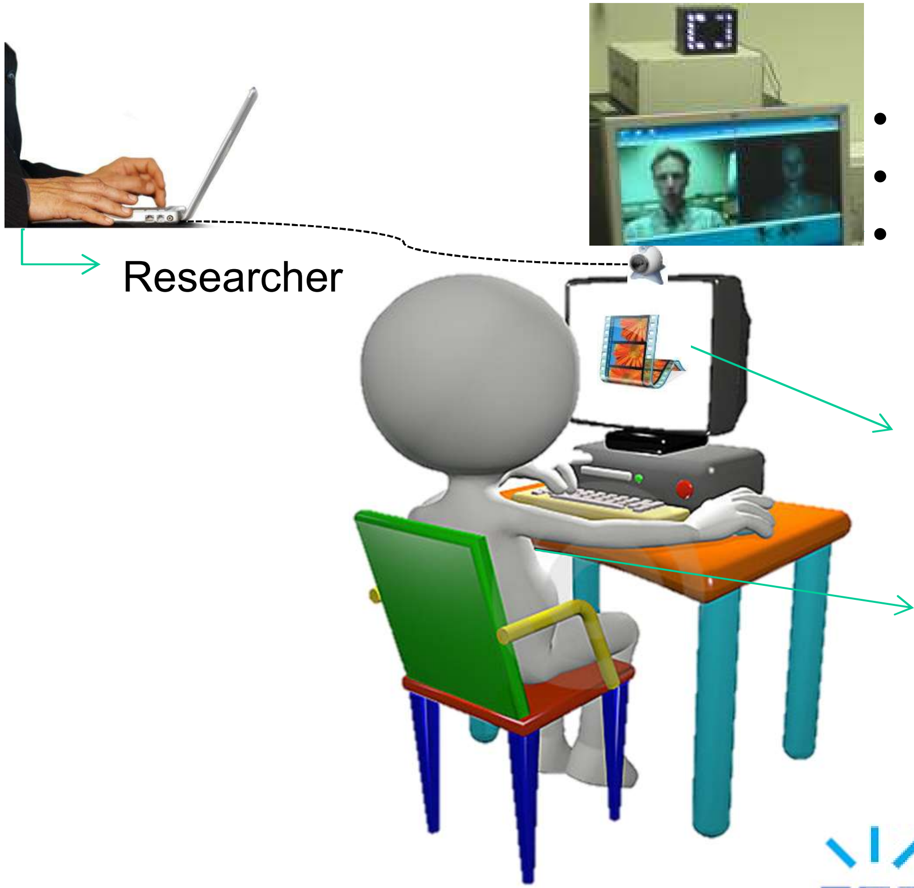
Two pathways

Prof. David Matsumoto: psychology professor at San Francisco State University

- One major nerve in the brain stem that enervates most of our facial muscles.
- A pathway, a neural pathway: from the subcortical areas that goes to that nerve that says fire the face when we're emotional - as well the rest of our physiology.
- Another pathway from the cortex in the motor strip that goes to that nerve that says control our expressions in a situation we want to control it.
- So there's two pathways going to that thing, and it's times of conflict of those two impulses that the micro-expressions are leaking out.

Micro-expression: starting with datasets

SMIC: Micro-expression Database

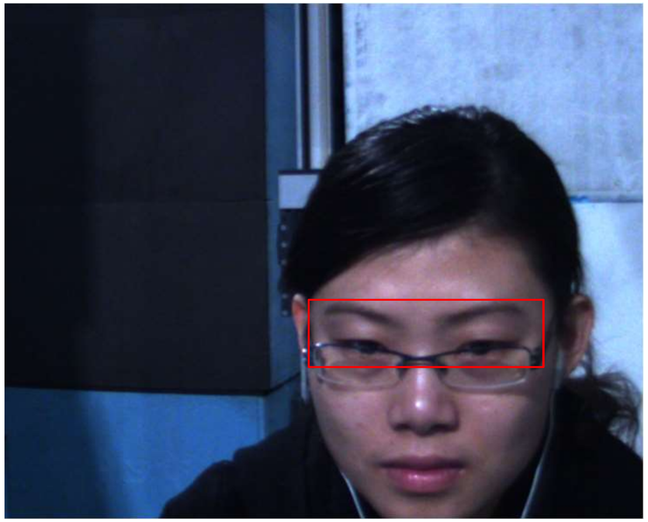


- High speed camera
- Ordinary speed RGB camera
- Ordinary speed NIR camera

Micro-expression: starting with datasets

Movie to induce Negative emotion:

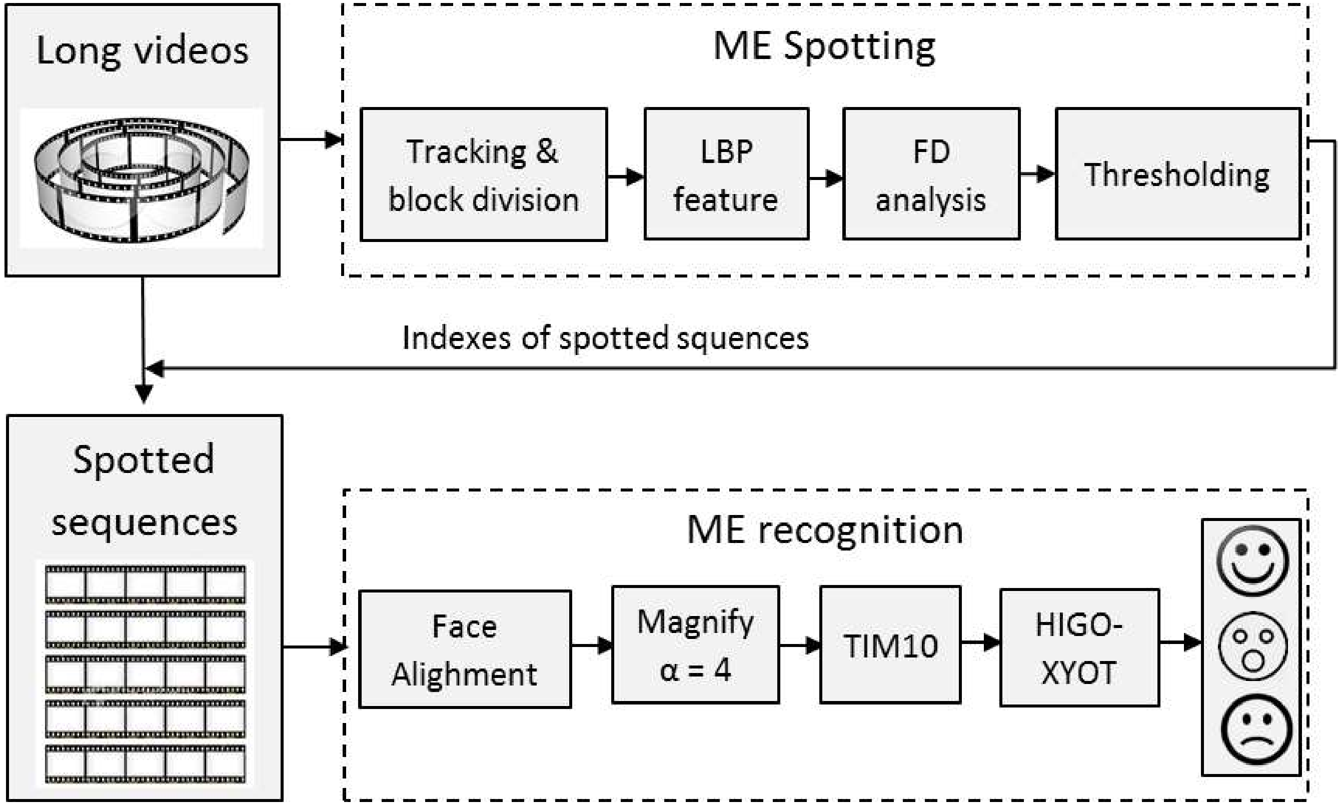
1/10 Speed



Micro-expression: starting with datasets

Database	Subjects		ME clips			Annotations	
	Number	MultiEth	Number	Resolution	Frame rate	Emotion category	AU category
SMIC (HS/NIR/VIS)	16/8/8	Y	164/71/71	640 × 480	100/25/25	*Pos* (51) *Neg* (70) *Sur* (43) / *Pos* (28) *Neg* (23) *Sur* (20) / *Pos* (28) *Neg* (24) *Sur* (19)	None
CASME (classA/classB)	19	N	195	640 × 480 1280 × 720	60	*Hap* (5) *Dis* (88) *Sad* (6) *Con* (3) *Fea* (2) *Ten* (28) *Sur* (20) *Rep* (40)	12+
CASME II	26	N	247	640 × 480	200	*Hap* (33) *Sur* (25) *Dis* (60) *Rep* (27) *Oth* (102)	11+
CAS(ME)^2	22	N	57	640 × 480	30	*Pos* (8) *Neg* (21) *Sur* (9) *Oth* (19)	28
SAMM	32	N	159	2040 × 1088	200	*Hap* (24) *Ang* (20) *Sur* (13) *Dis* (8) *Fea* (7) *Sad* (3) *Oth* (84)	ALL AUs
MEVIEW	16	N	29	720 × 1280	30	*Hap* (4) *Ang* (1) *Sur* (9) *Dis* (1) *Fea* (2) *Unc* (7) Con(5)	7
MMEW	36	N	300	1920 × 1080	90	*Hap* (36) *Ang* (8) *Sur* (89) *Dis* (72) *Fea* (16) *Sad* (13) *Oth* (66)	17
Composite ME	68	Y	442	640 × 480 1280 × 720 720 × 1280	200	*Pos* (109), *Neg* (250), and *Sur* (83)	27
Compound ME	90	Y	1050	640 × 480 1280 × 720 720 × 1280	200	*Neg* (233) *Pos* (82) *Sur* (70) *PS* (74) *N S* (236) *PN* (197) *NN* (158)	27

Micro-expression: combining Spotting and Recognition



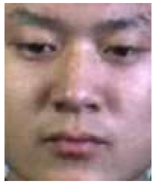
X. Li, X. Hong, A. Moilanen, X. Huang, T. Pfister, G. Zhao, and M. Pietikäinen. Towards Reading Hidden Emotions: A Comparative Study of Spontaneous Micro-expression Spotting and Recognition Methods. IEEE Transactions on Affective Computing, 9(4): 563-577, 2018.

Micro-expression: from same datasets to cross-datasets

- How methods work for the scenario where the training and testing samples belong to different micro-expression databases.
- Problem detail: Predict the micro-expression labels of unlabeled samples from Database A (e.g., SMIC) based on the labeled samples from Database B (e.g., CASME II) .
- and leads to a more challenging but interesting problem in micro-expression recognition:

Cross-Database Micro-Expression Recognition

Examples:



Training sample

Testing sample

Micro-expression: from same datasets to cross-datasets

- Domain Regeneration (DR): regenerating the source and target micro-expression samples such that they can abide by the same or similar feature distributions.

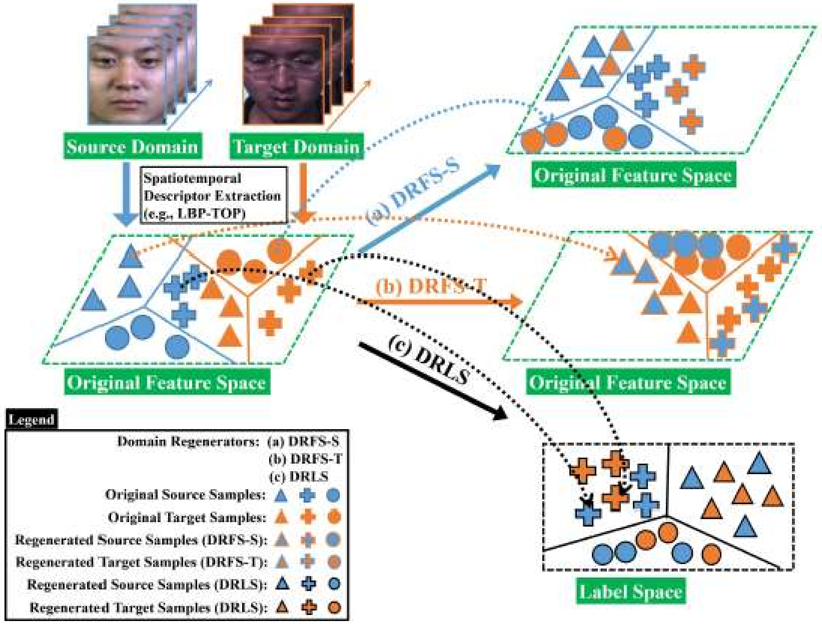


Fig. 1. Domain Regeneration (DR) framework for dealing with cross-database micro-expression recognition. Under this framework, we design three domain regenerators including (a) **DRFS-S**: Domain **R**egeneration in the original **F**eature **S**pace with unchanged **S**ource domain, (b) **DRFS-T**: Domain **R**egeneration in the original **F**eature **S**pace with unchanged **T**arget domain, and (c) **DRLS**: Domain **R**egeneration in the **L**abel **S**pace, for cross-database micro-expression recognition.

Yuan Zong, Wenming Zheng, Xiaohua Huang, Jingang Shi, Zhen Cui, Guoying Zhao. Domain regeneration for cross-database micro-expression recognition. IEEE Transactions on Image Processing, 27(5): 2484-2498, 2018.

Micro-expression: from video to apex

Snapshot taken at an point when the expression is at its apex can easily convey the emotion message

-Paul Ekman

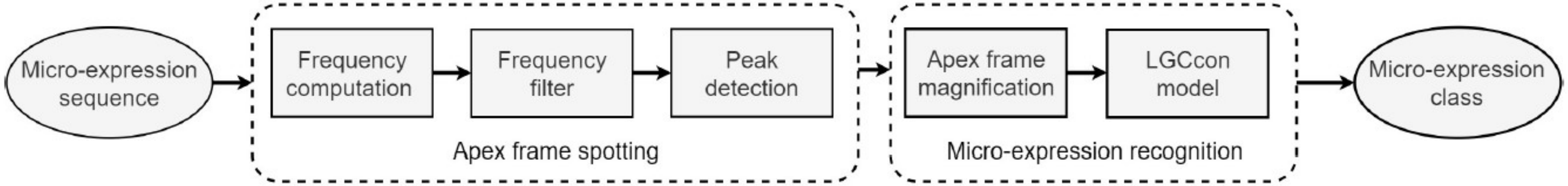
An example of micro-expression (Happiness) sequence

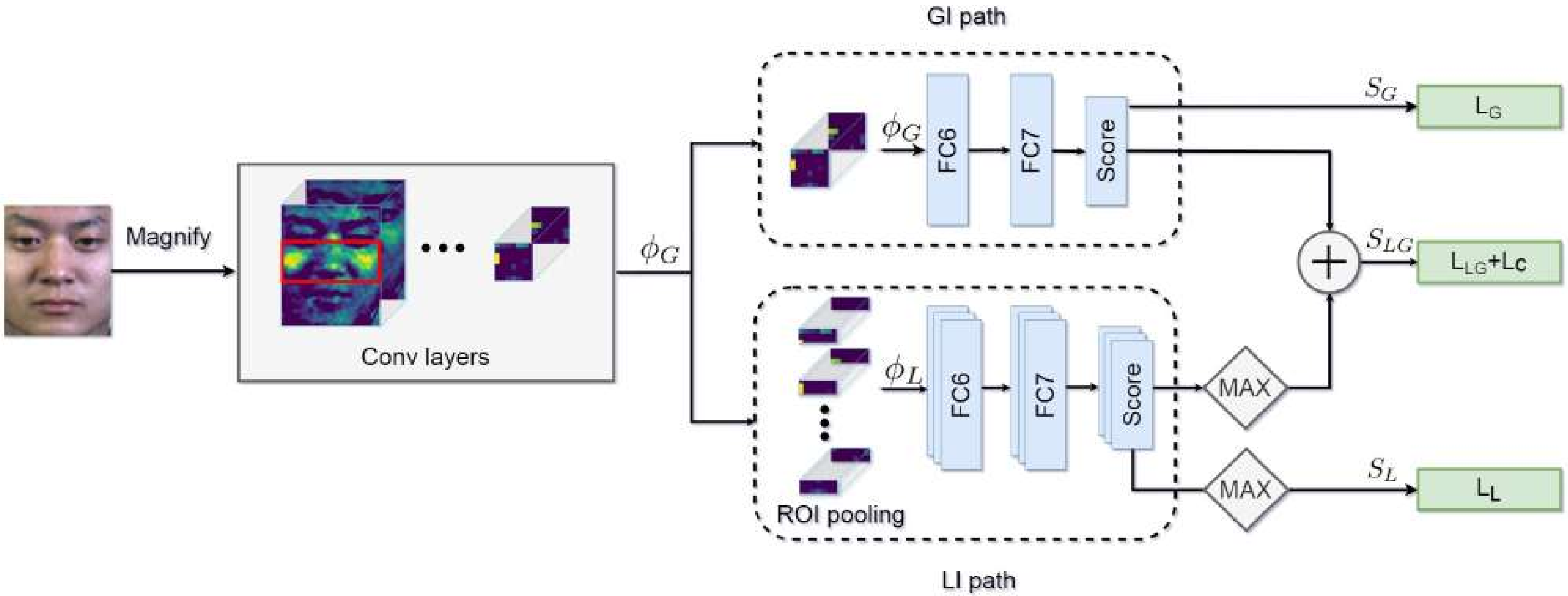
- ‘Does the apex frame in micro-expression also include the important information ?’
- ‘How is the recognition result based on the apex frame compared with the methods employing ME sequence?’

Infotech Oulu

UNIVERSITY OF OULU

Micro-expression: from video to apex





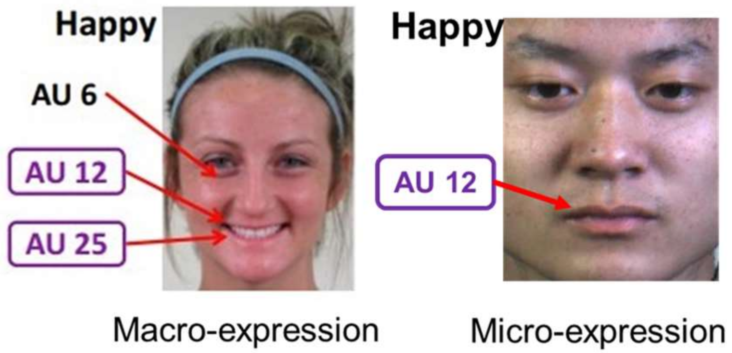
Yante Li, Xiaohua Huang, Guoying Zhao. Joint Local and Global Information Learning with Single Apex Frame Detection for Micro-expression Recognition. IEEE Trans. on Image Processing. 2020.

Micro-expression: from basic category to action units

- Facial action coding system is important for emotion recognition
- Few research on micro-expression action units (AUs) analysis

- **Difficulties**
  - Limited samples
  - Low intensity

Dataset	CASME II	CASME	SAMM	BP4D
Clips	247	195	159	328
Frames	<9,000	<2,500	<6,000	~140,000



Micro-expression: from basic category to action units

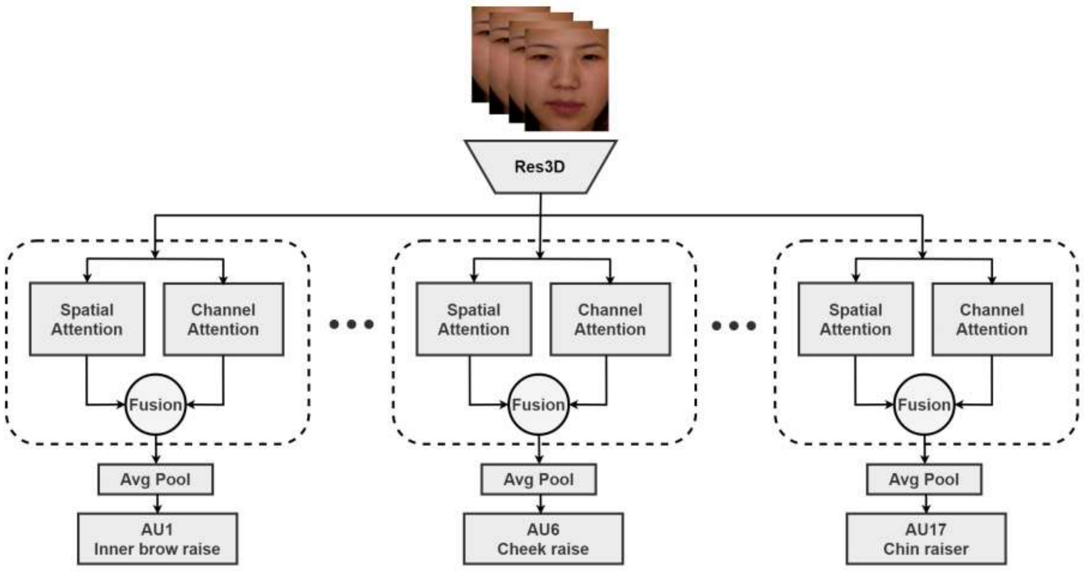
Contribution

- Low intensity

  Channel attention: capture micro regional changes by embedding the high-order statistics.

- Limited samples

  Spatial attention: leverage the relationship of individual regions



Yante Li, Xiaohua Huang, Guoying Zhao. Micro-expression Action Unit Detection with Spatial and Channel Attention. Neurocomputing. 2021.

Micro-expression: from basic category to action units

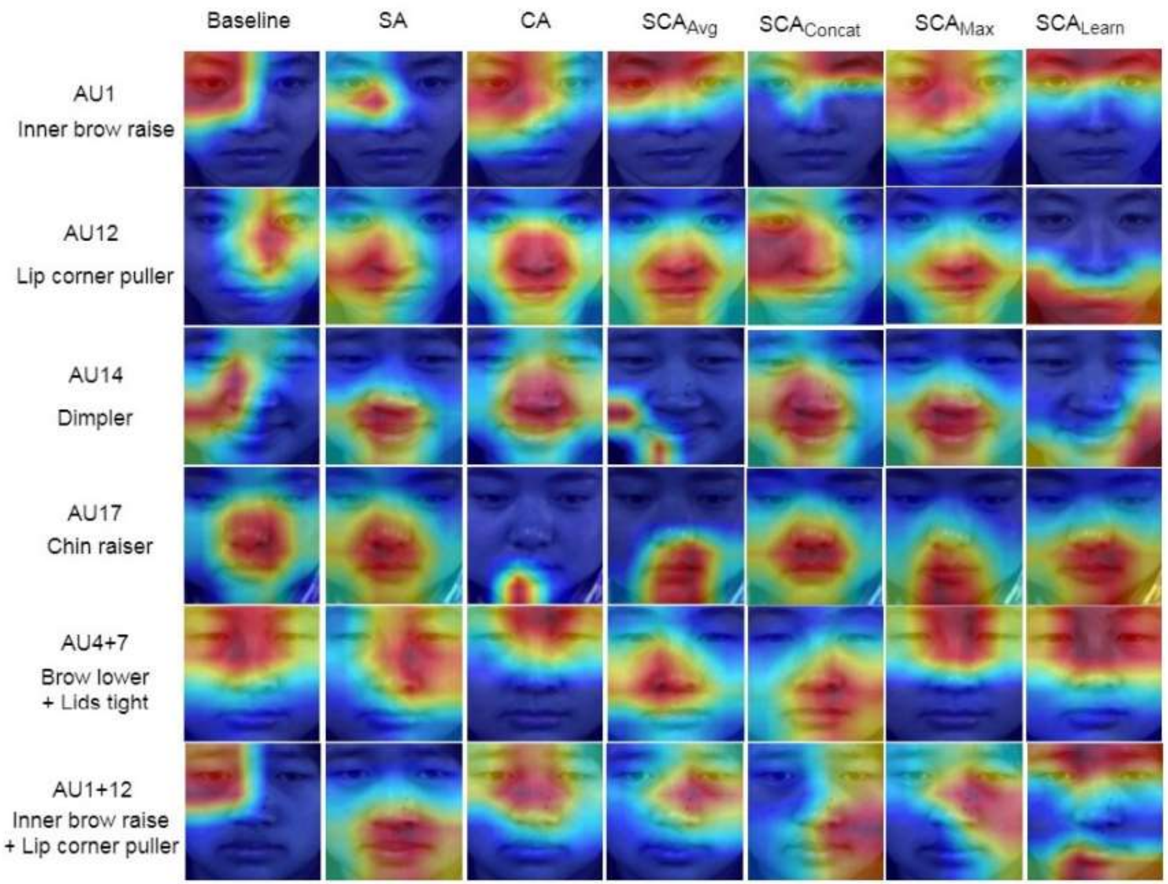
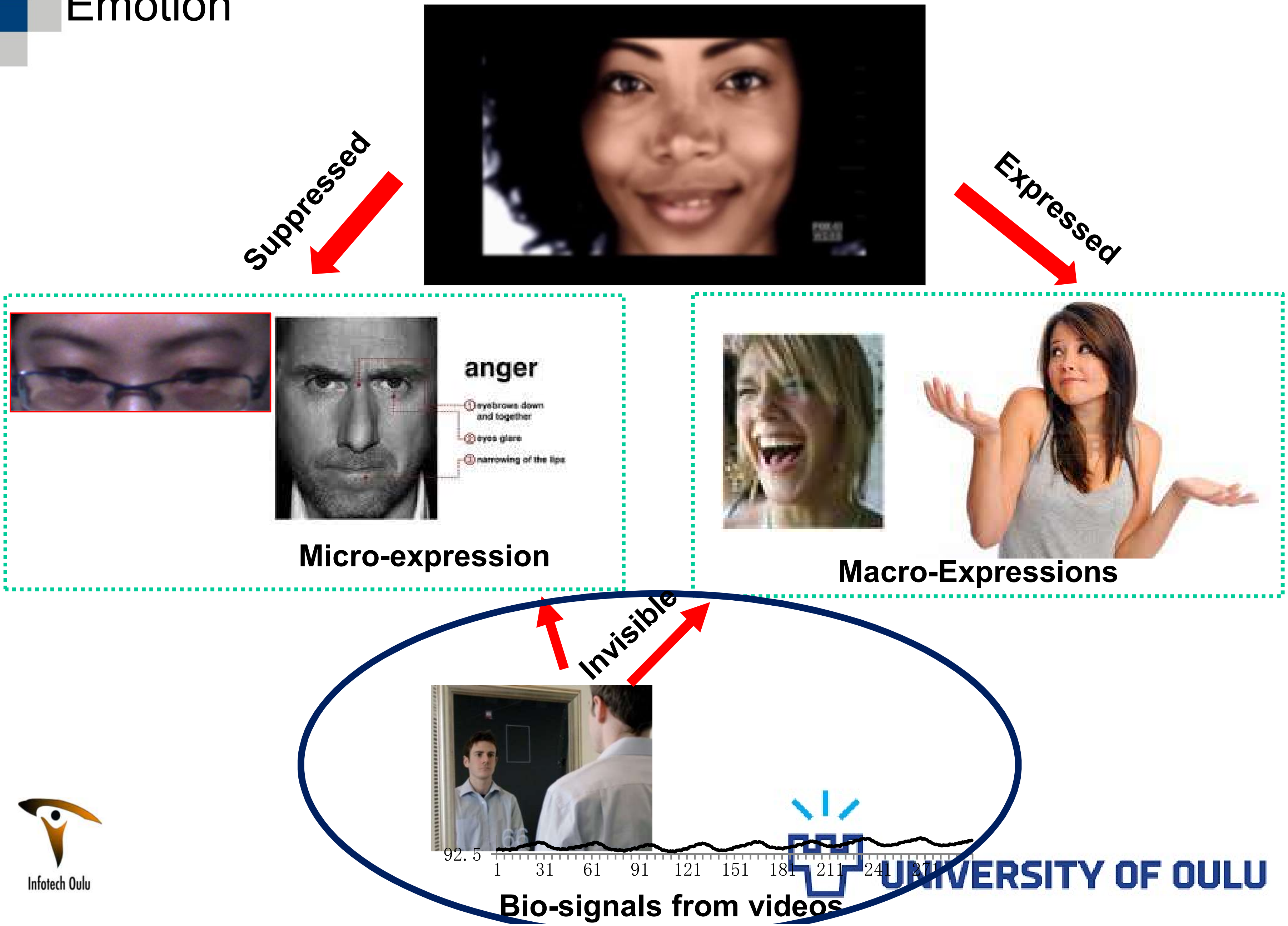


Figure: Micro-expression AU visualization. Some examples of class activation maps for baseline (Res3D18), SA, CA, and different fusion methods. SCAAvg, SCAConcat, SCAMax, and SCALearn represent fusion through maximum, average, feature concatenation and learning methods, respectively.

Infotech Oulu

UNIVERSITY OF OULU

Emotion



Infotech Oulu

UNIVERSITY OF OULU

We can:

From the chest

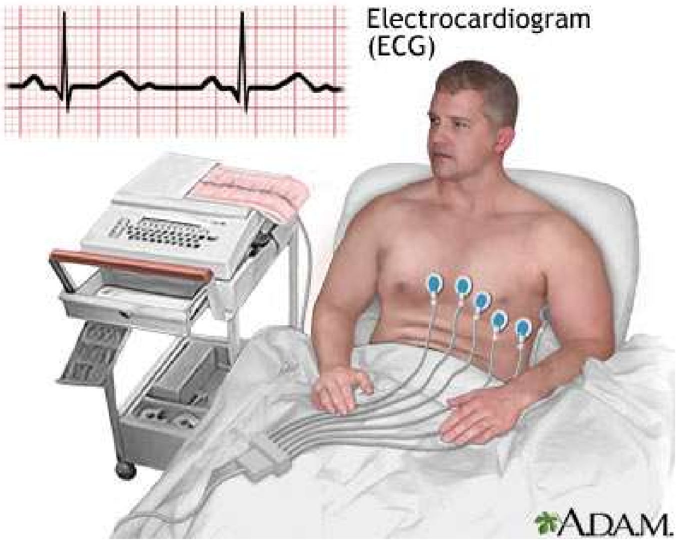
From other body parts

Infotech Oulu

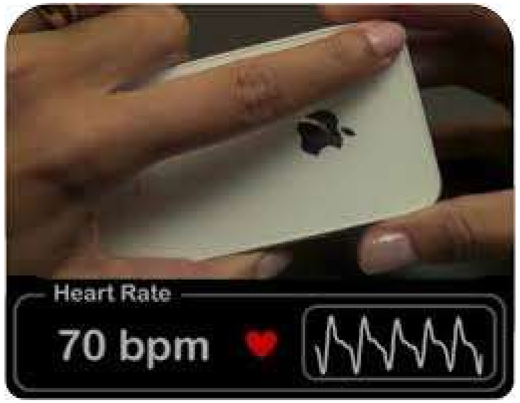
UNIVERSITY OF OULU

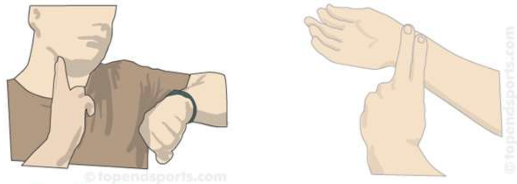
We can:

Traditional



Convenient





Infotech Oulu

UNIVERSITY OF OULU

Problems with all these products: special instrument & contact

Key words:

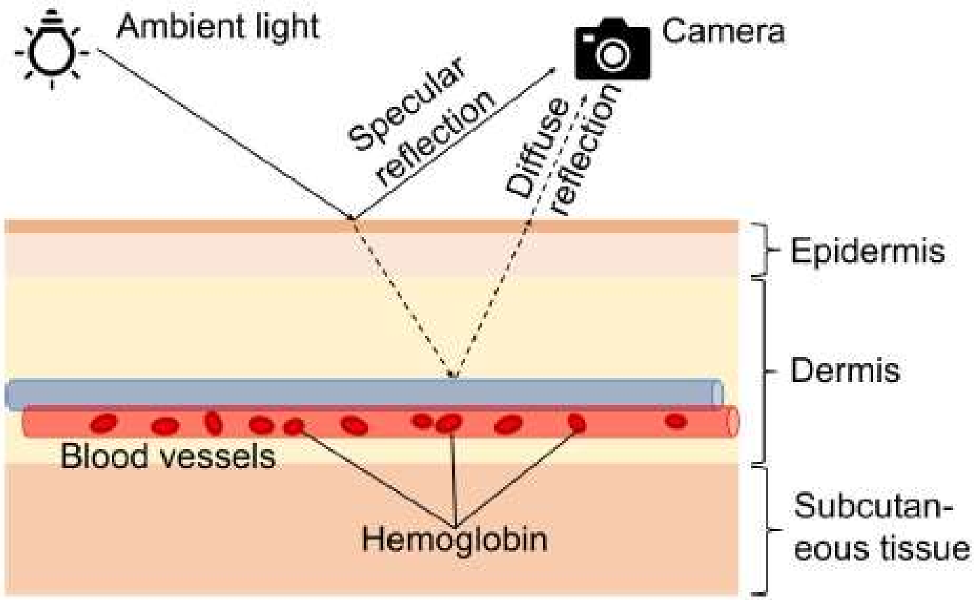
1. Remote
2. Camera

Infotech Oulu

UNIVERSITY OF OULU

Heart rate measuring from videos (remotely!)

The blood volume of micro-vascular changes together with cardiac pulse

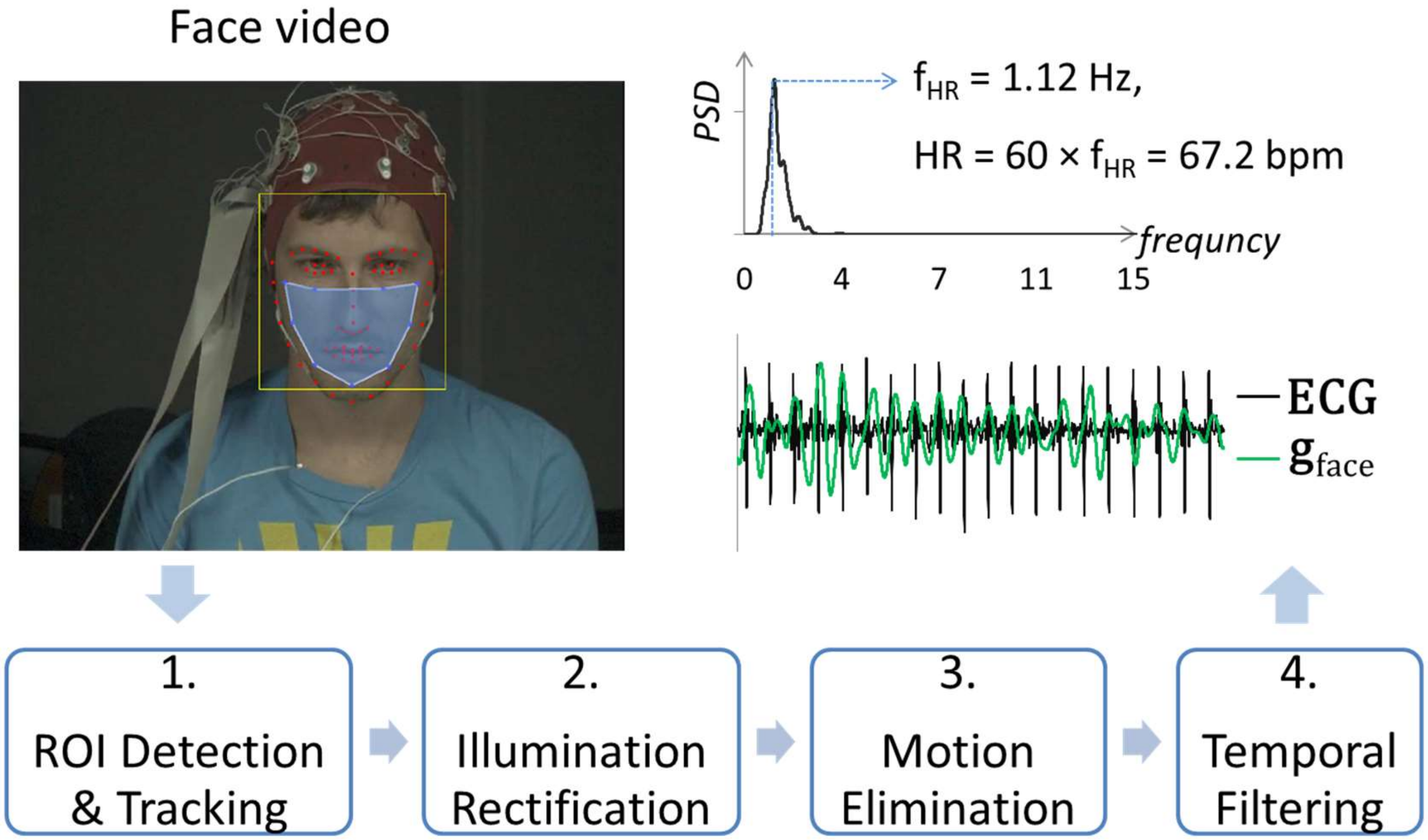


Wu et al. Eulerian Video Magnification for Revealing Subtle Changes in the World. SIGGRAPH 2012

Infotech Oulu

UNIVERSITY OF OULU

Heart rate measuring from videos (remotely! Invisible)



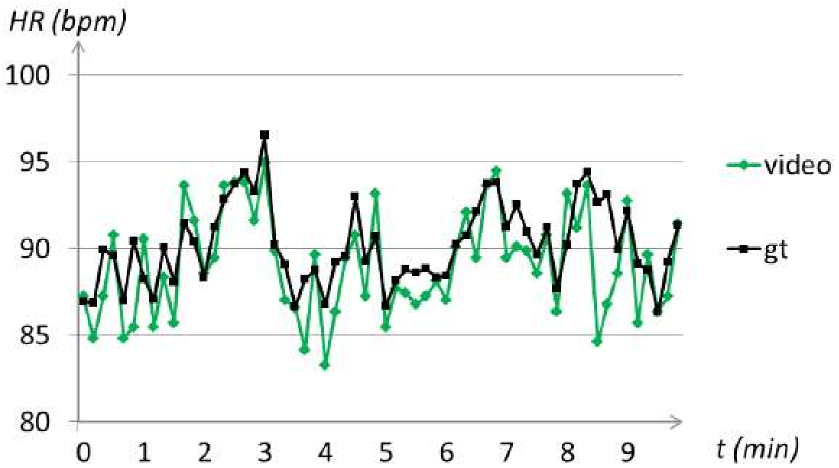
X. Li, J. Chen, G. Zhao, M. Pietikäinen. Remote Heart Rate Measurement From Face Videos Under Realistic Situations. CVPR 2014.

Heart rate measuring from videos (remotely!)

1st frame of input video

Game Playing:

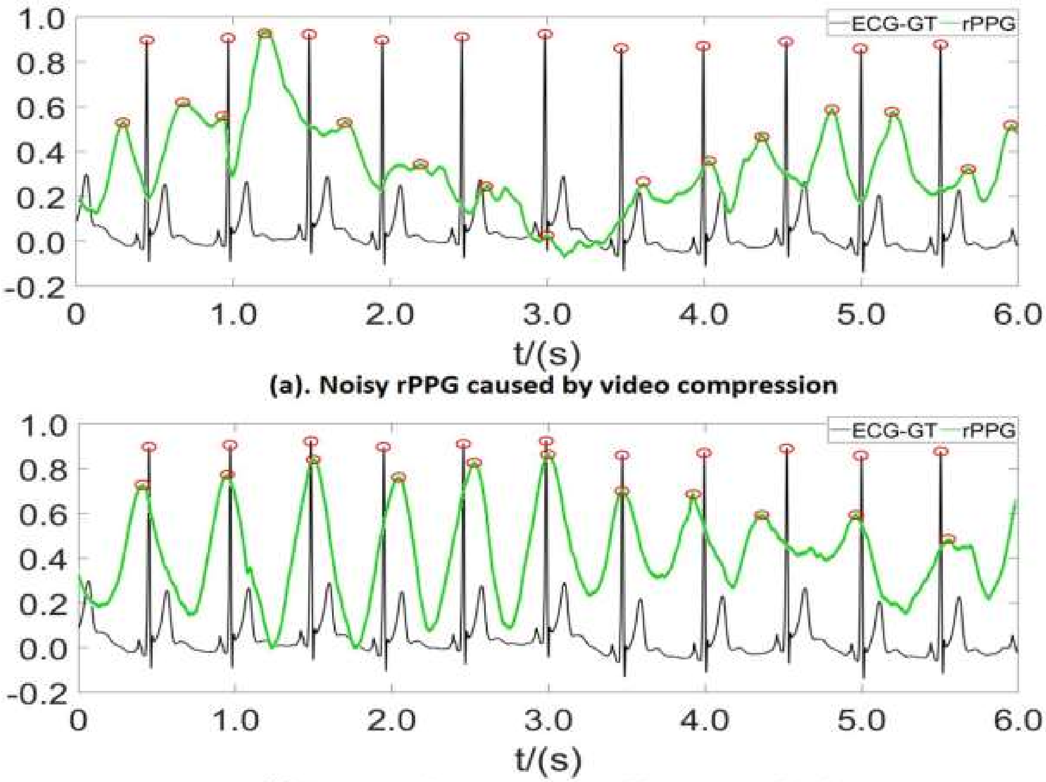
- HRs measured by using our framework has a mean error rate of 1.89%.
- Subject's HR changes as the content of the game progresses, which can be used for later analysis about the player's experience.



HR monitoring of one subject while playing a video game. The black curve is the ground truth HR measured by Polar system; the green curve is HR measured from video by using our framework.

From high quality to highly compressed videos

The first attempt to counter video compression loss and recover rPPG signals from highly compressed videos.



(a). Noisy rPPG caused by video compression

(b). Improved rPPG recovered by our method

From highly compressed video

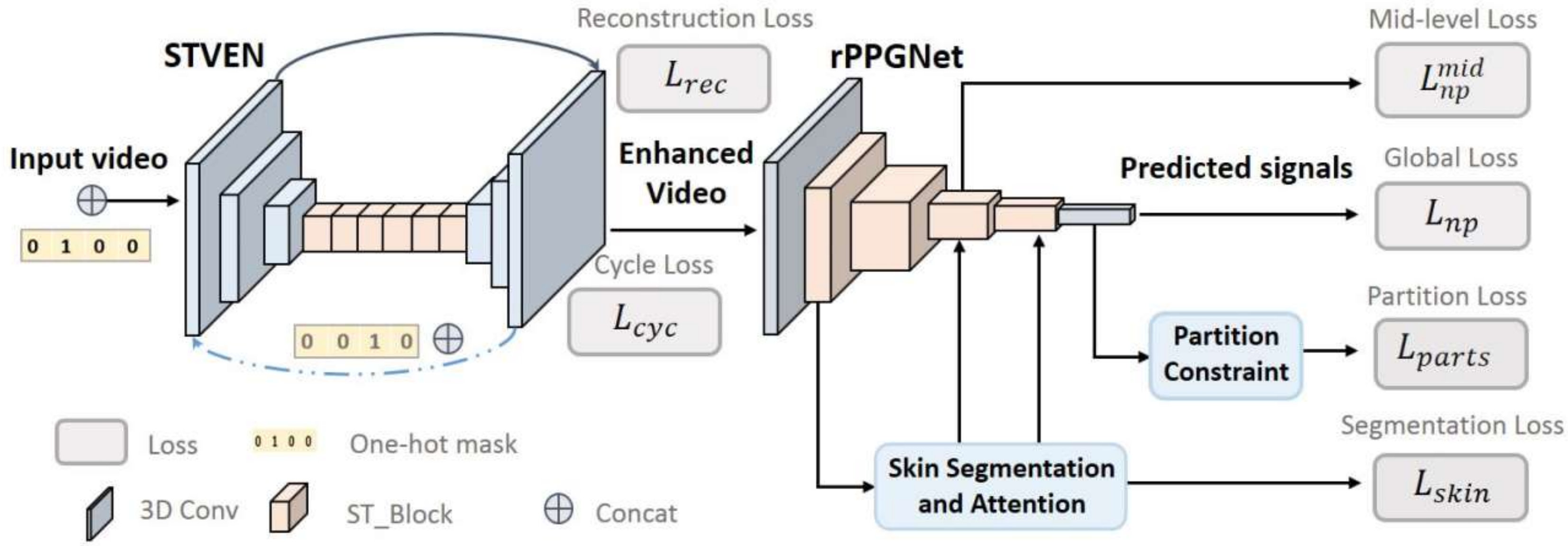
From quality enhanced video

Z. Yu, W. Peng, X. Li, X. Hong, G. Zhao. Remote Heart Rate Measurement from Highly Compressed Facial Videos: an End-to-end Deep Learning Solution with Video Enhancement. ICCV 2019. IEEE Finland Best Student Conference Award.

From high quality to highly compressed videos

- Spatio-Temporal Video Enhancement Network (STVEN) for video enhancement.

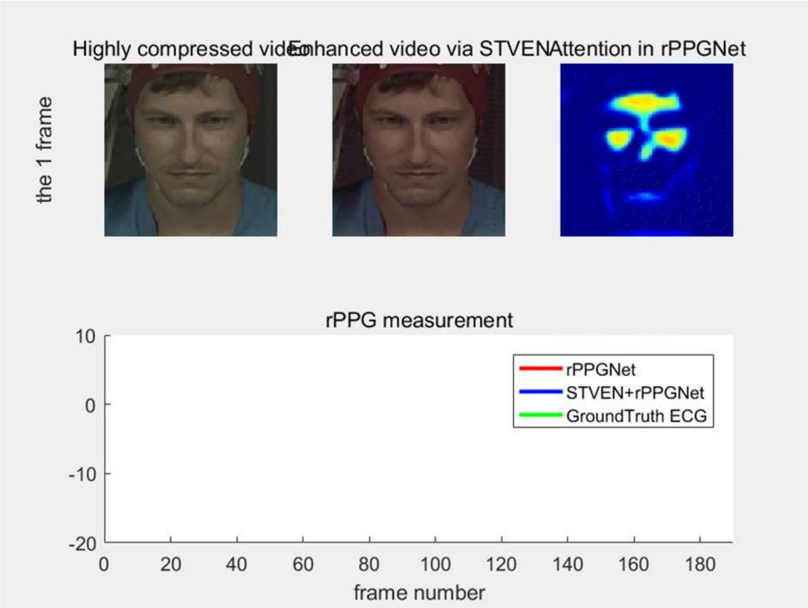
- rPPGNet for rPPG signal recovery.



Infotech Oulu

UNIVERSITY OF OULU

Our Method (Demo):



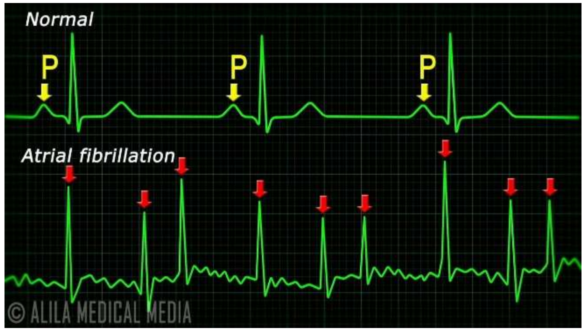
Infotech Oulu

UNIVERSITY OF OULU

Application to remote Atrial Fibrillation Detection

Heart diseases: Atrial Fibrillation

AF is ~~one of~~ the most common type of arrhythmia, characterized by rapid and irregular beating of the atria.



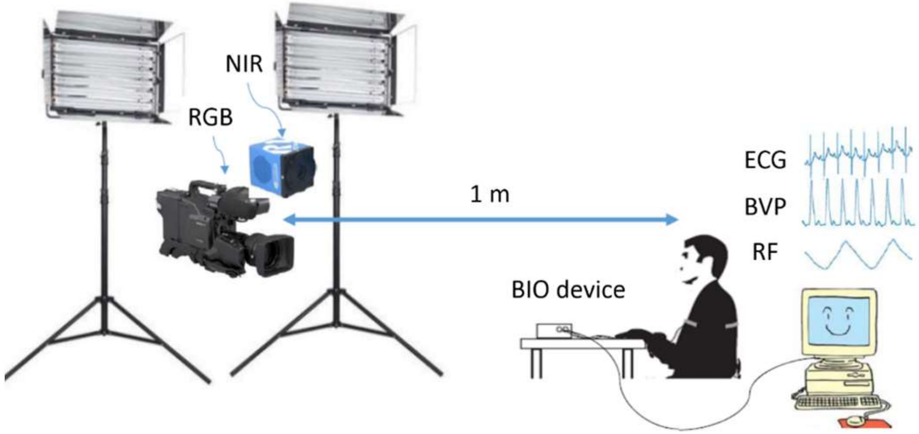
ECG sample of normal and AF patient.

- high incidence, affects 2%-3% of the population;
- may lead to blood clot, stroke, etc;
- early detection and intervention is crucial for good treatment outcome;
- "**I had no symptoms at all**. I discovered my AF at a regular check-up. I'm glad we found it early."

X. Li, I. Alikhani, J. Shi, T. Seppänen, J. Junttila, K. Majamaa-Voltti, M. Tulppo and G. Zhao. The OBF Database: A Large Face Video Database for Remote Physiological Signal Measurement and Atrial Fibrillation Detection. International Conference on Face and Gesture 2018.

Oulu BioFace (OBF) database

Setups:



Oulu BioFace (OBF) database

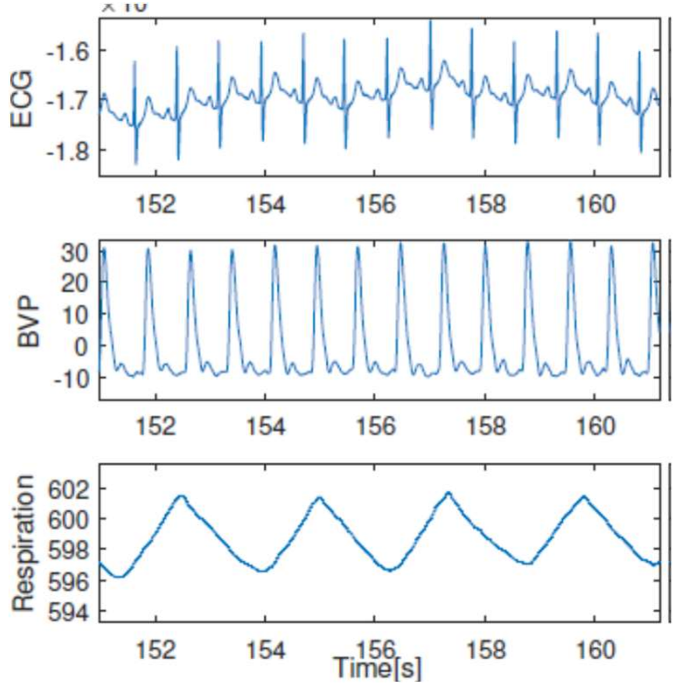
Equipment for OBF data recording

Device	Specifications	Settings	Output
Computer	HP EliteDesk	Windows 7 OS	N/A
RGB camera	Blackmagic URFA mini	ISO 400, FPS 60, HD 1920 x 1080	RGB video
NIR camera	Customized USB 2.0 Camera box (SN9C201&202)	640 x 480, FPS 30	NIR video
LED lights (2)	Aputure, LightStorm LS 1c	Brightness: 3500 lux, Temperature: 5500 k	N/A
Biosignal Acquisition	NeXus-10 MKII	N/A	N/A
ECG sensor	NX-EXG2B	256Hz	ECG Signal
Respiratory Belt	NX-RFP1B	32Hz	RF Signal
BVP sensor	NX-BVP1C	128Hz	BVP Signal

Participants: 100 healthy + 138 AF patients

Oulu BioFace (OBF) database

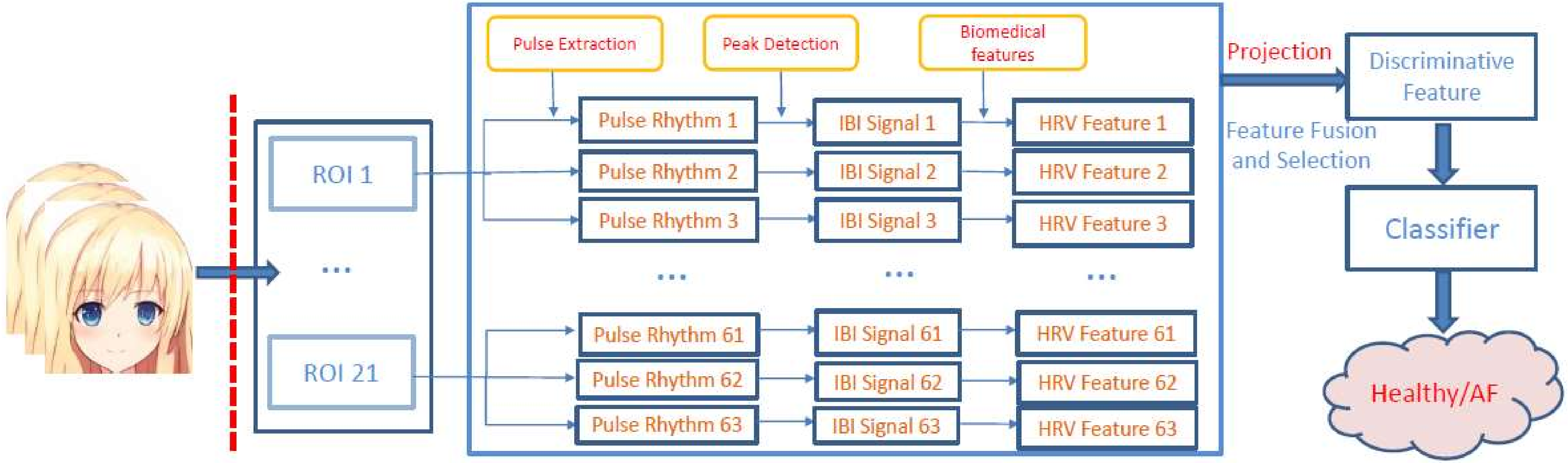
Data composition:



Modalities	Data length/person		Total (min)
	Healthy (n=100)	AF (n=138)	
RGB video, NIR video, ECG, BVP, RF	Resting state: 5 minutes	Prior-treatment: 5 minutes	2380
	Post-exercise: 5 minutes	Post-treatment: 5 minutes	

Infotech Oulu

UNIVERSITY OF OULU



- Utilize various rPPG algorithms to capture pulse rhythms from different regions on the face video: 21 facial ROIs with 3 different pulse extraction methods -> Totally 63 HRV features
- Investigate biomedical statistical methods to extract suitable features from each pulse signal
- Propose a feature fusion algorithm by learning a projection matrix to select and combine reasonable information from multiple physiological features

J. Shi, I. Alikhani, X. Li, Z. Yu, T. Seppänen and G. Zhao. Atrial Fibrillation Detection from Face Videos by Fusing Subtle Variations. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), 2019

<u>Application to face anti-spoofing</u> BY DETECTING *pulse* FROM FACE VIDEOS

THE PROBLEM

- Authentication system
  - Iris, finger prints…
  - **Face**
    - Registered user
    - Outsider
- Security risk: <u>Face Spoofing</u> attacks

1) Print/photo

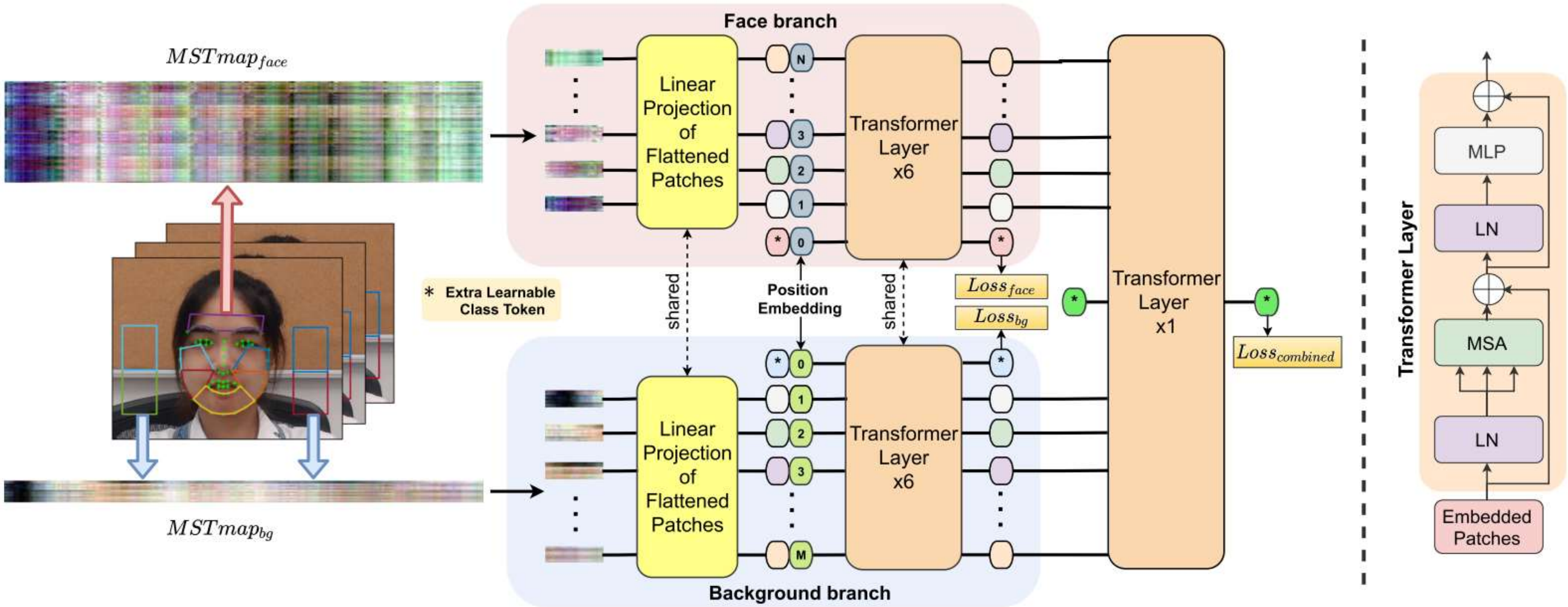
2) Video replay

3) Mask Attack:

- 3D printing
- Depth
- Eye movements

Infotech Oulu

Framework of TransRPPG



Given a face video, two MSTmaps are constructed from facial and background region, respectively. Then two-branch vision transformers (shared) are used to extract the respective rPPG and environmental features. Finally, the combined features are refined via an extra transformer for binary (bonafide or mask attack) prediction.

Zitong Yu, Xiaobai Li, Pichao Wang, Guoying Zhao. TransRPPG: Remote Photoplethysmography Transformer for 3D Mask Face Presentation Attack Detection. IEEE Signal Processing Letters. 2021.

RePSS 2020 challenge: The 1st Challenge on Remote Physiological Signal Sensing (RePSS) with CVPR 2020

https://competitions.codalab.org/competitions/22287

The training data of RePSS is randomly selected from VIPL-HR-V2 database. RGB videos of 500 subjects recorded with Realsense F200 camera at the average speed of 25 fps with resolution of 960 by 720 are used.
The testing data of RePSS challenge consists of two parts, that 100 subjects (no overlap with the training set) from the VIPL-HR-V2 database and 100 subjects (all from the healthy group) from the OBF databases are used.

VIPL-HR-V2

OBF

images of anonymized testing videos. The left one from VIPL-HR, and the right one from OBF.

Statistical information of VIPL-HR-V2 and OBF subjects.

	VIPL-HR-V2	OBF
Age (y)	35.4 ± 18.0, [6, 60]	31.6 ± 8.8, [18, 68]
Gender	49% M , 51% F	61% M , 39% F
Ethnic	Asian:100%	Caucasian:32% Asian:37%, Others:31%
Weight (Kg)	61 ± 12	71 ± 16
Wear eyeglasses	N/A	39%

RePSS 2021 challenge: The 2nd Challenge on Remote Physiological Signal Sensing (RePSS) with ICCV 2021

https://competitions.codalab.org/competitions/30855

1) Track1: inter-beat-interval (IBI) curve measurement from facial videos.

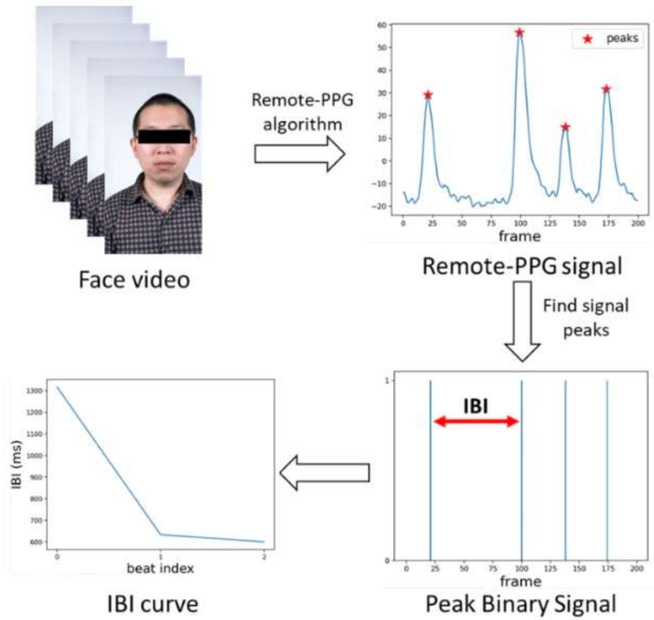


Fig.1 IBI curve and peak binary signal from a face video

2) Track2: respiration measurement from facial videos.

Emotions

- BMVC2019 (Valence, Arousal)

Z. Yu, X. Li, G. Zhao. Remote Photoplethysmograph Signal Measurement from Facial Videos Using Spatio-Temporal Networks. BMVC 2019.

- IJCAI 2021 (Pain)

R. Yang, Z. Guan, Z. Yu, X. Feng, J. Peng, G. Zhao. Non-contact Pain Recognition from Video Sequences with Remote Physiological Measurements Prediction. International Joint Conference on Artificial Intelligence (IJCAI) 2021

- TMM 2021 (Pain)

D. Huang, X. Feng, H. Zhang, Z. Yu, J. Peng, G. Zhao, Z. Xia. Spatio-Temporal Pain Estimation Network with Measuring Pseudo Heart Rate Gain. IEEE Trans. on Multimedia, 2021.

Summary

Machines with some emotional intelligence are emerging

Various powerful methods for emotion analysis were introduced: Expressed -> Supressed -> Invisible.

- Expressed: static (apex) -> dynamic (video), single person -> group; 2D -> 3D/4D;
- Supprssed: seperated recongtion and spotting -> combined; same datasets -> cross datasets; video -> apex; categores -> action unit;
- Invisible: handcrafted -> deep learning end to end; high quality -> highly compressed; HR/HRV measure -> AF detection and face anti-spoofing applications.

They provide a good basis for implementing affect-sensitve systems

Micro-expressions and heart rate provide useful invisible information

Infotech Oulu

UNIVERSITY OF OULU

Future topics

There are still many challenges to make major breakthroughs

- Analysis of spontaneous emotions in real-world conditions is a major challenge
- The context, e.g. what a person is doing and where, should be considered
- Much more (multi-modal) data is needed for learning to analyze natural expressions!

陈杰：北大；鹏城实验室

洪晓鹏：哈尔滨工业大学

黄晓华：南京工程学院

李晓白：芬兰Oulu大学

史金刚：西安交通大学

刘鑫：芬兰LUT

Infotech Oulu

UNIVERSITY OF OULU

Kevin

Mohammad

Qianru

Hanlin

Jukka

Henglin

Khanh

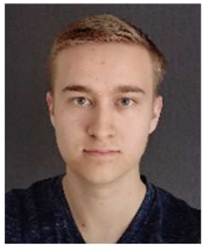
Yante

Haoyu

Wei

Muzammil

Zitong



Tuomas

Atif

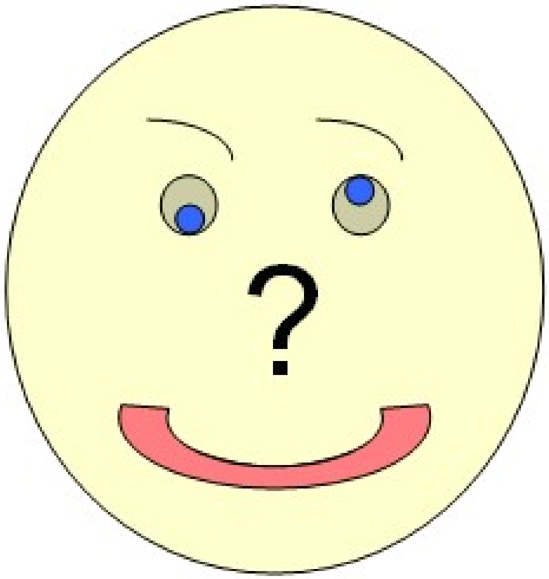
Marco

Collaborators

- 中科院计算所、心理所、自动化所；西北工业大学，西北大学，中南大学，东南大学，........
- 英国帝国理工大学，牛津大学；美国CMU，匹兹堡大学，马里兰大学；澳大利亚悉尼大学；荷兰**Delft**；......

Infotech Oulu

UNIVERSITY OF OULU



Thanks!

Infotech Oulu

UNIVERSITY OF OULU