TÉCNICO LISBOA

Social Robots and Human-Robot Interaction
MSc Course 2021 / 2022

Lecture on **Social Robots Recognising and Understanding the Others**

Invited Lecture by
Filipa Correia
Postdoc at LARSyS-ITI

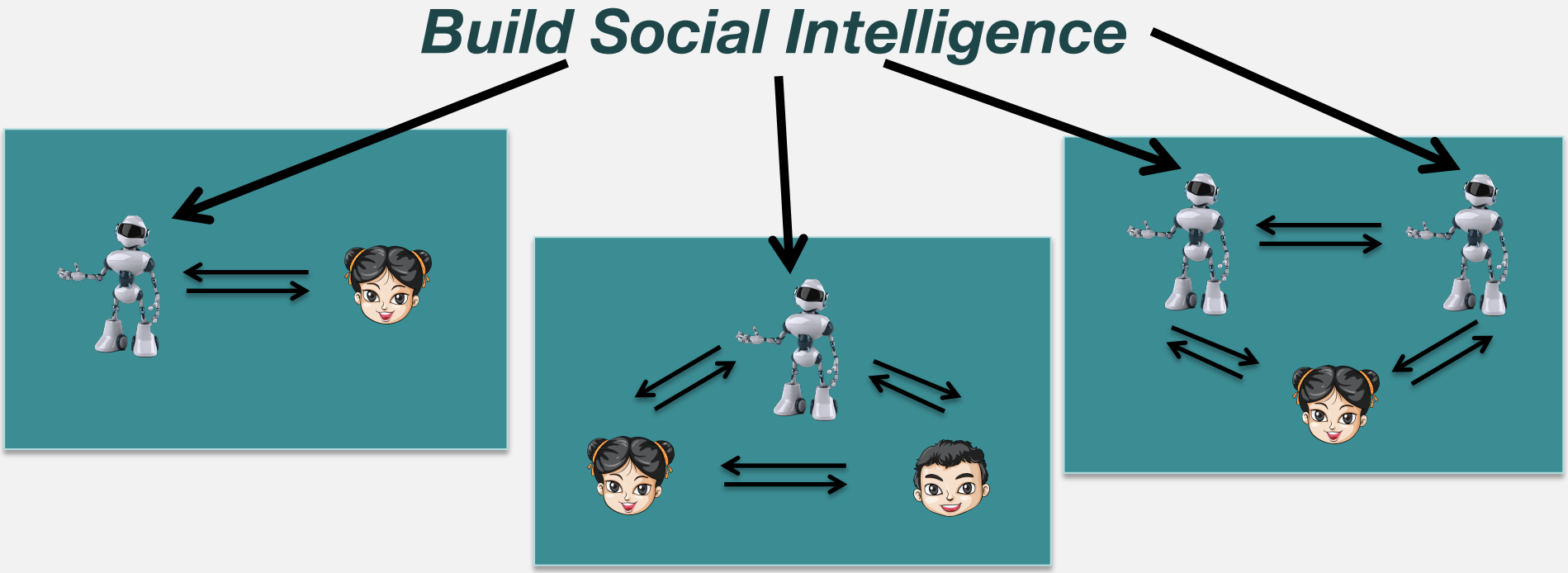
TÉCNICO LISBOA

Before starting…

Prof. Ana Paiva will discuss projects tomorrow!

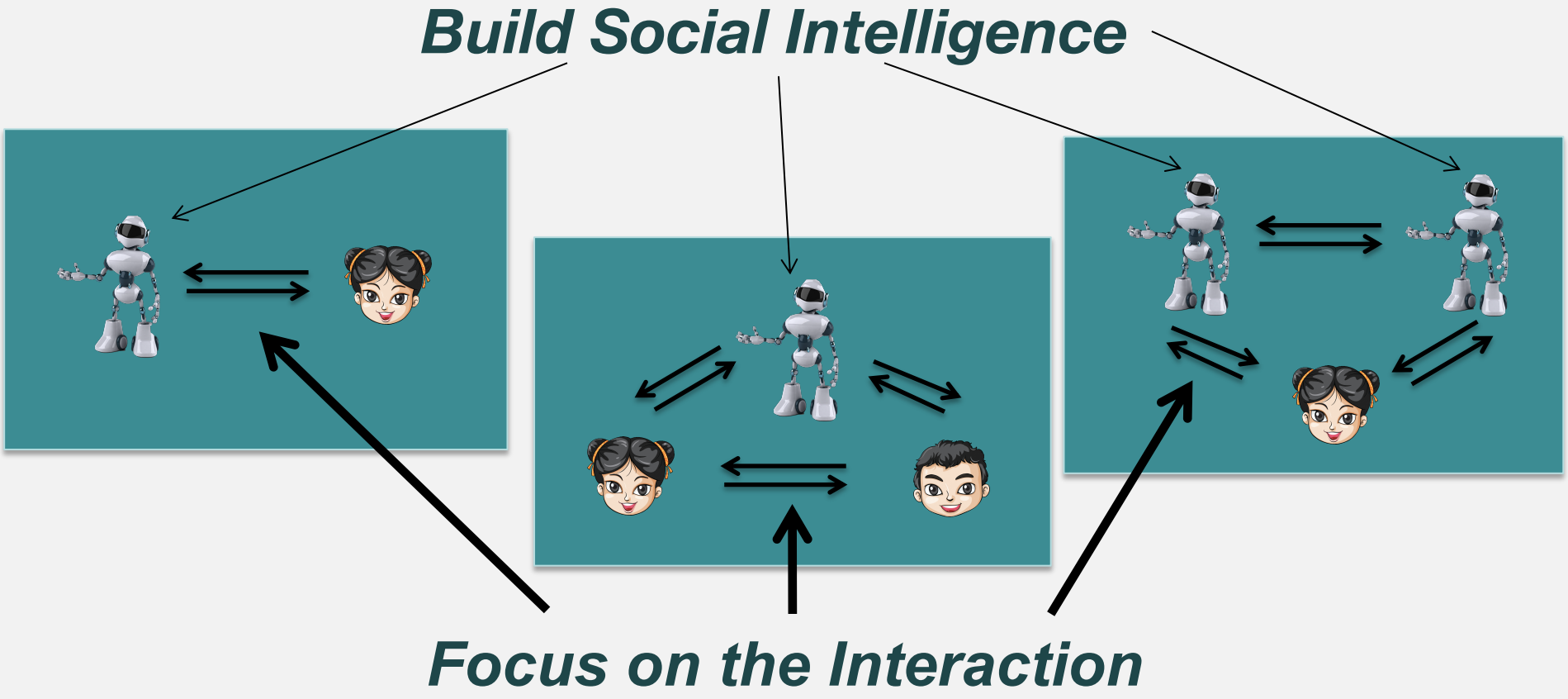
TÉCNICO LISBOA

Our goal…



TÉCNICO LISBOA

Scenarios we are interested…



TÉCNICO LISBOA

The problem

Based on the limited perception of a robot,

TÉCNICO LISBOA

The problem

Based on the limited perception of a robot,
how to build technology

TÉCNICO LISBOA

The problem

Based on the limited perception of a robot, ***how to build technology*** *to understand the social situation*

TÉCNICO LISBOA

The problem

Based on the limited perception of a robot, ***how to build technology*** *to understand the social situation and the user's (and other agents') affective, social, motivational and informational states,*

TÉCNICO LISBOA

The problem

Based on the limited perception of a robot, ***how to build technology*** *to understand the social situation and the user's (and other agents') affective, social, motivational and informational states, in order to respond in a socially appropriate manner.*

TÉCNICO LISBOA

The problem

Machine sensing and automatic understanding of human behaviors

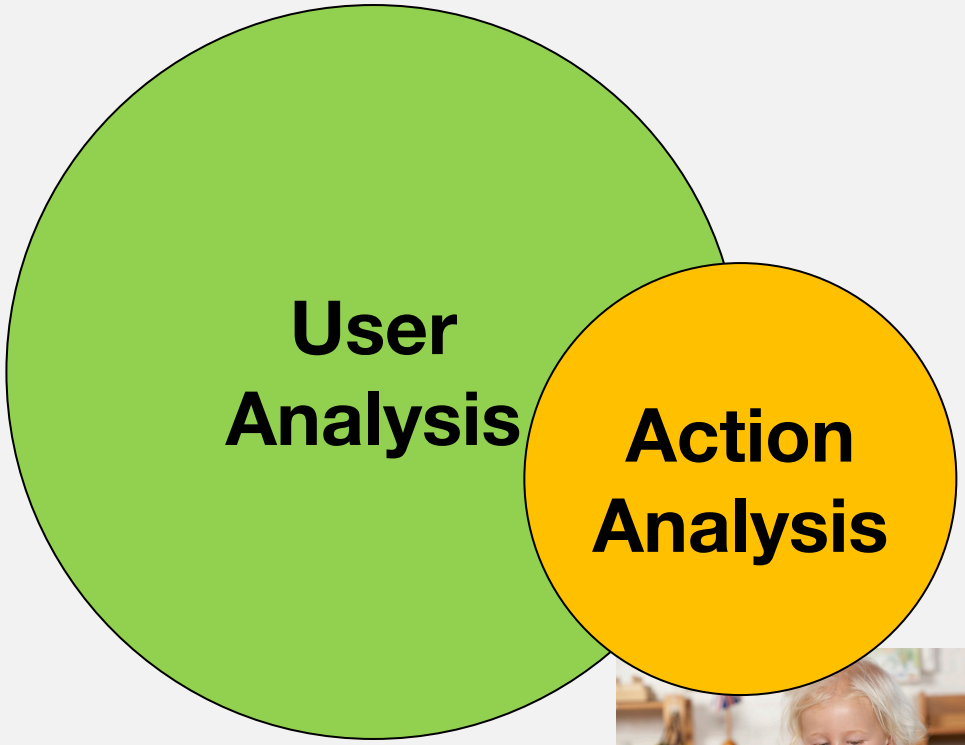
TÉCNICO LISBOA

Perception Levels

User Analysis

TÉCNICO LISBOA

Perception Levels



TÉCNICO LISBOA

Perception Levels

TÉCNICO LISBOA

Beyond the 4 W's

- ***Who?***
- ***Where?***
- ***What?***
- ***How?***

TÉCNICO LISBOA

Beyond the 4 W's

- ***Who?*** (Who is the user?)
- ***Where?***
- ***What?***
- ***How?***

TÉCNICO LISBOA

Beyond the 4 W's

- ***Who?*** (Who is the user?)
- ***Where?*** (Where is the user?)
- ***What?***
- ***How?***

TÉCNICO LISBOA

Beyond the 4 W's

- ***Who?*** (Who is the user?)
- ***Where?*** (Where is the user?)
- ***What?*** (What is the current task of the user?)
- ***How?***

TÉCNICO LISBOA

Beyond the 4 W's

- ***Who?*** (Who is the user?)
- ***Where?*** (Where is the user?)
- ***What?*** (What is the current task of the user?)
- ***How?*** (How is the information passed on? Which behavioral signals have been displayed?)

TÉCNICO LISBOA

Beyond the 4 W's

- ***Who?***
- ***Where?***
- ***What?***
- ***How?***

- ***When?***
- ***Why?***

TÉCNICO LISBOA

Beyond the 4 W's

- ***Who?***
- ***Where?***
- ***What?***
- ***How?***

- ***When?*** (What is the timing of displayed behavioral signals with respect to changes in the environment? Are there any co-occurrences of the signals?)
- ***Why?***

TÉCNICO LISBOA

Beyond the 4 W's

- ***Who?***
- ***Where?***
- ***What?***
- ***How?***

- ***When?*** (What is the timing of displayed behavioral signals with respect to changes in the environment? Are there any co-occurrences of the signals?)
- ***Why?*** (What may be the user's reasons to display the observed cues?)

TÉCNICO LISBOA

Scientific and Engineering Issues*

- ***Which types of messages are communicated (by behavioral/social signals of humans)?***

*Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.

TÉCNICO LISBOA

Scientific and Engineering Issues*

- ***Which types of messages are communicated (by behavioral/social signals of humans)?***
- ***Which human communicative cues convey information about a certain type of behavioral signals?***

*Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.

TÉCNICO LISBOA

Scientific and Engineering Issues*

- ***Which types of messages are communicated (by behavioral/social signals of humans)?***
- ***Which human communicative cues convey information about a certain type of behavioral signals?***
- ***How are various kinds of evidence to be combined to optimize inferences about shown behavioral signals?***

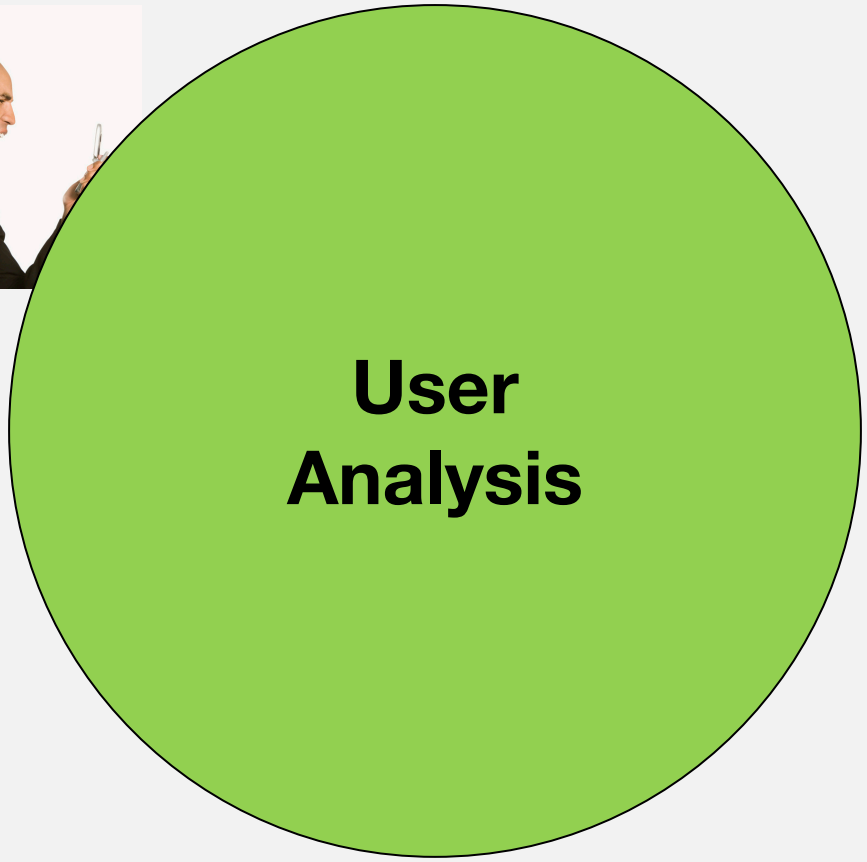
*Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.

TÉCNICO LISBOA

Perception Levels

TÉCNICO LISBOA

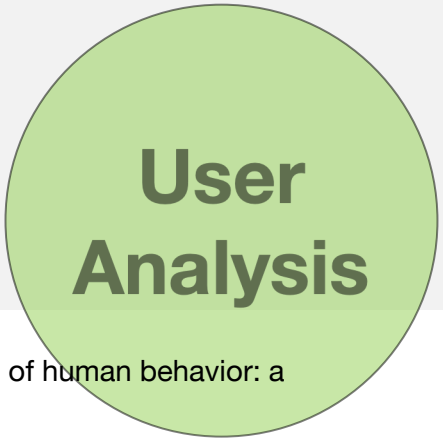
Perception Levels



Human Sensing*

Sensing human behavioral signals is essential in the human judgment of behavioral cues and involves a number of tasks:

- **Face**
  - Face recognition
  - Face detection and tracking
  - Facial expression analysis
  - Gaze tracking



*Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.

TÉCNICO LISBOA

Human Sensing*

Sensing human behavioral signals is essential in the human judgment of behavioral cues and involves a number of tasks:

- **Face**
  - Face recognition
  - Face detection and tracking
  - Facial expression analysis
  - Gaze tracking
- **Body**
  - Body detection and tracking
  - Hand tracking
  - Recognition of posture, gestures and activity

User Analysis

*Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.

TÉCNICO LISBOA

Human Sensing*

Sensing human behavioral signals is essential in the human judgment of behavioral cues and involves a number of tasks:

- **Face**
  - Face recognition
  - Face detection and tracking
  - Facial expression analysis
  - Gaze tracking
- **Body**
  - Body detection and tracking
  - Hand tracking
  - Recognition of posture, gestures and activity
- **Vocal nonlinguistic signals**
  - Estimation of auditory features such as pitch, intensity, and speech rate
  - Recognition of nonlinguistic vocalizations like laughs, cries, sighs, and coughs

User Analysis

*Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.

TÉCNICO LISBOA

Social Signals

- Five types of nonverbal behaviour:
  1. Emblems
     - Gestures directly translated to words
       - E.g. peace sign
     - Different meanings across cultures
       - E.g. giving the finger

User Analysis

- Ekman, Paul, and Wallace V. Friesen. "Nonverbal leakage and clues to deception." Psychiatry 32, no. 1 (1969): 88-106.
- Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.

TÉCNICO LISBOA

Social Signals

- Five types of nonverbal behaviour:
  2. Illustrators/Iconic gestures
     - Gestures or facial expressions that accompany speech to make it vivid, visual, or empathic
       - E.g. illustrating the size of something, throwing a ball, finger pointing, raised eyebrows

User Analysis

- Ekman, Paul, and Wallace V. Friesen. "Nonverbal leakage and clues to deception." Psychiatry 32, no. 1 (1969): 88-106.
- Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.

TÉCNICO LISBOA

Social Signals

- Five types of nonverbal behaviour:
  3. Regulators
     - Nonverbal behaviours used to coordinate conversation
       - E.g. head nodes
       - E.g. looking at/orienting the body towards someone

User Analysis

- Ekman, Paul, and Wallace V. Friesen. "Nonverbal leakage and clues to deception." Psychiatry 32, no. 1 (1969): 88-106.
- Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.

TÉCNICO LISBOA

Social Signals

- Five types of nonverbal behaviour:
  4. Self-adaptor
     - Unconscious behaviours that release nervous energy
       - E.g. touching face, tug hair, bite lips

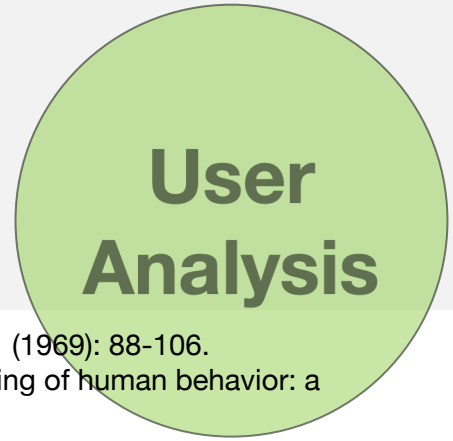
User Analysis

- Ekman, Paul, and Wallace V. Friesen. "Nonverbal leakage and clues to deception." Psychiatry 32, no. 1 (1969): 88-106.
- Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.

TÉCNICO LISBOA

Social Signals

- Five types of nonverbal behaviour:
  5. Displays of emotion
     - face, voice, body, touch



- Ekman, Paul, and Wallace V. Friesen. "Nonverbal leakage and clues to deception." Psychiatry 32, no. 1 (1969): 88-106.
- Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer Berlin Heidelberg.

TÉCNICO LISBOA

Social Signals

Table 2

SELF-DECEPTION: PATIENT A, WITHHOLDING INFORMATION ABOUT SEDUCTIVE, IMMATURE, IMPULSIVE BEHAVIOR, AND SIMULATING COOPERATIVENESS

Head Messages	% *Head*	% *Body*	*Body Messages*	% *Head*	% *Body*	*Head & Body Messages*	% *Head*	% *Body*
Talkative	68	30	Confused	48	83	Emotional	65	83
Alert	65	39	Awkward	47	78	Active	74	74
Cheerful	61	30	Excitable	42	78	Changeable	68	74
Cooperative	59	35	Restless	32	74	Nervous	65	74
Serious	52	22	Impulsive	39	65	Defensive	52	61
			High strung	29	65			
			Feminine	32	65			

Table 3

SELF-DECEPTION: PATIENT B, WITHHOLDING INFORMATION ABOUT CONFUSION, ANXIETY, AND DELUSIONS, SIMULATING WELL-BEING AND HEALTH

Head Messages	% *Head*	% *Body*	*Body Messages*	% *Head*	% *Body*	*Head & Body Messages*	% *Head*	% *Body*
Cooperative	85	36	Tense	18	68	Active	59	53
Friendly	81	25	Nervous	44	64	Changeable	55	53
Cheerful	70	11	Defensive	26	57	Alert	63	50
Sensitive	63	39	Confused	33	53			
Affectionate	59	28	Cautious	30	53			
Appreciative	59	18	Worrying	30	50			
Pleasant	59	11						
Warm	59	18						
Kind	55	32						
Talkative	55	21						
Considerate	52	25						
Good-natured	52	25						
Honest	52	28						

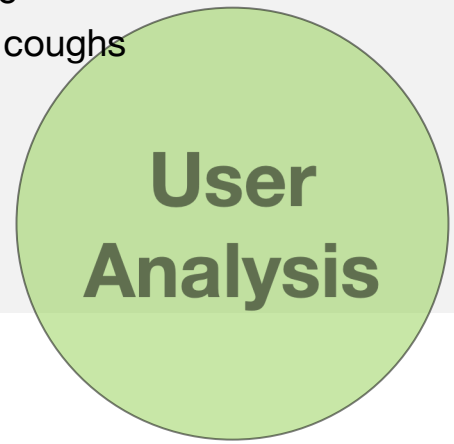
User Analysis

Ekman, Paul, and Wallace V. Friesen. "Nonverbal leakage and clues to deception." Psychiatry 32, no. 1 (1969): 88-106.

TÉCNICO LISBOA

Human Sensing

- **Face**
  - **Face recognition**
  - **Face detection and tracking**
  - Facial expression analysis
  - Gaze tracking
- Body
  - Body detection and tracking
  - Hand tracking
  - Recognition of posture, gestures and activity
- Vocal nonlinguistic signals
  - Estimation of auditory features such as pitch, intensity, and speech rate
  - Recognition of nonlinguistic vocalizations like laughs, cries, sighs, and coughs



TÉCNICO LISBOA

Face Recognition

Tracking and recognizing people are essential skills modern social robots have to be provided with.

User Analysis

TÉCNICO LISBOA

Face Recognition

Tracking and recognizing people are essential skills modern social robots have to be provided with. There are several ways:

1. Automatic face analysis

User Analysis

TÉCNICO LISBOA

Face Recognition

Tracking and recognizing people are essential skills modern social robots have to be provided with. There are several ways:

1. Automatic face analysis
2. People recognition using the human torso (e.g. using color histogram comparison)

User Analysis

TÉCNICO LISBOA

Face Recognition

Tracking and recognizing people are essential skills modern social robots have to be provided with. There are several ways:

1. Automatic face analysis
2. People recognition using the human torso (e.g. using color histogram comparison)
3. People recognition using full body traits (e.g. colour, height, and gait features)

User Analysis

TÉCNICO LISBOA

Face Recognition

Tracking and recognizing people are essential skills modern social robots have to be provided with. There are several ways:

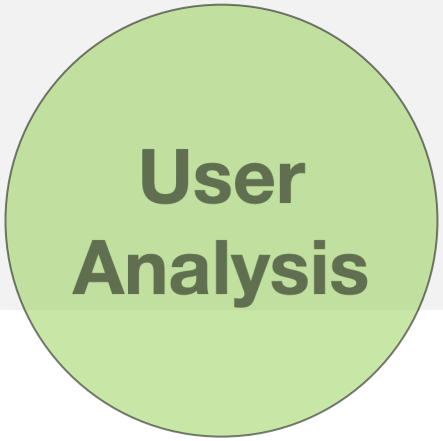
1. Automatic face analysis
2. People recognition using the human torso (e.g. using color histogram comparison)
3. People recognition using full body traits (e.g. colour, height, and gait features)
4. People recognition using the human fingerprint and touch

User Analysis

TÉCNICO LISBOA

Face Recognition

But how to automatize this process?



TÉCNICO LISBOA

Face Recognition

Tracking and recognizing people are essential skills modern social robots have to be provided with. There are several ways:

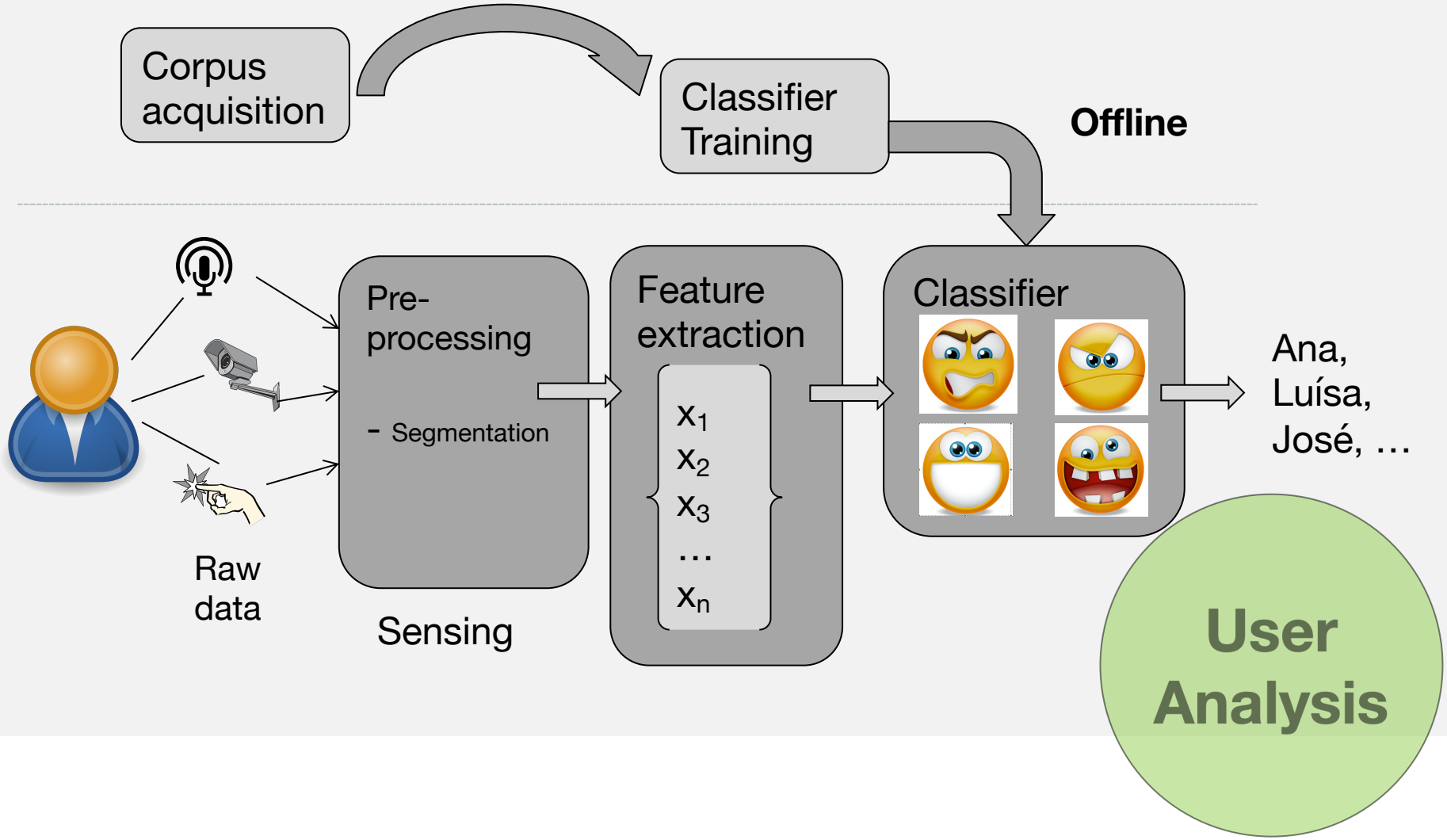
1. **Automatic face analysis** 
2. People recognition using the human torso (e.g. using color histogram comparison)
3. People recognition using full body traits (e.g. colour, height, and gait features)
4. People recognition using the human fingerprint and touch

User Analysis

Cohn, J. F., & De la Torre, F. (2015). Automated face analysis for affective computing. In R. A. Calvo, S. K. D'Mello, J. Gratch, & A. Kappas (Eds.), The Oxford handbook of affective computing (pp. 131–150). Oxford University Press.

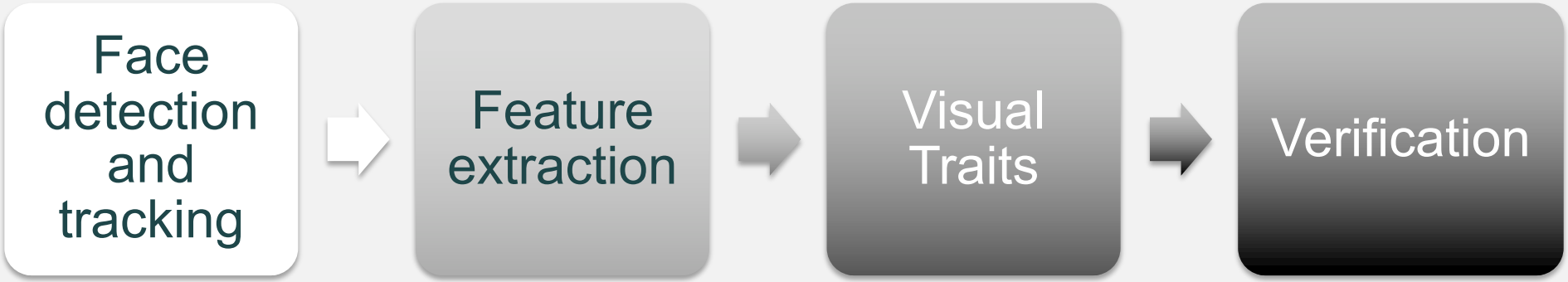
TÉCNICO LISBOA

Face Recognition - **Typical Process**



TÉCNICO LISBOA

Face Recognition

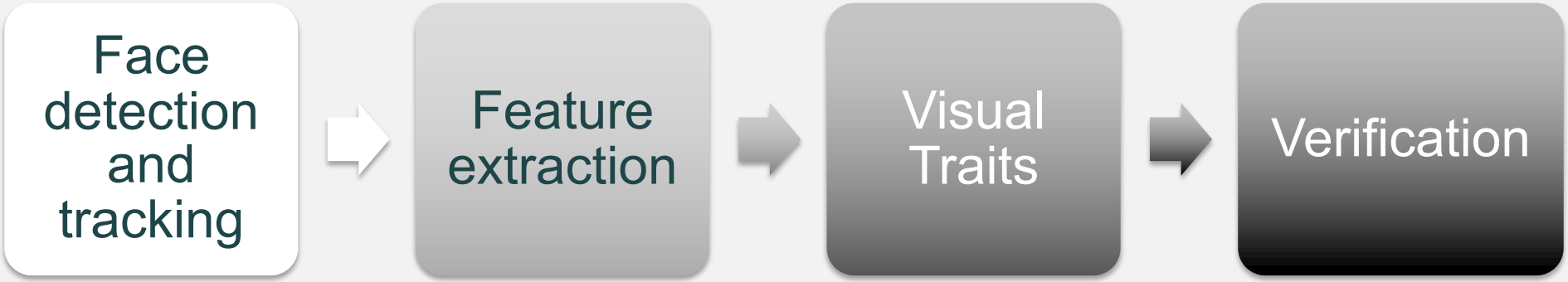


Inspired on different observational methods

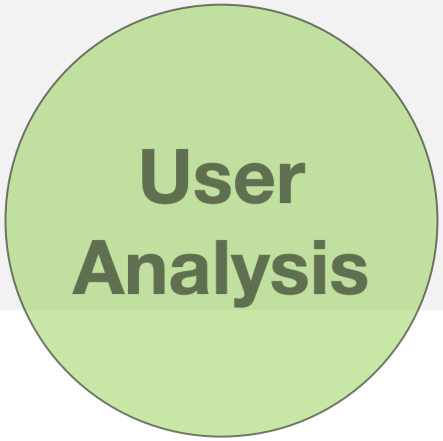
User Analysis

TÉCNICO LISBOA

Face Recognition

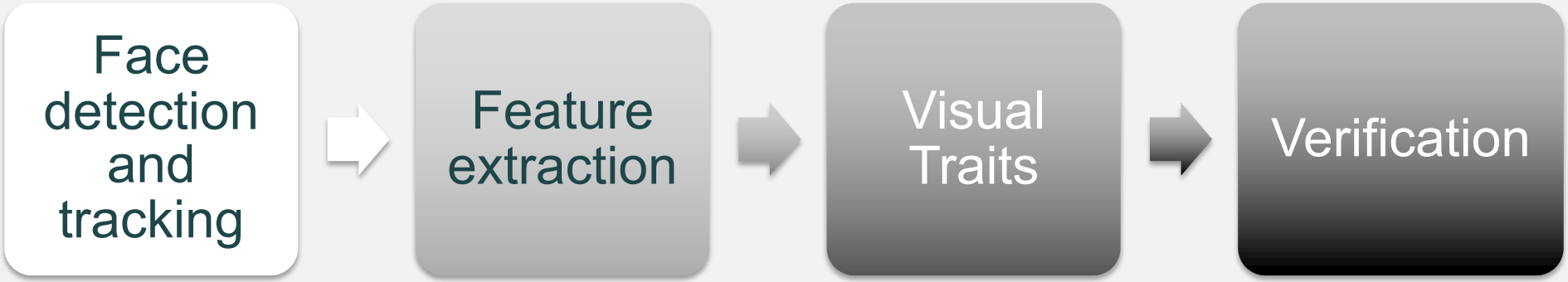


Extract Features: For each face image low-level features are extracted (for example normalized pixel values, image gradient directions) these vectors to form a large feature vector F(I).



TÉCNICO LISBOA

Face Recognition

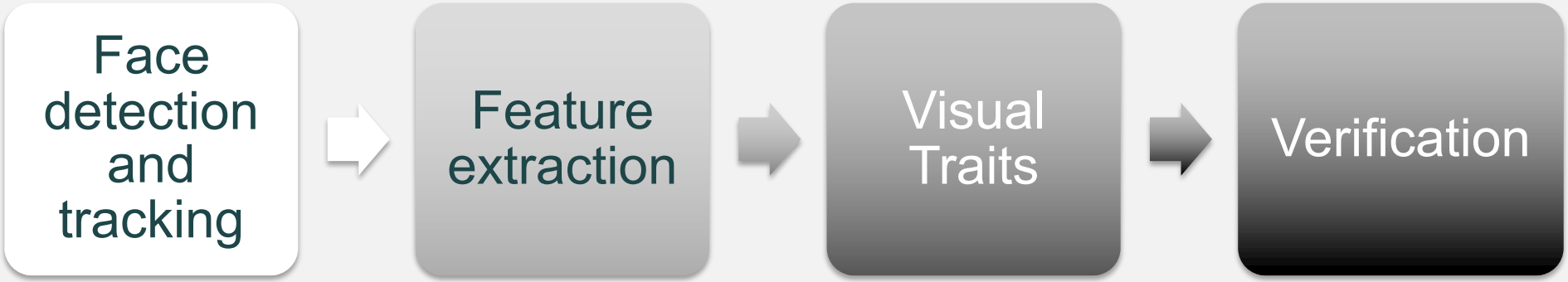


Visual Traits: For each extracted feature vector F(I), the output of *N* traits is calculated based on *N* classifiers in order to produce a "trait vector" C(I) for the face. These classifiers may be focused on attributes such as gender, age, and race, which provide strong cues about a person's identity.

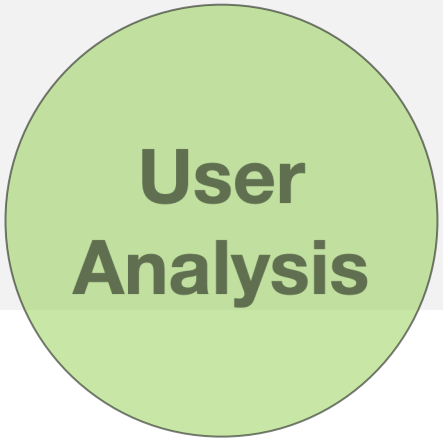
User Analysis

TÉCNICO LISBOA

Face Recognition



Verification: To decide if face matches one already in the system (calculating if the new user is he same person, can be done by comparing their trait vectors using a final classifier D.

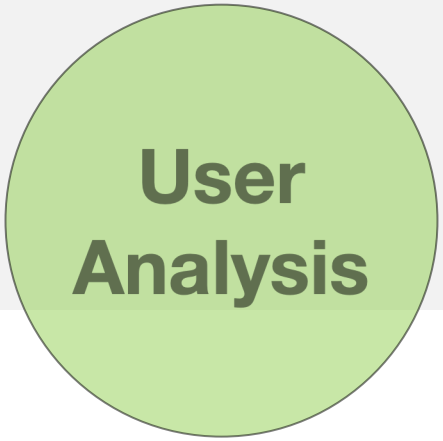


TÉCNICO LISBOA

Face Recognition

Tracking and recognizing people are essential skills modern social robots have to be provided with. There are several ways:

1. Automatic face analysis
2. People recognition using the human torso (e.g. using color histogram comparison)
3. **People recognition using full body traits (e.g. colour, height, and gait features)**
4. People recognition using the human fingerprint and touch



Recognition based on the body: an example

Using the colour of the clothes the height and the gait to do person identification

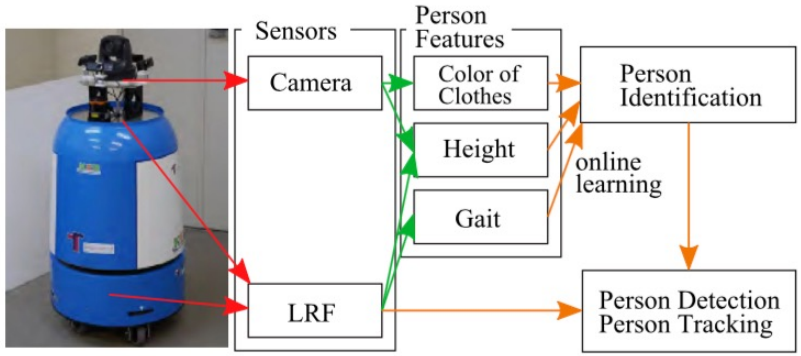


Fig. 3. Person tracking and identification system.

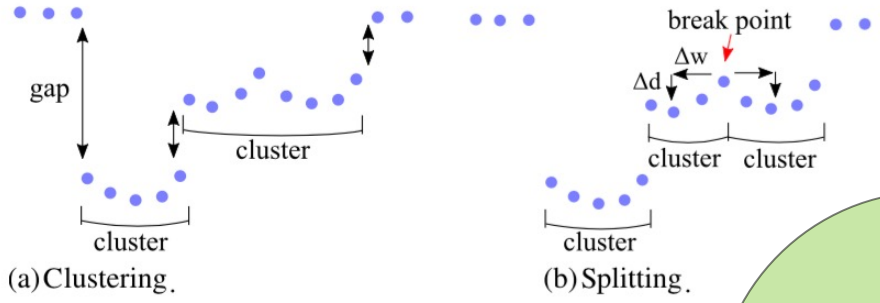


Fig. 4. Torso and leg detection procedure.

User Analysis

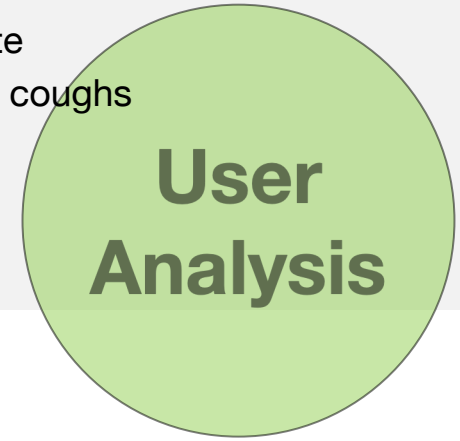
Koide, Kenji, and Jun Miura. "Identification of a specific person using color, height, and gait features for a person following robot." Robotics and Autonomous Systems 84 (2016): 76-87.

TÉCNICO LISBOA

Human Sensing

- **Face**
  - Face recognition
  - Face detection and tracking
  - **Facial expression analysis**
  - Gaze tracking

- Body
  - Body detection and tracking
  - Hand tracking
  - Recognition of posture, gestures and activity
- Vocal nonlinguistic signals
  - Estimation of auditory features such as pitch, intensity, and speech rate
  - Recognition of nonlinguistic vocalizations like laughs, cries, sighs, and coughs



TÉCNICO LISBOA

Emotions in face

User Analysis

TÉCNICO LISBOA

Facial Expressions

- What does a facial expression show?
  - The internal physical state of a person
  - An indication of what he/she is going to do next
  - The plans, expectations and memory
  - The emotional state

User Analysis

TÉCNICO LISBOA

Facial Expressions of Emotion

- Markers of Emotional Expression
  - **Involuntary muscle actions** that people cannot deliberately produce/suppress
  - **Usually last a couple of seconds**
    - smile with enjoyment - 10 seconds
    - polite smile without emotion (exceptionally brief ¼ second or it can be enforced during long periods of time)

User Analysis

Facial Expression of Emotion

- Markers of Emotional Expression
  - E.g. Duchenne smile

Non Duchenne

Duchenne

User Analysis

Facial Expression of Emotion

- Markers of Emotional Expression
  - E.g. Duchenne smile

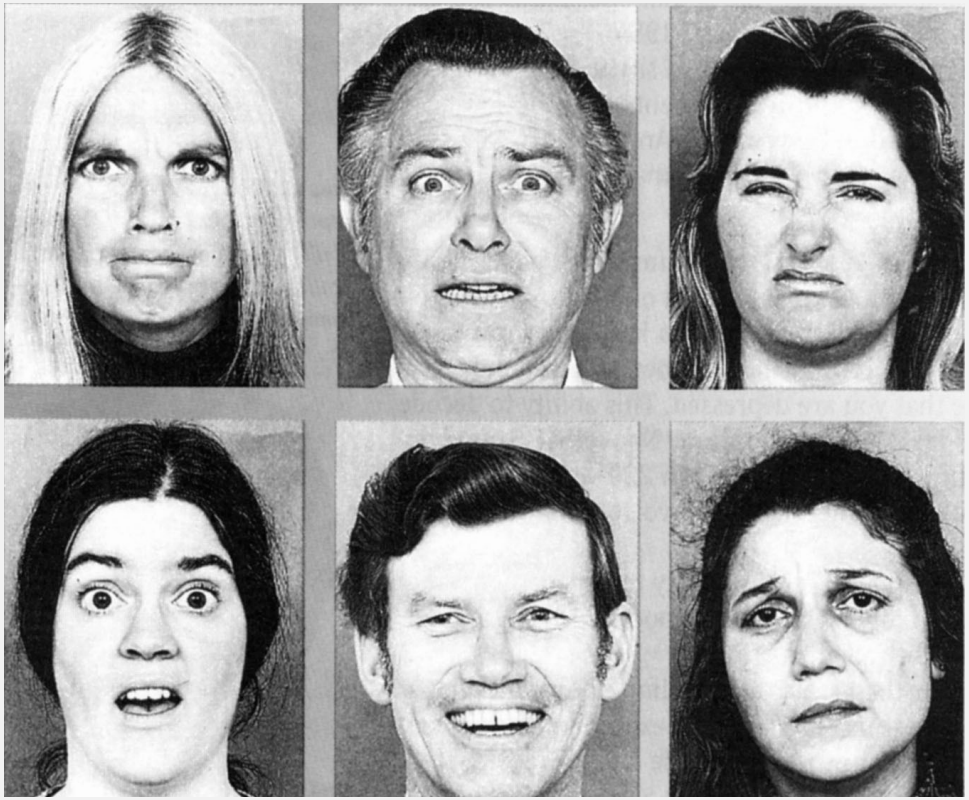
Non Duchenne

Duchenne

User Analysis

TÉCNICO LISBOA

Universality of Facial Expressions

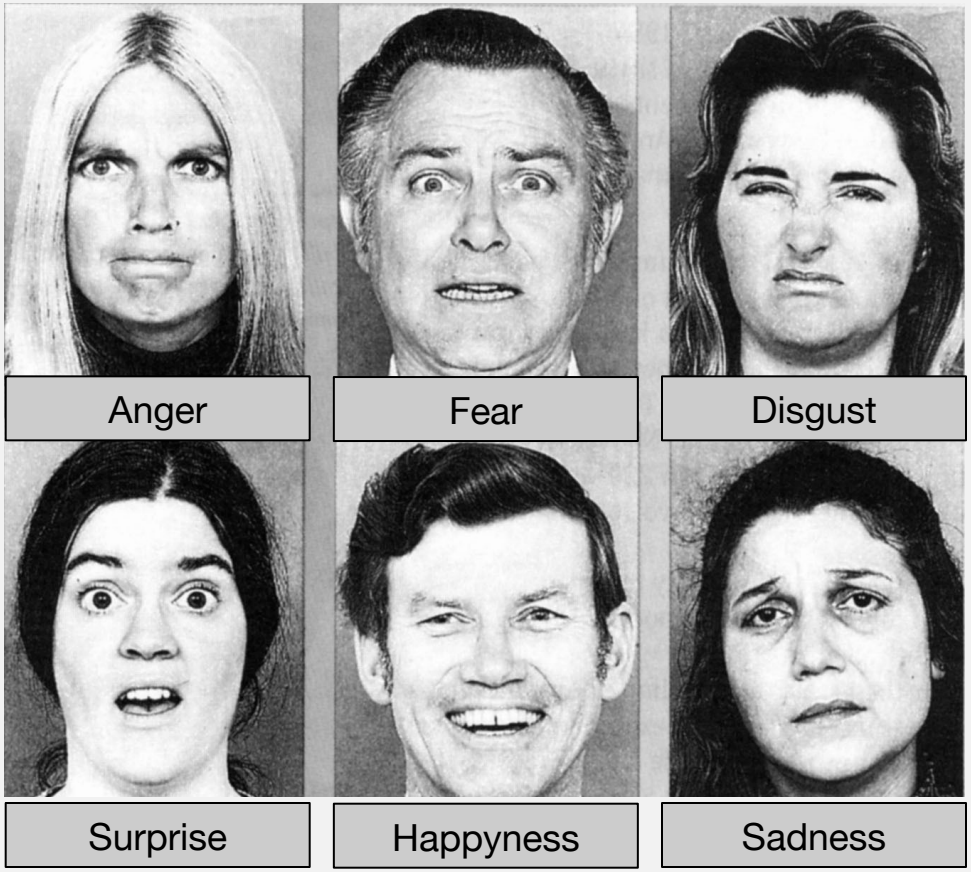


- First test of the universal hypothesis
  - 3000 photos of different people
  - 6 basic emotions

User Analysis

TÉCNICO LISBOA

Universality of Facial Expressions



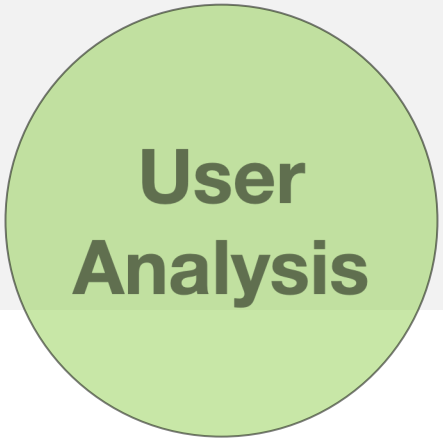
- First test of the universal hypothesis
  - 3000 photos of different people
  - 6 basic emotions

User Analysis

TÉCNICO LISBOA

Universality of Facial Expressions: Ekman's studies

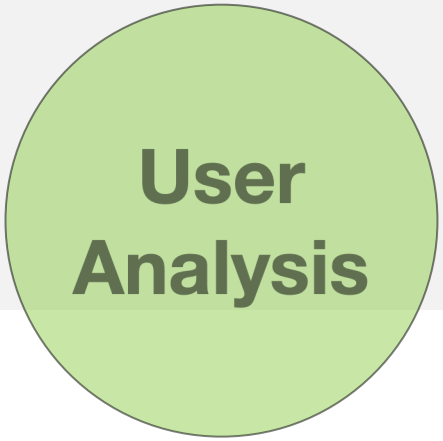
- First test of the universal hypothesis
  - Photos showed to participants across countries
    - Japan, Brasil, Argentina, Chile, U.S



TÉCNICO LISBOA

Universality of Facial Expressions: Ekman's studies

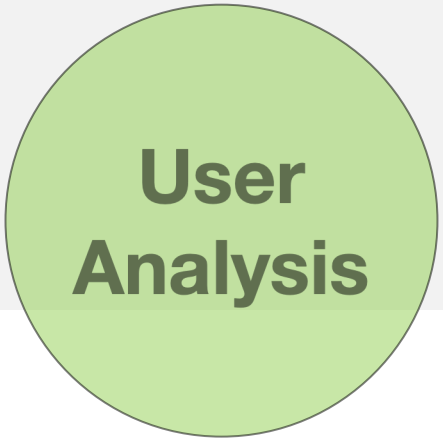
- First test of the universal hypothesis
  - Photos showed to participants across countries
    - Japan, Brasil, Argentina, Chile, U.S
  - Participants were asked to select the emotion term that better matched the displayed emotion
    - From a list of 6 terms



TÉCNICO LISBOA

Universality of Facial Expressions: Ekman's studies

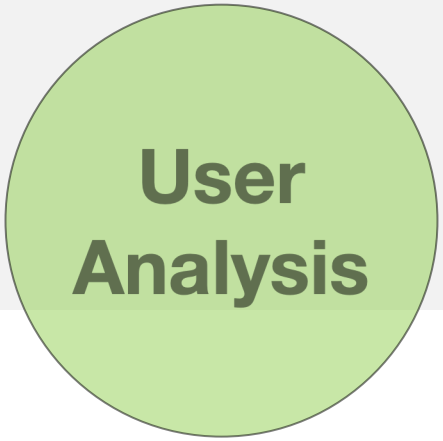
- First test of the universal hypothesis
  - Photos showed to participants across countries
    - Japan, Brasil, Argentina, Chile, U.S
  - Participants were asked to select the emotion term that better matched the displayed emotion
    - From a list of 6 terms
  - Accuracy rates of 80-90%
    - In all countries



TÉCNICO LISBOA

Universality of Facial Expressions: Ekman's studies

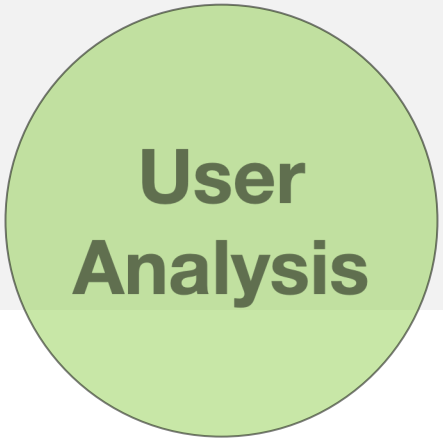
- Critics to this first experiment
  - Participants had seen U.S. television and movies and might have learned american labels for the expressions



TÉCNICO LISBOA

Universality of Facial Expressions: Ekman's studies

- Second experiment
  - Ekman travelled to Papua, new Guinea
    - Lived 6 months with people of the Fore tribe
      - did not see any movie or magazine
      - did not speak English
      - minimal exposure to westerners
  - Two tasks were designed to evaluate the universality hypothesis



TÉCNICO LISBOA

Universality of Facial Expressions: Ekman's studies

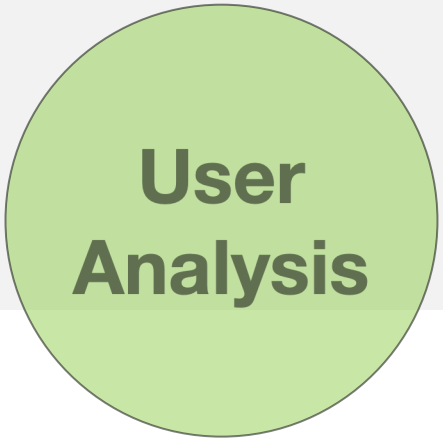
	Fore participants judging western photos		U.S students judging Fore expressions
	Adults	Children	
Anger	84	90	51
Disgust	81	85	46
Fear	80	93	18
Happiness	92	92	73
Sadness	79	91	68
Surprise	68	98	27

User Analysis

TÉCNICO LISBOA

Universality of Facial Expressions

- Other studies about universality of facial expressions further confirmed these results
  - Ekman, 1984, 1993
  - Elfenbein & Ambady 2002,2003
  - Izard 1971, 1994



TÉCNICO LISBOA

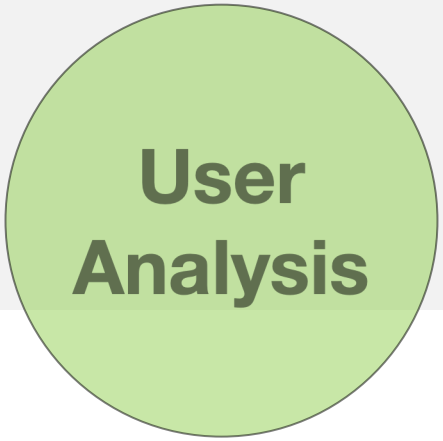
How to measure emotions in a face?

Sign-based measurements:

- Purely descriptive approach
- Relation between those signs and emotions

The most used method is

Facial Action Coding System (FACS)



TÉCNICO LISBOA

FACS

The FACS taxonomy describes 44 Action Units (AU) that can be coded binary (presence/absence) or with a number (corresponding its intensity).

Upper face action units					
AU1	AU2	AU4	AU5	AU6	AU7
					
Inner brow raiser	Outer brow raiser	Brow lowerer	Upper lid raiser	Cheek raiser	Lid tightener
*AU41	*AU42	*AU43	AU44	AU45	AU46
					
Lip droop	Slit	Eyes closed	Squint	Blink	Wink

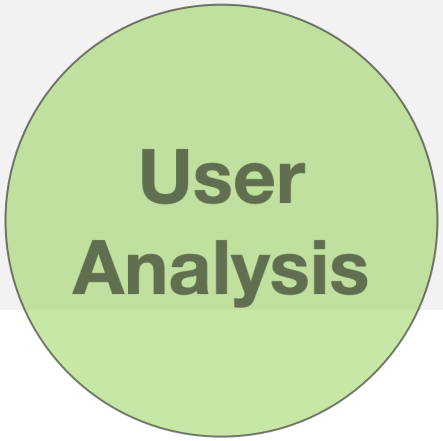
Lower face action units					
AU9	AU10	AU11	AU12	AU13	AU14
					
Nose wrinkler	Upper lip raiser	Nasolabial deepener	Lip corner puller	Cheek puffer	Dimpler

User Analysis

TÉCNICO LISBOA

FACS

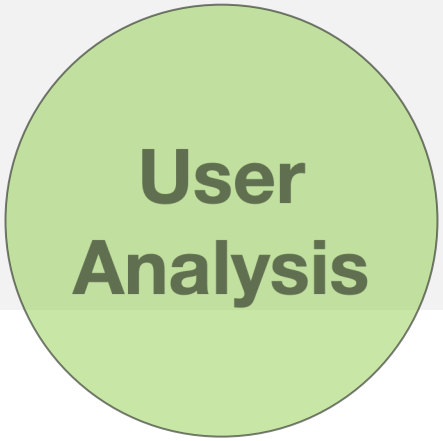
- Designed to detect subtle changes in facial features
- Viewing videotaped facial behaviour in slow motion, trained observers can manually FACS code all possible facial displays, which are referred to as Action Units (AU)



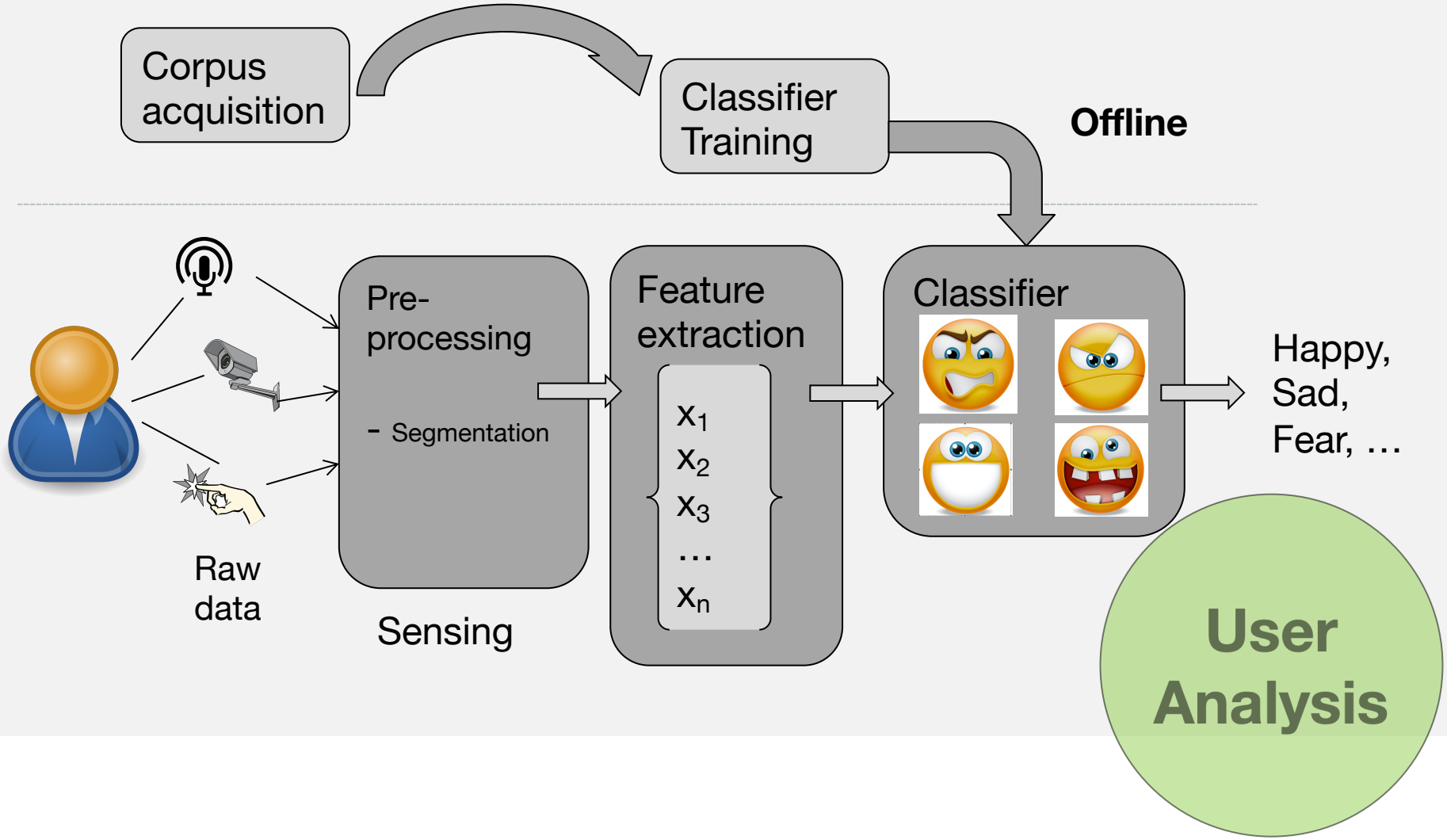
TÉCNICO LISBOA

Facial Expression of Emotion

But how to automatize this process?

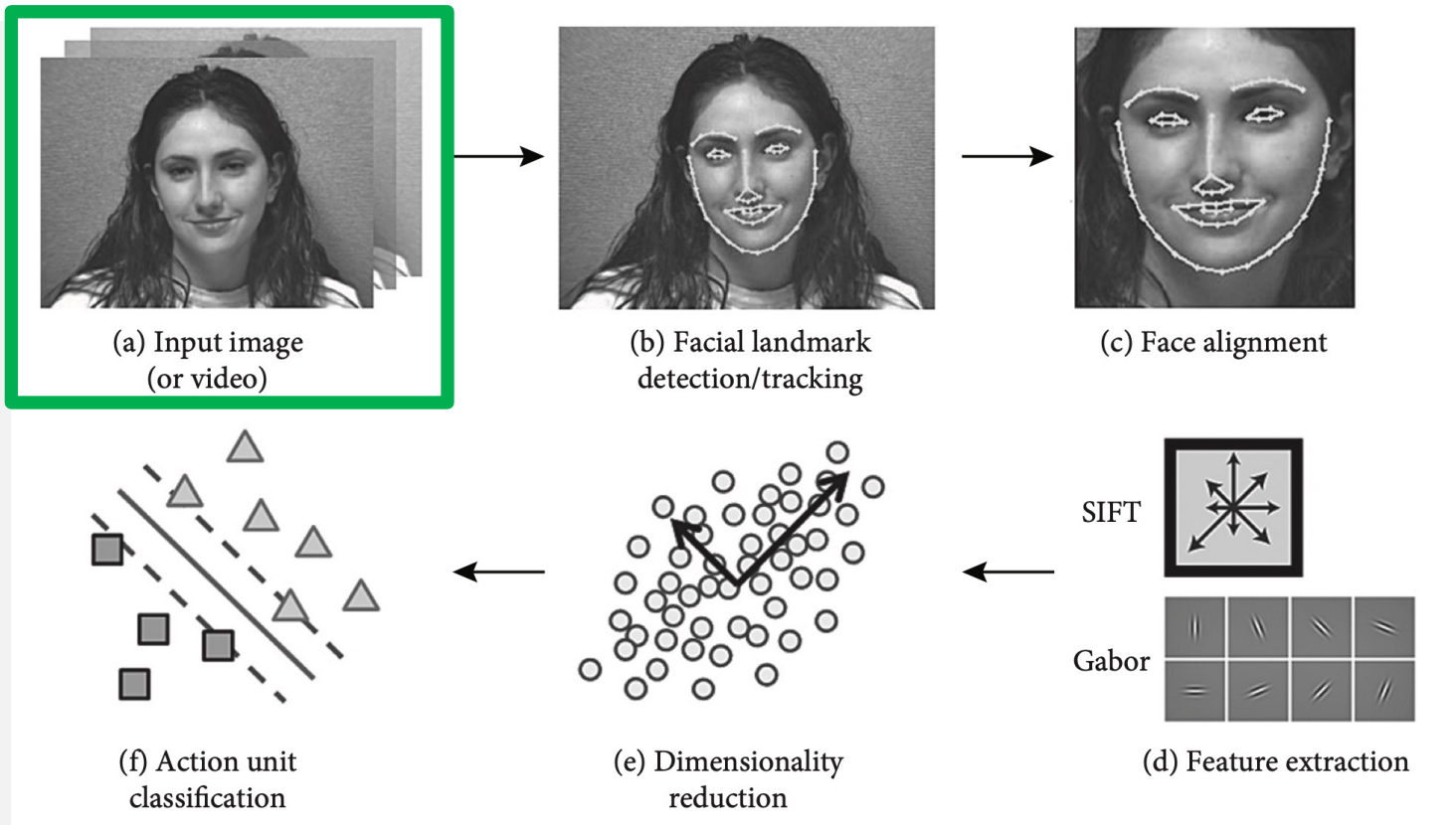


Facial Expression of Emotion - **Typical Process**



TÉCNICO LISBOA

Automated Face Analysis (AFA)

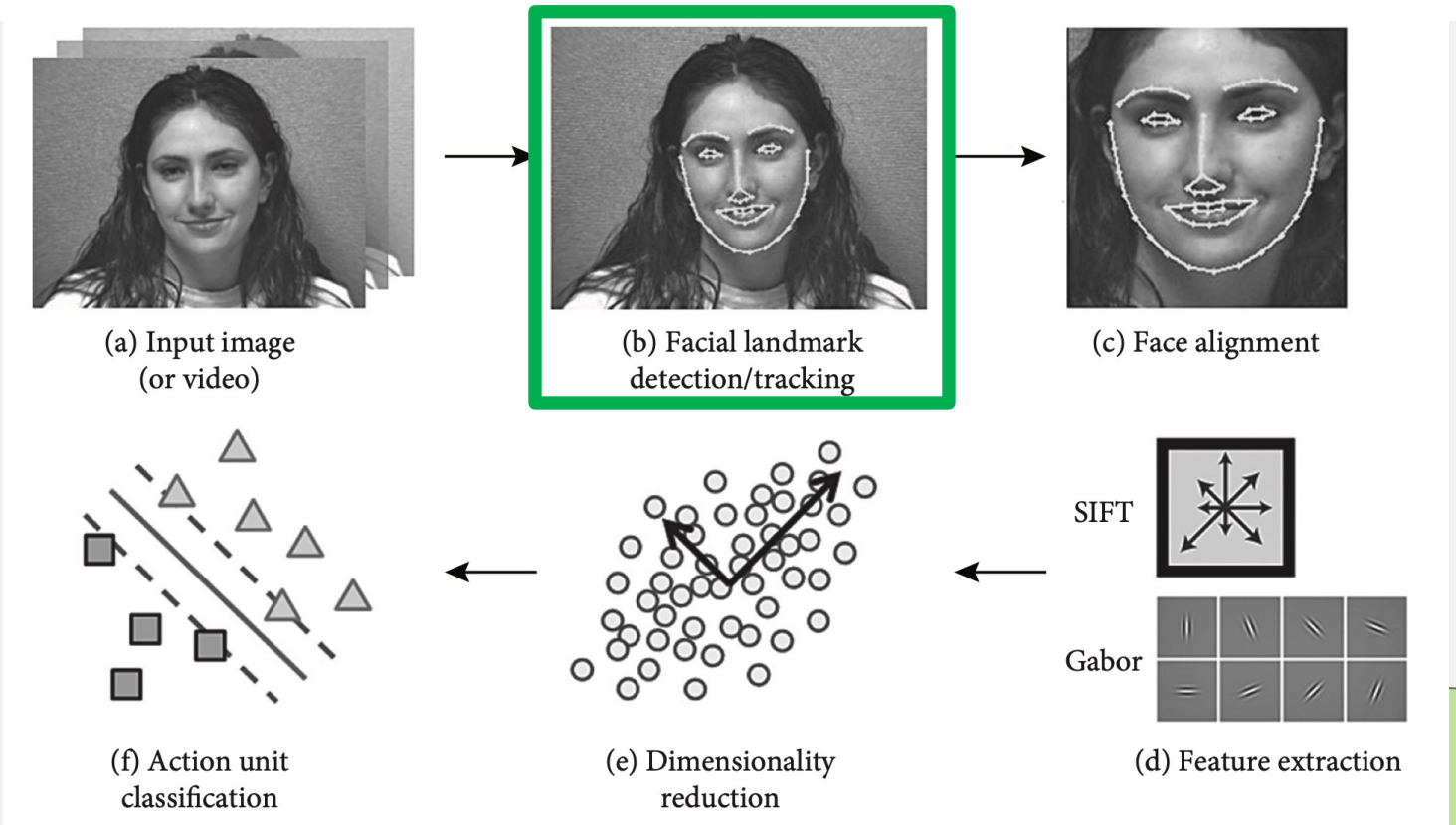


User Analysis

Cohn, J. F., & De la Torre, F. (2015). Automated face analysis for affective computing. In R. A. Calvo, S. K. D'Mello, J. Gratch, & A. Kappas (Eds.), The Oxford handbook of affective computing (pp. 131–150). Oxford University Press.

TÉCNICO LISBOA

Automated Face Analysis (AFA)



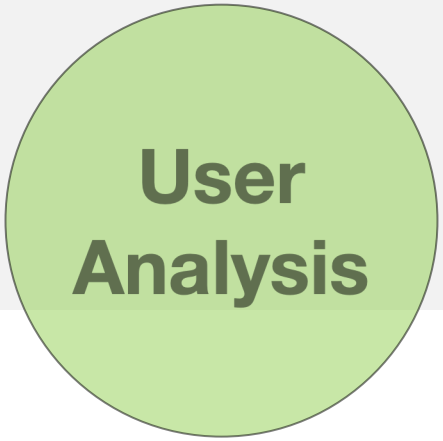
User Analysis

Cohn, J. F., & De la Torre, F. (2015). Automated face analysis for affective computing. In R. A. Calvo, S. K. D'Mello, J. Gratch, & A. Kappas (Eds.), The Oxford handbook of affective computing (pp. 131–150). Oxford University Press.

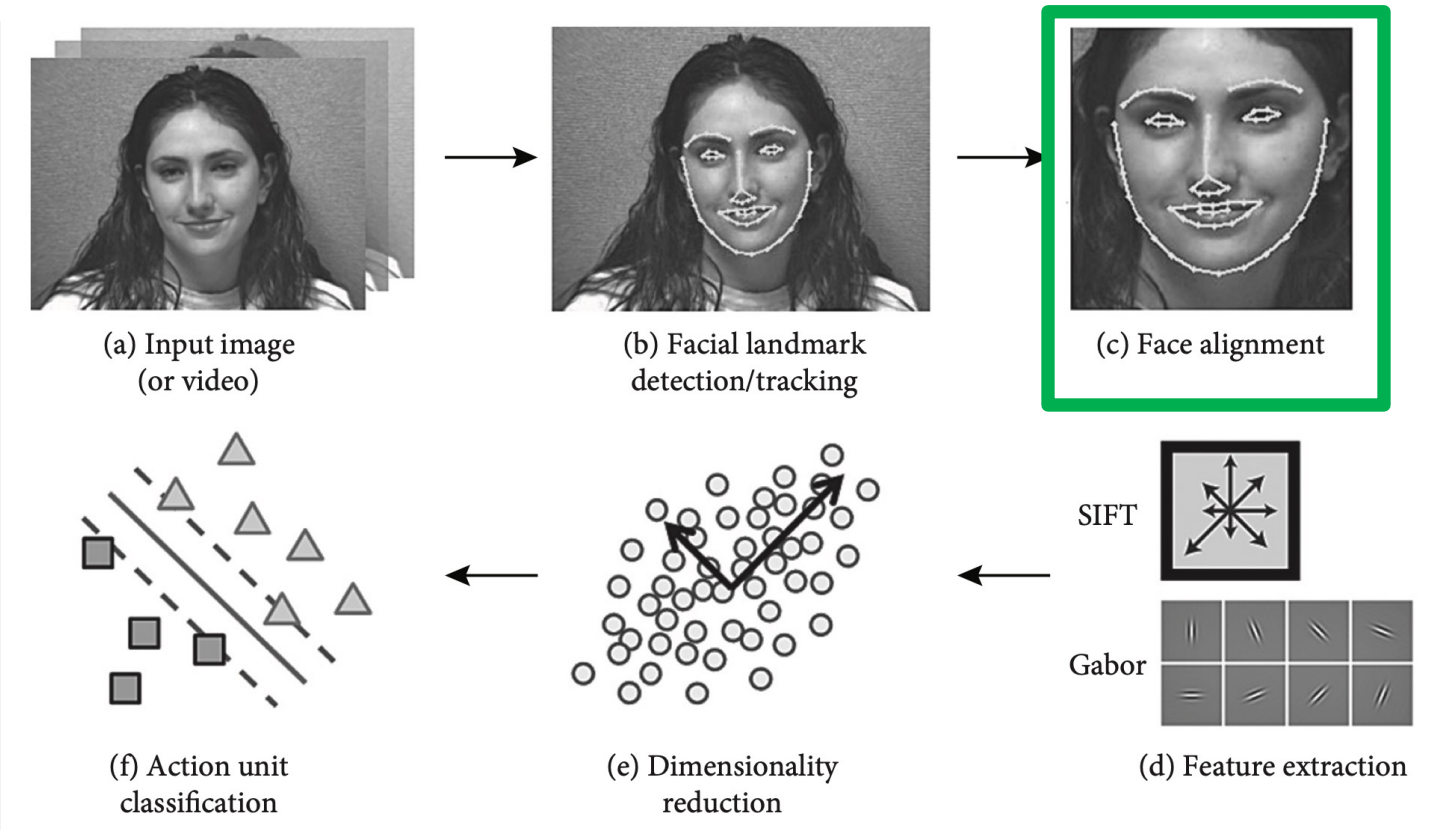
TÉCNICO LISBOA

Facial Landmark Detection / Tracking

- Track a dense set of facial features
- Decouple the “shape” and “appearance” of a face image



Automated Face Analysis (AFA)



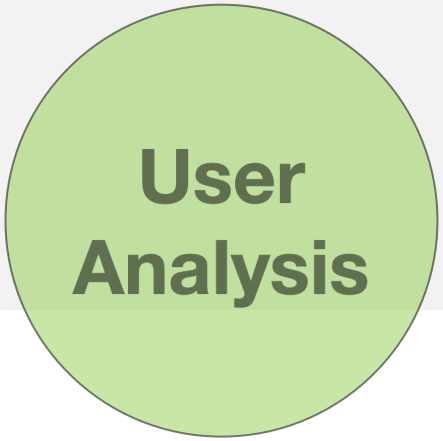
User Analysis

Cohn, J. F., & De la Torre, F. (2015). Automated face analysis for affective computing. In R. A. Calvo, S. K. D'Mello, J. Gratch, & A. Kappas (Eds.), The Oxford handbook of affective computing (pp. 131–150). Oxford University Press.

TÉCNICO LISBOA

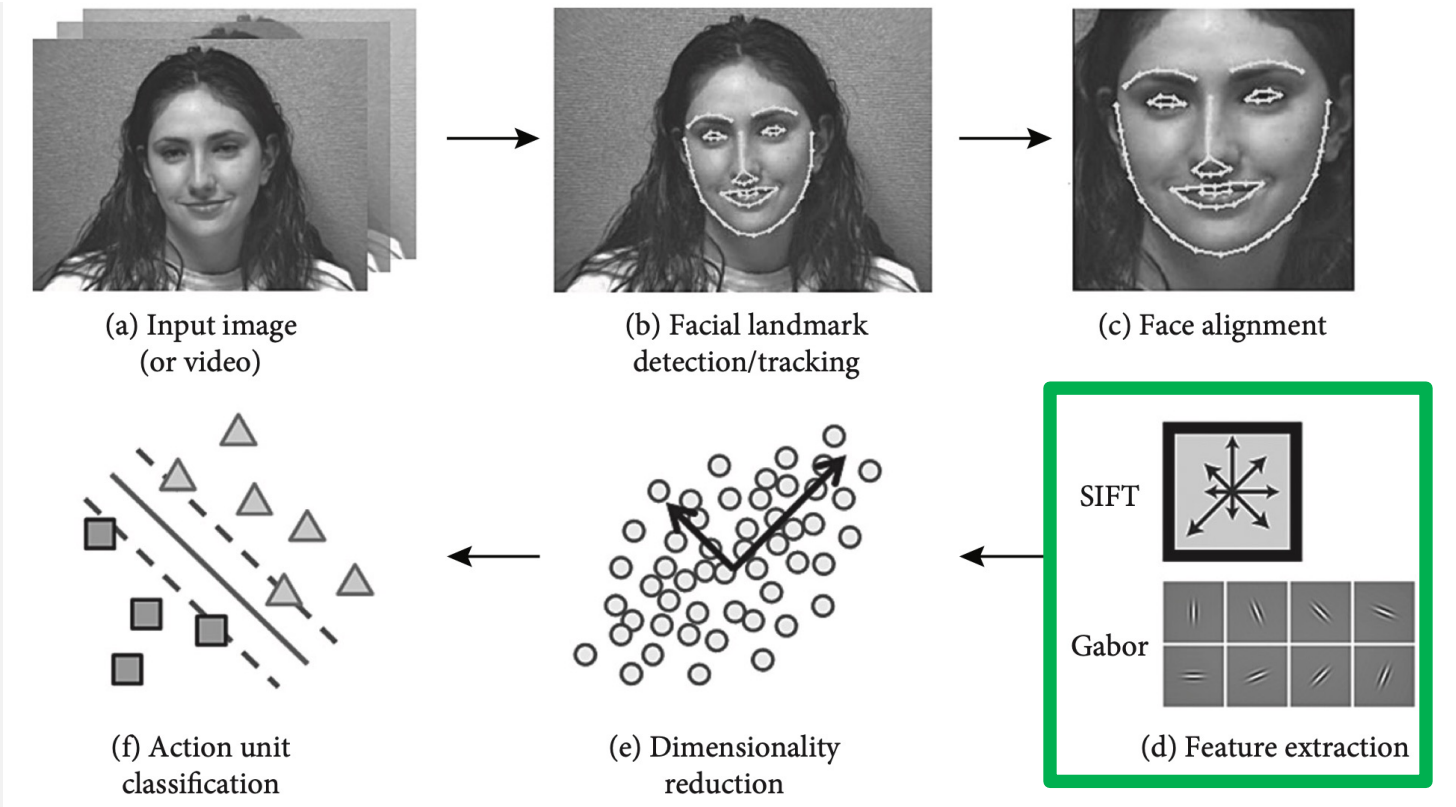
Face Alignment

- Remove the effects of spatial variations (position, rotation, proportions)
- Obtain the canonical size and orientation
- Align the shape model to an unseen image containing the face and facial expression of interest



TÉCNICO LISBOA

Automated Face Analysis (AFA)



User Analysis

Cohn, J. F., & De la Torre, F. (2015). Automated face analysis for affective computing. In R. A. Calvo, S. K. D'Mello, J. Gratch, & A. Kappas (Eds.), The Oxford handbook of affective computing (pp. 131–150). Oxford University Press.

TÉCNICO LISBOA

Feature Extraction: Geometric Features

- Combination of facial landmarks
- Relation between landmarks (distance and ratios)

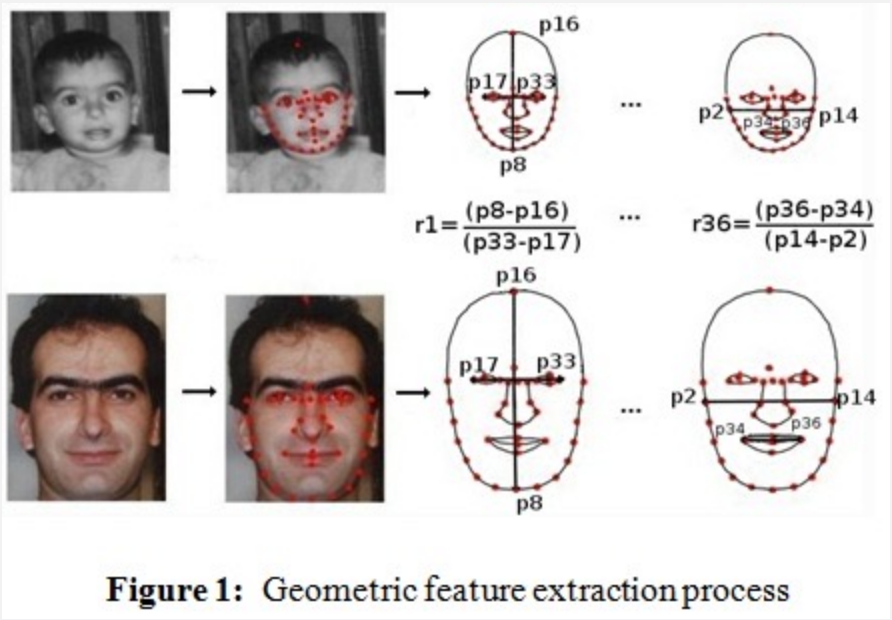


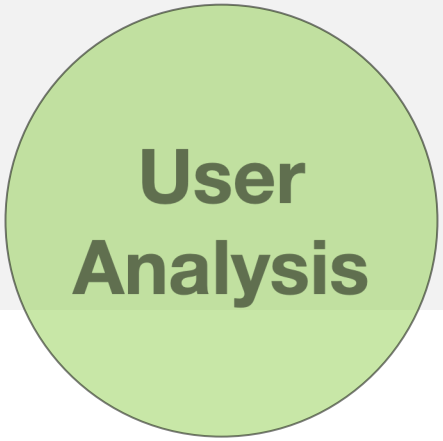
Figure 1: Geometric feature extraction process

User Analysis

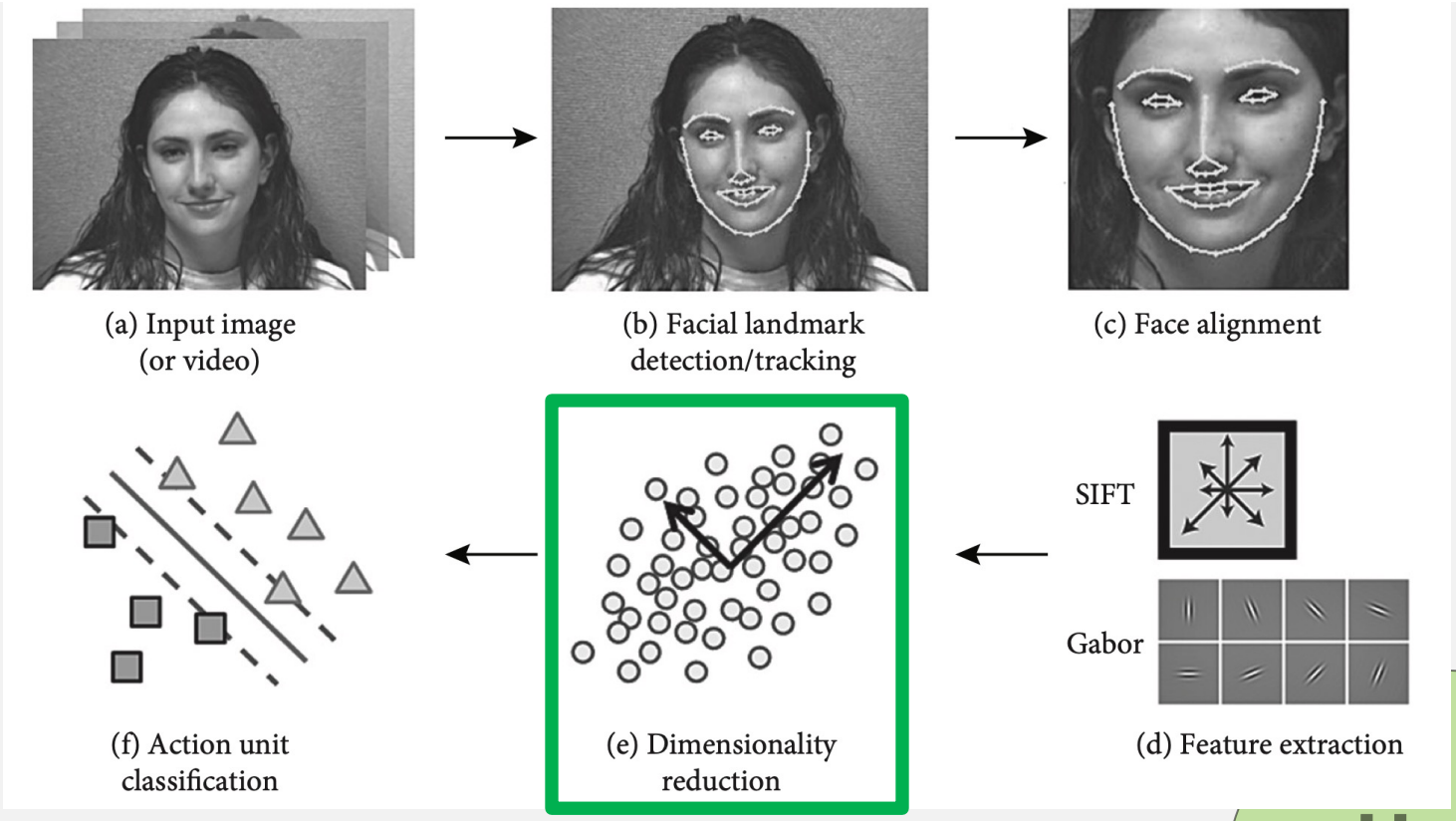
TÉCNICO LISBOA

Feature Extraction: Motion Features

- Optical flow (used to estimate activity in a subset of the facial muscle)
- Dynamic textures or motion history images (compress into one frame the motion over a number of consecutive ones)



Automated Face Analysis (AFA)



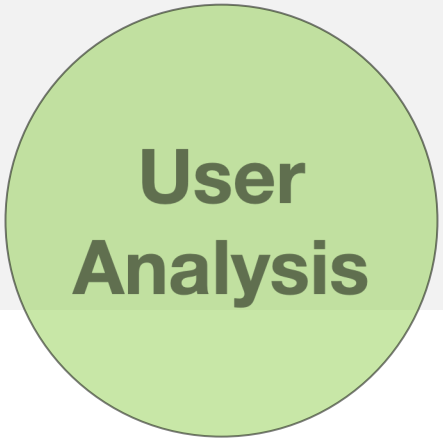
User Analysis

Cohn, J. F., & De la Torre, F. (2015). Automated face analysis for affective computing. In R. A. Calvo, S. K. D'Mello, J. Gratch, & A. Kappas (Eds.), The Oxford handbook of affective computing (pp. 131–150). Oxford University Press.

TÉCNICO LISBOA

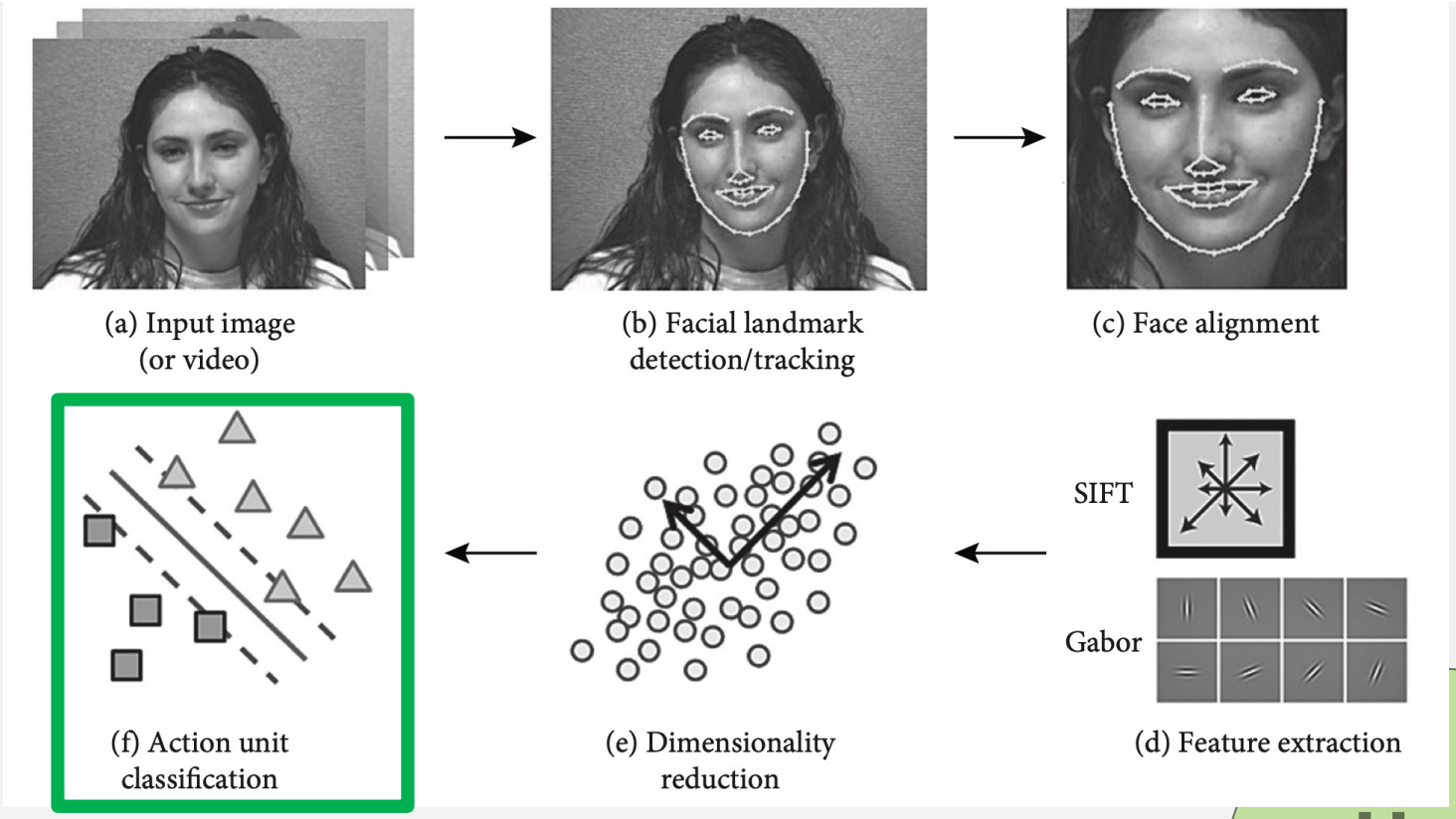
Dimensionality reduction

- Possible techniques and algorithms:
  - Linear techniques are principal components analysis (PCA), Kernel PCA, and independent components analysis
  - Nonlinear techniques include Laplacian eigenmaps and local linear embedding (LLE)



TÉCNICO LISBOA

Automated Face Analysis (AFA)



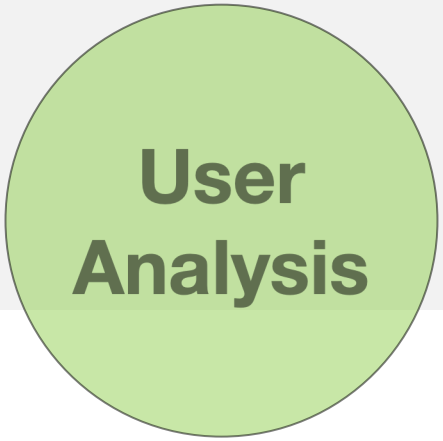
User Analysis

Cohn, J. F., & De la Torre, F. (2015). Automated face analysis for affective computing. In R. A. Calvo, S. K. D'Mello, J. Gratch, & A. Kappas (Eds.), The Oxford handbook of affective computing (pp. 131–150). Oxford University Press.

TÉCNICO LISBOA

Action Unit Classification

Most approaches use **supervised learning**, which means the categories (e.g. emotion labels or AUs) are obtained in advance, in labelled training data.

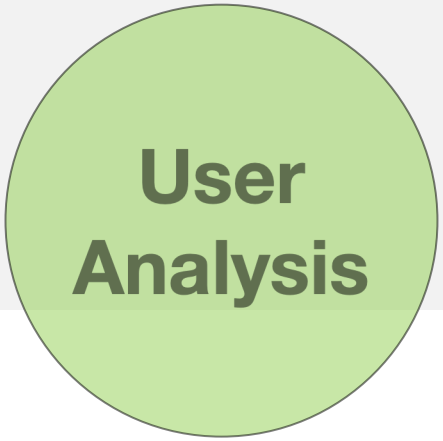


TÉCNICO LISBOA

Action Unit Classification

Two approaches to supervised learning are:

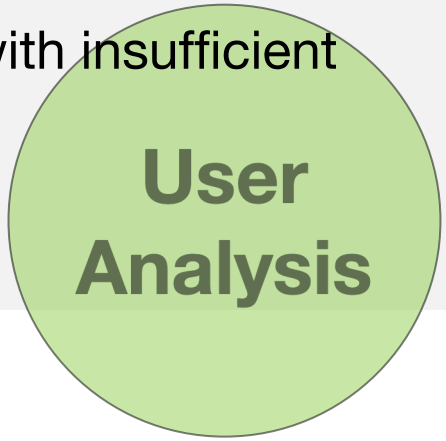
1. **Static modelling**—typically posed as a discriminative classification problem in which each video frame is evaluated independently
2. **Temporal modelling**—frames are segmented into sequences and typically modelled with a variant of dynamic Bayesian networks (e.g., hidden Markov models, conditional random fields).



TÉCNICO LISBOA

Facial Expression of Emotion - **Challenges**

1. Non-frontal pose and moderate to large head motion make facial image registration difficult;
2. Many facial actions are inherently subtle, making them difficult to model;
3. The temporal dynamics of actions can be highly variable;
4. Discrete AUs can modify each other's appearance (i.e., non-additive combinations);
5. Individual differences in face shape and appearance undermine generalization across subjects; and
6. Classifiers can suffer from overfitting when trained with insufficient examples.

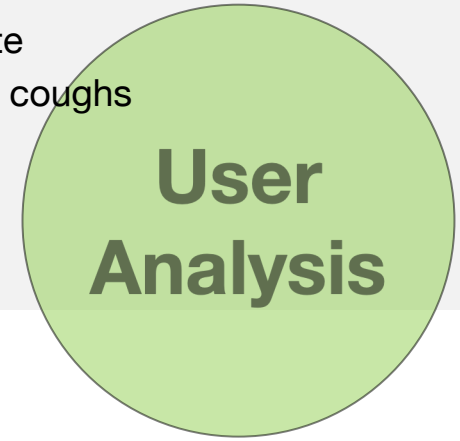


TÉCNICO LISBOA

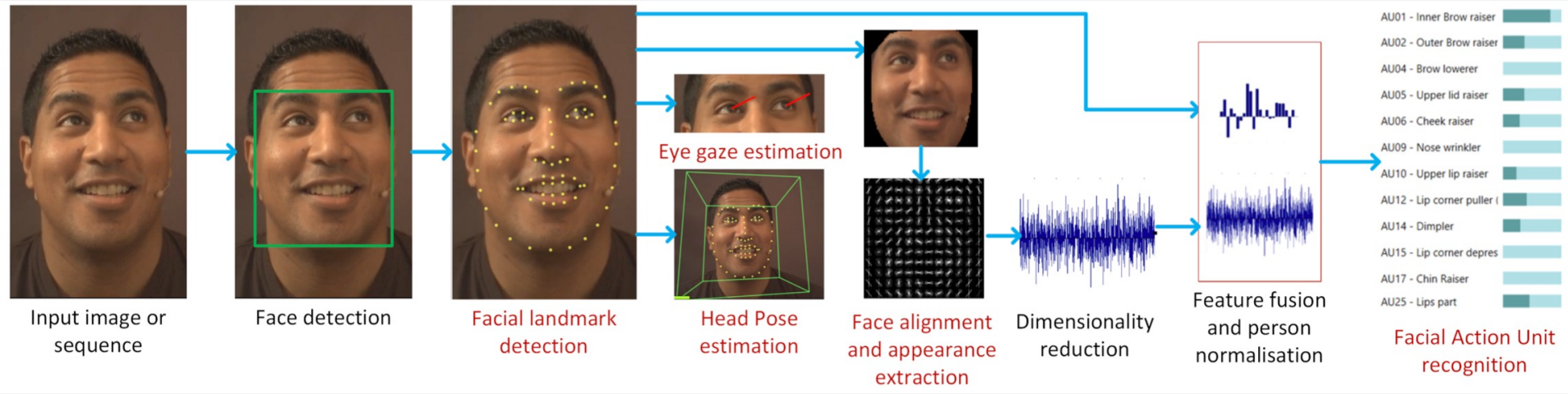
Human Sensing

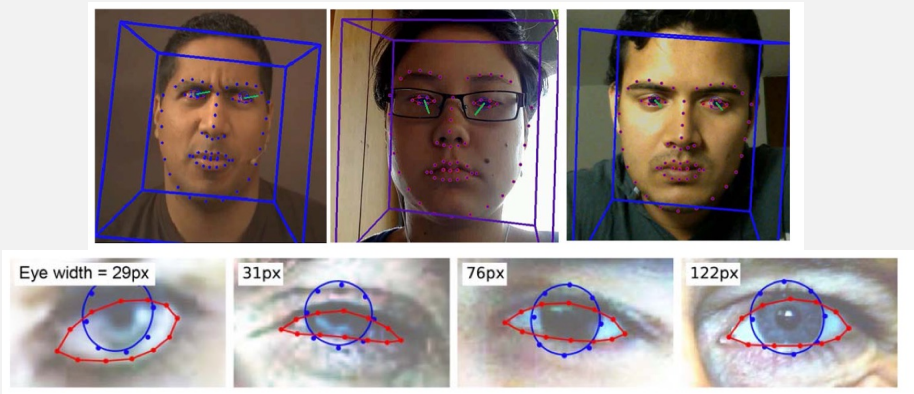
- **Face**
  - Face recognition
  - Face detection and tracking
  - Facial expression analysis
  - **Gaze tracking**

- Body
  - Body detection and tracking
  - Hand tracking
  - Recognition of posture, gestures and activity
- Vocal nonlinguistic signals
  - Estimation of auditory features such as pitch, intensity, and speech rate
  - Recognition of nonlinguistic vocalizations like laughs, cries, sighs, and coughs



Gaze Tracking – Example: OpenFace





Baltrusaitis, Tadas, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. "Openface 2.0: Facial behavior analysis toolkit." In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), pp. 59-66. IEEE, 2018.

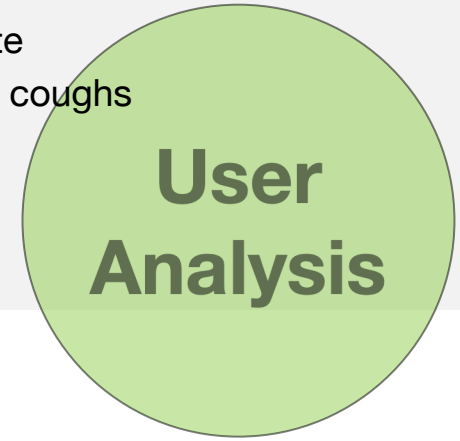
TÉCNICO LISBOA

Human Sensing

- **Face**
  - Face recognition
  - Face detection and tracking
  - Facial expression analysis
  - Gaze tracking

- Body
  - **Body detection and tracking**
  - **Hand tracking**
  - **Recognition of posture, gestures and activity**

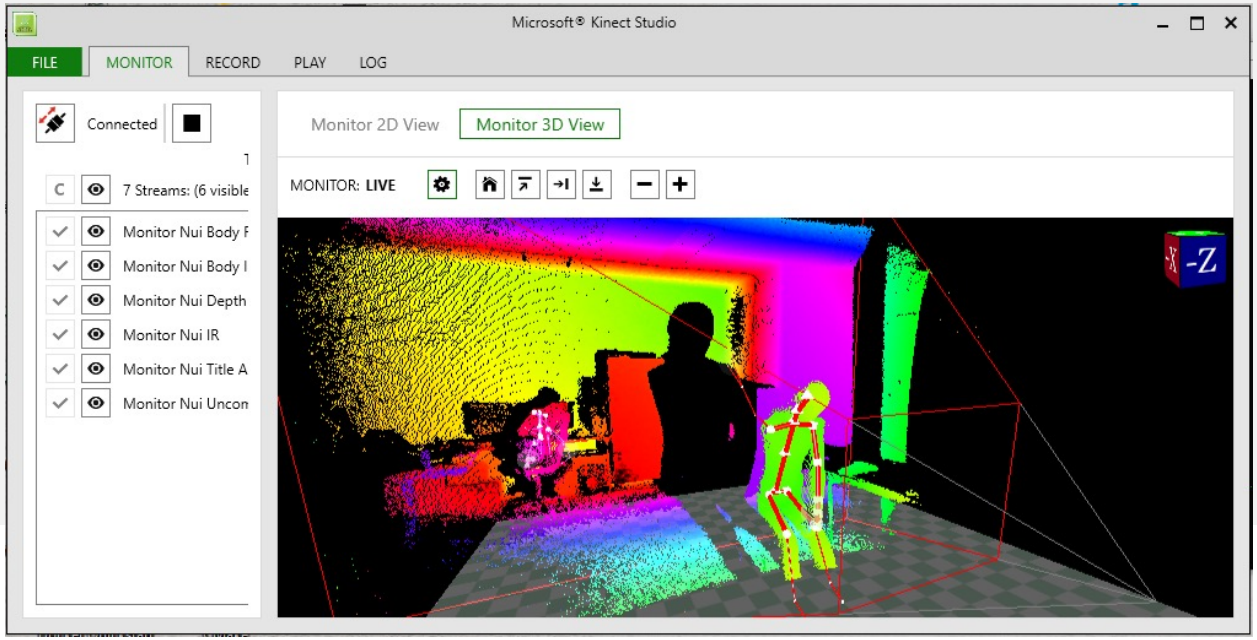
- Vocal nonlinguistic signals
  - Estimation of auditory features such as pitch, intensity, and speech rate
  - Recognition of nonlinguistic vocalizations like laughs, cries, sighs, and coughs



Recognizing Poses and Gestures

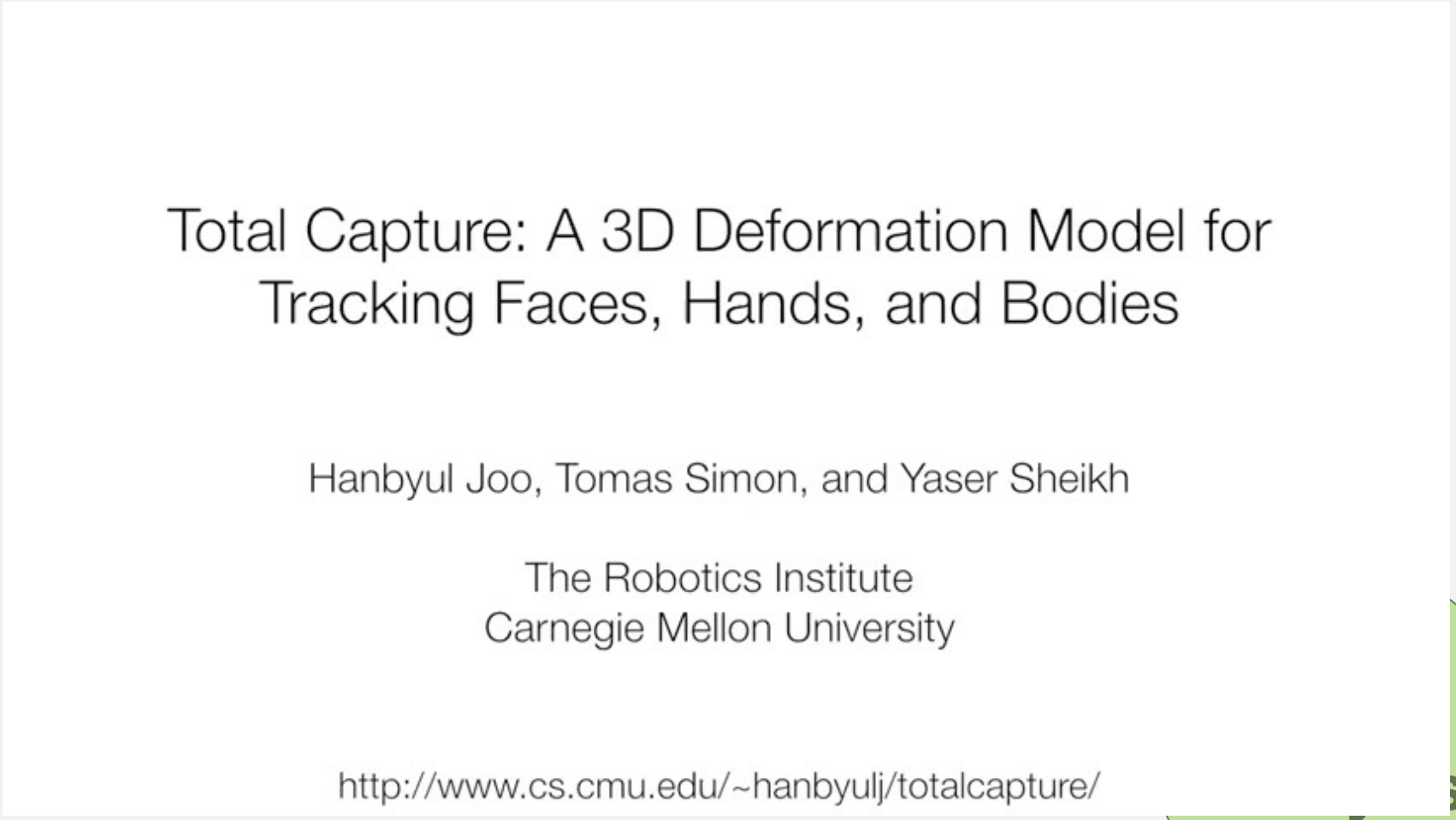
Kinect Sensor and Software Development Kit (SDK):

- Kinect is an array of sensors, including a camera and a depth sensor
- In addition to the raw depth image, Kinect extracts a 3D virtual skeleton of the body



User Analysis

Recognizing Poses and Gestures



https://www.cs.cmu.edu/~hanbyulj/totalcapture/

TÉCNICO LISBOA

Perception Levels

TÉCNICO LISBOA

Perception Levels

Action Analysis

TÉCNICO LISBOA

Identifying Human Actions

Human actions usually involve human-object interactions, where we can see articulated motions along complex temporal structures.

Actions are spatio-temporal patterns!

Action Analysis

TÉCNICO LISBOA

Identifying Human Actions

Issues in action recognition:

- Extraction and representation of suitable spatio-temporal features
- Modeling and learning of dynamical patterns

Action Analysis

TÉCNICO LISBOA

From Perceptions to Actions

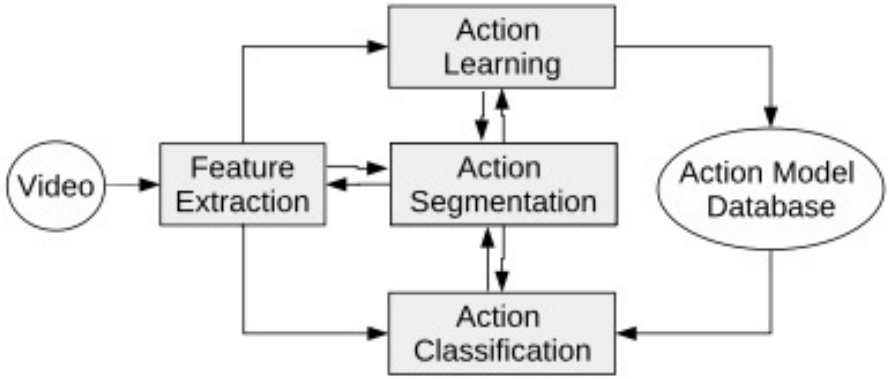
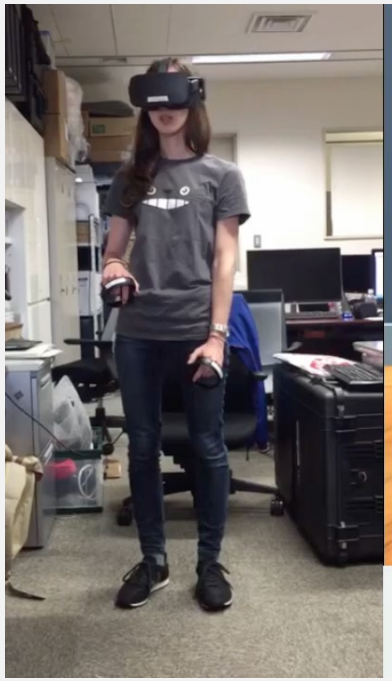
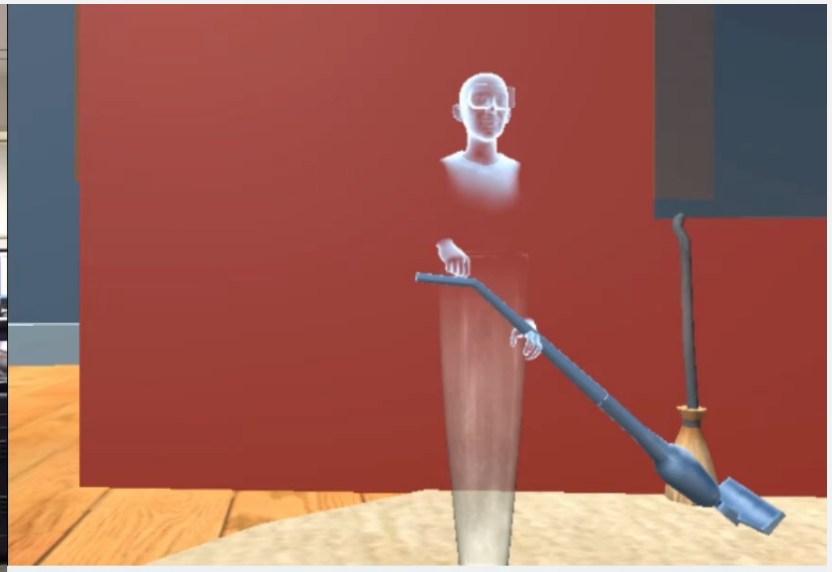


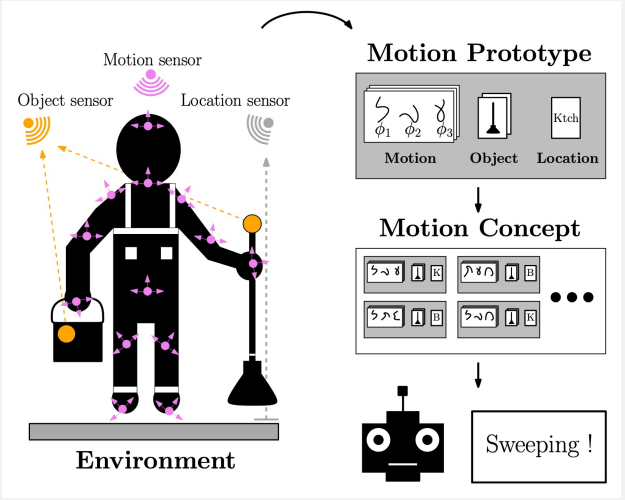
Figure 1: A typical data-flow for generic action recognition system comprises inter-dependent stages of feature extraction, learning, segmentation and classification.

Action Analysis

Human Actions – Multimodal Approach







Action Analysis

Vasco, Miguel, Francisco S. Melo, David Martins de Matos, Ana Paiva, and Tetsunari Inamura. "Learning multimodal representations for sample-efficient recognition of human actions." In 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4288-4293. IEEE, 2019.

TÉCNICO LISBOA

Perception Levels

TÉCNICO LISBOA

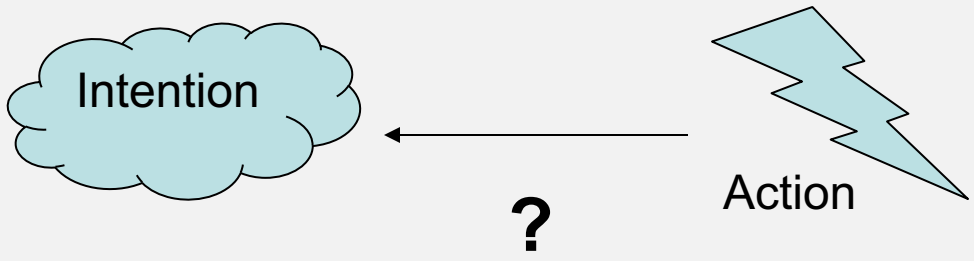
Perception Levels

Interaction Analysis

TÉCNICO LISBOA

From Actions to Intentions

- Goal: not only to figure out the action of the user but also, “why” the user is performing that action

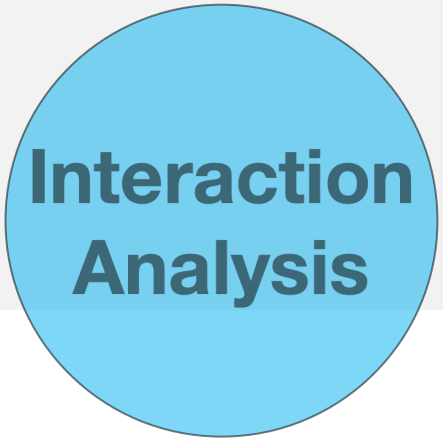


Interaction Analysis

TÉCNICO LISBOA

From Actions to Intentions

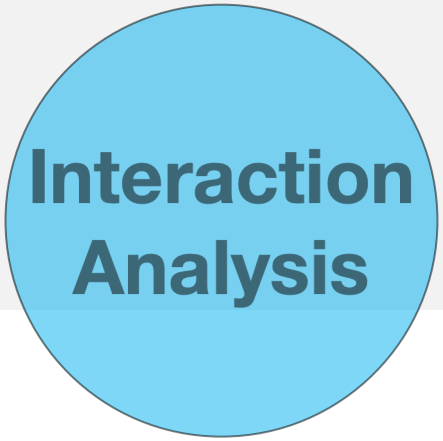
- Possible approaches:
  - Using object affordances to anticipate the human's next activity in order to enable the robot to plan ahead for a reactive response
  - Using simulation theory (and ToM) for intention recognition



Using Simulation Theory for Intention Recognition

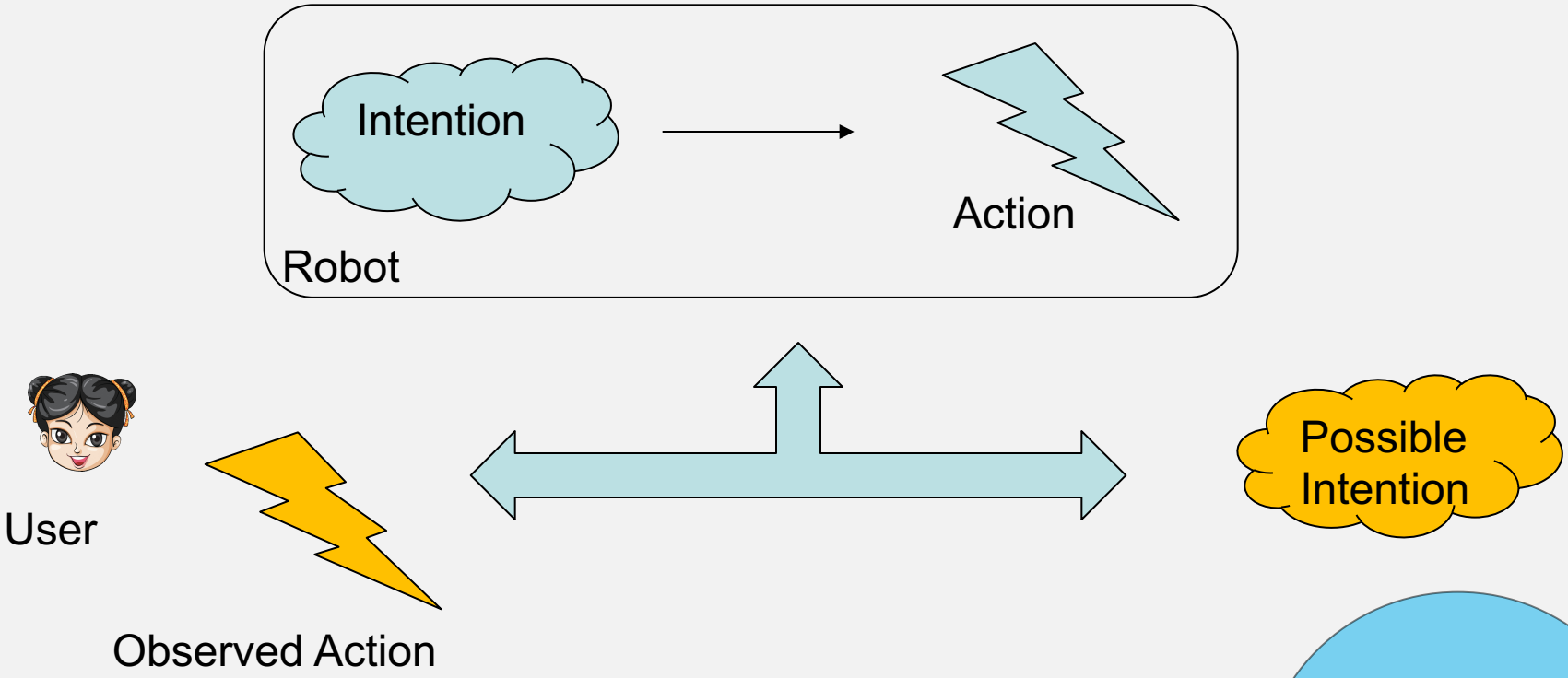
- Simulation theory - people attribute mental states using their own mental processes

- Taken off-line and used in simulation with states derived from taking the perspective of another person



TÉCNICO LISBOA

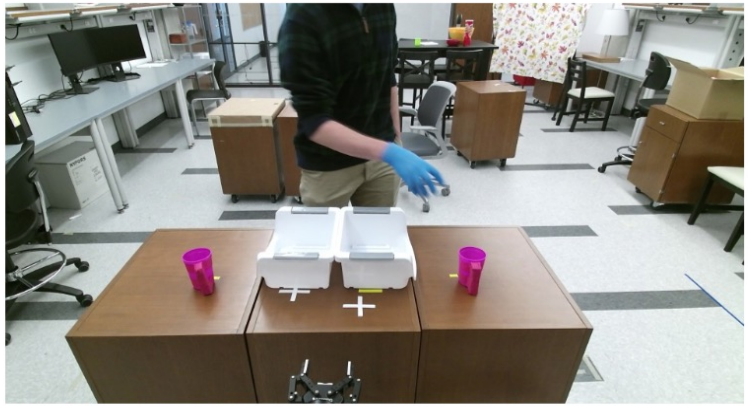
Using Simulation theory for intention recognition



TÉCNICO LISBOA

Intent Expression

(a) Object Intent



(b) Hand Motion Intent

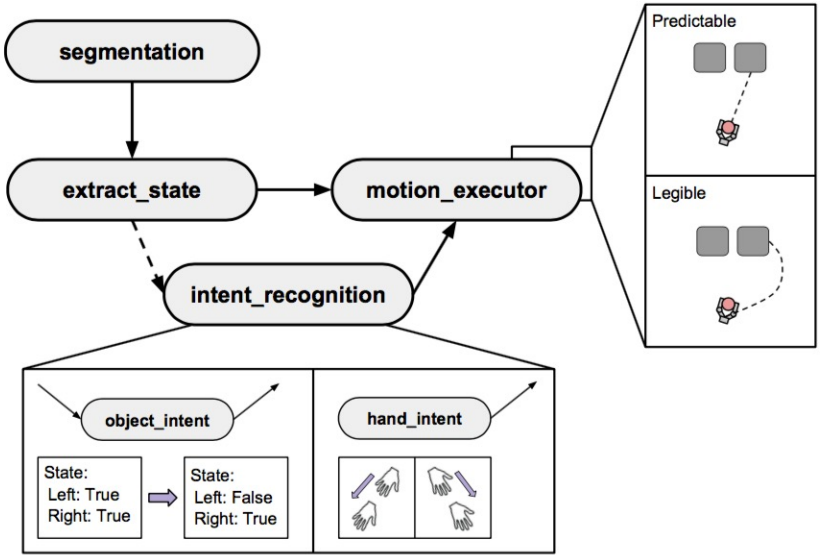


Fig. 2. The integrated intent recognition and generation system consists of four modules. The intent recognition and motion execution modules are set according to the experimental conditions.

Interaction Analysis

TÉCNICO LISBOA

2016

Towards Multi-Modal Intention Interfaces for Human-Robot Co-Manipulation

Luka Peternel, Nikos Tsagarakis and Arash Ajoudani

HRI^2 lab of Advanced Robotics department
Istituto Italiano di Tecnologia, Genoa, Italy

Interaction Analysis

https://www.youtube.com/watch?v=e3t5odKe6_c

TÉCNICO LISBOA

Summary

- User Analysis
  - (Face) Recognition of facial expressions
  - (Body) Recognition of poses and gestures
- Action Analysis
  - Recognition of dynamic motions
- Interaction Analysis
  - Recognition of intentions

What are the challenges?

TÉCNICO LISBOA

Social Signals Recognition - **Challenges**

Because it uses data-driven models most of time…

- There is a huge need for datasets!
  - Capture loads of data
  - Annotate loads of data

TÉCNICO LISBOA

Some extra examples...

TÉCNICO LISBOA

Case Study 1: User Analysis

Ginevra Castellano, Iolanda Leite, André Pereira, Carlos Martinho, Ana Paiva, Peter W. McOwan: Multimodal Affect Modeling and Recognition for Empathic Robot Companions. I. J. Humanoid Robotics 10(1) (2013)

TÉCNICO LISBOA

Case Study 2: Interaction Analysis

Bohus, Dan, Sean Andrist, Ashley Feniello, Nick Saw, Mihai Jalobeanu, Patrick Sweeney, Anne Loomis Thompson, and Eric Horvitz. "Platform for situated intelligence." *arXiv preprint arXiv:2103.15975* (2021).

TÉCNICO LISBOA

Discussion

My Work

My Work

- Human-Robot Teamwork
- Multiparty Interactions

Work on Perception

- Gaze perceptions in multiparty settings
- Should the robot establish mutual gaze?
- What about joint attention?
- And if the target of joint attention in a human teammate?

MSc Thesis Proposals

- **Gaze & Agency in Child-Robot Interaction**
  – Miguel Azinheira (in progress)

MSc Thesis Proposals

2nd semester 21 / 22 (starting in Feb / Mar)

- **Gaze Perceptions in "*The Mind*" Game**
- **Anthropomorphism of Minimal Embodiments**

MSc Thesis Proposals

1st semester 22 / 23 (starting in Sep / Oct)

- Gaze & Agency in Child-Robot Interaction
- Gaze Perceptions in "*The Mind*" Game
- Anthropomorphism of Minimal Embodiments

Questions?

filipacorreia@tecnico.ulisboa.pt
@PipzCorreiaz