Facial Expression Recognition of Neurologically Impaired Children

9 July 2015

Ciprian A. Corneanu

supervised by Sergio Escalera & Marc Oliu Simon

Contents

Introduction

Motivation # Challenges # Applications

Why Facial Expression Analysis?

Sculptures by Franz Messerschmidt.
18th century German-Austrian artist.

Paintings by Duarte Vitoria.
Contemporary Portuguese artist.

Why Facial Expression Analysis?

FEA is a part of Affective Computing. It is essential in building socially aware systems, improving Human Computer Interaction and helping understanding human emotion.

**Figures from courses.media.mit.edu,, cdm.depaul.edu, marshable.com*

Challenges

- Head-pose variations
- Illumination
- Registration errors
- Occlusions
- Identity bias (telling between person specific and expression specific features)
- Relating facial expressions to affective state

Applications

Potential Applications

- Detection of truthfulness or potential deception (HR, police)
- Socially aware systems (HCI)
- Pain detection (clinical context)

Commercial Applications

< I M R S V >

affectiva®

realeyes

EMOTIENT™

The Neurochild Project

General Presentation # Multimodal Approach # Capturing Data # Proposed Affective States

The Neurochild Project

- Joint project: ICA, UPM, Institut Guttmann.
- The goal is to develop a framework for patient progress assessment during rehabilitation sessions.

The Neurochild Project

INSTITUT GUTTMANN — Defining Affective States. Advising on Neurological/Psychological aspects.

CVC Centre de Visió per Computador — R&D for the Facial Expression Recognition Framework.

ICA — Project coordination. Web Integration.

UNIVERSIDAD POLITECNICA MADRID — R&D of the Eye Tracking Framework.

Contributions of different partners to Neurochild.

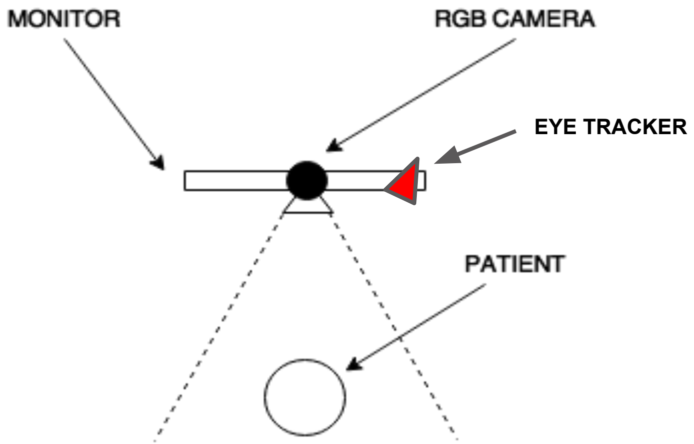
Project Overview

Neurochild will consist in:

- Multimodal approaches that combine Facial Expression Analysis with Eye Tracking and Head Pose to recognize complex affective states
- Designing affect stimulating contexts through customly built video games
- Integration into web based products

Capturing Data

- During sessions young patients (5-14 yrs) play specially designed games.
- RGB cameras capture their faces while playing
- Eye tracker for detecting eye gaze



Capturing data for Neurochild.

Proposed Affective States

Short Term		Medium Term
Emotional State	**Cognitive State**	**Mood**
Happy/Amused Angry Disgusted Suprised Afraid Sad	Interested Bored Concentrated Frustrated Sure/Unsure Agreeing/Disagreeing	Nervous Animated/Energetic Impulsive

Affective states to be targeted in the Neurochild automatic facial expression recognition framework.

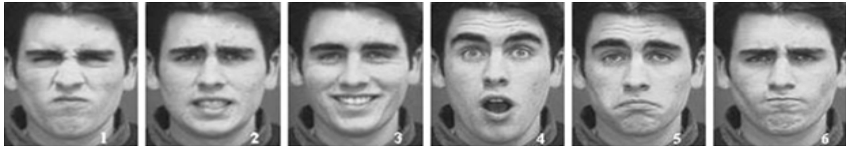
Building Automatic Facial Expression Analysis Systems

General Considerations # Automatic FE Analysis Systems

General Considerations

*Parametrizing facial expressions. The Facial Action Coding System.**

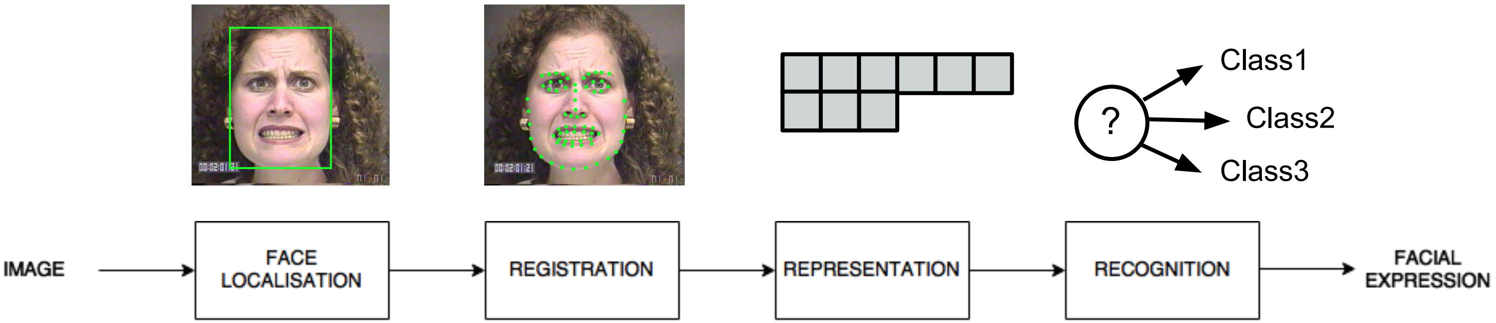
- FACS is the most used parametrization system
- Ekman's universal expressions of emotions constitute the basis of most of research on expressions of affect on the face.



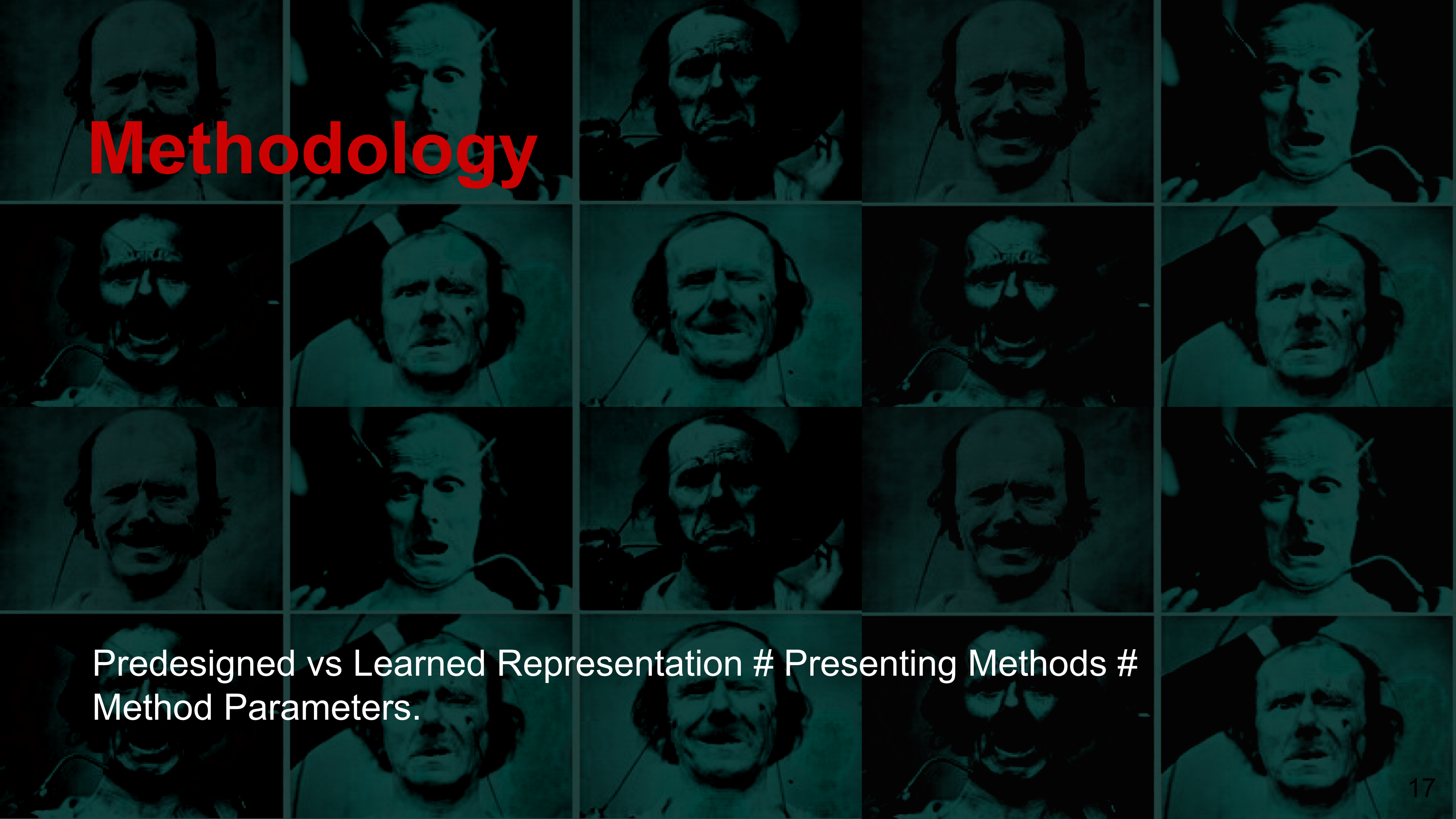
*Paul Ekman's Universal Facial Expression of Emotion. From left to right: Disgust, Fear, Happiness, Surprise, Sadness, Anger.**

*From http://what-when-how.com

Automatic FE Analysis Systems



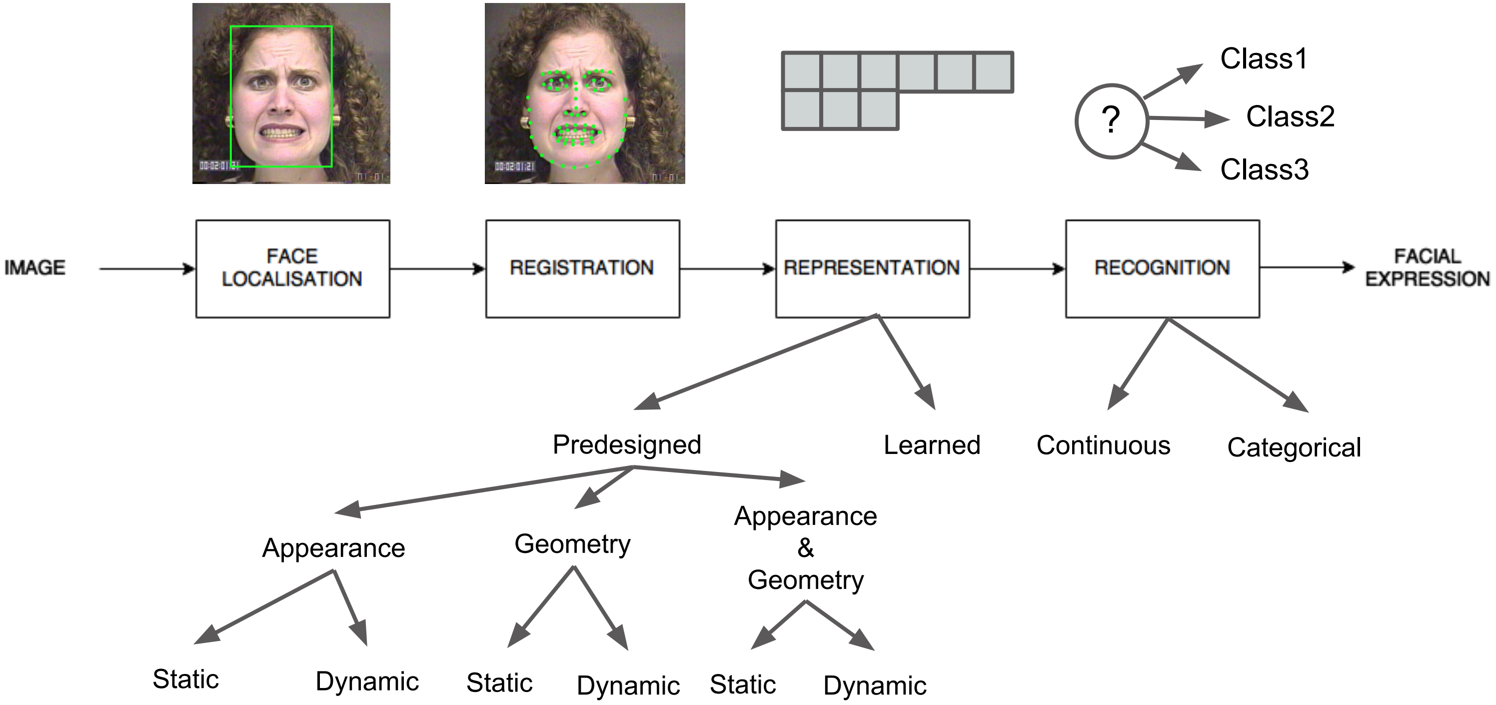
Structure of a facial expression recognition system.



Methodology

Predesigned vs Learned Representation # Presenting Methods # Method Parameters.

Predesigned vs. Learned Representation



Method classification according to Representation and Recognition.

Predesigned vs. Learned Representation

Predesigned

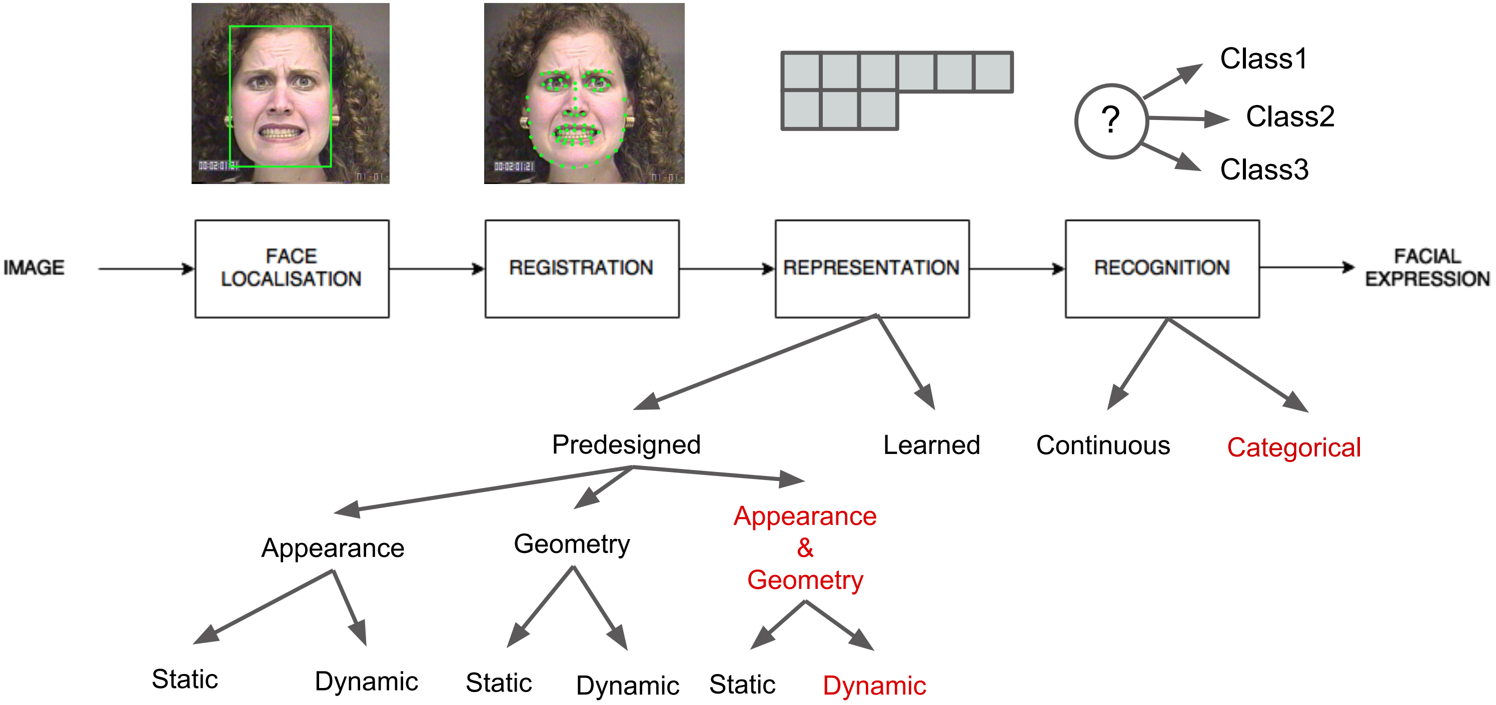
- Representation and Classification are not necessarily connected
- Choosing representation is empiric
- Tuning parameters can be cumbersome
- Smaller amounts of data needed

Learned

- The representation that optimizes classification is learned
- Long training times
- Large amount of labelled data is needed
- Choosing CNN topology can be tricky
- Higher performance

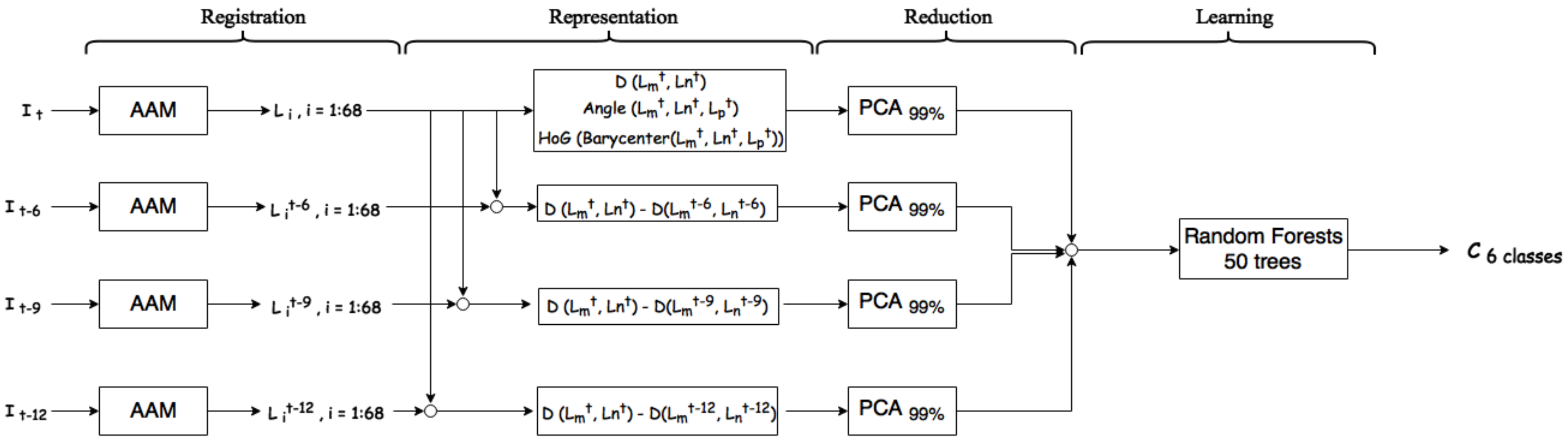
Comparison of predesigned and learned representation based methods.

Method 1



Method classification according to Representation and Recognition.

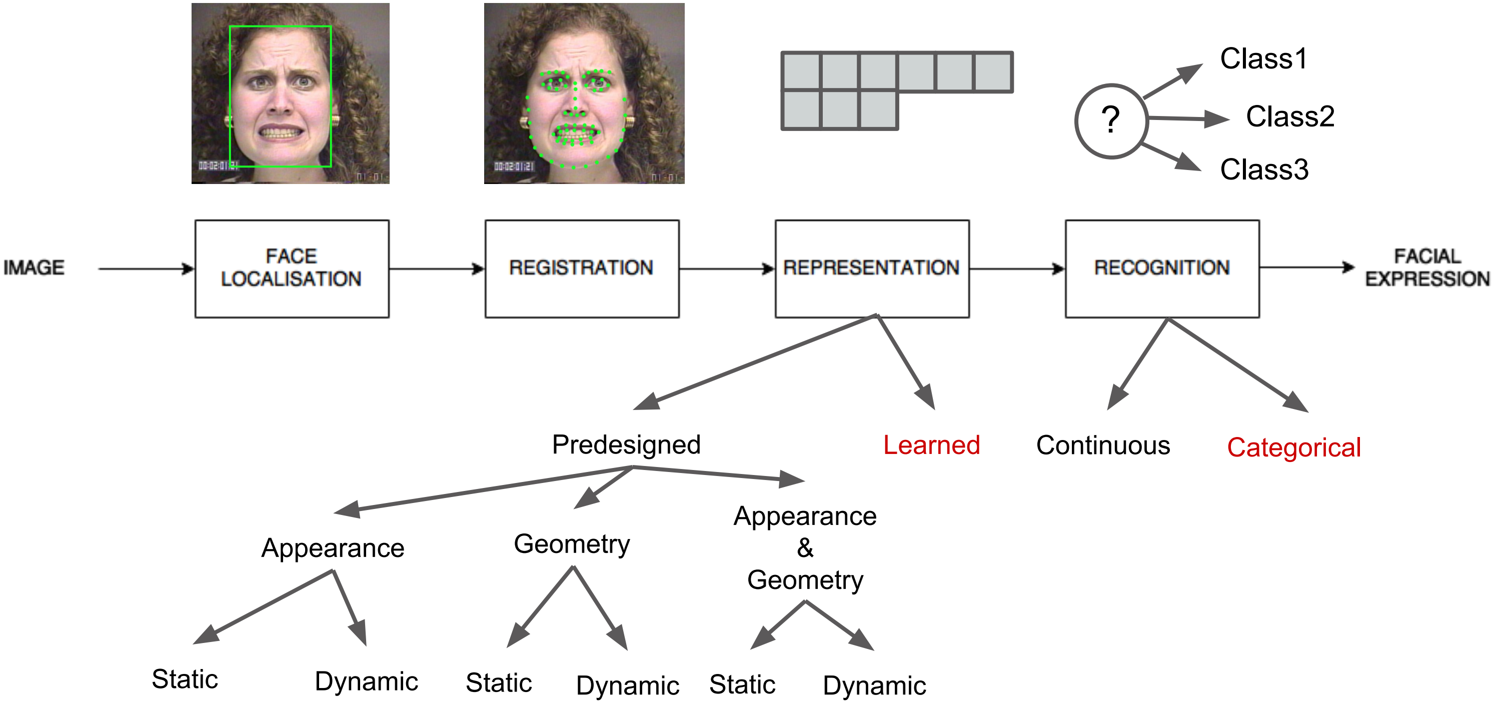
Method 1



Block diagram of Method 1.

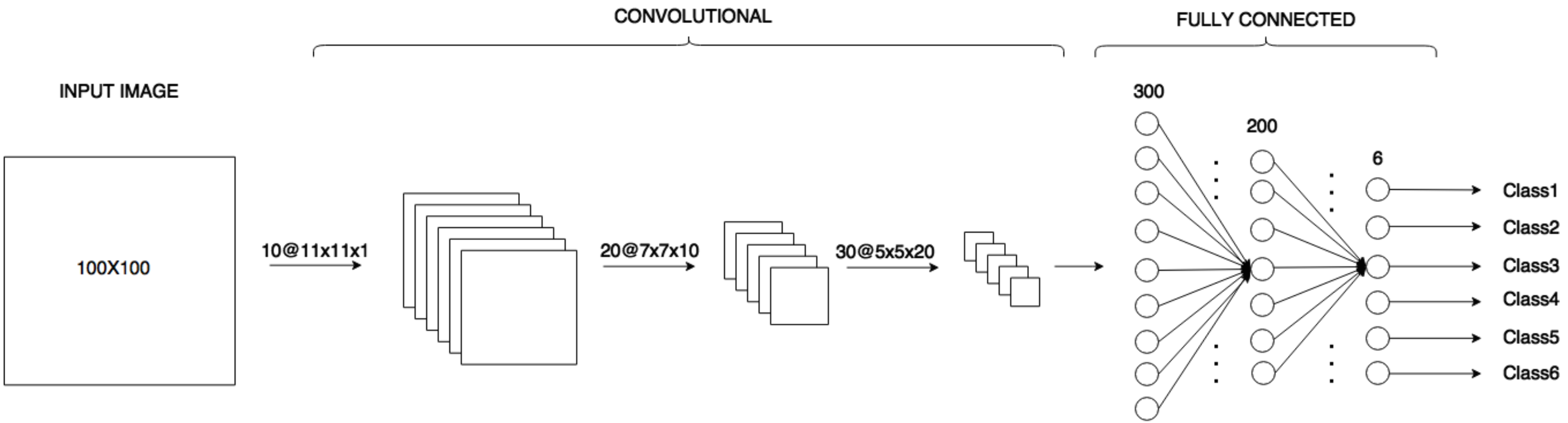
This method is based on: Dapogny, Arnaud, Kevin Bailly, and Séverine Dubuisson. "Dynamic facial expression recognition by joint static and multi-time gap transition classification."

Method 1

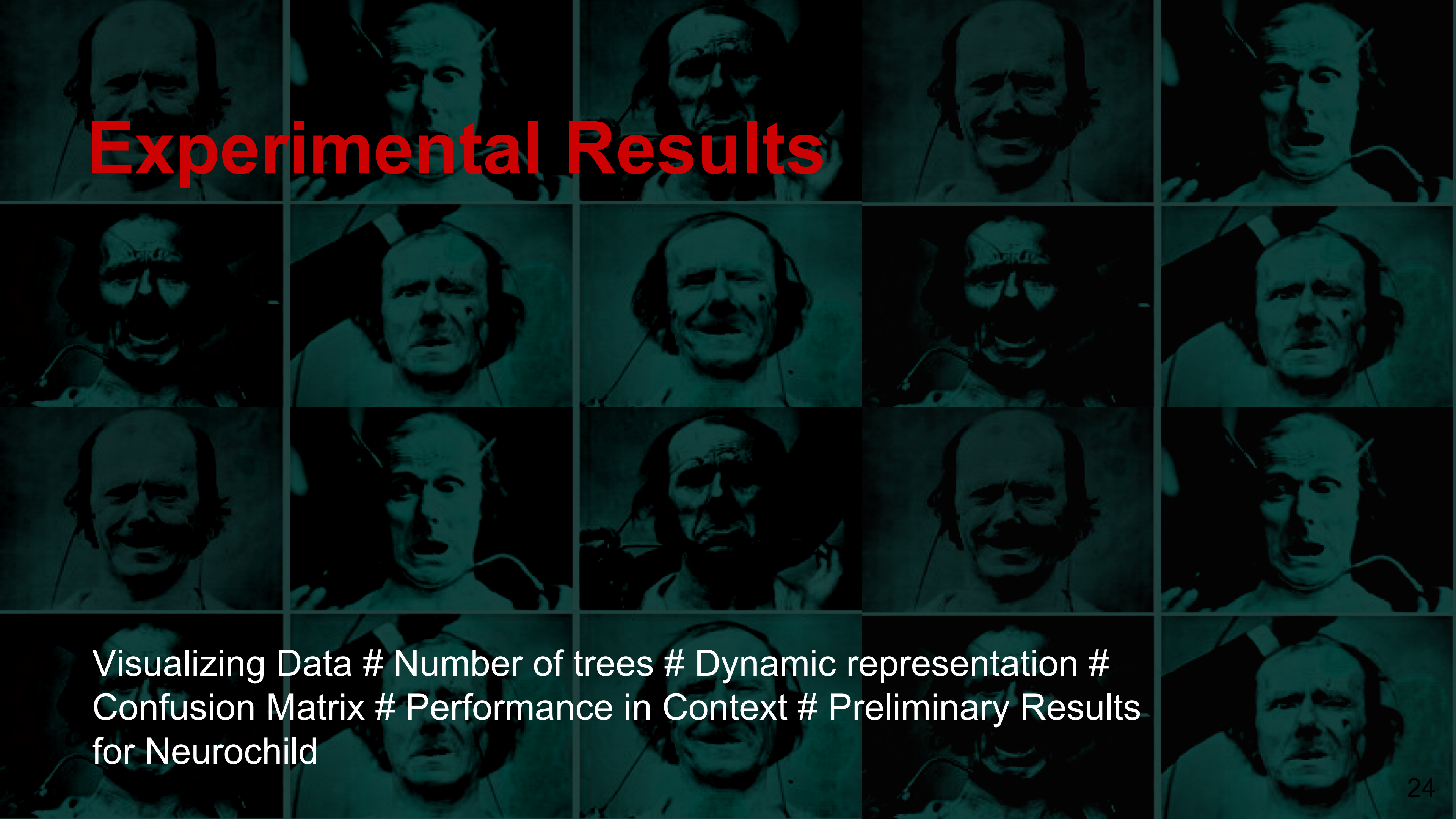


Method classification according to Representation and Recognition.

Method 2



CNN topology of Method 2.



Experimental Results

Visualizing Data # Number of trees # Dynamic representation # Confusion Matrix # Performance in Context # Preliminary Results for Neurochild

The CK+ Dataset

- Posed and spontaneous expressions
- Dynamic (neutral to apex)
- Captured in the lab
- Frontal, standard Illumination, no occlusions.
- 201 subjects
- Ethnic diversity
- Gender diversity: 31% males, 69% females.
- Age: 18-50 yrs

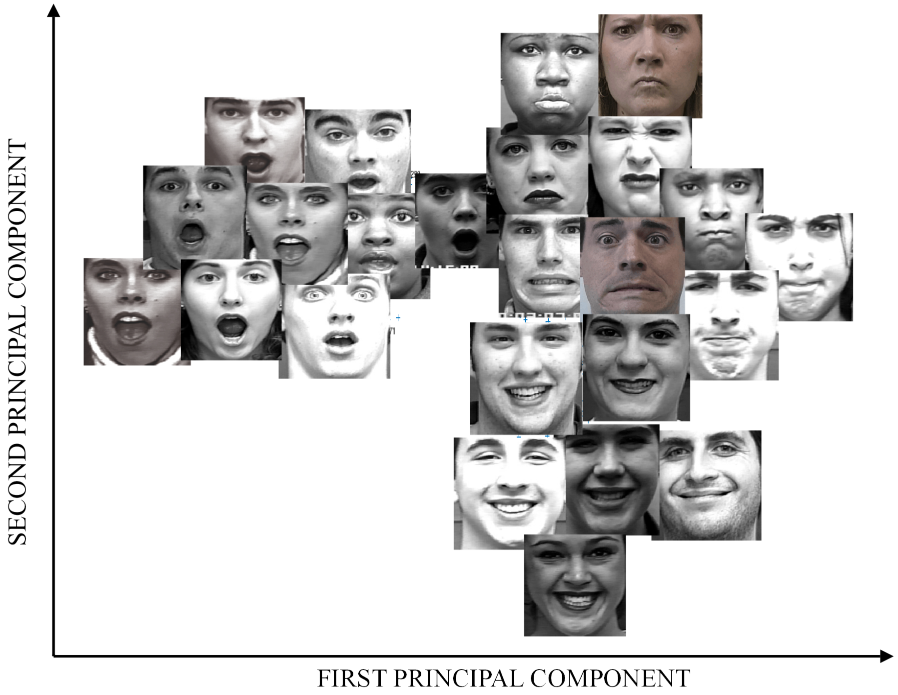
*Samples from CK+**

*CK+ contains facial expressions from neutral to apex.**

*Kanade, J. F. Cohn, and Y. Tian, “Comprehensive database for facial expression analysis,” in *FG*, 2000, pp. 46–53.

Visualizing data

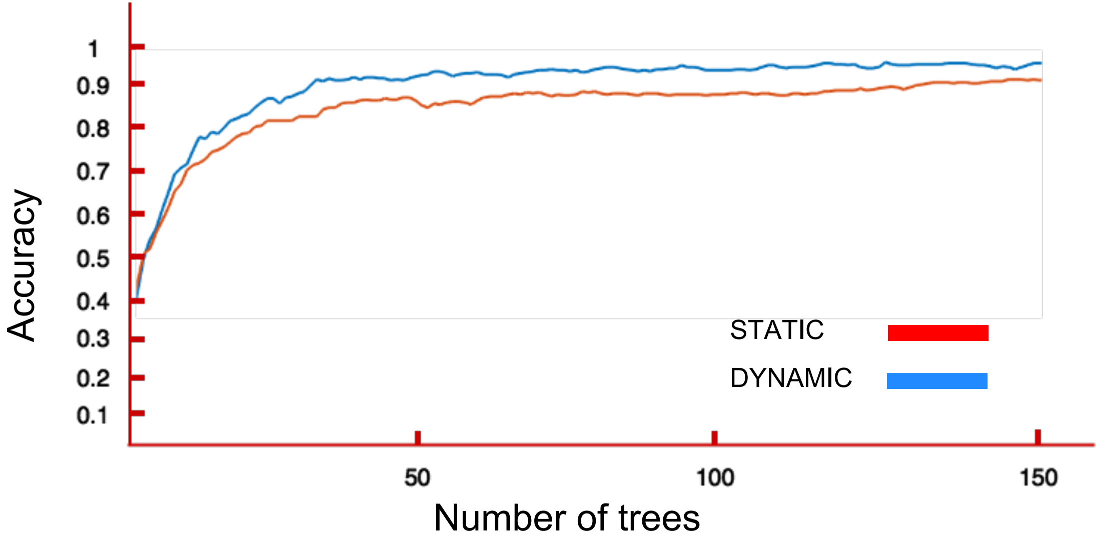
- Displaying samples along first two principal components shows the validity of the representation for clustering expression.
- First principal components codes large variations of the whole face while second principal component mostly codes opening of the mouth.



Data along first two principal components.

Method 1: Number of trees

50 is the optimum number of trees for the Random Forest



Accuracy dependence on number of trees in the Random Forest.

Method 1: Number of point tuples

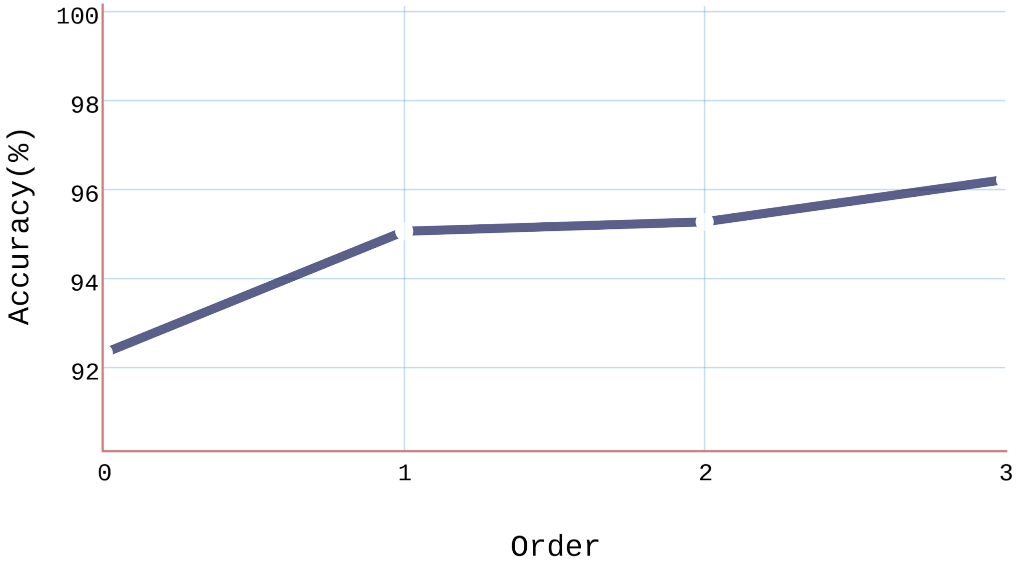
100 is the optimum number of pairs and triplets for the geometrical representation

# of tuples	Accuracy
30	93.80%
100	96.22%
300	93.34%

Accuracy dependence on the number of tuples for the geometrical representation.

Method 1: Dynamic representation

- The order is defined by how many transitions are added to the static representations.
- Dynamic representation considerably improves on static representation.
- Higher order achieves higher performance.
- Going to close to the reference frame did not provide relevant information
- Sequences in the CK+ dataset do not allow adding longer transitions.
- Order = 3 is optimal.



Accuracy according to dynamic order. Transitions represented at 6,9 and 12 frames in the past.

Method 1: Confusion Matrix (Static)

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	**88%**	1%	0%	0%	11%	0%
Disgust	5%	**91%**	1%	0%	3%	0%
Fear	0%	0%	**100%**	0%	0%	0%
Happiness	0%	0%	7%	**93%**	0%	0%
Sadness	9%	0%	2%	0%	**89%**	0%
Suprise	0%	0%	2%	0%	0%	**98%**

Confusion matrix for static representation accuracy.

Method 1: Confusion Matrix (Dynamic)

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	**85%**	5%	2%	0%	8%	0%
Disgust	1%	**99%**	0%	0%	0%	0%
Fear	0%	0%	**100%**	0%	0%	0%
Happiness	0%	0%	7%	**93%**	0%	0%
Sadness	0%	0%	0%	0%	**100%**	0%
Suprise	0%	0%	0%	0%	0%	**100%**

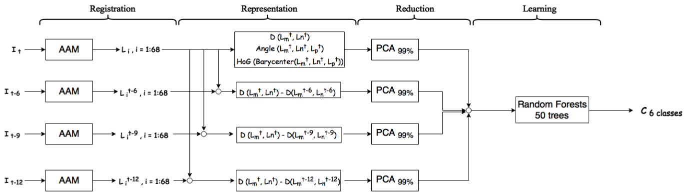
Confusion matrix for dynamic representation accuracy.

Method 2

- Data expansion: flip, rotation (-10, -5, +5, +10 degrees), sliding
- Data regularization: Gaussian White Noise
- Adaptive learning rate
- We have experimented with different kernel sizes in the convolutional layer and different number of units in the fully connected layer
- Best result achieved was 78% accuracy
- Difficult to obtain higher performance without overfitting data
- Larger amount of data is needed

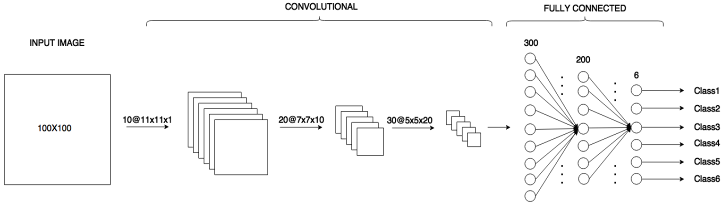
Final Configurations

Method 1



3 transitions (6,9,12 frames)
100 pairs and triplets of points
RF: 50 trees
Early Fusion btw. geometrical and appearance representations

Method 2



5 hidden layers
10@11x11x1
20@7x7x10
30@5x5x20
300/200/6 units
Adaptive Learning Rate

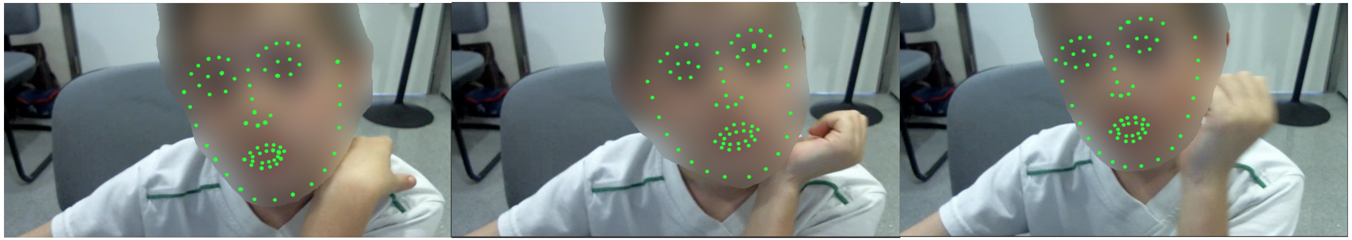
Performance in Context

Proposed methods performance compared with related methods in the same context.

Method	Accuracy
Method 2	**78%**
Ranzato ‘11	90.11%
Bartlett ‘03	93.3%
Littlewort ‘04	93.3%
Sebe ‘07	93.4%
Aleksic ‘06	93.6%
Method 1	**96.22%**
Liu ‘14	96.7%
Kotsia ‘07	99.7%

Preliminary Results for Neurochild

- Face Localization and Registration were performed on sequence captured from Neurochild.
- Occlusions and head rotations constitute major challenges.
- Next steps will include labelling and training proposed methods with captured data.



Preliminary results from the Neurochild Project.

Conclusions

Conclusions # Future Work # Acknowledgements

Conclusions

- We have done a survey of the State of the Art for determining most appropriate methods for solving the problem. This resulted in the paper: *C.Corneanu, M. Oliu, J. Cohn, S. Escalera, "Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Analysis: History, Trends, and Affect-related Applications ", TPAMI, 2015 (Second Revision).*
- In order to automatically assess patients progress in clinical contexts a generic affect recognition framework was proposed and preliminary data captured.
- Two approaches were tested for detecting facial expressions of primitive emotions with the main goal of comparing predesigned with learned representations.
- In the case of **Method 1**, we have showed how taking into account the dynamic improves facial expression classification.
- **Method 2** based on learning the representation is limited by the available amount of data.

Future Work

- Specially trained persons will label data captured for the Neurochild project.
- Methods proposed should be trained with the labelled data.
- Preliminary data shows that methods should be robust to head pose variation and occlusions. Improvements of methods in this directions will be necessary.
- For recognizing cognitive states and medium term affective states a multimodal approach should be used by integrating facial expression analysis with eye tracking and head pose.

Personal Contributions

- Coordinating collaboration with Prof. Jeffrey Cohn for building extended survey of the state of the art.
- For Method 1 significant improvements over Dapogny et al, 2015 by optimizing the number of represented transitions and changing the dynamic representation.
- Matlab and Python code developed for conducting experiments.
- Joint contribution together with Institut Guttmann for defining a taxonomy of targeted affective states.

Acknowledgements

We thank for the support of our consortium partners: UPM, Institut Guttmann, and ICA.

Special thanks to Prof. Jeffrey Cohn from CMU for his support in the definition of the state of the art survey on Facial Expression Analysis.

I am deeply grateful to Prof. Sergio Escalera for his suggestions, his artistic eye and the obsession for details we share and to Marc Oliu for his help.

Thank you!