Person Identification System in a Platform for Enabling Interaction with Individuals Affected by Profound and Multiple Learning Disabilities

Carmen Campomanes-Álvarez, Blanca R. Campomanes-Álvarez[1], Pelayo Quirós

[1]Ph.D. in Computer Science, Artificial Intelligence
Intelligent Data Analysis Department
CTIC Technology Centre - Gijón, Asturias, Spain
www.fundacionctic.org

ICCI*CC'19

Outline

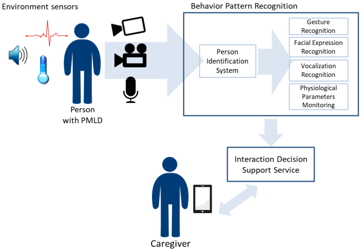
Introduction

People with Profound and Multiple Learning Disabilities (PMLD):

- present severe intellectual difficulties that affect their skill to communicate their feelings.
- can not be independent.
- express their needs with non-symbolic behaviors (non conventional reactions).
- need a constant support by a group of professional caregivers.

Introduction (II)

The INSENSION (H2020) project aims to design and develop an intelligent platform that enables people with PMLD to enhance the quality of their life with digital applications and services.



Based on the previous knowledge about each person, the platform will be able to associate the recognized expressions with their meaning in an individualized way.

Introduction (III)

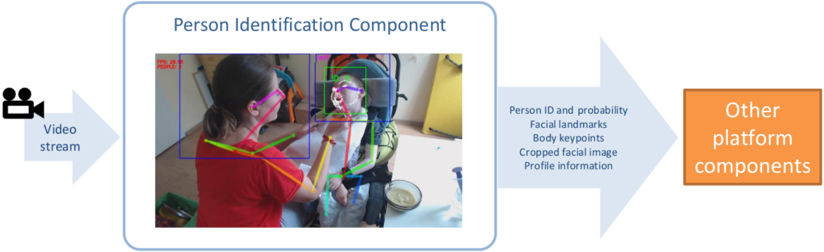
First step in the INSENSION platform

The aim is to implement an automatic person identification system for:

- recognizing correctly six people with PMLD that participate in the project.
- discerning them from the rest of individuals that could appear in the videos.
- integrating as the first stage of a real-time platform.

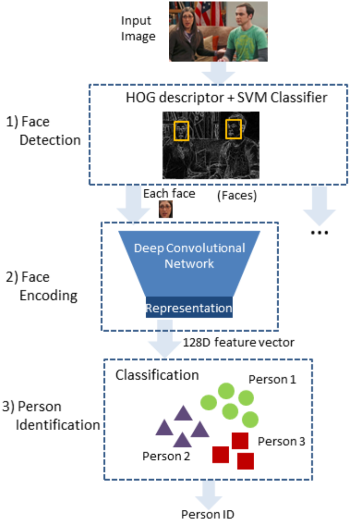
Introduction (IV)

Person identification component



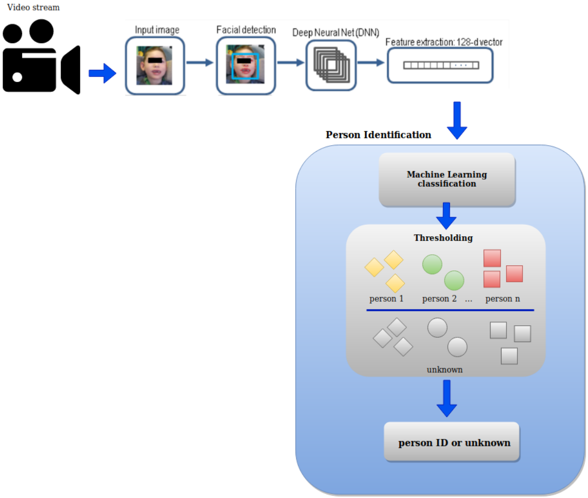
Existing methods for facial identification

Three phases



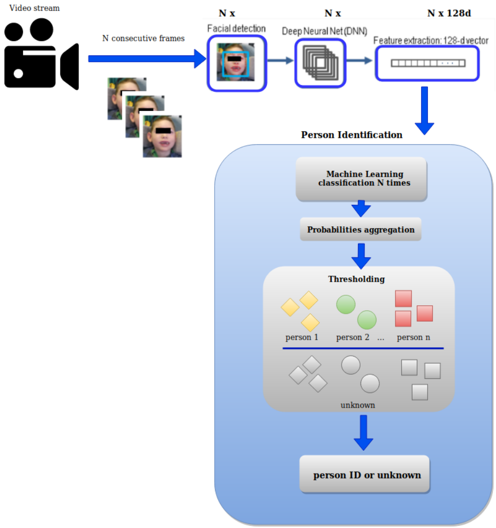
Our proposal: identification frame by frame

classification + **thresholding**



Improving our proposal: multi-frame identification

classification + **probabilities aggregation** + **thresholding**



Methodology

Designing the database

- PMLD-people: 3000 facial images of 6 individuals:
  - 5-fold cross-validation approach (100 images per person).
  - each fold 400 images for training, 100 for test (per person).
- Unknown-people: two datasets:
  - Related-Unknown People (RUP): 300 facial images, 20 per 15 people in same conditions.
  - General-Unknown People (GUP): 3303 facial images of 395 individuals.

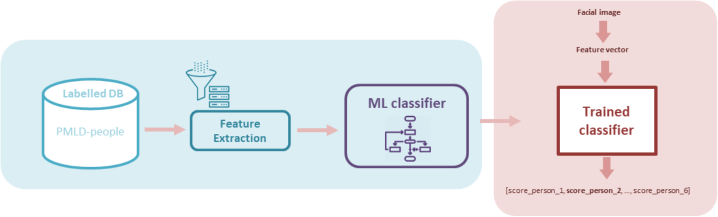
Methodology (II)

Classification methods considered

- Gaussian Naïve Bayes
- KNN
- SVM
- Logistic Regression
- Stochastic Gradient Descent (SGD)
- Neural Networks
- Random Forest
- XGBoost

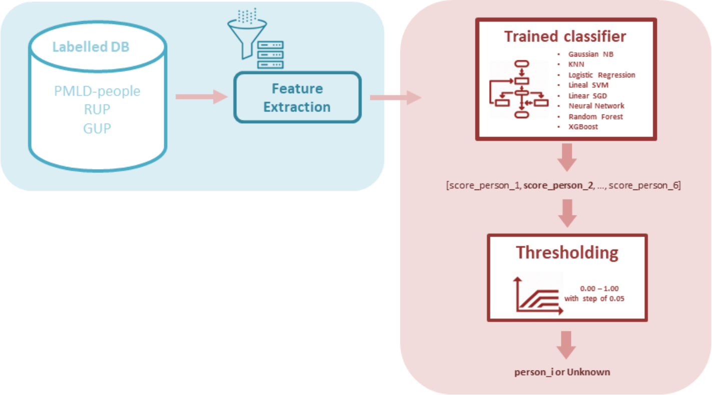
Methodology (III)

Classifiers generation



Methodology (IV)

Person identification using the previously trained classifiers



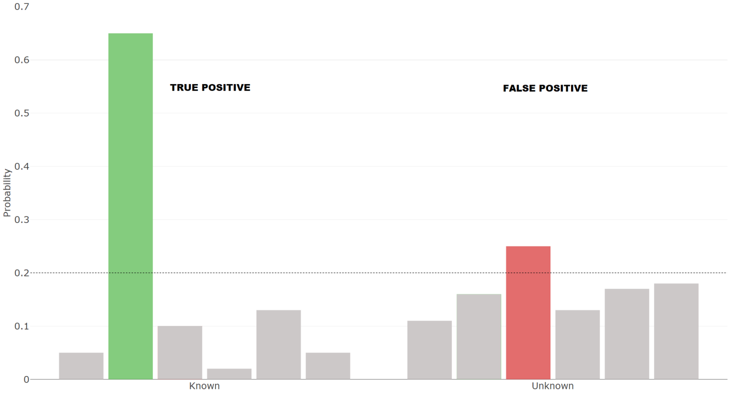
Methodology (V)

Thresholding

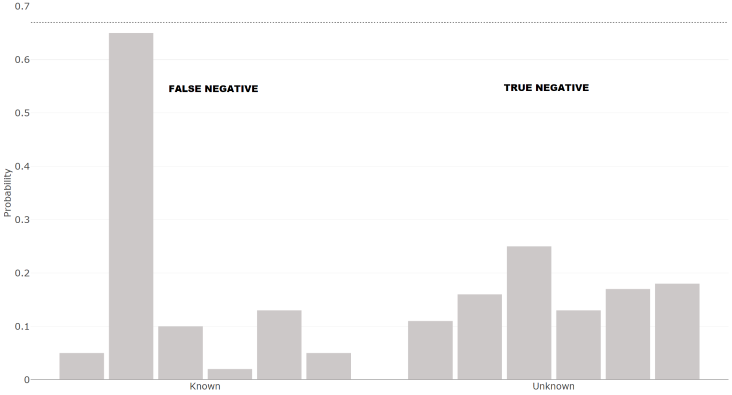
After obtaining the ID scores for each class (six PMLD individuals):

- A threshold is applied for determining if the obtained score is high enough to assign a PMLD-person identity.
- Probabilities lower than the selected threshold result in "unknown" person.
- Probability higher than the selected threshold results in a PMLD-person identity.

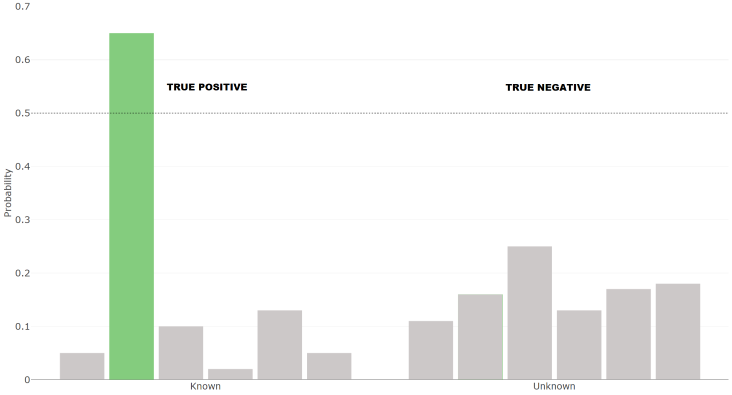
Methodology (VI)



Methodology (VII)



Methodology (VIII)



Methodology (IX)

How well does each model classify known and unknown people?

$$GS = \frac{1}{2} \cdot (TP + TN)$$

- TP are True Positives testing with PMLD-people dataset.
- TN are True Negatives testing with RUP and GUP datasets.
- As $TP, TN \in [0, 1]$, the closer GS to 1 the more accuracy.

Experiments and results

General setup

- Intel CoreTM i7-5820K CPU 3.30GHz x12 TITAN Xp Graphic Card, Ubuntu 16.04LTS.
- Facial image to 128 feature vector performed using the FaceNet approach, dlib.
- Python libraries scikit-learn and face recognition.
- Multi-frame approach: mean and max for aggregating probabilities.

Experiments and results (II)

Experiment 1. Inner classification.

Goal: Training the selected eight algorithms for obtaining eight identification classifiers.

- The PMLD-people dataset was used.
- The eight algorithms were tested using 5-cross validation.
- Metrics for evaluation: accuracy, precision, recall and F1.

Experiments and results (III)

Results for Experiment 1

Method	Accuracy	Precision	Recall	F1
Gaussian NB	0.975	0.977	0.975	0.975
KNN	0.993	0.993	0.993	0.993
Linear SGD	0.990	0.991	0.990	0.990
Linear SVM	0.992	0.992	0.992	0.992
Logistic Regression	0.988	0.988	0.988	0.988
Neural Network	0.991	0.991	0.991	0.991
Random forest	0.986	0.987	0.986	0.986
XGBoost	0.978	0.978	0.978	0.978

Experiments and results (IV)

Experiment 2. Outer classification, mono-frame approach.

Goal: Selection of the best method for correctly identifying known and unknown people.

- The eight obtained classifiers were evaluated.
- The PMLD-people dataset was used for testing and computing the *TP* rate for each classifier.
- RUP and GUP datasets were used for testing and computing the *TN* rate for each classifier.
- *GS* was computed.
- Different thresholds were considered.

Experiments and results (V)

Results for Experiment 2

Scores of the 10 best methods with thresholds, sorted in descending order by GS.

Threshold	Method	TP	TN_1	TN_2	GS
0.70	Logistic Regression	0.912	0.951	0.927	0.931
0.90	Linear SVM	0.939	0.911	0.911	0.925
0.75	Logistic Regression	0.865	0.981	0.970	0.923
0.65	Logistic Regression	0.939	0.892	0.859	0.914
0.95	Linear SVM	0.854	0.968	0.969	0.911
0.85	Linear SVM	0.959	0.844	0.843	0.901
0.95	Linear SGD	0.842	0.958	0.961	0.900
0.90	Linear SGD	0.920	0.871	0.879	0.896
0.80	Logistic Regression	0.772	0.996	0.996	0.884
0.95	Neural Network	0.943	0.823	0.816	0.883

Experiments and results (VI)

Experiment 3. Robust classification, including multi-frame approach.

Goal: Study the performance of the multi-frame approach.

- N= 5, 10 and 15 frames were selected as input.
- For each individual, N feature vectors were classified resulting in a N probabilities vector.
- Aggregation functions (mean and max) were used to obtain a probability value from the N probabilities vector.
- PMLD-people and RUP datasets were used for computing the TP and TN rates, respectively.
- GS and execution time (per input image) were computed.

Experiments and results (VII)

Results for Experiment 3

Score and execution time of the 10 best Logistic Regression multi-frame configurations and the two best mono-frame, sorted by GS in descending order.

Index	N. frames	Threshold	Agg.	GS	Execution Time (s)
1	15	0.65	mean	0.989	0.00160
2	10	0.80	max	0.985	0.00070
3	10	0.65	mean	0.983	0.00070
4	15	0.70	mean	0.982	0.00160
5	5	0.80	max	0.981	0.00010
6	15	0.80	max	0.976	0.00160
7	15	0.60	mean	0.974	0.00170
8	10	0.70	mean	0.973	0.00070
9	15	0.85	max	0.973	0.00160
10	5	0.65	mean	0.970	0.00010
...	...	...	...	...	...
23	1	0.65	—	0.912	0.00001
...	...	...	...	...	...
27	1	0.70	—	0.904	0.00001

Conclusions

- Multi-frame analysis achieves better results than the mono-frame approach.
- Logistic Regression outperforms the other methods regarding the detection of unknown individuals (using a threshold).
- The most suitable solution ($GS = 0.98$) corresponds to Logistic Regression with a multi-frame approach: 5 frames, aggregated with maximum, threshold of 0.80.
- The best configuration is selected for the person identification component within the INSENSION platform.

IEEE ICCI*CC 2019
18th IEEE International Conference on
COGNITIVE INFORMATICS & COGNITIVE COMPUTING
July 23-25, 2019

Thank you for your attention!

Blanca R. Campomanes Álvarez, Ph.D.
CTIC Technology Centre, Gijón, Asturias, Spain.

charo.campomanes@ctic.es

Given the proportion of correctly classified known individuals (TP), correctly classified unknown individuals (TN_1) and correctly classified unknown individuals that appeared at some point in the originally processed videos (TN_2), GS is obtained as:

$$GS = \frac{1}{2} \cdot \left(TP + \frac{|GUP| \cdot TN_1 + |RUP| \cdot TN_2}{|GUP| + |RUP|} \right) =$$

$$= \frac{1}{2} \cdot TP + \frac{|GUP|}{2 \cdot (|GUP| + |RUP|)} \cdot TN_1 +$$

$$+ \frac{|RUP|}{2 \cdot (|GUP| + |RUP|)} \cdot TN_2,$$

where $|\cdot|$ represents the size of the given set. It must be noted that, as $TN_1, TN_2, TP \in [0,1]$, and that their corresponding coefficients satisfy that:

$$\frac{1}{2} + \frac{|GUP|}{2 \cdot (|GUP| + |RUP|)} + \frac{|RUP|}{2 \cdot (|GUP| + |RUP|)} = 1,$$

FaceNet approach: make a selection of a face, encoding in a vector, train a DNN, obtain a model

For the model: Library face-recognition and encoding, trained by:

https://github.com/davisking/dlib-models