Online Face Recognition based on a Cognitive Learning Paradigm

Bogdan Raducanu

Computer Vision Center, Barcelona

Universitat Autònoma de Barcelona

Centre de Visió per Computador

"The purpose of the brain is to produce future."

Daniel Dennett

(Do-It Yourself Understanding –

Brainchildren, Essays on Designing Minds)

Outline

- 1. Introduction and Motivation
- 2. Review Batch NDA (Non-parametric Discriminant Analysis)
- 3. Incremental NDA
- 4. Experimental Results
- 5. Conclusions and Future Work

1. Introduction

- Learning is a matter of *extracting meaning* from our experiences for use in the future

- Classical approach = the system is trained with user-tailored data

  Drawback = for real-world applications, it is impossible to estimate beforehand the whole spectrum of possible situations

  (e.g. classification problem: we cannot know exactly the number of classes and the number of instances per class)

- Solution = machine learning should be a continuous process

- Cognitive development for biological systems

For humans, visual learning is an ongoing process:

- new-born babies come pre-wired with the ability of recognized moving stimuli that resemble face-like pattern

- later on, we are able to distinguish different subclasses among 'face' class:
  - gender (male/female)
  - age (young/mature/old)
  - identity

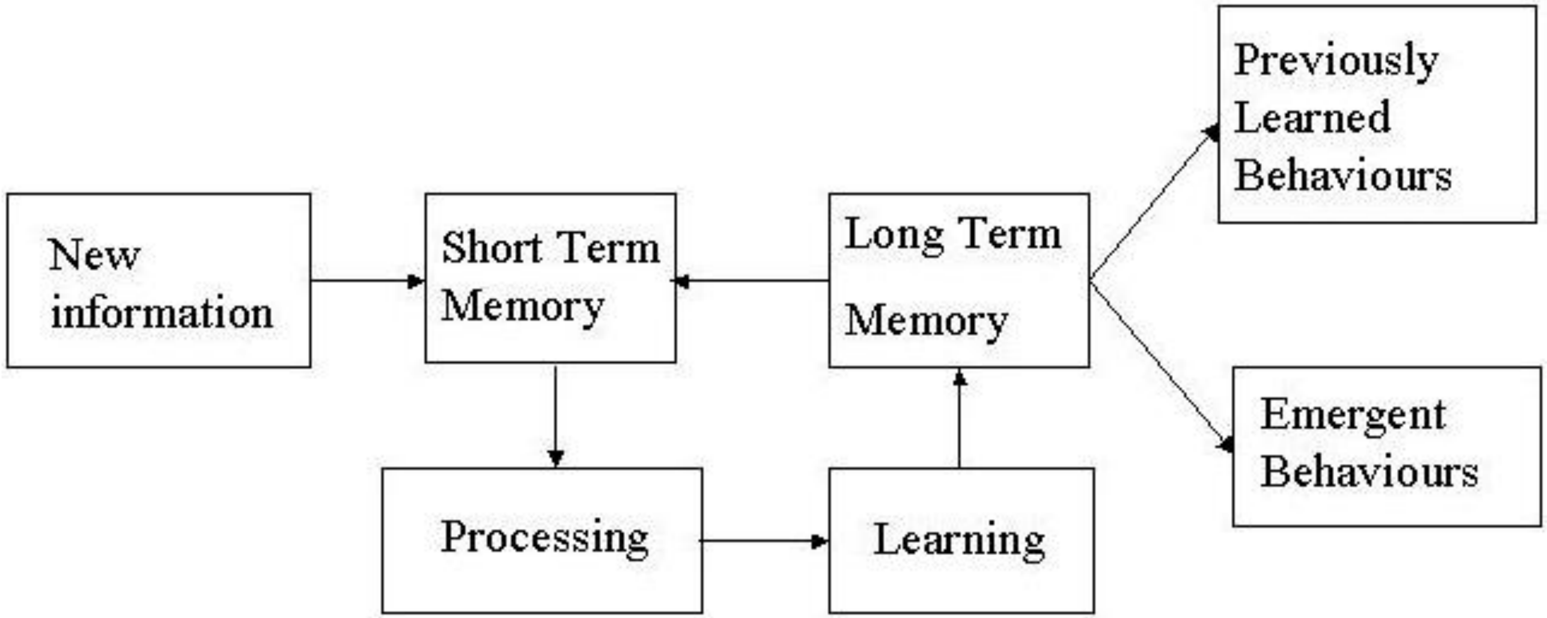
- Cognitive development for artificial systems

The learning process should manifest continuously during its 'life' period

The machine should develop skills through direct interaction with the environment (including humans)

It must develop the ability to learn new tasks which couldn't have been foreseen by the human creator in the design phase

- Cognitive development representation



Centre de Visió per Computador

Classical Paradigm

1. Centered on the 'creator'
2. The 'creator' possesses the knowledge
3. The 'creator' provides the resources
4. Batch training
5. The 'creator' teaches the system
6. Learning process is isolated and limited
7. The system carries 'creator's' vision

Cognitive Perspective

1. Centered on the system
2. The system discovers and build the knowledge
3. The system finds the resources
4. Online training
5. The system learns by itself
6. Learning process is contextual and continuous
7. The system develops its own reality

- We propose a simple way to model cognitive development through incremental learning with automatically derived

  discriminating features

- We will continuously update the knowledge representation of the NDA (Non-Parametric Discriminant Analysis) eigenspace

- The proposed solution is applied to the problem of online face recognition

2. Review Batch NDA

- Let's assume the data samples are divided in *N* classes, C_i, $i=1..N$, i.e.
- We define two scatter-matrices: S_w and S_b in order to express inter-class separability:

$$C_i = \left\{x_1^i, x_2^i, \ldots, x_{n_{C_i}}^i\right\}$$

$$S_w = \sum_{i=1}^{C_N} \sum_{j \in C_i} (x_j - \bar{x}^{C_i})(x_j - \bar{x}^{C_i})^T$$

- NDA $$S_b = \sum_{i=1}^{C_N} \sum_{j=1, j\neq i}^{C_N} \sum_{t=1}^{n_{C_i}} W(C_i, C_j, t)(x_t^i - \mu_{C_j}(x_t^i))(x_t^i - \mu_{C_j}(x_t^i))^T$$

$$\left(S_w^{-1} S_b\right)$$

where $\overline{x}^{C_i}$ is the mean of class C_i and

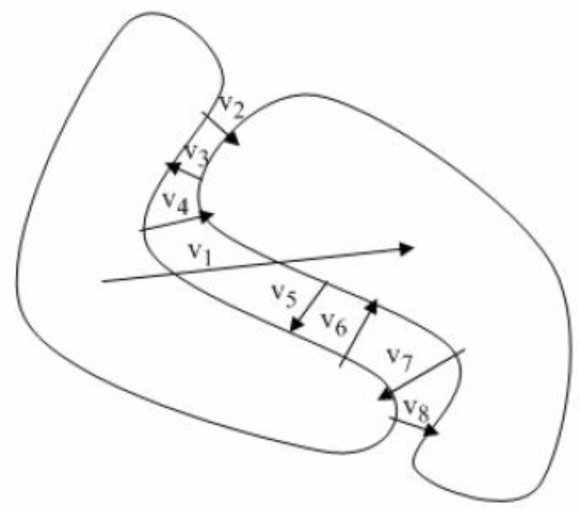
$$\mu_{C_j}(x_t^i) = \frac{1}{k}\sum_{p=1}^{k} NN_p(x_t^i, C_j)$$

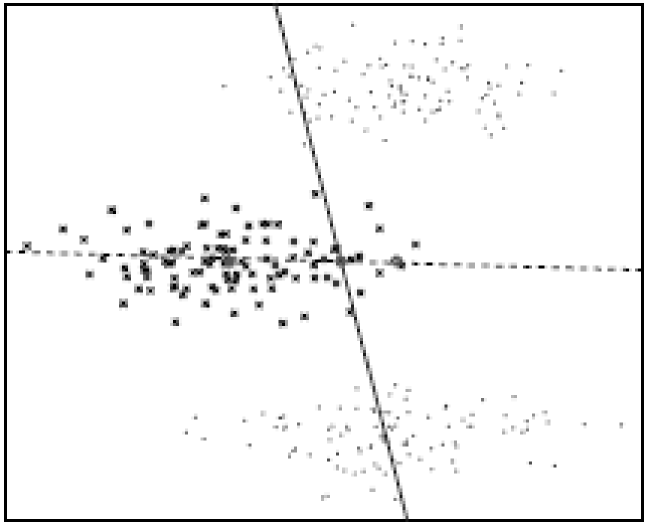
with $NN_p(x_t^i, C_j)$ ng the p-th nearest neighbor from the vector $\left(x_t^i\right)$ to the class C_j and

$$W(C_i, C_j, t) = \frac{min\{d^{\alpha}(x_t^i, NN_k(x_t^i, C_i)), (x_t^i, NN_k(x_t^i, C_j))\}}{d^{\alpha}(x_t^i, NN_k(x_t^i, C_i)) + d^{\alpha}(x_t^i, NN_k(x_t^i, C_j))}$$

- NDA vs. LDA

- being a non-parametric method, its application is not limited to Gaussian distributions
- it is more effective in capturing border information between classes
- it extracts those features which work well with NN classifier





Centre de Visió per Computador

3. Incremental NDA

3.1 The new pattern belongs to an existing class

($y \in C_L$, with $1<L<N$)

- S_b update

$$S_b' = S_b - S_b^{in}(C_L) + S_b^{in}(C_{L'}) + S_b^{out}(y^{C_L})$$

where

$$C_{L'} = C_L \cup \{y^{C_L}\}$$

$$S_b^{in}(C_L) = \sum_{j=1, j\neq L}^{C_N} \sum_{i=1}^{n_{C_j}} W(C_j, C_L, i)(x_i^j - \mu_{C_L}(x_i^j))(x_i^j - \mu_{C_L}(x_i^j))^T$$

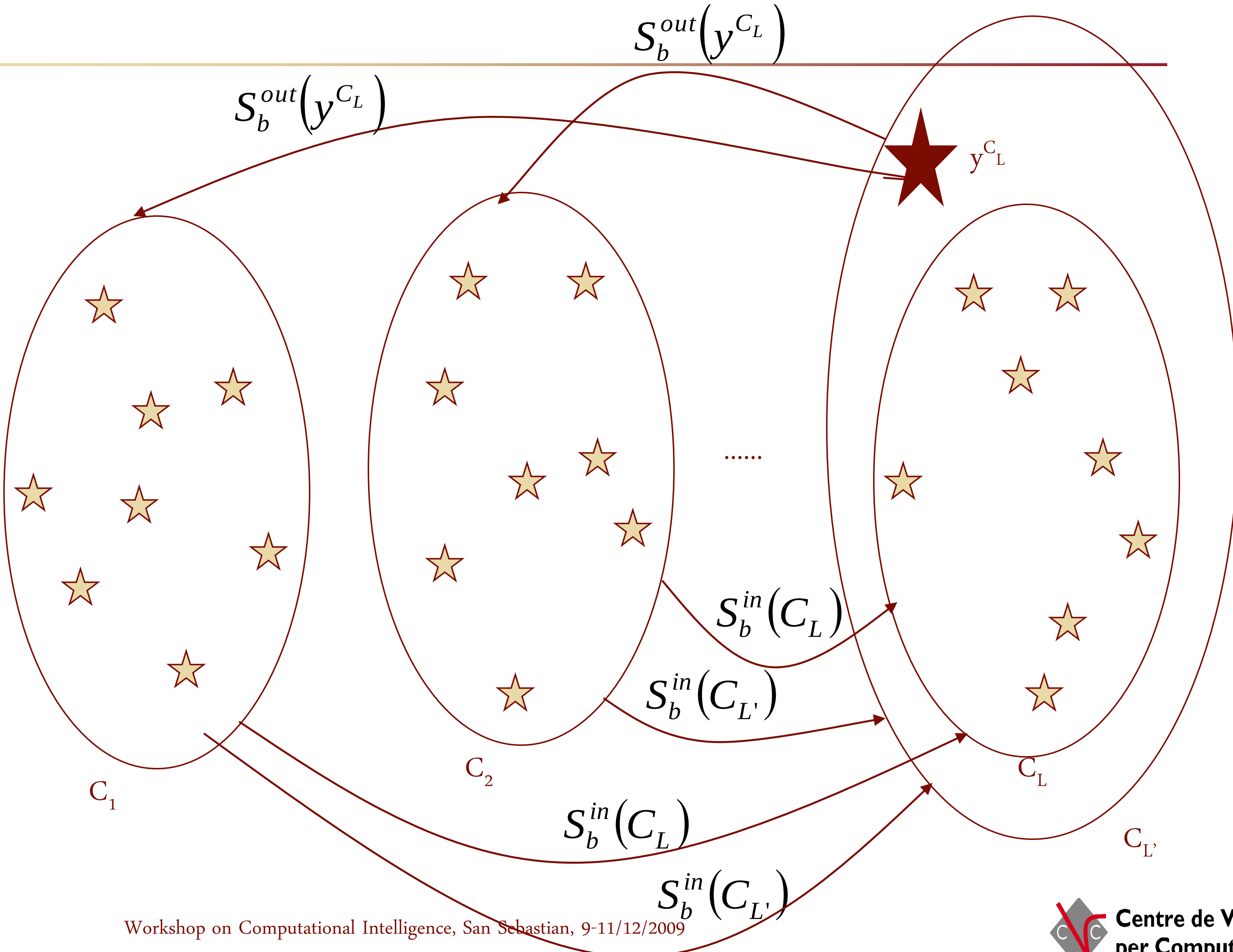
$$S_b^{out}(y^{C_L}) = \sum_{j=1, j\neq L}^{C_N} (y^{C_L} - \mu_{C_j}(y^{C_L}))(y^{C_L} - \mu_{C_j}(y^{C_L}))^T$$

- S_w update

$$S'_w = \sum_{j=1, j \neq L}^{C_N} S_w(C_j) + S_w(C_{L'})$$

where

$$S_w(C_{L'}) = S_w(C_L) + \frac{n_{C_L}}{n_{C_L}+1}(y - \bar{x}^{C_L})(y - \bar{x}^{C_L})^T$$



$S_b^{out}(y^{C_L})$
$S_b^{out}(y^{C_L})$
y^{C_L}
......
$S_b^{in}(C_L)$
$S_b^{in}(C_{L'})$
$S_b^{in}(C_L)$
$S_b^{in}(C_{L'})$
C_1
C_2
C_L
$C_{L'}$

3.2 The new pattern belongs to a new class ($y \in C_L$, with $L>N$)

- S_b update

$$S'_b = S_b + S_b^{out}(C_{N+1}) + S_b^{in}(C_{N+1})$$

where

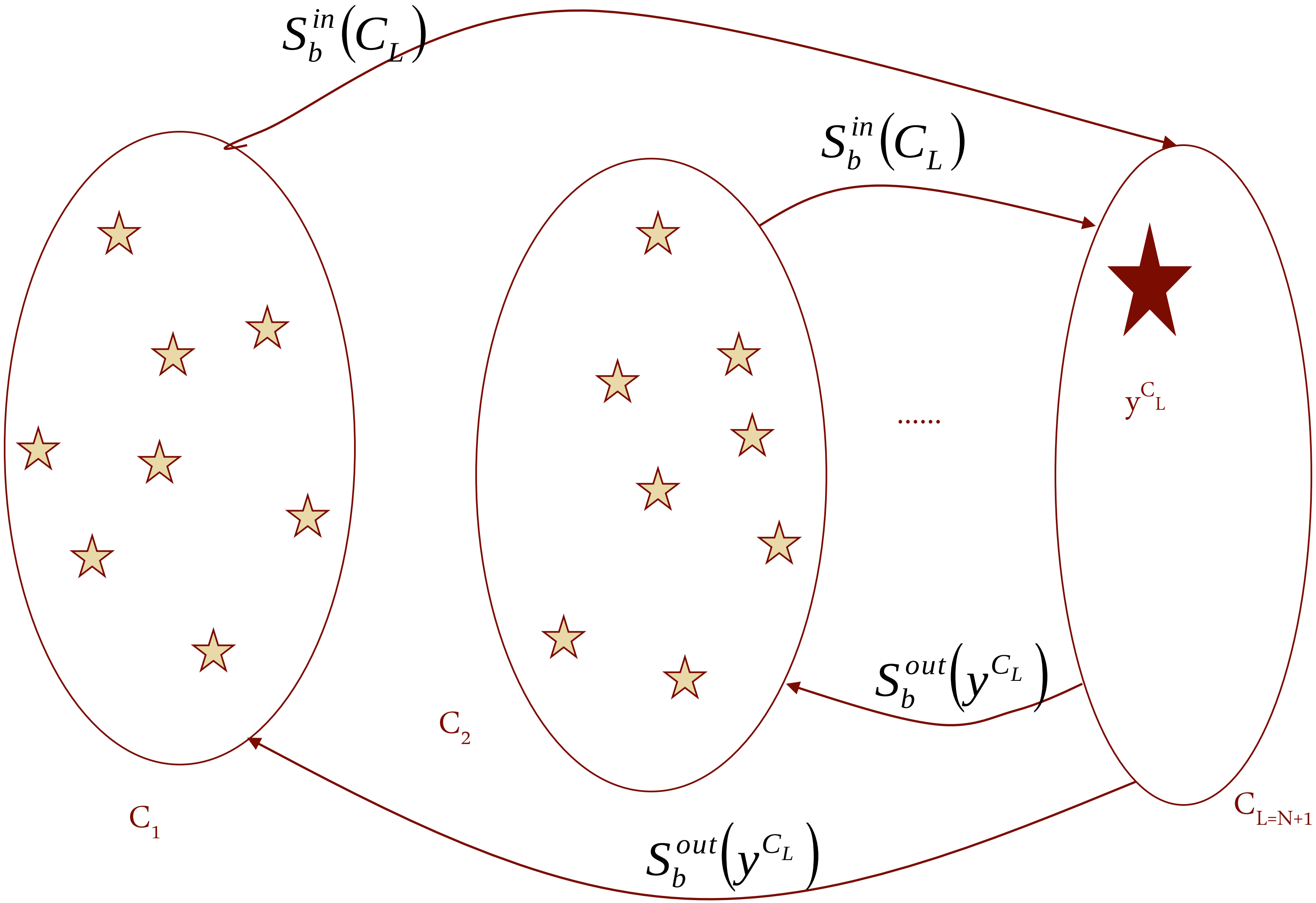
$$S_b^{out}(C_{N+1}) = \sum_{j=1}^{C_N} (y^{C_{N+1}} - \mu_{C_j}(y^{C_{N+1}}))(y^{C_{N+1}} - \mu_{C_j}(y^{C_{N+1}}))^T$$

$$S_b^{in}(C_{N+1}) = \sum_{j=1}^{C_N} \sum_{i=1}^{n_{C_j}} W(C_j, C_{N+1}, i)(x_i^j - \mu_{C_{N+1}}(x_i^j))(x_i^j - \mu_{C_{N+1}}(x_i^j))^T$$

- S_w update

$$S'_w = S_w$$

Centre de Visió per Computador



$S_b^{in}(C_L)$
$S_b^{in}(C_L)$
y^{C_L}
......
$S_b^{out}(y^{C_L})$
C_2
C_1
$C_{L=N+1}$
$S_b^{out}(y^{C_L})$

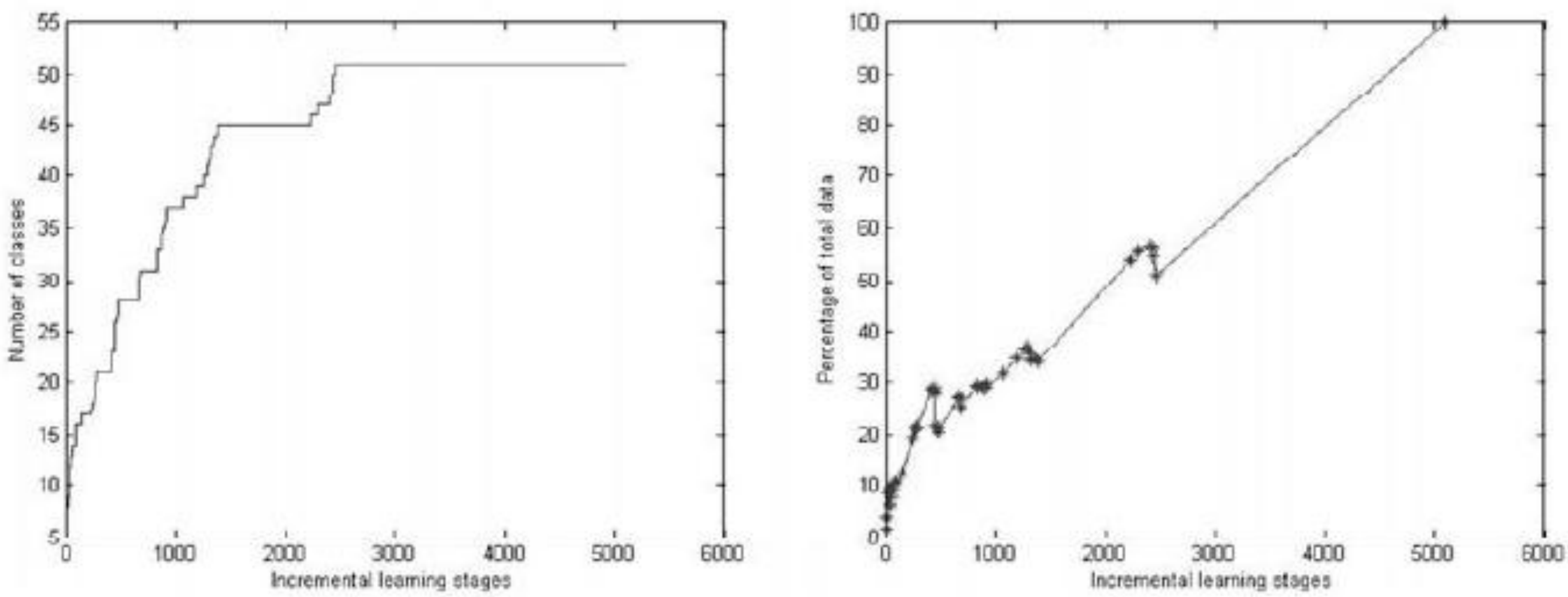
4. Experiments on Face Recognition

We use the CVC face database:

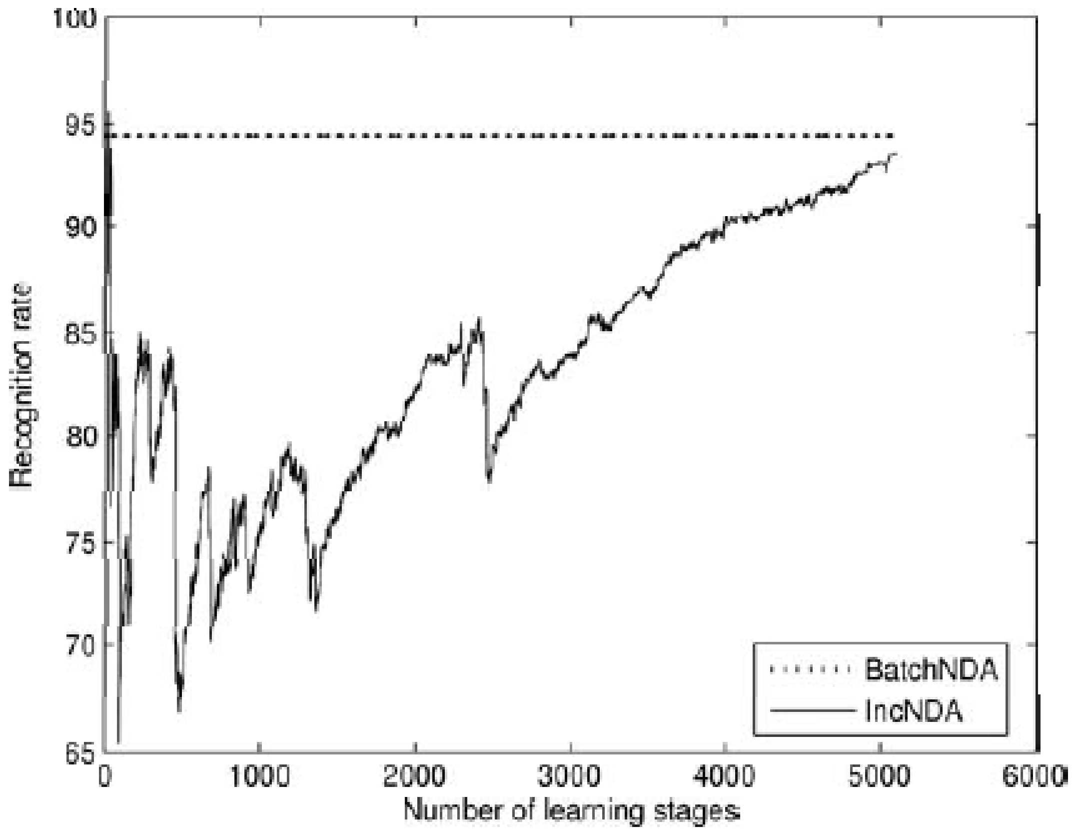
- consists of 6882 images of 51 people
- For the purpose of the current study, we preferred to have a reasonable number of classes with a lot of instances instead of having a large number of classes with very few instances
- images were taken in an uncontrolled environment during several weeks
- the number of instances per class varies between 20 and 400
- face segmentation has been performed automatically based on the Viola-Jones detector
- image size: 48x48 pixels
- no image preprocessing (face alignment or intensity normalization) has been performing, simulating an 'ad-hoc' application

Some instances of the faces used in our experiment

- Experimental details:
- we chose 90% of samples (about 6000 images) as training set and the remaining 10% as test set
- from the training set, we chose 15% of samples (about 900 images) to initialize the S_w and S_b matrices
- the remaining 5100 samples from the training set have been added sequentially in order to update the NDA eigenspace representation
- in order to avoid the singularity problem, a PCA has been applied beforehand
- for classification, we used the NN rule (considering 1,3,5,7 neighbors)
- recognition rate achieved: aprox. 95%



Learning process: evolution of the number of classes function of learning stages (left) and the percentage of training data function of learning stages (right)



In order to assess the accuracy of our approach, we compared it with BatchNDA. We can see that at the end of the learning process, IncNDA converges towards BatchNDA

Centre de Visió per Computador

Method	Number of samples										
	200	400	600	800	1000	1200	1400	1600	1800	2000	2200
BatchNDA	0.46	1.51	3.17	4.53	6.57	8.54	10.65	13.03	16.09	18.73	22.14
IncNDA	0.03	0.04	0.07	0.09	0.12	0.15	0.17	0.20	0.21	0.26	0.29

Comparison between BatchNDA and IncNDA in terms of computational

complexity: execution time tocalculate the S_b matrix (in seconds)

Some instances of misclassified faces

• Applications to Social Robotics

- oriented towards richer forms of communication between persons and robots

- the robots are expected to develop self-awareness capabilities, in order to be able to initiate, respond and maintain coherent interaction with the user and with the environment

- behave in a personalized manner when a certain user has been identified and whose preferences are known

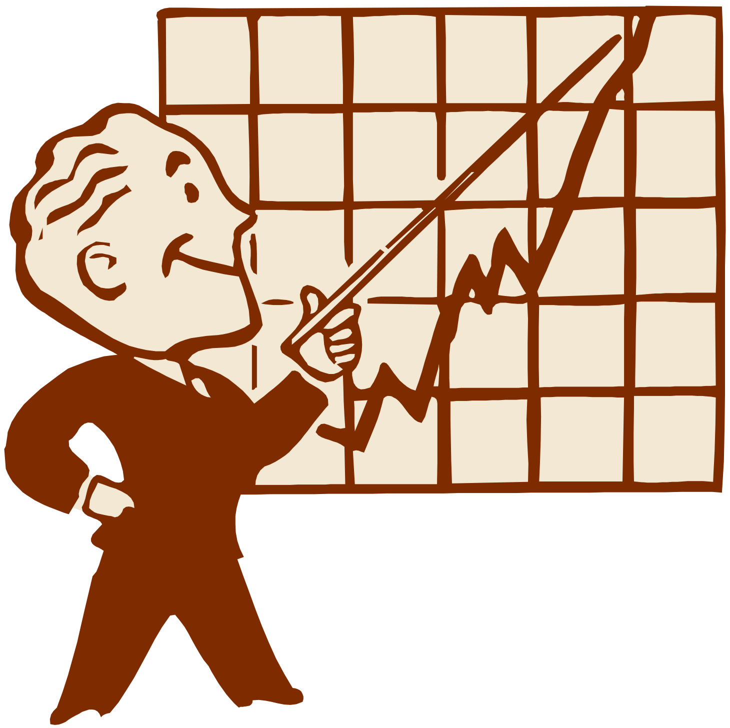
- most important areas of applications:

    - elderly people living alone (improve quality of life, but also monitor the state of the person)

    - pediatrics (studies on children with autism)

5. Conclusions and future work

- The current approaches using ‘batch’ learning strategies have proven their limitation for real-world applications
- We introduced an online version of NDA, which uses a sequential updating of the eigenspace representation
- We demonstrated that at the the end of the learning process, IncNDA converges towards BatchNDA
- In the future we want to perform chunk-updating of the NDA eigenspace
- We plan also to introduce the notion of ‘decremental learning’, to simulate the forgetting process of the long-term memory
- Implement as a real application for AIBOs: the expected result would be a personalized behavior, depending on the recognized person



Thank you very much for your attention!!!