# Incremental Subspace Learning for Cognitive Visual Processes

Bogdan Raducanu[1] and Jordi Vitrià[1,2]

[1] Computer Vision Center, Building "O" - Campus UAB
[2] Computer Science Dept., Autonomous University of Barcelona (UAB)
08193 Bellaterra, Barcelona
Spain
E-mail: {bogdan, jordi}@cvc.uab.es

**Abstract.** In real life, visual learning is supposed to be a continuous process. Humans have an innate facility to recognize objects even under less-than-ideal conditions and to build robust representations of them. These representations can be altered with the arrival of new information and thus the model of the world is continuously updated. Inspired by the biological paradigm, we propose in this paper an incremental subspace representation for cognitive vision processes. The proposed approach has been applied to the problem of face recognition. The experiments performed on a custom database show that at the end of incremental learning process the recognition performance achieved converges towards the result obtained using an off-line learning strategy.

## 1 Introduction

The human visual cognitive system is very robust among a large range of variations in environmental conditions. Opposite to this, a similar robustness of visual cognition is still far to be achieved with artificial systems. Despite of the progresses reported in areas like vision sensors, statistical pattern recognition and machine learning, what for humans represents a natural process, for machines is still a far-fetched dream. One of the factors that limit these performances is the learning strategy that has been used. Most of the nowadays approaches, require the intervention of the human operator to collect, store and segment hand-picked images and train pattern classifiers with them.[3]. It is unlikely that such a manual operation could meet the demands of many challenging cognition tasks that are critical for generating intelligent behavior, such as object recognition, in general, and face recognition, in particular. The desired goal is to enable machines to learn directly from sensory input streams while interacting with the environment, including humans. During the interaction, the human is not allowed to interfere in the internal state of the system [2].

The cognitive approach in generating intelligent behavior consists of an integrated, recursive process that aims at building a model of the 'world' and a

[3] In real world scenarios, it is unlikely to know beforehand the number of total classes or the exact number of instances per class

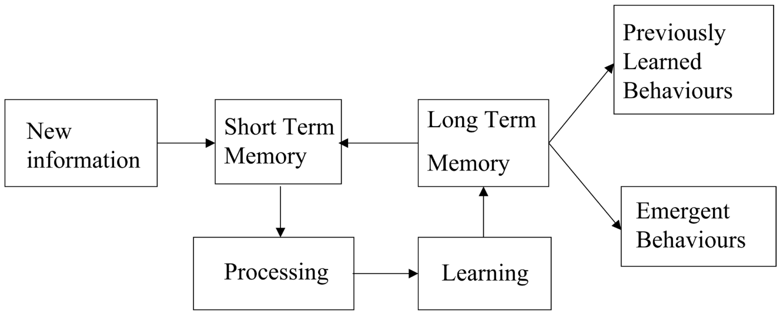


**Fig. 1.** The structure of cognitive processes

continuous adaptation of this model [8]. In consequence, it is the system itself which is responsible of how to analyze, interpret and represent the information. The system will learn new concepts (develop new competencies) based on previous data and the experience acquired over time. When a new piece of information becomes available, it is responsible to analyze it and in case it is relevant, should be added to the existing representation (at times, might be necessary a change of representation structure). This way, the system could present two classes of behaviors: one class, consisting of specifically learned behaviors and another one, corresponding to emergent behaviors. This cognitive strategy is depicted in figure 1.

A very important characteristic of cognitive processes is represented by information management. We have to distinguish between two types of memory: a short-term and a long-term memory. The short-term memory is responsible for maintaining the information for a very brief period of time (acting like a buffer), after which it is discarded. On the other hand, the long-term memory represents the knowledge database built over time. If the information passed from short-term memory is relevant, than the knowledge content of long-term memory is updated. The long-term memory is responsible for guaranteeing the system viability over large period of times (weeks, maybe months or years). A phenomenon that can affect the long-term memory is the 'forgetting' or 'degradation' (partial or total loss of some data). Sometimes, this process can be irreversible. These are also characteristics of human mind.

In the current paper we will focus on the long-term memory, i.e. how the knowledge database can be built incrementally. We introduce an online version of the non-parametric discriminant analysis (NDA)[6]. The proposed solution is applied to the problem of face recognition and is presented as an application for social robotics. The paper is structured as follows: in the next section, we will present a comparative between incremental learning in biological and artificial systems. Section 3 is dedicated to the introduction of the novel incremental non-parametric discriminant analysis (from now on referred as IncNDA). In section 4 we discuss the application of our approach to the problem of face recognition. We will show that at the end of the learning process, the recognition performance achieved converges towards the result obtained using an off-line version of the NDA (from now on referred as BatchNDA). Finally, section 5 contains our conclusions and the guidelines for future work.

## 2 Incremental Learning in Biological and Artificial Systems

Incremental learning is associated with evolutive processes where a standard learning mechanism is combined with or is influenced by stepwise adjustments during the learning process. These adjustments can be changes in the structure or parameters of the learning system or changes in the presentation or constitution of the input signals. For biological systems, the 'incremental learning' is codified in the genetical material. It starts to run at the time of conception of each entity. This 'program' is responsible for whatever can happen through the entire life span of that individual. Let's take as an example the development of human visual learning system. In [5], the authors claim that the newborn babies arrive to this world pre-wired with the ability to recognize face-like patterns. It looks like that they are attracted by moving stimuli which resemble human faces. Later on, and according with the evolution of our cognitive abilities, we learn to distinguish different subclasses within face class: males/females, young/mature/old, familiar/unfamiliar, etc. [7]

In its most general sense (by analogy with their biological counterpart), the 'incremental learning' for an artificial system should start to manifest at its 'birth'. This process enables the machine to develop skills through direct interactions with its environment through its perceptual mechanisms. For machines to truly understand the world, the environment must be the physical world, including humans and the machine itself. It must enable the machine with ability to learn new tasks that a human creator cannot foresee in the design phase. This implies that the representation of any task that the machine learns must be generated by the machine itself.

In the context of the current paper, we will refer to 'incremental learning' with the acceptance of 'online pattern training'. In this case, the initial representation of the knowledge is continuously updated, as new patterns become available. Visual learning in the case of artificial systems is often approached by the appearance based modelling of objects. Object modelling is often followed by a feature selection and extraction step. The outcome of this process consists of obtaining either an efficient data representation (through dimensionality reduction, when class labels are ignored) or an effective data discrimination (when besides the dimensionality reduction, we are focused also on class labels) [10]. For the latter, parametric and non-parametric forms have been proposed [4].

So far, several online knowledge representations have been proposed. In [3, 9, 1] the Incremental Principal Component Analysis (IPCA) is presented. The update of the covariance matrix is achieved through a residual procedure. They keep only the learned coefficients of the eigenspace representation and discard the original data. In the same context of IPCA, in [15] it is demonstrated that is possible to build incrementally an eigenspace representation without the need to compute the covariance matrix at all. On the other hand, some incremental versions of Linear Discriminant Analysis (ILDA) are proposed in [13] and [12]. In the next section we present a brief review of the classical NDA and introduce its online version. Our choice for NDA is motivated by the fact that being a non-parametric method, its application is not limited to gaussian distributions of data. Another advantage provided by this method is that it extracts those features which work well with the nearest-neighbor classifier [11].

## 3 Non-parametric Discriminant Analysis

As introduced in [6], the within-class scatter matrix $S_w$ and between-class scatter matrix $S_b$ are used as a measure of inter-class separability. One of the most used criteria is the one that maximize the following expression:

$$\zeta = tr(S_b S_w) \tag{1}$$

It has been shown that the $M \times D$ linear transform that satisfies the equation 2 optimizes also the separability measure $\zeta$:

$$\hat{W} = \arg \max_{W^T S_w W = I} tr(W^T S_b W) \tag{2}$$

This problem has an analytical solution and is mathematically equivalent to the eigenvectors of the matrix $S_w^{-1} S_b$.

### 3.1 BatchNDA

Let's assume that the data samples we have belong to $N$ classes $C_i, i = 1, 2, ..., N$. Each class $C_i$ is formed by $n_i$ samples $C_i = \{x_1^i, x_2^i, ..., x_{n_{C_i}}^i\}$. By $\bar{x}^{C_i}$ we will refer to the mean vector of class $C_i$. According to [6], the $S_w$ and $S_b$ scatter matrices are defined as follows:

$$S_w = \sum_{i=1}^{C_N} \sum_{j \in C_i} (x_j - \bar{x}^{C_i})(x_j - \bar{x}^{C_i})^T \tag{3}$$

$$S_b = \sum_{i=1}^{C_N} \sum_{j=1, j \neq i}^{C_N} \sum_{t=1}^{n_{C_i}} W(C_i, C_j, t)(x_t^i - \mu_{C_j}(x_t^i))(x_t^i - \mu_{C_j}(x_t^i))^T \tag{4}$$

where $\mu_{C_j}(x_t^i)$ is the local $K-$NN mean, defined by:

$$\mu_{C_j}(x_t^i) = \frac{1}{k} \sum_{p=1}^{k} NN_p(x_t^i, C_j) \tag{5}$$

where $NN_p(x_t^i, C_j)$ is the $p-$th nearest neighbor from vector $(x_t^i)$ to the class $C_j$. The term $W(C_i, C_j, t)$ which appears in equation 4 is a weighting function whose role is to emphasize the boundary class information. It is defined by the following relation:

$$W(C_i, C_j, t) = \frac{min\{d^\alpha(x_t^i, NN_k(x_t^i, C_i)), (x_t^i, NN_k(x_t^i, C_j))\}}{d^\alpha(x_t^i, NN_k(x_t^i, C_i)) + d^\alpha(x_t^i, NN_k(x_t^i, C_j))} \tag{6}$$

Here $\alpha$ is a control parameter that can be selected between zero and infinity. The sample weights take values close to 0.5 on class boundaries and drop to zero as we move away. The parameter $\alpha$ adjusts how fast this happens.

### 3.2 IncNDA

The shortcoming of the BatchNDA described in the previous section, is that assumes that all the data are available at the classification. This is not the case for real applications, when the data is coming over time, at random time intervals, and the representation of the data must be updated. Computing from the beginning the scatter matrices, each time a new sample arrives, is not computationally feasible, especially when the number of classes is very high and the number of samples per class increases significantly. For this reason, we propose the IncNDA technique, that can process sequentially later-on added samples, without the need for recalculating entirely the scatter matrices. In order to describe the proposed algorithm, we assume that we have computed the $S_w$ and $S_b$ scatter matrix from at least 2 classes. Let's now consider that a new training pattern $y$ is presented to the algorithm. We distinguish between two situations.

**The new training pattern belongs to an existing class** Let's assume, for instance, that $y$ belongs to one of the existing classes $C_L$ (i.e. $y^{C_L}$, where $1 < L < N$). In this case, the equation that updates $S_b$ is given by:

$$S_b' = S_b - S_b^{in}(C_L) + S_b^{in}(C_{L'}) + S_b^{out}(y^{C_L}) \tag{7}$$

where $C_{L'} = C_L \bigcup \{y^{C_L}\}$, $S_b^{in}(C_L)$ represents the covariance matrix between the existing classes and the class that is about to be changed, $S_b^{in}(C_{L'})$ represents the covariance matrix between existing classes and the updated class $C_{L'}$ and by $S_b^{out}(y^{C_L})$ we denote the covariance matrix between the vector $y^{C_L}$ and the other classes:

$$S_b^{in}(C_L) = \sum_{j=1, j\neq L}^{C_N} \sum_{i=1}^{n_{C_j}} W(C_j, C_L, i)(x_i^j - \mu_{C_L}(x_i^j))(x_i^j - \mu_{C_L}(x_i^j))^T \tag{8}$$

$$S_b^{out}(y^{C_L}) = \sum_{j=1, j\neq L}^{C_N} (y^{C_L} - \mu_{C_j}(y^{C_L}))(y^{C_L} - \mu_{C_j}(y^{C_L}))^T \tag{9}$$

In the case of $S_w'$ the update equation is the following:

$$S'_w = \sum_{j=1, j \neq L}^{C_N} S_w(C_j) + S_w(C_{L'}) \tag{10}$$

where

$$S_w(C_{L'}) = S_w(C_L) + \frac{n_{C_L}}{n_{C_L} + 1}(y - \bar{x}^{C_L})(y - \bar{x}^{C_L})^T \tag{11}$$

**The new training pattern belongs to a new class** Let's assume that $y$ belongs to a new class $C_{N+1}$ (i.e. $y^{C_{N+1}}$).

For this case, the updated equations for the scatter matrices are given by:

$$S'_b = S_b + S_b^{out}(C_{N+1}) + S_b^{in}(C_{N+1}) \tag{12}$$

where $S_b^{out}(C_{N+1})$ and $S_b^{in}(C_{N+1})$ are defined as follows:

$$S_b^{out}(C_{N+1}) = \sum_{j=1}^{C_N} (y^{C_{N+1}} - \mu_{C_j}(y^{C_{N+1}}))(y^{C_{N+1}} - \mu_{C_j}(y^{C_{N+1}}))^T \tag{13}$$

$$S_b^{in}(C_{N+1}) = \sum_{j=1}^{C_N} \sum_{i=1}^{n_{C_j}} W(C_j, C_{N+1}, i)(x_i^j - \mu_{C_{N+1}}(x_i^j))(x_i^j - \mu_{C_{N+1}}(x_i^j))^T \tag{14}$$

Regarding, the new $S'_w$ matrix, this one remains unchanged, i.e:

$$S'_w = S_w \tag{15}$$

## 4 Face Recognition: A Case Study

The incremental learning approach introduced in the previous section has been tested on a face recognition problem using a custom face database. The image acquisition phase was extended over several weeks and was performed in an automatic manner. For this purpose, we put the camera in an open space and snapshots were taken each time a person was passing in front of it. The face was automatically extracted from the image using the face detector based on [14]. We didn't impose any restrictions regarding ambient conditions.

Overall, our database consists of 6882 images of 51 people (both male and female)[4]. Since no arrangements were previously made, some classes contain only a handful of images (as much as 20), meanwhile, the largest of them contains over 400. Segmented faces were normalized at a standard size of 48x48 pixels. Because

[4] In the current study we put the accent in having a reasonable number of classes with a lot of instances rather having an excessive number of classes with very few instances

of the particularity of the acquisition process, face images reflect the changes in appearance suffered by subjects over time. Furthermore, since our application was thought to run in real-time (and to give it a more ad-hoc impression), we didn't perform any pre-processing step to face images before passing them to the classifier. That's why the faces used in the experiment show a certain degree of variation in pose and size and are not constrained to be exactly frontal. For the same reason, face images used to be a little wider than the face region itself. Some samples of these face images are presented in figure 2.

**Fig. 2.** Samples of face images from CVC custom database showing a certain degree of variation in illumination, pose and size

To test the IncNDA technique, we used 90% of the images (i.e. about 6000) as training set and the remaining ones as test set. From the training set, we used 15% of the images (belonging to 5 classes and representing 900 samples) to build the initial IncNDA eigenspace. In order to overcome the singularity problem, a PCA step was performed beforehand[5]. This way, data dimensionality was downsized from 2304 to 60. The remaining samples (5100) from the training set

[5] Because the dimensionality of a typical image (i.e. the number of pixels in it) is usually much larger than the number of available samples, the scatter matrices might be singular. In order to avoid this phenomenon, a dimension reduction procedure (PCA) is applied previously.

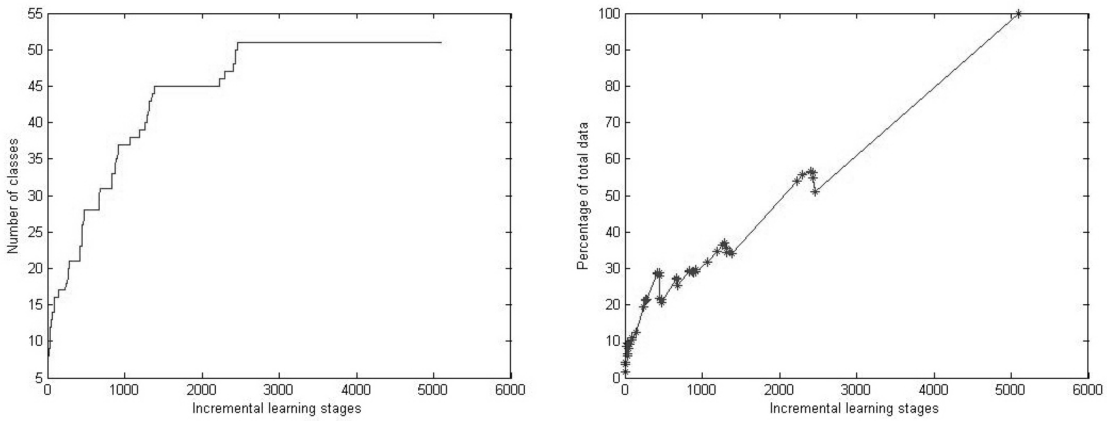


**Fig. 3.** Learning process: evolution of the number of classes function of learning stages (left) and the percentage of the training data function of learning stages (right)

were added later on in a sequential manner (the samples were drawn randomly) and this way the NDA-eigenspace was updated.

In figure 3 (left) we depicted the evolution of the learning process after each update (a new sample added) of the initial IncNDA eigenspace. In the early stages, there are a lot of new classes presented at very short intervals. It can be appreciated that, with almost 50% of the remaining training samples introduced, all classes have been represented. In figure 3 (right), we depicted the percentage of incremental training samples introduced so far (the stars represent the moment when a new class has been added). This graphic should be read in concordance with the above one.

As a final proof of accuracy, we compared IncNDA with the BatchNDA. In figure 4, we show that indeed the IncNDA is converging (at the end of the learning process) towards BatchNDA. The common recognition rate achieved is around 95%, which in our opinion is a very good result, taking into account the difficulty of the database. Both graphics were plotted after averaging the results obtained from a ten-fold cross-validation procedure (the training samples were chosen in a random manner in each run). We repeated the experiments considering different number of neighbors (1, 3, 5, 7) in computing the equation 4, but the best results obtained correspond to a number of neighbors equal to 3. The figure 4 corresponds to this case. The oscillation of the IncNDA in its early stages corresponds to the situation when a significant number of new classes have been added at very short intervals and only a very few samples of those classes were available. After some learning stages, when enough samples for each

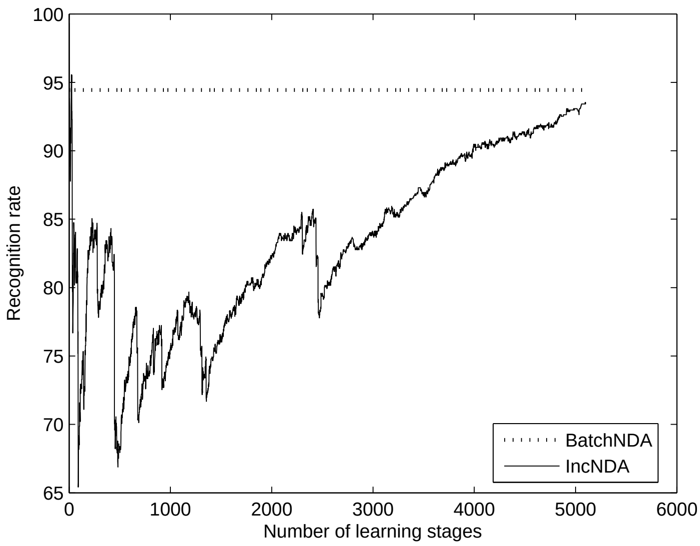


**Fig. 4.** IncNDA vs. BatchNDA curves. IncNDA converges towards BatchNDA at the end of the learning process

class became available, we can appreciate that the evolution curve regulates its tendency and becomes constantly ascending.

## 5 Conclusions and Future Work

In the current paper we presented some aspects regarding the cognitive development in biological and artificial systems. By using an incremental learning strategy, we showed how a knowledge representation can be continuously updated, with the arrival of new information. For this purpose, we introduced a novel approach represented by the online non-parametric discriminant analysis. This learning strategy has been tested on a face recognition problem. In the future, we will study the possibility to replace the sequential way of updating the knowledge representation by a parallel one, in which we present data chunks of variable size. Another research direction is represented by the analysis of decremental learning, which emulates the 'forgetting' process in humans: those patterns which became irrelevant are removed from the knowledge representation after a certain period of time.

## Acknowledgements

This work is supported by MEC Grant TIN2006-15308-C02, Ministerio de Educacin y Ciencia, Spain. Bogdan Raducanu is supported by the Ramon y Cajal research program, Ministerio de Educación y Ciencia, Spain.

## References

1. Artač, M., Jogan, M., Leonardis, A.: Incremental PCA for on-line visual learning and recognition. Proc. of 16th Intl. Conf. Pattern Recognition, **3** (2002) 781-784, Québec, Canada
2. Brooks, R.A.: Intelligence without Reason. Proc. of International Joint Conference on Artificial Intelligence (IJCAI), (1991) 569-595 Sydney, Australia
3. Chandrasekaran, S., Manjumath, B.S., Wang, Y.F., Winkler, J., Zhang, H.: An eigenspace update algorithm for image analysis. Graphical Models Image Processing, **59**(5) (1997) 321-332
4. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification (2nd Ed). John Wiley and Sons, New York, USA (2001)
5. Fischler, M.A., Elschlager, R.A.: The Representation and Matching of Pictorial Structures. IEEE Transactions on Computers, **COM-22** (1973) 67-92
6. Fukunaga, K: Introduction to Statistical Pattern Recognition (2nd Ed). Academic Press, Boston, USA (1990)
7. de Gelder, B., Rouw, R.: Beyond Localisation: A Dynamical Dual Route Account of Face Recognition. Acta Psychologica, **107** (2001) 183207
8. Grow, G.O.: A Cognitive Model of Learning. Part of the paper: Serving the Strategic Reader: Cognitive Reading Theory and Its Implications for the Teaching of Writing (1996). Available on-line at: *http://www.longleaf.net/ggrow*. Original paper available as Eric Documentation Reproduction Service No. ED 406 644.
9. Hall, P., Marshall, D., Martin, R.: Incremental Eigenanalysis for Classification. Proc. of British Machine Vision Conference, **1** (1998) 286295, Southampton, UK
10. Martinez, A.M., Kak, A.C.: PCA versus LDA. IEEE Trans. on Pattern Analysis and Machine Intelligence, **23**(2) (2001) 228-233
11. Masip, D., Kuncheva, L.I., Vitrià, J.: An Ensemble-based Method for Linear Feature Extraction for two-class problems. Patterns Analysis and Applications, **8** (2005) 227-237
12. Pang, S,, Ozawa, S., Kasabov, N.: Chunk Incremental LDA Computing on Data Streams. International Symposium on Neural Networks, vol. LNCS 3497 (Wang et al, Eds.), (2005) pp. 51-56, Chongqing, China
13. Skočaj, D., Uray, M., Leonardis, A:, Bischof, H.: Why to Combine Reconstructive and Discriminative Information for Incremental Subspace Learning. Proc. of Computer Vision Winter Workshop (Chum et al, Eds.), (2006) pp. N/A, Telč, Czech Republic
14. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. International Journal of Computer Vision, **57** (2004) 137-154
15. Weng, J., Zhang, Y., Hwang, W.-S.: Candid covariance-free incremental principal component analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence, **25**(8) (2003) 1034-1040