Why to Combine Reconstructive and Discriminative Information for Incremental Subspace Learning

Danijel Skočaj[1], Martina Uray[2], Aleš Leonardis[1], and Horst Bischof[2]

[1] University of Ljubljana, Faculty of Computer and Information Science, Tržaška 25, SI-1001 Ljubljana, Slovenia
danijel.skocaj@fri.uni-lj.si, ales.leonardis@fri.uni-lj.si

[2] Graz University of Technology, Institute for Computer Graphics and Vision, Inffeldgasse 16/II, 8010 Graz, Austria
uray@icg.tu-graz.ac.at, bischof@icg.tu-graz.ac.at

Abstract *In the paper we propose a novel method for incremental visual learning by combining reconstructive and discriminative subspace methods. This is achieved by embedding LDA learning and classification into the incremental PCA framework. The combined subspace consists of a truncated PCA subspace and a few additional basis vectors that encompass the discriminative information, which would be lost by the discarded principal vectors. As such it contains both sufficient reconstructive information to enable incremental learning, and the previously extracted discriminative information to enable efficient classification as well. We demonstrate that we are able to efficiently update the current model with new instances of the already learned classes as well as to introduce new classes.*

1 Introduction

Visual learning and recognition/categorization has become an important and popular research topic in the computer vision community. Several different methods have been proposed in recent years. Based on the type of object representations they use, most of them can be classified in one of two main categories: reconstructive or discriminative methods. The *reconstructive* representations strive to be as informative as possible in terms of well approximating the original data. Their goal is predominantly to encompass the variability of the training data and are as such not task-dependent. On the other hand, *discriminative* methods usually do not provide good reconstruction of the data, they are task-dependent, but spatially and computationally much more efficient and often give superior classification results compared to the reconstructive methods.

We will study the properties of these two types of methods from the perspective of *incremental learning*. Incremental learning is very often a desirable or even essential property of an artificial cognitive system. In contrast to the batch approaches, which process all training images simultaneously, incremental methods process one image after another. Thus, only one image (or maybe a few of them) is processed at each step, and only the *representations* of the previously encountered images are available; the original training images are discarded immediately after being processed. The

advantages of an incremental over the batch method are that not all training images have to be given in advance (enabling online learning), that less calculation time is needed (to update a model is less expensive than to build a new model from scratch) and that less storage is required (since only representations of the images are being kept). In the case of *reconstructive* methods, these representations can be used as a good approximations of the discarded training images but in general, this does not hold true for *discriminative* methods due to the lack of the information that would enable a good reconstruction. Thus, in order to enable incremental updating of discriminative representations as well, we have to combine them with the reconstructive methods.

This is the problem that we want to address in this paper. We will create a representation, which combines the reconstructive models and discriminative classifiers. The reconstructive property of such a representation will bring sufficient redundancy in the data to enable updating of the representations, while the discriminative property of the representation would still keep the representation efficient and effective.

In this paper we focus on the Principal Component Analysis (PCA) [11] and the Linear Discriminant Analysis (LDA) [4]. PCA is a well known reconstructive method, which encompasses the reconstructive task-independent information that can approximate the training data well. LDA, on the other hand, is a discriminative method, which keeps only the discriminative task-dependent information about input images. While the LDA is recognized to be superior over the PCA in recognition tasks, it is less suitable for incremental learning due to the reasons elaborated above. Therefore we propose to combine both methods to achieve the best of both worlds.

We thus embed the LDA learning and classification into the PCA framework facilitating incremental updating of the already learned representations. The combined subspace consists of a truncated PCA subspace and a few additional basis vectors that encompass the discriminative information, which would be lost by the discarded principal vectors. As such it contains both sufficient reconstructive information to enable incremental learning, and the previously extracted discriminative information to enable efficient classification as well.

The proposed method allows for two types of updating the current representation, thus coping with different aspects of incremental learning. It is possible to add **new instances** of known classes such that the representations of the classes improve and adapt to the new appearances of the known objects/subjects improving the classification results. Additionally it is possible to add **new classes** such that previously not observed classes can be introduced and their representations can be created and then maintained through the process of incremental learning. Both types of learning that fully exploit the reconstructive and discriminative nature of the proposed method are presented and experimentally evaluated demonstrating the advantages of the proposed approach.

The paper is organized as follows. First we discuss related work in Section 2. In Section 3 we introduce the notation and define the problem, while we describe the proposed method in Section 4. To verify our claims we present the experimental results in Section 5. Finally, we summarize the paper, expose the contributions, and outline some possible extensions.

2 Related work

Several research topics are directly or indirectly related to the method presented in this paper: combining the reconstructive and discriminative methods, incremental learning, and combining different subspace methods. In this section we will discuss some of them and expose the differences with respect to the method we are proposing.

Combination of generative and discriminative methods as well as integration of representative and discriminant models have gained a lot of attention, which resulted also in a plethora of methods published very recently [6, 10, 14]. Most of these methods aim at improving the classification results or increasing the robustness. They do not, however, consider incremental learning.

Incremental learning have been better studied in the domain of subspace methods, especially the reconstructive ones. Many methods for incremental building of principal subspaces have been proposed [3, 8] and different extensions have also been introduced, such as weighted [12] and robust incremental learning [13, 22]. These methods are unsupervised and do not take into account the prior information about object labels, thus they do not exploit all information, which is available for classification.

Several methods for incremental LDA have also been already proposed [9, 15, 20, 21, 24]. Most of these methods focus on the updating of the between class scatter matrix and the within class scatter matrix, thus keeping the discriminative information only. In contrast, our method keeps updating the current representation of the images encompassing the discriminative and reconstructive information as well. The richer representation allows for updating of the acquired knowledge in a more powerful way.

PCA and LDA have often been combined in the past. Even some of the methods for incremental LDA mentioned above involve the estimation of PCA subspaces. Also in many other discriminative approaches PCA is first used as a preprocessing step for dimensionality reduction or to avoid

singularity problems [1, 2, 23, 25]. In addition, many approaches aim at improving the classification power of discriminative methods by incorporating the PCA information in different ways [16, 17]. We rather focus on incremental aspects of the learning process; this is the main reason for combining LDA with PCA. In our approach the methods are tightly coupled in a principled way; the LDA-relevant information is being considered during the creation of the PCA subspace as well, resulting in a combined representation, which is the main novelty of the proposed approach.

3 Problem definition

Let n be the number of images in the training set, each of them containing m pixels, aligned in the columns of the matrix $\mathbf{X} \in \mathbf{R}^{m \times n}$, let $\mu \in \mathbf{R}^m$ be the mean image, and c the number of classes the images belong to. The goal of subspace methods is to find subspaces that transform the input data (images) in a way that enables efficient classification of novel images. Reconstructive and discriminative methods offer different solutions to this problem.

Reconstructive methods are designed to find a linear representation that best describes the input data, i.e.[1],

$$\mathbf{X} \approx \mathbf{U}_{\underline{k}}\mathbf{A}_{\underline{k}} + \mu \mathbf{1}_{1 \times n} \qquad (1)$$

where k vectors in the columns of $\mathbf{U}_{\underline{k}} = [\mathbf{u}_1, \dots, \mathbf{u}_k] \in \mathbf{R}^{m \times k}$ form the reconstructive basis and n vectors in the rows of $\mathbf{A}_{\underline{k}} = [\mathbf{a}_1^T, \dots, \mathbf{a}_n^T]^T \in \mathbf{R}^{k \times n}$ are referred to as coefficient vectors, i.e., the k-dimensional representations of the training images.

PCA looks for a low-dimensional representation of the data which minimizes the squared reconstruction error [11]. Therefore it guaranties the best possible representation of the input images in a linear subspace of a given dimension k. PCA is thus an unsupervised method, which does not look for differences between the images belonging to different classes, but rather tries to model each image as well as possible. Hence it keeps as much information about the training images as possible, and stores it in k-dimensional representations (usually $c \ll k \ll n \ll m$). Since the model is not built for a specific task, it is general and task-independent.

Discriminative methods are designed in a different way and are particulary suited for classification tasks. They assume that prior knowledge about the classes of the training data is available, which is integrated in the supervised learning process to produce a small number of hyperplanes that are capable of separating the training data. To be more specific, the objective of discriminative methods is to find a linear function

$$g(\mathbf{x}) = \mathbf{W}^T (\mathbf{x} - \mu) , \qquad (2)$$

where $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{(c-1)}] \in \mathbf{R}^{m \times (c-1)}$ is used for transforming the data into a lower-dimensional classification space upon which it is decided to which class a given sample $\mathbf{x}$ belongs.

LDA finds the projection directions on which the intraclass scatter is minimized whilst the inter-class scatter is

---

[1]$\mathbf{1}_{m \times n}$ denotes a $m \times n$ matrix of ones.

maximized [4]. That is, it finds $c - 1$ vectors that can be used for efficiently separating the images belonging to different classes. The model is thus very compact (c is usually very small) and efficient, but it is task-dependent. And since the projections of the images in the LDA space are very low-dimensional (($c-1$)-dimensional) and $\mathbf{W}$ is not particulary designed to encompass the reconstructive information, they can not be used for reconstruction of the training images.

The comparison between discriminative and reconstructive methods for classification tasks has been a subject of extensive research and testing [1, 18]. The general conclusion was that discriminative methods outperform the reconstructive methods. The explanation for this is rather obvious: the discriminative methods focus more on specific prior knowledge, which can thus be more efficiently integrated into the learning process. However, the latter observation can also be disadvantageous, when we want to put learning in an incremental framework. If the discriminative methods are too focused to specific discriminative features sufficient for classifying the given data, then many features, which may be useful for discriminating in the feature, get discarded, and the representations can not be adapted to new information. Since the images get discarded in the training process, and their representations are rather poor, there is not sufficient information, which would enable reconsidering the discarded training images. A new model, which would consider the discarded images and the new ones, can not be created. To overcome these deficiencies, the model should encompass also a certain level of reconstructive information.

4 Our approach

In this section we describe the algorithm for incremental updating of LDA representations (Algorithm 1). It takes the training images sequentially and computes the new representation from the current representation and the new input image.

Let n be the number of training images observed so far. The idea is to represent these n images in a way that includes both reconstructive *and* discriminative properties. Reconstructive representation is based on PCA [22]. As most of the visual variability of the images is contained in the first k principal vectors (where $k \ll n$), only the k-dimensional principal subspace is retained. However although the first k coefficients contain most of the reconstructive information, there is no guarantee that most of the discriminative information is present in them as well. In order not to lose discriminative information, we propose to augment the truncated principal subspace with $c - 1$ additional basis vectors, which keep all information relevant for LDA.

Now, let us suppose that we have already built an augmented PCA subspace (APCA subspace) from the first n images. The current augmented reconstructive model therefore consists of basis vectors[2] $\widehat{\mathbf{U}}^{(n)} \in \mathbb{R}^{m \times (k+c-1)}$, mean vector $\mu^{(n)} \in \mathbb{R}^m$, and coefficient vectors $\widehat{\mathbf{A}}^{(n)} \in \mathbb{R}^{(k+c-1) \times n}$. In step $n + 1$ we can calculate a new APCA subspace from the *representations* (coefficient vectors) of the first n input

images and a new image as proposed in [22]. Since the dimension of the APCA subspace is small, this update is computationally very efficient. The procedure for one update of the current APCA subspace is outlined in the first eight steps of Algorithm 1.

Once we updated the current representations of the images observed so far, we can perform LDA on these updated low-dimensional coefficient vectors[3] aligned in $\mathbf{A} \in \mathbb{R}^{(k+c) \times (n+1)}$. LDA yields the discriminative representation in the form of LDA vectors aligned in $\mathbf{V} \in \mathbb{R}^{(k+c) \times (c-1)}$.

Until now, no reconstructive nor discriminative information has been lost, since all the information contained in the novel image has been incorporated into the model. However, as a consequence, the model has grown; the dimension of the APCA space has increased by one. To keep the size of the model, we have to truncate the obtained matrix $\mathbf{U} \in \mathbb{R}^{m \times (k+c)}$ (and consequently $\mathbf{A}$ and $\mathbf{V}$) by one. We propose to truncate $\mathbf{U}$ in a way, which preserves the discriminative information, similarly to [5].

Note that the classification using the combination of the reconstructive and discriminative representations is performed as a two step procedure: first a novel image is projected into the augmented PCA basis and the obtained coefficient vector is then projected onto the low-dimensional LDA vectors. The classification function is thus $g(\mathbf{x}) = \mathbf{V}^T \mathbf{U}^T (\mathbf{x} - \mu^{(n+1)})$. Now we will show how to truncate $\mathbf{U}$ and $\mathbf{V}$ by one dimension and still keep the classification function unchanged.

Let us first divide the matrices $\mathbf{U}$, $\mathbf{A}$, and $\mathbf{V}$ on submatrices containing the first k dimensions we want to keep and the last c dimensions we want to truncate by one[4] (line 10 in Algorithm 1). Then let us orthonormalize $\mathbf{V}_{\underline{c}}$ and update the APCA basis, the coefficients and the LDA vectors (lines 11 to 15 in Algorithm 1). We will show that the obtained updated representation, which is of the same size as at the beginning of the update step ($\widehat{\mathbf{U}}^{(n+1)} \in \mathbb{R}^{m \times (k+c-1)}$, $\widehat{\mathbf{A}}^{(n+1)} \in \mathbb{R}^{(k+c-1) \times (n+1)}$, and $\widehat{\mathbf{V}}^{(n+1)} \in \mathbb{R}^{(k+c-1) \times (c-1)}$) preserves the discriminative information. To verify this, let us rewrite the new classification function $\widehat{g}(\mathbf{x}) := \widehat{\mathbf{V}}^{(n+1)T} \widehat{\mathbf{U}}^{(n+1)T} (\mathbf{x} - \mu^{(n+1)})$ as

$$
\begin{aligned}
\widehat{g}(\mathbf{x}) &= (\widehat{\mathbf{U}}^{(n+1)} \widehat{\mathbf{V}}^{(n+1)})^T (\mathbf{x} - \mu^{(n+1)}) = \\
&= \left(\left[\mathbf{U}_{\underline{k}}, \mathbf{U}_{\underline{c}} \widetilde{\mathbf{V}}_{\underline{c}} \right] \left[\begin{array}{c} \mathbf{V}_{\underline{k}} \\ (\mathbf{V}_{\underline{c}}^T \mathbf{V}_{\underline{c}})^{1/2} \end{array} \right] \right)^T (\mathbf{x} - \mu^{(n+1)}) = \\
&= \left[\mathbf{U}_{\underline{k}} \mathbf{V}_{\underline{k}} + \mathbf{U}_{\underline{c}} \mathbf{V}_{\underline{c}} \right]^T (\mathbf{x} - \mu^{(n+1)}) = \\
&= \left(\left[\mathbf{U}_{\underline{k}}, \mathbf{U}_{\underline{c}} \right] \left[\begin{array}{c} \mathbf{V}_{\underline{k}} \\ \mathbf{V}_{\underline{c}} \end{array} \right] \right)^T (\mathbf{x} - \mu^{(n+1)}) = \\
&= (\mathbf{U}\mathbf{V})^T (\mathbf{x} - \mu^{(n+1)}) = \\
&= g(\mathbf{x}) \,.
\end{aligned} \tag{3}
$$

---

[3]Note that also in the standard LDA (fisher space) approaches, LDA is performed on the vectors of the PCA coefficients and not on the original images to avoid the singularity problems LDA encounters when dealing with high-dimensional data such as images. However, the complete vectors of principal coefficients are used in these cases.

[4]We denote the first k *columns* of $\mathbf{U}$, the first k *rows* of $\mathbf{A}$, and the first k *rows* of $\mathbf{V}$ with $\mathbf{U}_k$, $\mathbf{A}_k$, and $\mathbf{V}_k$, respectively, and the last c *columns* of $\mathbf{U}$ and *rows* of $\mathbf{A}$ and $\mathbf{V}$ with $\mathbf{U}_{\underline{c}}$, $\mathbf{A}_{\underline{c}}$, and $\mathbf{V}_{\underline{c}}$, respectively.

---

[2]A superscript denotes the step which the data is related to ($\widehat{\mathbf{U}}^{(n)}$ denotes the values of $\widehat{\mathbf{U}}$ at the step n).

It is therefore equivalent to the original classification function $g(\mathbf{x})$, thus all the discriminative information has been preserved.

Algorithm 1 : ILDA – incremental LDA

Require: current augmented principal subspace (mean vector $\mu^{(n)}$, APCA vectors $\widehat{\mathbf{U}}^{(n)}$, APCA coefficients $\widehat{\mathbf{A}}^{(n)}$) and new input image $\mathbf{x}^{(n+1)}$.

Ensure: new augmented principal subspace (mean vector $\mu^{(n+1)}$, APCA vectors $\widehat{\mathbf{U}}^{(n+1)}$, coefficients $\widehat{\mathbf{A}}^{(n+1)}$), new low-dimensional LDA vectors $\widehat{\mathbf{V}}^{(n+1)}$ and class centers $\boldsymbol{\nu}^{(n+1)}$.

1: Project a new image $\mathbf{x}^{(n+1)}$ into the current eigenspace: $\mathbf{a} = \widehat{\mathbf{U}}^{(n)\top}(\mathbf{x}^{(n+1)} - \mu^{(n)})$.

2: Reconstruct the new image: $\mathbf{y} = \widehat{\mathbf{U}}^{(n)}\mathbf{a} + \mu^{(n)}$.

3: Compute the residual vector: $\mathbf{r} = \mathbf{x}^{(n+1)} - \mathbf{y}$.

4: Append $\mathbf{r}$ as a new basis vector: $\mathbf{U}' = \left[\begin{array}{cc} \widehat{\mathbf{U}}^{(n)} & \frac{\mathbf{r}}{\|\mathbf{r}\|} \end{array}\right]$.

5: Determine the coefficients in the new basis: $\mathbf{A}' = \left[\begin{array}{cc} \widehat{\mathbf{A}}^{(n)} & \mathbf{a} \\ \mathbf{0} & \|\mathbf{r}\| \end{array}\right]$.

6: Perform PCA on $\mathbf{A}'$. Obtain the mean value μ'' and the eigenvectors $\mathbf{U}''$.

7: Project the coefficient vectors to the new basis: $\mathbf{A} = \mathbf{U}''^{\top}(\mathbf{A}' - \mu''\mathbf{1}_{1\times(n+1)})$.

8: Rotate the subspace $\mathbf{U}'$ for $\mathbf{U}''$: $\mathbf{U} = \mathbf{U}'\mathbf{U}''$.

9: Perform LDA on $\mathbf{A}$. Obtain low-dimensional LDA vectors $\mathbf{V}$ and class centers $\boldsymbol{\nu}$.

10: Divide $\mathbf{U}$, $\mathbf{A}$, and $\mathbf{V}$ on submatrices: $\mathbf{U} = \left[\begin{array}{cc} \mathbf{U}_{\underline{k}} & \mathbf{U}_{\underline{c}} \end{array}\right], \mathbf{A} = \left[\begin{array}{c} \mathbf{A}_{\underline{k}} \\ \mathbf{A}_{\underline{c}} \end{array}\right], \mathbf{V} = \left[\begin{array}{c} \mathbf{V}_{\underline{k}} \\ \mathbf{V}_{\underline{c}} \end{array}\right]$.

11: Orthonormalize $\mathbf{V}_{\underline{c}}$: $\widetilde{\mathbf{V}}_{\underline{c}} = \mathbf{V}_{\underline{c}}(\mathbf{V}_{\underline{c}}^{T}\mathbf{V}_{\underline{c}})^{-1/2}$.

12: Update the mean: $\mu^{(n+1)} = \mu^{(n)} + \mathbf{U}'\mu''$.

13: Update the APCA basis: $\widehat{\mathbf{U}}^{(n+1)} = \left[\begin{array}{cc} \mathbf{U}_{\underline{k}} & \mathbf{U}_{\underline{c}}\widetilde{\mathbf{V}}_{\underline{c}} \end{array}\right]$.

14: Update the coefficients: $\widehat{\mathbf{A}}^{(n+1)} = \left[\begin{array}{c} \mathbf{A}_{\underline{k}} \\ \widetilde{\mathbf{V}}_{\underline{c}}^{T}\mathbf{A}_{\underline{c}} \end{array}\right]$.

15: New LDA vectors: $\widehat{\mathbf{V}}^{(n+1)} = \left[\begin{array}{c} \mathbf{V}_{\underline{k}} \\ (\mathbf{V}_{\underline{c}}^{T}\mathbf{V}_{\underline{c}})^{1/2} \end{array}\right]$.

16: New class centers: $\boldsymbol{\nu}^{(n+1)} = \boldsymbol{\nu}$.

Using Algorithm 1 we can thus update the current reconstructive and discriminative representations without losing a valuable discriminative information and without enlarging the reconstructive basis. Since at each step LDA is recalculated using low-dimensional APCA representations of the training images, the update step is fast, while still enabling various types of updating.

The model can be updated with an image of a known class making the representation of this class more reliable. Additionally an image of a novel class can be introduced. In this case, a new class is initialized, c is incremented by one, and consequently all the matrices keeping the representations are enlarged by one dimension. The performance of the proposed method for these approaches is evaluated in the next section.

5 Experimental results

In the experiments we focus on comparison of the proposed method, which combines both reconstructive and discriminative information, with the methods that exploit only one type of information. We also always show the performance of the standard batch LDA method (denoted as *batchLDA*) giving the best results because it processes all training images simultaneously, therefore it can find the hyperplane which is optimally suitable for the given data. Thus our aim is to achieve similar results with the incremental training.

The idea of incremental learning is to start with a given model (denoted as starting model) and update it when new information is available. In the following we compare three different approaches differing in the usage of discriminative and reconstructive information.

ILDAonK is the incremental LDA based on a truncated PCA basis keeping only k PCA-eigenvectors. It thus predominantly contains the reconstructive information, while some important discriminative information may be discarded. On the other hand, *ILDAonL* does not keep any additional reconstructive information. The training images are represented only by $(c-1)$-dimensional LDA coefficient vectors, which are propagated in the updating steps. Finally, *ILDAaPCA*, the proposed method, combines both types of representations keeping the reconstructive as well as the discriminative information.

In all experiments the dimension k of the truncated PCA space is fixed such that the starting model contains 80% of the energy (a fraction of the total variance). Since in the case of *ILDAaPCA* method the truncated principal subspace is augmented by $(c-1)$ basis vectors, we do not truncate the principal subspace in *ILDAonK* approach at k but rather at $k+(c-1)$ to enable a fair comparison. In this way both approaches produce representations of the same size.

In the following we will show that *ILDAaPCA* is actually capable of facing two challenges. It is possible to add images of already known categories and to add new categories.

We will test the above described methods on the pre-cropped Sheffield Face Database [7]. It consists of 20 persons with at least 19 images of each individual and the images cover poses from profile to frontal views. We took 9 images (every second one) of each person for training (e.g., see Figure 1) and 10 images for testing.

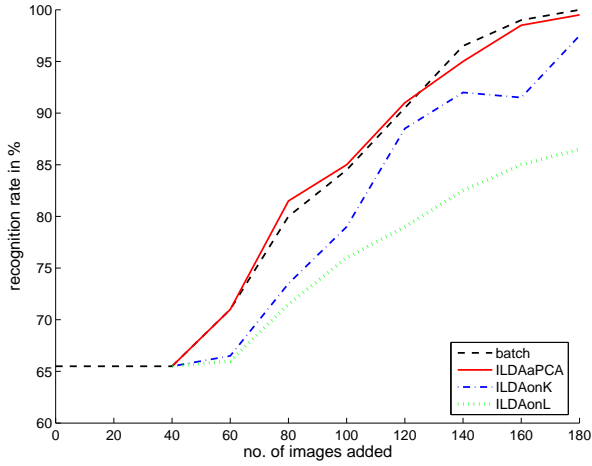**Figure 1:** Training images for one person in the Sheffield Face Database.

Figure 2: Comparison of the recognition rate on Sheffield Face Database of *batchLDA*, *ILDAaPCA*, *ILDAonK* and *ILDAonK*

Figure 3: Recognition rate for test images of already trained classes on Sheffield Face Database

Adding new instances: To build the starting model for the first task we took two images of each class, having 40 images at the beginning and added 140 images, 7 of each class, in the sequential updating steps.

For *ILDAaPCA*, *ILDAonK* and *ILDAonK* the eigenspace was updated after each image was presented, while for *batchLDA* the LDA space was always built from scratch using the current number of training images. In both cases the recognition rate was calculated after adding 20 images, one of each class, and repeated 7 times.

As can be seen in Figure 2, the recognition rate keeps growing with increasing number of training images. This demonstrates that new images bring additional knowledge in the model and improve the current representations resulting in a better performance of the classifier. It is also evident that *ILDAaPCA* clearly outperforms *ILDAonK* and *ILDAonK* being nearly as good as *batchLDA*. As one could expect, *ILDAonK* yields the worst results, since the very low-dimensional discriminative representations do not suffice for updating the model. We can also conclude that *ILDAonK* approach discards some discriminative information and produce results inferior to *ILDAaPCA* method, which preserves this discriminative information.

Adding new classes: Here we started with a basis created from all training images of two subjects (18 images altogether) and then added new faces one by one. The model was updated with all the training images of the new class before adding the next one. We classified only those test images for which the model of the corresponding class was already built.

The results are displayed in Fig. 3. As expected, the recognition rate drops a little bit in all approaches by increasing the number of classes, since it is more difficult to discriminate between 20 classes then between only a few of them. However, the results again clearly demonstrate that the proposed *ILDAaPCA* method outperforms *ILDAonK* and *ILDAonK* approaches.
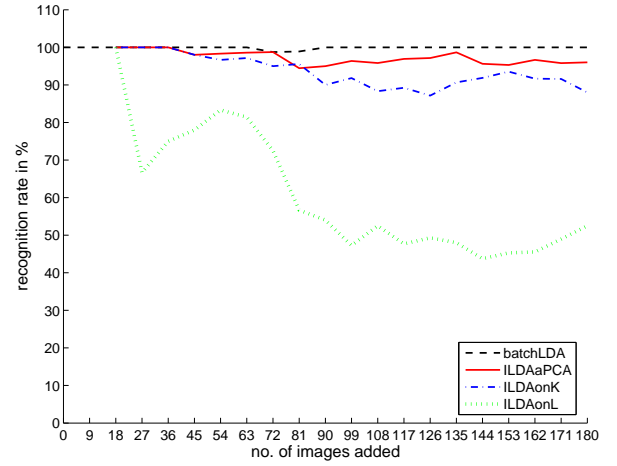
To demonstrate that our method works for different tasks too, we present the results of an object recognition task. The experiment was performed on the Columbia image database COIL20 [19]. It consists of 20 objects with 72 gray scale images of views from 0 to 360 degrees in 5 degree steps. For our tests we took 14 images for training and the remaining 58 for testing.

We started again with a basis created from two images of each class, and then added the remaining images in $12 * 20$ update steps. As can be seen in Figure 4 the behavior of the curves is similar to the experiment on faces, again showing that *ILDAaPCA* achieves the highest recognition rate.
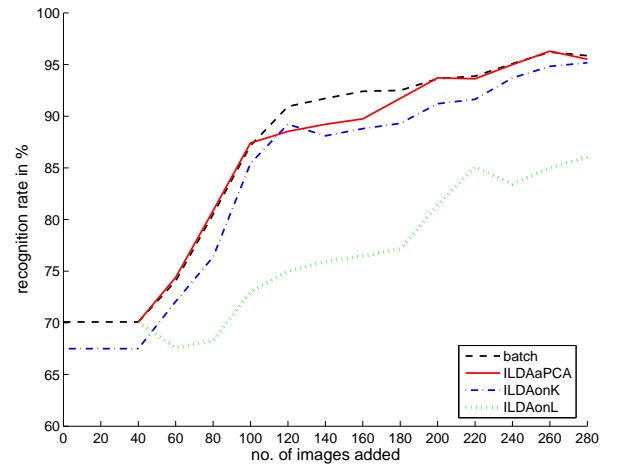

Figure 4: Comparison of the recognition rate on COIL20 database

6 Conclusion

In this paper we proposed a method that combines reconstructive models and discriminative classifiers to enable updating of the already learned representations. To achieve that, we enrich the discriminative LDA representations with reconstructive information. This is realized by embedding the LDA learning and classification into an augmented

5

PCA subspace enabling incremental updating of the already learned representations without discarding significant discriminative information.

The augmented PCA subspace thus contains sufficient reconstructive information, which enables incremental learning, and the previously extracted discriminative information, which enables efficient classification. In this way we are able to efficiently update the current model with new instances of the already learned classes and to introduce new classes. In addition, this method could in principle also enable updating the current model to new tasks. Moreover, the reconstructive representation would also enable detection of outliers [5, 22], thus the proposed method could be further extended in a robust approach for incremental learning of LDA representations. This technique could also be applied to other reconstructive and discriminative linear subspace methods (Independent Component Analysis, Non-negative Matrix Factorization; Canonical Correlation Analysis, Support Vector Machines). The combination of reconstructive and discriminative methods thus offers a promise to achieve best of both worlds; to enable successful discrimination using efficient task-dependent discriminative representations, while at the same time enabling robustness, and adaptation to new images using the reconstructive representations.

Acknowledgement

References

[1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenspaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE PAMI*, 19(7):711–720, 1997.

[2] M. Borga and H. Knutsson. Canonical correlation analysis in early vision processing. In *Proc. of the 9th European Symposium on Artificial Neural Networks*, 2001.

[3] S. Chandrasekaran, B. S. Manjunath, Y. F. Wang, J. Winkeler, and H. Zhang. An eigenspace update algorithm for image analysis. *Graphical Models and Image Processing*, 59(5):321–332, 1997.

[4] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. Wiley, 2001.

[5] S. Fidler, D. Skočaj, and A. Leonardis. Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. *IEEE PAMI*, 28(3):337-350, 2006.

[6] M. Fritz, B. Leibe, B. Caputo, and B. Schiele. Integrating representative and discriminative models for object category detection. In *ICCV05*, II:1363–1370, 2005.

[7] D. B. Graham and N. M. Allinson. In *Face recognition: From theory to applications*, *NATO ASI Series F, Computer and Systems Sciences*, 163:446 – 456. 1998.

[8] P. Hall, D. Marshall, and R. Martin. Merging and splitting eigenspace models. *IEEE PAMI*, 22(9):1042–1048, 2000.

[9] K. Hiraoka, S. Yoshizawa, K. Hidai, M. Hamahira, H. Mizoguchi and T. Mishima. Convergence analysis of online linear discriminant analysis. In *Proc. of the IEEE–INNS–ENNS International Joint Conference on Neural Networks (IJCNN 2000)*, 3:387 – 391, 2000.

[10] A. Holub, M. Welling, and P. Perona. Combining generative models and fisher kernels for object recognition. In *ICCV05*, I:136–143, 2005.

[11] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.

[12] A. Levy and M. Lindenbaum. Sequential karhunen-loeve basis extraction and its application to images. *IEEE Trans. on Image Processing*, 9(8):1371–1374, 2000.

[13] Y. Li. On incremental and robust subspace learning. *Pattern recognition*, 37(7):1509–1518, 2004.

[14] Y. Li, L. Shapiro, and J. Bilmes. A generative/discriminative learning algorithm for image classification. In *ICCV05*, II:1605–1612, 2005.

[15] R. Lin, M. Yang and S. E. Levinson. Object tracking using incremental fisher discriminant analysis. In *ICPR04*, 2:757 – 760, 2004.

[16] X. Lu, Y. Wang, and A.K. Jain. Combining classifiers for face recognition. In *ICME 2003*, III:13 – 16, 2003.

[17] G. L. Marcialis and F. Roli. Fusion of PCA and LDA for face verification. In *Proc. of Post-ECCV Workshop on Biometric Authentication (BIOMET)*, pp. 30 – 37, 2002.

[18] A. M. Martinez and A. C. Kak. PCA versus LDA. *IEEE PAMI*, pages 23(2):228 – 233, 2001.

[19] S. A. Nene, S. K. Nayar and H. Murase. Columbia object image library (COIL-20). Technical Report CUCS-005-96, Columbia University, 1996.

[20] S. Pang, S. Ozawa and N. Kasabov. Chunk incremental LDA computing on data streams. In *Advances in Neural Networks ISNN 2005*, 3497:51 – 56, 2005.

[21] S. Pang, S. Ozawa and N. Kasabov. Incremental linear discriminant analysis for classification of data streams. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 35(5):905 – 914, 2005.

[22] D. Skočaj and A. Leonardis. Weighted and robust incremental method for subspace learning. *ICCV03*, II:1494–1501, 2003.

[23] J. Yang and J.-Y. Yang. Why can LDA be performed in PCA transformed space?. *Pattern recognition*, 36:685–690, 2003.

[24] J. Ye, Q. Li, H. Xiong, H. Park, R. Janardan and V. Kumar. IDR/QR: An incremental dimension reduction algorithm via QR decomposition. In *Proc. of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004)*, pp. 364 – 373, 2004.

[25] W. Zhao, A. Krishnaswamy, R. Chellappa, D. Swets, and J. Weng. Discriminant analysis of principal components for face recognition. *Face Recognition: From Theory to Applications*, pp. 73 – 85, 1998.