Online Learning for Human-Robot Interaction

Bogdan Raducanu
Computer Vision Center
Edifici "O" - Campus UAB
08193 Bellaterra (Barcelona), Spain
bogdan@cvc.uab.es

Jordi Vitrià
Computer Vision Center
CS Dept., Autonomous Univ. of Barcelona
08193 Bellaterra (Barcelona), Spain
jordi@cvc.uab.es

Abstract

This paper presents a novel approach for incremental subspace learning based on an online version of the Non-parametric Discriminant Analysis (NDA). For many real-world applications (like the study of visual processes, for instance) there is impossible to know beforehand the number of total classes or the exact number of instances per class. This motivated us to propose a new algorithm, in which new samples can be added asynchronously, at different time stamps, as soon as they become available. The proposed technique for NDA-eigenspace representation has been applied to the problem of online face recognition for human-robot interaction scenario.

1. Introduction

The human visual system is very robust among a large range of variations in environmental conditions. Opposite to it, a similar performance is impossible to be achieved by any artificial vision system. 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 goal. One of the factors that impede to achieve these performances is the learning strategy that is 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. It is unlikely that such a manual operation could meet the demands of many challenging tasks that are critical for generating intelligent behavior, such as object recognition, in general, and face recognition, in particular. Furthermore, it is assumed that all the available data is all that a machine need to 'learn'. Paradoxically, by using this 'off-line' or 'batch' approach the development of such a system is denied by the creator of the system itself.

The purpose of the present study is to show how image feature extraction can be addressed in terms of a developmental learning strategy for artificial vision systems. Feature extraction and selection is a common pre-processing step in any pattern classification problem. Eigenspace-based methods are the most common techniques used to obtain a dimensionality reduction of the original data. The point coordinates in this eigenspace represent the coefficients of the projected input images. 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) [12]. For the latter, parametric and non-parametric forms have been proposed [6]. The shortcomings of parametric discriminant analysis (PDA) are twofold. On the one hand, it assumes that the samples present a normal distribution. On the other hand, it fails to capture the boundary class information, creating a generative model of data. Because of these limitations, methods based on parametric discriminant analysis show a serious performance degeneration in real-world applications when data present multi-modal densities and classes are not linearly separable.

Opposed to this case, non-parametric discriminant analysis (NDA) [7] is more effective when dealing with general data distributions and capturing properly the structural information between class boundaries. Despite its undeniable advantages, the NDA has received little attention within pattern recognition community. In [2], the authors introduced a non-parametric form of the within-class scatter matrix. This way, the matrix is normalized: instead of assuming a gaussian distribution on the points of the same class, it normalizes the distances between each point and their nearest neighbors, which has been shown to benefit the nearest neighbor rule. In [13], simple (1D) projections are combined such that together can provide more separability on the whole data set. From the Adaboost perspective, each simple 1D projection is treated as a "weak classifier", so the Adaboost algorithm is used to select among a large set of 1D projections the ones that best separate the training data. The nonparametric nature of their approach is guaranteed

by the fact that doesn't require any global statistic measure of the input data. In [11], two kinds of nonparametric subspace analysis, which complement each other, are proposed. First of them is the principal nonparametric subspace analysis and is used to extract nonparametric discriminating features within the principal subspace of within-class scatter matrix. The second one is called null-space nonparametric subspace analysis and is based on the null space of the within-class scatter matrix.

Typical implementations for the above mentioned techniques assume that all the data is provided in advance and learning is carried out in one step (for this reason, we will refer as these as batch techniques). However, in real-world applications, this is not the case, since it is unlikely that all the data is available from the very beginning. For this situation, a new learning strategy is required. One-pass incremental algorithms, which performs sequentially the update of the eigenspace representation, are the solution we are seeking. So far, several approaches have been proposed. In [4, 8, 1] the Incremental Principal Component Analysis (IPCA) is introduced. 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 [17] 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 [15] and [14].

In this paper we propose an online version for NDA technique (referred for the rest of the paper as IncNDA). More concrete, we introduce a sequential update of the NDA-eigenspace representation. The proposed solution for online learning 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 briefly review the classical NDA algorithm (from now on referred as BatchNDA). 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 for human-robot interaction scenario. We will show that at the end of the learning process, the recognition performance achieved converges towards the result obtained using the BatchNDA. Finally, section 5 contains our conclusions and the guidelines for future work.

2. Classical Nonparametric Discriminant Analysis (BatchNDA)

As introduced in [7], 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 ζ:

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

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

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 [7], 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 α is a control parameter that can be selected between zero and infinity and $d(u, v)$ is the distance between two vectors u and v. The sample weights take values close to 0.5 on class boundaries and drop to zero as we move away. The parameter α adjusts how fast this happens.

3. Incremental Nonparametric Discriminant Analysis (IncNDA)

The shortcoming of the BatchNDA described in the previous section comes from the assumption 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:

- y belongs to one of the existing classes C_L (y^{C_L}, where $1 < L < N$).

In this case, the equation that updates S_b is given by:

$$S'_b = S_b - S_b^{in}(C_L) + S_b^{in}(C_{L\prime}) + S_b^{out}(y^{C_L}) \quad (7)$$

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

$$S_b^{in}(C_L) = \sum_{i=1, i\neq L}^{C_N} \sum_{t=1}^{n_{C_i}} W(C_i, C_L, t)(x_t^i - \mu_{C_L}(x_t^i))(x_t^i - \mu_{C_L}(x_t^i))^T \quad (8)$$

$$S_b^{out}(y^{C_L}) = \sum_{i=1, i\neq L}^{C_N} (y^{C_L} - \mu_{C_i}(y^{C_L}))(y^{C_L} - \mu_{C_i}(y^{C_L}))^T \quad (9)$$

In the case of S'_w the update equation is the following:

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

where

$$S_w(C_{L\prime}) = S_w(C_L) + \frac{n_{C_L}}{n_{C_L}+1}(y - \bar{x}^{C_L})(y - \bar{x}^{C_L})^T \quad (11)$$

- y belongs to a new class C_{N+1} ($y^{C_{N+1}}$).

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

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

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

$$S_b^{in}(C_{N+1}) = \sum_{i=1}^{C_N} \sum_{t=1}^{n_{C_i}} W(C_i, C_{N+1}, t) (x_t^i - \mu_{C_{N+1}}(x_t^i))(x_t^i - \mu_{C_{N+1}}(x_t^i))^T \quad (13)$$

$$S_b^{out}(C_{N+1}) = \sum_{i=1}^{C_N} (y^{C_{N+1}} - \mu_{C_i}(y^{C_{N+1}}))(y^{C_{N+1}} - \mu_{C_i}(y^{C_{N+1}}))^T \quad (14)$$

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

$$S'_w = S_w \quad (15)$$

4. Face Recognition for Human-Robot Interaction: A Case Study

Detecting and responding to human presence is an issue of great interest for the area of human-robot interaction. The current trend in robotics is represented by social-oriented robots, i.e. robots which are enabled with perceptual capabilities in order to make communication with humans more natural. This can include robots responsive to hand or head gestures, head pose orientation, voice recognition, etc. [3, 10, 5, 9]. Our goal is to build an application whose aim is to have an AIBO behaving in a personalized manner, depending on the frequency it sees a certain person. Thus, we expect the robot to develop a 'friendlier' attitude towards persons who are frequently seen, meanwhile to act more 'reserved' in front of a person who has been seen less frequent. In order to achieve this goal, 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 performed in an automatic way. For this purpose, we put an AIBO robot in an open space and snapshots were taken each time a person was passing in front of the camera. In figure 1 we extracted some frames from the face acquisition process. The face was automatically extracted from the image using the face detector based on [16]. We didn't impose any restrictions regarding ambient conditions.

Overall, our database consists of 6882 images of 51 people (both male and female)[1]. 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 of the particularity of the acquisition process, face images reflect the changes

[1] 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

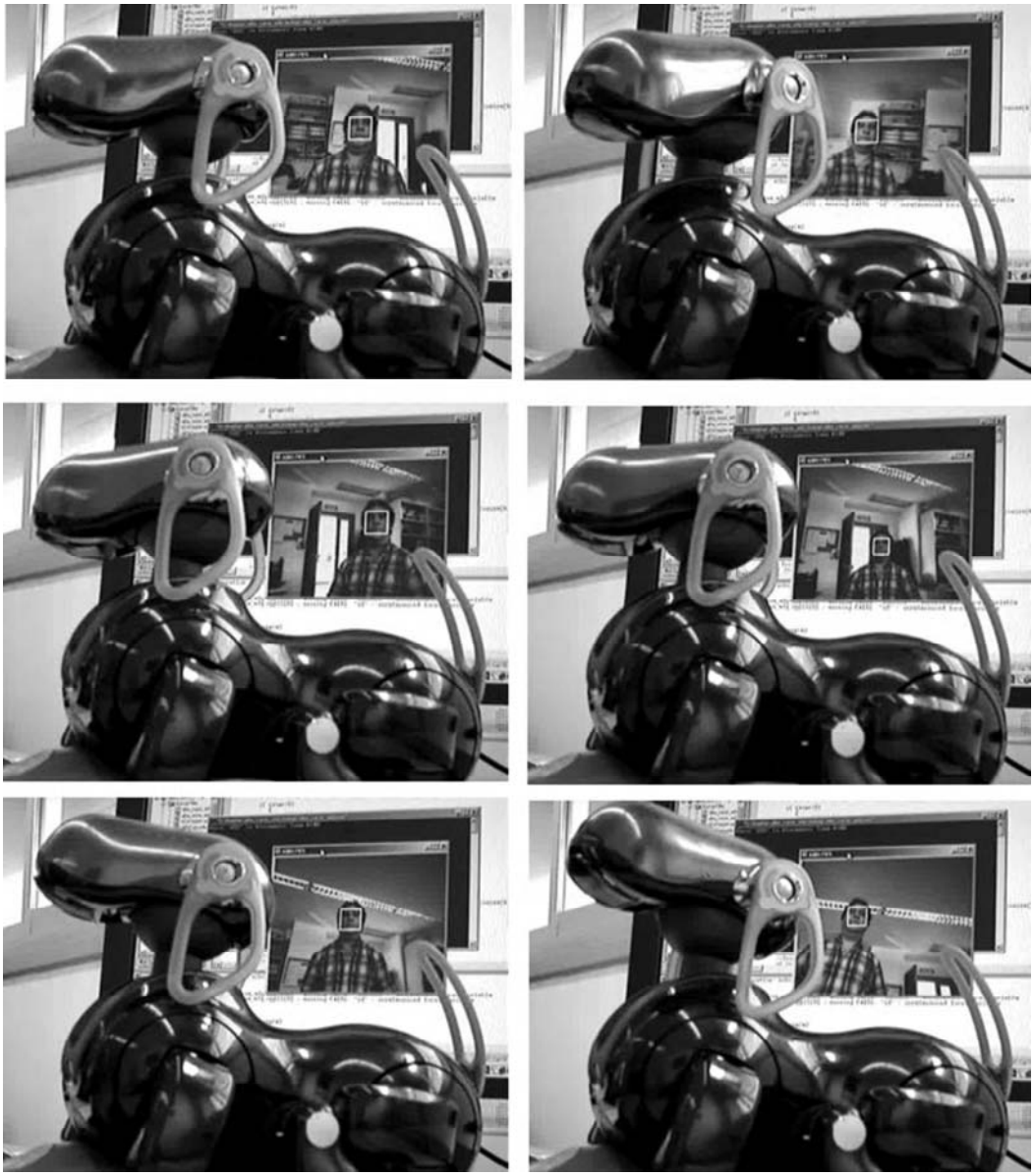

Figure 1. Real-time face detection and tracking by an AIBO robot.

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.

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. This way, data dimensionality was downsized from 2304 to 60. The remaining samples (5100) from the training set were added later on in a sequential manner (the samples were drawn randomly) and this way the NDA-eigenspace was updated[2]. The classification accuracy was evaluated based on the nearest-neighbor rule.

In figure 3 (above) we depicted the evolution of the learn-

[2]We considered our learning strategy a supervised one, so the class label of the new added sample from the training set is known

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

ing 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 (below), we depicted the percentage of incremental training samples introduced so far (the stars represent the moment when a new class has been added). The graph falls a couple of times, because the percentage of total data is actually computed relative to the number of classes presented up to a given moment, not with the total number of classes. For this reason, 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

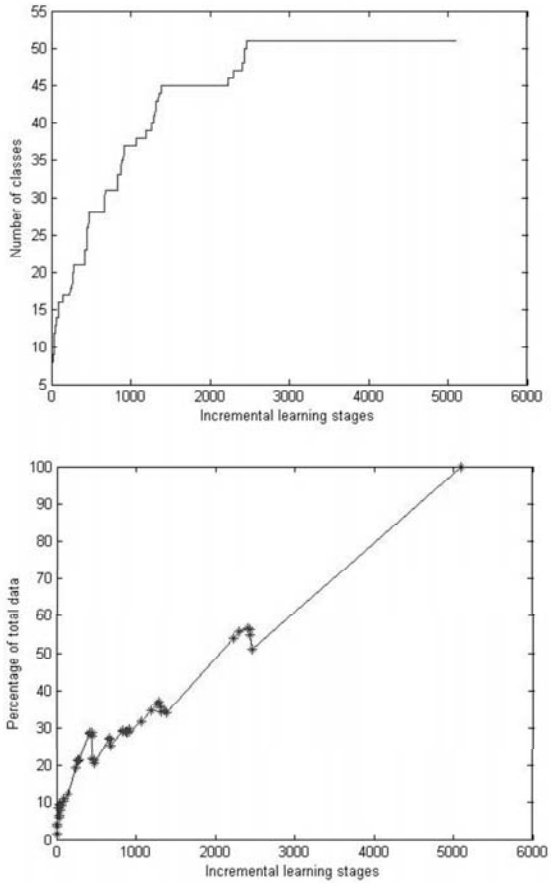

Figure 3. Learning process: evolution of the number of classes function of learning stages (above) and the percentage of the training data function of learning stages (below)

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

Some instances of misclassified faces are represented in figure 5. From the experiments performed, we arrive at the following conclusion. The misclassification occurs in three situations: when there are too few face instances per class, when there are too few instances of a particular head pose/illumination conditions and when the image presents a high level of distortion (the 'blurring' effect due to person movement).

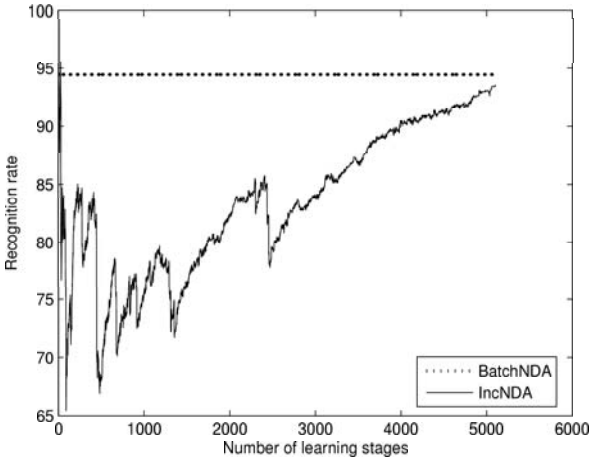

Figure 4. IncNDA vs. BatchNDA curves. IncNDA converges towards BatchNDA at the end of the learning process

Figure 5. Some instances of misclassified faces

5. Conclusions and Future Work

For real-world applications, one-step (batch mode) learning techniques prove to be inadequate. For this reason, we proposed in this paper an online version of the Nonparametric Discriminant Analysis. We start to build the NDA-eigenspace representation in an incremental way, by adding sequentially new data. This new approach has been applied to the problem of face recognition for human-robot interaction scenario. The tests performed on a custom face database confirm the robustness of IncNDA and the fact that, at the end of the learning process, it converges towards BatchNDA. In the future, we plan to extend the current approach, by allowing the update of the NDA-eigenspace in terms of data chunk.

Acknowledgements

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

References

[1] M. Artač, M. Jogan, and A. Leonardis. Incremental pca for on-line visual learning and recognition. In *Proceedings of 16th International Conference on Pattern Recognition*, pages 781–784, 2002.

[2] M. Bressan and J. Vitrià. Nonparametric discriminant analysis and nearest neighbor classification. *Pattern Recognition Letters*, 24(15):2743–2749, 2003.

[3] L. Brèthes, F. Lerasle, and P. Danès. Data fusion for visual tracking dedicated to human-robot interaction. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 2087–2092, 2005.

[4] S. Chandrasekaran, B. Manjumath, Y. Wang, J. Winkler, and H. Zhang. An eingenspace update algorithm for image analysis. *Graphical Models Image Processing*, 59(5):321–332, 1997.

[5] J. R. del Solar, R. Verschae, P. Vallejos, and M. Correa. Face analysis for human computer interaction applications. In *Proceedings of the 2nd VISAPP International Conference on Computer Vision Theory and Applications*, pages 23–30, 2007.

[6] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Jown Wiley and Sons, New York, USA, 2001.

[7] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, Boston, USA, 1990.

[8] P. Hall, D. Marshall, and R. Martin. Incremental eingenanalysis for classification. In *Proceedings of British Machine Vision Conference*, pages 286–295, 1998.

[9] J.-H. Hong, Y.-S. Song, and S.-B. Cho. A hierarchical bayesian network for mixed-initiative human-robot interaction. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 3819–3824, 2005.

[10] T. Kee, S.-K. Park, and M. Park. A new facial features and face detection method for human-robot interaction. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 2075–2080, 2005.

[11] Z. Li, W. Liu, D. Lin, and X. Tang. Nonparametric subspace analysis for face recognition. In *Proceedings of 2005 IEEE Conference on Computer Vision and Pattern Recognition*, pages 961–966, 2005.

[12] A. Martinez and A. Kak. Pca versus lda. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2001.

[13] D. Masip and J. Vitrià. *Boosted Linear Projections for Discriminant Analysis*, pages 45–52. IOS Press, Amsterdam, The Netherlands, 2005.

[14] S. Pang, S. Ozawa, and N. Kasabov. *Chunk Incremental LDA Computing on Data Streams*, volume LNCS 3497, pages 51–56. Springer-Verlag, New York, 2005.

[15] D. Skočaj, M. Uray, A. Leonardis, and H. Bischof. Why to combine reconstructive and discriminative information for incremental subspace learning. In *Proceedings of Computer Vision Winter Workshop*, page N/A, 2006.

[16] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57:137–154, 2004.

[17] J. Weng, Y. Zhang, and W.-S. Hwang. Candid covariance-free incremental principal component analysis. *IEEE Transactions on Pattern Recognition*, 25(8):1034–1040, 2003.