Learning Kernel Subspace Classifier

Bailing Zhang[1], Hanseok Ko[2], and Yongsheng Gao[3]

[1] School of Computer Science and Mathematics
Victoria University, VIC 3011, Australia
bailing.zhang@vu.edu.au
[2] School of Electronics and Computer Engineering
Korea University, Seoul, 136-713, Korea
hsko@korea.ac.kr
[3] School of Engineering
Griffith University, QLD 4111, Australia
yongsheng.gao@griffith.edu.au

Abstract. Subspace classifiers are well-known in pattern recognition, which represent pattern classes by linear subspaces spanned by the class specific basis vectors through simple mathematical operations like SVD. Recently, kernel based subspace methods have been proposed to extend the functionalities by directly applying the Kernel Principal Component Analysis (KPCA). The projection variance in kernel space as applied in these earlier proposed kernel subspace methods, however, is not a trustworthy criteria for class discrimination and they simply fail in many recognition problems as we encountered in biometrics research. We address this issue by proposing a learning kernel subspace classifier which attempts to reconstruct data in input space through the kernel subspace projection. While the pre-image methods aiming at finding an approximate pre-image for each input by minimization of the reconstruction error in kernel space, we emphasize the problem of how to estimate a kernel subspace as a model for a specific class. Using the occluded face recognition as examples, our experimental results demonstrated the efficiency of the proposed method.

1 Introduction

Subspace classifier is a traditional pattern recognition method that has been broadly applied in signal processing and computer vision for performing various types of recognition. The essence of subspace classifier is classifying an unlabeld pattern based on its distance from different subspaces that represent known classes. One of the first subspace classification algorithms was CLASFIC (class feature information compression)[1], in which the basis vectors spanning the subspaces of each class are determined by a principal component analysis (PCA) of the patterns belonging to that class. The principal assumption behind subspace classifiers is that the vector distribution in each class lies in a lower-dimensional subspace of the original feature space and it is optimal from a reconstruction point of view. And the earlier subspace classifier has been extended in many

S.-W. Lee and S.Z. Li (Eds.): ICB 2007, LNCS 4642, pp. 299–308, 2007.

ways. For example, it was found a better performance could be gained if the subspaces are modified in an error-driven way, which was termed as Learning Subspace Method (LSM)[2].

Subspace classifiers are well suited for the classification problems with high dimensional input space. The best reconstruction property accompanized with the linear transformation offers an efficient way of handling missing pixels and occlusions that frequently appear in practices for many image recognition problems. The linear subspace methods, however, are limited in performance if non-linear features are involved. Furthermore, PCA encodes the data based on second order statistics and ignores higher-order dependencies, which may contain important discriminant information for recognition. As a solution, kernel representations can be introduced by projecting the input attributes into a high dimensional feature space, through which the complex nonlinear problems in the original space will more likely be formulated as near-linear ones [6,3,9].

In the past, several works have been reported on combining kernel method with subspace classifiers [4-5]. These earlier works, however, shared the same idea of applying the kernel principal component analysis (KPCA) and establishing the classifier based on the projection variance in kernel space. More specifically, the kernel trick is used to map each class of input data into their respective implicit feature space F and then PCA is performed in F to produce the nonlinear subspace of the corresponding class. A test data is then projected to all of the nonlinear subspaces and the projection variance in kernel space is used as classification criteria. Our extensive experiment on applying this idea to some face recognition problems, however, turned out that such a simple "kernel subspace" classifier does not work.

In addition to [4-5], the application of KPCA in pattern classification has also been extensively discussed in recent years. For example, KPCA de-noising and the pre-image problem [7-8] is very close to the kernel subspace classifier. For a given pattern, the pre-image algorithms attempt to find the reconstructed one in input space by minimizing the reconstruction error in kernel space F with gradient descent. Pre-image algorithms, however, do not measure the discrepancy between an input vector and its reconstruction in input space. This is more important in kernel subspace classifier as the reconstruction error in input space directly indicates the discrimination capability of a data class with its corresponding kernel subspace. We attempted to solve the problem by formulating the objective as best reconstructing input data from the kernel principal component projections and the new method is termed as *learning kernel subspace classifier*. On robust face recognition problems, our experiments showed its superiority over the pre-image algorithms and some complex occlusion robust face recognition methods.

2 Subspace Classifier

Assume that each of the data classes forms a lower-dimensional linear subspace distinct from the subspaces spanned by other data classes [1-2], then the subspace

representing a class can be defined in terms of basis vectors spanning the subspace. And a testing data item is classified based on the lengths of its projections onto each of the subspaces or, alternatively, on the distances of the test vector from these subspaces.

Let $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_N)$ be the training data matrix belongs to a class $\omega^{(c)}, c = 1, \cdots, C$, where $\mathbf{x}_i$ is a training data vector. A set of orthonormal vectors $\mathbf{p}_i$ can be obtained by, for example, the principal component analysis of the correlation matrix $\mathbf{X}^T\mathbf{X}$, *i.e.*, $\mathbf{p}_i^T\mathbf{p}_j = \delta_{ij}$. The basis vectors $\mathbf{p}_i \in R^n, i = 1, \cdots, d$ $(d < n)$, span a subspace for the class, which can be expressed as $L : L = L(\mathbf{p}_1, \cdots, \mathbf{p}_d)$. Denote P the matrix whose column vectors are $\mathbf{p}_i$, $P = [\mathbf{p}_1, \cdots, \mathbf{p}_d]$. When an unlabeled sample $\mathbf{x}$ is classified by the subspace classifier, the distance between $\mathbf{x}$ and each of the subspace is calculated by the projection $y = P^T\mathbf{x}$,. Then, $\mathbf{x}$ is classified to the class with the smallest distance. The distance between $\mathbf{x}$ and L is described as

$$d = ||\mathbf{x}||^2 - ||P^T\mathbf{x}||^2. \tag{1}$$

Since the first term is independent from the class, the discriminant function indicating the membership of $\mathbf{x}$ belonging to $\omega^{(c)}$, can be written as

$$f_c(\mathbf{x}) = ||P_c^T\mathbf{x}||^2, \qquad c = 1, \cdots, C \tag{2}$$

In training stage, the sum of the squared distances between the training samples and the subspace is minimized, *i.e.*, the reconstruction of $\mathbf{x}$

$$\hat{\mathbf{x}} = \sum_{i=1}^{m} y_i\mathbf{p}_i = \sum_{i=1}^{m}(\mathbf{x}^T\mathbf{p}_i)\mathbf{p}_i \tag{3}$$

will be minimized. This is also equivalent to the maximization of the projection variance in Eqn. (2) and the standard solution is the principal component analysis on the correlation matrix.

3 Kernel PCA and Kernel-Based Subspace Classifier

Subspace classifiers combined with kernel methods have been proposed in [4-5], which are all based on the direct application of Kernel Principal Component Analysis (KPCA) in feature space. KPCA maps all data samples to a higher-dimensional feature space via the so-called kernel trick and then finds the subspace in this transformed space through the PCA for each class separately.

Suppose a high dimensional feature space, F, is related to the input space by the (nonlinear) map $\mathbf{\Phi}(\mathbf{x}) : R^n \to F$. The map $\mathbf{\Phi}$ and the space F are determined implicitly by the choice of a kernel function k, which computes the dot product between two input examples $\mathbf{x}$ and $\mathbf{y}$ mapped into F via

$$k(\mathbf{x}, \mathbf{y}) = \mathbf{\Phi}(\mathbf{x}) \cdot \mathbf{\Phi}(\mathbf{y}) \tag{4}$$

where $(\cdot)$ is the vector dot product in F. The most commonly used kernel is:

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{||\mathbf{x} - \mathbf{y}||}{2\sigma^2}\right) \tag{5}$$

where σ is the width of the kernel. And the space F is called a reproducing kernel Hilbert space (RKHS) generated by k [6]. The input space is then mapped to F in the way that a sample $\mathbf{v}$ is transformed to the kernel function centered on $\mathbf{v}$:

$$\mathbf{v} \rightarrow \mathrm{k}(\mathbf{x}, \mathbf{v})$$

For a set of N patterns $\mathbf{x}_\mathrm{i}$, $i = 1, 2, ...N$ in R^n, the $N \times N$ kernel matrix K can be formed:

$$K_{ij} = k(\mathbf{x}_\mathrm{i}, \mathbf{x}_\mathrm{j}) \tag{6}$$

The kernel matrix K should then be centralized with the result as the estimate of the covariance matrix of the new feature vector in F. Then the linear PCA is simply performed on it by finding a set of principal components in the span of vectors $\{\Phi(\mathbf{x})_\mathrm{r}\}$, which represents the principal axes in the kernel space.

Let $\alpha^k = [\alpha_1^k, \cdots, \alpha_N^k]^T$ be the normalized eigenvectors and $\lambda_1 \leq \cdots \leq \lambda_N$ be the eigenvalues of the matrix K such that $\lambda_k(\alpha^\mathrm{k}, \alpha^\mathrm{k}) = 1$ for all $k = 1, \ldots, N$ where $\lambda_p > 0$. It can be shown that the eigenvectors in F can be expressed as linear combinations of the mapped training samples:

$$\mathbf{v}_\mathrm{k} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_\mathrm{i}^\mathrm{k} \Phi(\mathrm{x}_\mathrm{i}) \tag{7}$$

with known coefficients α_i^k. For a test data point $\mathbf{x}$ with image $\Phi(\mathbf{x})$ in kernel space, the projection of a mapped point $\Phi(\mathbf{x})$ on the eigenvector $\mathbf{v}_\mathrm{k}$ is therefore given by:

$$\beta_k = (\mathbf{v}_\mathrm{k}, \Phi(\mathbf{x})) = \sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_\mathrm{i}^\mathrm{k} \mathrm{k}(\mathbf{x}_\mathrm{i}, \mathbf{x}) \tag{8}$$

In RKHS, the conventional subspace classifier can be simply performed by replacing the inner product in Eqn (2) by the one from RKHS [6]. The discriminant function in RKHS can then be described as follows:

$$f_c(\mathbf{x}) = ||\sum_{\mathrm{k}=1}^{\mathrm{r}} \beta_\mathrm{k}||^2 = \sum_{\mathrm{k}=1}^{\mathrm{m}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_\mathrm{i}^\mathrm{k} \mathrm{k}(\mathbf{x}_\mathrm{i}, \mathbf{x}), \tag{9}$$

where r is the number of principal components in F.

4 Learning Kernel Subspace Classifier in Input Space

The kernel subspace classifier based on Eqn (8) means performing PCA in F with optimal reconstruction of $\Phi(\mathbf{x})$ (the map of a test point $\mathbf{x}$ in F) based on its projections, i.e.,

$$\rho(\mathbf{x}) = ||\mathrm{P_r}\Phi(\mathbf{x}) - \Phi(\mathbf{x})||^2 \tag{10}$$

is minimized for a mapped test point with its projection onto the subspace spanned by the first r eigenvectors:

$$P_r\Phi(\mathbf{x}) = \sum_{\mathrm{k}=1}^{\mathrm{r}} \alpha_{\mathrm{k}}\mathbf{v}_{\mathrm{k}} \tag{11}$$

where P_r is the projection operator in F.

However, distance in Eqn (10) does not give the reconstruction of $\mathbf{x}$ in input space. Consequently, kernel subspace classifier based on Eq (9) or Eq. (10) can not work well, particularly for many classifications in biometrics where the small sample size (SSS) difficulties make the situation worse. For the KPCA to be efficient in data classification, the reconstructed pre-image of $\Phi(\mathbf{x})$ should be as close to $\mathbf{x}$ as possible, following the same principle of PCA.

The subject of data reconstruction has been discussed in the past with the name *data de-noising* or *pre-image* of KPCA [7-9]. That means, we are looking for an explicit vector $\mathbf{z} \in \mathrm{R^n}$ satisfying $\Phi(\mathbf{z}) = \mathrm{Pn}\Phi(\mathbf{x})$. In other words, pre-image concerns the best reconstruction of mapped data in the kernel space and the solution can be approximated by minimizing the squared distance $\rho(\mathbf{z})$ between the Φ-image of a vector $\mathbf{z}$ and the reconstructed pattern in F:

$$\rho(\mathbf{z}) = ||\Phi(\mathbf{z}) - \mathrm{P_n}\Phi(\mathbf{x})||^2 \tag{12}$$

For kernels satisfying $k(\mathbf{x}, \mathbf{x}) = \mathrm{const}, \forall \mathbf{x}$, an optimal $\mathbf{z}$ can be determined by an iterative update scheme as follows

$$\mathbf{z}_{\mathrm{t}+1} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} \gamma_{\mathrm{i}} \exp(-||\mathbf{z}_{\mathrm{t}} - \mathbf{x}_{\mathrm{i}}||^2/\mathrm{c})\mathbf{x}_{\mathrm{i}}}{\sum_{\mathrm{i}=1}^{\mathrm{N}} \gamma_{\mathrm{i}} \exp(-||\mathbf{z}_{\mathrm{t}} - \mathbf{x}_{\mathrm{i}}||^2/\mathrm{c})} \tag{13}$$

The popular kernel type which satisfies $k(x; x) = \mathrm{const}$ are, *e.g.*, the RBF kernels. Though Eqn (13) seems applicable with the kernel subspace classifier paradigm, we will empirically prove in Section 5 that it does not work.

While pre-image of KPCA addresses the minimization of reconstruction error in kernel space F, we emphasize the data reconstruction in input space after the KPCA projection, as this will explicitly express the representation capability of the kernel subspace for the data class. We formulated the problem as *learning kernel subspace*, with the objective of minimization of reconstruction error for the input data. The objective can be simply solved based on the kernel principal component regression [3], which define the data reconstruction as a the following regression problem from kernel space:

$$\hat{\mathbf{x}} = \mathbf{\Phi}\xi + \epsilon \tag{14}$$

where $\mathbf{\Phi}$ is an matrix composed of vector $\Phi(\mathbf{x}_{\mathrm{i}})$, ξ is a vector of regression coefficients and ϵ is the error term. Performing PCA on $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$ will result in M eigenvalues $\{\lambda_j\}_{j=1}^{M}$ and corresponding eigenvectors $\{\mathbf{V}^{\mathrm{j}}\}_{\mathrm{j}=1}^{\mathrm{M}}$. The projection of

the $\mathbf{\Phi}(\mathbf{x})$ onto the k-th principal components is given by Eqn (6). By projecting all the $\mathbf{\Phi}(\mathbf{x})$ onto the principal component, the above equation becomes

$$\hat{\mathbf{x}} = \mathbf{\Psi}\mathbf{w} + \epsilon \tag{15}$$

where $\mathbf{B} = \mathbf{\Phi}\mathbf{V}$ is an $(n \times M)$ matrix and $\mathbf{V}$ is an $(M \times M)$ matrix with $\mathbf{V}^{\mathrm{k}}$ as its $k-$th column. The least squares estimate of the coefficients $\mathbf{w}$ becomes:

$$\hat{\mathbf{w}} = (\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{x} = \mathbf{\Lambda}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{x} \tag{16}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_{\mathrm{M}})$.

The proposed model (Eqn. (16)) has been discussed earlier by the author from the point of view of auto-associator model [18], which is a direct result of applying the kernel principal component regression [3]. The classification scheme has proved its efficiency in general face recognition problem [18] and in cancer classification [17]. Formulating the methodology in the framework of kernel subspace classifier not only justifies the model theoretically but also clarifies some confusions arose from recent works on kernelization of subspace classification.

In summary, the kernel subspace classifier model provides a description of the nonlinear relationships between input and features from the kernel space. The model building involves two operations. The first is the kernel operation which transforms an input pattern to a high-dimensional feature space. The second is the mapping of the feature space back to the input space. The proposed kernel subspace classifier proves satisfactory performance on some benchmarking robust face recognition problems, as explained in next section.

5 Experiments

5.1 Experiment with the AR faces

AR faces consist of frontal facial images of 135 subjects (76 male and 59 females), with 26 different images for each subjects. For each subject, the images were recorded in two different sessions separated by two weeks, each session consisting of 13 images. Each image is with 768×576 pixels.

Following the practice in [21, 12], we used the images of 50 subjects (the first 25 males and 25 females). In the pre-processing step, the original images were converted to gray scale, aligned by the eyes, resized, and cropped to size 104×85. In our experiment, the non-screaming and non-occluded images from both sessions were used for the training of each subject's kernel subspace classifier and the remaining occluded images by sunglasses and scarf and images of the screaming expression were used for testing. The first row of Fig. 1 gives examples of training images of the third subject from the AR database, while the second row of Fig 1 contains the test images for the same subject.

Recently, occlusion robust face recognition has attracted much attention and several algorithms have been published. In [15] a Selective Local Nonnegative Matrix Factorization (SL-LNMF) technique was proposed, which includes occlusion detection step and the selective LNMF-based recognition step. Paper

Fig. 1. Sample images from the AR database. First row: training images. Second row: test images with occlusion by sunglasses/scarf, and with screaming expression.

[13] proposed a *Face-ARG matching* scheme in which a line feature based Face-ARG model is used to describe face images. Based on robust estimation, [14] propounded a classification method that combines reconstructive and discriminative models. For brief, we term it as Reconstructive and Discriminative Subspace model (RDS). These published recognition performances on the AR face are compared in the Table 1. It is worthy to note that the experiment settings from these publications are not exactly same as us except the RDS in [14]. Therefore the comparison can only give an intuitive meaning.

Table 1. Comparison of the recognition accuracies

	RDS	IS-ICA	S-LNMF	Face-ARG	Pre-image	Kernel Subspace
sunglasses	84%	65%	90%	80.7%	50%	**92%**
scarf	93%	NA	92%	85.2%	13%	**51%**
scream	87%	NA	44%	66.7%	43%	**95%**

Figure 2 further explains the works of the proposed method. The first column displays the probe images from sunglasses/scarf occluded faces and the screaming face. The images from second column to sixth column are the first five best reconstructed images from the corresponding probe by applying the kernel subspace classifier. It can be observed that for sunglasses occluded face

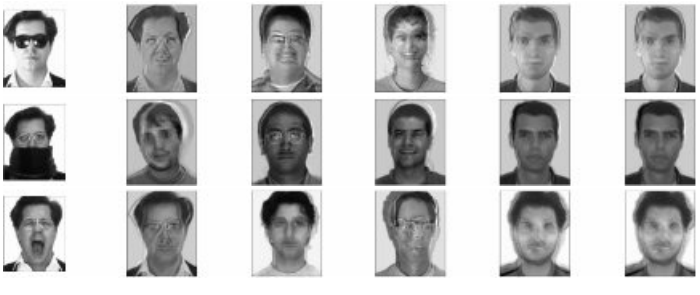

Fig. 2. Reconstruction of probe images from the kernel subspace classifier. First column: probe images; Second column to sixth column are the first five best reconstructed images from the corresponding probe image.

and screaming face, the kernel subspace classifier gives reasonably good reconstructions, thus yielding high recognition accuracies as shown in Table 1. For the scarf occluded face, however, the reconstruction is pretty poor, which is consistent with the low accuracy 51%.

As the pre-image problem of KPCA is relevant to the kernel subspace classifier, we also applied it to the AR faces with result shown in Table 1. The poor performance is in agreement with the visualization of the reconstructed images from the three kind of probe face images, as illustrated in Fig. 3.

Fig. 3. Reconstruction of probe images from the pre-image algorithm Eqn. (13). First column: probe images; Second column to sixth column are the first five best reconstructed images from the corresponding probe image.

5.2 Experiment with the UPC Faces

In the second experiment, we used the UPC faces data provided by Universitat Politecnica de Catalunya [16], which was created for the purpose of testing the robustness of face recognition algorithms against strong occlusion, pose and illumination variations. This database includes a total of 18 persons with 27 pictures per person which correspond to different pose views. In our experiment, we chose 8 near-front images per person for training while used occluded images for testing, with occlusions from sunglasses or hands, as illustrated in the following Figure 4.

We tested the recognition performances on two different occlusions. The first is sunglasses occlusion which is similar to the AR face scenario. The second is occlusion by hand as shown in right of Fig. 4. The recognition accuracies from our proposed kernel subspace are 80% and 86% respectively. Figure 5 illustrated the corresponding reconstructed images from the two probe faces. As a comparison, the recognition accuracies from the pre-image algorithm are 43% and 47%, which shows again its unacceptability in kernel subspace classification.

Fig. 4. Left: training samples from the UPC data set; Right: some of the testing mages

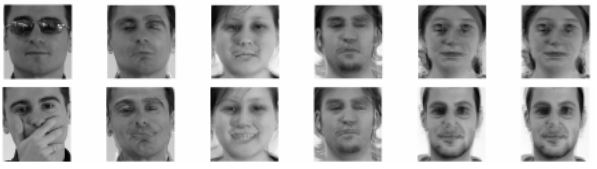

Fig. 5. First column: probe images; Second column to sixth column are the first five best reconstructed images from the corresponding probe

Table 2. Comparison of the recognition accuracies

	Pre-image	Kernel Subspace
sunglass	43%	**80%**
others	47%	**86%**

6 Conclusion

In this paper, a new kernel subspace classifier algorithm is proposed which is based on the KPCA image reconstruction in input space after the KPCA projection. With the objective of minimizing the reconstruction error in the input space, the least square regression is applied to map the KPCA projection from the implicit feature space to the input space. Our experiments on some occluded face recognition problems using the AR face and UPC face show very encouraging performance, which also compare favorably with some very complex occlusion robust face recognition methods proposed in recent years.

Acknowledgement

This research was supported by the MIC(Ministry of Information and Communication), Korea, Under the ITFSIP (IT Foreign Specialist Inviting Program) supervised by the IITA (Institute of Information Technology Assessment).

References

1. Oja, E.: Subspace Methods of Pattern Recognition. Research Studies Press, Letchworth and J.Wiley (1983)
2. Laaksonen, J., Oja, E.: Subspace Dimension Selection and Averaged Learning Subspace Method in Handwritten Digit Recognition. In: Vorbrüggen, J.C., von Seelen, W., Sendhoff, B. (eds.) Artificial Neural Networks - ICANN 96. LNCS, vol. 1112, pp. 227–232. Springer, Heidelberg (1996)
3. Rosipal, R., Girolami, M., Trejo, L., Cichocki, A.: Kernel PCA for Feature Extraction and De-Noising in Non-linear Regression. Neural Computing & Applications 10, 231–243 (2001)
4. Tsuda, K.: Subspace classifier in the Hilbert Space. Pattern Recognition Letters 20, 513–519 (1999)

5. Maeda, E., et al.: Multi-category Classification by Kernel based Nonlinear Subspace Method. ICASSP 1999 2, 1025–1028 (1999)
6. Scholkopf, B., Smola, A.: Muller K-R. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10, 1299–1319 (1998)
7. Bakir, G.H., Weston, J., Scholkopf, B.: Learning to Find Pre-Images. In: Advances in Neural Information Processing Systems, vol. 16, pp. 449–456. MIT Press, Cambridge, MA, USA (2004)
8. Mika, S., Scholkopf, B., Smola, A., Muller, K., Scholz, M., Ratsch, G.: Kernel PCA and De-Noising in Feature Spaces. In: Proc.1998 conference on Advances in neural information processing systems II, pp. 536–542. MIT Press, Cambridge, MA, USA (1998)
9. Scholkopf, B., Smola, A., Muller, K.: Kernel principal component analysis. In: Scholkopf, B., Burges, C.J.C., Smola, A.J. (eds.) Advances in Kernel Methods - SV Learning, pp. 327–352. MIT Press, Cambridge, MA. USA (1999)
10. Rosipal, R., Trejo, L.: Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space. Journal of Machine Learning Research 2, 97–123 (2001)
11. Kim, J., Choi, J., Yi, J., Turk, M.: Effective Representation Using ICA for Face Recognition Robust to Local Distortion and Partial Occlusion. IEEE Trans. PAMI 27, 1977–1981 (2005)
12. Martinez, A., Kak, A.: PCA versus LDA. IEEE Trans. PAMI 23, 228–233 (2001)
13. Park, B., Lee, K., Lee, S.: Face Recognition Using Face-ARG Matching. IEEE Trans. PAMI 27, 1982–1988 (2005)
14. Fidler, S., Skocaj, D., Leonardis, A.: Combining Reconstructive and Discriminative Subspace Methods for Robust Classification and Regression by Subsampling. IEEE Trans. PAMI 28, 337–350 (2006)
15. Oh, H., Lee, K., Lee, S.: Occlusion invariant face recognition using selective LNMF basis images. In: Narayanan, P.J., Nayar, S.K., Shum, H.-Y. (eds.) ACCV 2006. LNCS, vol. 3851, pp. 120–129. Springer, Heidelberg (2006)
16. UPC Face Database: `http://gps-tsc.upc.es/GTAV`
17. Zhang, B.: Cancer Classification by Kernel Principal Component Self-regression. In: Australian Conf. on Artificial Intelligence 2006, Horbat, Australia, pp. 719–728 (2006)
18. Zhang, B.: Kernel Auto-associator from Kernel Principal Component Autoregression with Application to Face Recognition. In: Proc. Int'l Conf. Comput. Inte. for Modeling, Control & Automation (CIMCA) 2005, Vienna, pp. 15–19 (2005)