Czech Pattern Recognition Workshop 2000, Tomáš Svoboda (Ed.)
Peršlák, Czech Republic, February 2–4, 2000
Czech Pattern Recognition Society

Robust Eigenspace Construction*

Horst Wildenauer and Horst Bischof
Pattern Recog. and Image Proc. Group
Vienna University of Technology
Favoritenstr. 9-1832
A-1040 Vienna, Austria

Aleš Leonardis
Faculty of Computer and Info. Science
University of Ljubljana
SI-1001 Ljubljana, Slovenia

Abstract *In this paper we extend our previous work on constructing multiple eigenspaces from a set of training images based on the Minimum Description Length (MDL) principle. To enhance the robustness we propose a robust framework for calculating the eigenspaces. We also propose a mechanism for incrementally adding new eigenspaces to increase the computational efficiency.We illustrate the robust performance of the new algorithm on 2D point data, and demonstrate the ability to form object specific eigenspaces which enhances the classification accuracy on the ORL face recognition database.*

1 Introduction

In the past years we have seen an extensive use of methods applying eigenspace methods for recognition of objects [14, 9]. Most commonly, all the training images are encoded by a single eigenspace whose dimension is determined on the basis of the maximum allowable reconstruction error. In some cases (see for example [10, 9]), the appearances of individual objects are encoded by separate eigenspaces. However, the fact that the correlation between the images (even between the images of views of a single object) may not be significant, has not been taken into account in any of these cases. Such eigenspaces ignore the locally low-dimensional structure of the data. Also, the generalization property of such eigenspaces is usually rather poor. Recently, there has been work on constructing multiple eigenspaces using mixture models e.g., [2, 3, 1]. However, all these approaches require that both the number of mixture components and the dimensionality of the PCA subspaces is *a priori* given. This is a major difference from the approach proposed by Leonardis and Bischof [7], where these two parameters are automatically determined by the MDL principle.

Leonardis & Bischof [7] have proposed a novel approach for construction of (possibly) multiple eigenspaces from a set of training images based on the Minimum Description Length principle (MDL). The main idea is to systematically build a redundant set of eigenspaces by *eigenspace-growing*. These eigenspaces are treated as hypotheses that are subject to a selection procedure (*eigenspace-selection*), which selects a subset of the eigenspaces based on the MDL principle. Eigenspace-growing and eigenspace-selection are iteratively combined to yield a numerically feasible algorithm. In this paper we extend the approach by using a robust formulation of the eigenspace growing procedure. In addition, we propose an incremental hypothesis generation step which decreases the number of generated hypotheses, thus improving the computational complexity.

The paper is organized as follows: First, we briefly review our MDL approach for constructing multiple eigenspaces. In section 3 we present the robust eigenspace construction and the algorithm for adding new eigenspaces. The experimental results are shown in section 4. We conclude with a summary and outline the work in progress.

2 Multiple Eigenspaces by MDL

Let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_n | \mathbf{x}_i \in \mathbb{R}^N\}$ be a set of images. The aim is to compress the set of images $\mathcal{X}$ in terms of a set of low-dimensional eigenspaces. Each image from the set $\mathcal{X}$ can then be represented as a linear combination of eigenimages from an eigenspace:

$$\mathbf{x}_i = \sum_{j=1}^{m} \mathcal{I}_j{}^{(i)} \sum_{k=1}^{d_j} c_{jk}^{(i)} \mathbf{e}_{jk}^{(i)} \quad . \tag{1}$$

$\mathcal{I}_j{}^{(i)}$ is a variable which is 1 for the j-th eigenspace which encodes the image $\mathbf{x}_i$ and 0 otherwise, m is the number of all eigenspaces, d_j is the dimension of the j-th eigenspace, and c_{jk} and $\mathbf{e}_{jk}$ are the corresponding coefficients and eigenimages, respectively. Often, only p_j, $p_j < d_j$, eigenimages are needed to represent an image to a sufficient degree of accuracy, yielding an image approximation $\hat{\mathbf{x}}_i$. p_j is called the *effective dimension* of the j-th eigenspace.

The goal is to partition the set of images $\mathcal{X}$ into a set of subsets $\mathcal{G}_i$, ($\mathcal{G}_i \subset \mathcal{X}, |\mathcal{G}_i| = k_i$), which contain images that lie on a linear subspace of low dimensionality and can thus be encoded efficiently in terms of eigenspaces. k_i denotes the number of images in the subset $\mathcal{G}_i$. Since the goal is to achieve a compact representation, the MDL principle [11] was chosen as a global optimization criterion.

* A. L. acknowledges the support from the Ministry of Science and Technology of Republic of Slovenia (Projects J2-0414 and J2-8829). H. B. was supported by a grant from the Austrian National Fonds zur Förderung der wissenschaftlichen Forschung (No. S7002MAT).

To minimize the description length the following iterative approach (for a detailed derivation see[7]) was used:

1. **Initialization:** An initial set of eigenspace hypothesis is formed by taking a small subset of images $\mathcal{G}_i^{(0)}$, $|\mathcal{G}_i^{(0)}| = k_i$, $(k_i << n)$ from the set $\mathcal{X}$, calculating the eigenspace $\mathcal{E}_i^{(0)}(\mathcal{G}_i^{(0)})$, and determining the effective dimension $p_i^{(0)}$[1]. The dimension of the eigenspace is found by minimizing the length of encoding of the data in terms of the eigenspace (denoted by $L(\mathcal{G}_i^0)$),

$$\begin{aligned} L(\mathcal{G}_i^0) &= \mathrm{K}_1 p_i^{(0)} + \mathrm{K}_2 |\mathcal{G}_i^0| p_i^{(0)} + f(\sum_{j=p_i^{(0)}+1}^{d_i^{(0)}} \lambda_j) \\ &\approx \mathrm{K}_1 p_i^{(0)} + \mathrm{K}_2 |\mathcal{G}_i^0| p_i^{(0)} + \mathrm{K}_3 \sum_{j=p_i^{(0)}+1}^{d_i^{(0)}} \lambda_j . \end{aligned} \quad (2)$$

   Here λ_i denotes the eigenvalues of the corresponding eigenspace, and K_1, K_2, and K_3 are constants. The first term on the right hand side is the length of encoding of p_i eigenimages, the second one is the length of encoding of p_i coefficients for each image in $\mathcal{G}_i$, and the third one is the length of encoding of the deviations between the data and the model[2].

   The initial set of images $\mathcal{G}_i^{(0)}$ can be chosen on the basis of proximity of images (spatial or temporal), on some a priori knowledge, or randomly.

2. **Eigenspace Growing:** To *grow* the eigenspace $\mathcal{E}_i^{(t)}$, a search for images that are *compatible* with the current eigenspace has to be performed and can thus be added to the subset $\mathcal{G}_i^{(t)}$. Each image is checked for the reconstruction error ($||\mathbf{x} - \hat{\mathbf{x}}||$) and if it is bellow a compatibility threshold Θ the image is included in the current eigenspace, then the effective dimension of the new eigenspace is determined. These steps lead to a new hypothesis.

3. **Eigenspace Selection:** The problem of eigenspace selection is solved in the framework of the MDL principle. It is defined as an optimization problem which minimizes the length of encoding of $\mathcal{X}$, i.e., $L(\mathcal{X})$, which is equivalent to maximizing the savings in the length of encoding. The objective function has the following form

$$F(\mathbf{h}) = \mathbf{h}^T \mathbf{C} \mathbf{h} = \mathbf{h}^T \begin{bmatrix} c_{11} & \dots & c_{1R} \\ \vdots & & \vdots \\ c_{R1} & \dots & c_{RR} \end{bmatrix} \mathbf{h} \ . \quad (3)$$

   Vector $\mathbf{h}^T = [h_1, h_2, \dots, h_r]$ denotes a set of hypotheses (eigenspaces), where h_i has the value 1 for the presence and 0 for the absence of the hypothesis (eigenspace) i in the resulting description. The diagonal terms of the matrix $\mathbf{C}$ express the cost-benefit value for a particular hypothesis (eigenspace) i, i.e., $c_{ii} = S(\mathcal{E}_i(\mathcal{G}_i))$. Where $S(\mathcal{E}_i(\mathcal{G}_i))$ denotes the *savings* in the length of encoding for each eigenspace $\mathcal{E}_i$,i.e., the difference between coding the images in $\mathcal{G}_i$ individually or in terms of the eigenspace.

$$S(\mathcal{E}_i(\mathcal{G}_i)) = \mathrm{K}_0 |\mathcal{G}_i| - (\mathrm{K}_1 p_i + \mathrm{K}_2 |\mathcal{G}_i| p_i + \mathrm{K}_3 \chi_i) \ . \quad (4)$$

   where χ_i denotes the error over $\mathcal{G}_i$, and K_0, K_1, K_2, and K_3 are constants. K_0 is related to the average cost of describing an image (in bits) in the absence of the eigenspace, K_1 is related to the cost of encoding an eigenimage, K_2 is related to the average cost of specifying a coefficient, and K_3 is related to the average cost of specifying the error.

   Note that for images $\mathrm{K}_0 >> \mathrm{K}_2 p_i$, therefore the term $\mathrm{K}_2 |\mathcal{G}_i| p_i$ can be neglected; also it can usually be assumed that $\mathrm{K}_0 \approx \mathrm{K}_1$.

   The off-diagonal terms c_{ij} take into account that the intersections between image subsets $\mathcal{G}_i$, $\mathcal{G}_j$ may not be empty, i.e., that the same images are included in more than one subset. Here only pairwise intersections are considered, however, intersections including multiple subsets could also be taken into account.

$$\begin{aligned} c_{ij} &= (-\mathrm{K}_0 |\mathcal{G}_i \cap \mathcal{G}_j| + \mathrm{K}_2 |\mathcal{G}_i \cap \mathcal{G}_j| \max(p_i, p_j) \\ &\quad + \mathrm{K}_3 \chi_{ij})/2 \end{aligned} \quad (5)$$

   χ_{ij} is the maximal error of the images in the intersection of $\mathcal{G}_i \cap \mathcal{G}_j$ with respect to the $\mathcal{E}_i$ and $\mathcal{E}_j$. To maximize the objective function $F(\mathbf{h})$, Eq. (3) Tabu search [4]is used.

   Note that we can extend the above framework, such that it is possible that some of the images from $\mathcal{X}$ are not included in any of the eigenspaces and will be encoded separately—these images are called *outliers*. Let us denote their length of encoding as $L(\mathcal{O})$.

4. **Termination:** If the eigenspaces have converged (i.e. no new images are included) stop, otherwise goto 2.

3 Incremental and Robust Eigenspace growing

The algorithm outlined in the previous section can be improved in at least two ways:

1. The number of initial hypotheses: To guarantee (in probabilistic terms) that no seeds leading to a good solution are missing, a large number of hypotheses has to be generated. Many of these hypotheses are redundant and will be removed by the selection. This results in a high computational complexity which is further increased in case of outliers in the training set.

[1] Superscript $^{(t)}$ refers to the t-th iteration.

[2] The term $\sum_{j=p_i+1}^{d_i} \lambda_j$ can be calculated efficiently without the need to calculate all eigenimages (eigenvalues) [8].

2. The parameter Θ which determines when an image is included in an eigenspace: If this parameter is set too low, the eigenspaces will not grow. If it is set too high, a highly overlapping set of eigenspaces will result, with the tendency of increasing their dimension.

In order to avoid these problems we propose two mechanisms which will be explained in the sequel.

3.1 Incrementally Adding Eigenspaces

The following observation can be used to drastically decrease the necessary number of initial hypotheses: Since the selection procedure compares the eigenspaces on a relative basis, new eigenspaces can be generated at any time. If we generate new eigenspaces using the initialization procedure on currently unexplained images $\mathcal{X} - \bigcup \mathcal{G}_i$ (i.e., images not covered by any eigenspace) we increase the likelihood of finding eigenspaces that have not yet been explored. The idea is very similar to the *data driven exploration* proposed in [13].

3.2 Robust Eigenspaces

In order to achieve a more robust updating of the eigenvectors and to be less sensitive on the parameter Θ, we use a weighted eigenspace formulation in the spirit of robust estimators like hard/soft redecenders[5, 12].

The weight of an image is determined by the distance from the eigenspace, i.e. $w_i \approx \frac{1}{||\mathbf{x}_i - \hat{\mathbf{x}}_i||}$. In particular, we use a trapezoidal function $f :\mapsto [0..1]$ as a weighting function. In order to calculate the "weighted" eigenspace we can use Singular Value Decomposition (SVD). We can premultiply each image with the corresponding weight prior to SVD (since SVD assumes zero mean data, this has also to be considered in the calculation). This process can be iteratively repeated, thereby downweighting (ignoring) outliers. Through this scheme we assure that images with a high reconstruction error have less or no influence on the eigenspace calculation of $\mathcal{G}_i$, therefore they can do not influence the final result.

Fig. 1 demonstrates the robustness of this approach. The original dataset is corrupted with 5 outliers. One can see that after 3 iterations the solution has converged to the original eigenspace corresponding to the original dataset without outliers. Several experiments have shown that this procedure converges reliably and fast (usually less than five iterations are sufficient).

3.3 The Algorithm

The two above mentioned improvements can now be integrated into the complete algorithm as depicted in detail in Fig. 2. The algorithm can be roughly divided into four stages:

1. Initialization
2. Robust Eigenspace Growing
3. Eigenspace Selection
4. Adding Hypotheses

which are described in the following.

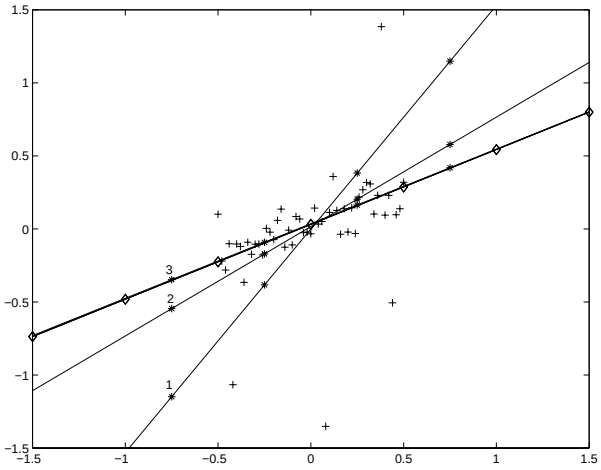


Figure 1: Robust construction of eigenspaces on a data set disturbed with 5 outliers. Robust PCA (lines marked with $\star$), converges in 3 iterations towards the Eigenvector of the non-spoiled data set (line marked with $\diamond$).

1. **Initialization:** This step is equivalent to the initialization step of the original Eigenspace algorithm (section 2). The initial set of hypothesis is formed by taking a small subset of images $\mathcal{G}_i^{(0)}$, $|\mathcal{G}_i^{(0)}| = k_i$, $(k_i << n)$ from the set $\mathcal{X}$, calculating the eigenspace $\mathcal{E}_i^{(0)}(\mathcal{G}_i^{(0)})$, and determining the effective dimension $p_i^{(0)}$.

   Note that due to the possibility to add hypotheses there is no longer a need to initialize a large set of hypotheses to ensure a sufficient exploration of the input space.

2. **Robust Eigenspace Growing:** First, for each eigenspace a subset of compatible images $\mathcal{G}_i^{(t)}$ is determined in the *subset selection step*. For these subsets the new eigenspaces (hypotheses) and their effective dimension Eq. 2 is calculated (*Eigenspace Updating*) by the proposed robust Eigenspace algorithm. If the new MDL-rating Eq.(4) is better then MDL-rating of the previous hypothesis, the updated eigenspace $\mathcal{E}_i^{(t)}$ replaces the eigenspace $\mathcal{E}_i^{(t-1)}$ of the previous iteration. Otherwise a pre-chosen percentage of the least compatible images are excluded from the hypotheses (*subset decrementation*) and the MDL-rating is compared again. If $||\mathcal{G}_i^{(t)}||$ is smaller or equal to the original number of images (before the eigenspace growing step) the original eigenspace remains unaltered. Through this growing scheme we can assure that:

   - Images which tend to decrease the quality of the eigenspace in terms of the MDL-rating are not included in a new hypothesis, and
   - Good hypotheses found in an early iterations can be propagated unaltered throughout the iterations of the algorithm.

3. **Eigenspace Selection:** In the eigenspace selection stage the valid hypotheses for the next robust eigenspace growing iteration are chosen. This is done by maximizing the

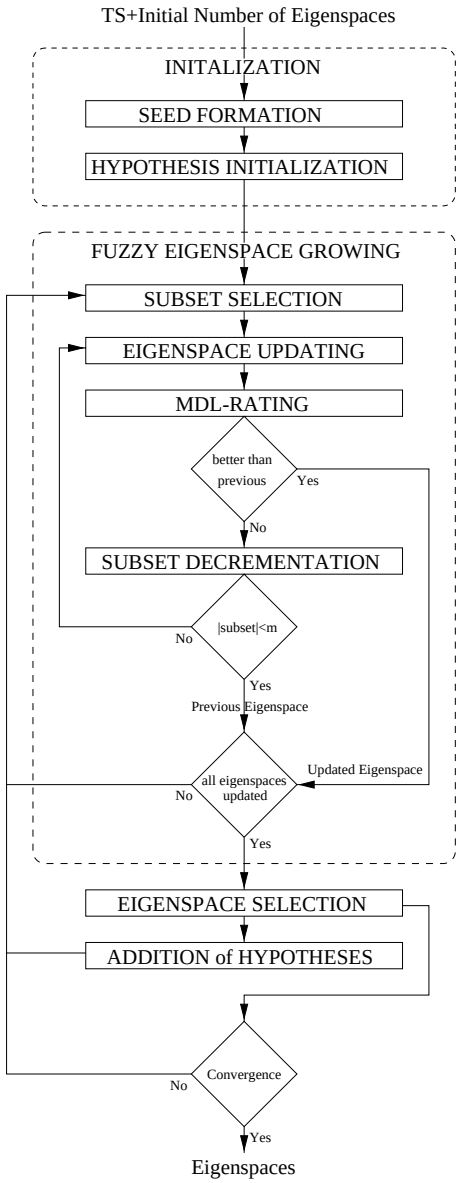


Figure 2: Structure and data flow of the Robust Eigenspaces algorithm

objective function $F(\mathbf{h})$, Eq. 3 using Tabu search. For details see section 2.

4. **Addition of Hypotheses**: For images which are currently not encompassed by any eigenspace $\mathcal{X} - \bigcup \mathcal{G}_i$ new eigenspaces are generated by using the initialization procedure. Through this scheme we increase the likelihood of finding eigenspaces that have not yet been explored.

   Note that after a hypotheses has been added at least one eigenspace growing and selection step has to be performed in order to ensure that no newly initialized hypothesis can be part of the final solution without comparing it to other hypotheses.

   In order to prevent the algorithm from adding and deleting the same hypotheses over and over again each image has an counter, which incremented when the image is chosen to be part of a new hypotheses, if this counter is above a threshold for all images the algorithm terminates.

4 Experimental Results

First we present a 2-D example to demonstrate the main features of the method. Fig. 3 shows a set of 240 points in a 2-D space. 200 of these points originating from four lines are perturbed by additive Gaussian noise with a $\sigma = 0.02$. The remaining 40 points, which are considered as outliers (clutter), are generated by a uniform distribution over $[0..1] \times [0..1]$.

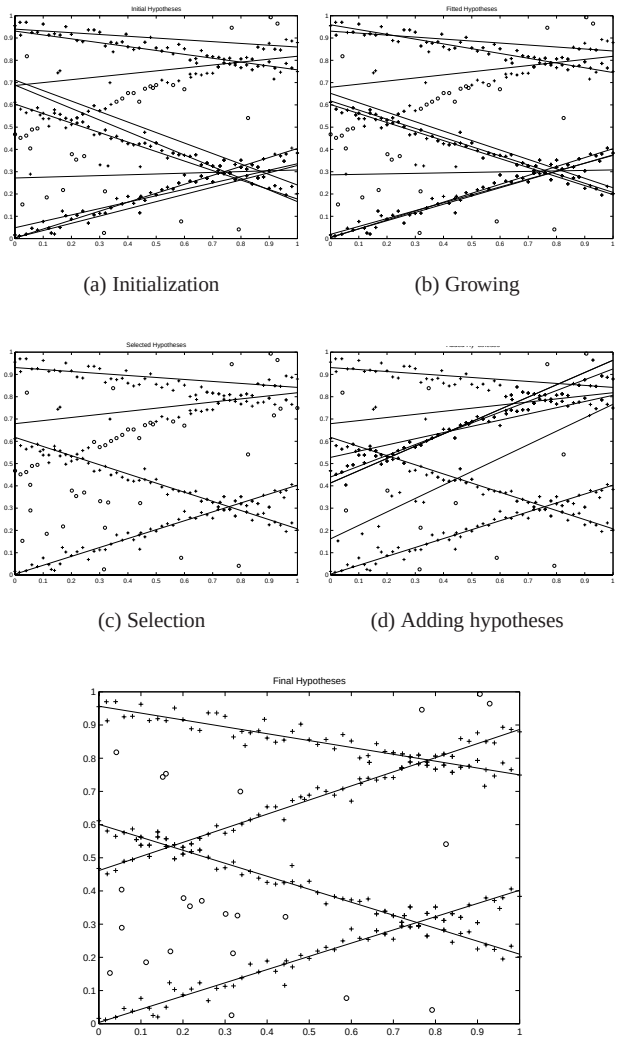
We applied our method to the data and obtained four eigenspaces, each having one dominant dimension ($\lambda_{1_i} >> 0, \lambda_{2_i} \approx 0$). Fig. 3 shows the intermediate stages and the final result. Fig. 3(a) shows the initial hypotheses; one can see that one line is not covered by any hypotheses. In the growing step (Fig. 3(b)) these initial hypotheses are considerably improved (i.e. they cover more points). The redundant hypotheses are eliminated by the selection step (Fig. 3(c)). In the next step new hypotheses are generated for the unexplained points (Fig. 3(d)). An additional growing step (not shown) improves these hypotheses and a further selection step gives the final result (Fig. 3(e)). All four lines have been successfully recovered and the uniformly distributed points have been classified as outliers.

In addition to the efficient representation (measured in bits), we can now discriminate the points that lie close to the four lines (i.e., they have a low reconstruction error) from those that lie far away from them. This would not be possible having a single eigenspace.

Next, we have evaluated our algorithm extensively on the ORL face database. This database contains for each individual 10 face images (different head orientations and facial expressions). In total there are 40 individuals in this database.

Fig. 4 shows four typical clusters obtained by our method. Note that the algorithm found automatically and unsupervised for each individual a separate eigenspace. For most clusters the dimension of the eigenspace is 1, for a few clusters the dimension is 2 (when the pose is varying significantly). Table 1 compares the mean reconstruction error, its variance, and the number of eigenimages needed for representing the images in the following cases: a single eigenspace, one eigenspace for each individual (i.e. partitioning is done manually), and the four eigenspaces produced with our method (for the first two cases the dimensionality of the eigenspace is determined as capturing 95% of the variance). While using the same number of eigenimages for representing the images, our method yields the lowest mean reconstruction error of all three cases. The difference between the eigenspaces produced by our method and the four individual eigenspaces is due to the robust eigenspace calculation, i.e., for the individual eigenspaces each face is weighted by 1, which is not necessarily true for our method.

In the next experiment we compared the recognition rate (whether faces of different individuals are correctly recognized) for the eigenspaces obtained with our method and the object-specific eigenspaces. The images to be recognized are corrupted by replacement noise (i.e., we selected



(a) Initialization (b) Growing

(c) Selection (d) Adding hypotheses

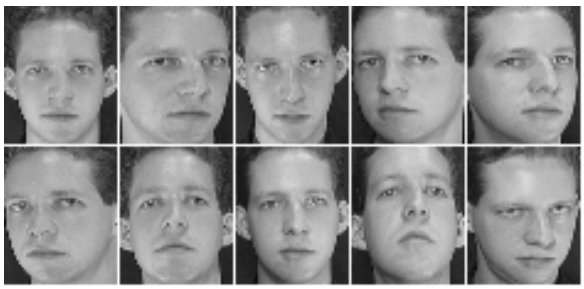
(e) Final result

Figure 3: Illustration of our method on a set of points in 2-D. Points encompassed by any eigenspace are represented by $+$. 'Outliers' are marked with $\circ$.

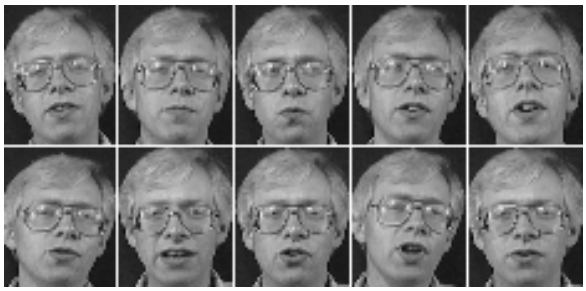
randomly 10%–90% of the pixels of an image and replaced them by uniformly distributed random numbers between $[0...255]$). One can see that the eigenspaces constructed with our method yield a higher recognition rate (Fig. 5(a)). It is also interesting to note that, when we vary the rejection threshold we can achieve a higher recognition rate with our eigenspaces than with the object-specific eigenspaces (Fig. 5(b) depicts this situation for a noise level of 60%).

5 Conclusions and summary

We have presented a novel approach to constructing robust eigenspaces from a set of training images based on the MDL principle. The results indicate that using the robust eigenspaces obtained by our method is more powerful (in terms of recognition rate) than using a single (all-encompassing) eigenspace, or multiple eigenspaces based on ad-hoc partitioning of the input images. Thus, the method should be of interest to all researchers using eigenspace

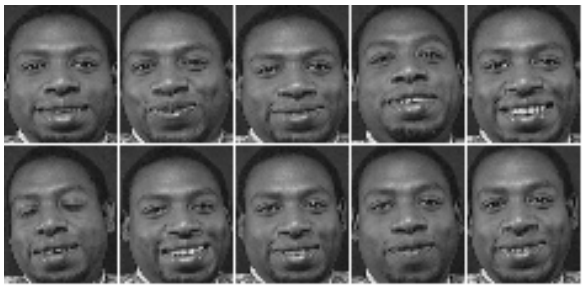


(a) Subset of images which define the 1st eigenspace



(b) Subset of images which define the 2nd eigenspace

(c) Subset of images which define the 3rd eigenspace

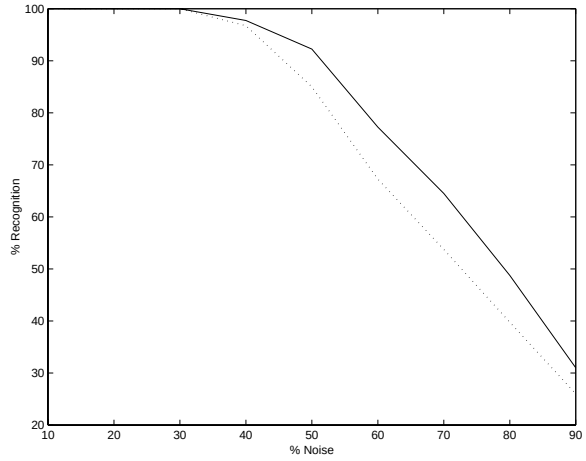


(d) Subset of images which define the 4th eigenspace

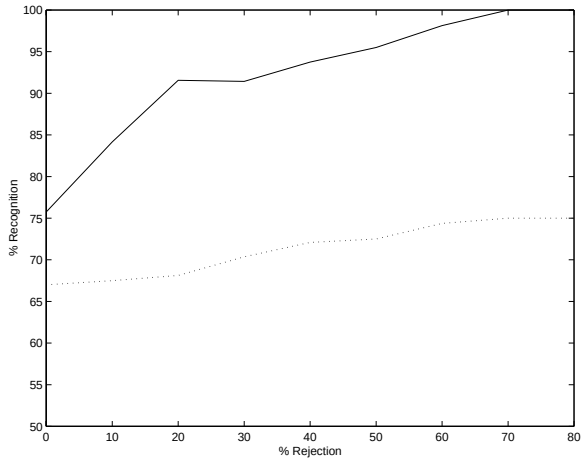
Figure 4: Images encompassed by four eigenspaces for the set of face images.

Table 1: Comparison of different eigenspaces.

Method	Mean Error	Variance	# Eigenimages
One ES	$4.03 * 10^4$	$7.58 * 10^7$	4
Face-1 ES	$5.71 * 10^4$	$3.20 * 10^7$	1
Face-2 ES	$3.70 * 10^4$	$1.15 * 10^7$	1
Face-3 ES	$3.67 * 10^4$	$2.86 * 10^7$	1
Face-4 ES	$3.54 * 10^4$	$3.59 * 10^7$	1
Face-1 RFES	$4.91 * 10^4$	$9.18 * 10^7$	1
Face-2 RFES	$3.13 * 10^4$	$5.39 * 10^7$	1
Face-3 RFES	$3.00 * 10^4$	$4.72 * 10^7$	1
Face-4 RFES	$2.95 * 10^4$	$6.11 * 10^7$	1



(a) Recognition rate versus the % of replacement noise



(b) Recognition rate versus rejection rate for a noise level of 60%

Figure 5: Classification results for object-specific eigenspaces (dotted line) and for eigenspaces obtained with our method (full line).

methods for recognition.

There are various possible extensions of the method proposed in this paper. Particularly interesting is the extension to mixture PCA models [2, 6, 3], which would allow an automatic selection of the number of mixtures and the dimensionality of the individual PCA spaces. Basically, we can use a similar selection procedure also for these models, we only need to change the growing phase which has to be replaced by an EM-like algorithm.

References

[1] C. Bishop and M. Tipping. A hierarchical latent variable model for data visualization. *IEEE Trans. on PAMI*, 20(3):281–293, 1998.

[2] C. Bregler and S. M. Omohundro. Nonlinear image interpolation using manifold learning. In G. Tesauro, D. Touretzky, and T. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7, pages 973–980. The MIT Press, 1995.

[3] B. Frey, A. Colmenarez, and T. Huang. Mixtures of local linear subspaces for face recognition. In *Proc. CVPR98*, pages 32–37. IEEE Computer Society, 1998.

[4] F. Glover and M. Laguna. Tabu search. In C. R. Reeves, editor, *Modern heuristic techniques for combinatorial problems*, pages 70–150. Blackwell Scientific Publications, 1993.

[5] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics - The Approach Based on Influence Functions*. John Wiley & Sons, 1986.

[6] G. E. Hinton, M. Revow, and P. Dayan. Recognizing handwritten digits using mixtures of linear models. In G. Tesauro, D. Touretzky, and T. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7, pages 1015–1022. The MIT Press, 1995.

[7] A. Leonardis and H. Bischof. MDL-Based construction of multiple eigenspaces. In M. Vincze, editor, *Robust Vision for Industrial Applications 1999*, volume 128 of *OCG Schriftenreihe*, pages 79–88. Oldenbourg, 1999.

[8] B. Moghaddam and A. Pentland. Probabilistic visual learning for object detection. In *Int. Conference on Computer Vison ICCV-95*, pages 786–793. IEEE Computer Society, 1995.

[9] H. Murase and S. K. Nayar. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, 14:5–24, 1995.

[10] A. Pentland, B. Moghaddam, and T. Straner. View-based and modular eigenspaces for face recognition. Technical Report 245, MIT Media Laboratory, 1994.

[11] J. Rissanen. A universal prior for the integers and estimation by minimum description length. *Annals of Statistics*, 11(2):416–431, June 1983.

[12] P. J. Rousseuw and A. M. Leroy. *Robust Regression and Outlier Detection*. Wiley, New York, 1987.

[13] M. Stricker and A. Leonardis. ExSel++: A general framework to extract parametric models. In V. Hlavac and R. Sara, editors, *6th CAIP'95*, number 970 in Lecture Notes in Computer Science, pages 90–97, Prague, Czech Republic, September 1995. Springer.

[14] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.