

The Global Dimensionality of Face Space

Penio S. Penev*

Laboratory of Computational Neuroscience
The Rockefeller University
1230 York Avenue, New York, NY 10021
<http://venezia.rockefeller.edu/>
PenevPS@IEEE.org

Lawrence Sirovich

Laboratory for Applied Mathematics
Mount Sinai School of Medicine
One Gustave L. Levy Place, New York, NY 10029
chico@camelot.mssm.edu

Abstract

Low-dimensional representations of sensory signals are key to solving many of the computational problems encountered in high-level vision. Principal Component Analysis (PCA) has been used in the past to derive such compact representations for the object class of human faces. Here, with an interpretation of PCA as a probabilistic model, we employ two objective criteria to study its generalization properties in the context of large frontal-pose face databases. We find that the eigenfaces, the eigenspectrum, and the generalization depend strongly on the ensemble composition and size, with statistics for populations as large as 5500, still not stationary. Further, the assumption of mirror symmetry of the ensemble improves the quality of the results substantially in the low-statistics regime, and is also essential in the high-statistics regime. We employ a perceptual criterion and argue that, even with large statistics, the dimensionality of the PCA subspace necessary for adequate representation of the identity information in relatively tightly cropped faces is in the 400–700 range, and we show that a dimensionality of 200 is inadequate. Finally, we discuss some of the shortcomings of PCA and suggest possible solutions.

1. Introduction

In a typical face-recognition application, images are digitized as a set of pixel intensities on a 2D grid. Faces in them are localized, normalized, and subjected to further processing, such as feature extraction and/or recognition. Key in

that process is an accurate estimate of the *probability density* of the set of faces in the the space of all images.

Necessarily, the parameters of this probability density are estimated from a finite set of training examples. It is paramount to keep the number of these parameters, or the *dimensionality* of the representation, low in order to have a reliable estimate, and avoid *over-learning* [8]. Also, both the speed and accuracy of many high-level algorithms critically depend on the dimensionality of their inputs.

Principal Component Analysis (PCA) [9], or the Karhunen-Loève expansion [11, 13], is a well-established method for the derivation of optimal low-dimensional representations. “Eigenfaces” were initially derived by the application of PCA to ensembles of facial regions [23] and full faces [12]; subsequently, the low-dimensional subspace spanned by the strongest eigenfaces was called “face space” [25]. The hypothesis was made that the Euclidean length of the of projection to face space of the pixel-level difference between two facial images is an adequate distance measure for face discrimination, under broad variations in the imaging conditions [25].

Later, other holistic distance measures for face recognition were studied, based on multiple classes [24, 1, 4], and on two classed [14, 15, 17]; under the general ideas of *Linear/Fisher Discriminant Analysis* [7], which all employ an initial projection to face space. Besides of identity, classification and perception of race and sex [16] has been studied using PCA. Also, the PCA representation has been shown to be robust against known gaps in the data [5], as well as uncorrelated noise [18].

Given the wide applicability of PCA, both as a representation and a probabilistic model, it is somewhat surprising that the questions of the dimensionality of the face space and the size of the statistics necessary for robust estimation of the model parameters are still open. Here we propose a framework for approaching those questions and provide bounds on their answers. We also study the rôle of the mirror symmetry for the quality of the results.

*The major part of this research was made possible by the William O’ Baker Fellowship, so generously extended to, and gratefully accepted by, PSP. He is also indebted to M. J. Feigenbaum for his hospitality and support—scientific and otherwise. The quality of the paper was improved substantially by the numerous suggestions from B. W. Knight. We are also thankful to L. G. Jordanov for useful discussions, and to the three anonymous referees, for their insightful comments.

2. The face space

A properly registered and normalized face¹ will be represented by the image intensity values $\phi(\mathbf{x})$, where $\{\mathbf{x}\}$ is a pixel grid that contains V pixels. An *ensemble* of T faces will be denoted by $\{\phi^t(\mathbf{x})\}_{t \in T}$. Briefly (see, e.g., [22, 17] for details), its PCA representation is given by

$$\phi^t(\mathbf{x}) = \sum_{r=1}^M a_r^t \sigma_r \psi_r(\mathbf{x}) \quad (1)$$

where $M = \min(T, V)$ is the rank of the ensemble, $\{\sigma_r\}$ (in non-increasing order) is the *eigenspectrum* of the two correlation matrices

$$\begin{aligned} R(\mathbf{x}, \mathbf{y}) &\equiv \frac{1}{T} \sum_t \phi^t(\mathbf{x}) \phi^t(\mathbf{y}) = \sum_{r=1}^M \psi_r(\mathbf{x}) \sigma_r^2 \psi_r(\mathbf{y}) \\ C^{tt'} &\equiv \frac{1}{V} \sum_{\mathbf{x}} \phi^t(\mathbf{x}) \phi^{t'}(\mathbf{x}) = \sum_{r=1}^M a_r^t \sigma_r^2 a_r^{t'} \end{aligned} \quad (2)$$

and $\{\psi_r(\mathbf{x})\}$ and $\{a_r^t\}$ are their respective orthonormal eigenvectors.² The pictures $\{\psi_r\}$ are the *eigenfaces* of the ensemble [12].

The *average signal power* of the ensemble is

$$\frac{1}{TV} \sum_{\mathbf{x}, t} |\phi^t(\mathbf{x})|^2 = \text{tr } \mathbf{R} \equiv \text{tr } \mathbf{R}_M \equiv \sum_{r=1}^M \sigma_r^2. \quad (3)$$

PCA is optimal in the sense that, among all N -dimensional subspaces ($N < M$), the subset of eigenfaces $\{\psi_r\}_{r=1}^N$ (2) span the one which captures the most signal power, $\text{tr } \mathbf{R}_N$ [13, 9]. For a given dimensionality N , the coefficients of the PCA *representation* of an arbitrary face, $\phi(\mathbf{x})$ are

$$\sigma_r a_r = \frac{1}{V} \sum_{\mathbf{x}} \psi_r(\mathbf{x}) \phi(\mathbf{x}) \quad (4)$$

while the respective *reconstruction* and *error* are

$$\phi_N^{rec} = \sum_{r=1}^N a_r \sigma_r \psi_r \quad \text{and} \quad \phi_N^{err} = \phi - \phi_N^{rec}. \quad (5)$$

With the standard *multidimensional Gaussian* model for the probability density $\mathcal{P}[\phi]$ [14, 17], the *information content*, or the *log-likelihood*, of the reconstruction (5) is

$$-\log \mathcal{P}[\phi_N^{rec}] \propto \sum_{r=1}^N |a_r|^2. \quad (6)$$

¹For this study, *frontal pose* faces were: registered by semi-automatically pinpointing the centers of the eyes; normalized by an affine transformation that makes the inter-eye line horizontal, centered, and 28 pixels wide; and cropped to a $V = 64 \times 60 = 3840$ grid that includes 17 pixels above the eyes—a somewhat tighter cropping than is typically used in the literature, chosen to include as little background and hair as possible, but still preserve all facial identity and expression information (cf. Fig. 1).

²When $M = T < V$, the diagonalization of \mathbf{C} is easier; this is called the *snapshot method* [22].

3. Dimensionality and generalization

What is the dimensionality of face space, how many terms should be kept in (5), what value of N captures “enough” information about a face?

Eigenfaces. One possibility to answer those questions is to look at the eigenfaces of the ensemble, keep the ones that are “face-like,” and throw away the “noise-like.” Such an analysis, of Ensemble FERET, which is comprised by the examples without eye glasses from the “development set” of the FERET database [19] ($T = 1038$), is shown in Fig. 1. Notably, the initial regime captures sources of variability that are intuitively understandable: ψ_1 is the “typical face;” ψ_2 , the figure-ground separation; ψ_3 through ψ_8 are various lighting-related modes. The next regime, up to about ψ_{50} , is obviously related to the variability in the facial composition, and the rest can be attributed to noise. It is tempting to conclude that $N \in [50, 100]$ would be reasonable truncations in (5). Notably though, the remnants of facial structure are strong in ψ_{100} , decay very slowly, and persist up to the end of the eigenface sequence.

Perceptual quality. The idea that classification can be performed directly on the first N PCA coefficients (4) relied on the hypothesis that the reconstruction (5) is “adequate” [25]. In order to verify that, one can look at the *perceptual quality* of the reconstructions, ϕ_N^{rec} (5), and residual errors, ϕ_N^{err} , as well as their *signal to noise ratios*.

Evidently in Fig. 2, ϕ_{100}^{rec} , with SNR ≈ 6.5 octaves, is perceptually inadequate—almost all identity information is contained in the error. This is also true for ϕ_{200}^{rec} , with SNR ≈ 7 , although there is a noticeable improvement there. Next, there is a regime of transfer of identity and expression information from the error to the reconstruction, with most of the identity already captured by the reconstruction with $N = 500$, SNR ≈ 7.5 ; and only tiny, but significant details develop all the way to SNR ≈ 8 , notably the eyebrow midpoint and tops, the width of the nose, and the structures around the eyes. Although, supposedly for most purposes, $N \approx 500$ would be a reasonable reconstruction of this example in the context of Ensemble FERET, even at $N = 1000$, the details of the expression and the eyebrows are not fully developed.

Relative entropic cost. The finding that ϕ_{100}^{rec} is inadequate, in the sense that it captures almost none of the identity information, is somewhat surprising. On one hand, the modes in the regime $N \in [100, 500]$ are needed for adequate reconstruction; on the other, evidently in Fig. 1, they are global, non-intuitive, noisy, with weak facial content.

One possibility for this discrepancy would be that the ensemble does not generalize well—in its context, some re-

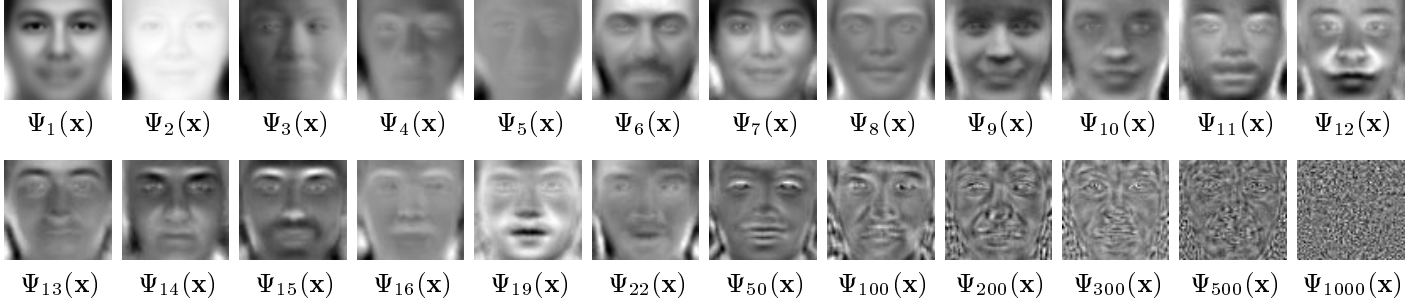


Figure 1. The first 16 eigenfaces of Ensemble FERET ($T = 1038, V = 3840$) and 8 of the rest.

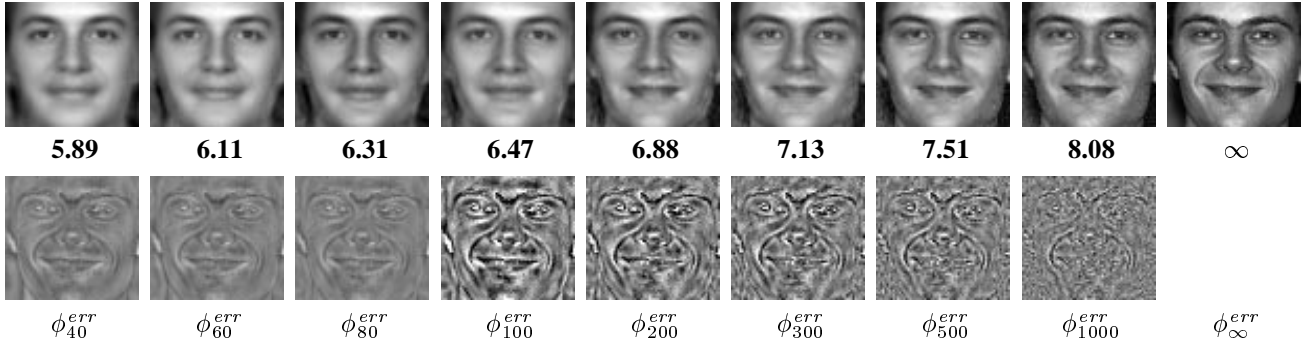


Figure 2. Successive reconstruction in the cross-over regime of Ensemble FERET

The reconstructions ϕ_N^{rec} (5) of Example 1 (an out-of-sample example for all ensembles) are labeled with the logarithm of their *signal to noise ratios* in octaves (= 3 dB), $SNR \equiv \log_2(\|\phi\|^2/\|\phi_N^{err}\|^2)$, and shown on the top row; the respective residual errors, ϕ_N^{err} , are shown on the bottom. For display purposes, the first three errors are amplified $1\times$ and the last— $4\times$.

constructions ϕ_N^{rec} (5) of the out-of-sample example ϕ could be very *improbable* (6). Fig. 3 shows the information content, or the *entropic cost*, and the quality of the reconstruction, parametrically as a function of the dimensionality, N .

Four regimes are evident: a rapid SNR gain up to $N \sim 50$ when $SNR \approx 6$ is achieved at very low entropic cost; a gradual increase of the cost per component until $SNR \approx 6.5$, $N \sim 100$; a plateau, up to $SNR \approx 7.2$, $N \sim 250$; and finally, an “explosion” in the entropic cost.

Although it is obviously “unwise” to keep terms in (5) much more than $N = 250$ in the context of Ensemble FERET, evidently from Fig. 2, it is “necessary;” this supports the conclusion that this ensemble ($T = 1038$, $V = 3840$) does not generalize well in the $N > 250$ regime.

Eigenspectrum. The same regimes in the generalization properties of the ensemble are also evident from the eigenspectrum of the ensemble, shown in Fig. 4. Initially, there is a *power-law* regime, where the ensemble generalizes well, up to $N \approx 100$; then there is a *cross-over* regime, $N \in [100, 300]$, in which some genuine information about the face space is still available; and finally, there is an *exponential-decay* regime, related to noise and artifacts.

4. Stationarity and large statistics

Section 3 provides evidence that, with Ensemble FERET ($T = 1038, V = 3840$), although $N \approx 500$ PCA coefficients are needed in (3) for adequate reconstruction, at most 100 of them are truly genuine, and at least 200, truly artifactual. This observation opens the question how many examples, T , are needed in the ensemble, for the reconstruction to be entirely in a good-generalization regime.

Here we consider Ensemble 2 ($T = 5627$), which additionally contains examples from a large, internally-developed face database. Evidently from Fig. 5, its eigenfaces are cleaner; the “vertical lighting” mode (ψ_5) is stronger and purer; the old “background lighting” modes (ψ_4 and ψ_5 , Fig. 1) have moved to ψ_7 and ψ_8 ; the old facial-feature modes ($\psi_9, \psi_{10}, \psi_{12}, \psi_{19}$, Fig. 1) are cleaner, more localized, and have moved slightly. We will refer to this phenomenon as *mode renumbering*. Also, some novel features have developed ($\psi_9, \psi_{10}, \psi_{14}, \psi_{17}, \psi_{23}, \psi_{56}$). Notably, there is clearly facial information in ψ_{1000} , and the last mode, ψ_{3840} , contains just pixel noise. Obviously, the regime of genuine information extends further than before.

This conclusion is reinforced by the generalization curves in Fig. 3—Ensemble 2 has an almost constant regime

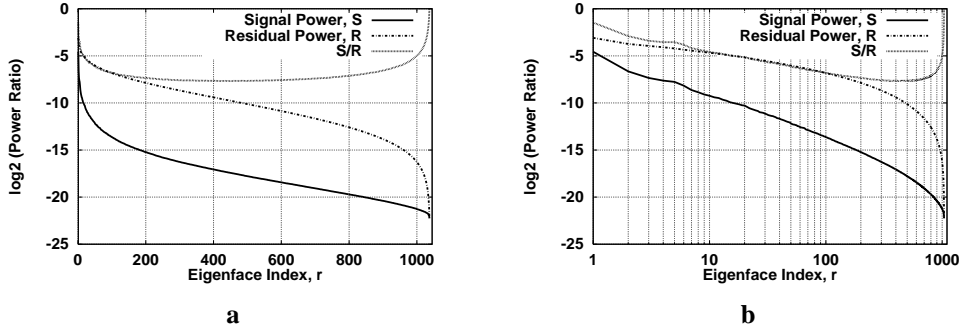


Figure 4. log-linear and log-log Spectra of Ensemble FERET ($T = 1038, V = 3840$)

$\log_2(\sigma_r^2 / \text{tr } \mathbf{R})$ —the ratio of the signal power in the r -th mode, σ_r^2 , and the total signal power, $\text{tr } \mathbf{R}$ (3)—is shown with a solid line; of the residual signal power after truncation (3) and the total signal power, $\log_2((\text{tr } \mathbf{R} - \text{tr } \mathbf{R}_r) / \text{tr } \mathbf{R})$, with a dashed line; of the eigenface signal power and the residual signal power, with a dotted line. When the power spectrum, S , is exponentially decaying, so is its integral—the residual energy, R ; then their ratio, S/R , is a constant. When S is a power law, so is R ; then S/R is again a power law, with a power of -1 ; this is exactly true for the eigenspectrum of the ensemble of natural images, which is not constrained to include only well-framed objects [20]. The spectrum until $r \approx 100$ is in the power-law, and after $r \approx 300$, in the exponential-decay regime, with a cross-over in between.

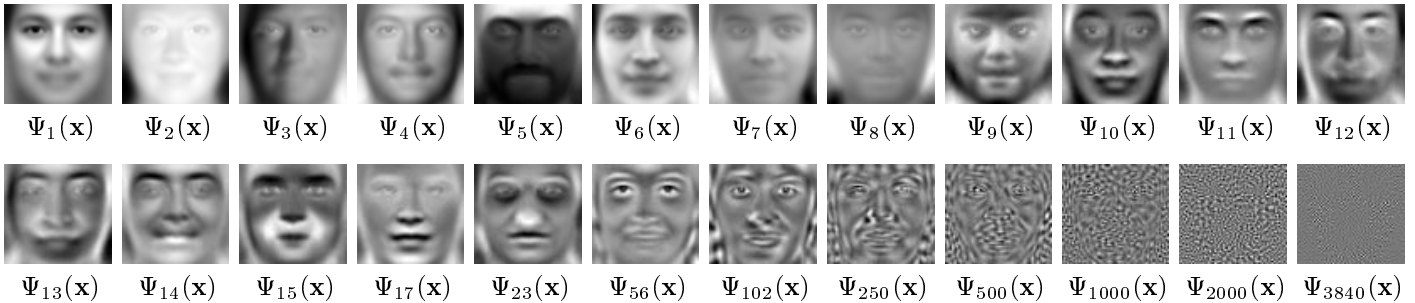


Figure 5. The first 15 eigenfaces of Ensemble 2 ($T = 5627, V = 3840$) and 9 of the rest.

for $N \in [150, 400]$, corresponding to the regime $N \in [100, 250]$ of Ensemble FERET; it correlates with the cross-over regimes of their spectra (cf. Fig. 10). Evidently, larger ensembles can probe deeper into the structure of face space and achieve better SNR with “genuine” modes; also, for a given SNR, less terms are needed in the reconstruction (5) which, additionally, come at a lower entropic cost.

5. Mirror symmetrization

In Section 4 it was shown that large statistics are paramount to getting high-quality results with PCA. Here we explore an additional method for the enrichment of the statistics, which was pioneered in [23]. There is *a priori* knowledge of a global mirror symmetry of the ensemble—if a picture is a face, its mirror image is also a face (possibly not of the same individual). Hence, we construct the *mirror-symmetrization* of an ensemble with T examples—by including also their mirror images—for total statistics with $\hat{T} \equiv 2T$.

Small statistics. A comparison of the reconstruction quality with the symmetrized and non-symmetrized ensembles is shown in Fig. 6. Notably, any finite unsymmetrized ensemble is not exactly symmetric. Also, a typical probe example ϕ is not mirror-symmetric either. Hence, one of the mirror versions of ϕ is more probable in such asymmetric context—the one that is more symmetric in the same direction as the ensemble itself. Intuitively, the entropic cost in the context of the symmetrized ensemble has to fall somewhere between those of the “correctly” and “wrongly” asymmetric ones. Nevertheless, evidently from Fig. 6, the symmetrized reconstructions have higher SNRs after $N = 200$ and lower entropic costs, after $N = 300$, than even the “correctly-”asymmetric ones.

Large statistics. The mirror symmetrization, crucial for ensembles with poor statistics, is also very useful for richer ones, such as Ensemble 2. The regime of fast entropic growth, $\text{SNR} \approx 9$ and beyond, for the non-symmetrized ensembles on Fig. 7a, has transformed to only moderate

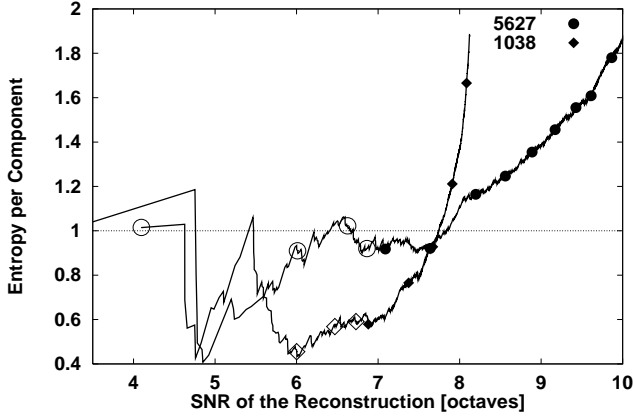


Figure 3. Generalization analysis of the reconstructions with Ensemble FERET ($T = 1038$) and Ensemble 2 ($T = 5627$). The relative entropic cost, $(1/N) \sum_{r=1}^N |a_r|^2$ (6), and signal-to-noise ratio (cf. Fig. 2) are shown as a function of the dimensionality N for the reconstructions of Example 1 (cf. Fig. 2). To illustrate the dependence on N , the curves are marked with open points for $N \in \{1, 50, 100, 150\}$, and with filled, for $\{N \bmod 200 = 0\}$. Notably, the ensemble compositions are not identical, and since no special normalization has been performed, the overall scales of their entropies are slightly different; what can safely be compared is the shape of the curves, as well as the SNR, which is absolute. In contrast, exactly the same examples participate in all three ensembles in Fig. 6, and also their entropies can be compared directly. In addition, the full generalization analysis for Ensemble 2 is shown in Fig. 7a, curve *ensemble 5627*.

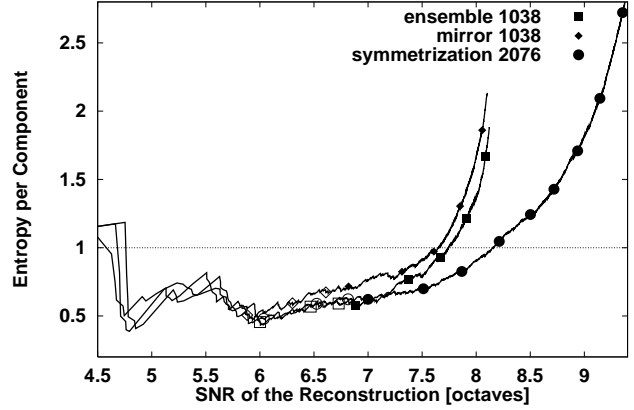


Figure 6. Generalization analysis (cf. Fig. 3) of reconstructions with Ensemble FERET ($T = 1038$) (*ensemble*), its mirror image (*mirror*), and symmetrization ($\tilde{T} = 2076$).

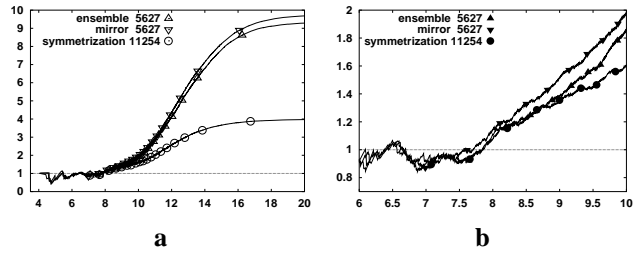


Figure 7. Generalization analysis (cf. Fig. 6) of reconstructions with Ensemble 2 ($T = 5627$) (*ensemble*), its mirror image (*mirror*), and symmetrization ($\tilde{T} = 11254$), shown in full in (a) and in part in (b).

growth for the symmetrized one. Remarkably, the doubling of the statistics—from $T = 5627$ to $\tilde{T} = 11254$ —has improved the generalization properties of the ensemble, which signifies that $T = 5627$ is not yet in the stationarity regime.

The perceptual quality of the successive reconstructions with the symmetrized ensemble ($\tilde{T} = 11254$), shown in Fig. 8, is markedly better than that of Ensemble FERET ($T = 1038$) (cf. Fig. 2). Nevertheless, although eigenfaces in the regime $N \in [400, 700]$ are necessary, they are still not in the plateau-, or the good generalization, regime.

Eigenfaces. The effect of the mirror symmetrization on the eigenfaces of large ensembles is also significant; Fig. 9 shows the eigenfaces of the symmetrized Ensemble 2.

The most striking effect is that now all the modes are, necessarily, either even or odd. The symmetrization has forced several of the leading principal components into intuitively more understandable incarnations. Typical examples are $\psi_2, \psi_5, \psi_8, \psi_{11}$, and many others; there is no *a priori* reason they should not be (a)symmetric, and here they are.

Notable is the liberation of the asymmetric modes, such as $\psi_3, \psi_7, \psi_{12}, \psi_{94}$, and others (not shown); they were previously obscured by the tendency to mix into the symmetric ones which have similar eigenvalues (average energies). We will refer to this phenomenon as *mode mixing*.

Notably, the face content is much more pronounced and cleaned up in the regime $r \in [250, 1000]$, also evident from the lower entropic cost regime in Fig. 7b, and the extended power-law and cross-over regimes of the spectrum Fig. 10.

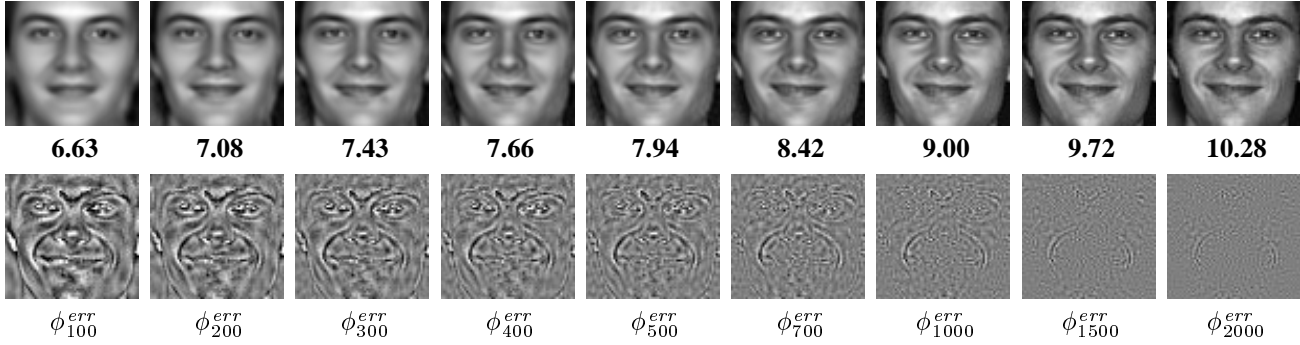


Figure 8. Successive reconstruction (cf. Fig. 2) with the symmetrization of Ensemble 2 (cf. Fig. 7).

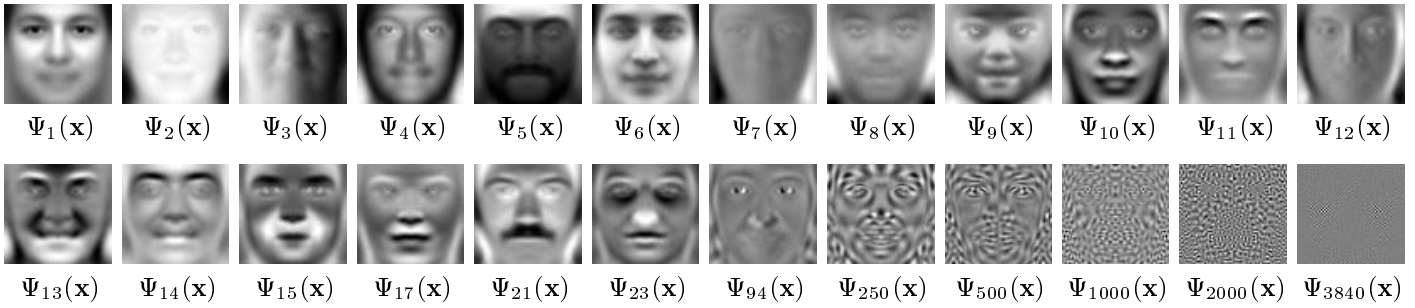


Figure 9. Eigenfaces of the symmetrized Ensemble 2 ($\tilde{T} = 11254, V = 3840$).

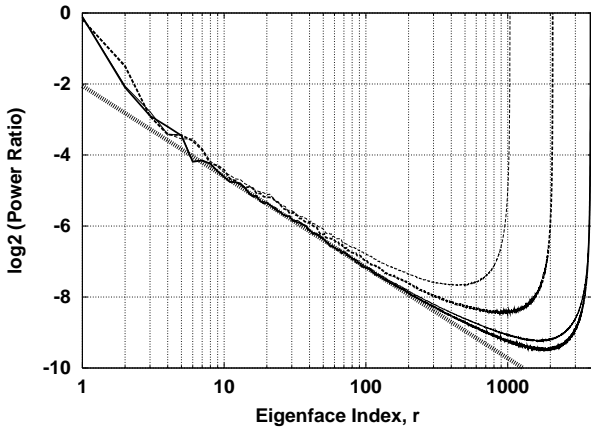


Figure 10. The ratio of the power of an eigenface with index r and the residual signal power, S/R (cf. Fig. 4), is shown for Ensemble FERET (*dashed*) and Ensemble 2 (*solid*) with thin curves, and for their respective symmetrizations, with thick curves. The straight line is a power law with a power of $-0.77 \neq -1$, which signifies that the spectrum, S , is decreasing faster than a power law, but slower than an exponential.

6. Discussion

The PCA basis of face space (2) serves as an efficient, low-dimensional representation for the execution of higher-level algorithms, both of the facial images themselves (5), and of their probabilities (6). Here we have studied the successive regimes in the structure of face space as a function of the dimensionality: they begin with global facial features followed by feature localization, feature mixing, ripples of decreasing scale, and pixel noise at the end (cf. Fig. 9). Three regimes are evident in the eigenspectrum: a putative power law, a cross-over, and an exponential decay (cf. Fig. 10); they correspond to rapid quality increase, plateau, and entropic explosion in the representation of the energy and entropy of facial images (cf. Fig. 7).

The exact shape of the final regime of the spectrum is a known function of T/V , which is a manifestation of the limited statistics [21]; this knowledge can be used to recover the true spectrum of the signal in the cross-over regime through Bayesian estimation [6]. Evidently from Fig. 10, with increasing statistics, the initial regimes of the spectra tend to an universal curve; the departures therefrom signal the end of the genuine regime and the beginning of the transition to the artifactual one. Notably, the universal spectrum of this ensemble of *natural objects* is decaying faster than a power-law, in contrast with that of *natural images* [20].

Evidently from Fig. 8 (cf. Fig. 9), the first $N \approx 200$

eigenfaces capture such global variables as race, sex, lighting, small scale- and pose variations, overall shape and general structure of the face, and general expression. At least twice that many seem to be necessary for such minor, but identity-distinguishing details as the exact shape of the eyebrows, and the detailed structure of the nose and the eyes.

In the light of these findings, it is not surprising that initially encouraging face-recognition results from methods that project to a subspace with too low a dimensionality have resisted successful scaling-up to large databases [17].

Several factors can influence the dimensionality of face space, N , and the size of the statistics necessary for robust estimation, T . They will both increase when additional sources of variability are present: eyeglasses; pose variability; larger number of pixels, V , due both to more liberal cropping and/or higher resolution; and less stringent registration. N and T will decrease with a tighter cropping, preferably non-rectangular [12, 14], and better registration, such as, based on non-linear warping, *e.g.*, [3, 10].

Nevertheless, there are phenomena that cannot be dealt with naturally in the context of PCA. The ambiguity of *mode mixing*, described in Section 5, which is due to eigenfaces with similar second-order cumulants (2), could be resolved on the basis of their fourth-order cumulants, in the context of *Independent Component Analysis* [2].

Another possibility is, while retaining all of the representational power of PCA, to build on top of it an explicit sparse, flexible feature-template representation, such as *Local Feature Analysis* [18, 17]; it reduces dimensionality additionally [17], and has been found to be very suitable for face recognition [19].

References

- [1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces—recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 19(7):711–720, July 1997.
- [2] J.-F. Cardoso. Blind identification of independent components with higher-order statistics. In *Proc. Workshop on Higher-Order Spect. Anal., Vail, Colorado*, pages 157–160, 1989.
- [3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In H. Burkhardt and B. Neumann, editors, *5th European Conf. Computer Vision*, volume 2, pages 484–498. Springer, 1998.
- [4] K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America A*, 14(8):1724–1733, Aug. 1997.
- [5] R. Everson and L. Sirovich. The Karhunen-Loève transform for incomplete data. *J. Opt. Soc. Am. A*, 12(8):1657–1664, 1995.
- [6] R. M. Everson and S. J. Roberts. Inferring the eigenvalues of covariance matrices from limited, noisy data. *IEEE Trans. Sig. Proc.*, 2000. in press.
- [7] R. Fisher. The use of multiple measurements in taxonomic problems. *Ann. Eugenics*, 7(II):179–188, 1936.
- [8] S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias / variance dilemma. *Neural Computation*, 4:1–58, 1992.
- [9] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York Berlin Heidelberg Tokio, 1986.
- [10] M. J. Jones and T. Poggio. Model-based matching by linear combination of prototypes. Technical Report A.I. Memo No. 1583, MIT A.I. Lab, Nov. 1996. available at <ftp://publications.ai.mit.edu/>.
- [11] K. Karhunen. Zur Spektraltheorie Stochastischer. *Prozesse Ann. Acad. Sci. Fennicae*, 37, 1946.
- [12] M. Kirby and L. Sirovich. Application of the Karhunen-Loève procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, January 1990.
- [13] M. Loève. *Probability Theory*. Van Nostrand, Princeton, N.J., 1955.
- [14] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):669–710, July 1997.
- [15] B. Moghaddam, W. Wahid, and A. Pentland. Beyond eigenfaces: Probabilistic matching for face recognition. In *3rd IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, pages 30–35, Nara, Japan, 1998. IEEE.
- [16] A. J. O'Toole, H. Abdi, K. Deffenbacher, and J. Bartlett. Classifying faces by race and sex using an autoassociative memory trained for recognition. In K. Hammond and D. Gentner, editors, *Proc. of the Thirteenth Annual Conf. of the Cognitive Science Society*, pages 847–851. 1991.
- [17] P. S. Penev. *Local Feature Analysis: A Statistical Theory for Information Representation and Transmission*. PhD thesis, The Rockefeller University, New York, NY, May 1998. also available at <http://venezia.rockefeller.edu/penev/thesis/>.
- [18] P. S. Penev and J. J. Atick. Local Feature Analysis: A general statistical theory for object representation. *Network: Computation in Neural Systems*, 7(3):477–500, 1996.
- [19] P. J. Phillips, P. Rauss, and S. Der. FERET (face-recognition technology) recognition algorithm development and test report. Technical Report 995, U.S. Army Research Laboratory, 1996.
- [20] D. L. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Phys. Rev. Lett.*, 73(6):814–817, Aug 1994.
- [21] J. W. Silverstein. Eigenvalues and eigenvectors of large-dimensional sample covariance matrices. *Contemporary Mathematics*, 50:153–159, 1986.
- [22] L. Sirovich. Turbulence and the dynamics of coherent structures. *Q. Appl. Math.*, XLV:561–590, 1987.
- [23] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A*, 4:519–524, 1987.
- [24] D. L. Swets and J. J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8):831–836, August 1996.
- [25] M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.