

Image Analysis, A Tutorial

Part 1. Basics

Lawrence Sirovich

August 11, 2003

1 Introduction

Generally, in these notes, we will be interested in image data. (However, the underlying analysis and mathematical apparatus encompasses more general frameworks.) As an illustration, and one which we will return to for inspiration, consider an ensemble of pictures of human faces. Six exemplars are shown below (of a collection of 280 photos that are located in <http://camelot.mssm.edu/imaging.html>). Symbolically, we can represent the ensemble of such images by

$$f = f(t, \mathbf{x}), \tag{1}$$

where t is the index of the image (this might be the *timestamp* of when the photo was taken), \mathbf{x} the pixel location and f is the *gray level* of image t at pixel location \mathbf{x} . Figure 1 exhibits $f(t, \mathbf{x})$ for $t = 1, \dots, 6$. A word about notation: In what follows, I will be cavalier about variables being either continuous or discrete - virtually everything will be independent of this distinction. This gives us the option of thinking of a continuous or discrete model, whichever is more convenient.

Each image in figure 1 contains roughly $128 \times 128 = O(10^4)$ pixels, and therefore each face in the ensemble is a point in a vector space of $O(10^4)$ dimensions - the point location tells us, through the function,(1), what the gray level at each pixel location in this vector space. One says that the *state space* of the ensemble of faces has $O(10^4)$ dimensions.

The issue that I want to take up is whether there is a *best* space in which to represent the faces. What is meant by best will emerge below, but for the moment we take this to mean that we seek ways of identifying or locating any member of the *face space*, or for that matter any face by means more concise than individual pixel specification. This challenge is referred to as the *Rogue's Gallery Problem*.¹

¹This problem was introduced and solved in 1987 (Sirovich and Kirby 1987), see also (Kirby and Sirovich 1990), and was originally intended as a *hobby horse* problem for more complicated problems in turbulence. Subsequent face recognition technology has its origin in these two papers.

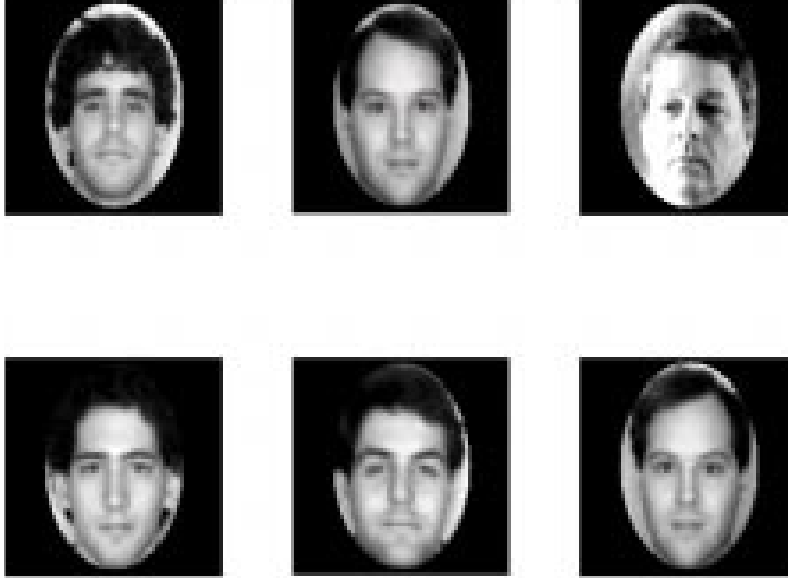


Figure 1: Six chosen from a collection of 280 faces. Note that the sixth face is the mirror image, in the vertical midline, of the second.

2 Mathematical Preliminaries

There are diverse ways of viewing the problem and arriving at its solution, and in the spirit that "more is better" I will indicate some of these. We begin with the observation that the state space description treats each pixel independently, and thus is not sensitive to the fact that adjacent pixels are likely to be correlated. We can imagine that a more efficient description will be attained if globally smooth fitting functions are used, instead of specifying individual pixels, since such imagefunctions carry implied correlations.

To start, suppose we have T images in (1), *i.e.*, $t = 1, \dots, T$. Then we look for an image $\phi(\mathbf{x})$ of unit length

$$1 = \|\phi\|_{\mathbf{x}}^2 = \sum_x (\phi(\mathbf{x}))^2 = (\phi, \phi)_{\mathbf{x}} \quad (2)$$

which gives a *best fit* to the ensemble (1). By best fit we shall mean that the criterion function

$$C(\phi) = \langle (\phi, f)_{\mathbf{x}}^2 \rangle_t = \frac{1}{T} \sum_t (\phi, f)_{\mathbf{x}}^2 \quad (3)$$

is maximized, *i.e.* ϕ 's projection on the images, is on average a maximum. The normalization condition (2) is taken care of by a Lagrange multiplier, denoted by λ , and we have the usual condition for a maximum,

$$\frac{\delta}{\delta\phi}\{C(\phi) - \lambda(\phi, \phi)_{\mathbf{x}}\} = 0. \quad (4)$$

Namely, that the first variation vanishes. To implement condition (4) suppose the solution to (4) is denoted by ϕ^0 then for arbitrary $\varphi(\mathbf{x})$

$$\phi(x) = \phi^0(\mathbf{x}) + \epsilon\varphi(\mathbf{x}), \quad (5)$$

is a candidate trial function. In such terms condition (4) can be recast as

$$\frac{d}{d\epsilon}\{C(\phi^0 + \epsilon\varphi) - \lambda\|\phi^0 + \epsilon\varphi\|^2\}|_{\epsilon=0} = 0. \quad (6)$$

Observe

$$C(\phi^0 + \epsilon\varphi) = \frac{1}{T} \sum_t \left(\int (\phi^0(\mathbf{x}) + \epsilon\varphi(\mathbf{x}))f(t, \mathbf{x})d\mathbf{x} \right) \left(\int (\phi^0(\mathbf{y}) + \epsilon\varphi(\mathbf{y}))f(t, \mathbf{y})d\mathbf{y} \right) \quad (7)$$

and

$$\|\phi^0 + \epsilon\varphi\|^2 = \int (\phi^0(\mathbf{x}) + \epsilon\varphi(\mathbf{x}))^2 d\mathbf{x}. \quad (8)$$

(In keeping with our policy of blurring the distinction between discrete and continuous variables we informally treat summation and integration as the same operation $\int d\mathbf{x} \Leftrightarrow \sum_{\mathbf{x}}$.) Then if we carry out the differentiation indicated in (6), and then we set $\epsilon = 0$, we get

$$\sum_{\mathbf{x}} \varphi(\mathbf{x}) \left\{ \sum_{\mathbf{y}} \frac{1}{T} \sum_t f(t, \mathbf{x})f(t, \mathbf{y})\phi^0(\mathbf{y})d\mathbf{y} - \lambda\phi^0(\mathbf{x}) \right\} = 0 \quad (9)$$

Since $\varphi(\mathbf{x})$ is arbitrary this implies that the solution is given by

$$K\phi = \sum_{\mathbf{y}} K(\mathbf{x}, \mathbf{y})\phi^0(\mathbf{y}) = \lambda\phi^0(\mathbf{x}) \quad (10)$$

where

$$K(\mathbf{x}, \mathbf{y}) = \frac{1}{T} \sum_t f(t, \mathbf{x})f(t, \mathbf{y}) = K(\mathbf{y}, \mathbf{x}), \quad (11)$$

the spatial covariance of the ensemble, is symmetric. (It is left as an exercise to show that arbitrary $\varphi(x)$ implies (10).)

There are several things to note about (10). First it is an eigenfunction problem, and from this we can expect many eigenvalues and eigenfunctions. In

fact since K is symmetric, (11), we know that a complete orthonormal set is generated from (10). If we take the inner product of (10) with $\phi^0(x)$ we obtain

$$\lambda = (\phi^0, K\phi^0)_{\mathbf{x}} = \frac{1}{T} \sum_t (\phi^0, f)^2. \quad (12)$$

Therefore $\lambda \geq 0$, and obviously, since (3) is to be maximized, we want the eigenfunction ϕ^0 correspond to the largest eigenvalue.

To proceed to the next step, we might seek a function orthogonal to ϕ^0 , which satisfies (13) and maximizes C , and so on. It should come as no surprise that all such functions will satisfy (10). This is not an appealing approach, and also leaves open the question of whether we might do better with some finite sum instead of the successive approach just indicated. With this in mind suppose for any integer N , $\{\varphi_n(\mathbf{x})\}, n = 1, \dots, N$ is a collection of, as yet unknown, functions for the description of the faces. Take these to be orthonormal,

$$(\varphi_n, \varphi_m)_{\mathbf{x}} = \sum_{\mathbf{x}} \varphi_n(\mathbf{x})\varphi_m(\mathbf{x}) = \delta_{nm}. \quad (13)$$

To pose the problem we ask for what collection of orthonormal functions, $\{\varphi_n(\mathbf{x})\}$, is there a best fit to $f(t, \mathbf{x})$ in the sense that we can choose constants $\{\alpha_n(t)\}$ so that

$$\langle \|f(t, \mathbf{x}) - \sum_{n=1}^N \alpha_n(t)\varphi_n(\mathbf{x})\|_{\mathbf{x}}^2 \rangle_t \quad (14)$$

is a minimum for any fixed integer N . Clearly if we can find such sets $\{\varphi_n(\mathbf{x})\}$ and $\{\alpha_n\}$ we have obtained a best set in the sense that for any N we commit the smallest error, ϵ_N , in writing $f(t, \mathbf{x})$ as

$$f(t, \mathbf{x}) = \sum_{n=1}^N \alpha_n(t)\varphi_n(\mathbf{x}) + \epsilon_N(t, \mathbf{x}). \quad (15)$$

As a first step in solving the posed problem, we first observe that the projection of f onto the space spanned by $\{\varphi_n\}, n = 1, \dots, N$ is given by

$$P_N f = \sum_{n=1}^N (\varphi_n, f)\varphi_n. \quad (16)$$

Next, it is straightforward that the quantity that appears under the average in (14) satisfies the Pythagorean relation

$$\begin{aligned} \|f - \sum_{n=1}^N \alpha_n \varphi_n\|_{\mathbf{x}}^2 &= \|f - P_N f + P_N f - \sum_{n=1}^N \alpha_n \varphi_n\|_{\mathbf{x}}^2 \\ &= \|f - P_N f\|_{\mathbf{x}}^2 + \|P_N f - \sum_{n=1}^N \alpha_n \varphi_n\|_{\mathbf{x}}^2, \end{aligned} \quad (17)$$

and we leave this as an *Exercise*. The first *leg of the triangle* in the second form of (17) is independent of the choice of the $\{\alpha_n\}$ and depends only on f . From this it follows that for any orthonormal set $\{\varphi_n\}$ the best choice of the $\{\alpha_n\}$ is one for which the second term of (17) vanishes, and hence such that

$$\alpha_n = (\varphi_n, f) \quad (18)$$

In other words the best fit to f is $P_N f$, the projection of f onto the space spanned by $\{\varphi_n\}$, a fact which is surely intuitive.

In the Appendix to this section it is shown that the sought after set $\{\varphi_n\}$, $n = 1, \dots, N$ is given by the first N eigenfunctions of $K(\mathbf{x}, \mathbf{y})$,

$$K\varphi_n = \lambda_n \varphi_n, \quad (19)$$

under the ordering convention

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \quad (20)$$

Under standard theorems of linear algebra (or analysis) since K is symmetric, (11), we are guaranteed that the $\{\varphi_n\}$ exist and can be taken as orthonormal.

Still Another Approach

Suppose $\{\varphi_n(\mathbf{x})\}$ are an unspecified complete set of orthonormal functions (we use the symbol φ to denote these for good reason) and $\{a_n(t)\}$ a set of orthonormal functions in t -space,

$$(a_n, a_m)_t = \sum_{t=1}^T a_n(t)a_m(t) = \delta_{mn}. \quad (21)$$

Next we require that a_n and φ_n be such that

$$f(t, \mathbf{x}) = \sum_n \mu_n a_n(t) \varphi_n(\mathbf{x}) \quad (22)$$

where $\{\mu_n\}$ represent a still to be determined set of constants. The reader should note that such an expansion, *i.e.* a sum of products of orthogonal sets of functions, generally cannot be assumed to exist a priori, and might even be regarded as remarkable. If (21) and (13) are applied to (22) we obtain

$$\mu_n a_n = (\varphi_n, f)_{\mathbf{x}} \quad (23)$$

and

$$\mu_n \varphi_n = (a_n, f)_t. \quad (24)$$

If we multiply (24) by μ_n and back substitute (23) into it we obtain

$$\mu_n^2 \varphi_n = ((\varphi_n, f)_{\mathbf{y}}, f(t, \mathbf{x}))_t. \quad (25)$$

This is just (19) with $\lambda_n = \mu_n^2/T$. Thus we have shown that the assertion of (22) leads to the same eigenfunction framework as the optimization principle posed earlier.

Numerically, the eigenfunction problem posed by (19) appears to be formidable. Formally, the order of the *matrix* K is equal to the number of pixels in a typical image. Thus for the modest case of 128×128 pixels, the order of K is $O(10^4)$, and diagonalization becomes computationally challenging. Fortunately other considerations can simplify the problem. From (11) we have that

$$K(\mathbf{x}, \mathbf{y}) = \frac{1}{T} \sum_{t=1}^T f(\mathbf{x}, t) f(\mathbf{y}, t). \quad (26)$$

A kernel of an integral equation which is such a sum of products is said to be degenerate and can be reduced considerably. In particular substitution of (26) into (19)

$$\varphi_n(x) = \frac{1}{\lambda_n T} \sum_{t=1}^T (f(t, y), \varphi_n(y)) f(t, x) \quad (27)$$

shows that every eigenfunction of K is an admixture of the T snapshots, $f(t, \mathbf{x})$. It immediately follows from this that the eigenfunction problem reduces to the diagonalization of a $T \times T$ matrix, instead of the $O(10^4)$ matrix K . Thus if $T = 1000$ a nominal amount of images, the eigenfunction problem is reduced by two orders of magnitude. This reduction is known as the *snapshot method* (Sirovich 1987). Instead of developing this we show it next by an essentially equivalent procedure.

In place of (25) we can back substitute (24) into (23) in the above to obtain

$$C a_n = \sum_{s=1}^T C(t, s) a_n(s) = \lambda_n a_n(t) \quad (28)$$

where

$$C = C(t, s) = (f(t, \mathbf{x}), f(s, \mathbf{x}))_{\mathbf{x}}, \quad (29)$$

which is a form proportional of the auto-covariance. Clearly the symmetric non-negative matrix, $C(t, s) = C(s, t)$ generates an orthonormal eigenbasis and is of order no more than T . Once $a_n(t)$ is determined, we can obtain $\varphi_n(\mathbf{x})$ from (24).

Since both (19) and (28) lead to symmetric operators, and from this it can be proven that $\{a_n\}$ and $\{\varphi_n\}$ satisfying (21) and (13) exist and hence the assertion that we can expand in the form (22) has been demonstrated. We note that only one of the two eigenfunction problems (28) or (19) need be solved. For once one set of eigenfunctions is determined the complementary set is determined from (23) or (24), whichever is appropriate.

Appendix

The goal of the optimization (3) is to achieve a minimal error in the expansion of $f(t, \mathbf{x})$, i.e., we wish to force the error

$$\langle \|f(t, \mathbf{x}) - \sum_{n=1}^N \alpha_n(t) \phi'_n(\mathbf{x})\|_{\mathbf{x}}^2 \rangle_t \quad (30)$$

to be a minimum. Within the framework of an arbitrary orthonormal set, $\{\phi'_n\}$, $(\phi'_n, \phi'_m)_{\mathbf{x}} = \delta_{nm}$, as shown in section 2, this is achieved by determining the coefficients, $\alpha_n(t)$, through projection,

$$\alpha_n = (\phi'_n, f)_{\mathbf{x}}. \quad (31)$$

The minimization of (30) is therefore equivalent to maximizing,

$$\| \sum_{n=1}^N \alpha_n(t) \phi'_n(\mathbf{x}) \|_{\mathbf{x}, t}^2 = \sum_t \left(\sum_{n=1}^N \alpha_n^2 \right), \quad (32)$$

subject to (31) and

$$\| \phi'_n \|_{\mathbf{x}}^2 = 1; \text{ for all } n \quad (33)$$

This can be put into the form of a variational problem with the criterion function,

$$\mathcal{F} = \sum_t \left(\sum_{n=1}^N \alpha_n^2 + \sigma_n(t) (\alpha_n - (\phi'_n, f)_{\mathbf{x}}) \right) - \sum_n \lambda_n (\| \phi'_n \|_{\mathbf{x}}^2 - 1), \quad (34)$$

with the Lagrange multipliers, $\sigma_n(t)$ and λ_n . The α_n -variation, $\delta\mathcal{F}/\delta\alpha_k = 0$, yields

$$2\alpha_k + \sigma_k = 0, \quad (35)$$

and the ϕ'_k -variation, $\delta\mathcal{F}/\delta\phi'_k = 0$, yields

$$-(\sigma_k, f)_t = 2\lambda_k \phi'_k. \quad (36)$$

Thus if (35) is substituted into (36) we get

$$(\alpha_k, f)_t = \lambda_k \psi'_k. \quad (37)$$

If the identifications $\alpha_k/\sqrt{\lambda_k} \leftrightarrow a_k$ and $\sqrt{\lambda_k} = \mu_k$ are made, then (37) and (31) are identical to (23) and (24), thus establishing our assertion. It is interesting to note that the orthogonality of the ϕ'_n does not appear in (14). It results as a consequence of the fact that $\{\phi'_n\}$ (or $\{\phi_n\}$) are the eigenfunctions of a symmetric operator, (11).

3 Rogues Gallery Problem

We illustrate the procedure by returning to the problem introduced earlier. An ensemble of faces can be found at <http://camelot.mssm.edu/imaging.html> (A description of these as well as instructions for accessing the database is available on the website.) These images have been prepared so that eye location has been put in a fixed position, lighting has been normalized and so forth (Sirovich and Kirby 1987), (Kirby and Sirovich 1990) . As is clear from Figure 1 the 128×128 pixels furnish an adequately resolved image of a face.

The number of faces in the ensemble is 280, but this number is in reality double an original set of 140 faces. To explain the doubling, we denote the original set of faces by

$$f(t, \mathbf{x}) = f_n(x, y), n = 1, \dots, 140. \quad (38)$$

Here y denotes a vertical variable and x denotes the horizontal variable measured from the midline of the face (The perpendicular bisector of the line segment between the eyes.) The original ensemble has been doubled by including $f_n(-x, y)$, the *mirror image* of each face. As noted in the caption the sixth face of Figure 1 is the mirror image of the second. Obviously each image generated in this way is an admissible face, and one might believe that in a large enough population of faces, something close to a mirror image of any face might eventually be realized. Aside from doubling our original ensemble this maneuver has the interesting consequence that it makes every eigenfunction either even in the midline

$$\varphi_n(x, y) = \varphi_n(-x, y) \quad (39)$$

or odd in the midline

$$\varphi_n(-x, y) = -\varphi_n(x, y). \quad (40)$$

And even though the ensemble size has been doubled to 280 it can be seen that we need only have to diagonalize two separate order 140 matrices.

Exercise: For the above described ensemble in which for each face the mirror image face also appears, prove that the eigenfunctions are either odd or even in the mid-line.

(Hint:

$$\begin{aligned} K &= \sum_n f_n(x, y) f_n(x', y') \\ &= \frac{1}{2} \sum_n (f_n(x, y) + f_n(-x, y))(f_n(x', y') + f_n(-x', y')) \\ &\quad + \frac{1}{2} \sum_n (f_n(x, y) - f_n(-x, y))(f_n(x', y') - f_n(-x', y')) \\ &= K^+ + K^- \end{aligned} \quad (41)$$

Any eigenfunction of K can be found from $K^\pm \varphi^\pm = \lambda \varphi^\pm$.

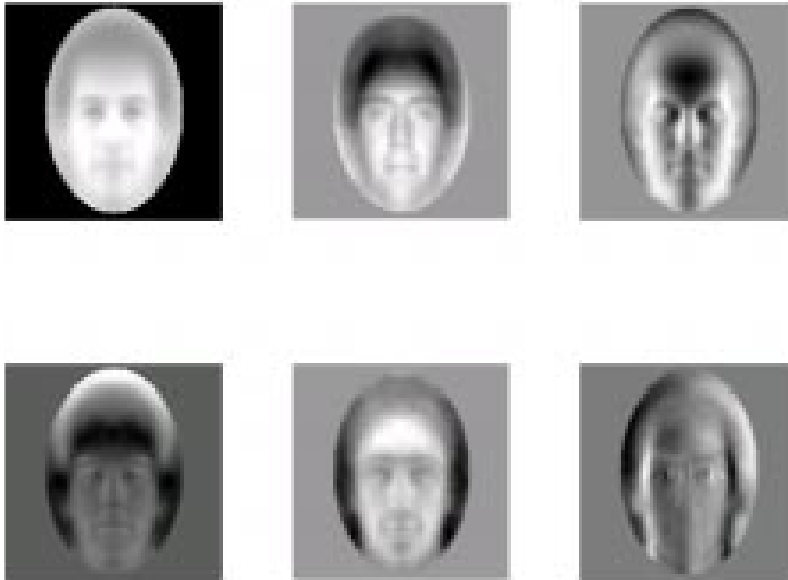


Figure 2: The first 6 eigenfaces arranged in descending order of eigenvalues, λ_n , for the ensemble specified in Figure 1.

All eigenfaces are arranged in descending order of λ_n , and the first six eigenfunctions are shown in Figure 2. We observe that the first five are even, and only at the sixth eigenfunction does oddness appear.

Of interest is the range of values of the eigenvalues. It is clear from the expansion (22) that λ_n measures the degree of importance of the corresponding a_n . Another way of seeing this comes from the observation that

$$\lambda_n = T \langle \|(\varphi_n, f)_{\mathbf{x}}\|^2 \rangle_t, \quad (42)$$

so that λ_n is a clear measure of the average *energy* of the corresponding eigenfunction.

It is natural to associate the eigenfunctions, at the end of the range with *noise*, and in figure 3 we show the last three eigenfunctions, all of which are odd in the mid-line. The fact that these eigenfunctions have remnants of a face (look for example at the teeth) shows the shortcomings of dealing with the relatively small population under consideration.

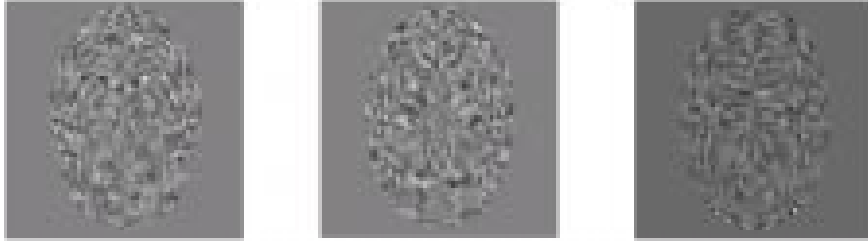


Figure 3: The last 3 eigenfaces arranged in descending order of eigenvalues, λ_n , for the ensemble specified in Figure 1.

In figure 4 we plot λ_n versus n for the Rogues Gallery problem. It might reasonably be assumed that noisy eigenfunctions are not important. To judge how many such *negligible* eigenfunctions are present we run a straight line through the log of the *noisy* eigenvalues and note that it departs from the plot at an index of $n \approx 100$. Thus each face of the ensemble should be well represented by a sum truncated at $N < 100$,

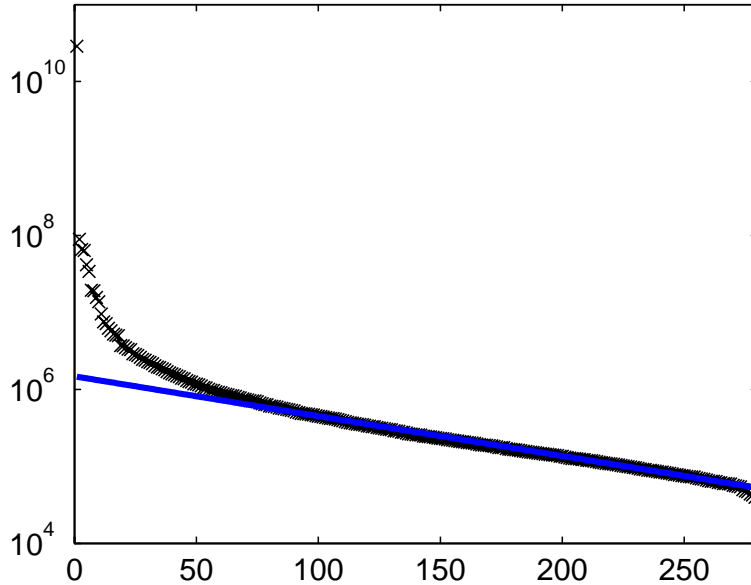


Figure 4: The eigenvalues as a function of their index, for the ensemble specified in Figure 1.

$$f(t, \mathbf{x}) \approx \sum_{n=1}^{100} \alpha_n(t) \varphi_n(t). \quad (43)$$

In figure 5 we show a typical approximate representation of an in-ensemble face. This approximation may be given somewhat precise form. Let us consider the relative error in taking only N terms,

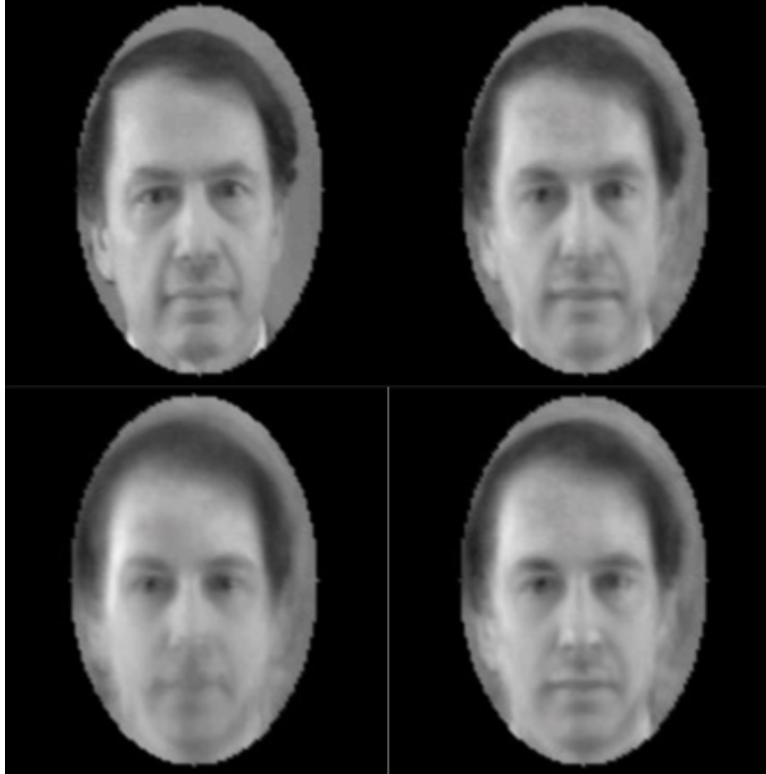


Figure 5: A typical in-ensemble face appears in the upper left. Its approximation by 100, 66 and 33 eigenfunctions appear in clockwise order.

$$\epsilon_N = \frac{\|f - \sum_{n=1}^N \alpha_n(t) \varphi_n(\mathbf{x})\|_{\mathbf{x}}^2}{\|f(t, \mathbf{x})\|_{\mathbf{x}}^2} = \frac{\|\sum_{n=N+1}^T \alpha_n(t) \varphi_n(\mathbf{x})\|_{\mathbf{x}}^2}{\|f(t, \mathbf{x})\|_{\mathbf{x}}^2} \quad (44)$$

where T , as before, is the total number in the ensemble. As a plausible hypothesis we assume that each image has the same power

$$\|f(t, \mathbf{x})\|_{\mathbf{x}}^2 \approx \frac{1}{T} \sum_t \|f(t, \mathbf{x})\|_{\mathbf{x}}^2 = \frac{1}{T} \sum_{n=1}^T \lambda_n \quad (45)$$

and therefore

$$\langle \epsilon_N \rangle_t \approx \frac{\sum_{n=1}^T \lambda_n}{\sum_{n=1}^{N+1} \lambda_n}. \quad (46)$$

For the Rogues Gallery problem, with $T = 280$ and $N = 100$, we get

$$\sqrt{\langle \epsilon_N \rangle_t} \approx .03 \quad (47)$$

so that there is an average 3% error in signal.

The ensemble of 280 faces is a relatively modest size population and the resulting eigenfunctions, $\{\varphi_n\}$, cannot be regarded as being robust. We can nevertheless try these to fit an out-of-ensemble face. The result is shown in figure 6. Thus while the fit, unlike the in-ensemble faces, is equivocal at $N = 100$, it is reasonably good if the full set of eigenfunctions is used.



Figure 6: A non-ensemble face appears at the upper left. Its approximation by all 280 eigenfunctions appears next and by 100 at the right.

Roughly speaking, we see that the empirical eigenfunctions reduces the dimensionality of *face space* from $O(10^4)$ to $O(10^2)$. This reduction is accomplished at a cost. A great many objects can be imaged with an assembly of 100×100 pixels. But only faces can be constructed with the $O(100)$ or so eigenfaces that are significant. This is illustrated in the next figure in which we try to fit a monkey face with the set of eigenfaces.



Figure 7: A monkey face appears at the left. Its approximation by all 280 eigenfunctions appears next and by 100 at the right.

Remarks

It has been seen that for the dataset $f(t, \mathbf{x})$ of T images and P pixels, that there are at most T eigenvalues. (All of which are non-negative.) However if we solve in pixel space there would appear to be a far greater number of eigenvalues, namely P in number.

The apparent discrepancy is easily resolved. For in pixel space each image $f(t, \mathbf{x})$, for t fixed, can be thought of as a vector, and thus the entire database is composed of T vectors. Clearly there must be $P - T$ vectors orthogonal to our database and hence K has a large nullspace, i.e. $\lambda_n = 0$, a degeneracy of at least order $P - T$.

4 Dynamics

An issue that deserves some more discussion is the role of the *time-like* index t . As a little reflection reveals, if the sequences of facial images is shuffled, the eigenpictures $\varphi_n(\mathbf{x})$ remain the same. This invariance of φ_n under shuffling is of course not true for the $a_n(t)$, each of which in fact undergoes a predictable shuffled *time course* under the transformation. For the Rogue's Gallery Problem the order in which we acquire pictures is of no consequence, and hence the coefficients, $\{\alpha_n\}$, are of dubious value. However, in many applications the time course is important since it may reflect the dynamics underlying the phenomenon. In such cases the dynamical evolution of the *model coefficients* $\{\alpha_n(t)\}$ takes an importance equal to the *modes* $\{\varphi_n\}$ themselves.

To illustrate situations in which dynamics is important consider Figure 8 which shows a succession of images of a cross-section of the human heart. A heartbeat takes place roughly every second, and for the transesophageal records shown 30 frames/sec are captured. For simplicity every other image, fifteen in

all, are shown. Therefore in the short span of a minute and a half about 2500 frames are accumulated.

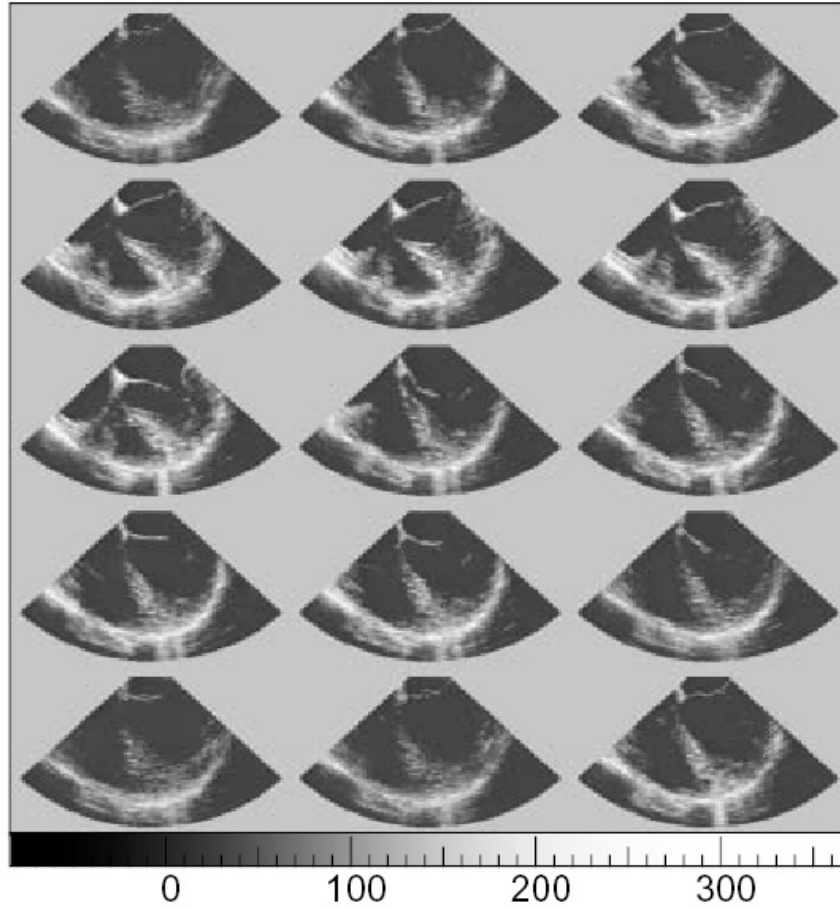


Figure 8: Echo-cardiogram snapshots. Fifteen ECG snapshots spanning approximately 1 second (i.e., about a single beat). Time runs along rows from top to bottom. The lunate region at the left represents the right ventricle. It is separated from the right atrium, which is just above it by the tri-cuspid valve. Similarly, to its right we see the left ventricle and above it the left atrium, with the mitral valve connecting the two.

The eigenfunction analysis of this collection of frames produces the spectrum of eigenvalues shown in Figure 9

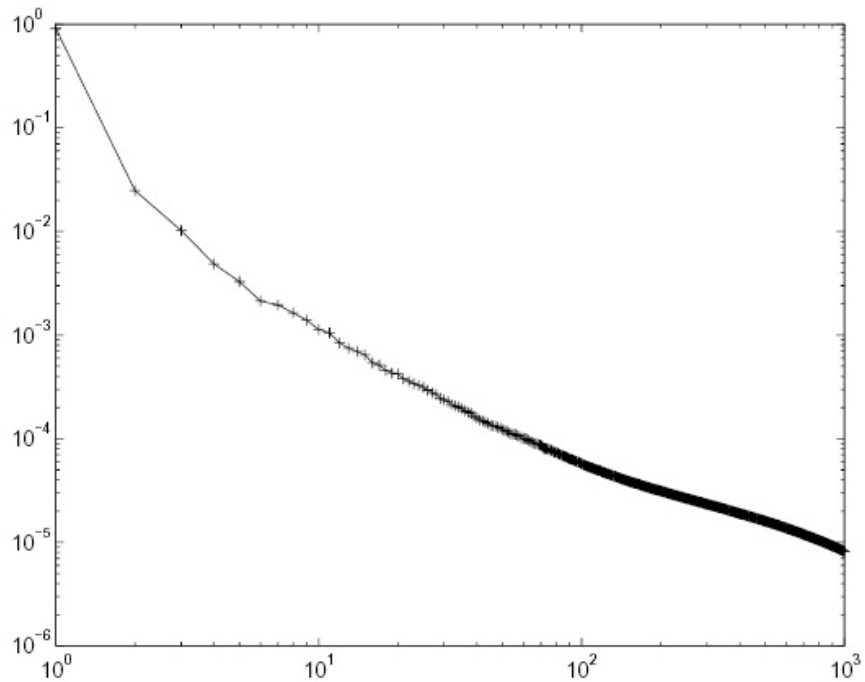


Figure 9: Eigenvalues, λ_n , corresponding to the eigenfunctions. The index of the eigenvalues is given as the abscissa and the eigenvalues are normalized so that $\sum_n \lambda_n = 1$.

and from the figure we can conclude that only 10-20 eigenfunctions are significant. The first six eigenfunctions are next shown in Figure 10. Some observations about the features are indicated in the legend, and no further comment will be given here.

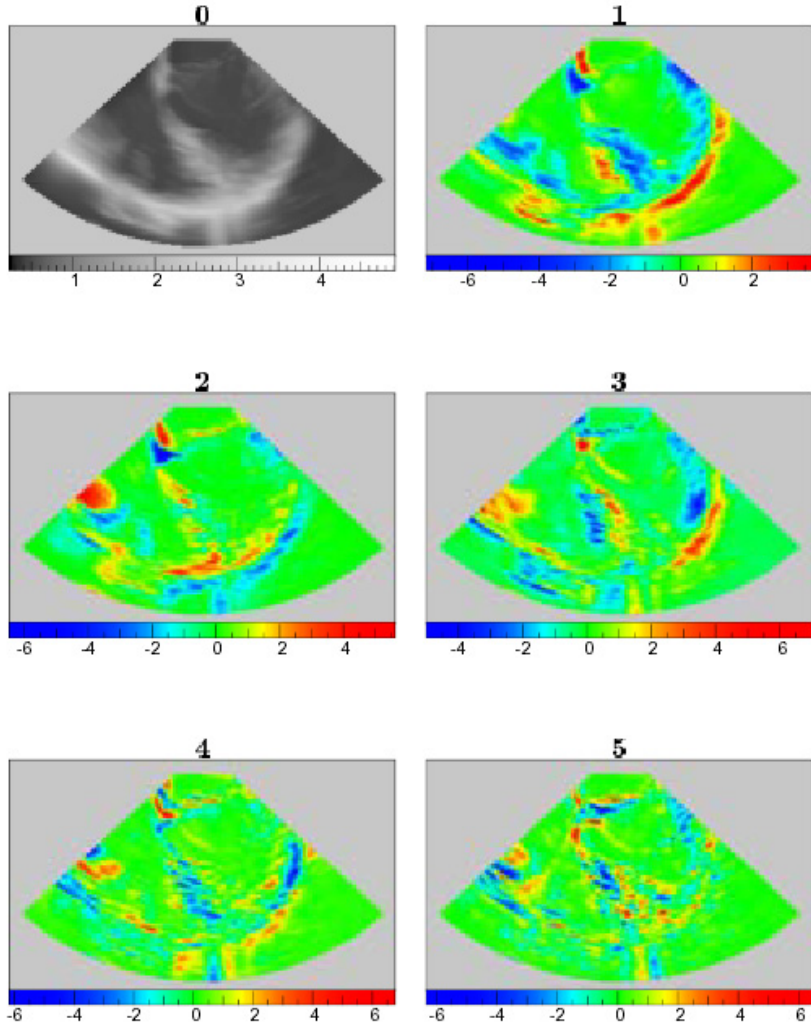


Figure 10: Principal six eigenfunctions, $\psi_n(\mathbf{x})$ from an ECG record.

Of principal interest in this section are the projections of the eigenfunction on the data. If we denote the full record of echo-cardiogram records by $f(t, x)$ then

$$\alpha_n(t) = (f(t, x), \psi_n)_x \quad (48)$$

records of the time histories of the coefficients $\{\alpha_n\}$. Figure 11 displays the time histories of the coefficients corresponding to the six eigenfunctions depicted in

Figure 10.

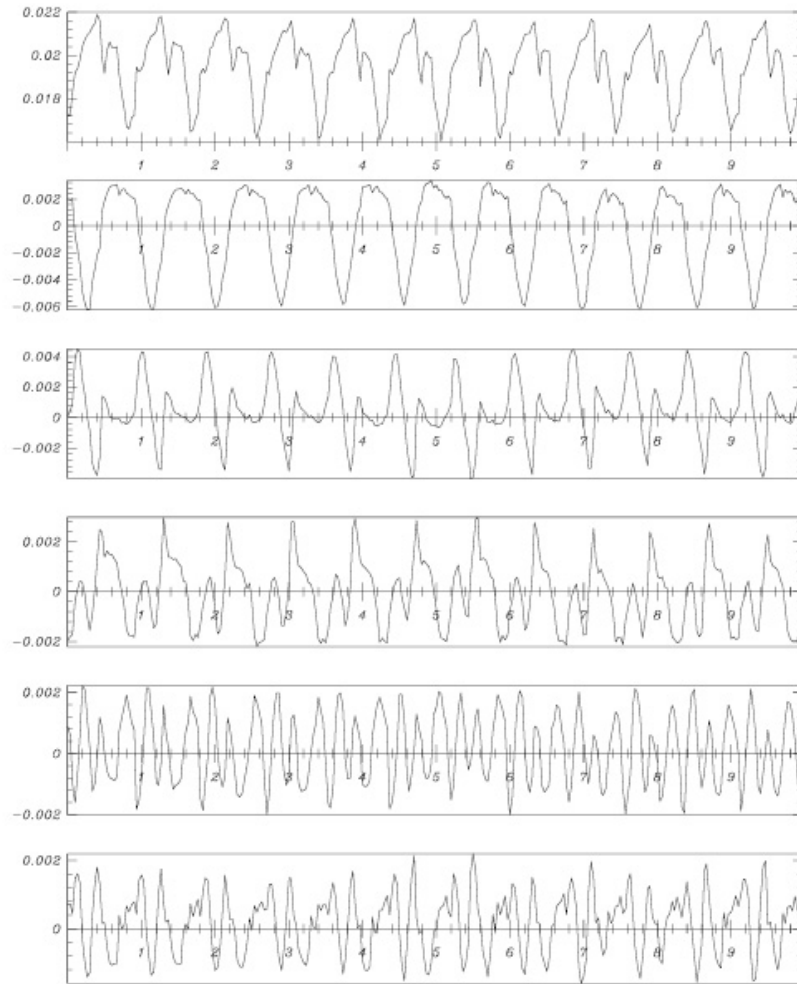


Figure 11: Modal coefficients, $a_n(t)$, corresponding to the principal 6 eigenfunctions. The first 10 seconds of an 85 second record are shown

Note that the time histories, $a_n(t)$, show a resemblance to EKG records normally recorded in clinical heart studies. However, these two types of records are fundamentally different. EKGs, by their very nature are temporal records of electrical activity inherent in heart activity. By contrast, the records of Figure

11 depict the time histories of modes that describe the mechanical activity of the heart. EKG records and those shown in Figure 11 are complementary in content.

5 An Optical Imaging Example

Consider the particular instance of a collection of images acquired in response to gratings drifting at various angles denoted by θ , and denoted by

$$F = F(\theta, \mathbf{x}) \tag{49}$$

where for example

$$F(\theta, \mathbf{x}) = \langle f(t, \theta', \mathbf{x}) \rangle_{t|\theta'=\theta}. \tag{50}$$

Next we form

$$F_{\Delta}(\theta, \mathbf{x}) = F(\theta, \mathbf{x}) - F(\theta + \pi/2, \mathbf{x}). \tag{51}$$

This construction is meant to eliminate background signal, and rests on the experimental observation that a cell tuned for θ is silent for $\theta + \pi/2$. This construction is illustrated in Figure 12.

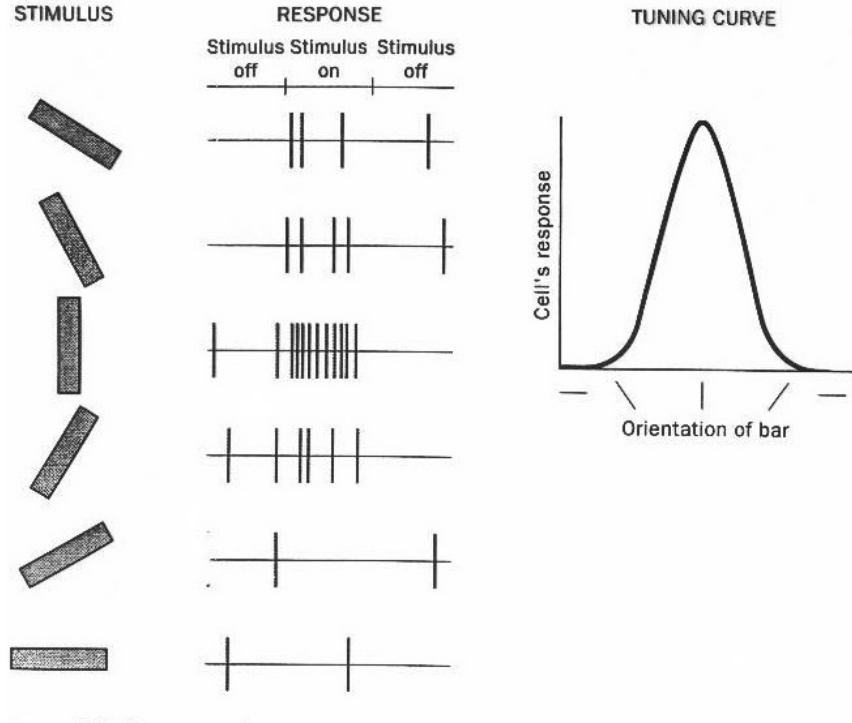


Figure 12: Illustration of a cell tuned for some angle $\theta + \pi/2$.

Next we can expand F_Δ in empirical eigenfunctions

$$F_\Delta(\theta, \mathbf{x}) = \sum_n a_n(\theta) \mu_n \phi_n(\mathbf{x}) \quad (52)$$

with $\{a_n\}$ and $\{\phi_n\}$ are each orthonormal sets. The set $\{a_n\}$ are eigenfunctions of

$$\int C(\theta, \theta') a_n(\theta') d\theta' = \lambda_n a_n(\theta) \quad (53)$$

where

$$C(\theta, \theta') = (F_\Delta(\theta, \mathbf{x}), F_\Delta(\theta', \mathbf{x}))_{\mathbf{x}} \quad (54)$$

which might be regarded as an auto-correlation.

But orientation selectivity can be regarded as *democratic* (Sirovich et al. 1996). Therefore, there is no preferred orientation axis, so that

$$C(\theta, \theta') = C(\theta + \theta_0, \theta' + \theta_0) \quad (55)$$

for any θ_0 . In particular if $\theta_0 = -\theta'$

$$C = C(\theta - \theta'). \quad (56)$$

This therefore implies that

$$a_n = \frac{e^{2\pi i n \theta}}{\sqrt{2\pi}}. \quad (57)$$

Thus

$$F_{\Delta}(\theta, \mathbf{x}) = \sum_n \frac{e^{2\pi i n \theta}}{\sqrt{2\pi}} \mu_n \phi_n(\mathbf{x}). \quad (58)$$

Note that $\mu_0 = 0$. The set

$$F_{\Delta}^O(\theta, \mathbf{x}) = \sum_{n=even} e^{2\pi i n \theta} \mu_n \phi_n(\mathbf{x}) \quad (59)$$

governs orientation, and

$$F_{\Delta}^d(\theta, \mathbf{x}) = \sum_{n=odd} e^{2\pi i n \theta} \mu_n \phi_n(\mathbf{x}) \quad (60)$$

governs direction. Figure 13 shows the result of such a construction.

Orientation Preference Map

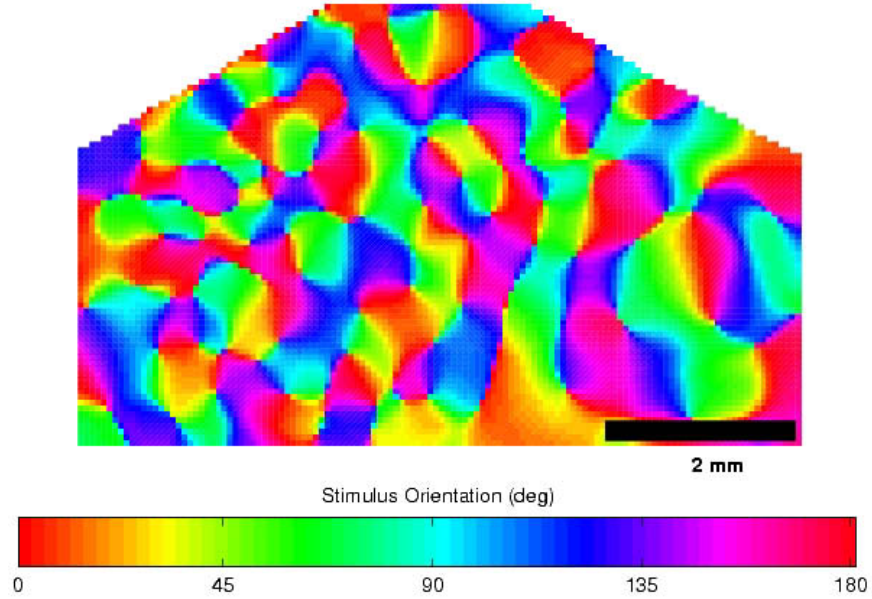


Figure 13: Map of orientation tuning

6 Singular Value Decomposition

We can go back to the decomposition

$$f(t, \mathbf{x}) = \sum_{n=1}^T a_n(t) \mu_n \varphi_n(\mathbf{x}) \quad (61)$$

and cast it into the framework of matrix theory. For each t we construct a vector of pixels by concatenating the rows of the image. The resulting matrix, denoted by \mathbf{M} , is

$$\mathbf{M} = \begin{bmatrix} f(1, 1) & f(1, 2) & \dots & f(1, P) \\ f(2, 1) & f(2, 2) & \dots & f(2, P) \\ \vdots & & & \vdots \\ f(T, 1) & \dots & \dots & f(T, P) \end{bmatrix}, \quad (62)$$

and hence each row of the matrix is an image of the ensemble. We perform a like operation on the $a_n(t)$

$$\mathbf{A}_n = \begin{bmatrix} a_n(1) \\ \vdots \\ a_n(T) \end{bmatrix} \quad (63)$$

and also on the $\varphi_n(\mathbf{x})$ and write

$$\mathbf{V}_n = \begin{bmatrix} \varphi_n(1) \\ \vdots \\ \varphi_n(P) \end{bmatrix}. \quad (64)$$

In such terms (61) can be re-expressed as

$$\mathbf{M} = \sum_{n=1}^T \mathbf{A}_n \mu_n \mathbf{V}_n^\dagger, \quad (65)$$

or if we assemble the matrices

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_T \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}, \mathbf{U} = \begin{bmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_T \end{bmatrix} \quad (66)$$

and

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 & \dots & \mathbf{V}_T \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \quad (67)$$

then (65) and hence (61) can be written as

$$\mathbf{M} = \mathbf{A} \mathbf{U} \mathbf{V}^\dagger \quad (68)$$

in which the columns of \mathbf{A} are orthonormal, as are the columns of \mathbf{V} . This is known as the singular value decomposition (SVD) of the matrix \mathbf{M} .

The SVD is a general result for arbitrary matrix arrays. The assertion that an arbitrary matrix \mathbf{M} , has the decomposition (65) with the sets $\{\mathbf{A}_n\}$ and $\{\mathbf{V}_n\}$, each orthonormal, leads to their construction. This, in essence, is what was proven in section 2. SVD is also a standard numerical tool and may be found for example as a function in Matlab.

It is interesting to note that any image itself may be regarded as a matrix. Thus any face can be put into the form

$$\mathbf{F} = \begin{bmatrix} f(x_1, y_1) & f(x_2, y_1) & \dots & f(x_N, y_1) \\ f(x_1, y_2) & f(x_2, y_2) & \dots & f(x_N, y_2) \\ \vdots & \vdots & & \vdots \\ f(x_1, y_M) & & \dots & f(x_N, y_M) \end{bmatrix} \quad (69)$$

in terms of the gray scales at each of the $M \times N$ pixel locations. If SVD is applied to \mathbf{F} we get, say (suppose $M \leq N$)

$$\mathbf{F} = \mathbf{a}\mathbf{u}\mathbf{v}^\dagger \tag{70}$$

where \mathbf{a} is $M \times N$, \mathbf{u} is $M \times M$ and \mathbf{v} is $N \times M$.

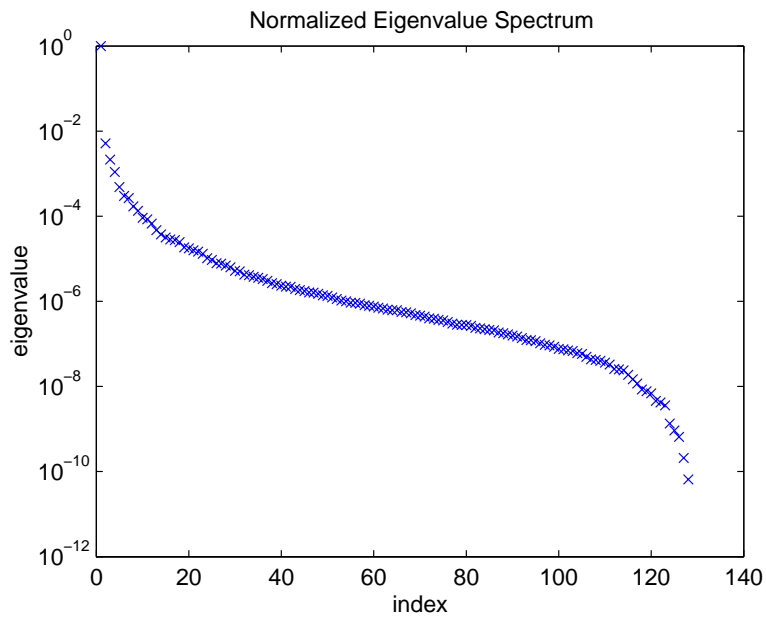


Figure 14: Spectrum of a face regarded as a matrix.

To illustrate the use of SVD on an image regarded as a matrix we consider another in-ensemble face . Viewed as a matrix this has the SVD eigenvalue spectrum shown in Figure 14. the precipitous fall off of the eigenvalues suggests that we can compress an image by taking fewer than M elements from \mathbf{a} and \mathbf{v} . This is illustrated in figure 15.

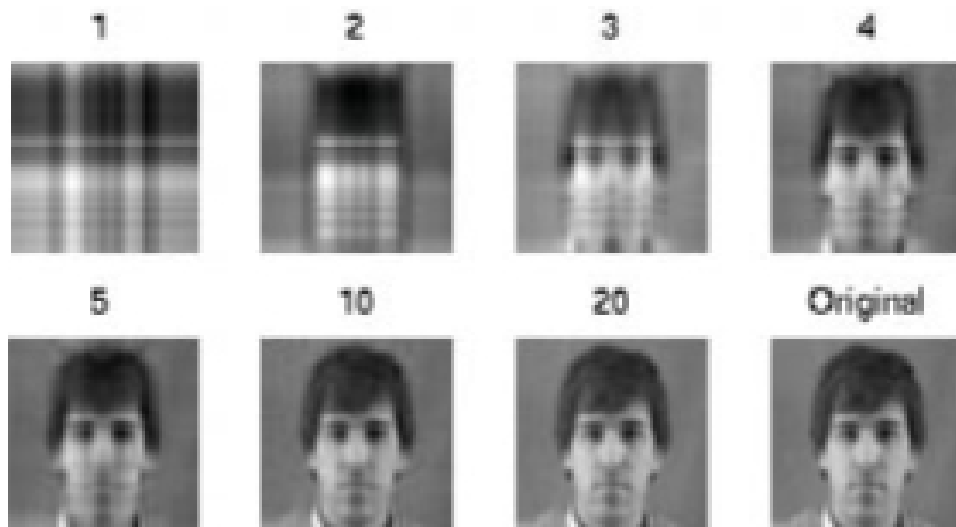


Figure 15: A typical face, lower right, is approximated by 1,2,3,4,5,10 and 20 principal components as indicated in the figure.

Acknowledgement. I am grateful to Takeshi Yokoo and Andrew Sornborger for help and suggestions in the preparation of these notes.

References

- Kirby, M. and L. Sirovich (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(1), 103–108.
- Sirovich, L. (1987). Turbulence and the dynamics of coherent structures, parts i, ii, and iii. *Quarterly of Applied Mathematics* XLV(3), 561–590.
- Sirovich, L., R. Everson, E. Kaplan, B. Knight, E. O’Brien, and D. Orbach (1996). Modeling the functional organization of the visual cortex. *Physica D* 96, 355–366.
- Sirovich, L. and M. Kirby (1987). Low-dimensional procedure for the characterization of human faces. *J. Optical Society of America* 4, 519–524.