Symmetry, probability, and recognition in face space

Lawrence Sirovich[1] and Marsha Meytlis

Laboratory of Applied Mathematics, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, New York, NY 10029

The essential midline symmetry of human faces is shown to play a key role in facial coding and recognition. This also has deep and important connections with recent explorations of the organization of primate cortex, as well as human psychophysical experiments. Evidence is presented that the dimension of face recognition space for human faces is dramatically lower than previous estimates. One result of the present development is the construction of a probability distribution in face space that produces an interesting and realistic range of (synthetic) faces. Another is a recognition algorithm that by reasonable criteria is nearly 100% accurate.

face dimension | probability distributions | face recognition

Fig. 1. Face elements. (*A*) Example of a typical face. (*B*) A cameo representation of a typical face, where the background is removed with a mask. (*C*) The average or mean face. (*D*) *B* − *C*; the face in *C* is subtracted from the face in *B*, resulting in a caricature, i.e., the departure of the face from the mean.

$\mathbf{V}$isual space may be conveniently regarded as a tableau of gray levels assigned to a square of $O(10^5)$ pixels. Human faces belong to a subspace of this high-dimensional space, and the representation/identification of faces termed the Rogue's Gallery problem (1, 2) can be carried out in the reduced space, defined by the masks of Fig. 1. Dimension might be halved by exploiting the latent symmetry of faces and associating each face with its midline symmetrized counterpart (see Fig. 1*A*). Recent laboratory experiments strongly imply that the primate visual system adopts this right/left symmetry, and, moreover, does so for a full range of poses (3); in fact, right and left poses equally angled from the frontal position give rise to equal neuronal response. As will be demonstrated, this thread leads to an exceptional dimension reduction. After a correction for the illumination artifact, a probabilistic exploration of face space leads to a Gaussian probability description in terms of appropriately transformed variables. One result of this is that synthetic faces, i.e., faces drawn from the probability distribution, result in realistic human faces that pass the test of visual inspection. This is illustrated (see Fig. 5) by a small range of reasonable faces, drawn from the probability distribution and are, hence, faces of people we will never encounter.

Mounting evidence indicates that in the primate brain, appreciable resources involving many cortical areas are dedicated to face perception (3–7). As emphasized in the last reference, (see also ref. 8), this provides compelling evidence for the modular character of face recognition in support of the single "domain-specific" side in the cognitive science debate on whether mechanisms are specific or general. As a consequence, as has been observed in many studies faces are our biologically most significant stimuli. For this reason, the role of the Rogue's Gallery problem takes on an importance well beyond its practical applications. Another consequence of these deliberations impacts on the face-recognition debate on whether a face should be parsed or viewed holistically (9). In early physiological studies on macaque, it was found that elements of incomplete faces elicited little or no response, in contrast to the robust response produced by a complete face (10). This is supported by a recent human psychophysical investigation (11) showing dramatic responses to full-face stimuli and virtually no effect arising from partial faces. These studies provide evidence that the human visual system chooses the holistic approach taken here, although for purposes of machine face recognition, parsing may yet prove more efficient.
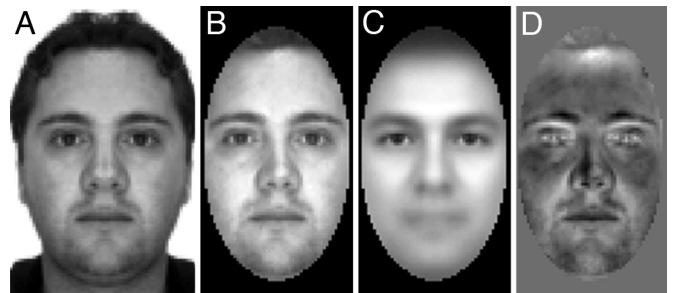
On a practical note, this investigation leads to a recognition algorithm which, by reasonable criteria, gives nearly perfect frontal identification even under naturally diverse lighting conditions.

Background

Face-recognition investigations are largely based on derivatives of pixel image representations in the form of an empirical eigenfunction decomposition. Specifically, if $\{f_n(\mathbf{x})\}$ denotes an ensemble of faces in terms of gray level f at pixel locations $\mathbf{x}$, then a face can generally be represented by some orthonormal function set $\{\phi_n(\mathbf{x})\}$, so that

$$f_j(\mathbf{x}) = \sum_{n=1}^{N} a_n^j \phi_n(\mathbf{x}), \qquad [1]$$

where the coefficients for the jth face are given by the projections

$$a_n^j = (\phi_n, f_j)_{\mathbf{x}} = \int \phi_n(\mathbf{x}) f_j(\mathbf{x}) \, d\mathbf{x}. \qquad [2]$$

If the set $\{\phi_n\}$ are the eigenfunctions of the covariance operator of the ensemble, these are termed eigenfaces, and it can be shown that for any fixed N, Eq. **1** produces the minimal average Euclidean error. These considerations, which were first applied to the Rogue's Gallery problem (1, 2), have become standard (see, e.g., cited references and www.face-rec.org/new-papers/).[†] Based on a tolerable average error of ≈3%, one can fix N and then regard this as a dimension estimate. For the limited early ensemble (1, 2) of ≈125 faces, this dimension estimate was estimated to be ≈100. A subsequent study with a larger ensemble suggested $O(500)$ (12). Another approach with origins in signal

---

[1]To whom correspondence should be addressed. E-mail: lawrence.sirovich@mssm.edu.

[†]In the following, faces have been normalized so that eye separation is standardized and so that the total reflectivity is the same for each face. Hence, each face distantly viewed is the same gray blur. Unless said otherwise, f_j has the mean face subtracted.
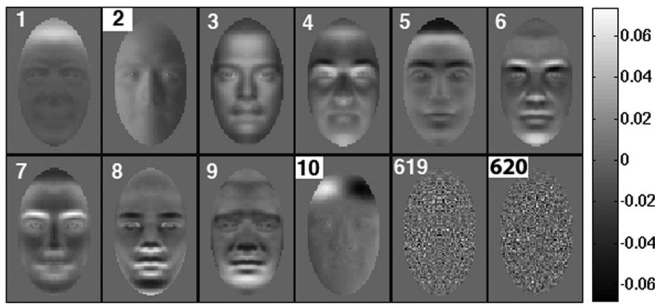
Fig. 2. Typical eigenfaces. The first 10 eigenfaces that are in the signal range and 2 of the eigenfaces that are in the noisy range are shown. It should be noted that the odd eigenfaces (2, 10, and 620) are labeled with a white square around the index number. Eigenfaces are sorted according to decreasing eigenvalue.

Fig. 3. Covariance spectrum. A log–log depiction of the covariance spectrum is shown. Dashed lines are regression fits to signal and noise ranges. The intersection of the 2 dashed lines furnishes the estimate of the dimension as $\approx$ 75. Eigenvalues have been normalized by their sum and thus appear as probabilities. Note that the eigenvalues corresponding to odd and even eigenfaces are shown as different symbols.

filtering considers the eigenvalue spectrum of the covariance operator as being divided into signal and noise, with each approximately obeying a different power law (13). Applying this in the present case leads to a dimension of $N \approx 175$. Such deliberations are based on Euclidean measures and, as such, lie within the framework of a machine recognition perspective. A recent psychophysical study furnished a cortical estimate of face space dimension and found this to be $O(100)$ (13), which as we show, can be further reduced. An interesting feature of that study was the subjective variation in dimension. Perhaps not surprisingly, some subjects appeared to be exceptional in recognizing a face from significantly less information, thus leading to a significantly smaller individual dimension estimate.

Informally, it is reasonable to estimate a tolerable standard deviation in determining a single dimension as $\sigma = 0.1$. On estimating this by $1/\sqrt{n}$, for a sample size of N implies $O(10^2)$ images per dimension. Each dimension may be regarded as independent because eigenfaces decorrelate dimension. Thus, even a modest estimate of 10^2 dimensions would require $O(10^4)$ members for the underlying ensemble. It therefore becomes compelling that we pursue our investigation within the framework of a homogeneous population, and only for this reason we restrict attention to Caucasian males, without facial hair, glasses, and so forth. Experience has taught us that merging this population with a like female population increases the dimension by at least 40% (13). Although there is no like estimate for inclusion of images with facial hair, eyeglasses, etc., or for that matter, racial/ethnic classes, consideration of a diverse population produces a compounding effect that quickly elevates the needed ensemble size to unattainable proportions. Thus, the reasons for restricting attention to 1 homogeneous population appears to be compelling in initial studies.

Results

A typical face from the population is shown in Fig. 1A. In Fig. 1B, background was removed with a mask; for comparison, the mean face of the ensemble is shown in Fig. 1C; and when the mean is removed from Fig. 1B, the caricature, Fig. 1D, results. Representative eigenfunctions are shown in Fig. 2. The 2 high-index examples show the noisy quality at the tail of the spectrum. The actual population contains 329 individuals, which exploiting the latent symmetry of faces (2), is doubled by including the midline mirror image faces. Recent remarkable experiments (3) strongly suggests that the cortex also makes use of this latent symmetry (also see ref. 14). For sound mathematical reasons, the use of symmetry renders all eigenfaces as even or odd, in the midline (2). This dichotomy is evident in Fig. 2, where odd eigenfaces are indicated by white background numbers.
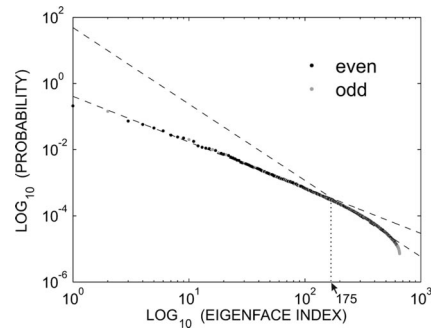
The corresponding eigenvalue spectrum, in log–log form, is displayed in Fig. 3, and if one is not too exacting, the spectrum can approximately be separated into 2 power law regimes. Low indices (low spatial frequency) are characteristic of signal, and high indices (high spatial frequency) are characterized by noise. The eigenfaces are sorted by the decreasing size of eigenvalues and are found to be well-correlated with increasing spatial frequency. The crossing point at $N \approx 175$, furnishes an estimate of face space dimension. According to our earlier deliberations, we might need >17,000 exemplars to confidently deal with this dimensionality. Thus, with the available ensembles, we are only sparsely sampling face space. Therefore we explore leveraging the limited data by means of a probabilistic description.

Probability Considerations

The goal is to obtain a representative probability distribution function (pdf) that will reasonably describe face space, so that a face selection based on the pdf results in a corresponding (synthetic) face that is reasonable by the standard of visual inspection. To pursue this thread, we adopt the following null hypothesis: that gray levels of pixels are independent identically distributed random variables.

In Eq. **1**, the coefficients are determined by the inner product (Eq. **2**), and hence from the form of the eigenfaces, as illustrated in Fig. 2, each coefficient is a sum over many pixels, and therefore under the null hypotheses and the central limit theorem (15) we infer that coefficients $\{a_n^j\}$, (n fixed), should obey a Gaussian pdf.

a_n^j is calculated as the nth coefficient of the expansion (Eq. **1**) of the jth face in the population (Eq. **2**), and we can estimate the pdf of a_n by

$$p^*(a_n) = \frac{1}{N} \sum_{j=1}^{N} \delta(a_n - a_n^j). \qquad [3]$$

As pointed out in ref. 16, this is the basis of a maximum-likelihood estimate, a zero-bias estimate and a bootstrap procedure. For our purposes Eq. **3** is ideally suited to dealing with the difficulties of relatively small amounts of data. We denote the cumulative pdf of Eq. **3** by C^*, and use the zero-bias estimate to construct the cumulative Gaussian pdf denoted by C_G. The following variant of the Kolmogorov–Smirnov (KS) statistic,

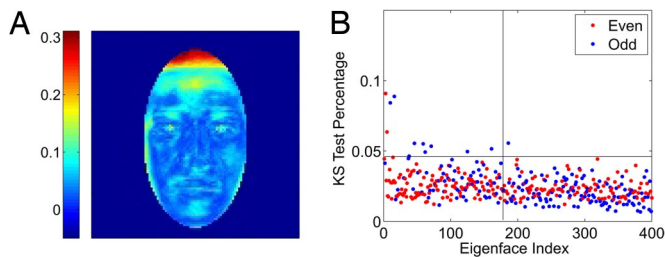$$KS = \sqrt{\frac{\|C_G - C^*\|^2}{\|C_G\|^2}} \qquad [4]$$

Fig. 4. KS statistic. (*A*) The KS statistic is calculated for each pixel over the population of all faces. (*B*) The KS statistic is calculated in coefficient a_n space; this can also be referred to as PCA space. For region of significance $n < 140$, 61 of the eigenfaces are odd in the midline. The absence of red or blue dots for KS ≈ 0 is purely a result of sampling. It can be shown that $(6N)^{-0.5}$ is a diffuse boundary for KS, where N is the number of samples.

is taken as the criterion statistic for normality of the pdf. For the ensemble under consideration the results are depicted in Fig. 4. In Fig. 4*B*, the KS statistics of the coefficients (Eq. **2**) are presented, with even (red dots) and odd coefficients (blue dots) distinguished.

Fig. 4 contains diverse statistical information. For example in Fig. 4*B* the KS statistic (Eq. **4**) is displayed for even (red) and odd (blue) coefficients. Clearly for KS small normality of the pdf is implied, and the question of a threshold arises. In this connection, we mention that the appearance of odd eigenfaces at low indices was greeted with surprise in the investigation leading to ref. 1 until it was realized that this was due to the daily variation of sunlight as images were acquired.

It is our contention that if face images could be acquired under ideal conditions of totally diffuse light, then odd eigenfaces would be relegated to the high indices especially if hair is removed, as we have with the mask shown in Fig. 1. Therefore the presence of odd eigenfunctions at low indices is an indicator of the lighting artifact. It is clear from Fig. 4*B* that after the first red point, the ratio of cumulative blue points to cumulative red points climbs to a sharp maximum as KS is decreased. After this the ratio falls rapidly into basic parity with the red dots. At this point KS = 0.046, and this is taken as the criterion point for normality of the pdf. This is indicated by the horizontal line shown in Fig. 4*B*. For the relevant range of $n < 175$ just 2 even eigenface coefficients are non-Gaussian. Thus the null hypothesis is refuted for a large portion of the odd eigenfaces and only for 2 even eigenfaces.

Some insight into this result is provided by Fig. 4*A* which furnishes a pixel map as measured by the KS statistic. Essential to the central limit theorem is that the summed quantity be independent and random. Application of the theorem in this instance depends on correlation lengths, random light sources, pixel size, and ensemble size. Highly correlated regions such as hair and eyes are clearly nonrandom, as indicated in Fig. 4*A*. Other sources that lack randomness are the mask boundaries, where the vector normal to the original face surface can be nearly perpendicular to incident light.

As has been pointed out in (17) illumination variation very likely exceeds actual face variations, and from our investigation illumination appears to be the greatest source of non-random behavior. In most instances lighting for capture of a facial image is frontal, and as Fig. 4*A* indicates the forehead is a large area of highly non-random pixels. The forehead behaves almost as a mirror which, along with the hair artifact, is largely accounted for by eigenfaces 3 and 5 of Fig. 2 (second and fourth even eigenfaces), the 2 non-Gaussian cases.

Synthetic Faces

In keeping with the contention that odd eigenfaces are largely the result of the lighting artifact, we explore dropping these from
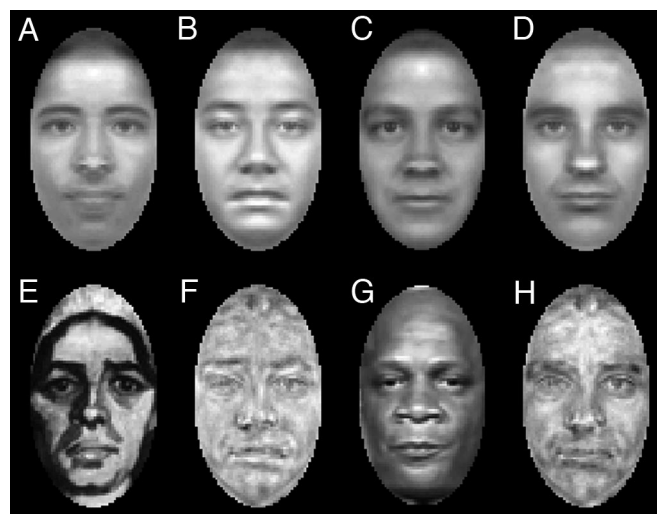

Fig. 5. Synthetic and other faces. (*A–D*) Male synthetic faces generated from male even eigenfaces. Each face is of the form (4), with the mean and the constant contribution from ϕ_3 and ϕ_5, as indicated in *Synthetic Faces*. The latter gives the impression that there is lighting from behind the camera. (*E–H*) *B* and *D* are the same as *F* and *H* except that all odd and even eigenfaces are used in constructing the faces, and all coefficients are chosen from a Gaussian, based on variances; *E* and *G* are discussed in *Synthetic Faces*.

consideration. In addition we drop the 2 special even eigenfaces. The remaining coefficients are regarded as normally distributed with zero mean, and variance proportional to the eigenvalue of the covariance as given in Fig. 3. We can then construct synthetic faces out of just the even eigenfaces by drawing even coefficients $\{a_n^e\}$ from the specified appropriate normal probability distribution, thus generating the form

$$S(\mathbf{x}) = \sum_{n \neq 2,4} a_n^e \phi_n^e (\mathbf{x}). \qquad [5]$$

Four exemplars of this construction are exhibited in Fig. 5 *A–D*, where for viewing purposes, the mean has been added to each representative as well as $(\sqrt{\lambda_2}\phi_2^e + \sqrt{\lambda_4}\phi_4^e)$, so as to give the impression of being a frontal face. Each example passes the visual test of being a face, but being generated at random from a probability distribution, it is unlikely that we will meet an individual with such a face. A small fraction of the so-generated synthetic faces appeared to be androgenous or even feminine, as typified by Fig. 5*A*. This image appears to bear a resemblance to van Gogh's "Kop van een vrouw" Fig. 5*E*, which shows some of the range of faces that can appear based on Caucasian males. Along these lines we observe that Fig. 5*C* bears at least a superficial resemblance to the great baseball star of the 1980s New York Mets, Darryl Strawberry, a black athlete, shown in Fig. 5*G*.

We note in passing that the idea of synthetic faces appeared earlier (18), where less realistic "pseudofaces" are used to explore cortical organization of faces.

If the full probability distribution, comprising all odd and even eigenfaces is used in a Gaussian fit, then grotesque images can emerge. To illustrate this we have taken the images of Fig. 5 *B* and *D* and added to them the contribution of terms excluded from Eq. **5**, using an appropriate Gaussian distribution for the coefficients. These are shown in Fig. 5 *F* and *H*, and they are clearly distorted and unnatural. The mottled appearance of the lower row of Fig. 5 can be traced to the odd eigenfaces, which carry localized variations in lighting.

Table 1. Summary of results for recognition experiment

Eigenbasis	Errors, n	Errors, %
Full	82	87
Even	21	22
Even with 2nd and 4th even dropped	4	4
Even with 2nd and 4th even dropped normalized by SD	1	1

Fig. 6. Recognition results. (*A–D*) The 4 out-of population faces with lighting artifact that were incorrectly identified by using the unweighted Euclidean distance metric. (*E–H*) Their in-population representors. Only *C* is misidentified under the weighted distance described in the text.

Rogue's Gallery Problem

The even rendering of a face results in a visually recognizable face. This transformation was applied to Fig. 1*A* clearly without loss of identity. Thus, ignoring odd components removes the lighting artifact without cost to identification. To follow up on these observations, we next explore the face recognition problem. In anticipation of this, additional male images were acquired from the 3 face databases (19–21). Each of these collections contained 3 images of each individual: 1 frontal image with center lighting and 2 frontal images that are marred by variable lighting from the right or left. All center-lit images, 47 in number, were included as in-population faces of the male population.[‡] The corresponding 94 male images with pronounced lighting variations were reserved for testing recognition and were not included in the population.

Recognition was based on the L_2 norm of the 175 coefficients that were deemed to carry the signal. Table 1 contains the results leading to what we deem to be an optimal procedure. Less optimal results are included for their informational content. If all coefficients are considered, 82 (of 94) errors are made, an 87% error rate. If only the subspace of even coefficients are considered, the error rate drops to 22%, and if the subspace is further reduced by dropping a_2^e and a_4^e, the error rate drops to 4%. The 4 errors in this case are depicted in Fig. 6, and in each instance, one sees that contrary to the instructions from the ''photographer,'' the subject changed his pose for the marred image.

For the procedure thus far, coefficients carry their natural weighting, which because $\langle a_k^2 \rangle = \lambda_k$ says only that a coefficient's weight diminishes with index. Under the ideal of an ensemble of unlimited number, each eigenface would be perfectly resolved, which would then argue for weighting the a_k term in the L_2 norm by the reciprocal variance, λ_k^{-1}, the so-called whitened norm. In our case, this leads to negligible improvement, 1 less error. We adopt a geometric compromise and weight each a_k term in the L_2 norm by $1/\sqrt{\lambda_k}$, the reciprocal of the singular value that is proportional to the standard derivation. The result as indicated in Table 1 is just 1 error, namely Fig. 6*C*. In anecdotal terms, in resolving Fig. 6*C* by eigenfaces, for the fit, the procedure attempts to open the closed eyes. As a result, the fit wanders too far from the correct result. Earlier authors (20, 21) noted that dropping some of the early eigenfaces improved

recognition, and our discussion provides an objective and specific basis for the observation as well as another norm. On a cautionary note, the face recognition algorithm does not do as well for faces with artificial vertical illumination, artifactual hair display, facial distortion, and similar impediments to recognition.

Discussion

The data for frontal poses marred by lighting artifacts is somewhat limited but has been acquired from 3 different databases (19–21). The sole error of recognition occurred because the subject did not follow instructions and closed his eyes. If on this basis, we exclude this case, recognition becomes 100%.

Face recognition was accomplished by using the even subspace minus the 2 non-Gaussian eigenfaces of the estimated 175-dimensional face space, and is therefore of dimension ≈110. This form of recognition takes place using an L_2 norm and therefore lies in the framework of machine recognition. An earlier psychophysical study (13), involving a merged population of male and female faces, demonstrated that human observers require on average only dimension ≈ 100 (of N ≈ 200) for recognition. However, the reduction to even eigenfaces was not applied in that study. With the estimate that >30% of the relevant eigenfaces are odd, we can revise the earlier estimate (13) to <70 dimensions for the human recognition system. This framework is consistent with the physiological studies (3, 6) discussed in the introduction.

[‡]This was not essential, and we were able to achieve the same degree of recognition accuracy without including these in the base population.
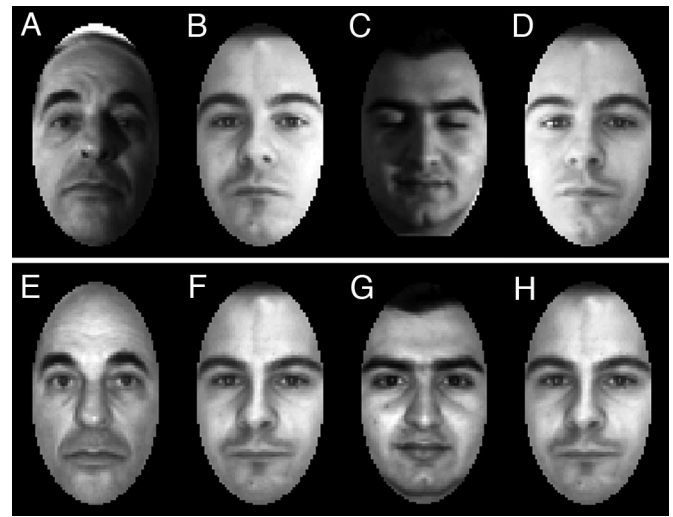
1. Sirovich L, Kirby M (1987) Low-dimensional procedure for the characterization of human faces. *J Opt Soc Am* 4:519–524.
2. Kirby M, Sirovich L (1990) Application of the Karhunen–Loeve procedure for the characterization of human faces. *IEEE Trans Pattern Anal Machine Intell* 12:103–108.
3. Freiwald W, Tsao D (2006) Single-unit recording in fMRI-identified macaque face patches: Evidence for a second domain-specific module. *Society for Neuroscience Abstracts online. Program NO. 438.8.* Available at: www.abstractsonline.com/viewer/viewAbstractPrintFriendly.asp?CKey={11A9C9EE-3DA4-4585-8C2E-8A19558C318A}&SKey={F8DC3D3D-E02F-46BE-870C-EBA5C9E47285}&MKey={D1974E76-28AF-4C1C-8AE8-4F73B56247A7}&AKey={3A7DC0B9-D787-44AA-BD08-FA7BB2FE9004}. Accessed on March 12, 2009.
4. Tanaka K, Saito H, Fukada Y, Moriya M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66:170–189.
5. Kanwisher N, McDermott J, Chun M (1997) The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
6. Tsao D, Freiwald W, Tootell R, Livingstone M (2006) A cortical region consisting entirely of face-selective cells. *Science* 311:670–674.
7. Kanwisher N, Yovel G (2006) The fusiform face area: A cortical region specialized for the perception of faces. *Philos Trans R Soc London Ser B* 361:2109–2128.
8. Kanwisher N (2006) What's in a face? *Science* 311:617–618.
9. Wechsler, H (2006) *Reliable Face Recognition Methods: System Design, Implementation and Evaluation* (Springer, New York).

10. Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71:856–867.

11. Hershler O, Hochstein S (2005) At first sight: A high-level pop out effect for faces. *Vision Res* 45:1707–1724.

12. Penev P, Sirovich L (2000) The global dimensionality of face space. In *Proc 4th Intl Conf Automatic Face and Gesture Recognition, IEEE CS* (IEEE, New York), pp 264–270.

13. Meytlis M, Sirovich L (2007) On the dimensionality of face space. *IEEE Trans Pattern Anal Machine Intell* 29:1262–1267.

14. Everson R, *et al.* (1998) Representation of spatial frequency and orientation in the visual cortex. *Proc Natl Acad Sci USA* 95:8334–8338.

15. Papoulis A (1991) *Probability, Random Variables and Stochastic Processes.* (McGraw–Hill, New York).

16. Lupton R (1993) *Statistics in Theory and Practice* (Princeton Univ Press, Princeton).

17. Moses Y, Adini Y, Ullman S (1994) Face recognition: The problem of compensating for changes in illumination direction. In *Lecture Notes in Computer Science, Computer Vision–ECCV'94*, ed Eklundh JO (Springer-Verlag, Berlin Heidelberg), Vol. 800, pp 286–298.

18. Wilson H, Loffler G, Wilkinson F (2002) Synthetic faces, face cubes, and the geometry of face space. *Vision Res* 42:2909–2923.

19. Georghiades A, Belhumeur P, Kriegman D (2001) From few too many illumination cone models for face recognition under variable lighting and pose. *IEEE Trans Pattern Anal Machine Intell* 23:643–660.

20. Belhumeur P, Hespanha J, Kriegman D (1997) Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans Pattern Anal Machine Intell* 19:711–720.

21. Martinez A, Benavente R (1998) The AR face database. *CVC Tech Report No. 24.* Available at: http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html. Accessed on March 9, 2009.

APPLIED MATHEMATICS