

On the Dimensionality of Face Space

Marsha Meytlis and Lawrence Sirovich

Laboratory of Applied Mathematics

Mount Sinai School of Medicine

1 Gustave L. Levy Place

New York, NY 10029

Abstract

The dimensionality of *face space* is measured objectively in a psychophysical study. Within this framework we obtain a measurement of the dimension for the human visual system. Using an eigenface basis, evidence is presented that talented human observers are able to identify familiar faces that lie in a space of roughly 100 dimensions, and the average observer requires a space of between 100 and 200 dimensions. This is below most current estimates. It is further argued that these estimates give an upper bound for *face space* dimension, and this might be lowered by better constructed "eigenfaces", and by talented observers.

I. INTRODUCTION

The general face recognition problem, seeks the identification of a person in a scene, based on known faces in a stored database. This problem may be regarded in three stages [1]: face detection [2], [3]; feature extraction [4]; face identification [5], [6], [7], [8]. The present study focuses on the third stage, which includes representation [5], [6] and recognition [7], [8].

Implicit to the present approach is the fact that the image of a face is dealt with in its entirety and not in terms of its parts or elements, thus in essence this approach is holistic. This procedure has its origin in a body of related mathematical methods having the acronyms SVD, PCA, KL, EOF, POF and so forth, an approach that has been discovered and rediscovered many times over [9]. Common to all of these is a reliance upon empirical data to generate a mathematically

optimal coordinate system that is intrinsic to the data. The coordinates are eigenfunctions of a data generated operator, and thus may be referred to in general as empirical eigenfunctions. We avoid the fine points of the various acronyms, although SVD best describes the approach used here. The introduction of this approach for face recognition first appears in [5], where it was shown to lead to a low-dimensional description of *face space*. The appropriate dimension of this space is a central concern of this paper. Mention should also be made of other holistic-based methods [10], [11], [12], [13], [14], [15], [16], [17].

The eigenface approach, perhaps the first face recognition model aimed at extracting information important for recognition, was based on the premise that a relatively small number of elements or 'features' could be efficiently used for face recognition [7], [8]. Dimension reduction should be expected since faces share the same basic configuration and shape. Eons of evolution suggest that the human brain possesses an optimal algorithm for face recognition. Given the sheer size of an individual's lifetime database, it is astonishing that as little as 75 milliseconds of viewing time is needed for face identification and only 800 milliseconds to solve the face recognition problem [18]. Sorting through a lifetime database in this short span of time is inconceivable without some sort of compact representation. The eigenfunction approach [5], [6], [7], [8] was the first successful attempt to find global 'features' that are important for discriminating faces and to model a compact representation in an automated face recognition setting.

The dimension of *face space* may be reasonably defined as an acceptable threshold number of dimensions necessary to specify an identifiable face. Performance of various face recognition algorithms is largely evaluated in terms of recognition accuracy as a function of dimension. Recognition performance is calculated as the proportion of test images that are correctly identified. In the common scheme, eigenfaces are constructed from a *training* set of face images and particular test faces are recognized by comparing the eigenface weights [8]. Correlation of two sets of weights may be used as a comparison measure.

Recognition performance will suffer from insufficient information if dimensionality is underestimated. On the other hand, an over-estimate of dimension will introduce noisy components

which also reduces performance [19], [20]. The number of dimensions of *face space* also figures in the speed and accuracy of processing large face databases. In addition, the dimensionality question is critical to performance of schemes based on dimensionality reduction [20], [21], [22]. In all this it is important to note that *face space* is specific to the representation of human faces, and any image which diverges from a conventional human face, e.g. a monkey face, [23], is not as well captured by *face space*.

A typical approach to the determination of dimensionality is to examine the variance captured by the eigenfaces and to search for the location of significant drop-off in the variance (eigenvalue) spectrum [5], [6], [21]. The question of dimensionality within a probabilistic framework looks at the *signal-to-noise ratio* (SNR) of the reconstructions [20]. Both methods allow us to determine how much pixel information has been captured in a reconstruction, from an original image; however, they do not inform us whether or not information critical for the identification of a face has been captured.

While the empirical eigenfunction approach is both objective and optimal, like the less efficient pixel representation, it is *man-made*. It is plausible that the human face recognition system is using some other *algorithm*. This is relevant since, in this study, we hope to gain some sense of the representation used by human observers, albeit with the eigenface approach.

[20] implicitly suggest that perceived image quality can be used to measure the dimensionality of *face space*, however no objective assessment of this aspect of images was studied. Here we report on a human psychophysics experiment which interrogates the human visual system to estimate the number of dimensions necessary for recognition of a face. By measuring, the recognition threshold at which enough information is present for human observers to classify a face as familiar, we can come closer to the intrinsic dimensionality of human *face space*. A related but complementary, psychophysical approach to face classification was recently considered by [24] (also see [25]).

The use of error or variance as a determinant of identification, may be required for machine identification, and in such terms a dimension of $O(500)$ has been contemplated [20]. From another perspective there is the intrinsic *dimension* of the *algorithm* used by the human visual

system. Our results indicate a dimension that is closer to 100. In determining this estimate, we have used the eigenface framework, a *man-made* coordinate system. It is possible, and probable, that our visual apparatus uses a more efficient 'basis', perhaps some sort of sparse coding of faces [26], or other efficient coding [27]. From this perspective our estimate of $O(100)$ dimensions for identification is an upper bound for the *intrinsic* dimension. It should also be noted that non-holistic methods which parse a face into component parts offer another approach which may further reduce the dimension. In particular Gabor transforms and Gabor jets which focus on fiducial facial locations are appealing [28], [29], [30], [31].

II. BACKGROUND

In the experiment, facial images viewed by human observers appeared on a computer monitor. All face images were converted to gray scale levels, and then normalized by an affine transformation that set the centered inter-eye line horizontal, and 24 pixels apart. Images were cropped to 128×128 pixels, and an oval mask, the shape of an average face, was applied to each image. Pixels outside the mask were set to 0. There were 3,496 pixels inside the mask. Every image was normalized so that the reflectance of cheek areas were on average the same [5].

The basic ensemble of facial images which was used contained 993 frontal images from the 2003 U.S. Army Face REcognition Technology (FERET) [32] color database, and this set was augmented by 40 images acquired by us. As described in [6] there are a number of advantages in doubling the ensemble by including mirror images of each ensemble member. On so doubling the augmented FERET database we obtained a *training* set of 2,066 images. An additional 40 images were sequestered for later use, and not included in the basic population. SVD of the *training* set ensemble of 2,066 faces was performed and resulted in the eigenfunctions (eigenfaces), $\psi_n(\mathbf{x})$ and the corresponding eigenvalues λ_n , (see Appendix).

Eigenvalues are non-negative and the ratio $\lambda_n/\sum_n \lambda_n$ furnishes the average probability that a corresponding eigenface $\psi_n(\mathbf{x})$ appears in the representation of a face. In Figure 1 shows the probability spectrum of the FERET augmented database from index 2 to 1,500 components. The first eigenface lies close to the average face, and thus has a disproportionately large eigenvalue,

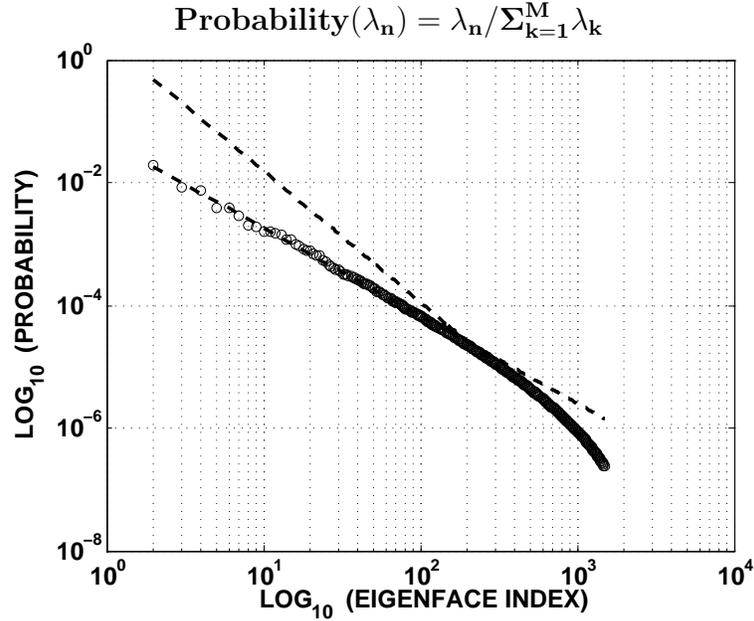


Figure 1. Log-log plot of probability projection along the eigenface directions for the first 1500 components. The curve is fit with 2 dotted lines. We drop out the first point in this plot because the first eigenface is close to the average face of the population. The two power law fits cross at $n \approx 200$.

which we leave out of this plot. As might be supposed, and is supported by the eigenfaces seen in Figure 2, high index eigenfaces capture *noise*, and early indices relate to the actual components of a face, and thus indicate *signal*. As shown in Figure 1 the spectrum is well fit by two straight lines, suggesting two power law regions. The line which captures the *signal* components has a slope of -1.43 and the line that captures the *noise* components has the slope -2.14. In [20] it is observed that the remnants of facial structure in the eigenfaces decay slowly after the first 100 components, as seen in Fig. 2. The cross-over point of Figure 1 lies at roughly $n = 200$, which in [5] and [6] was used as a basis for establishing a dimension estimate. Similarly, the deliberations of [20] are in large part based on the error incurred in reconstructing a face from its projection onto a truncated number of eigenfaces. Since, the norm of the ratio of the reconstruction, to the residual error was used as a measure, this was termed *signal-to-noise ratio* (SNR), a usage we follow. SNR is a measure of error in the reconstruction, i.e., the amount of variance that has been captured in the reconstruction. Thus, a reconstruction which is a better fit to the original face will have a larger SNR. [20] suggested that most face identify information necessary for

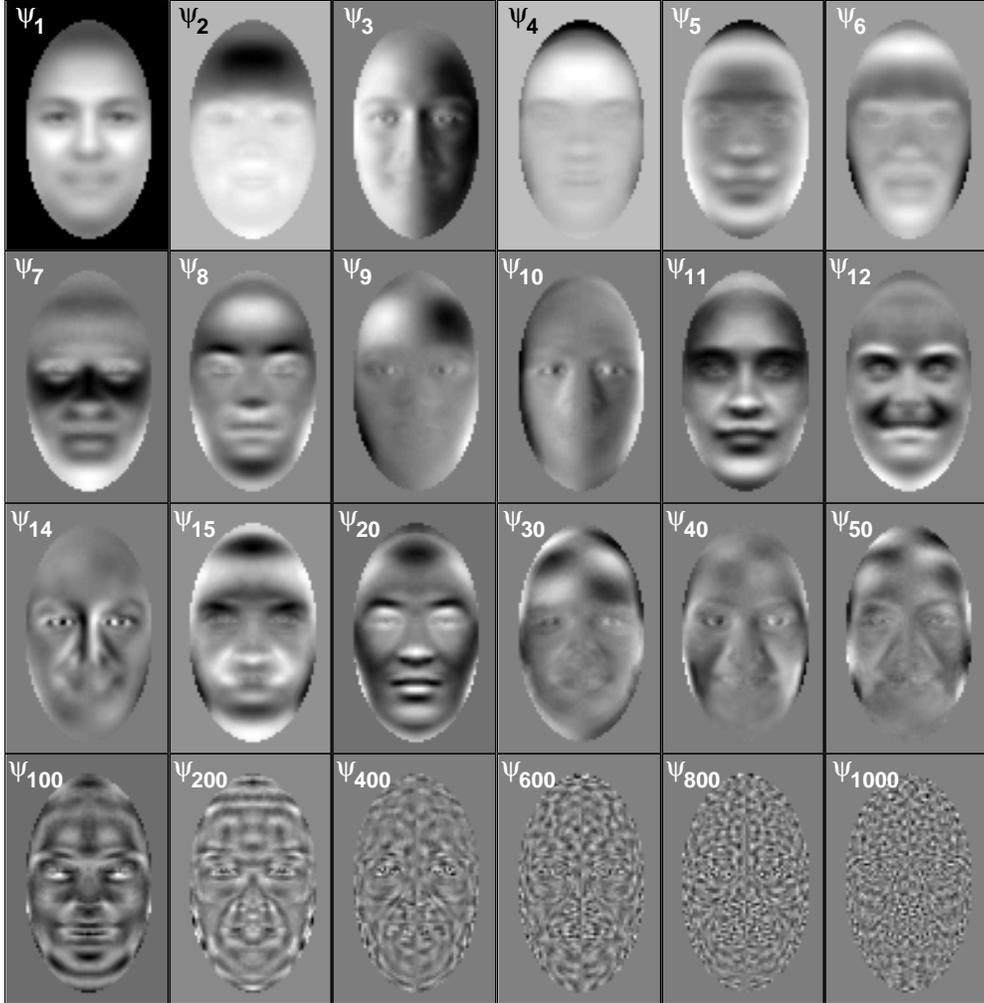


Figure 2. The first 16 and 8 other eigenfaces of the *training* set.

recognition is captured within a SNR span of approximately 7-7.5 octaves, which in their case could require $O(500)$ components. The SNR and related issues are reviewed in the Appendix.

It is important to realize that the dimensional estimates of [5], [6], [20] and related studies are based on Euclidean distance measured in the pixel space of face images. Clearly, this distance, and in particular SNR does not inform us about error tolerance, and hence dimension, for the human visual system that monitors face recognition. A goal of this paper is the exploration of the issue of dimension by measuring this error tolerance.

III. EXPERIMENT

The goal of the experiment was to arrive at an estimate of the dimension of *face space* as it might appear in the *algorithm* for recognition used by the human visual system. To achieve this goal we designed an experiment in which human observers were shown partial reconstructions of faces and asked whether there was recognition.

Ten healthy volunteers participated in the experiment (5 men and 5 women, mean age 27, range 20-35, all right handed): all had normal or corrected-to-normal vision. The observers viewed images on an optimally placed computer monitor.

In more detail the experiment was performed in two parts:

The first part sought to assess a baseline for the observers' knowledge of *familiar* faces. Images of 46 people (3 images of each) deemed to be *familiar*, i.e., popularly recognizable faces, were shown to observers. Participants were asked to respond with one of the following options: high familiarity, medium familiarity, low or no familiarity. None of the images used, in this preliminary part of the experiment, figured in the *training* set or in later stages of the experiment. In addition, six of the baseline people were left out of second part altogether, so that observers would not be able to use the process of elimination in classification. Baseline familiarity ratings for each observer are plotted in Figure 7, which we discuss later.

In the second part of the experiment the observers viewed the truncated versions of 80 faces, referred to as test faces: 20 familiar and 20 unfamiliar images, that were added to the FERET training set, an additional 20 familiar and 20 unfamiliar images not in the training set that were added in order to monitor differences in recognition ability of truncated images not in the *training* set. Sex of the test faces was balanced across familiarity and *training* set inclusion. In the interest of simplifying the problem we chose only Caucasian faces, and hence a relatively homogeneous population was used in classification. In the same vein, test faces with facial hair, extreme features or any other distinguishing characteristics were excluded. In order to match the unfamiliar and famous test faces in attractiveness, we used photographs of unknown models and actors for the unfamiliar test faces.

In Figures 3a, 3b and 3c an unfamiliar and two familiar test faces from outside the *training*

population are reconstructed to successively higher degrees. The number of eigenfaces appears on the top and the SNR on the bottom. These three, out-of-population faces appear to need no more than 200 elements, as suggested by Figure 1, however, different observers make varying judgements. To estimate the error tolerance of the human face recognition system we performed an experiment asking observers to discriminate between familiar/unfamiliar faces reconstructed as a function of varying SNR.

Observers first viewed all 80 test faces in a random sequence reconstructed to SNR of 5. Then, in the same manner, images reconstructed for each subsequent SNR were viewed. SNR was incremented in even steps of 0.5 until 10 was reached, eleven steps in all.

In reconstructing an image faces were incremented in equal steps of SNR instead of count by eigenfaces. This ensures that, with each new face stimulus, equal amounts of variance are captured, a procedure intrinsic to the face. Thus faces are reconstructed at slightly different rates depending on how distinct they are from the faces that are used to generate the eigenfaces [33]. Observers viewed 880 images in this part of the experiment.

Half of the test faces classified by the observers in the psychophysics experiment were in-population, i.e., they were part of the *training* set used in the construction of the eigenfaces, and the other half were out-of-population. This furnished a baseline comparison of reconstruction error. Mean SNR of each group is plotted as a function of component count in Figure 5, confirming the fact that in-population faces are better reconstructed. The standard error of the mean (μ/\sqrt{n} , where μ is the standard deviation from the mean and $n = 40$ is the sample size) shows small variation in the rate of reconstruction, and remains constant as a function of dimensionality. Figure 5 gives the specific relation $SNR = f(n)$, i.e., SNR as a function of eigenface index. Both curves are well fit by power laws $SNR \propto n^p$, with $p \approx 1/7 - 1/6$. In passing we remark that in the limit of an unbounded *training* set the two curves of Figure 5 should converge to each other, thus Figure 5 also measures the completeness of the eigenface basis.

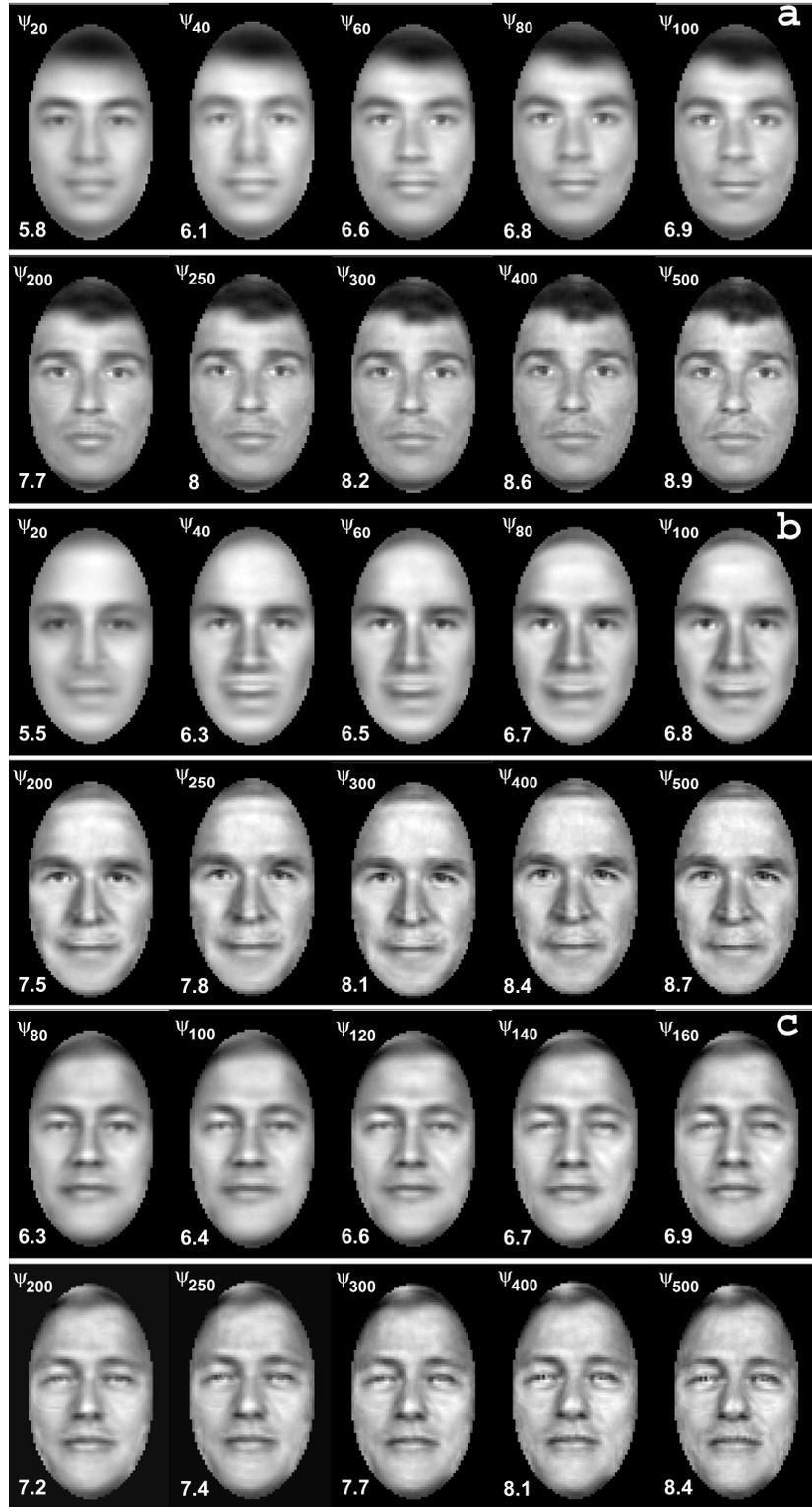


Figure 3. An example of the reconstructions of an unfamiliar(a) and familiar(b, c) out-of-population faces. Each reconstructed image is labeled with the number of components, ψ_n on the top, and the SNR on the bottom.

IV. ANALYSIS OF DATA

Data gathered in the experiments were analyzed using Receiver Operating Characteristic (ROC) curves [34], [35], [36], [37], [38] to classify familiar versus unfamiliar faces. An ROC curve is essentially a plot of false positives versus true positives. ROC analysis, invented for dealing with noisy radio transmissions, deals with threshold effects in the trade-off between false positives and false negatives. It has been used in studies of object detection [38], edge detection [37] and machine learning [34]. In the present psychophysical study the "device" experiencing uncertainty is a human observer who supplies a response based on an internal decision on face recognition. Typical ROC analysis factor the variation of thresholds of certainty. In our study, the human observers, distinguish the degree to which a face is familiar or unfamiliar. Observers were asked to respond with one of the following options: (1) high certainty a face is unfamiliar; (2) medium certainty a face is unfamiliar; (3) low certainty a face is unfamiliar (4); low certainty a face is familiar; (5) medium certainty a face is familiar; (6) high certainty a face is familiar. This is a standard rating procedure in psychophysics [39], [40], [41]. This 6 point response is transformed into a binary for recognition, based on five different thresholds for the observer's responses, $r > (5)$, $r > (4)$, $r > (3)$, $r > (2)$, and $r > (1)$. Thus $r > (5)$ may be regarded as the probability that the observer is certain that he/she is viewing a familiar face, given that a familiar face is indeed being viewed. $r > (4)$ is this probability plus the probability of medium certainty, and so forth with corresponding cumulative probabilities. Thus, an image which received a score above a specific threshold was classified as *familiar* and otherwise it was classified as *unfamiliar*. The proportion of true positive responses was determined as the percentage of familiar faces that were classified as *familiar* at a particular threshold, while the proportion of false positive responses was determined as the percentage of unfamiliar faces that were classified as *familiar* at a particular threshold. Each threshold setting corresponds to one point on the ROC curve.

To illustrate the use of this procedure consider Figure 4 which shows the series of ROC curves for a single observer. Clearly pure chance is described by a 45° line, shown as a dashed line. Each SNR appears at the five thresholds and these appear as we move from left to right in the plots, going from low false positives to high false positives. The data for $\text{SNR} < 7.5$ hovers

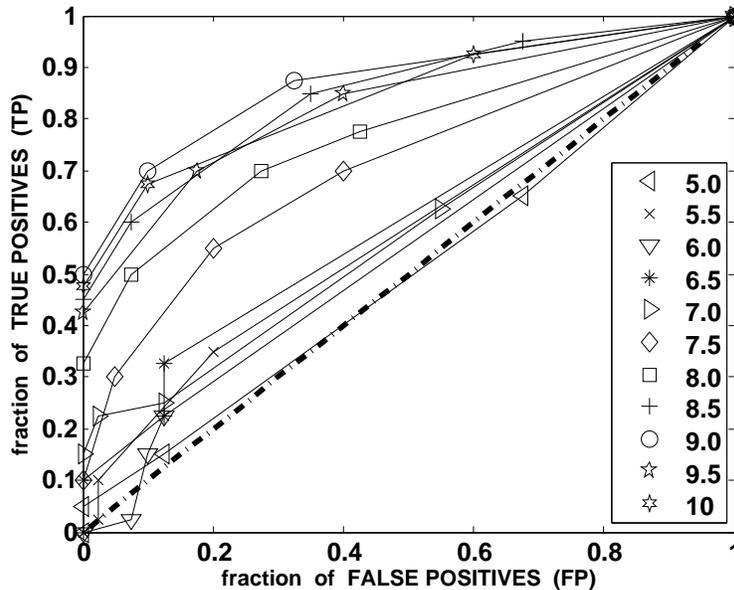


Figure 4. A sample ROC plot for a single observer. The area between each curve and the 45° line corresponds to the observer’s classification accuracy (identification probability) for a specific SNR, as labeled in the legend.

near the 45° line, and might be regarded as being noisy in contrast to $\text{SNR} \geq 7.5$, which carry high signal. Evidently, the area between each curve, and 45° line corresponds to classification accuracy, an increasing function of SNR. We follow common practice, and use the area under the ROC curve which simply adds a baseline value of 0.5 to the numerical classification of accuracy.

V. RESULTS

The plot of face classification accuracy (identification probability) as a function of SNR is referred to as the observers’ psychometric function [43]. Following standard practice, the psychometric function is fit by the Weibull distribution: $p(\text{SNR}) = 1 - 0.5 * \exp(-(\text{SNR}/\alpha)^\beta)$, where p is the proportion of faces identified computed as a function of SNR, parameter values are given in Table 1. Figure 6 shows the average face classification accuracy of 10 observers as a function of SNR. Classification accuracy of 1.0 indicates *perfect* stimulus detection and conventionally, the point at which there is a 50% improvement over chance in classification accuracy (signal detection probability of 0.75) is chosen as the detection threshold [42], [43].

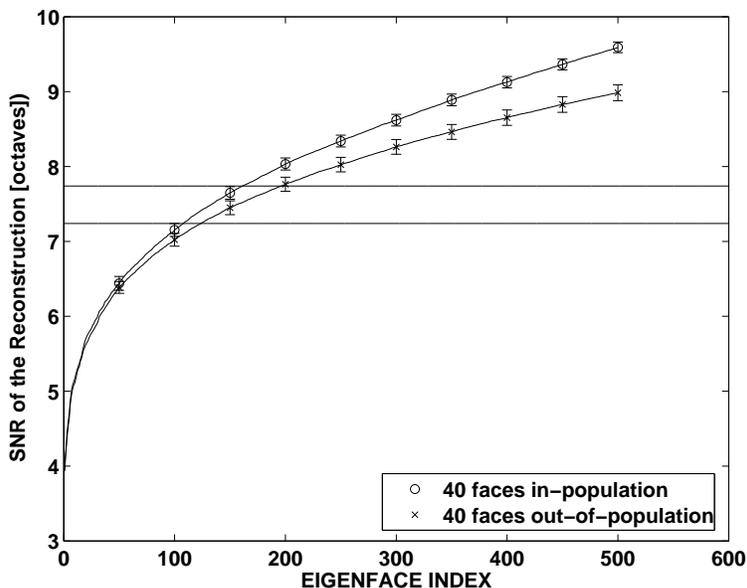


Figure 5. The average *signal-to-noise ratio* (see Appendix for definition) and the standard error of the mean for faces in- and out-of-population is plotted as a function of dimensionality for the first 500 components. The horizontal lines are plotted at an SNR of 7.74 and 7.24, which correspond to the threshold of recognition in Figure 6.

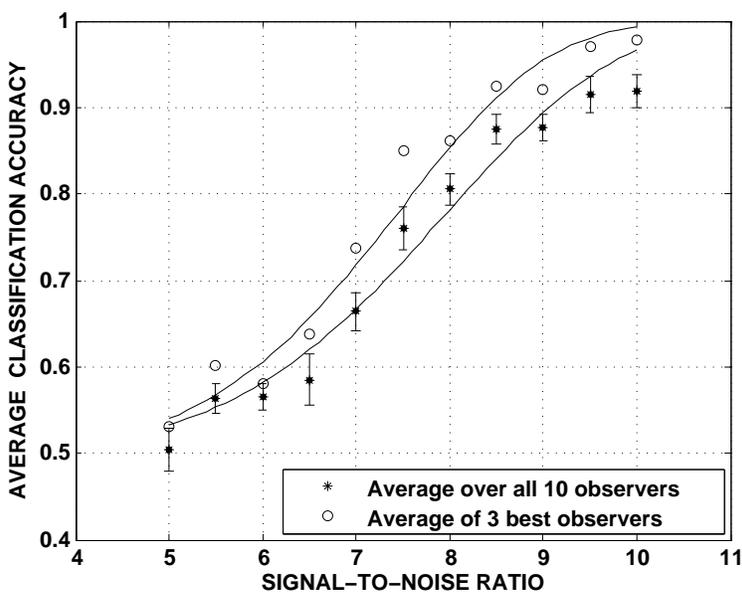


Figure 6. Average psychometric function for face classification. Average classification accuracy is plotted vs. *signal-to-noise ratio*. Using ROC analysis we determined the average classification accuracy for 10 human observers (asterisks) and fit this average to the Weibull distribution [43]. From the fitted psychometric function we calculated the lowest *signal-to-noise ratio* giving a classification accuracy of 0.75 (See Table 1). Similarly the average classification accuracy was determined for the 3 'best observers' (circles) who were most familiar with the familiar faces, selected based on familiarity ratings in Figure 7.

	SNR where $p(\text{SNR})=0.75$	α	β
Average (Fig. 6)	7.74	8.29	5.31
3 Highest Familiarity Scores (Fig.6)	7.24	7.72	5.68
OUT Faces (Fig. 8)	7.66	8.2	5.41
IN Faces (Fig. 8)	7.9	8.42	5.72

Table 1. Parameter values for the Weibull distribution [43] in Figures 6 and 8.

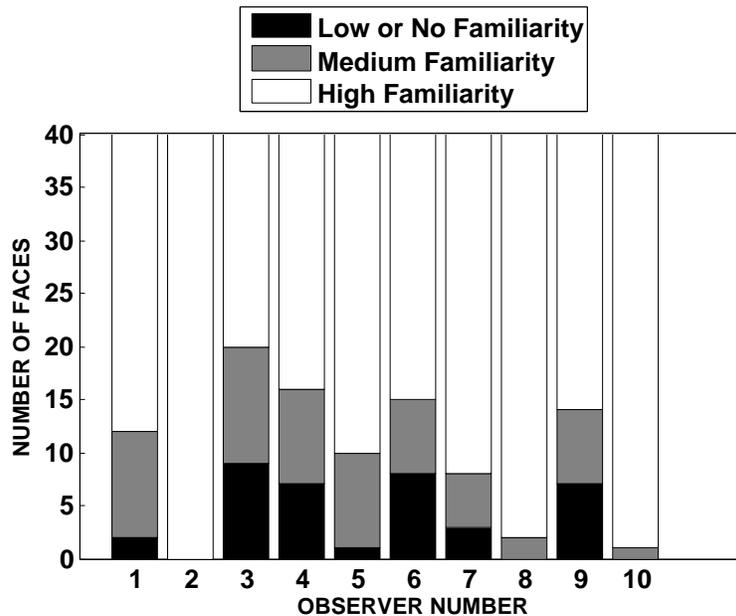


Figure 7. Famous face familiarity rating. The combined histogram for each observer adds to 40, the total number of familiar test faces.

It can be seen in Table 1 (see Appendix) that the accepted standard of 0.75 for classification accuracy threshold [42], [43] for the average of all observers is reached at an SNR of 7.74. Thus from Figure 5 we can conclude that 161 in-population and 196 out-of-population eigenfaces are sufficient for recognition. These observations are in agreement with earlier remarks and with Figure 1.

Next we consider the important issue of the relationship between dimensionality and how well faces are stored in memory. Unlike most machine face recognition *training* sets, the human *lifetime database of faces* is drawn from a virtually limitless reservoir of images. We can expect that some familiar faces are better coded for than others, especially since there is a more general *eigenface basis*, or counterpart to it. In addition the number of exemplars can play a role, e.g., the database of a person who watches the news frequently may contain more exemplars of

particular politicians, such as Bill Clinton or George Bush. To address this issue, at the outset we asked observers to subjectively rate their familiarity with 3 different baseline photographs. Once again we emphasize that these images were not used in the second part of the experiment. In Figure 7 it can be seen that not all observers were equally familiar with the faces. Observers 2, 8 and 10 appeared to have a particularly good representation of the familiar faces in memory, as evidenced by their familiarity ratings in Figure 7. The average psychometric function of these three observers is plotted separately in Figure 6, with circles. These observers reached a 0.75 perceptual recognition threshold at a SNR of 7.24, which when converted by Figure 5 yields to an in-population dimensionality of 107 and an out-of-population dimensionality of 124. The dimensionality measure based on observers that have the highest baseline familiarity ratings is significantly lower than the estimate based on the average observer. This indicates that a person's measure of dimensionality might be dependant upon how well these familiar faces are coded in memory, amongst other possibilities.

In Figure 8 we plot the average psychometric function for faces that were in- and out-of-population separately. We can see that there is no significant difference in classification accuracy as a function of *signal-to-noise ratio*, as displayed by the error bars, and therefore there is little reason to regard this as a confounding issue.

VI. DISCUSSION

Factors that could affect our estimates of dimensionality are differences in face image composition such as facial expression, facial hair, lighting, race, sex and presence or absence of glasses. The FERET database is rather heterogeneous, containing images with an assortment of these attributes, and our measure of dimensionality could be influenced by such attributes. In selecting the test faces to use for the psychophysics experiment we endeavored to be consistent in controlling for these factors. All test faces were centrally lit, had neutral facial expression, didn't have glasses or facial hair, were equally split between the sexes, and were Caucasian.

We have determined that on average the dimension of *face space* is in the range of 100-200 eigenfeatures. This estimation was made within a *face space* as parameterized by eigenfaces. In past works the question of how much error the human face recognition system can tolerate within

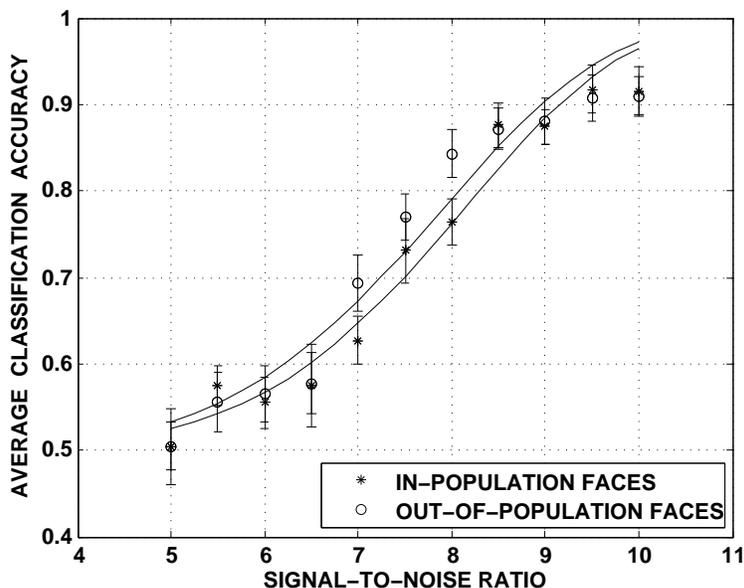


Figure 8. Psychometric function for face classification accuracy of all observers (same as Figure 6), but plotted separately for faces that were in- and out-of-population.

this eigenface framework was touched upon [20], [5], [6], but not really settled. By performing a human psychophysics experiment we have obtained a measure of this error tolerance for the human visual system. However, the human face recognition norm was arrived at within the eigenfunction framework. There is no reason to assume that the human perceptual system uses a Euclidean norm, nor that it uses eigenfunctions. As mentioned earlier the visual system may parse a face into its parts. Thus, it is plausible that evolution has improved on our man-made construction, in which case ours is an upper bound for face dimension.

We also found an indication that the error tolerance of observers may be related to an observer's prior familiarity with the familiar faces, in which case it might be supposed that such observers had somehow incorporated more *training* samples of these faces in their *lifetime database of faces*, or had a better *basis set* in their memory. Observers who were better acquainted with the familiar faces performed better. These observers would appear to provide an indication that there exists a range of dimensions, reflecting a range of talents. This may be a reflection of what is anecdotally referred to as a "good memory for faces." Thus the best thresholds, 124 for out-of-population faces and 107 for in-population faces, might be diminished by better "eigenfunctions" and by more "talented" observers.

It should also be emphasized that our estimates of dimension given above and also those that follow, should be interpreted in the somewhat narrow framework of our experiment. These dimension estimates should in no way be regarded as applying to the entire entity of human faces. It is rather the case that we are presenting a perspective and a guide for dealing with the issue of *face space* dimension, viewed as an innate measure in contrast to a metrical construct.

ACKNOWLEDGMENT

The authors are grateful to Ehud Kaplan, Denis Pelli, and Robert Uglesich for helpful discussions and advice. Marsha Meytlis would also like to express her gratitude to Mount Sinai School of Medicine and its Dean, Diomedes Logothetis, for support of this research project.

REFERENCES

- [1] Zhao, R., Chellappa, A., Rosenfeld, P., and Phillips, P. (2003). Face recognition: A literature survey. *ACM Computing Surveys*, pages 399–458.
- [2] Hjelms, E. and Low, B. (2001). Face detection: A survey. *Computer Vision and Image Understanding*, **83**, 236–274.
- [3] Yang, M., Kriegman, D., and Ahuja N. (2002). Detecting faces in images: A survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **24**, 34–58.
- [4] Chellappa, R., Wilson, C., and Sirohey, S. (1995). Human and machine recognition of faces, a survey. *Proceeding of the IEEE*, **83**, 705–740.
- [5] Sirovich, L. and Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *J. Optical Soc. Am.*, **4**, 519–524.
- [6] Kirby, M. and Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **12**(1), 103–108.
- [7] Turk, M. and Pentland, A. (1991a). Eigenfaces for recognition. *J. Cog. Neurosci.*, **3**, 71–86.
- [8] Turk, M. and Pentland, A. (1991b). Face recognition using eigenfaces. In *Proceedings of Computer Vision and Pattern Recognition, IEEE*, pages 586–591.
- [9] Stewart, G. (1993). On the early history of singular value decomposition. *SIAM Review*, **35**, 551–566.
- [10] Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, Part II: 179–188.
- [11] Brunelli, R., and Poggio, T. (1993). Face recognition: features vs. templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **15**, 1042–1053.
- [12] Swets, D. and Weng, J. (1996). Using discriminant eigenfeatures for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **18**, 831–836.
- [13] Moghaddam, B. and Pentland, A. (1997). Probabilistic visual learning for object representation. *Proceedings of IEEE International Conference on Pattern Analysis and Machine Intelligence*, **19**, 696–710.

- [14] Belhumeur, P., Hespanha, J., and Kriegman, D. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **19**, 711–720.
- [15] Etemad, K. and Chellappa, R. (1997). Discriminant analysis for recognition of human face images. *J. Optical Soc. Am. A*, **14**, 1724–1733.
- [16] Moghaddam, B., Wahid, W., and Pentland, A. (1998). Beyond eigenfaces: Probabilistic matching for face recognition. Nara, Japan. 3rd IEEE International Conference on Automatic Face and Gesture Recognition.
- [17] Bartlett, M., Lades, H., and Sejnowski, T. (1998). Independent component representation for face recognition. *In Proceedings of the SPIE: Conference on Human Vision and Electronic Imaging III*, **3299**, 528–539.
- [18] Tanaka, J. (2001). The entry point of face recognition: evidence for face expertise. *Journal of Experimental Psychology*, **130**, 534–543.
- [19] Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural networks and the bias variance dilemma. *Neural Computation*, **4**, 1–58.
- [20] Penev, P. and Sirovich, L. (2000). The global dimensionality of face space. *In Proc. 4th Int'l Conf. Automatic Face and Gesture Recognition, IEEE CS*, pages 264–270.
- [21] Shakhnarovich, G. and Moghaddam, B. (2004). *Face Recognition in Subspaces*. Springer-Verlag.
- [22] O'Toole, A., Abdi, H., Deffenbacher, K., and Valentin, D. (1993). Low-dimensional representation of faces in higher dimensions of face space. *In J. Optical Soc. Am. A*, **10** pages 405–411.
- [23] Everson, R. and Sirovich, L. (1995). Karhunen-loeve procedure for gappy data. *J. Optical Soc. Am.*, **12**, 1657–1664.
- [24] Graf, A., Wichmann, F., Bulthoff, H., and Scholkopf, B. (2006). Classification of faces in man and machine. *In Neural Computation*, **18** pages 143–165.
- [25] O'Toole, A., Deffenbacher, K., Valentin, D., McKee, K., Huff, D., and Abdi, H. (1998). The perception of face gender. The role of stimulus structure in recognition and classification. *In Memory and Cognition*, **26** pages 146–160.
- [26] Olshausen, B., and Field, D. (2004). Sparse coding of sensory inputs. *In Current Opinion in Neurobiology*, **14** pages 481–487.
- [27] Simoncelli, E. (2003). Vision and the statistics of the visual environment. *In Current Opinion in Neurobiology*, **13** pages 144–149.
- [28] Marcelja, S. (1980). Mathematical Description of the Responses of Simple Cortical Cells. *J. Optical Soc. Am.*, **70**, 1297–1300.
- [29] Daugman, J. (1985). Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Cortical Filters. *J. Optical Soc. Am.*, **2**, 1160–1169.
- [30] Liu, C. and Wechsler, H. (2001). Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model (EMF) for Face Recognition. *IEEE Trans. on Image Processing*, **11**, 467–476.
- [31] Liu, C. and Wechsler, H. (2003). Independent Component Analysis of Gabor Features for Face Recognition. *IEEE Trans. on Neural Networks*, **14**, 919–928.
- [32] Phillips, P., Wechsler, H., Huang, J., and Rauss, P. (1998). The FERET Database and Evaluation Procedure for Face-Recognition Algorithms. *Image and Vision Computing*, **16** pages 295–306.

- [33] Hancock, P., Burton, A., and Bruce, V. (1996). Face processing: Human perception and principal components analysis. *Memory and Cognition*, **24**, 26–40.
- [34] Bradley, A. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, **30**, 1145–1159.
- [35] Provost, F., Fawcett, T., and Kohavi, R. (1998). The case against accuracy estimation for comparing induction algorithms. *Proc. Int'l Conf. Machine Learning*, 445–453.
- [36] Duda, R., Hart, P., and Stork, D. (2000). *Pattern Classification*. John Wiley and Sons. 48–50.
- [37] Konishi, S., Yuille, A., Coughlan, J., and Chun Zhu, S. (2003). Statistical edge detection learning and evaluating edge cues. *IEEE Trans. of Pattern Analysis and Machine Intelligence*, **25**, 57–74.
- [38] Carmichael, O. (2004). Shape-based recognition of wiry objects. *IEEE Trans. of Pattern Analysis and Machine Intelligence*, **26**, 1537–1552.
- [39] Green, D. and Swets, J. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
- [40] Macmillan, N. and Creelman, C. (1991). *Detection theory: A users guide*. Cambridge University Press, New York.
- [41] Pelli, D., and Farell, B. (1994). Psychophysical methods. In: Bass, M., Van Stryland, E., Williams, D., and Wolfe W. (Eds.). *Handbook of Optics*, 2nd ed., I (pp. 29.21-29.13). New York: McGraw-Hill.
- [42] Wolfe, J., Kluender, K., and Levi, D. (2006). *Sensation & Perception*. Sinauer Associates, Inc.
- [43] Quick, R. (1974). A vector magnitude model of contrast detection. *Kybernetik*, **16**, 65–67.
- [44] Strang, G. (1988). *Linear Algebra and its Applications*. Wellesley-Cambridge Press.

APPENDIX I
EIGENFACES

A face which has been normalized as described in 2.2.1 can be represented by image intensity values $f(\mathbf{x})$ where \mathbf{x} is the pixel location and V is the total number of pixels in the image. We consider the ensemble of T pictures $\{f(t, \mathbf{x})\}_{t \in T}$ with its SVD representation [43] denoted by

$$f(t, x) = \sum_{n=1}^M a_n(t) \sigma_n \psi_n(\mathbf{x}), \quad (1)$$

where $M = \min(T, V)$ is the rank of the ensemble. Following conventional notation, we have the orthonormality conditions.

$$(a_n, a_m)_t = \sum_t a_n(t) a_m(t) = \delta_{nm} \quad = \quad \sum_x \psi_n(\mathbf{x}), \psi_m(\mathbf{x}), \quad (2)$$

The *weighting* constants σ_n , are referred to as the *singular values*.

One can easily show that

$$((f(t, \mathbf{x}), f(t, \mathbf{y}))_t, \psi_n(\mathbf{y}))_y = \lambda_n \psi_n(\mathbf{x}) \quad (3)$$

where $\lambda_n = \sigma_n^2$ is said to be the eigenvalue and

$$((f(t, \mathbf{x}), f(s, \mathbf{x}))_x, a_n(s)) = \lambda_n a_n(t) \quad (4)$$

The reader should be cautious in comparing these deliberations, with their counterparts in [23] since the latter uses an unconventional notation.

The *pictures* $\{\psi_n\}$ are the eigenfaces of the ensemble. For a given dimensionality N the image reconstruction and the image error are given by

$$f_N^{rec} = \sum_{n=1}^N a_n \sigma_n \psi_n \quad \text{and} \quad f_N^{err} = f - f_N^{rec}, \quad (5)$$

respectively, The *signal to noise ratio* (SNR) is defined as

$$SNR = \log_2(\|f\|^2 / \|f_N^{err}\|^2) = \log_2(\sum_{n=1}^M a_n^2 / \sum_{n=N+1}^M a_n^2) \quad (6)$$