Perception, 1997, volume 26, pages 75–84

Sex classification is better with three-dimensional head structure than with image intensity information

Alice J O'Toole
School of Human Development, GR4.1 The University of Texas at Dallas, Richardson,
TX 75083-0688, USA; e-mail: otoole@utdallas.edu.
Thomas Vetter, Nikolaus F Troje, Heinrich H Bülthoff
Max-Planck-Institut für biologische Kybernetik, Spemannstrasse 38, D 72076 Tübingen, Germany;
e-mail: vetter@mpik-tueb.mpg.de
Received 17 April 1996, in revised form 22 November 1996

Abstract. The sex of a face is perhaps its most salient feature. A principal components analysis (PCA) was applied separately to the three-dimensional (3-D) structure and graylevel image (GLI) data from laser-scanned human heads. Individual components from both analyses captured information related to the sex of the face. Notably, single projection coefficients characterized complex differences between the 3-D structure of male and female heads and between male and female GLI maps. In a series of simulations, the quality of the information available in the 3-D head versus GLI data for predicting the sex of the face has been compared. The results indicated that the 3-D head data supported more accurate sex classification than the GLI data, across a range of PCA-compressed (dimensionality-reduced) representations of the heads. This kind of dual face representation can give insight into the nature of the information available to humans for categorizing and remembering faces.

1 Introduction

In recent computational work, principal components analysis (PCA)(1) has been applied widely to analyzing the information in images of human faces. The information captured by PCA has been shown to be reliable, in purely computational terms, for recognizing faces (O'Toole et al 1993; Turk and Pentland 1991) and for classifying them by race and sex (O'Toole et al 1991). Additionally, different components of PCA have been shown to relate reliably to human performance on some of these same tasks (O'Toole et al 1994).

The human head, however, is a complex three-dimensional (3-D) object with a characteristic shape and an associated surface image (ie surface marking information) from which individual faces, and categories of faces, vary. In the present study, we applied PCA to a more complete physical model of the human face than is available from a facial image. This model included both a 3-D structure and a component based on graylevel image (GLI). Recently, Hancock et al (1996) highlighted the importance of considering separately image and 'shape-based' components in modeling human performance on a face recognition task. They used images of faces to create a 'shape-free' face by hand-selecting facial landmarks and morphing the faces to an average shape (Craw and Cameron 1991). They then applied PCA separately to the shape-free image of the face and to the deviation of the individual faces from the average shape. Their results indicated that different components of human recognition performance related to the structure and GLI-based information.

A more explicit representation of 3-D head information than that available from morphing between 2-D images can be obtained from laser scan technology, which provides both the 3-D coordinates of the head (structure) and a wrap-around image

---

(1) For related analyses, see also Pearson (1901), Hotelling (1933), Karhunen (1946), and Loève (1955).

(GLI) that maps point for point onto the head surface.[(2)] While this representation lacks the advantageous feature of established correspondences between a reduced set of facial features [eg 34 facial landmarks such as the corners of the mouth and eyes (Hancock et al 1996)], it has the important advantage of including much more detailed information about subtle variations in 3-D structure than that retained with a relatively small set of interfacial landmark distances. The purpose of the present study was to compare the overall utility of the structure versus GLI information for predicting the sex of a face. We did this in two steps. First, we applied PCA separately to the head structure and GLI data taken from a large number of male and female faces. Second, we carried out a series of simulations to compare the quality of the information available in the head structure versus GLI data for categorizing the faces by sex.

2 Principal component analysis

2.1 *Stimuli*

Laser scans (CyberwareTM) of 130 heads of young adults (65 male and 65 female) were used as stimuli. The subjects were scanned wearing bathing caps, which were removed digitally prior to the PCA, consequently eliminating most of the hair region of the head. Additionally, further preprocessing of the head scans was done by making a vertical cut behind the ears, and a horizontal cut to remove the shoulders.

The laser scans provide two kinds of information. First, the head structure data consisted of the lengths of 512 × 512 radii from a vertical axis in the middle of the subject's head to 'sample' points on the surface of the head. This is a cylindrical representation of the head surface, with surface points sampled at 512 equally-spaced angles around the circular slices of the cylinder, and at 512 equally spaced vertical distances along the long axis of the cylinder. The radii in our data set varied from a minimum of 4 cm (neck area) to a maximum of 12 cm (nose area). The resolution of the scanner was 16 μm.

The graylevel data consisted of 512 × 512 256-graylevel image intensities that mapped point for point onto the 3-D sample points of the head. The 3-D head models and GLI maps were aligned so that regions around the nose and eyes roughly coincided spatially. In both cases, the mean of the data was subtracted prior to the application of the PCA.

In summary, the head structure data provide an object-centered representation of the 3-D coordinates of a head. This representation is, therefore, inherently independent of viewpoint. Likewise, the 'wrap-around' GLI maps point for point onto the 3-D head structure and so can be considered, also, as a viewpoint independent representation.

2.2 *Analysis procedure and results*

Principal components analysis was applied separately to the structure and GLI data. Figure 1 shows the average head structure in the top row, and the two eigenvectors (the first and the sixth), that related most reliably to face sex for the head data in the second and third rows, respectively. The eigenvectors are displayed by adding (subtracting) them to (from) the mean head. In both cases, distinctive shape differences between male and female heads can be captured simply by adding versus subtracting these single components to/from the average head. This demonstration replicates a similar finding for images (O'Toole et al 1993) and indicates that, in this representation

---

[(2)] We use the term 'graylevel image' (GLI) to refer to this complete wrap-around image. This wrap-around image is distinct from a view-based image of the head (ie an image taken from a particular view), since it contains image intensity information about the entire head surface. In computer graphics, this graylevel wrap-around image is referred to as a 'texture map' in contrast to the pure geometry of an object. We do not use the term 'texture' here, because in the computer vision literature it generally connotes a repetitive pattern.

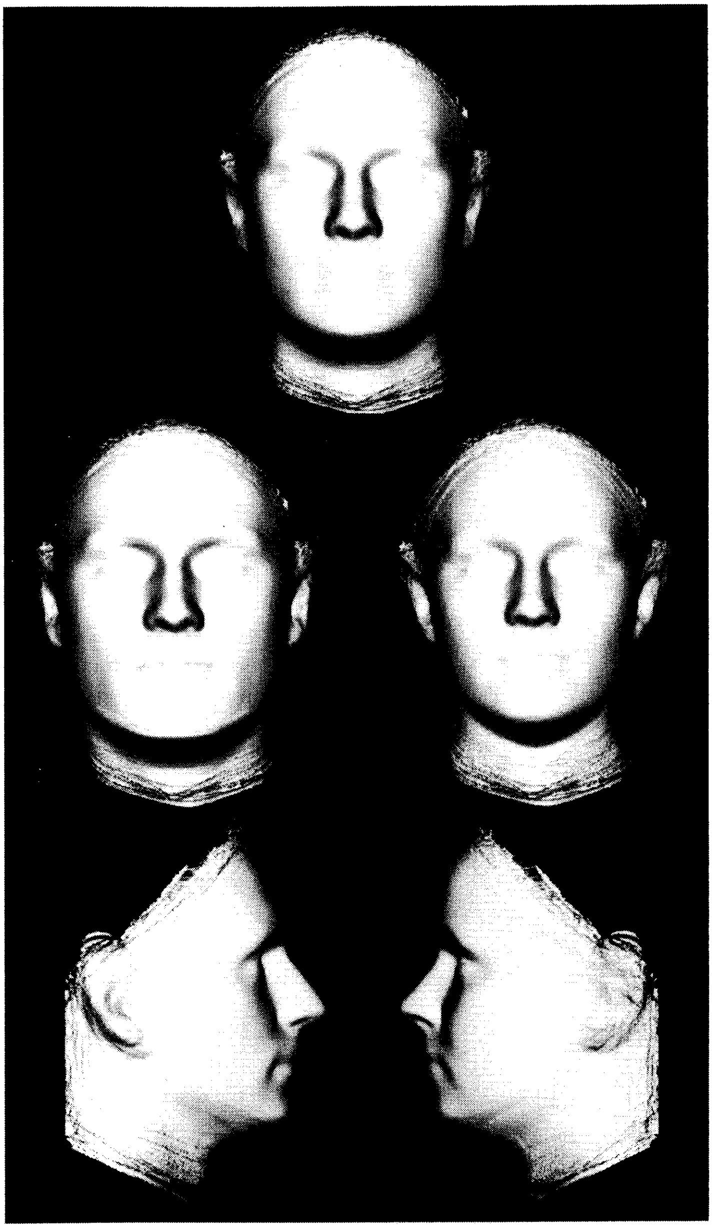

Figure 1. *Row 1:* average of 130 heads (65 male and 65 female).
Row 2: left head constructed by adding the shape defined by the first eigenvector to the average head; right head constructed by subtracting the shape defined by the first eigenvector from the average head. More precisely, the display shows the average head combined linearly with $\pm 2\sigma$ times the first eigenvector, where σ is the standard deviation of the projection coefficients on this eigenvector computed across all faces. It is interesting to note that a single coefficient describes a compelling and complex structural transformation between male and female heads, including aspects of brow protrusion and jaw shape.
Row 3: the same as row 2 but with the sixth eigenvector, viewed from the side where it is easiest to see that it captures differences in nose size and shape between male and female heads.

as well, the facial characteristics related to face sex explain relatively large proportions of variance in the face set. The first eigenvector captured global differences in the shapes of male versus female heads, including the structure of the jaw and brow. The sixth eigenvector captured differences in nose/brow size and shape for male and female heads. More formally for the prediction of face sex, point biserial correlations(3) between the projection of individual faces onto the first and sixth eigenvectors and the

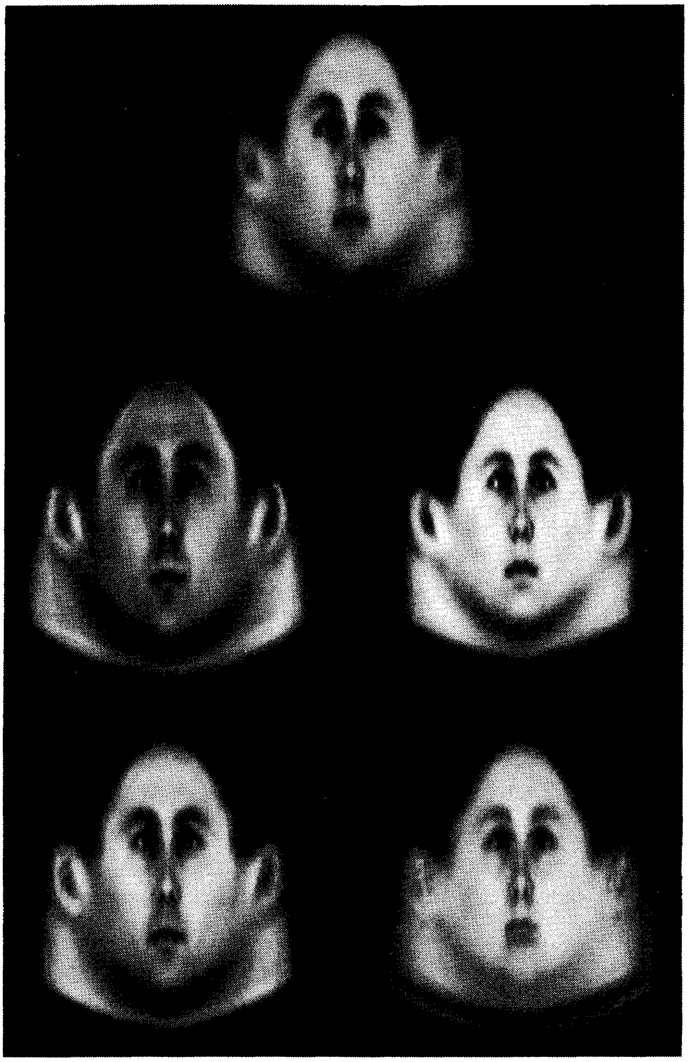

Figure 2. This figure duplicates the illustration in figure 1 for the graylevel image data, with the use of the second and fifth eigenvectors. The second eigenvector is easily interpretable as contrasting the characteristic graylevel image shadowing of male facial hair against its absence in female faces and also captures some differences in the size and shape of the image-based data, including a difference in the neck width; the fifth eigenvector also contains information about facial hair shadowing patterns and overall differences in the size/shape of the outer contour.

---

(3) A point biserial correlation is simply a standard correlation computed between two variables, one measured on a continuous interval scale and a second measured on a dichotomous nominal scale, eg male and female, defined arbitrarily as 0 and 1 (Klugh 1974). This correlation can be interpreted in the same way as a Pearson *r*.

face sex (defined as 1 or 0) were statistically reliable ($r = 0.66$, $p < 0.001$; $r = 0.30$, $p < 0.001$, respectively).(4)

For GLI data also, the PCA revealed several eigenvectors that contrasted male versus female GLIs. The two eigenvectors that most related to face sex were the second and fifth. A point biserial correlation between face sex and the projection coefficient for the second eigenvector was statistically reliable ($r = 0.71$, $p < 0.0001$). For the fifth eigenvector, the correlation was reliable at the 0.01 level ($r = 0.25$, $p < 0.01$), though not at the more conservative 0.001 level appropriate for the number of correlations we carried out. The mean GLI map is displayed in the top row of figure 2, and the second and fifth eigenvectors appear in the second and third rows, respectively, again adding (subtracting) them to (from) the mean GLI map. The second eigenvector is easily interpretable as contrasting the characteristic GLI shadowing of male facial hair against its absence in female faces. Additionally, this eigenvector captures some differences in the relative size and shape of the head outline, especially the width of the neck. The fifth eigenvector also seems to capture aspects of the pattern of facial hair shadowing and the size/shape of the outer contour. It is perhaps worth noting that because this GLI representation can capture information about size/shape variations of the outer contour of the heads, at least some information it contains should be redundant with the head structure data. The technique developed by Craw and Cameron (1991) for normalizing head shape might be a way to eliminate some of this redundancy, though, as noted previously, some of the subtle details of the 3-D shape would be lost. Other very recently developed techniques for putting face images into pixel-wise correspondence might be exploited in the future for achieving both of these ends simultaneously (cf Poggio and Beymer 1995; Vetter and Poggio 1996).

3 Sex classification simulations

Given that some of the individual components appeared to capture salient categorical information about faces, and that individual correlations provide only a preliminary look at the reliability and generality of the relationship between projection coefficients and face sex, we carried out a more formal analysis as follows. The PCA enables a dimensionally reduced representation of individual heads/GLIs in terms of their projection coefficients. The utility of different low-dimensional representations for specifying categorical information about faces, such as sex, can be analyzed in this abbreviated representation much more easily than in full GLI head data. Abdi et al (1995) compared raw image and PCA preprocessed image representations and showed that PCA preprocessing had tangible advantages over a purely image-based representation in terms of the generalizability of sex information it captured when used as input to simple sex classification algorithms. We applied the methods developed by Abdi et al to compare the quality of the GLI versus structure information in human faces for classifying faces by sex. Three simulations were carried out. In the first we analyzed the head structure data, and in the second we analyzed the GLI map data. In a third simulation we combined the two representations for predicting face sex.

3.1 *Methods*

3.1.1 *Stimuli.* Separate representations of the stimuli in terms of their head structure versus GLI data were created. In both cases, each stimulus was represented by its projection coefficients onto the eigenvectors extracted from the PCA of the appropriate data, ie either the 3-D structure data or GLI map data.

(4) It is worth noting that the alpha level for statistical significance when carrying out multiple correlations should be defined as 0.05 (the standard value used in most applications of this sort) divided by the number of correlations. In the present application, we have considered only the first 50 coefficients, making this level 0.001.

3.1.2 *Procedure.* The derived face codings were used to train simple perceptron sex classification networks, which were then tested for the generalizability of the information learned for classifying novel faces by sex. The perceptron was chosen since previous work comparing several sex classification network models of face images indicated that it performed roughly equivalently to a radial basis function (RBF) network (Abdi et al 1995). The perceptron was chosen over the RBF network since, mathematically and conceptually, it is the simpler of the two models. Further, from a mathematical point of view, it can be shown to optimize the placement of a hyperplane within the representation space to separate male and female faces in the learned set. As we will show shortly, the model performs nearly perfectly with a relatively small set of projection coefficients, indicating that this simple linear classifier was sufficient for the task.

To test the generalizability of the sex classification performance of the network, we applied a standard 'jackknife' procedure, which operated as follows. A perceptron model was trained with all possible combinations of the $n-1$ faces, where n is the number of available faces (130 in this case). Each trained network was then used to classify the single unlearned face of the set. The error rate was taken as the percentage of these novel faces correctly classified across the 130 perceptrons in each subspace simulation.

Additionally, we varied the size of the subspace, to determine the minimum subspace required in the two representations to achieve maximally general sex classification, as was done with images previously by Abdi et al (1995).

3.2 *Results*

3.2.1 *Separate head structure and graylevel image (GLI) representations.* The results of these simulations appear in figure 3, plotted as the generalization accuracy of the sex classification network for the GLI and structure-based representations as a function of the size of the subspace. Three points are worth noting. First, both the structure-based

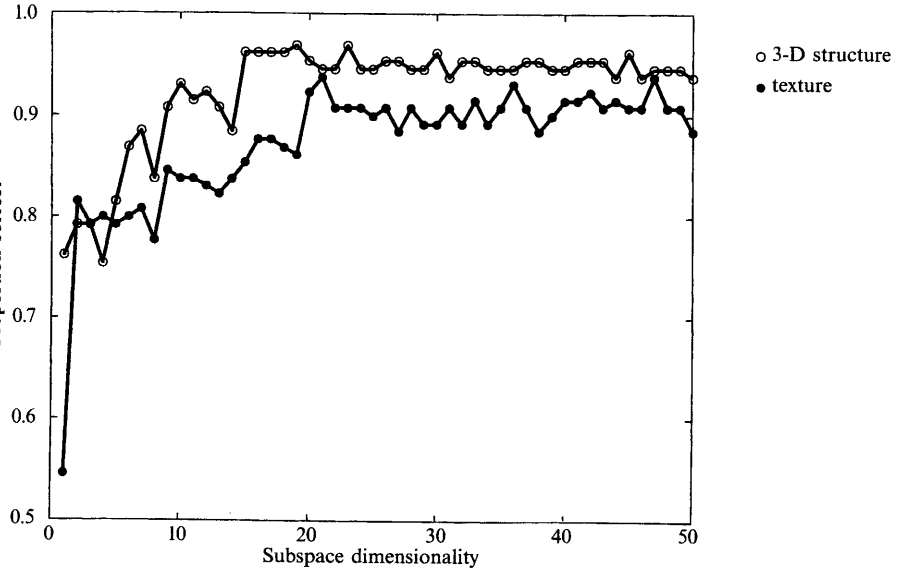

Figure 3. The generalization accuracy of the sex classification network for the graylevel image and structure-based representations as a function of the size of the subspace. In general, three-dimensional structure data supported more accurate sex classification than did graylevel image data. Though interestingly, one exception occurred due to the very strong predictive power of the second eigenvector of the graylevel image analysis, which enabled better classification in this very-low-dimensional subspace.

and GLI-based representations provided reliable information for determining face sex. Performance for even very small subspaces was well above chance. Second, across nearly the entire range of subspaces tested, the structure data supported better sex classification than the GLI data. For the GLI data, a peak generalization performance of 93.8% correct sex generalization was achieved with a minimum subspace of 20 projection coefficients—a rate comparable to that reported previously with a perceptron and images with hair cues to sex present, but better than that reported previously with images more comparable to our stimuli [ie excluding hair (Abdi et al 1995)]. While it is difficult to make precise performance comparisons with other previous models of automated image-based sex classification (owing to important differences in stimulus sets, face representations, and classification algorithms), in general, the present results for GLI data compare quite favorably with these other models (Brunelli and Poggio 1993; Burton et al 1993; Fleming and Cottrell 1990; Golomb et al 1991; Gray et al 1995; see the last paper for more precise model comparisons). For the 3-D head data in the present study, the results showed generalization performance that was even better than that obtained with GLI, peaking at 96.9% correct with a minimum subspace dimensionality of 17 projection coefficients.

3.2.2 *Combined head structure and graylevel image (GLI) representations.* In a third simulation, we varied the subspace dimensionality using a face representation that combined the information from the 3-D head structure and GLI data. We did this to assess the extent to which the structure and image-based data may be capturing nonredundant information about the faces. We increased the subspace, one coefficient at a time, alternating between the 3-D structure and GLI data. We found a peak performance of 97.7% correct with a minimum subspace of 32 coefficients (half structure-based and half GLI-based). This performance was better than that obtained with either the structure or image data alone and suggests the structure and image representations may be capturing nonredundant information for sex classification. Some caution is required, however, in comparing the minimum dimensionality for peak performance in this combined representation with the minimum dimensionalities found for the structure or image data alone. This is due to the fact that there are a number of reasonable ways to combine the coefficients, which may give somewhat different results. Thus, it is difficult to make absolute statements about minimum dimensionality for the combined data case.

Finally, it is worth noting that in all cases the most reliable information for sex classification was found in eigenvectors with relatively larger eigenvalues, a finding consistent with previous work on images (Abdi et al 1995; O'Toole et al 1993). The nearly perfect classification performance of the perceptron at its peak indicates that male and female faces, represented with only a relatively small number of projection coefficients onto the structure-based or image-based principal components, can be separated almost completely by a hyperplane dividing these low-dimensional representations (cf also Gray et al 1995, for a comparison between a perceptron and a multilayer perceptron for a sex classification).

4 General discussion

The results of the present study indicated that the PCA of the separated structure and GLI data captured information relevant for determining the sex of a face. Additionally, the quality of the sex-related information differed in structure versus GLI representations of the faces. This finding highlights the importance of considering the nature of the information available for the different tasks we carry out with faces. Previous work on human perception of purely shape-based data on faces (again from laser scans) has indicated generally that human observers find it very difficult to extract the information useful for identifying a face from this representation (cf Bruce et al 1991; Bruce and Langton 1994; though the data from both studies need to be interpreted with caution given the very small number of head scans used, varying between 3 and 8 for the recognition and identity tasks reported). Consistent with this finding, Troje and Bülthoff (1996), using a large number of heads, showed that human observers matched the identity of depth-rotated (ie view-changed) heads with GLI data more quickly than of heads without GLI data.

The difficulty human observers experience recognizing faces from laser scanned heads, however, does not suggest that 'head/face shape', presented in a more natural way, is not used by human observers for recognizing faces. For example, Hancock et al (1996) showed that at least one component of human face recognition performance, the hit rate, was best predicted by separate analyses of the 2-D shape (ie the configuration of landmark feature points) and GLI information, subsequently brought together. Further, comparing recognition performance for own-race versus other-race faces, O'Toole et al (1994) found that the distinctiveness of face shape related strongly to the recognizability of own-race faces. An additional important component of face recognizability was the presence or absence of local, distinctive features. Interestingly, the recognizability of 'other-race' faces (specifically, Japanese faces for Caucasian observers), was more related to small-scale distinctive features than to the global shape-based distinctiveness of faces. O'Toole et al (1994) speculated that this might be due to the generally lesser experience we have with the complex 3-D structure of faces of another race.

For sex classification, there is good evidence from several sources that human observers also make use of a very broad range of cues, including aspects of the structure/shape of the face and a variety of image-intensity-based cues. For example, Bruce et al (1993) showed accuracy decrements in sex classification using 'graylevel imageless' laser scanned heads. They surmised that these imageless heads might have been missing important cues for sex classification such as eyebrow bushiness and beard stubble. Additionally, however, they found that sex classification performance declined with manipulations of (a) face inversion, which they interpreted in terms of the importance of the global configuration, and (b) photographic negation, which they interpreted in terms of the importance of 3-D shape-from-shading information. Using a more comprehensive combination of inversion and negation with laser scanned heads and photographs, Bruce and Langton (1994) found rather different patterns of performance decrements for identification and sex classification tasks—again stressing the importance of considering the kinds of information available for different tasks. For identification, their results indicated rather less importance of the shape-from-shading information relative to other cues. For sex classification, however, their results were again consistent with the importance of shape-from-shading information (see also Burton et al 1993).

Finally, to our knowledge, the role of viewpoint has not been considered explicitly in psychophysical studies of sex classification performance. Viewpoint may be an important variable in determining the quality of information useful for sex classification in the image intensity versus 3-D structure data we consider. As noted previously, the representations we have analysed in this study are viewpoint invariant, and hence some caution is required in comparing the simulation data based on these viewpoint-invariant representations with psychological data based on a single view of a face.

More generally, the extent to which humans rely on 3-D structure information derived from objects/faces as opposed to 2-D image-based information is currently a much-debated point in the psychology literature (Biederman 1987; Bülthoff and Edelman 1992) and is being investigated actively by neuroscientists interested in the neurophysiological substrates of objects and face recognition (Logothetis et al 1995; Perrett et al 1992). At present, we consider the application of PCA to different representations of faces as a very useful tool for quantifying the information available in these representations. Applied to 3-D heads, the analysis provides a useful low-dimensional quantification of the information that would be available to human observers were they able to mentally generate such representations. Likewise, the application of this analysis of GLI data versus view-dependent images can also provide a baseline measure of the information inherent in these different representations. The comparison of human performance on simple tasks like sex classification and recognition, with the quality of information available in different representations (3-D structure, complete GLIs, and single view-based representations) can serve as a powerful tool for learning about the nature of human representations of objects.

Acknowledgements. The order of the first two authors is arbitrary. Many thanks are due to Harald Volz for work on stimulus processing and display tools and to Guy Wallis for useful advice on the project. We thank also B Edelman and P Assmann and two anonymous reviewers for helpful comments on a previous version of this manuscript. Alice O'Toole gratefully acknowledges the support of the Alexander von Humboldt Foundation and the hospitality of the Max-Planck-Institut für biologische Kybernetik, Tübingen, Germany, as well as support from NIMH grant 1R29MH5176501A1.

References

Abdi H, Valentin D, Edelman B, O'Toole A J, 1995 "More about the difference between men and women: Evidence from linear neural networks" *Perception* **24** 539 – 562

Biederman I, 1987 "Recognition by components: A theory of human image understanding" *Psychological Reviews* **94** 115 – 147

Bruce V, Burton A M, Hanna E, Healey P, Mason O, Coombes A, Fright R, Linney A, 1993 "Sex discrimination: How do we tell the difference between male and female faces?" *Perception* **22** 131 – 152

Bruce V, Healey P, Burton M, Doyle T, Coombes A, Linney A, 1991 "Recognising facial surfaces" *Perception* **20** 755 – 769

Bruce V, Langton S, 1994 "The use of pigmentation and shading information in recognizing the sex and identities of faces" *Perception* **23** 803 – 822

Brunelli R, Poggio T, 1993 "Caricatural effects in automated face perception" *Biological Cybernetics* **69** 235 – 241

Bülthoff H H, Edelman S, 1992 "Psychophysical support for a two-dimensional view interpolation theory of object recognition" *Proceedings of the National Academy of Sciences of the USA* **89** 60 – 64

Burton M, Bruce V, Dench N, 1993 "What's the difference between men and women? Evidence from facial measurement" *Perception* **22** 153 – 176

Craw I, Cameron P, 1991 "Parameterising images for recognition and reconstruction", in *Proceedings of the British Machine Vision Conference* Ed. P Mowforth (London: Springer) pp 367 – 370

Fleming M, Cottrell G W, 1990 "Categorization of faces using unsupervised feature extraction" *Proceedings of the International Joint Conference on Neural Networks IJCNN-90* volume 2 (Ann Arbor, MI: IEEE Neural Networks Council) pp 65 – 70

Golomb B A, Lawrence D T, Sejnowski T J, 1991 "Sexnet: A neural network identifies sex from human faces", in *Advances in Neural Information Processing Systems* Volume 3 Eds R P Lippman, J Moody, D S Touretsky (San Mateo, CA: Morgan Kaufmann) pp 572 – 577

Gray M, Lawrence D T, Golomb B A, Sejnowski T J, 1995 "A perceptron reveals the face of sex" *Neural Computation* **7** 1160 – 1164

Hancock P J B, Burton A M, Bruce V, 1996 "Face processing: human perception and principal components analysis" *Memory & Cognition* **24** 26 – 40

Hotelling H J, 1933 "Analysis of statistical variables into principal components" *Educational Psychology* **24** (September) 417 – 441, 498 – 520

Karhunen K, 1946 "Über lineare Methoden in der Wahrscheinlichkeitsrechnung" *Annales Academiae Scientarum Fennicae, Series A1, Mathematica Physica* **37** 1 – 79

Klugh H E, 1974 *Statistics: The Essentials for Research* 2nd edition (New York: Wiley)

Loève M M, 1955 *Probability Theory* (Princeton, NJ: Von Norstrand)

Logothetis N K, Pauls J, Poggio T, 1995 "Shape recognition in the inferior temporal cortex of monkeys" *Current Biology* **5** 552 – 563

O'Toole A J, Abdi H, Deffenbacher K A, Bartlett J, 1991 "Classifying faces by race and sex during an autoassociative memory trained for recognition" *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* Eds K J Hammond, D Gentner (Hillsdale, NJ: Lawrence Erlbaum) pp 847–851

O'Toole A J, Abdi H, Deffenbacher K A, Valentin D, 1993 "Low dimensional representation of faces in higher dimensions in the face space" *Journal of the Optical Society of America A* **10** 405–411

O'Toole A J, Deffenbacher K A, Valentin D, Abdi H, 1994 "Structural aspects of face recognition and the other-race effect" *Memory & Cognition* **22** 208–224

Pearson K, 1901 "On lines and planes of closest fit to systems of points in space" *Philosophical Magazine* 6th Series **2** 557–572

Perrett D I, Hietanen J K, Oram M W, Benson P J, 1992 "Organisation and functions of cells responsive to faces in the temporal cortex" *Philosophical Transactions of the Royal Society London B* **335** 23–30

Poggio T, Beymer D, 1995 "Learning networks for face analysis and synthesis", in *Proceedings of the International Workshop on Face and Gesture Recognition* Ed. M Bichsel (Zürich: University of Zürich Multimedia Laboratory) pp 160–165

Troje N F, Bülthoff H H, 1996 "Face recognition under varying pose: The role of graylevel image and shape" *Vision Research* **36** in press

Turk M, Pentland A, 1991 "Eigenfaces for recognition" *Journal of Cognitive Neuroscience* **3** 71–86

Vetter T, Poggio T, 1996, "Image synthesis from a single example image" *Proceedings of the European Conference on Computer Vision* (Cambridge, UK: Cambridge University Press) pp 652–659