Max-Planck-Institut
für biologische Kybernetik
Arbeitsgruppe Bülthoff

Spemannstraße 38 • 72076 Tübingen • Germany

Technical Report No. 23 December 10, 1995

The role of shape and texture information in sex classification

Alice J. O'Toole, Thomas Vetter, Heinrich H. Bülthoff & Nikolaus F. Troje

Abstract

The sex of a face is perhaps its most salient feature. A principal components analysis (PCA) was applied separately to the three-dimensional structure and texture 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 structural differences between three-dimensional male and female heads and between male and female texture maps. In a series of simulations, we compared the quality of the information available in the head versus texture data for predicting in the sex of the face. The results indicated that the three-dimensional structural data supported more accurate sex classification than the texture 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.

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. A.O. gratefully acknowledges the support of the Alexander von Humboldt Foundation and the hospitality of the Max Planck Institut für biologische Kybernetik, Tübingen, Germany.

This document is available as `/pub/mpi-memos/TR-23.ps.Z` via anonymous ftp from `ftp.mpik-tueb.mpg.de` or from the World Wide Web, `http://www.mpik-tueb.mpg.de/projects/TechReport/list.html`.

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, Abdi, Deffenbacher & Valentin, 1993; Turk & Pentland, 1991) and for classifying them by race and sex (O'Toole, Abdi, Deffenbacher & Bartlett, 1991). Additionally, different components of PCA have been shown to relate reliably to human performance on some of these same tasks (e.g., O'Toole, Deffenbacher, Valentin, & Abdi, 1994). The human head, however, is a complex three-dimensional object with a characteristic shape and texture, 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 three-dimensional structure- and texture-based component. Recently, Hancock, Burton and Bruce (in press) highlighted the importance of considering separately texture 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 & Cameron, 1991). They then applied PCA separately to the shape-free texture map 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 texture-based information.

A more explicit representation of three-dimensional head information than that available from morphing between two-dimensional images can be obtained from laser scan technology, which provides both the three-dimensional coordinates of the head (structure) and a wrap-around image (texture) 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 (34 facial landmarks such as the corners of the mouth and eyes, Hancock et al.), it has the important advantage of including much more detailed information about subtle variations in three-dimensional structure than that retained with a relatively small set of inter-facial landmark distances. The purpose of the present study was to compare the overall utility of the structure versus texture information for predicting the sex of a face. We did this in two steps. First, we applied PCA separately to the head structure and texture 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 texture data for categorizing faces by gender.

[1] For related analyses, see also Pearson, 1901; Hotelling, 1933; Karhunen, 1946; Loéve, 1955.

[2] We use the term "texture" to refer to this complete wrap-around image, which is distinct from view- based images of the heads.

1.1 PCA Analysis

1.1.1 Methods

Stimuli. Laser scans (Cyberware^{TM}) 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 pre-processing of the heads 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 512x512 radii from a vertical axis centered in the middle of the subject's head to the head surface. This provides a representation of the three-dimensional coordinates of a head. The texture data consisted of a 512x512 gray level image that maps point-for-point onto the three-dimensional head scan. The three-dimensional head models and texture 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.

Analysis Procedure and Results. Principal components analysis was applied separately to the structure and texture data. Figure 1 shows the average head in the top row, and the two eigenvectors (the first and 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 (cf., Figure 2 for several views of the first eigenvector demonstration). This demonstration replicates a similar finding for images (O'Toole et al., 1993) and indicates that in this representation 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 predition of face sex, point-by-serial correlations between the projection of individual faces onto these eigenvectors and the face sex (defined as 1 or 0) were statistically reliable ($r = .66$, $p < .0001$; $r = .30$, $p < .001$, respectively).

For the texture data also, the PCA revealed several eigenvectors that contrasted male versus female face textures. The two eigenvectors that most related to face sex were the second and fifth. Again, point-by-serial correlations between projection coefficients and sex were statistically significant ($r = .71$, $p < .0001$; $r = .25$, $p < .01$; respectively). The mean texture map is displayed in the top row of Figure 3, and the second and fifth eigenvectors appear in the second and third rows, respectively, again adding (subtracting) them to (from) the mean texture map. The second eigenvector is easily interpretable as contrasting the characteristic texture shadowing of male facial hair against its absence in female faces. The fifth eigenvector seems to reflect overall differences in the size/shape of the outer contour, with an interesting relative size difference especially apparent for neck width.

1.2 Gender Classification Simulations

Given that some of the individual components appeared to capture salient categorical information about faces, we carried out a more formal analysis as follows. The PCA enables a dimensionally-reduced representation of individual heads/textures in terms of their projection coefficients. The utility of different low dimensional representations for specifying categorical information about faces, such as gender, can be analyzed in this abbreviated representation much more easily than in full texture/head data. Abdi, Valentin, Edelman and O'Toole (1995) compared raw image and PCA-pre-processed image representations and showed that PCA pre-processing 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 texture versus structure information in human

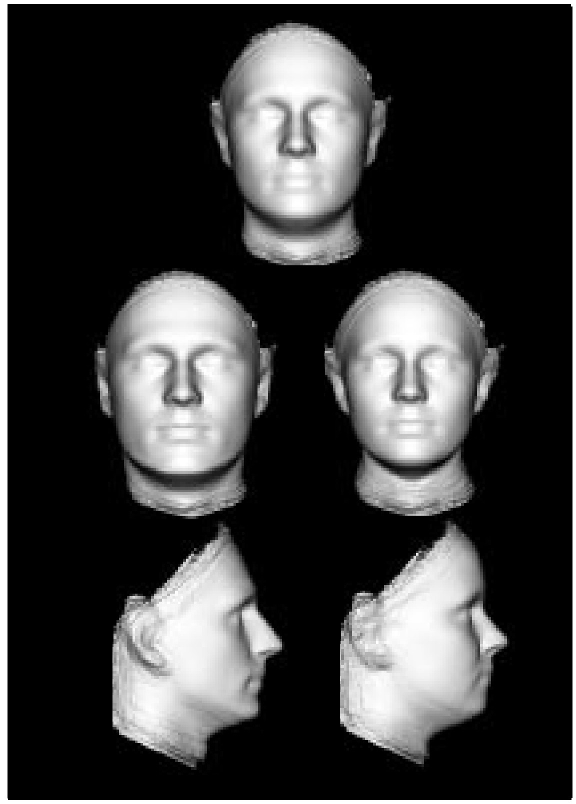

Figure 1: Top center: average of 130 heads (65 male and 65 female). Row 1: left head constructed by adding the shape defined by the first eigenvector to the average head; right head constructed by subtracting the shape defined by the first eigenvector from the average head. More precisely, the display 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 2: the same as Row 1 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.

faces for classifying faces by sex. Two simulations were carried out. In the first, we analyzed the head structure data and in the second we analyzed the texture map data.

1.2.1 Methods

Stimuli. Separate representations of the stimuli in terms of their head structure versus texture 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, i.e., either the three-dimensional structure data or texture map data.

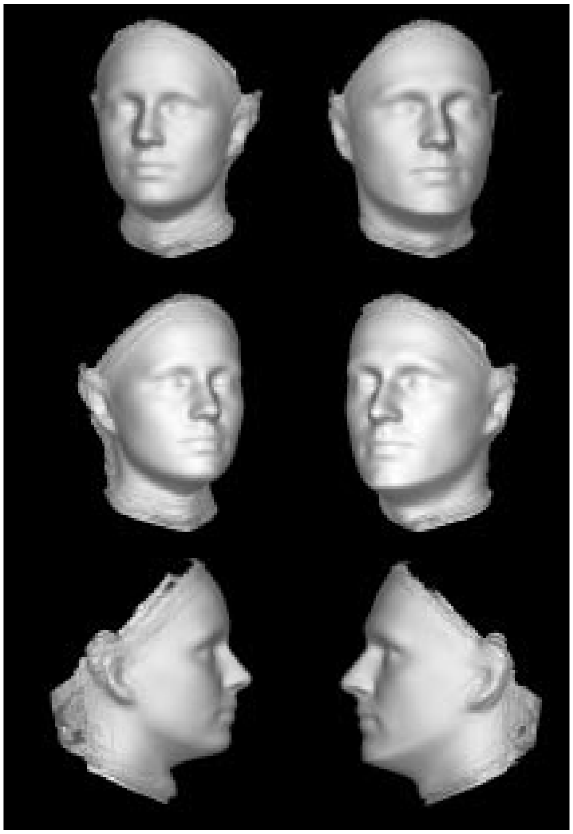

Figure 2: Several views of the male-female contrast demonstration using the first eigenvector for shape.

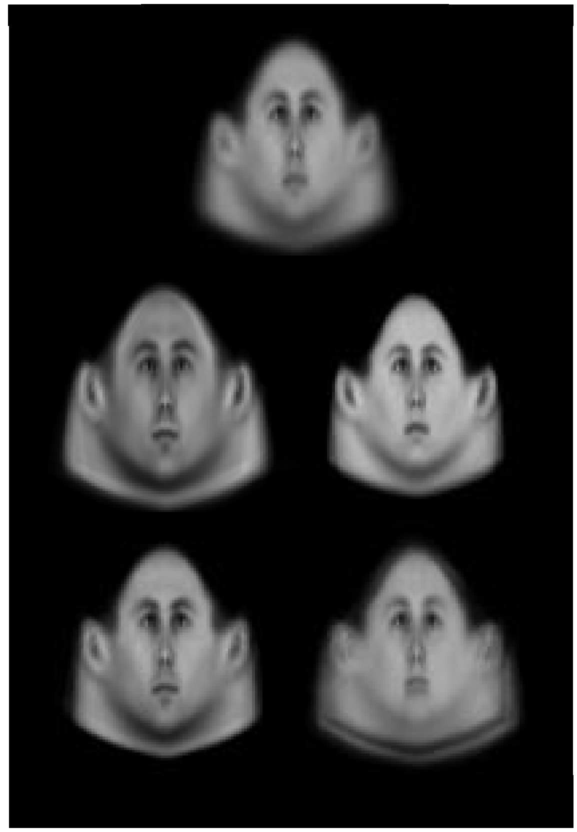

Figure 3: This figure duplicates the illustration in Fig. 1 for the texture data, using the second, and fifth eigenvectors. The second eigenvector is easily interpretable as contrasting the characteristic texture shadowing of male facial hair against its absence in female faces; the fifth eigenvector seems to reflect overall differences in the size/shape of the outer contour, with an interesting relative size difference especially apparent for neck width.

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, both of which outperformed all other models tested on a generalization sex classification task (Abdi et al., 1995). The perceptron was chosen over the RBF network since it is the simpler of the two models. To test the generalizability of the sex classification performance of the network, we applied a standard "jack-knife" 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).

Results. The results of these simulations appear in Figure 4, plotted as the generalization accuracy of the sex classification network for the texture and structure- based representations as a function of the size of the subspace. Three points are worth noting. First, both the structure- and texture-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 texture data. For the texture data, a peak generalization performance of 93.8 percent 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 (i.e., 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 (due to important differences in stimulus sets, face representations, and classification algorithms), in general, the present results for texture data compare quite favorably with these other models (Brunelli & Poggio, 1993; Burton, Bruce, & Dench, 1993; Fleming & Cottrell, 1990; Golomb, Lawrence, & Sejnowski, 1991; Gray, Lawrence, Golomb, & Sejnowski, 1995; see the last paper for more precise model comparisons). For the three-dimensional head data in the present study, the results showed generalization performance that was even better than that obtained with texture, peaking at 96.9 percent correct with a minimum subspace dimensionality of 17 projection coefficients.

Finally, the most reliable information for sex classification was found in eigenvectors with relatively larger eigenvalues, a finding consistent with previous work on images O'Toole et al. 1993; Abdi, et al., 1995).

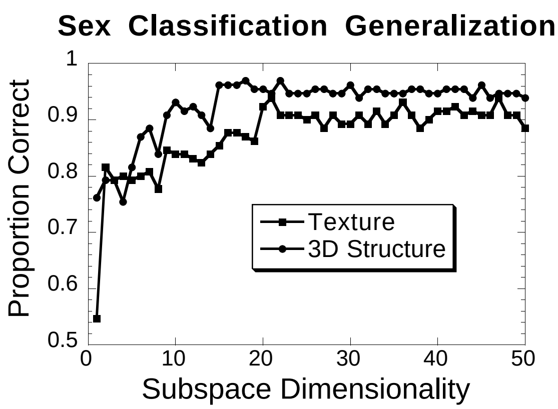

Figure 4: The generalization accuracy of the sex classification network for the texture and structure-based representations as a function of the size of the subspace. In general, three-dimensional structure data supported more accurate sex classification than did texture data. Though interestingly, one exception occurred due to the very strong predictive power of the second eigenvector of the texture analysis, which enabled better classification in this very low-dimensional subspace.

2 General Discussion

The results of the present study indicated that the PCA of the separated structure and texture data captured information relevant for determing the sex of a face. Additionally, the quality of the sex-related information differed in structure versus texture 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 that has considered 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, Healey, Burton, Doyle, Coombes & Linney, 1991; Bruce & 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, using a large number of heads, Troje and Bülthoff (in press) showed that human observers matched the identity of depth-rotated (i.e., view-changed) heads with texture data more quickly than heads without texture data. On a face recogntion task, however, Hancock et al. (in press) showed that at least one component of human performance, the hit rate, was best predicted by separate analyses of the shape and texture information, subsequently brought together.

Interestingly, where sex classification is concerned, there is evidence from several sources that human observers make use of a very broad range of cues, including aspects of the the structure/shape of the face and a variety of texture cues. For example, Bruce, Burton, Hanna, Healey, Mason, Coombes, Fright and Linney (1993) showed accuracy decrements in sex classification using "textureless" laser scanned heads, which they surmised carried important cues for sex classification (e.g., bushiness of eyebrows, and beard stubble). Additionally, however, they found decremental effects of sex classification performance with the manipulations: (I) of face inversion, which they interpreted in terms of the importance of the global configuration; and (II) photographic negation, which they interpreted in terms of the importance of three-dimensional 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 a more diverse set of cues (see also Burton et al. 1993).

More generally, the extent to which humans rely on three-dimensional structure information derived from objects/faces as opposed to two-dimensional image-based information is currently a much-debated point in the psychology literature (Biederman, 1987; Bülthoff & Edelman, 1992) and is being investigated actively by neuroscientists interested in the neurophysiological substrates of object and face recognition (Logothetis, Pauls, Poggio, 1995; Perrett, Hietanen, Oram, & Benson, 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 three-dimensional 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 to texture data versus view-dependent images can also provide a baseline measure of the information inherent in these diffferent representations. The comparison of human performance on simple tasks like sex classification and recognition, with the quality of information available in different representations (e.g., three-dimensional structure, complete textures, and single view-based representations) can serve as a powerful tool for learning about the nature of human representations of objects.

References

[1] 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.

[2] Biederman I, (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, *94*, 115-147.

[3] 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.

[4] Bruce V, Healey P, Burton M, Doyle T, Coombes A, Linney A, (1991). Recognising facial surfaces. *Perception*, *20*, 755-769

[5] Bruce V, Langton S, (1994) The use of pigmentation and shading information in recognizing the sex and identities of faces. *Perception*, *23*, 803-822.

[6] Brunelli R, Poggio T, (1993). Caricatural effects in automated face perception. *Biological Cybernetics*, *69*, 235-241.

[7] Bülthoff H H, Edelman S., (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Natl. Acad. Sci. USA*, *89*, 60-64.

[8] Burton M, Bruce V, Dench N, (1993). What's the difference between men and women? Evidence from facial measurement. *Perception*, *22*, 153-176.

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

[10] Fleming M, Cottrell G W, (1990). Categorization of faces using unsupervised feature extraction. *Proceedings of IJCNN-90*, *Vol 2* (Ann Arbor, MI:IEEE Neural Networks Council) pp 65-70.

[11] Golomb B A, Lawrence D T, Sejnowksi T J, (1991). Sexnet: A neural network identifies sex from human faces. In *Advances in Neural Information Processing Systems*, *Vol 3* Eds R P Lippman, J Moody, D S Touretsky (San Mateo, CA: Morgan Kaufmann) pp 572-577.

[12] Gray M, Lawrence D T, Golomb B A, Sejnowski T J, (1995). A perceptron reveals the face of sex. *Neural Computation*, *7*, 1160-1164.

[13] Hancock P J B, Burton A M, Bruce V, (in press). Face proccessing, human perception and principal components analysis. *Memory & Cognition.*

[14] Hotelling, H. J, (1933) Analysis of statistical variables into principal components. *Educational Psychology Sept.*

[15] Karhunen K, (1946). Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae, A1, Math. Phys. 37.*

[16] Loéve M M, (1955). *Probability Theory.* (Princeton, NJ: Von Norstrand).

[17] Logothetis N K, Pauls J, Poggio T, (1995). Shape recognition in the inferior temporal cortex of monkeys. *Current Biology*, *5*, 552-563.

[18] O'Toole A J, Abdi H, Deffenbacher K A, Bartlett J, (1991). Classifying faces by race and sex using 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.

[19] O'Toole A J, Abdi H, Deffenbacher K A, Valentin D, (1993). Low dimensional representation of faces in higher dimensions of the face space. *Journal of the Optical Society of America A*, *10*, 405-411.

[20] O'Toole A J, Bülthoff H H, Troje N F, Vetter T, (1995). Face Recognition across large changes in view. In *Proceedings of the International Workshop on Face and Gesture Recognition.* Ed M Bichsel (Zürich: University of Zürich Multimedia Laboratory) pp 326-331.

[21] 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.

[22] Pearson K, (1901). On lines and planes of closest fit to systems of points in space. *Phil. Mag. 6th Series.*

[23] 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.

[24] Troje N F, Bülthoff H H, (in press). Face recognition under varying pose: The role of texture and shape. *Vision Research.*

[25] Turk M, Pentland A, (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*, 71-86.