PERGAMON

Vision Research 39 (1999) 3145–3155

Vision Research

www.elsevier.com/locate/visres

Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: an application of three-dimensional morphing

Alice J. O'Toole [a,]*, Thomas Vetter [b], Volker Blanz [b]

[a] *The University of Texas at Dallas, School of Human Development, GR 4.1 Richardson, TX 75083-0688, USA*
[b] *Max Planck Institut für biologische Kybernetik, Spemannstrasse 38, Tübingen 72076, Germany*

Received 18 June 1998; received in revised form 7 January 1999

Abstract

We measured the three-dimensional shape and two-dimensional surface reflectance contributions to human recognition of faces across viewpoint. We first divided laser scans of human heads into their two- and three-dimensional components. Next, we created *shape-normalized* faces by morphing the two-dimensional surface reflectance maps of each face onto the average three-dimensional head shape and *reflectance-normalized* faces by morphing the average two-dimensional surface reflectance map onto each three-dimensional head shape. Observers learned frontal images of the original, shape-normalized. or reflectance-normalized faces, and were asked to recognize the faces from viewpoint changes of 0, 30 and 60°. Both the three-dimensional shape and two-dimensional surface reflectance information contributed substantially to human recognition performance, thus constraining theories of face representation to include both types of information. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Face recognition; Three-dimensional *morphing*; Representation

1. Introduction

As a visual stimulus. the human face consists of a three-dimensional surface with an overlying reflectance function at each point on the surface. The three-dimensional information is determined by the structure of the human skull and by the shape and texture of the overlying skin and tissue. The reflectance function at any given point on the surface is simply a measure of how efficiently the skin at that point reflects light of various wavelengths[1]. The information that reaches one's eye from this stimulus is, therefore, a complicated function of the three-dimensional structure of the facial surface, the reflectance function of the face at each point, and the illumination and viewpoint conditions.

Despite the complicated nature of the information in faces and the complexity of the tasks required to achieve some constancy in representing this information, human observers are remarkably good at recognizing and categorizing faces. A primary question of interest for psychologists is to understand the nature of the information humans use in accomplishing these tasks. In recent years, both in the object and face recognition literatures, much attention has been paid to the issue of whether the human representation of faces and objects is based more predominantly on the two- or three-dimensional features of the stimulus (e.g. Biederman, 1987; Bülthoff & Edelman, 1992). Although there is reasonably good evidence that many aspects of the two-dimensional image-based structure of faces relate to human performance recognizing (O'Toole, Deffenbacher, Valentin & Abdi, 1994; Hancock, Burton & Bruce, 1996) and categorizing faces (O'Toole, Deffen-

* Corresponding author. Fax: + 1-972-8832491.

E-mail address: otoole@utdallas.edu (A.J. O'Toole)

[1] The term reflectance is not always used in precisely the same way. Horn (1986) in his book on robot vision states, 'Often reflectance properties can be described in terms of the product of two factors: a geometric term expressing the dependence on the angles of light reflection, and another term that is the fraction of light reemitted by the surface. This latter is called the *albedo* (pp 224).' In fact. it is the albedo that we are interested in here. though we have not used this term because it may be unfamiliar to may readers and is not used more consistently than reflectance across different literatures.

PII: S0042-6989(99)00034-6

bacher, Valentin, McKee, Huff & Abdi, 1998) from within a single viewpoint, the problem of generalizing recognition across viewpoint shifts seems to require better information about the three-dimensional shape of the face (though cf. O'Toole & Edelman, 1996; Valentin & Abdi, 1996, for a discussion of the issues).

In previous work, experiments aimed at understanding the nature of representations, used by human observers for objects and faces, have often measured the extent to which recognition and perceptual matching tasks generalize over viewpoint. Although good viewpoint generalization (view invariance) is considered support for a three-dimensionally-based representation, and poor viewpoint generalization (view dependence) is considered support for a two-dimensionally-based representation, some ambiguity still remains. For example, view dependence can be consistent with a three-dimensional representation if we make the assumption that the creation of an accurate three-dimensional representation of a face requires a great deal of experience with the face. Likewise, view invariance can be consistent with a two-dimensional representation if we make the assumption that familiarity with a face comes after we have experienced it from many viewpoints.

In the present study we have taken a more direct approach to the representation issue by varying the information in the stimulus and by measuring its effects on human performance in a face recognition task. In fact, the major problem encountered in studying the extent to which human observers encode faces in terms of their two- or three-dimensional features is that it is difficult to isolate the three-dimensional shape information in faces from the two-dimensional images commonly available[2]. It is equally difficult to get a pure measure of the reflectance information in a face from an image of the face[3]. Further, to investigate the psychological validity of a posited representation of faces, one must be able, not only to isolate this information, but also to selectively manipulate it in ways that enable an estimate of the extent to which human observers rely on two- versus three-dimensional information for successfully completing different face processing tasks. Thus, we have a two-part problem that consists of separating the three-dimensional shape and two-dimensional surface reflectance components of faces and selectively manipulating these components for use in an experiment.

1.1. Separating the surface and reflectance components of faces

The recently available technology of laser scanning enables a solution to the problem of separating the two- and three-dimensional information in faces, but does not solve the problem of the selective manipulation of these two kinds of information in faces. More precisely, commercially available laser scanners operate by simultaneously sampling the three-dimensional surface of the face and the pixel-based reflectance values (i.e. usually *rgb*) at these same points. The scanner we used rotates horizontally around the head, sampling the surface at each of 512 equidistant steps. Each such sample is made along a vertical line projected onto the head. The three-dimensional surface structure and reflected light along these vertical line samples are also measured in 512 equidistant steps. Thus, for both the surface and the reflectance data we have a 512×512 map of measurements. More formally, this sampling process transforms shape and reflected light into a cylindrical coordinate system with an imaginary vertical axis at the center of the head. The three dimensional surface code, then, consists of the lengths of radii from the central vertical axis of the cylinder to the surface points on the face. We will refer to this three-dimensional geometrical representation as the *surface map*. The reflectance code consists of the *rgb* values at these same surface sample points. We will refer to this representation as the *reflectance map*.

It is worth noting that the laser scan measuring process leads to some limitations because the sampling is performed only perpendicular to the surface of a cylinder. Consequently, the geometrical structure of occluded areas behind the ears and sometimes beneath the chin cannot be resolved. Rarely, however, are the internal facial features such as the nose adversely affected by this sampling. The reason for this is that variations of the reflected light at steps of the structure perpendicular to the surface of the cylinder are averaged and so, the tiny internal regions of the faces that might be susceptible to this kind of sampling artifact, are generally not problematic.

An example of the surface and reflectance data from a laser scan, separately and in rendered combination, appear in Fig. 1.

Before proceeding, we must digress briefly to note some difficulties in our choice of terminology for this paper. As noted, the intensity and spectral composition of the light that reaches the eye from a face is determined by the three-dimensional structure of the head, the spectral reflectance function of the skin, hair, cornea, etc., and the relative positions of the light

[2] Directly recovering the three-dimensional surface information from an image is a notoriously difficult problem in computer vision. The problem is classically ill-posed and has never been solved in a completely general way. However, when prior knowledge can be used, e.g. as for faces, more specific but satisfactory solutions can be found (Vetter & Blanz, 1998).

[3] This latter problem is due to the fact that photographs of faces confound two-dimensional reflectance information with information about the three-dimensional surface in the form of shape-from-shading cues, which depend on viewpoint and illumination conditions.

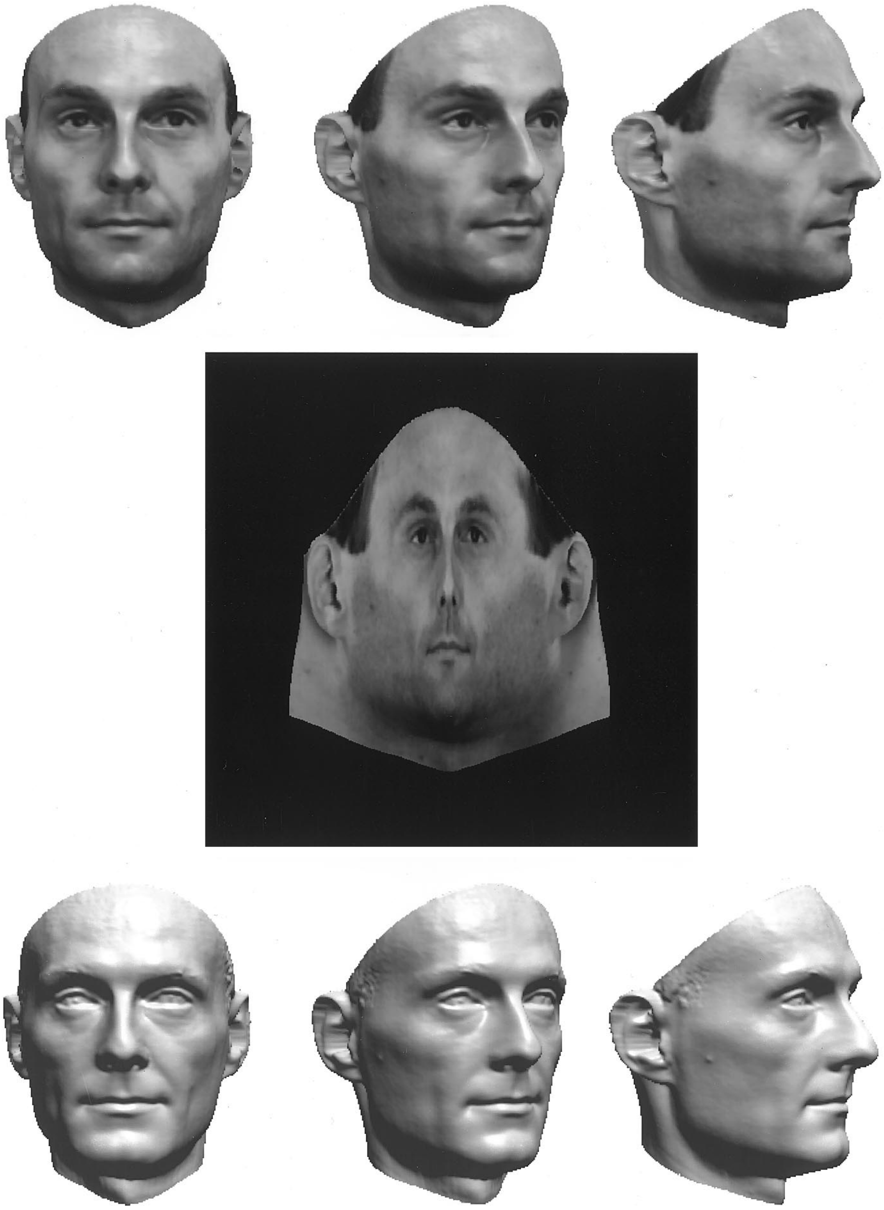

Fig. 1. Output of the laser scan consists of a three-dimensional shape combined with its reflectance map, which is rendered from three viewpoints (row 1), a reflectance map, unrolled here so that all sides of the face are seen (row 2), and the pure three-dimensional shape rendered from three viewpoints (row 3).

source and the viewer. In short, what reaches the eye from a face at any given point in time is a view or two-dimensional image of the face that confounds all three kinds of information. The three-dimensional surface map from the laser scanner is a relatively pure measure of the three-dimensional structure of the *entire* object (i.e. more than one can see with a single view). Likewise, the two-dimensional surface reflectance map from the laser scanner is a relatively pure measure of the spectral reflectance function of the *entire* face[4]. These representations exist, therefore, independent of particular viewpoints. Indeed, software that wraps the reflectance map over the surface map of the face and rotates the face model to a particular viewpoint is needed to render a standard image of the face, based on the combined surface and reflectance data that would be visible from that viewpoint. Illumination conditions would, of course, need to be specified in the rendering software (see again Fig. 1 for a face rendered with and without its reflectance map).

[4] It should be noted that in computer science applications this reflectance component of the laser scan data is sometimes referred to as a texture map. The computer science use of the term texture is totally unrelated to the use of the term by vision scientists.

A final note on terminology concerns the role of the shape information that is normally confounded with the surface reflectance information in an image of a face. First, because the laser scanner's reflectance component comes from a set of localized images, one might expect that information about the surface orientation (shape) would be confounded in the reflectance map. The use of ambient illumination by the laser scanner eliminates (or at least strongly limits) the contribution of surface orientation to the reflectance component of the laser scan data. Thus, the reflectance map we refer to here contains relatively pure information about the two-dimensional reflectance properties of the facial surface. Although the term reflectance map is still not a perfect description of the information we are referring to, we have tried to define it here as precisely as we can.

The ability of laser scanner technology to solve the problem of separating the surface and reflectance information in a face has been put to good use in a recent experiment by Hill, Bruce and Akamatsu (1995) who assessed observers' ability to make sex and race judgments about faces using different components of the laser scan stimuli. They found that both kinds of information contributed to these categorizations. They used pure 3D surface information and unrolled color reflectance maps[5] from laser scans and found that the surface information dominated for the race decisions and that the color reflectance information dominated for the sex decisions. Combined, the relative advantages of shape and color varied with viewpoint. The color information was more important for the frontal images and the shape information was more important for the angled views.

1.2. *Selectively manipulating the surface and reflectance components of faces*

The problem of selectively *manipulating* the surface and reflectance information can be solved, in theory, by simply exchanging the surface and reflectance maps among arbitrary pairs of faces, or in the present case, between individual faces and the average face. However, this presents a technical problem because the reflectance map of one face may not fit properly onto the shape of another face[6]. Thus, for example, it may be that the pupils of the eyes are captured at different sample points for two faces. This would yield a composite face in which the reflectance values of the pupils get mapped onto incorrect parts of the face surface, for example, onto the eyelid region. In short, before one can selectively manipulate the surface and reflectance information from the laser scan data, one must put both the surface and reflectance data into a comparable coordinate system.

The same kind of problem must be solved, also, before applying a morphing algorithm to two images. Because morphing is currently a very popular technique for blending images of objects, we use it as an analogy for understanding the approach we have taken with the laser scans. In order to morph images of two faces together, one must first locate a set of corresponding points on the faces. These points include the fiducial points but are often supplemented with additional comparable points to obtain a high quality morph. In standard two-dimensional morphing software these points are located and marked on each face image by a human operator, prior to applying the morphing procedure.

In the present work, we have taken a somewhat different approach. We have made use of an automated correspondence algorithm for solving the problem of putting the surface and reflectance information from laser scan data on faces into a comparable coordinate system for all of the faces (Vetter & Blanz, 1998). In this algorithm, the problem of matching is cast into the more general computer vision form in which one attempts to match all of the data points in two images/surfaces, rather than just a subset of the facial landmark points[7]. This is the approach taken most commonly in solving the classical correspondence problems in stereopsis and motion analyses. Although this problem is far from solved in a perfectly general form, a great deal of progress has been made recently on the problem with faces. Specifically, several methods have been applied successfully to the task of automating a correspondence finding procedure for images of human faces (Beymer, Shashua & Poggio, 1993; Beymer & Poggio, 1996; Lanitis, Taylor & Cootes, 1997; Vetter & Poggio, 1997). These approaches have been extended successfully to the laser scan data by Vetter and Blanz (1998), (cf. also O'Toole, Vetter, Volz & Salter, 1997 for an application). Full details of how the correspondence algorithm works can be found in Vetter and Blanz (1998). We also include the implementation details of this algorithm as they apply to the present study in Appendix A.

[5] Hill et al. (1995) refer to these as colour data and the process of adhering these to the facial surface as texture mapping.

[6] The data formats of laser scans are obviously pre-determined to take the same number of surface and reflectance samples from faces, and so by fit we do not mean sample points per se but rather mean the (non)correspondence between the features captured at spatially identical places in the surface code for one face and the reflectance code for another face.

[7] This complete correspondence approach is a difference of degree rather than of kind to the hand-placing of points on the faces. Software to do three-dimensional morphing on laser scans using hand-placed corresponding points is not, to our knowledge, commercially available. The morphs we have made are of a higher resolution and quality than those that would result from hand-placing a subset of points on three-dimensional surface, but would not be qualitatively different in any way.

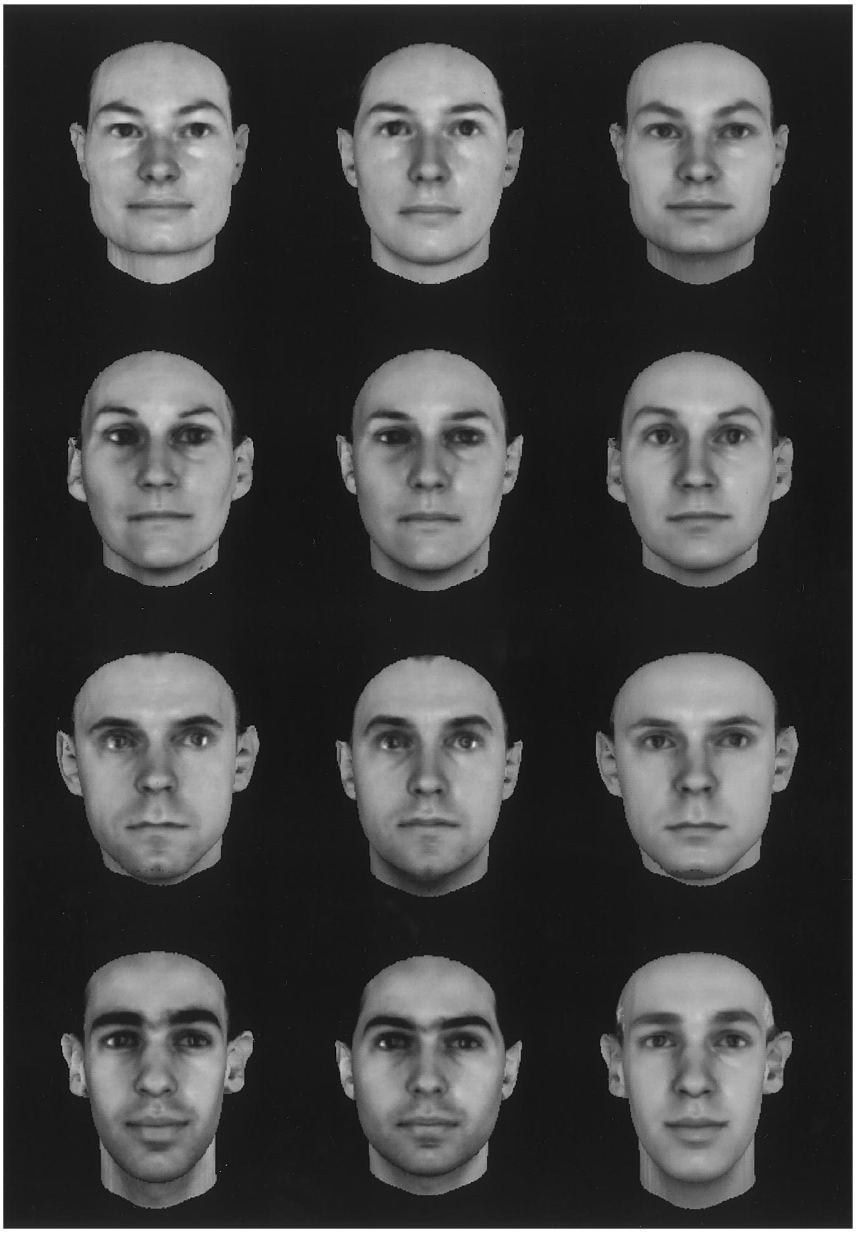

Fig. 2. Column 1 contains four normal faces. Column 2 contains the shape-normalized versions of the faces, and column 3 contains the reflectance-normalized versions of the faces.

For present purposes, once a comparable coordinate system is established for the surface and the reflectance data for the laser scans, it is then easy to exchange the surface and reflectance data for arbitrary pairs of faces. Thus, we are now able both to separate the surface and reflectance information in faces and to selectively manipulate these two components in real faces.

The purpose of the present study was very straightforward. We wished to measure the relative contributions of three-dimensional shape versus two-dimensional surface reflectance information in a task of face recognition across viewpoint change. We did this by creating stimuli that varied exclusively in either three-dimensional shape or in two-dimensional surface reflectance information. The simplest approach was to make two sets of stimuli from the original face scans by: (a) Mapping each face's reflectance map onto the average shape; and by (b) mapping the average reflectance map onto the shapes of each of the available faces. Samples of these stimuli appear in Fig. 2. The left column contains four normal faces, the middle column contains the shape-normalized versions of the faces and the right column shows the reflectance normalized versions of the faces. As can be seen, the head shapes of the faces in column 2 are identical and the reflectance information varies. In column 3, the head shapes vary, but the reflectance information remains the same. Readers may notice that the shape information, e.g.

thick lips etc., are still visible here. The morphing algorithm fits the average reflectance map to the shapes of the original faces, by warping where necessary to adhere properly to the correct parts of the face (e.g. being certain not to map eye reflectance information onto the forehead or cheek).

Finally, although not part of our original purpose, interesting differences between male and female faces were evident in the data, and so we have included this as a factor.

We first report a control experiment using the original faces to set a baseline for recognition performance across viewpoint with these faces. Next, we measured recognition performance as a function of the stimulus type (reflectance-normalized versus shape-normalized faces), viewpoint generalization conditions (0, 30, and 60° viewpoint change), and sex of the face. From the observer data we address the following questions: (1) Do observers rely more on shape or reflectance variations in recognizing faces?; and (2) does the relative reliance on shape versus reflectance information change as a function of the added complexity of the viewpoint generalization requirements? For example, it may be the case that one kind of information proves more reliable when no viewpoint change is involved, and the other when viewpoint changes of varying degrees occur between learning and test.

2. General methods for stimulus creation

We present first a brief description of the stimulus creation methods that are common to both experiments.

2.1. Description of laser scan head stimuli

Laser scans (Cyberware™) of 100 heads of young adults (50 male and 50 female) were used as stimuli. The mean age of faces in the data base was 26.9 years (standard deviation = 4.7 years). The subjects were scanned wearing bathing caps, which were removed digitally. The laser scans provided surface map data consisting of the lengths of 512 × 512 radii from a vertical axis centered 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. 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. A subset of 48 (24 male and 24 female) faces was selected randomly from this data base to serve as stimuli in the experiment.

2.2. The correspondence problem

The procedures applied to solving the correspondence problem for this particular set of laser scan stimuli are complex but the basic principles are described in detail in Vetter and Blanz (1998). Additionally, as noted previously, to make this manuscript self-contained, we describe the implementation details of the algorithm in Appendix A. For present purposes, we give only a basic overview of the procedure and representation achieved here and refer interested readers both to the Appendix A and to the sources cited.

The basic idea behind the procedure used here is to match the data points in each individual face with the corresponding feature points in the average face and hence to represent each face as a deformation field from the average. Thus each data point in the face representation would contain a pointer to the analogous data point in the average. This was done by applying optic flow algorithms optimized in this case to deal with the continuous surface and reflectance data found in faces.

2.3. Stimuli

Two sets of stimuli were made from the original surface and reflectance maps of 48 faces. *Reflectance normalized faces* were created by wrapping the average reflectance map onto the surface map of each individual face. *Shape normalized faces* were made by mapping the reflectance maps of each individual face onto the average shape. Thus, all faces existed in three versions: (1) *Normal* with the original shape and reflectance information intact; (2) *shape normalized faces* with the three-dimensional shape set to the average shape and the original reflectance information intact; and (3) *reflectance normalized faces* with the two dimensional reflectance information set to the average reflectance and the original three-dimensional shape information intact. From these versions of the face models, we created stimuli for our experiment by using computer graphic software to render each face in each of its three versions (normal, shape normalized, and reflectance normalized) from three viewpoints (0, 30, and 60°). An example stimulus appears in Fig. 3.

3. Experiment 1: baseline recognition for normal faces

This first experiment was carried out to provide baseline data on the normal faces and to be sure that we could replicate the common finding of viewpoint dependency found for face recognition with these stimuli.

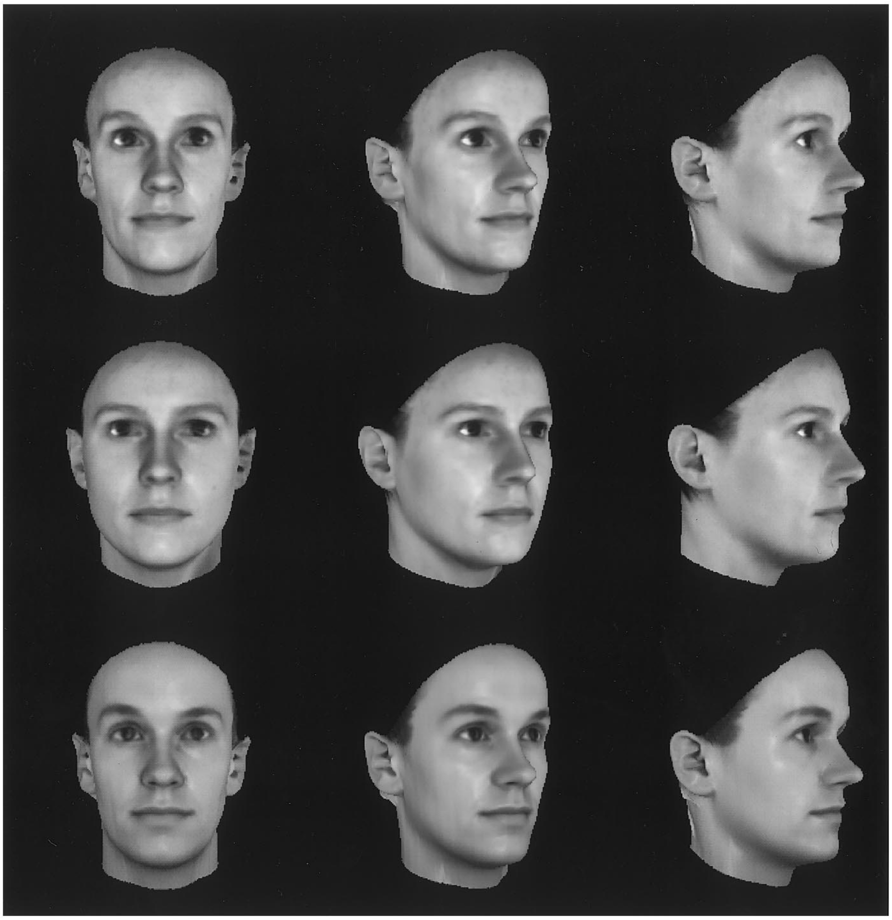

Fig. 3. A sample face in its normal (row 1), shape-normalized (row 2) and reflectance-normalized (row 3) versions rendered at 0, 30, and 60°.

3.1. Observers

A total of 30 volunteers from the University of Texas at Dallas community participated in the experiment. Some of the participants were students who were compensated with a research credit in a core course in the psychology program.

3.2. Apparatus

All experimental events were controlled by a Macintosh computer programmed with PsyScope (Cohen, McWhinney, Flatt & Provost, 1993).

3.3. Procedure

Observers read standardized instructions that indicated the purpose and course of the experiment. These instructions indicated that the test faces might be seen from a different viewpoint than the learned faces, and that the observers were to respond old to any views of the people pictured in the learning part of the experiment. In the learning session. observers viewed 24 frontal faces (half male and half female) presented on a computer screen for 5 s each. After a short break, they viewed all 48 faces, as in a total of 16 from each of the three view conditions, (0, 30, and 60°). Of the 16 faces in each view, eight were old and eight were new. Each face was presented on the computer screen until the observer responded old or new by pressing the appropriately labeled key on the computer keyboard. Counterbalancing was done so that across the complete set of observers, all faces appeared equally often as old and new, and all faces were tested equally often in each of the view conditions. The order of presentation of the learning and test faces was randomized individually for each observer.

3.4. Results

For each observer in each condition, a d' for discriminating learned from novel faces was computed. The overall pattern of these data appears in Fig. 4. These data were submitted to a two-factor analysis of variance with test orientation (0, 30, 60°) and the sex of the faces as within subjects factors. We found a main effect of face orientation, $F(2, 29) = 16.7$; $P < 0.01$. As expected, the effect indicated that the accuracy declined as a function of the orientation change from the learned view replicating previous work (e.g. Krouse, 1981; Logie, Baddeley & Woodhead, 1987; Valentin & Abdi, 1996; O'Toole, Edelman & Bülthoff, 1998; see also Troje & Bülthoff, 1996, who used three-dimensional shape models of heads and the combined shape and reflectance head models). The sex of the face was not significant, $F(1, 29) = 2.63$; $P = 0.11$. There was no interaction between sex of the face and its orientation at test, $F(2, 58) = 1.82$; $P = 0.17$. However, as Fig. 4 suggests, a simple main effect of the face sex variable at the frontal face orientation was significant, $F(1, 58) = 5.49$; $P < 0.05$, indicating that female faces were better recognized than male faces in the frontal view condition. This somewhat unexpected finding made us wonder if the sex of the observer might also have had an effect. We were not interested originally in this effect and so, unfortunately, did not equate the number of male and female observers. In fact, the proportion of male and female observers in this study strongly favored female observers, which makes a formal analysis difficult. We did, however, look at the pattern of data divided by the sex of the observer and saw only minor differences.

3.5. Discussion

These data replicate previous findings indicating view-dependent performance in recognition.

4. Experiment 2: recognizing shape- versus reflectance-normalized faces

In this experiment we tested the ability of observers to recognize faces over changes in viewpoint when the shape- versus reflectance-information in the faces was normalized.

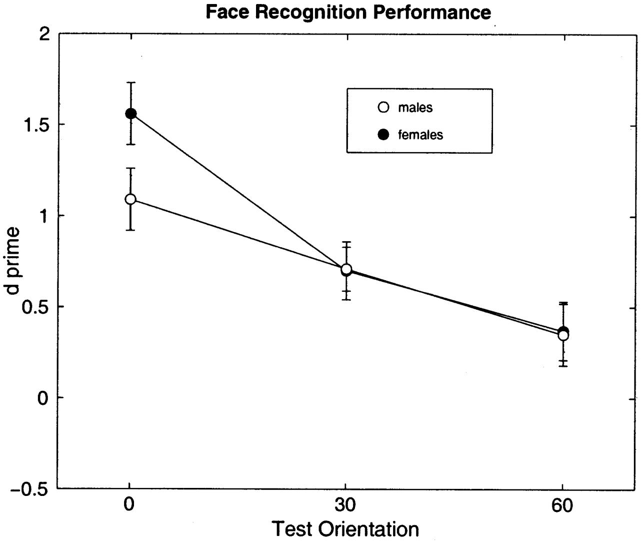

Fig. 4. Recognition scores for the normal male and female faces across the three viewpoints.

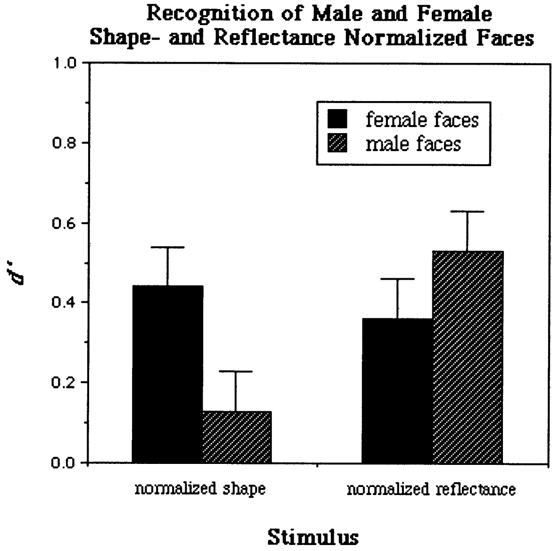

Fig. 5. Recognition scores for the shape- and reflectance-normalized male and female faces.

4.1. Observers

A total of 60 volunteers from the University of Texas at Dallas community participated in the experiment. Some of the participants were students who were compensated with a research credit in a core course in the psychology program.

4.2. Procedure

Observers were assigned randomly to the shape- versus reflectance-normalized face condition. The procedure was identical to the procedure used in Experiment 1 with the exception that the observers learned and were tested on either the shape- or reflectance-normalized faces.

4.3. Results

For each observer in each condition, a d' was computed[8]. These data were submitted to a three-factor analysis of variance with face type (shape-normalized versus reflectance-normalized) as a between subjects factor, and test orientation (0, 30, 60°) and the sex of the faces as within-subjects independent variables. We found a main effect of face orientation, $F(2, 114) = 18.21$; $P < 0.001$, indicating a decline in accuracy with increasing viewpoint change from the learned view-

[8] We eliminated one observer. who scored d' of less than -3.0 in one condition. The elimination of this observer changed none of the ANOVA results.

point. Neither the sex of the face, $F(1, 57) = 1$, nor the face type, $F(1, 57) = 2.06$; $P > 0.05$, proved significant. There was, however, a strong and significant interaction between the sex of the face and the face type, $F(1, 114) = 7.08$; $P < 0.01$. This can be seen in Fig. 5, which indicates that shape information is more important for recognizing male faces than for recognizing female faces. For the female faces we see a reversed but more balanced trend for the shape and reflectance information. Some caution in interpreting the generality of these results, however, can be seen in Fig. 6, which displays the results across viewpoint. Although the three-factor interaction among face type, face sex, and view was not significant, $F(2, 114) = 1.34$; $P = 0.33$, it can be seen qualitatively in this figure that the advantage of the shape information for recognizing male faces suggested by the two-factor interaction, is not entirely consistent across viewpoint. Specifically, the means for the shape- and reflectance-normalized male faces for the 30° view are nearly identical. In other words, it is likely that this two-factor interaction is being carried by the frontal and 60° views.

5. Summary and discussion

To summarize the results of these experiments, several points are worth noting. At the most general level, these data indicate that the ability to recognize faces relies both on information that uniquely specifies the three-dimensional shape of the face *and* on information that uniquely specifies the two-dimensional surface reflectance properties of the face. Neither the two- nor three-dimensional information alone can provide a complete account of the performance levels we obtained with the normal faces, which vary in both their two- and three-dimensional structure. This finding is consistent with Hill et al.'s (1995) finding that both shape and color information are important for race and sex categorizations. The results also make an interesting contrast to recent work by Troje, Huber, Loidolt, Aust and Fieder (1999). They used a two-dimensionally based model for exchanging reflectance and shape information in faces and showed that pigeons rely almost exclusively on the reflectance information for classifying faces by gender.

At a more specific level, the sex of the face was also a mediating variable for understanding the importance of three-dimensional shape- versus two-dimensional surface-reflectance information for face recognition. The pattern of data for female faces is very easy to interpret. The shape and reflectance information are roughly equally informative for recognition—an effect that remains stable across the three viewpoint change conditions.

For the male faces, a more complicated pattern is seen. First, the significant interaction between the sex of face and the face type is due the relatively poor performance seen for shape-normalized versus reflectance-normalized male faces. It is perhaps worth speculating that some of this difference may relate to the fact that the unaltered female faces were better recognized than the unaltered male faces from the frontal viewpoint. This difference may indicate that the male faces simply varied less in their two-dimensional surface reflectance properties than did the female faces. One possibility is that the trace of a 5-o'clock shadow in some of the scans hides part of the male facial surface limiting the surface reflectance variations that can occur there. In any case, additional data collected on more diverse

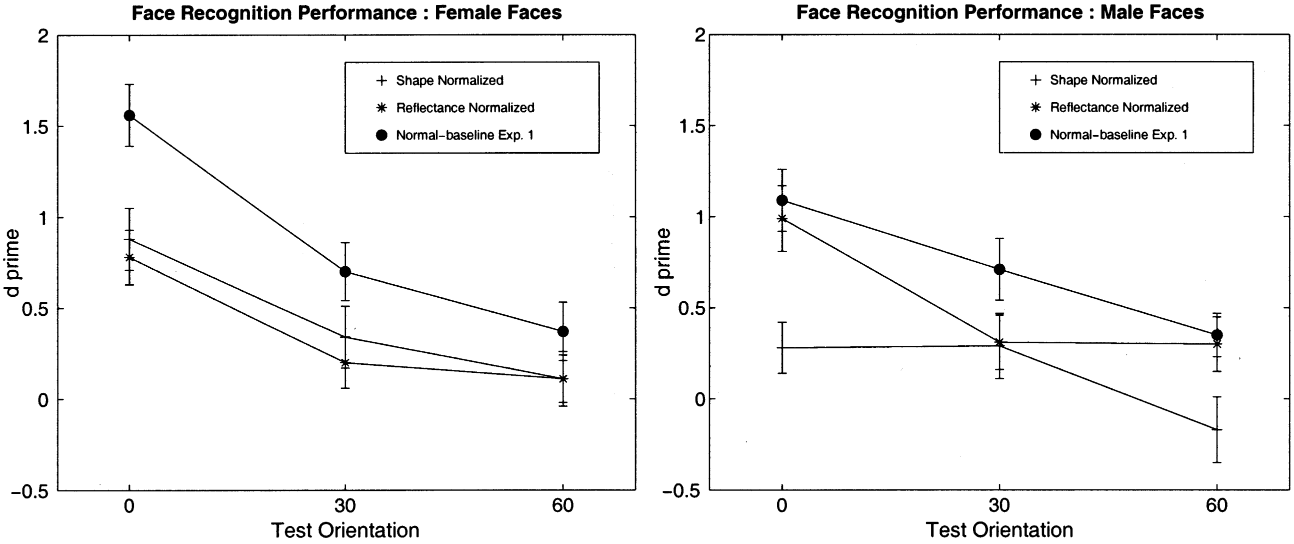

Fig. 6. Recognition scores for the shape- and reflectance-normalized female and male faces across the three viewpoints. The data from Experiment 1 are plotted for reference.

faces (e.g. more variations in age and ethnicity) would be required to assess the generality of these effects.

Finally, at a more theoretical level, although questions of representation contain some of the most important keys to current theories of face/object recognition, they also pose some of the most difficult challenges to experimental methods. The approach we have taken here is based on the recent technical advances provided by laser scanners, which give a direct measurement of the three-dimensional shape information in faces and also provide an independent measure of the two-dimensional reflectance information in faces (i.e. one not confounded with the three-dimensional surface structure). By using morphing procedures and automatic correspondence algorithms with these data, we have been able to selectively eliminate three-dimensional surface versus two-dimensional surface reflectance variation in a set of faces. The present results indicate that human observers can, and do, make use of the available variations in both three-dimensional shape and two-dimensional surface reflectance information in faces for recognition. Although these data do not tell us about the exact nature of the human representation of this information, they do at least definitively inform us that our theories of representation must include more than an exclusively structural coding of faces, and more than an exclusively image-based coding.

Acknowledgements

A. O'Toole gratefully acknowledges support from NIMH grant lR29MH5176501a4 and the Alexander von Humboldt Foundation. Thanks are due also to Niko Troje for the stimulus creation and processing and to Shimon Edelman and Heather Wild for helpful comments on this manuscript.

Appendix A. 3D correspondence algorithm

For matching points on the surfaces of two three-dimensional objects we modified an existing optical flow algorithm developed for two-dimensional images. Although establishing correspondence between surfaces of three-dimensional objects has been considered only recently, matching corresponding features in two-dimensional images has been studied for many years.

A.1. Optical flow algorithm

In video sequences, in order to estimate the velocities of scene elements with respect to the camera, it is necessary to compute the vector field of optical flow, which defines the displacements $(\delta \mathrm{x}, \delta y)=(x_2-x_1, y_2-y_1)$ between points $p_1=(x_1, y_1)$ in the first image and corresponding points $P_2=(x_2, y_2)$ in the second image. A variety of different optical flow algorithms have been designed to solve this problem (for a review see Barron, Fleet & Beauchemin, 1994). Unlike temporal sequences taken from one scene, a comparison of images of completely different scenes or faces may violate a number of important assumptions made in optical flow estimation. However, some optical flow algorithms can still cope with this more difficult matching problem, opening up a wide range of applications in image analysis and synthesis (Beymer et al., 1993).

In previous studies (Vetter & Poggio, 1997), we computed correspondence between face images using a coarse-to-fine gradient-based method (Bergen, Anandan, Hanna & Hingorani, 1992) applied to the Laplacians of the images and followed an implementation described in Bergen and Hingorani (1990). The Laplacians of the images were computed from the Gaussian pyramid adopting the algorithm proposed by Burt and Adelson (1983). For every point x, y in an image $I(x, y)$, the algorithm attempts to minimize the error term $E=\Sigma(I_x\delta x+I_y\delta y-\delta I)^2$ for $\delta x, \delta y$, with I_x, I_y being the spatial image derivatives of the Laplacians and δI the difference of the Laplacians of the two compared images. The coarse-to-fine strategy starts with low resolution images and refines the computed displacements when finer levels are processed. The final result of this computation $(\delta x, \delta y)$ is used as an approximation of the spatial displacement of each pixel between two images.

A.2. Three-dimensional face representations

The adaptation and extension of this optical flow algorithm to the three-dimensional head data is straightforward due to the fact that the cylindrical representation of a head surface is analogous to images: Instead of gray-level values in image coordinates x, y, here we store the radius values and the color values for each angle ϕ and height h. A parameterization of a three-dimensional head in cylindrical coordinates, therefore. consists of two images, one representing the geometry of the head and the other containing the texture information. In order to compute the correspondence between different heads, both texture and geometry were considered simultaneously. The optical flow algorithm as described earlier had to be modified in the following way. Instead of comparing a scalar gray-level function $I(x, y)$, our modification of the algorithm attempts to find the best fit for the vector function:

$$\vec{F}(H, \phi) = \begin{pmatrix} radius\ (h, \phi) \\ red\ (h, \phi) \\ green\ (h, \phi) \\ blue\ (h, \phi) \end{pmatrix}$$

in a norm $\left\| \begin{pmatrix} radius \\ red \\ green \\ blue \end{pmatrix} \right\|^2$

$$= w_1 \times radius^2 + w_2 \times red^2 + w_3 \times green^2 + w_4 \times blue^2.$$

The coefficients $w_1 \ldots w_4$ correct for the different contrasts in range and color values, assigning approximately the same weight to variations in shape as to variations in all color channels taken together.

For representing the geometry. radius values can be replaced by other surface properties such as Gaussian curvature or surface normals.

The displacement between corresponding surface points is captured by a correspondence function

$$C(h, \phi) = \begin{pmatrix} \mathrm{d}\, h(h, \phi) \\ \mathrm{d}\, \phi(h, \phi) \end{pmatrix}. \quad (1)$$

After all individual faces of the training set have been matched to a reference face, their average three-dimensional shape as well as the average surface reflectance map can be computed. Additionally, corresponding values of surface reflectance of different faces can be exchanged.

References

Barron, J. L, Fleet, D. J., & Beauchemin, S. S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision*, *12.1*, 43–77.

Bergen J. R., Anandan P., Hanna K., & Hingorani R. (1992). Hierarchical model-based motion estimation. In *Proceedings of the European Conference on Computer Vision* (pp. 237–252). Santa Marherita Ligure, Italy.

Bergen, J. R., & Hingorani, R. (1990). Hierarchical motion-based frame rate conversion. In *Technical Report*. Princeton, NJ: David Sarnoff Research Center.

Beymer, D., & Poggio, T. (1996). Image representations for visual learning. *Science*, *272*, 1905–1909.

Beymer D., Shashua A., & Poggio T. (1993). Example-based image analysis and synthesis. *A.I. Memo No. 431*. Cambridge MA: Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Biederman, I. (1987). Recognition by components: a theory of human image understanding. *Psychological Review*, *94*, 115–147.

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

Burt, P., & Adelson, E. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, *31*, 532–540.

Cohen, J. D., McWhinney, B., Flatt, N. I., & Provost, J. (1993). PsyScope: a new graphic interactive environment for designing psychology experiments. *Behavior Research Methods*, *Instruments and Computers*, *25*, 257–271.

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

Hill, H., Bruce, N., & Akamatsu, S. (1995). Perceiving the sex and race of faces: the role of shape and colour. *Proceedings of the Royal Society B*, *261*, 367–373.

Horn, B. K. P. (1986). *Robot vision*. Cambridge, MA: MIT.

Krouse, F. L. (1981). Effects of pose, pose change, and delay. *Journal of Applied Psychology*, *66*, 651–654.

Lanitis, A., Taylor, C. J., & Cootes, T. F. (1997). Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*, 743–756.

Logie, R. H., Baddeley, A. D., & Woodhead, M. M. (1987). Face recognition, pose, and ecological validity. *Applied Cognitive Psychology*, *1*, 53–69.

O'Toole A. J., & Edelman S. (1996). Face distinctiveness in recognition across viewpoint change: an analysis of the statistical structure of face spaces. *Proceedings of the international workshop on automatic face and gesture recognition*. Los Alamitos, CA: IEEE Computer Society Press.

O'Toole, A. J., Edelman, S., & Bülthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research*, *38*, 2351–2363.

O'Toole, N. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory and Cognition*, *22*, 208–221.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., McKee, K., Huff, D., & Abdi, H. (1998). The perception of face gender: the role of stimulus structure in recognition and classification. *Memory and Cognition*, *26*, 146–160.

O'Toole, A. J., Vetter, T., Volz, H., & Salter, E. M. (1997). Three-dimensional cariactures of human heads: distinctiveness and the perception of facial age. *Perception*, *26*, 719–732.

Troje, N., & Bülthoff, H. H. (1996). Face recognition under varying pose: the role of texture and shape. *Vision Research*, *12*, 1761–1771.

Troje, N. F., Huber, L., Loidolt, M., Aust, U., & Fieder, M. (1999). Categorical learning in pigeons: the role of texture and shape in complex static stimuli. *Vision Research*, *39*, 353–366.

Valentin, D., & Abdi, H. (1996). Can a linear autoassociator recognize faces from new orientations. *Journal of the Optical Society of America A*, *13*, 717–724.

Vetter T., & Blanz V. (1998). Estimating coloured 3D face models from single images: an example based approach. In H. Burkhardt, & B. Neumann, *Proceedings of the fifth European conference on computer vision* (*ECCV '98*) (pp 499–513). Freiburg, Germany.

Vetter, T., & Poggio, T. (1997). Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern analysis and Machine Intelligence*, *19*, 33–742.