Pergamon

0042-6989(95)00230-8

Vision Res., Vol. 36, No. 12, pp. 1761–1771, 1996

Printed in Great Britain
0042-6989/96 $15.00 + 0.00

Face Recognition Under Varying Poses: The Role of Texture and Shape

NIKOLAUS F. TROJE,*† HEINRICH H. BÜLTHOFF*

Received 28 February 1995; in revised form 17 August 1995

Although remarkably robust, face recognition is not perfectly invariant to pose and viewpoint changes. It has long been known that both profile and full-face views result in poorer recognition performance than a 3/4 view. However, little data exist which investigate this phenomenon in detail. The present work provides such data using a high angular resolution and a large range of poses. Since there are inconsistencies in the literature concerning these issues, we emphasize the different roles of the learning view and the testing view in the recognition experiment. We also emphasize the roles of information contained in the texture and in the shape of a face. Our stimuli were generated from laser-scanned head models and contained either the natural texture or only Lambertian shading and no texture. The results of our same/different face recognition experiments are: (1) only the learning view but not the testing view affects recognition performance. (2) For textured faces the optimal learning view is closer to the full-face view than for the shaded faces. (3) For shaded faces, we find a significantly better recognition performance for the symmetric view. The results can be interpreted in terms of different strategies to recover invariants from texture and from shading.

Face recognition Viewpoint invariance Shading Texture Symmetry

INTRODUCTION

From personal experience we know that the human face recognition system shows a remarkable degree of robustness against rotation about the vertical axis. If we become acquainted with a person's face from only a single photograph, it is nevertheless not very difficult to recognize that person even from different views which we have never seen before. Although there are several promising approaches (Beymer *et al.*, 1993; Lades *et al.*, 1993) human-like invariance to pose changes has not been achieved by any artificial recognition system up to now. This illustrates that we are still far from understanding viewpoint invariant face recognition (but see O'Toole *et al.*, 1995). An important prerequisite for a study of face recognition is exact data about human performance when generalizing to novel views.

The present study aims to provide such data. Before describing our experiments we want to briefly summarize previous work dealing with face recognition under different pose changes. The summary will motivate our approach of investigating the role of learning view and testing view, and of separating the influences from the texture‡ and from the shape of the presented faces (Fig. 1).

Several studies in the past two decades have dealt with face recognition tasks involving changes in pose (Patterson & Baddeley, 1977; Davies *et al.*, 1978; Krouse, 1981; Logie *et al.*, 1987; Bruce *et al.*, 1987; Wogalter & Laughery, 1987; Schyns & Bülthoff, 1994). Although different designs were used, the basic recognition experiment is similar in all these studies: subjects had to decide whether two sequentially presented images of faces showed the same person or not. In some of these studies, the two images were presented in immediate succession. In other studies, the experiment was divided into a training phase in which all the learning images were shown and a testing phase in which learned and novel faces were shown.

We define the term "learning image" to mean the first of the two corresponding images and "testing image" the second image. The object centred term "pose" is used synonymously with the subject centred term "view". The views used in the learning and the testing images are denoted by "learning view" and "testing view", respectively. The view itself is expressed in terms of the angle between the symmetry plane of the head and the viewing direction. In addition, we use the terms "full-face view" and "profile view" as synonyms for the 0 deg and the 90 deg view, respectively. The term "3/4 view" is widely used in the literature but as far as we know there

*Max-Planck-Institut für biologische Kybernetik, Spemannstraße 38, D-72076 Tübingen, Germany.
†To whom all correspondence should be addressed.
‡Throughout this paper, the term "texture" is used as in computer graphics, where the texture of an object is meant to be its color or grey-level map.

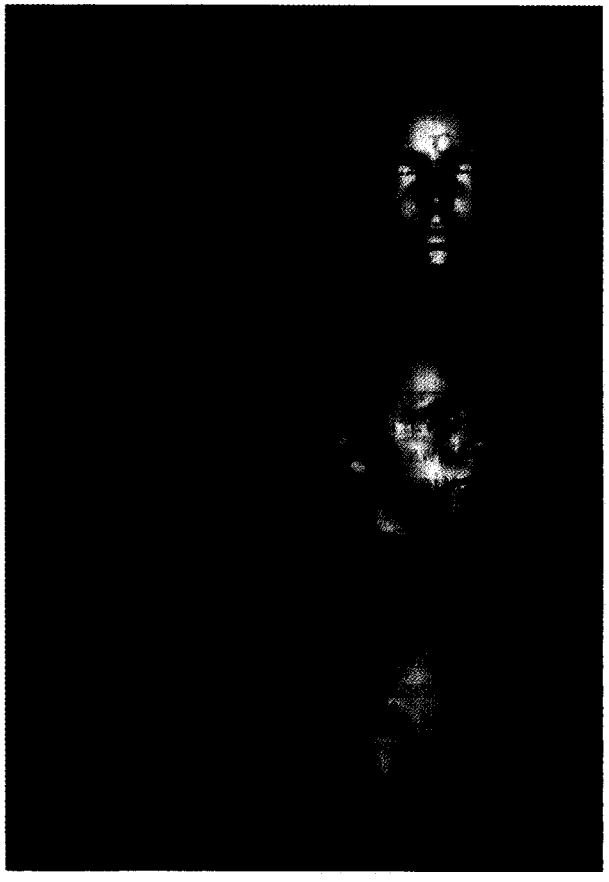

FIGURE 1. Examples for the face stimuli used in the experiments. The left column shows three different views (0, 45 and 90 deg view) of the same face with its natural texture. The right column shows the same face without any texture but applying a Lambertian shading model. The views were generated from laser-scanned three-dimensional head models.

is no exact definition for it. We assume that it corresponds to the 45 deg view but we use it only when we discuss studies of authors who use that term. Finally "pose change" indicates the angular difference between learning and testing views.

In several investigations, effects of pose and of pose changes were intermixed with changes in expression (Patterson & Baddeley, 1977), presentation mode (Davies *et al.*, 1978; Wogalter & Laughery, 1987), or the presence or absence of very obvious features like a beard or a wig (Patterson & Baddeley, 1977). The first author who really focused on pose changes was Krouse (1981). She used a memory task in which subjects were trained with 16 faces, half of them in full face pose and half of them in 3/4 pose. In the testing phase, subjects had to decide which of four simultaneously presented faces had been seen before (four alternative forced choice). The mean correct responses were evaluated and the effects of two factors were investigated: the first factor was the pose of the face shown in the learning image and had the two levels "full-face" and "3/4 view". The second factor was the pose change (i.e. the difference between learning and testing view). This factor had the two levels "matched" (i.e. learning and testing view were the same) and "unmatched". Krouse found significant effects for both factors. The 3/4 view yielded better recognition than the full-face view and the matched condition better recognition than the unmatched condition. Krouse did not test the influence of the testing view. However, her data can be re-arranged to get the mean correct responses for cases in which the testing view was either full-face or 3/4 view. The mean correct response for the full-face pose was slightly larger than for the 3/4 view, suggesting a full-face advantage rather than a 3/4 view advantage.

While Krouse used only two different poses Bruce *et al.* (1987) used face stimuli in full-face, 3/4 view, and profile view. Their experiment consisted of single trials in which two views were presented sequentially. Subjects were asked to decide whether or not the two views were taken from the same person and the latency of the response was measured and evaluated. Again there were two within-subject factors. The first factor was the pose of the face. However, in this study it was not the pose from the first face but the one from the second face. The factor "pose" thus corresponds to the testing view. The second factor was the pose change and included angles of 0, 45 and 90 deg between learning and testing view. In addition, a between-subject factor was introduced coding for the subject's familiarity with the faces. Half of the subjects were familiar with the presented faces (because the images were taken from members of the department staff and the subject group consisted of students of the same department) while the other half was naive. The authors reported a significant effect of the factors "pose", "pose change" and their interaction. Furthermore, the factor "familiarity" interacted with the factor "pose", due to differences in the preferred pose. In the group of subjects unfamiliar with the presented faces, there was a clear advantage for the 3/4 view. The group which was familiar with the faces, however, showed slightly better recognition for the full-face view. All statements concerning the pose of a face refer to the pose of the second view, the testing view. Bruce *et al.* did not search for an effect of the learning view and the design of their experiment does not allow a re-evaluation to obtain that information.

The effects of both the learning and the testing view were, however, the focus of another study by Logie *et al.* (1987). The authors used three levels for both factors: "full-face", "3/4 view", and "profile". The 3/4 view and the profile view always showed the right side of the face. Unlike Bruce *et al.*, they evaluated hit rates and false alarm rates instead of reaction times. This study found a pronounced effect of the learning view and no effect of the testing view. However, if the pose of learning and testing view were different, recognition performance decreased. There was no difference between pose changes of 45 and 90 deg.

The most recent study about the effect of different poses in face recognition is that of Schyns and Bülthoff (1994). Instead of using realistic faces, their stimuli consisted of shaded surface models of laser scanned

heads without texture information. These heads were presented in five different poses, deviating −36, −18, 0, +18 and +36 deg from the full-face view. The study found effects of the learning view but no effects of the testing view. However, there was an interaction between learning and testing view. The study of Schyns and Bülthoff (1994) is the only one that systematically used views from both sides of the face and the authors suggest that the symmetry between learning and testing view plays an important role. However, their data are not sufficient to entirely prove that suggestion.

In summary, the above investigations provide evidence for an overall advantage of the 45 deg view in the learning part of a recognition experiment. A similar advantage for the 45 deg view during testing was only found by Bruce *et al.*, while the other studies do not provide evidence for such an effect. Most of the studies only contain information about the effect of either the learning or the testing view. A clear conceptual distinction between the role of these two views is often missing. Such a distinction, on the other hand, might be crucial, since the two views serve different purposes. The learning view has to provide information to establish an appropriate representation to be memorized. The testing view, however, has to be used to recall the stored information supporting the recognition process itself.

Only two of the studies above investigated the role of both the learning and the testing view simultaneously. Logie *et al.* used three different views which were all taken from the same side of the head, so generalization to the contralateral side could not be studied. The investigation of Schyns and Bülthoff (1994) used views from both sides, but they differed only in the range of ± 36 deg from the full-face view. Additionally, the stimuli used were not comparable with the faces used in other studies since their shaded surface models lacked any texture information.

The present work aims to investigate the effect of both the learning and the testing view on face recognition in more detail, using a higher angular resolution, a greater range of poses, and both naturally textured faces and shaded surface models. We used a paradigm in which pairs of face-views were presented only very briefly. The faces did not include the hair and the back of the head. Also, features like beards, scars, glasses or earrings were excluded so that only the face itself could be used for recognition.

METHODS

Stimuli

For the generation of the face images, we used a data base of three-dimensional head models which had been collected by means of a Cyberware™ laser scanner. The scanner records shape and texture of a face simultaneously and with the same resolution, so that each surface coordinate is registered with exactly one texture pixel. The data base contains about 100 heads of caucasian people aged between 20 and 40 yr who volunteered to be scanned in our laboratory. The data contain the three-dimensional shape of the heads (i.e. the surface model) as well as the texture. If the texture is mapped onto the surface model, two-dimensional projections result in images that look like coloured photographs taken from the chosen angle. Shaded images can be produced by using only the shape data together with models about surface reflection and illumination. Further details appear later.

Before generating the images, the scanned head models were processed as follows. The hair region and the whole back of the head were removed completely. The ears, however, remained visible. The shoulders were also removed and only a small part of the neck was left. Since we did not want the skin colour to be a cue for recognition the texture map was normalized to provide similar skin colour in all faces. This was done by calculating the mean colour value in defined regions on the cheeks and the forehead. The derived skin colour values were then averaged over all faces to obtain the mean skin colour. Finally, the whole texture map of each face was scaled in a way to yield this mean skin colour.

The texture map was only used for the textured images. For the shaded images we applied a shading model assuming Lambertian reflection properties of the surface, parallel light from the direction of the camera and a small amount of additional ambient light. The images were generated by means of a raytracing algorithm implemented in the Wavefront™ rendering package.

Nine different views were calculated assuming perspective projection. The simulated pin-hole camera was set at a distance of 120 cm from the face. The views covered the whole range from the left profile to the right profile with an angle between two successive views of 22.5 deg. The faces were always presented in front of a black background. Figure 1 shows an example of three views of a face. The left column shows images from a textured face and the right column shows the same face using the shading model.

Procedure

Each subject was presented with 45 pairs of successively displayed views of faces. The faces were shown on a computer monitor. Their size was approximately 2/3 of their natural size and the viewing distance to the monitor was 80 cm, simulating a face seen at a distance of 120 cm.

Each subject saw either only textured faces or only shaded faces. The surface property thus was handled as a between-subject factor. Each pair of images was preceded by a short tone in order to draw the attention of the subject. Five hundred milliseconds later the first image appeared briefly (see below), followed immediately by a randomly coloured mask. The mask served to avoid retinal after-images. After another second the testing view was displayed for the same short period as the first image. It was also followed by a mask. The subject then had to decide whether or not the two views were from the same person and the answer was recorded.

TABLE 1. ANOVA for the error rates of the experiments with textured faces

Factor	d.f.	Sum of square	Mean square	F-Value	P
1. LV	4	3.48	0.869	$F(4,196) = 4.343$	<0.001
2. TV	8	0.92	0.115	$F(8,392) = 0.662$	>0.05
3. SJ*LV	196	39.23	0.200		
4. SJ*TV	392	68.10	0.174		
5. LV*TV	32	6.63	0.207	$F(32,1568) = 1.137$	>0.05
6. SJ*LV*TV	1568	285.46	0.182		

LV, learning view; TV, testing view; SJ, subject.

TABLE 2. ANOVA for the error rates of the experiments with shaded faces

Factor	d.f.	Sum of square	Mean square	F-value	P
1. LV	4	3.14	0.786	$F(4,156) = 6.408$	<0.001
2. TV	8	1.33	0.166	$F(8,312) = 0.981$	>0.05
3. SJ*LV	156	19.12	0.123		
4. SJ*TV	312	52.86	0.170		
5. LV*TV	32	13.90	0.434	$F(32,1248) = 2.470$	<0.001
6. SJ*LV*TV	1248	219.44	0.176		

Abbreviations as in Table 1.

The presentation time of the images was chosen so that the mean error rate of the responses was about 0.25. We did not perform systematic experiments to measure the dependence of the error rate on the presentation time. Pilot experiments gave similar error rates of 0.25 with presentation times of 165 msec for the textured faces and 1200 msec for the shaded faces.

Design

As within-subject factors we investigated the pose of the first face (the learning view, LV) and the pose of the second face (the testing view, TV). The learning view had five levels, namely 0, 22.5, 45, 67.5 and 90 deg (with respect to the full-face view). In half the cases the faces were shown from the left side and in the other half from the right side (see below). The testing view had nine levels, namely −90, −67.5, −45, −22.5, 0, 22.5, 45, 67.5 and 90 deg. The sign of TV does not reflect the absolute orientation (left or right), but its relationship to LV. Positive values for TV indicate that this is a view from the same side of the face as LV. Negative values for TV indicate that the face is shown from the contralateral side with respect to LV. In half of the cases a view with LV = 0 deg was treated as a left side view and in the other half as a right side view, so that even in this case the sign of TV did not correlate with the absolute orientation. Each subject had to perform 45 single trials which covered all possible combinations of the levels of LV and TV. Their order as well as the position of a particular face in the experiment were randomized according to the following constraints:

1. In half the trials, the first view was from the left side of the face and in the other half it was from the right side.
2. Half the trials actually showed views of the same person while the other half used a new, unknown face.
3. Half the trials consisted of pairs of male faces, while the other half used female faces. In order to avoid confusion of face recognition and gender classification, learning and testing view always showed faces of the same gender.
4. In all, 68 different heads were used (45 as targets and 23 as additional new faces). Each head appeared in only one trial.

In all 90 subjects, who were not familiar with the people in our database, participated in the experiment. Fifty subjects were tested with textured faces and 40 different subjects were tested with shaded faces. We used fewer subjects for the shaded faces because these results were clearer and more pronounced than for the textured faces.

Evaluation

We evaluated the mean error rate, which is the mean of the false alarm rate and the miss rate. A two-factor analysis of variance (ANOVA) was applied separately to the experiments using textured faces and shaded faces. Each design was balanced with respect to the five-level factor "learning view" (LV) and the nine-level factor "testing view" (TV).

RESULTS

Presentation time

The presentation times necessary to yield an average error rate of about 0.25 were estimated from pilot experiments. Textured faces were shown for 165 msec and shaded faces for 1200 msec, i.e. seven times as long. The average error rates calculated after the experiment were almost the same in both cases. The overall error rate

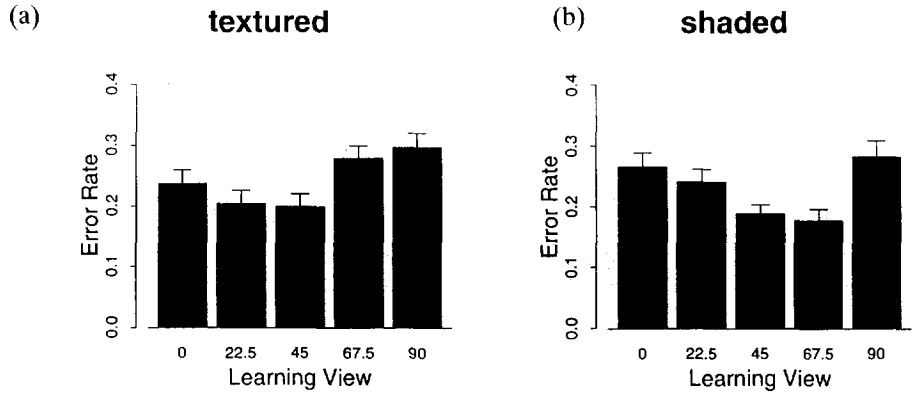

FIGURE 2. Mean error rates and the corresponding standard errors for the five different learning views. (a) Shows the results from experiments with textured faces and (b) the results from experiments with shaded faces.

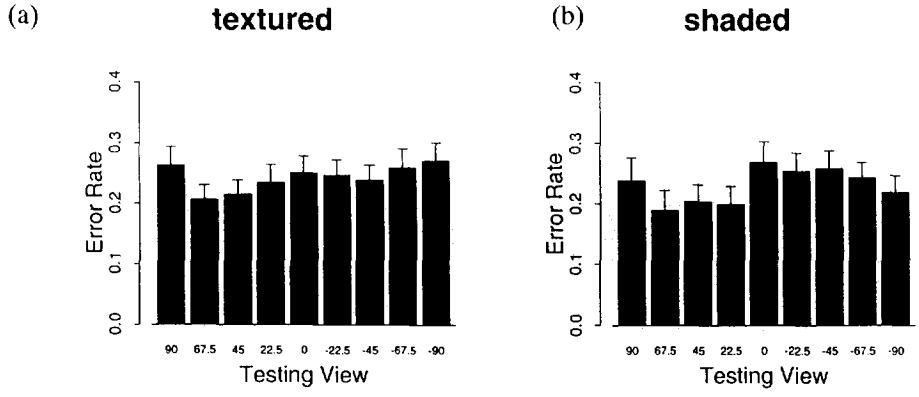

FIGURE 3. Mean error rates and standard errors for the nine levels of the testing view in the experiments using textured faces (a) and shaded faces (b).

for the textured faces was 0.244 and the one for shaded faces was 0.232.

Analysis of variance

The ANOVA tables are shown in Tables 1 and 2. For both the textured and the shaded faces there is an effect of the learning view but no effect of the testing view. In the case of the textured faces there is no interaction between learning and testing view, while for the shaded faces a strong interaction was found.

Effect of the learning view

The diagrams in Fig. 2(a) and (b) show the dependence of the mean error rate on the learning view for the textured and shaded faces, respectively. The learning view which resulted in the lowest error rate lies between 22.5 and 67.5 deg. The exact location of this optimal learning view, however, depends on the surface properties of the faces. For the textured faces this view is closer to the full-face view than for the shaded faces. We performed pairwise student t-tests between the mean error rates. For the textured faces the error rate corresponding to the 45 deg view differs significantly from the error rate corresponding to the 67.5 deg view ($t = 2.79$, $P < 0.005$), but it does not differ from the error rate corresponding to the 22.5 deg view ($t = 0.150$). This is quite different for the shaded faces. Here the error rate of the 67.5 deg view is still very low and does not differ significantly from the error rate of the 45 deg view ($t = 0.479$). The error rates of the 67.5 and the 90 deg view, however, are significantly different ($t = 3.349$, $P < 0.001$). The same holds if we compare the error rates of the 22.5 and the 45 deg view ($t = 2.12$, $P < 0.05$).

Effect of the testing view

Figure 3 illustrates the dependence of the error rate on the nine levels of the testing view. The diagrams show what the ANOVA already revealed. The error rate is independent of the testing view. The diagrams suggest a slight increase in the error rate from smaller to larger pose changes which, however, as shown by the ANOVA, is not significant. In Fig. 4, each bar combines the error rates corresponding to the respective symmetrical testing views, making the diagrams comparable to Fig. 2.

Effects of the pose change

Any effect of pose change should appear as an interaction between LV and TV. Figure 5 shows the error rates for all possible combinations of LV and TV in terms of intensity values. The low contrast in Fig. 5(a) reflects the absence of a significant interaction between LV and TV for the textured faces. Nevertheless, the

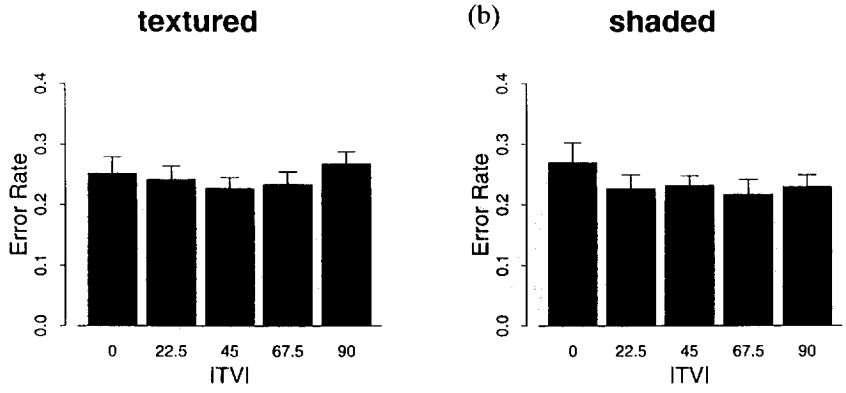

FIGURE 4. These diagrams result from those of Fig. 3 by combining columns corresponding to symmetrical testing views, i.e. error rates are plotted not with respect to the testing view itself but with respect to the absolute value of the testing view. This allows a better comparison with the dependence of the learning view shown in Fig. 2.

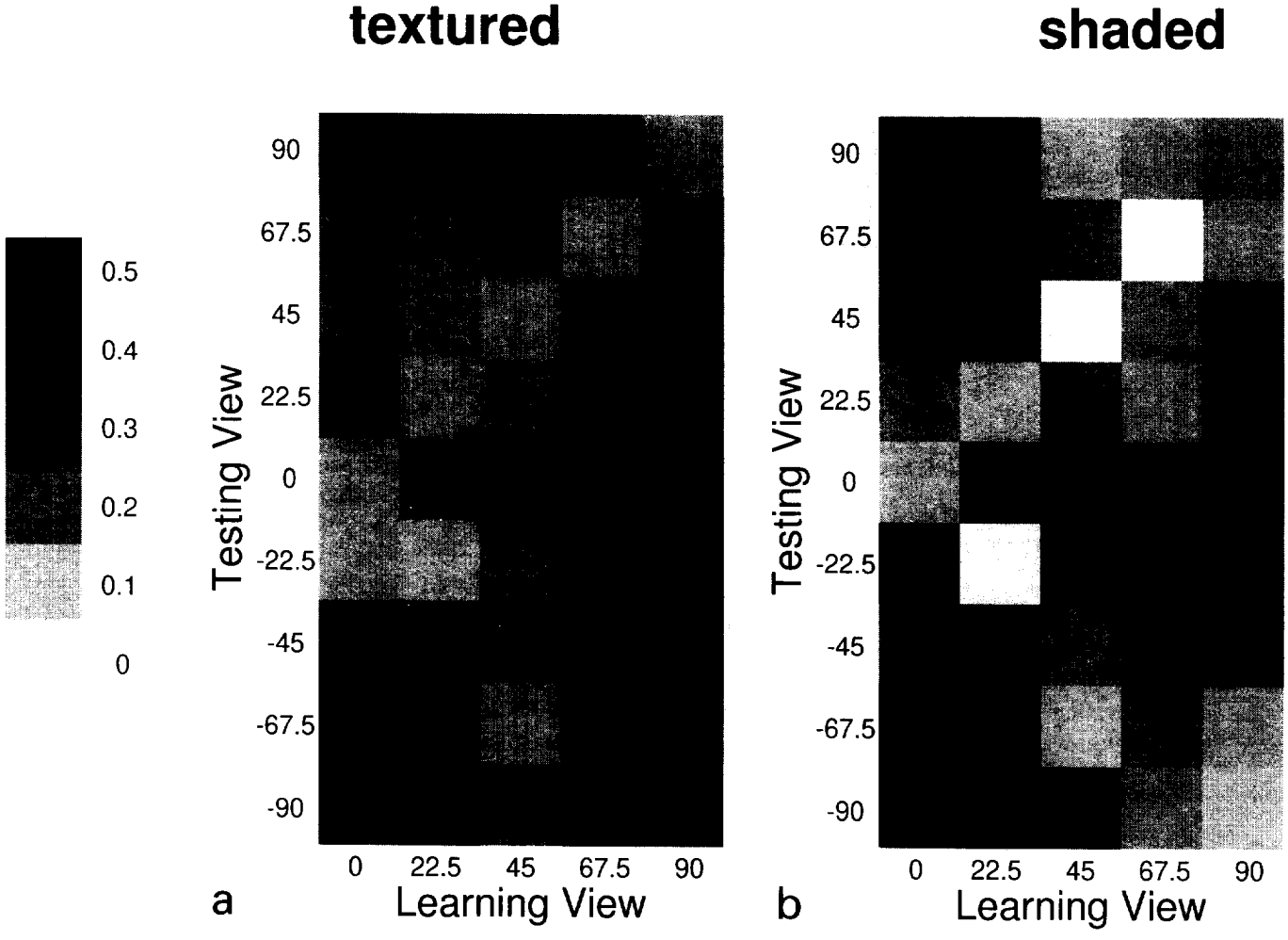

FIGURE 5. Error rates for each of the 45 combinations of LV and TV. Light patches denote low error rates and dark patches high error rates.

lowest error rates occur along the diagonal, which corresponds to the trials in which learning and testing view were identical.

Since we did not find an interaction for the textured faces we will now concentrate on the experiments using shaded faces, although we will continue to present the corresponding diagrams for the textured faces, as well. The pronounced interaction between LV and TV for the shaded faces is reflected by the higher contrast of Fig. 5(b). However, before investigating this interaction in detail, we have to note that neither LV nor TV are themselves really independent of pose change. The mean

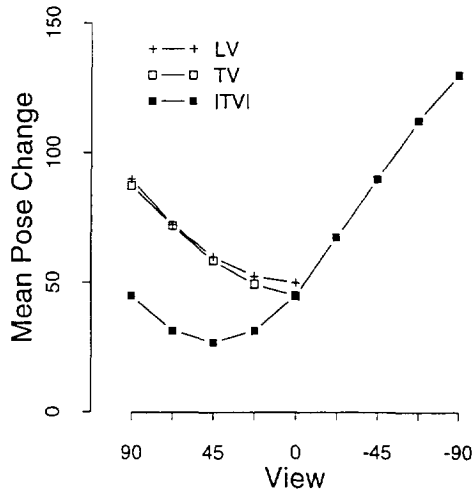

FIGURE 6. The mean pose change corresponding to the different levels of LV, TV and the absolute amount of TV.

pose change is not at all constant for the different levels of LV and TV. This is a consequence of the fact that the factors LV and TV were both balanced, and every combination of the two factors appeared exactly once. In Fig. 6 the mean pose change is plotted against the levels of LV, TV and the absolute value of TV. The strong increase in pose change associated with TV is mainly due to the fact that the negative sign always accounted for the contralateral view with respect to the learning view. The absence of an effect of TV therefore not only indicates that the orientation of the face in the testing image does not affect recognition performance, but also argues for the absence of an effect of pose change.

In Fig. 7 the error rate is plotted against the pose change. In fact there is no linear correlation between error rate and pose change. This is consistent with the absence of an effect of TV, but raises the question as to where the pronounced interaction between LV and TV for the shaded faces comes from. The shape of the histogram of Fig. 7(b) hints at an answer to this question.

Role of symmetric views

From Fig. 7(b) it is apparent that the recognition performance after pose changes of 180 deg is almost as good as in the cases without any pose change (0 deg pose change). The only trial in the experiment, however, that corresponds to a pose change of 180 deg is the one in which the learning view is the profile view and the testing view is the contralateral profile view. Does that mean that symmetric views are treated as being similar? If this were true, we would expect that the difference between the

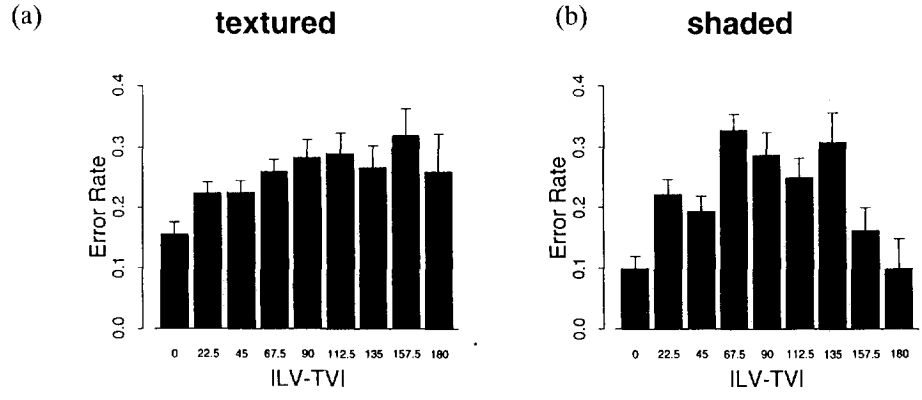

FIGURE 7. Dependence of the error rate on the pose change (i.e. the absolute amount of the difference of LV and TV). Note that in these and all the following diagrams the number of trials contributing to each bar is no longer constant (e.g. each subject performed only one trial with a pose change of 180 deg, but five trials with pose change of 0 deg). This is reflected in the standard error bars.

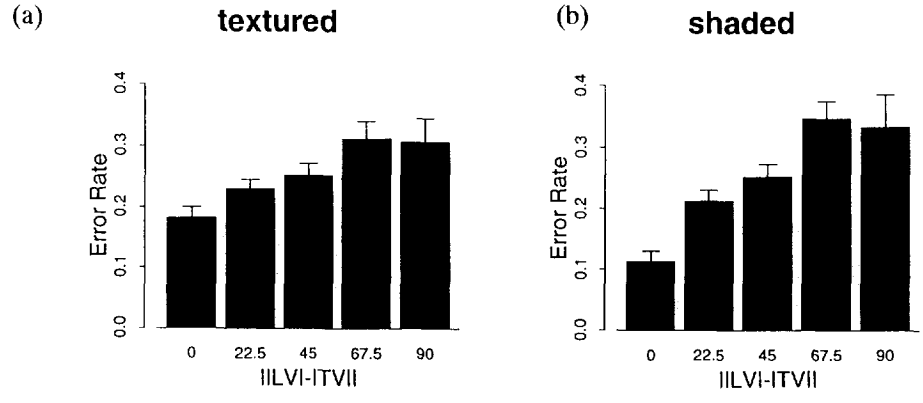

FIGURE 8. Error rates corresponding to the "symmetry corrected pose change", which is the difference between the learning view and the absolute amount of the testing view. For example the 22.5 deg bar contains trials with LV = 45 and TV = 67.5, but also trials with LV = 45 and TV = −67.5.

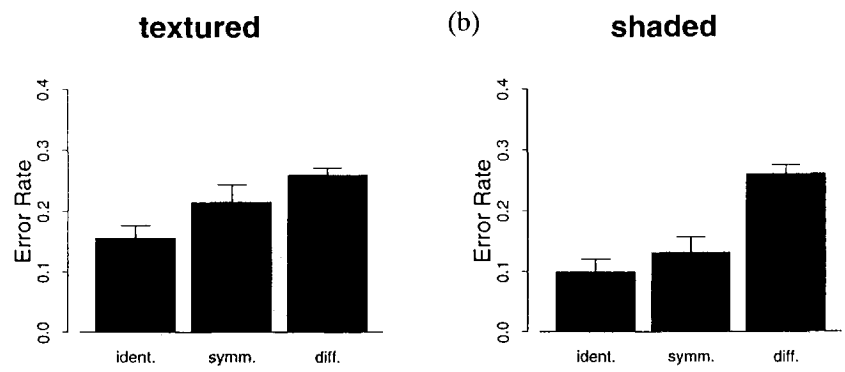

FIGURE 9. This diagram combines all trials with either identical views (LV=TV), symmetric views (LV = −TV) or otherwise different views (LV≠|TV|). The mean pose change for the trials with symmetric views is 112.5 deg and for the trials with otherwise different views 69 deg.

learning view and the *absolute value* of the testing view, rather than the pose change itself, would influence the recognition performance. We will call this quantity the "symmetry corrected pose change". In Fig. 8(b) the error rate is plotted against ||LV|−|TV||. In fact the error rate strongly increases with the symmetry corrected pose change (slope = 0.00254 deg^{-1}, $r = 0.952$, $P < 0.01$). We observe the same tendency for the textured faces (slope = 0.00146 deg^{-1}, $r = 0.962$), but it is less pronounced and since an interaction between LV and TV was missing, we have to treat the "symmetry effect" for textured faces with care.

Figure 9 illustrates the same phenomenon in a different way. The histograms show the mean error rates over all trials in which learning and testing view consisted of either identical views, two symmetric views, or otherwise different views. For the shaded faces the error rate for the symmetric views is statistically indistinguishable from the error rate corresponding to trials with similar views. The error rate for otherwise different views is significantly higher ($t = 4.48$, $P < 0.001$). We can see a similar but weaker trend for the textured faces. However, a t-test reveals a distinct difference only between the error rates corresponding to the same and to otherwise different views ($t = 4.50$, $P < 0.001$). The difference between the error rate corresponding to identical view and the error rate for the symmetric views is just at the border of significance [$t = 1.69$, $P(0.05, 98) = 1.66$] and the difference between the symmetric views and the otherwise different views is not at all significant ($t = 1.45$, $P > 0.05$). It should be considered in that context that the mean pose change corresponding to the trials using symmetric views is much larger (112.5 deg) than for the trials corresponding to otherwise different views (69 deg).

DISCUSSION

The results suggest pronounced differences between the role of learning and testing view. In addition, they are strongly dependent on whether we used naturally textured faces or faces without texture but with shading information. Symmetry also seems to play an important role.

Generalization from single learned views to novel views depends on the learning view but not on the testing view. The construction of a face-representation seems to be viewpoint dependent but the recall of the memorized information is characterized by a high degree of viewpoint invariance. The necessary presentation times for the textured faces were much shorter than for the shaded faces. Even if the presentation times were adjusted to yield similar overall error rates, the generalization to novel views is much better for the textured faces than for the shaded faces. This can be seen from the lack of interaction in the ANOVA, the lower contrasts in Fig. 5(a), and the flatter diagrams of Figs 7(a), 8(a) and 9(a).

The difference in generalization and viewpoint dependence for textured vs shaded faces has interesting parallels in other work on object recognition. If we compare the viewpoint dependence at different levels of recognition we find in many cases viewpoint independent recognition at the basic level (Biederman, 1987) and a viewpoint dependence at the subordinate level that is more or less pronounced, depending on the respective object class (Bülthoff & Edelman, 1992; Biederman & Gerhardstein, 1993). As discussed in Bülthoff *et al.* (1995) recognition processes based on localized features are rather viewpoint dependent and allow only for limited generalization. Other features (like colour or facial hair) are diffuse and should therefore be independent of the viewpoint. The shaded faces are lacking many diffuse features. Features that are maintained are often local and therefore viewpoint dependent (e.g. the profile of the nose).

While it does not depend on the testing view, recognition performance clearly depends on the learning view. The best generalization performance corresponds to neither the full-face view nor to the profile view but is found somewhere between 20 and 70 deg. The best angle for the learning view, however, depends on the stimulus type and is smaller for the textured faces than for the shaded faces. Since three-dimensional shape information is inherently viewpoint independent (apart from occlusion), it would be advantageous to recover as much as possible of the three-dimensional structure. The different optimal viewpoints for textured and shaded faces might have to do with different strategies employed to recover

three-dimensional structure from the images with textured and with shaded faces.

One possible strategy has to do with the fact that faces are more or less bilaterally symmetric. The crucial processes in the context of our experimental design certainly take place at the subordinate level of recognition (Rosch *et al.*, 1976) and we can assume that the observer already knows that he is confronted with a face. He could now use the general knowledge about the bilateral symmetry of faces to generate the view symmetric to the one he is actually seeing (Vetter *et al.*, 1994). This "virtual view" can than be used as a second view of the face to recover the three-dimensional structure. For a complete recovery, two views are only sufficient if the "camera parameters"—i.e. the viewing distance and the angles between the two views—are known. But even without complete information about the view geometry the interpretation of the seen object can be restricted to an invariant called the "affine structure" in the case of orthographic projection (Ullman & Basri, 1991; Koenderink & van Doorn, 1991; Vetter & Poggio, 1994) or "projective structure" (Shashua, 1993) in the case of perspective projection. However, a prerequisite is that enough feature points are visible in both images and that the correspondence problem can be solved. Since the second image in this case is only the "virtual" symmetric view, this requires that the features have to be visible in both halves of the face. In the full-face view all visible features can be seen in both halves of the face, but the virtual view would be identical with the original one. The viewing angle should not be too small to provide two different images. On the other hand, if the angle exceeds 30–40 deg, one eye is obscured by the nose and other features, such as the corner of the mouth, remain visible only in one half of the face. In our experiments the optimal learning view for the textured faces was somewhere between 25 and 40 deg. This might reflect that an algorithm based on an additional virtual view is used to extract invariants contained in the three-dimensional structure.

The shaded faces do not provide sharp contrasts that could be used as exactly localizable features. In general—i.e. if the direction of the illumination does not coincide with the symmetry plane of the head—shading is not symmetric in the two halves of the face. The shading itself, however, provides another cue for recovering three-dimensional shape (Horn & Brooks, 1989). Shape-from-shading algorithms are well known to be used in human perception (e.g. Bülthoff & Mallot, 1988; Todd & Reichel, 1989). Which viewpoint would be optimal for recovering three-dimensional invariants from a face by means of shape-from-shading information? Due to occlusion, only part of a head can be seen in a single view. Assuming bilateral symmetry, a full-face view or any view from a small angle with respect to the symmetry plane would be disadvantageous since it contains redundant information. A view from a large angle does not contain so much redundancy, but instead provides more details from the side of the head and the ears. This might be the reason for the larger optimal learning view for the shaded faces.

A very interesting point concerns the generalization to the symmetric view. Schyns and Bülthoff (1994), who also used non-textured surface models of faces, have already suggested that generalization to the symmetric view is better than to other views. Our experiment also showed that although generalization to novel views is generally relatively poor for shaded faces, the symmetric view is recognized almost as well as the identical view (Fig. 9). This is not the case for the textured faces—at least not to the same extent. A possible interpretation for that finding is the following: although faces are approximately bilaterally symmetric, this symmetry is never perfect. The images taken from symmetric views thus cannot be expected to be perfectly mirror symmetric as well. Although in our experiments we only used faces from young people without any unusual features, there still might be slight asymmetries due to scars, pimples or careless shaving, in the data base. An asymmetry of the eyes arises when a person does not look straight ahead but somewhere to the side. These asymmetries, however, would be more pronounced in the texture of the faces than in their shape. The two symmetric views of a shaded head therefore lead to images which are approaching symmetry better than the two symmetric views of a textured head. The comparatively bad generalization to the symmetric view for the textured heads might reflect asymmetries in the texture of the displayed heads.

Another, slightly different interpretation is not based on the possible asymmetries in the texture of the used head models. Asymmetries in the texture of the displayed head models might be negligible but, nevertheless, asymmetries are generally more pronounced in texture than in shape. The difference in the recognition performance for symmetric views of shaded vs textured faces might reflect that we do not use symmetry in the texture because it is less reliable than symmetry in the shape of a head.

If the first hypothesis were true, generalization to the symmetric view would increase if we would have shown the mirror symmetric image of the same view instead of the symmetric view of a textured three-dimensional head model. In the case that the second hypothesis is true, the replacement would not alter the results.

A question that is closely related to the one discussed above is whether the generalization to the symmetric view is based on the symmetry of the three-dimensional object or rather on the mirror symmetry of the corresponding images. Our experiments cannot distinguish between these possibilities. If faces were completely symmetric, the images of the two symmetric views would be perfectly symmetric too. For the generation of the shaded images, we simulated directional light coming from the direction of the camera. If we use light that deviates from the camera direction, the images of the symmetric views will no longer be mirror symmetric. On that basis, we are currently designing experiments that will be able to distinguish whether the observed

phenomenon is based on two-dimensional or three-dimensional symmetry.

The finding that generalization to the symmetric view is substantially better than to other views might explain a contradiction between previous studies. All the investigations discussed in the introduction are consistent in claiming an advantage of the 3/4 view during learning. However, there are differences concerning the effect of the pose of the testing view. Although all other studies did not find an effect of the pose of the second view, Bruce *et al.* (1987) reported that effect in their experiment. This might be due to the fact that their paradigm differed from the others, since the presentation times were long enough to reduce virtually any recognition errors, and rather than error rates or correct response rates the response reaction time was evaluated. However, here we want to propose another explanation. As described above, Bruce *et al.* used a 3×3 design to investigate the effects of the pose of the testing view and the pose change. For each level of the pose, they tested three pose changes, namely 0, 45 and 90 deg. The 90 deg pose change was realized as a change from full-face to profile in the case of the "profile" level of the testing pose, and from profile to full-face in the case of the "full-face" level. For the case of the 3/4 view, the 90 deg pose change was realized as a change from the 3/4 view to the symmetric, contralateral 3/4 view. In that combination, recognition performance is much better than for the other 90 deg pose changes. In fact, this data point accounts for a great part of the performance, associated with the 3/4 view level of the factor "pose". In terms of the "symmetry corrected pose change" (Fig. 8) the three pose changes used for the "3/4 view" level of factor "pose" correspond to the three values 0, 45, 0 deg instead of 0, 45, 90 deg for the "pose" levels "full-face" and "profile". The stimuli used by the authors were photographs, i.e. they are comparable to our textured stimuli, but—depending on the illumination used—probably also contained additional shading information. From Figs 8(a) and 9(a) it can be seen that even for the textured faces, there is a tendency that the symmetric views show better recognition performance then otherwise different views.

In contrast to some of the discussed studies (Krouse, 1981; Bruce *et al.*, 1987), we could not find a significant interaction between learning and testing view for the textured faces. The other studies used only two or three levels for the testing view, whereas in our experiment we used nine different levels. On the other hand, we have less data for each single level and smaller changes between neighbouring levels. If we pool the pose changes producing only the two levels "matched" and "unmatched" as in the work of Krouse (1981) the effect of the new two-level factor becomes highly significant even for the textured faces [$F(1,49) = 28.5$, $P < 0.001$]. If we introduce the "symmetry corrected pose change" of Fig. 8 as a factor, we also get significant effects [$F(4,196) = 5.56$, $P < 0.001$]. However, since these factors are introduced *post hoc* and their levels are not at all balanced, these results have to be treated with care. Nevertheless, they provide valuable cues for designing further experiments, which are more sensitive to parameters such as symmetry and symmetry corrected pose change.

REFERENCES

Beymer, D., Shashua, A. & Poggio, T. (1993). *Example based image analysis and synthesis* (pp. 1–20). Cambridge, MA: M.I.T. Artificial Intelligence Laboratory, Memo No. 1431.

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

Biederman, I. & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for 3D viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 1162–1182.

Bruce, V., Valentine, T. & Baddeley, A. D. (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology, 1*, 109–120.

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 USA, 89*, 60–64.

Bülthoff, H. H., Edelman, S. & Tarr, M. J. (1995) How are three dimensional objects represented in the brain? *Cerebral Cortex, 5*, 247–260.

Bülthoff, H. H. & Mallot, H. A. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America A, 5*, 1749–1758.

Davies, G., Ellis, H. & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology, 63*, 180–187.

Horn, B. K. P. & Brooks, M. J. (Eds) *Shape from shading*. Cambridge, MA: MIT Press, 1989.

Koenderink, J. & van Doorn, A. (1991). Affine structure from motion. *Journal of the Optical Society of America A, 8*, 377–385.

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

Lades, M., Vorbrüggen, J. C., Buhmann, J., Lange, J., Malsburg, v.d., C., Würtz, R. P. & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers, 42*, 300–311.

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., Bülthoff, H. H., Troje, N. F. & Vetter, T. (1995) Face recognition across large viewpoint changes. *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition* (pp. 326–331). Zürich: Multimedia Lab.

Patterson, K. E. & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 3*, 406–417.

Rosch, E., Merwis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Schyns, P. G. & Bülthoff, H. H. (1994). Viewpoint dependence and face recognition. *Proceedings of the XVI Meeting of the Cognitive Science Society*, 789–793.

Shashua, A. (1993). Projective depth: A geometric invariant for 3D reconstruction from two perspective/orthographic views and for visual recognition. *Proceedings of the International Conference on Computer Vision (ICCV)*, 583–590.

Todd, J. F. & Reichel, F. D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review, 96*, 643–657.

Ullman, S. & Basri, R. (1991). Recognition by linear combination of

models. *IEEE Transactions on Pattern and Machine Intelligence*, *13*, 992–1006.

Vetter, T. & Poggio, T. (1994). Symmetric 3D objects are an easy case for 2D object recognition. *Spatial Vision*, *8*, 443–453.

Vetter, T., Poggio, T. & Bülthoff, H. H. (1994) The importance of symmetry and virtual views in three-dimensional object recognition. *Current Biology*, *4*, 18–23.

Wogalter, M. S. & Laughery, K. R. (1987). Face recognition: Effects of study to test maintenance and change of photographic mode and pose. *Applied Cognitive Psychology*, *1*, 241–253.