Using 3D computer graphics for perception: The role of local and global information in face processing

Adrian Schwaninger[*,1,2] Sandra Schumacher[†,2] Christian Wallraven[‡,1]
Heinrich Bülthoff[§,1]

[1] Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2] Department of Psychology, University of Zürich, Switzerland

Abstract

Everyday life requires us to recognize faces under transient changes in pose, expression and lighting conditions. Despite this, humans are adept at recognizing familiar faces. In this study, we focused on determining the types of information human observers use to recognize faces across variations in viewpoint. Of specific interest was whether holistic information is used exclusively, or whether the local information contained in facial parts (featural or component information), as well as their spatial relationships (configural information) is also encoded. A rigorous study investigating this question has not previously been possible, as the generation of a suitable set of stimuli using standard image manipulation techniques was not feasible. A 3D database of faces that have been processed to extract morphable models (Blanz & Vetter, 1999) allows us to generate such stimuli efficiently and with a high degree of control over display parameters. Three experiments were conducted, modeled after the inter-extra-ortho experiments by Bülthoff & Edelman, 1992. The first experiment served as a baseline for the subsequent two experiments. Ten face-stimuli were presented from a frontal view and from a 45° side view. At test, they had to be recognized among ten distractor faces shown from different viewpoints. We found systematic effects of viewpoint, in that the recognition performance increased as the angle between the learned view and the tested view decreased. This finding is consistent with face processing models based on 2D-view interpolation. Experiments 2 and 3 were the same as Experiment 1 expect for the fact that in the testing phase, the faces were presented scrambled or blurred. Scrambling was used to isolate featural from configural information. Blurring was used to provide stimuli in which local featural information was reduced. The results demonstrated that human observers are capable of recognizing faces across different viewpoints on the sole basis of isolated featural information and of isolated configural information.

Keywords: psychophysics, viewpoint generalization, face recognition

1 Introduction

We can recognize familiar and unfamiliar faces reasonably well despite transient changes in their pose (orientation relative to the viewer), expression and lighting conditions (e.g. Moses, Ullman, & Edelman, 1996; Troje & Bülthoff, 1996). The extent of these capabilities is evidenced by the fact that, despite several promising attempts in the last few years, human-like invariance to different conditions has not been achieved by any artificial recognition system. There is still a long way to go before completely understanding the basic abilities underlying face recognition.

Obtaining a better understanding of face recognition processes is important, not only to the scientific community but also for everyday life. One important application of such knowledge is in the field of automatic face identification. Biometric methods are useful for this purpose as biometric characteristics are inseparably linked to their owner and as such, cannot be lost and cannot be forgotten. As a result, they are simple to use and highly reliable. Automatic face-identification, along with iris recognition and fingerprints, is one of the most common biometric methods used today. The advantage of automatic face identification over other biometric methods is that it can recognize people from a greater distance. Iris recognition and fingerprints can only be used when a person is within close contact with the sensor. In contrast, automatic face recognition enables us to assess one's identity without disturbing or annoying them in their everyday lives by close contact. Automatic face-detection systems could therefore play a useful role in the context of criminal prevention and criminal prosecution. However, artificial face identification and verification is still at an early stage, despite great progress in the last few years. Currently, it is possible that security problems may arise, not because face-recognition systems are unavailable, but because existing face-recognition systems are highly inaccurate. For instance, a face-recognition system that is too sensitive can paralyze a security system with a vast numbers of false alarms. Conversely, an inaccurate face-recognition system that is too insensitive can lull people into a false sense of security. In short, much work remains to be done in the field of artificial face recognition.

How does artificial face recognition relate to human face recognition? As mentioned above, human face recognition is a highly complex process with which we are able to recognize faces reasonably well across a wide variety of viewing conditions. Understanding human face recognition and its underlying processes will not only help to better understand human perception, but can also provide important insights into the refinement of automatic face recognition systems. See also the recent publication by Sinha et al., 2006 which attempts to gather several findings from perceptual research for the benefit of developing better computer vision algorithms. It seems clear that any attempt to explain the structural encoding process of face recognition, whether from the standpoint of computer science or from that of cognitive psychology, needs to address how a system can match currently viewed and previously stored faces despite different presentation conditions (Bruce, Burton, & Craw, 1992; Bruce, 1994). This study focuses specifically on determining the type of information human observers use to recognize faces across *variations in viewpoint*. In particular, we are interested in evaluating whether holistic information is used exclusively, or whether local information about facial parts (featural or component infor-

---

[*] e-mail: a.schwaninger@psychologie.uzh.ch
[†] e-mail: s.schumacher@psychologie.uzh.ch
[‡] e-mail: christian.wallraven@tuebingen.mpg.de
[§] e-mail: heinrich.buelthoff@tuebingen.mpg.de

APGV 2007, Tübingen, Germany, July 26–27, 2007.
© 2007 ACM 978-1-59593-670-7/07/0007 $5.00

mation) as well as their spatial relationship (configural information) is also encoded. Before describing the current experiments, we first review related work within this area. Broadly speaking, we can distinguish between two groups in the face perception literature: studies which investigate face recognition across different viewpoints and studies which focus on the particular types of information (e.g. local parts, spatial relationship between features, etc.) that human observers use for face recognition.

1.1 Face recognition across different viewpoints

Many face recognition experiments are motivated by view-based theories. Common to all view-based theories is the assumption that objects are stored in memory as a collection of discrete views (e.g., Bülthoff & Edelman, 1992). Such theories may differ however regarding the mechanisms by which new object views are compared with stored views, and similarly, how information from stored views generalizes to novel views. Some theories, for example, are based on an alignment to a 3D representation (e.g., Lowe, 1987). These theories assume that object recognition is more or less viewpoint-independent. In contrast, according to the linear combination approach (Ullman & Basri, 1991), new views are recognized based on linear combination of stored 2D views. Theories based on recognition by 2D view interpolation (e.g., Bülthoff & Edelman, 1992; Wallraven, Schwaninger, Schumacher, & Bülthoff, 2002) assume that recognition performance increases as the angle between the stored view and the new view decreases. In two earlier studies (Wallraven et al., 2002; Schwaninger, Schumacher, Wallraven, & Bülthoff, submitted), we evaluated which of the three approaches described above is the most adequate for face processing. In both studies we found a strong viewpoint effect for face recognition, such that generalization from a learned view decreased with an increasing angle of rotation. This finding is consistent with a *linear interpolation model.*

Another effect related to recognition performance across viewpoint changes is the symmetry effect (e.g. Vetter, Poggio, & Bülthoff, 1994; Troje & Bülthoff, 1998; Hill, Schyns, & Akamatsu, 1997). This effect is based on object-specific characteristics, and more specifically the bilateral symmetry of some objects. An important study investigating the recognition of symmetrical objects was conducted by Vetter, Poggio and Bülthoff (1994). Using tube-like objects, they were able to demonstrate that left-right symmetrical objects show a peak in recognition performance for the opposite symmetric view. Based on their findings, they developed a theory of ”virtual views”. This theory postulates that, if an object is known to be left-right symmetric and the symmetry point can be inferred from the viewing direction, a symmetric view around the axis of symmetry can be internally generated. This generated view has been called the ”virtual view”, in contrast to the ”real view” that the subject has actually visually previewed. Hill, Schyns and Akamatsu (1997) have shown that a similar effect also exists for symmetric views of faces. In their study, they found that a learned three-quarter view generalizes better to the opposite three-quarter view than to other unseen views.

A third effect often mentioned in connection with object recognition across different viewpoints relates to what is known as the canonical view. A canonical view of an object can be named ”better”, faster and more accurately than other views of the same object (Palmer, Rosch, & Chase, 1981). In face perception research, it has long been speculated as to whether canonical views also exist for faces similar to those found for objects. A candidate for a possible canonical view is the 3/4 view (e.g. O’Toole, Edelman, & Bülthoff, 1998), located approximately midway between a full-face (or frontal) view and a profile view. However, there is no clear evidence showing that one specific canonical view exists for face recognition. Rather, there seems to be a range of views which are canonical or optimal. In a recent study (Schwaninger et al., submitted), we found that the effective views may range more widely along horizontal than along vertical meridians.

1.2 Holistic, part-based or configural information

Several studies have suggested that face recognition is based on holistic processing. This view is supported by findings showing that humans are slower and less accurate at recognizing the top half of one face presented together with the bottom half of another face when the two parts are upright and aligned than when the two parts are inverted or offset laterally - manipulations that disrupt holistic processing (Young, Hellawell, & Hay, 1987). Further, human observers are more accurate in recognizing the identity of a facial feature when it is presented in the context of the whole face compared to when presented as an isolated feature (Tanaka & Farah, 1993). This does not mean, however, that face processing is exclusively holistic. Other processes, like the local processing of facial features (”component information”) and the processing of information about the spatial layout or configuration of these features (”configural information”) seem to play an important role in the face recognition process for recent reviews see Schwaninger, Carbon, & Leder, 2003; Schwaninger, Wallraven, Cunningham, & Chiller-Glaus, 2006). A still controversial issue in the literature is the question of the relative contribution of featural processing and configural processing to face recognition (e.g., Cabeza & Kato, 2000). Some studies take an extremely holistic view, even assuming that featural processing is irrelevant or does not exist. Such an extreme view can be found, for example, in Tanaka & Farah (1993). In other studies, a primarily featural view is proposed. Rakover and Teucher (1997), for example, have proposed a model that predicts satisfactorily upright face recognition performance and inversion effects based on facial features only. Several other studies have indicated that both featural and configural processing are important in face processing (e.g. Diamond & Carey, 1986; Schwaninger, & Mast, 1999; Leder, Candrian, Huber, and Bruce, 2001). Clear support for this assumption comes from a study conducted by Schwaninger, Lobmaier and Collishaw (2002), who demonstrate that both featural and configural types of information are important for frontal, same-view face recognition. Their results suggest that configural and featural information can be encoded and stored independently and in parallel of one another. Similar results have been reported recently by Hayward, Rhodes, and Schwaninger (in press) for own and other race face recognition.

Overview of the current experiments
In the experiments reported in this paper, we investigated the question of whether human observers use holistic information exclusively to recognize faces across variations in viewpoint, or whether they encode and store also the local information contained in facial parts (featural or component information) and their spatial relationships (configural information). As shown by Schwaninger et al. (2002), human observers are able to recognize frontal faces when learned and tested from the same view because of featural as well as configural information which can be used independently during recognition. In this study, we wanted to examine the question of whether the same processes that we find for same-view recognition also underlie *different-view recognition*. Furthermore, we wanted to investigate whether the effects that we found in connection with whole faces (recognition by view interpolation, symmetry effect and range of canonical views along the horizontal meridian, Schwaninger et al. submitted) can also be found in connection with featural processing and configural processing. In contrast to previous studies, we used the method of Schwaninger et al. (2002) which does not alter configural or featural information, but instead, systematically eliminates one or the other. Scrambling faces into

constituent parts (such as eyes, nose, mouth, etc.) is used to isolate local part-based information. Previous studies have often attempted to directly alter the facial features or their spatial positions. The effects of such manipulations are not always perfectly selective. For example, altering featural information by replacing the eyes and mouth with those of another face could also change their spatial relations (configural information; as noted in Rhodes, Brake, & Atkinson, 1993). Also, Rakover (2002) has indicated that altering configural information by increasing the inter-eye distance could also indirectly induce a change in featural information. For instance, the bridge of the nose might appear wider. To exclude such problems, we used scrambling and blurring procedures in our experiments, which allowed us to investigate the roles of featural and configural information separately. The experiments reported in this study were modeled after the inter-extra-ortho experiment by Bülthoff and Edelman (1992). In order to generate the stimuli for these experiments, it was necessary to scramble faces into their facial components (eyes, mouth, nose, etc.), for details on scrambling in 2D see Schwaninger et al. (2002). In the following experiments, we made use of a database of 3D faces (http://faces.kyb.tuebingen.mpg.de) for which:

- it was easy to specify the facial components for all faces as they were in correspondence
- generation of different views was simply a matter of rotating the 3D model (see also the study by Wallraven et al. (2002)
- standard computer graphics rendering techniques ensured that different faces and views were rendered using controlled lighting and poses

2 Experiments

2.1 Experiment 1: whole-face condition

As stated above, to the goal of this study was to determine which kind of information human observers use to recognize faces across variations in viewpoint; specifically, whether they use holistic information exclusively, or whether they also encode and store the local information contained in facial parts (featural or component information) as well as their spatial relationship (configural information). The second aim was to investigate the question of whether the effects we found in connection with whole faces recognition (recognition by view interpolation, symmetry effect and range of canonical views along the horizontal meridian) can also be found in connection with featural processing and configural processing. To compare featural and configural processing with whole-face processing, we needed a reference condition. Therefore, Experiment 1 served as a baseline for the subsequent experiments.

2.1.1 Participants, Materials and Procedure

Ten right-handed undergraduate students of the University of Zürich volunteered for this experiment. All reported normal or corrected-to-normal vision. The stimuli were presented on a 17" CRT screen. The viewing distance of 60 cm was maintained by a head rest so that the face stimuli covered approximately 6° of the visual field.

The stimuli were created using the MPI face-database (http://faces.kyb.tuebingen.mpg.de), which consists of high-resolution laser scans of 200 Caucasian individuals (100 male, 100 female). Each laser scan has both full three-dimensional information (X,Y,Z coordinates for 70,000 points) and texture information (R,G,B values for a texture map of 512x512 pixels). Using a technique developed by Blanz & Vetter (1999), all 200 scans are brought into correspondence, such that, e.g., vertex 22,345 describes the tip of the nose in each laser scan. For this experiment, we used 20 male faces from the database rendered in a frontal pose (i.e. facing into the virtual camera), which was placed at a distance of 1.35m to the face. In addition, each face was illuminated using a stationary point-light source coming from slightly above and to the right. This particular illumination was chosen to enhance the depth perception of the stimuli. Ambient light with a strength of 20% white was then added to the scene in order to avoid hard shadows on the face. Each face was rendered on a black background with an image size of 512x512 pixels. All faces were then rescaled to 246x246 pixels.

The experiment consisted of a learning phase and a testing phase. Ten faces were randomly selected as distractors and another ten faces were selected as targets. In the learning phase, the target faces were shown oscillating horizontally and vertically $\pm 5^\circ$ around the 0° front and the 45° right view (see Figure 1). The views of the motion sequence were separated by 1°. All faces were first shown from the frontal view, followed by a 45° side view. The oscillations around 0° always started and ended with the 0° view. The oscillations around 45° always started and ended with the 45° right view. Half of the target faces were first oscillated two full cycles horizontally and then two full cycles vertically, while the other half of the target faces were first oscillated vertically and then horizontally. These conditions were counterbalanced across participants. Both motion sequences (0° and 45°) lasted 7.5 sec each. The sequence of the 10 faces was counterbalanced across the 10 participants. The learning phase was split into four blocks. In each block, the faces were shown twice from the frontal view and twice from the side view. The sequence of the faces was the same in each repetition. Between the second and the third block, there was a short break of 15 minutes. In the testing phase, we presented the subjects with static views of the 10 target and the 10 distractor faces. The faces were shown in blocks of 20 trials in which each face was presented once in a random order. The test phase contained 260 trials, each face being presented once from each of the 13 angles depicted in Figure 1. Every trial started with a 1000 ms fixation cross followed by the presentation of a face. Participants were instructed to respond as quickly and accurately as possible. They had to decide whether the presented face had been shown in the learning phase (i.e. whether it was a target) or whether it was a distractor by pressing the left or right mouse button. In each trial, the faces were presented until the button press occurred. The assignment of buttons to responses was counterbalanced across participants.

2.1.2 Results and Discussion

Face recognition performance was measured using signal detection theory (Green & Swets, 1966) by calculating d' using an old-new recognition task. The relevant measure is $d' = z(H) - z(FA)$, where H represents the hit rate (i.e. the proportion of correctly identified targets) and FA represents the false alarm rate (i.e. the proportion of incorrectly reporting that a face had been learned in the learning phase). H and FA are converted into z-scores (i.e. to standard deviation units).

We first carried out a one sample t-test (one-tailed) for each of the 13 angles in order to test the group means against chance performance (i.e. $d' = 0$). For all 13 test angles, recognition performance differed significantly from chance ($p < .05$). In a next step, individually calculated d' values were subjected to a two-factor analysis of variance (ANOVA) with condition (extra, inter, ortho (up), ortho (down)) and amount of rotation (0°, 15°, 30°, 45°) as within-subjects factors. Mean values are shown in Figure 2. Recognition d' was found to be dependent on the condition, as indicated by the main effect of condition $F(3, 27) =$

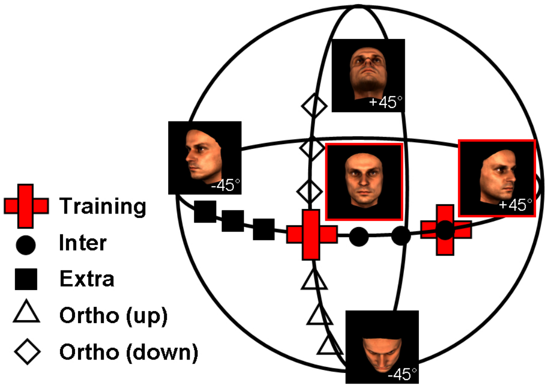

Figure 1: *Training occurred at 0° (frontal view) and at 45° (side view) oscillating 5° horizontally and vertically. Testing was performed for 13 views separated by 15°. The four testing conditions are labeled (inter, extra, ortho (up), ortho (down)).*

$22.91, p < .001$. There was also a main effect of amount of rotation, $F(3, 27) = 29.35, p < .001$. As illustrated by Figure 2, the effect of amount of rotation differed across conditions. This is also indicated by the interaction between amount of rotation and condition, $F(9, 81) = 7.29, p < .001$.

The four conditions were compared to each other using pairwise comparisons of the SPSS repeated measures ANOVA procedure. There was no significant difference in recognition between the inter and the extra condition ($p = .083$), nor did the two ortho conditions differ significantly ($p = .080$). Recognition performance in the inter and the extra conditions, however, was better than in either of the ortho conditions ($p < .01$).

In summary, the intact faces were recognized significantly better than chance from all 13 tested angles. This is a clear indication that human observers are capable of generalizing from stored views of an intact face to new views of the same intact face, at least within a certain range. We also found a systematic effect of viewpoint in that the recognition performance increased with decreasing angle of rotation. This is consistent with 2D view interpolation models (e.g., Bülthoff & Edelman, 1992; Wallraven et al., 2002; Schwaninger et al., submitted). The results also showed enhanced recognition performance for the opposite 45° view as suggested by the symmetry effect and an advantage in recognition performance for the horizontal views over the vertical.

2.2 Experiment 2: scrambled-face condition

As discussed above, Schwaninger et al. (2002) have shown that human observers store featural information independently of configural information. They also showed that, at least in a frontal same-view task, the two types of information can be recalled independently of each other (see also Hayward, et al., in press). In Experiment 2, we wanted to answer the question of whether these findings also account for different-view recognition; more specifically, whether isolated featural information can be processed for different-view recognition, or whether in this case featural and configural information can only be recalled dependently of each other or holistically. In addition, we investigated whether mechanisms such as recognition by view interpolation, the symmetry effect, or the range of canonical views along the horizontal meridian, which were found in connection with whole-face recognition, could also account for isolated featural information. To isolate the featural information, we broke up the faces into their constituent parts and scrambled them. These manipulations destroyed the configural information while leaving the featural information intact.

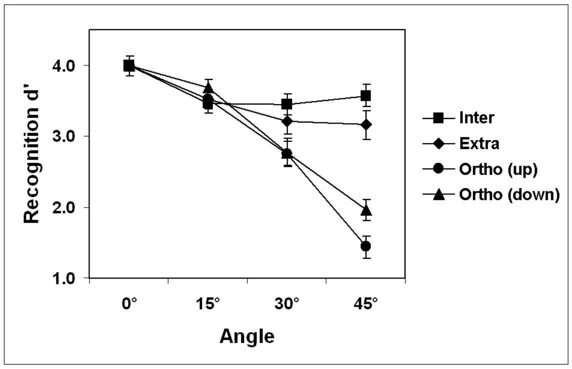

Figure 2: *Human recognition performance for whole-faces in the four rotation conditions inter, extra, ortho (up), ortho (down) across viewpoint (0° is the frontal view).*

2.2.1 Participants, Materials and Procedure

Ten undergraduates of the University of Zurich volunteered to take part in this experiment. All reported normal or corrected-to-normal vision. The procedure was the same as in Experiment 1. Again, the same twenty male faces served as stimuli - in the testing phase, however, the faces were presented scrambled (see Figure 3b). In order to create scrambled stimuli we used the correspondence property of the database: as each scan is in full correspondence to all other scans, it is sufficient to specify the extent of each part for only one scan - all other scans will then produce the same anatomical parts. It is important to note that this is a three-dimensional procedure - the resulting parts are defined on the three-dimensional surface of the scan. In order to reduce artifacts from sharp boundaries, each part was blurred around its edge during the rendering step with a small radius of around 4 pixels. All scrambled faces were then rescaled to 246x246 pixels. The number of parts was defined by a preliminary free listing experiment, in which 41 participants listed all parts of a face. The following parts were named by more than 80% of the participants and were used in this study: eyes, eyebrows, nose, forehead, cheeks, mouth, and chin.

2.2.2 Results and Discussion

As a first step we combined the d' data from experiments 1 and 2 to conduct a three-factor ANOVA with the within-subjects factors condition (extra, inter, ortho (up), ortho (down), and amount of rotation (0°, 15°, 30°, 45°), and the between-subjects factor stimulus type (intact face vs. scrambled face). We found a main effect of stimulus type $F(1, 18) = 67.41, p < .001$. In addition, we found main effects for the within-factors condition $F(3, 54) = 18,02, p < .001$ and amount of rotation, $F(1.89, 33.97) = 37.99, p < .001$. There was also interaction between stimulus type and condition, $F(3, 54) = 3,69, p < .05$, and an interaction between stimulus type and amount of rotation, $F(1.89, 33.97) = 4.34, p < .05$. We also found an interaction between amount of rotation and condition $F(5.14, 92.47) = 4.89, p < .001$ and a three-way interaction between amount of rotation, condition and stimulus type $F(5.14, 92.47) = 4.58, p < .001$. A paired-sample t-test between the two stimulus types showed a significantly better overall recognition performance for the whole-face compared to the scrambled-face condition ($p < .001$).

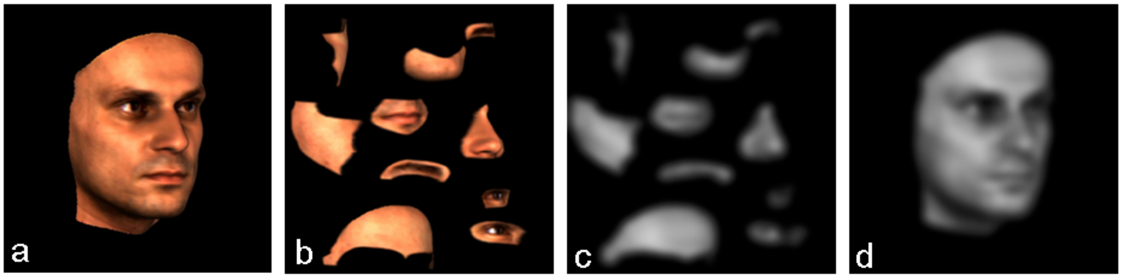

Figure 3: *Test conditions. a) whole-face, b) scrambled-face, c) scrambled-blurred-face, d) blurred-face*

To check if the recognition performance for scrambled faces is still higher than chance performance, we conducted one-sample t-tests (one-tailed) for each of the 13 test angles in order to test the group means against chance performance ($d' = 0$). In this case, the t-tests revealed a significant difference from chance ($p < .05$) in 10 of the 13 tested views. Face recognition performance did not differ significantly from chance in the 45° left view, the 45° up view and the 45° down view. Figure 2 shows that the two vertical 45° conditions for whole-face performance also had the lowest recognition performance of all 13 views. We suspect that the low recognition performance for some of the views is a consequence of the lower overall performance for scrambled faces compared to blurred faces. All in all, we can say that isolated featural information can also be processed for different-view recognition, but that it leads to a drop in recognition performance. To obtain a more exact picture of the differences between featural processing and whole face processing, we also conducted a two-factor ANOVA including the factors of condition (extra, inter, ortho (up), ortho (down) and amount of rotation (0°, 15°, 30°, 45°). For this calculation, we only took into account the data from the second experiment. Mean values are shown in Figure 4. Again, main effects of condition (inter, extra, ortho (right) or ortho (left) $F(3, 27) = 3.24, p < .05$ and amount of rotation (0°, 15°, 30°, 45°) $F(1.60, 14.42) = 10.01, p < .01$ were found. In this case, there was no interaction between amount of rotation and condition $F(4.21, 37.85) = 1.30, p > .28$. Pairwise-comparison t-tests between the four conditions showed no significant difference in recognition performance between the inter condition and the extra condition ($p > .09$). There was also no significant difference between the extra condition and the two ortho conditions (ortho (up) $p > .57$; ortho (down) $p > .11$) and between the two ortho conditions ($p > .83$). A significant difference was found between the inter condition and the two ortho conditions ($p < .05$).

In summary, our results show that isolated featural information can be used, not only for same-view processing, but also for different-view recognition. Human observers are therefore capable of recognizing faces across different viewpoints by using isolated featural information. However, the comparison of the results of Experiment 1 (whole-face recognition) and Experiment 2 (feature recognition) shows that isolated featural information leads to a considerably lower recognition performance. Thus, face recognition is based on other information sources and processes as well, such as configural or holistic information.

In the current experiment we also assessed whether mechanisms such as recognition by view interpolation, the symmetry effect and the range of canonical views along the horizontal meridian, which were found with whole-face recognition, also account for isolated featural information. We found a strong viewpoint effect for scrambled faces; such that one's capacity to generalize from initially learned views to subsequently tested views decreases with increasing angle of rotation (see Figure 4). Thus not only whole-face processing, but also featural processing can be explained by a linear interpolation model. No conclusive answer was found for the question of a symmetry effect in scrambled-face recognition. There was no significant difference between the inter and the extra conditions. While Figure 4 shows a clear trend towards such a difference, further investigation is necessary to provide a definite answer. However, Experiment 2 clearly showed that the effect of a horizontal benefit seems to disappear with scrambled faces.

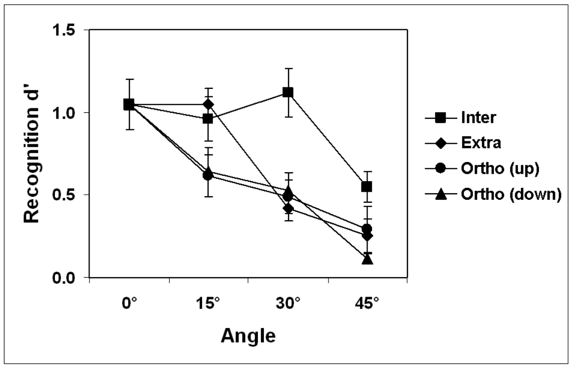

Figure 4: *Human recognition performance for scrambled-faces in the four rotation conditions (inter, extra, ortho (up), ortho (down)) across viewpoint (0° is the frontal view).*

2.3 Experiment 3: blurred-face condition

As we have discussed above, the experiments conducted by Schwaninger et al. (2002) show that, in frontal same-view recognition, both featural and configural information play an important role. In Experiment 2, we have seen that isolated featural information can be used not only for same-view recognition, but also for different-view recognition. Note, however, that recognition performance was substantially lower than in the whole face condition (Experiment 1). This is a clear indication that everyday face recognition is based on more than featural processing alone. In Experiment 3, we wanted to examine the role of configural processing. That is, we were interested in whether configural information can also be used for both same-view processing and different-view processing.

To isolate configural processing, we used a method first reported by Schwaninger et al. (2002). They had participants study intact faces, and then gave them a recognition test for one of three versions of the original faces:

- scrambled, in which all components were cut out and then rearranged
- blurred, in which the intact faces were blurred with a low-pass filter
- scrambled-blurred, in which the same low-pass filter was applied to the scrambled components

The blur-level had been determined in previous testing by determining the blurring strength at which performance in the scrambled-blurred condition was at chance. The idea behind this manipulation was that in the scrambled condition, all configural (and holistic) information was disrupted so that faces could only be recognized on the basis of individual components. When these rearranged components were blurred, participants were simply guessing, meaning that all discriminative information from the components was removed. When applied to the intact faces, this level of blur therefore removes information about fine detail required for component processing but retains the first- and, especially, second-order relational information (Diamond & Carey, 1986) thought to be the basis for configural face processing (Collishaw & Hole, 2000; Sergent, 1984). Schwaninger et al. (2002) showed that recognition of both scrambled (component-based) and blurred (configural-based) faces was above chance, adding to previous evidence that routes to recognition exist for both types of information. In addition, both yielded better performance for familiar than unfamiliar faces.

2.3.1 Participants, Materials and Procedure

Twenty right-handed undergraduate students from the University of Zurich volunteered to participate. All reported normal or corrected-to-normal vision. The participants were randomly assigned to one of two groups. Each group contained 10 participants and was tested in either the blurred testing condition or the scrambled and blurred testing condition. The procedure was the same as in Experiments 1 and 2. The same 20 male faces served as stimuli. In this experiment, the faces were shown blurred or scrambled-blurred in the testing phase. Blurred stimuli of both whole and scrambled faces were created by first rescaling the images to 246x246 pixels and then applying a standard Gaussian filter repeatedly to the resulting image. The parameters of the filter were: a radius of 3 pixels and a standard deviation of 3 pixels. In total 6 blur levels were used, which are specified by the number of times the Gaussian filter was applied (1, 2, 3, 4, 5, 6). For the experiments reported in this study blur level 4 was used. All blurred faces were then converted to gray-scale using Photoshop 6 (Figure 3d). Whole intact faces were also rescaled to 246x246 pixels.

2.3.2 Results and discussion

For a preliminary evaluation we first analyzed the data from the control condition (scrambled-blurred faces) to ensure that the low-pass filter eliminated the featural information effectively. We conducted one-sample t-tests (one-tailed) for each of the 13 angles in order to test the group means (M) against chance performance ($d' = 0$). In all 13 test angles, recognition performance did not differ significantly from chance ($p > .05$). This is a clear indication that the low pass filter, combined with the elimination of the color information, led to a destruction of the featural information.

Subsequently, we tried to investigate the effect of configural information on face-recognition performance. For this analysis, only the data from the blurred condition was included. Mean values are shown in Figure 5. Again, we conducted a one-sample t-test (one-tailed) to see whether recognition performance differed significantly from chance performance. In this case, only the 45°-down test angle did not differ significantly from chance performance. All other 12 test angles differed significantly from 0 ($p < .05$). The results show that human observers are capable of recognizing faces on the sole basis of configural information for almost all viewpoints.

In a second step, we compared the data with the data from the first baseline experiment. To do this, we calculated a three-factor ANOVA with the within-subjects factors of condition (extra, inter, ortho (up), ortho (down)) and amount of rotation (0°, 15°, 30°, 45°) and the between-subjects factor of stimulus type (intact face vs. blurred face). We found a main effect of stimulus type, $F(1, 18) = 26.30, p < .001$ and main effects of the within-factors condition $F(3, 54) = 44.16, p < .001$ and amount of rotation $F(1.66, 29.86) = 25.20, p < .001$. There was also an interaction effect between stimulus type and condition, $F(3, 54) = 6.06, p < .01$, but no interaction effect between stimulus type and amount of rotation, $F(1.66, 29.86) = 1.08, p > .34$. We also found an interaction effect between amount of rotation and condition $F(4.60, 82.71) = 12.52, p < .001$ and a three-way interaction between amount of rotation, condition and stimulus type $F(4.60, 82.71) = 2.78, p < .05$. A paired-sample t-test between the two conditions (whole face vs. blurred face) showed a significantly better overall recognition performance for the whole-face testing condition than for the blurred-face testing condition ($p < .001$).

In a next stage, we calculated paired-sample t-tests between whole-face testing condition, scrambled testing condition, and blurred testing condition. The results show a significantly better overall recognition performance for the whole-face testing condition than for the blurred face testing condition ($p < .001$) and a significantly better overall recognition performance for the blurred-face testing condition than for the scrambled-face testing condition ($p < .001$). We also calculated a two-factor ANOVA including the factors of condition (extra, inter, ortho (up), ortho (down)) and amount of rotation (0°, 15°, 30°, 45°), in order to focus on the qualitative characteristics of configural processing. Again, we took into account only the data from the blurred condition. In this case, main effects of condition (inter, extra, ortho (right) or ortho (left)) $F(1.57, 14.17) = 25.98, p < .001$ and amount of rotation (0°, 15°, 30°, 45°) $F(1.50, 13.48) = 6.89, p < .05$ were also found. There was also an interaction between the amount of rotation and condition $F(3.50, 31.46) = 7.96, p < .001$. A pairwise-comparison t-test between the four conditions showed that recognition performance in the inter condition was significantly better than in the other three conditions (extra $p < .05$; ortho $p < .001$). There was no significant difference between recognition performances in the two ortho conditions ($p > .11$). The extra condition showed a significantly better recognition performance than the ortho (down) condition ($p < .001$), but did not differ significantly form the ortho (up) condition ($p > .09$).

In summary, it is clear that human observers are capable of recognizing faces across different viewpoints on the sole basis of isolated configural information. This supports the idea that configural information is important for face recognition and that it can also be processed independently of other information. A comparison between the first two experiments shows that isolated configural information leads to a considerably lower recognition performance than whole-face processing but also to considerably higher recognition performance than featural processing. This finding is consistent with that of Schwaninger et al. (2002), who also reported a better recognition performance for configural processing than for featural processing. The current results confirm that the human perceptual system is capable of recognizing faces on the basis of isolated information such as configural or featural information, but they also demonstrate that integration from different information sources leads to the best recognition performance. Again, we found a systematic effect of viewpoint, in that recognition performance increased with a decreasing angle between a learned view and a tested view. As Figure 5 illustrates, there is enhanced recognition performance for the opposite 45° view similar to the results found in Experiment 1 using whole intact faces. Since this effect disappeared for scrambled faces in Experiment 2, we can conclude that view generalization for symmetrical views relies on configural information.

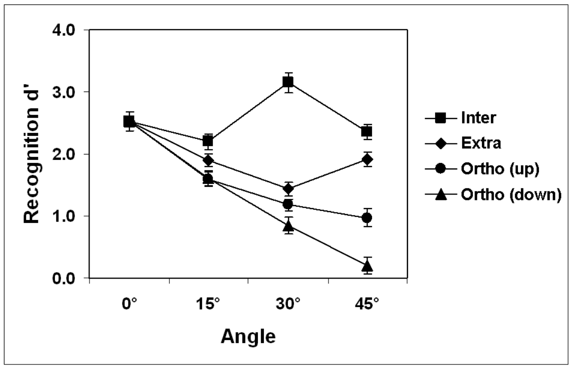

Figure 5: *Human recognition performance for blurred-faces in the four rotation conditions inter, extra, ortho (up), ortho (down) across viewpoint (0° is the frontal view).*

3 General Discussion

In this paper, we have investigated the fundamental question of how humans process local featural and global configural information in faces as a function of viewpoint. It is only through the use of state-of-the-art computer graphics techniques (most notably the morphable model by Blanz & Vetter, 1999) that we have been able to generate well-controlled, highly realistic stimuli.

At the beginning of this paper, we raised the question of what type of information human observers use to recognize faces across variations in viewpoint. We were specifically interested in whether they use only holistic information, or whether they also encode and store the local information contained in facial parts (featural or component information) as well as their spatial relationships (configural information). The results of this study clearly show that the visual system is capable of recognizing faces across different views based on isolated featural information as well as isolated configural information. This is a clear indication that the two information types can be encoded, stored and retrieved independently of one another. However, we can give no clear answer to the question of what role the holistic information plays for face recognition across different views, whether it can be encoded and stored independent of isolated configural and featural information and how important interaction processes between the different information types are. Earlier findings, however, suggest that face recognition is based on more than isolated configural and featural processes. It has been demonstrated, for example, that subjects are more accurate in recognizing the identity of a feature when it is presented in the context of the whole face rather than as an isolated feature (Tanaka & Farah, 1993; Tanaka & Sengco, 1997). What we can say, however, is that there are systematic effects of viewpoint in that the recognition performance increased with decreasing angle between learned view and tested view. This finding is consistent with the hypothesis that face recognition is mainly based on recognition by 2D-view interpolation and provides further evidence in favour of a view-based account of human processing (Bülthoff, Edelman, 1992; Tarr, Bülthoff, 1998).

We have seen that faces can be recognized based on different processes, but recognition performance across different viewpoints is dependent on the information that is available. The overall recognition performance based on complete information (whole-face condition), for example, is better than the overall recognition performance based on configural information (blurred condition). The lowest overall recognition performance was found for featural information (scrambled condition). Moreover, the symmetry effect and the range of canonical views along the horizontal meridian can be found in whole-face processing and, to a lesser degree, also in configural processing. We have also seen that they are not observable in connection with featural processing. A possible explanation for this finding is that different types of information are used for different tasks in face recognition. Evidence for such an assumption can be found in earlier studies. Schyns, Bonnar and Gosselin (2002) demonstrated that, depending on the face-recognition task (identity, gender, expression or not), different information was used as diagnostic. Thus, in a task in which the participants had to determine the identity of a face, high pass filtered information proved to be quite important. In a task in which the participants were required determine the gender of a face or to decide if a face smiled or not, high pass filtered information was of minor importance. By contrast, low pass filtered information was of great importance in all three tasks. A possible reason for these findings could be that isolated featural information plays an important role especially in a situation in which configural or holistic information is not sufficiently distinctive. This is the case when, for example, two faces look almost identical or we have to differentiate between two nearly identical expressions of a face, one of them with a real smile and the other with a feigned smile. This would also explain the relatively low recognition performance in the scrambled condition. The findings of Schyns et al. (2002) consistent with our results suggest that configural information is used as basic information for an initial classification step such as, for example, to decide if something is a face or not, which orientation the face has, if it is male or female, as well as for determining its identity. This might offer an explanation for the relatively high recognition performance for blurred faces: on most occasions, configural information is sufficient to identify a particular face and only when several faces look very similar, further information, such as high-frequency featural information is needed to disambiguate its identity.

Another interesting finding from our experiments is that the symmetry effect and the range of canonical views along the horizontal meridian were only found in connection with configural information, but not with featural information. This is consistent with the assumption that configural information is used as basic information for initial classification, and that featural information is used for further differentiation between individuals. In everyday life, faces are seen mostly from frontal or side views. Up or down views are less common, and hence less important. For the visual system, it is important to make a quick and reliable estimate in order to assess whether the object is relevant or whether further information is needed - this could be done by the configural processing route as we found good generalization performance along the horizontal axis. For featural processes, such a benefit for the horizontal axis is less essential as this type of processing might only be required when further, more detailed information is needed.

Acknowledgements

We would like to thank Jenny Campos and three anonymous reviewers for valuable comments on an earlier draft of this manuscript.

References

BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3d faces. In *Siggraph99*, 187–194.

BRUCE, V., BURTON, A. M., AND CRAW, I. 1999. Modeling face recognition. In *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, vol. 335(1273), 121–128.

BÜLTHOFF, H. H., AND EDELMANN, S. 1992. A morphable model for the synthesis of 3d faces. In *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89(1), 60–64.

CABEZA, R., AND KATO, T. 2000. Features are also important: contributions of featural and configural processing to face recognition. In *Psychol Sci*, vol. 11(5), 429–433.

COLLISHAW, S. M., AND HOLE, G. J. 2000. Featural and configurational processes in the recognition of faces of different familiarity. In *Perception*, vol. 29(8), 893–909.

DIAMOND, R., AND CAREY, S. 1986. Why faces are and are not special: An effect of expertise. In *Journal of Experimental Psychology: General*, vol. 115, 107–117.

GREEN, D. M., AND SWETS, J. A. 1966. In *Signal detection theory and psychophysics*, New York: Wiley, 187–194.

HAYWARD, W. G., RHODES, G., AND SCHWANINGER, A. in press. An own-race advantage for components as well as configurations in face recognition. In *Cognition*.

HILL, H., SCHYNS, P. G., AND AKAMATSU, S. 1997. Information and viewpoint dependence in face recognition. In *Cognition*, vol. 62(2), 201–222.

LEDER, H., CANDRIAN, G., HUBER, O., AND BRUCE, V. 2001. Configural features in the context of upright and inverted faces. In *Perception*, vol. 30, 73–83.

LOWE, D. G. 1987. 3-dimensional object recognition from single two-dimensional images. In *Artificial Intelligence*, vol. 31(3), 355–395.

MOSES, Y., ULLMAN, S., AND EDELMAN, S. 1996. Generalization to novel images in upright and inverted faces. In *Perception*, vol. 25(4), 443–461.

O'TOOLE, A. J., EDELMAN, S., AND BÜLTHOFF. 1998. Stimulus-specific effects in face recognition over changes in viewpoint. In *Vision Research*, vol. 38(15-16), 2351–2363.

PALMER, S., ROSCH, E., AND CHASE, P. 1981. Canonical perspective and the perception of objects. In *Attention and Performances IX*, Lawrence Erlbaum: Hillsdale, NJ, J. L. . A. Baddeley, Ed.

RAKOVER, S. S., AND TEUCHER, B. 1997. Facial inversion effects: Parts and whole relationship. In *Perception & Psychophysics*, vol. 59(5), 752–761.

RAKOVER, S. S. 2002. Featural vs. configurational information in faces: A conceptual and empirical analysis. In *British Journal of Psychology*, vol. 93, 1–30.

RHODES, G., BRAKE, S., AND ARKINSON, A. P. 1993. Whats lost in inverted faces. In *Cognition*, vol. 47(1), 25–57.

SCHANINGER, A., WALLRAVEN, C., CUNNINGHAM, D. W., AND CHILLER-GLAUS, S. 2006. Processing of identity and emotion in faces: A psychophysical, physiological and computational perspective. In *Progress in Brain Research*, vol. 156, 321–343.

SCHWANINGER, A., AND MAST, F. 1999. Why is face recognition so orientation-sensitive? psychophysical evidence for an integrative model. In *Perception (Suppl.)*, vol. 28, 116.

SCHWANINGER, A., LOBMAIER, J., AND S M. COLLISHAW. 2002. Role of featural and configural information in familiar and unfamiliar face recognition. In *Lectures Notes in Computer Science*, vol. 2525, 643–650.

SCHWANINGER, A., CARBON, C. C., AND LEDER, H. 2003. Expert face processing: Specialization and constraints. In *Development of face processing*, Göttingen: Hogrefe, G. Schwarzer and H. Leder, Eds., 81–97.

SCHWANINGER, A., SCHUMACHER, S., WALLRAVEN, C., AND BÜLTHOFF, H. H. submitted. View-based recognition of faces: Generalization across viewpoint is orientation-sensitive. In *Vision Research*.

SCHYNS, P. G., BONNAR, L., AND GOSSELIN, F. 2002. Show me the features! understanding recognition from the use of visual information. In *Psychological Science*, vol. 13(5), 402–409.

SERGENT, J. 1984. An investigation into component and configurational processes underlying face recognition. In *British Journal of Psychology*, vol. 75, 221–242.

SINHA, P., BALAS, B. J., OSTROVSKY, Y., AND RUSSELL, R. 2006. Face recognition by humans: 19 results all computer vision researchers should know about. In *Proceedings of the IEEE*, vol. 94, 1948–1962.

TANAKA, J. W., AND FARAH, M. J. 1993. Parts and wholes in face recognition. In *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, vol. 46(2), 225–245.

TANAKA, J. W., AND SENGCO, J. A. 1997. Features and their configuration in face recognition. In *Memory & Cognition*, vol. 25(5), 583–592.

TARR, M., AND BÜLTHOFF, H. H. 1998. Image-based object recognition in man, monkey and machine. In *Cognition*, vol. 67, 1–20.

TROJE, N. F., AND BÜLTHOFF, H. H. 1996. Face recognition under varying poses: The role of texture and shape. In *Vision Research*, vol. 36(12), 1761–1771.

TROJE, N. F., AND BÜLTHOFF, H. H. 1998. How is bilateral symmetry of human faces used for recognition of novel views? In *Vision Research*, vol. 38(1), 79–89.

ULLMAN, S., AND BASRI, R. 1991. Recognition by linear-combinations of models. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13(10), 992–1006.

VETTER, T., POGGIO, T., AND BÜLTHOFF, H. H. 1994. The importance of symmetry and virtual views in 3-dimensional object recognition. In *Current Biology*, vol. 4(1), 18–23.

WALLRAVEN, C., SCHWANINGER, A., SCHUMACHER, S., AND BÜLTHOFF, H. H. 2002. View-based recognition of faces in man and machine: Re-visiting inter-extra-ortho. In *Lectures Notes in Computer Science*, vol. 2525, 651–660.

YOUNG, A. W., HELLAWELL, D., AND HAY, D. C. 1987. Configurational information in face perception. In *Perception*, vol. 16(6), 747–759.