Available online at www.sciencedirect.com
ScienceDirect
Vision Research 47 (2007) 157–165
Vision Research
www.elsevier.com/locate/visres

The utility of surface reflectance for the recognition of upright and inverted faces

Richard Russell [a,b,*], Irving Biederman [c,d], Marissa Nederhouser [c], Pawan Sinha [a]

[a] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[b] Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA
[c] Department of Psychology, University of Southern California, HNB 316, Hedco Neurosciences Building, Los Angeles, CA 90089, USA
[d] Neuroscience Program, University of Southern California, HNB 316, Hedco Neurosciences Building, Los Angeles, CA 90089, USA

Received 26 January 2006; received in revised form 8 August 2006

Abstract

The variation among faces can be partitioned into two sources: (a) shape and (b) surface reflectance. To compare the utility of shape and reflectance for face recognition, we created two sets of faces, with individual exemplars differing only by shape in one set and only by reflectance in the other set. Grayscale and full color versions of the stimuli were used in separate experiments; the physical variation between exemplars was equated across the two sets with the grayscale but not the full color stimuli. Subjects performed a matching task in which both the target and distractor were drawn from the same set, so that only shape or only reflectance information could be used to perform the task. With the grayscale stimuli, performance was better in the shape condition, but with the color stimuli, performance was better in the reflectance condition. Inversion of the faces disrupted performance with the shape and reflectance sets about equally, suggesting that the inversion effect is not caused specifically by the spacing of facial features, or even by shape information more generally. These results provide evidence that facial identity is a function of reflectance as well as shape, and place important constraints on explanations of why inversion impairs face recognition.

Keywords: Face recognition; Shape; Pigmentation; Color; Inversion; Surface; Representation; Reflectance

1. Introduction

The visual information available to distinguish faces can be divided into two broad classes—shape and surface reflectance.[1] Though shape is believed to be the dominant cue for the recognition of objects (Biederman & Ju, 1988; Palmer, 1999; Tanaka, Weiskopf, & Williams, 2001; Ullman, 1996), there is evidence pointing to the importance of reflectance (sometimes called 'pigmentation') for face recognition. Unlike basic-level objects, faces are notoriously difficult to recognize from line drawings (Bruce, Hanna, Dench, Healey, & Burton, 1992; Davies, Ellis, & Sheperd, 1978; Leder, 1999; Rhodes, Brennan, & Carey, 1987). Because line drawings do not contain reflectance information, this suggests that reflectance plays a role in face recognition. Similarly, recognition is not as good with representations of faces that have uniform (and hence non-diagnostic) reflectance (Bruce et al., 1991). Also, there is evidence that faces with negated contrast are difficult to recognize largely because the perception of reflectance is disrupted (Bruce & Langton, 1994; Liu, Collin, Burton, & Chaurdhuri, 1999; Russell et al., 2006; Vuong, Peissig, Harrison, & Tarr, 2005).

* Corresponding author.
E-mail address: rrussell@fas.harvard.edu (R. Russell).

[1] Though reflectance is sometimes referred to as 'albedo', 'color', or 'texture', we believe these terms to be problematic because they can also be used to refer to specific subsets of the broader light transfer function of a surface—the fraction of light emitted by the surface in the case of 'albedo', greater reflectance of specific wavelengths in the case of 'color', and spatially variegated reflectance in the case of 'texture'. We use the term 'reflectance' to refer to the complete light transfer function of the surface, including subsurface scattering, which is an important attribute of human skin (Debevec et al., 2000). Elsewhere we have used the term 'pigmentation' to refer to the exact same concept (Russell, Sinha, Biederman, & Nederhouser, 2006).

0042-6989/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2006.11.002

But is reflectance actually as important as shape for recognizing faces? O'Toole and colleagues investigated this question by determining in an old/new task whether subjects could recognize laser-scanned faces for which only shape or only reflectance was diagnostic (O'Toole, Vetter, & Blanz, 1999). They used a laser-scanning process that records in separate files the three-dimensional depth (shape) and surface reflectance properties of a face. These files can be manipulated independently and subsequently re-combined. One group of subjects saw a set of faces that differed from each other in terms of shape but not reflectance, in that they all had the same reflectance as one another (the average of all the faces scanned of the appropriate gender). Another group of subjects viewed a set of faces that all had the same shape (again the average of all the faces scanned) but differed from one another in terms of their reflectance. Both groups of subjects performed the old/new task about as well as each other, but not as well as a third group of subjects who viewed faces that differed in terms of both shape and reflectance. Overall, the findings pointed to approximately equal utility of three-dimensional shape and reflectance information for face recognition.

The evidence points to reflectance and shape both being important for face recognition. Yet there remains a group of findings, related to the inversion effect, that could be interpreted as evidence that reflectance is relatively unimportant for face recognition. Because inversion (rotation in plane) of 180° makes faces much more difficult to recognize (Valentine, 1988; Yin, 1969), one might infer that the information that is disrupted by inversion is of outsized importance for recognition, and that information not disrupted by inversion is of lesser importance. For this reason, the nature of the information that is disrupted by inversion is germane to the question of the utility of reflectance for face recognition. This decline in face recognition performance caused by inversion has been attributed to impairment in the perception of "second-order relations"—the distances or spacing between features and contours (Diamond & Carey, 1986), which is a component of face shape. A group of studies investigated sensitivity to manipulations of the spacing of features or to manipulations of the details of the features themselves (e.g. the reflectance or shape of the eyes and mouth), and found evidence suggesting that inversion disrupts sensitivity to spacing more than sensitivity to the feature details (Barton, Keenan, & Bass, 2001; Freire, Lee, & Symons, 2000; Leder & Bruce, 2000; LeGrand, Mondloch, Maurer, & Brent, 2001). However, two recent studies have found that when presentation of the conditions is randomized rather than blocked, and performance in the upright conditions is equated, perception of spacing is not disrupted more by inversion than perception of features (Riesenhuber, Jarudi, Gilad, & Sinha, 2004; Yovel & Kanwisher, 2004). The broader question of whether shape or reflectance is disrupted by inversion has not been investigated. Does inversion only (or primarily) disrupt the employment of shape information for face discrimination? Or does inversion also disrupt the employment of reflectance information?

The overarching goal for our study was to investigate the utility of reflectance for face recognition. We approached this larger goal through two routes. First, we sought to compare the utility of shape and reflectance for face recognition, extending the findings of O'Toole et al. (1999) by replicating them with a different task and stimuli, and by controlling for visual similarity across conditions. Second, we sought to determine whether face inversion disrupts the perception of reflectance as well as shape. We turn now to describe at length the direct comparison of shape and reflectance, and return in the final paragraph of the introduction to the investigation of inversion.

To investigate the relative roles of shape and reflectance for face recognition, we created sets of faces in which the exemplars differed from one another only in terms of their shape or only their reflectance. We then assessed performance in a matching task in which both target and distractor faces were drawn from the same set, so that the faces in a given trial could be distinguished only by shape or by reflectance. This approach was similar to that of the O'Toole et al. study, but with some differences that we outline here. Both experiments used 'unfamiliar recognition' tasks (subjects did not know the people whose faces served as stimuli) but our experiment employed a two-alternative forced-choice delayed match-to-sample task rather than an old/new task. We used a within-subjects design, so that each subject saw all the stimuli, and did not know which type of information would be useful on a given trial. Perhaps most importantly, our stimuli were based on photographs rather than laser scans, and this involved a different separation of shape and reflectance.

Rather than creating stimuli that were matched for their three-dimensional shape, we created stimuli that were matched for their two-dimensional shape in the image plane. Shape was determined by the locations in the image of the outlines of the face, mouth, nose, eyes, irises, and eyebrows—those outlines that are part of the shared common configuration of all faces. In this two-dimensional definition of face shape, borders defined by sharp luminance gradients are considered attributes of face shape when they are common to all faces (e.g. the outline of the iris or the outline of the eyebrow), but not when they are unique to individual faces (e.g. a mole or freckles). This is different than the method used by O'Toole et al. (1999), for which the borders of the eyebrows or irises would be considered attributes of reflectance, though the resulting stimuli are largely similar in their appearance. Our method does have limitations—in particular the images that vary in terms of reflectance also vary slightly in terms of their shape. This is caused by a combination of imperfect image warping and the utility of shading and specularity as cues to shape. Nonetheless, the reflectance variation of these images is much greater than the shape variation, and so they can be considered to vary almost entirely in terms of reflectance. Overall, we believe that our method does a good job of capturing the subjective sense of shape and reflectance/pigmentation in a face image.

Differences in performance with the ‘Shape’ or ‘Reflectance’ sets could be due either to differences in the amount of information that each provides (i.e. how similar the faces are in terms of shape or in terms of reflectance) or to differences in observers’ ability to utilize the different kinds of information. To address this issue, we sought to equate the similarity of the images in the Shape and Reflectance sets, so that a link could be established between the amount of information provided by a given source, i.e., reflectance or shape, and the utility of that source for recognition. Toward this end, we computationally measured the similarity of the face images in the Shape and Reflectance sets using the Gabor-jet model developed by von der Malsburg and colleagues (Lades et al., 1993). This model computes the activation produced by columns (“jets”) of multi-scaled, multi-oriented Gabor filters, with each jet centered on a given region of the visual field, roughly corresponding to a V1 simple cell hypercolumn. An evenly spaced rectangular grid of these jets covers the image and, unlike a more recent version of the Gabor-jet model (Wiskott, Fellous, Kruger, & von der Malsburg, 1997), remains fixed in the same position regardless of the image contents. The system is sensitive to contrast, but not to whether the contrast is caused by shape or reflectance. This metric correlates well with human performance in matching faces (Biederman & Kalocsai, 1997; Hancock, Bruce, & Burton, 1998; Kalocsai, Biederman, & Cooper, 1994). We refer to the similarity values derived from this metric as “Gabor-jet similarity”. We measured the distribution of Gabor-jet similarities for all the pairs of images within each condition to the distributions of the other two conditions, to determine whether the conditions differed in terms of the physical similarity of their exemplar images. By manipulating the background color of the images, we were able to equate the similarity of the Shape and Reflectance sets.

Previous work investigating the utility of reflectance information has only used grayscale images. Yet removing color from an image to make it grayscale selectively removes reflectance cues without removing shape cues, changing the relative utility of the two classes of cues. As a coarse evaluation of this contribution of color, in a separate experiment we also compared the utility of shape and reflectance with colored images. The Gabor-jet system that we used to measure image similarity has been designed and validated for use with grayscale images, and cannot produce meaningful similarity measures of color images. This is why the evaluation with color is termed “coarse,” insofar as only in the experiment with grayscale images were the shape and reflectance sets of stimuli matched for image similarity. Because the color versions of the images could not be matched for image similarity, the results of the two experiments are not directly comparable in terms of the information content of the images. However, aside from the presence of color, the images are exactly the same. Because the grayscale Shape and Reflectance sets were equated for similarity, we can reasonably assume that the images of the color Reflectance set are somewhat less similar than the images in the color Shape set.

To evaluate whether face inversion disrupts the perception of reflectance as well as shape, we compared performance on the task when faces were inverted as well as upright. If it is specifically the use of second-order relations that is disrupted by inversion, or even the use of shape in general, we should find that inversion is more disruptive of performance with the Shape set than with the Reflectance set.

2. Methods

We ran two experiments using the same design. In Experiment 1 stimuli were presented in grayscale, and in Experiment 2 stimuli were presented in full color.

2.1. Subjects

Thirty-three subjects participated in the experiments—17 in Experiment 1, and 16 in Experiment 2. After data from two subjects in Experiment 1 and one subject in Experiment 2 were excluded for failure to exceed chance performance in all conditions, the data from 15 subjects were analyzed for each experiment. Subjects in Experiment 1 had a mean age of 23 years (SD 5 years), and those in Experiment 2 had a mean age of 21 years (SD 4 years). There was not a significant difference between the ages of the subjects in the two experiments, as assessed by an independent samples t-test, $t_{28} = 1.3$, $p = .2$. In both experiments, there were seven male and eight female subjects. All subjects were contacted through the MIT Brain and Cognitive Sciences subject pool, were naïve to the purpose of the study, and had normal or corrected-to-normal vision.

2.2. Stimuli

Frontal facial photographs were taken of eight male and eight female faces that were clean-shaven and wearing no cosmetics. The people were all Caucasian, and ranged in age from 18 to 25, with a mean age of 20. All of the faces were photographed under the same illumination in the same room, so that variation between images of the faces was caused by shape or reflectance but not illumination. The heights of the camera and lamps were fixed, and the chair used by the photographic subjects was adjusted such that their heads were all at the same height. To minimize shading cues the faces were illuminated by two studio lamps with large diffusing heads, in a small room with white walls (for greater ambient illumination). These two lamps, the camera and tripod, and the chair on which the photographic subjects sat, were kept in locations that were fixed with respect to each other and the room. The lights were centered at 0° elevation (level with the head), to eliminate cast shadows and minimize the effects of shading cues (Liu et al., 1999). Frontal illumination also allowed the direction of lighting relative to the face and relative to the viewer to remain constant whether the image was upright or inverted.

The images were cropped to remove hair, ears, and neck. For each sex, the eight original faces were morphed together using Morph Man 3.0 (Stoik Imaging) to produce an average face. We created the stimuli for the Shape condition by warping this average face into the shape of each of the original faces, producing new faces with the same (average) reflectance, but distinct shape. Similarly, we created the stimuli for the Reflectance condition by warping each of the original faces into the shape of the average face, producing new faces with the same (average) shape, but distinct reflectance. The original faces, which differed from one another in terms of both their shape and their reflectance, comprised the stimuli for the Shape + Reflectance condition. By ‘morphing’ we refer both to moving pixels in the image plane and to the averaging of the pixel intensities of different images, while by ‘warping’ we refer only to moving pixels in the image plane. In order to achieve very high fidelity between images (to ensure, for example, that the shape of the images in the reflectance set was as nearly identical as possible), we used approximately 250 reference points per face to perform the morphing and warping. Image warping and morphing are described in greater detail Appendix A. These stimuli were previously used in Russell et al. (2006). In Experiment 1, these stimuli were presented in grayscale (Fig. 1), and in Experiment 2 the same stimuli were presented in full color.

We used the Gabor-jet system to compare the similarity of all the pairs of faces in one condition to the similarity of the pairs of faces in each of the other conditions (with eight male and eight female faces in each condition, and each of the faces paired with every other face of its same sex, there were 56 distinct pairings of faces per condition—28 per sex). The relative similarity of the two sets of images was manipulated by manipulating the background color. It may seem odd to equate the similarity of a set of face images by manipulating the background rather than the faces themselves. However, the entire image is relevant to performance on any unfamiliar face recognition task. Yet the choice of background in such experiments is usually arbitrary and not given careful attention. For the present purposes, the choice of background was particularly relevant because it affects the relative similarity of the Shape and Reflectance sets. As the background color becomes closer to the average skin tone of the faces, the images of the Shape set become more similar to each other while those of the Reflectance become set less similar. This is because the outline of the faces becomes less distinct, making the Shape set less similar, while the differences in overall skin tone of the Reflectance set become more distinct, making the Reflectance set less similar. In this way, increasing or decreasing the contrast between the background and the average skin tone has the opposite effect on the relative similarity of the Shape and Reflectance sets. The stimuli for all three conditions were given the same background (as in Fig. 1), which was chosen because it resulted in equivalent distributions of similarity values for the pairs of images in the Shape and Reflectance sets.

Gabor-jet similarity ranges from 0 to 1, with higher numbers indicating greater similarity. The Gabor-jet similarities of face pairs in the Shape set had a mean of 0.930 and a standard deviation of 0.019, while those in the Reflectance set had a mean of 0.932 and a standard deviation of 0.012. A paired-samples t-test found the Shape and Reflectance sets to be matched for Gabor-jet similarity, $t_{55} < 1$. The Gabor-jet similarities of face pairs in the Shape + Reflectance set had a mean of 0.888 and a standard deviation of 0.020. Paired-samples t-tests found the Gabor-jet similarities of the Shape + Reflectance set to be significantly lower than those of either the Shape set $t_{55} = 24.0$, $p < .001$ or the Reflectance set $t_{55} = 20.8$, $p < .001$. The lower similarity among the images in the Shape + Reflectance set was expected, as they differed along more dimensions than did the images in either the Shape or Reflectance sets. The Gabor-jet system is designed to operate on grayscale (single channel) stimuli, and so the color versions of the stimuli could not be equated for similarity.

2.3. Procedure

Subjects performed a delayed match-to-sample, two-alternative forced-choice task in a darkened room. In each trial of this task, the observer saw a sample face for 141 ms, then a visual noise mask for 200 ms, and then a blank screen for 1000 ms. Next, two faces drawn

Fig. 1. Examples of stimuli. The top row contains faces from the Shape condition. The middle row contains faces from the Reflectance condition. The bottom row contains four faces from the Shape + Reflectance condition, which are photographs of actual people. Similarity among the faces in the 'Shape' and 'Reflectance' conditions was equated according to the Gabor-jet system (Lades et al., 1993).

from the same set were presented side by side in the center of the screen for 306 ms. One of the two faces (the sample) had been presented just previously, and the other was a distractor face that was drawn from the same set and same sex as the target face. Thus, when the target and distractor faces were drawn from the Shape set, only shape was available to perform the task, when the faces were drawn from the Reflectance set, only reflectance was available, and when the faces were drawn from the Shape + Reflectance set, both kinds of information were available. The task was to decide which of the two faces matched the sample and to press the corresponding key as quickly as possible. Subjects were not informed before the experiment as to how the faces would differ. Trials from the different conditions were randomized so that subjects could not adopt a strategy of attending exclusively to shape or reflectance, and the left–right ordering of target and distractor was counterbalanced. All the faces in a given trial were presented either upright or inverted. Each of the eight faces of either sex was paired with every other face of its same sex for each condition, and in each orientation, resulting in 28(pairs) × 2(sex of face) × 3(condition) × 2(orientation) × 2(left–right counterbalance) = 672 trials per observer. Stimuli were presented in Matlab (The MathWorks, Inc.) using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

3. Results

Percent correct responses for both experiments are shown in Fig. 2, with the results of Experiment 1 (with grayscale images) on the left, and those of Experiment 2 (with colored images) on the right. There was no significant effect of the sex of the subject in either experiment, and so this factor was excluded from further analysis. For each experiment, we ran an ANOVA with condition (Shape, Reflectance, Shape + Reflectance), orientation (upright, inverted), and the sex of the face (male, female) as fixed factors, and percent correct responses as the dependent variable. Response times (RT) appear in Fig. 3, with the results of Experiment 1 (with grayscale images) on the left, and those of Experiment 2 (with colored images) on the right. Each subject's RT data was cleaned by removing responses that were more than three standard deviations slower than the subject's mean response time. We ran the same ANOVA with the RT data as with the performance data.

3.1. Experiment 1—grayscale images matched for Gabor-jet similarity

Performance in all conditions, including Reflectance, was above chance (Fig. 2a). There was a significant main effect of condition, $F_{2,28} = 43.4$, $p < .001$, and subsequent Tukey's HSD post hoc comparisons indicated that performance with each of the conditions was significantly different at the 0.05 level. Performance was better in the Shape condition than in the Reflectance condition, and best in the Shape + Reflectance condition. There was also a significant main effect of orientation, $F_{1,14} = 43.2$, $p < .001$, with performance worse with inverted than upright faces. The interaction between condition and orientation was not significant, $F_{2,28} = 1.1$, $p > .1$, which suggests that orientation did not disrupt one type of information more than the other. There was no significant main effect of the sex of the face being matched, $F_{1,14} = 1.9$, $p > .1$, though the interaction between the sex of the face and orientation showed a trend toward significance, $F_{1,14} = 4.3$, $p = .057$, with greater effects of inversion for the female faces. There was a similar but weaker pattern of results with the response times (Fig. 3a), with significant main effects of condition, $F_{2,28} = 8.1$, $p < .01$, and orientation, $F_{1,14} = 27.1$, $p < .001$, but not of the sex of the face being matched, $F_{1,14} < 1$, and there was a trend toward significance of the interaction between the sex of the face and orientation, $F_{1,14} = 3.8$, $p = .071$, with greater effects of inversion for the female faces.

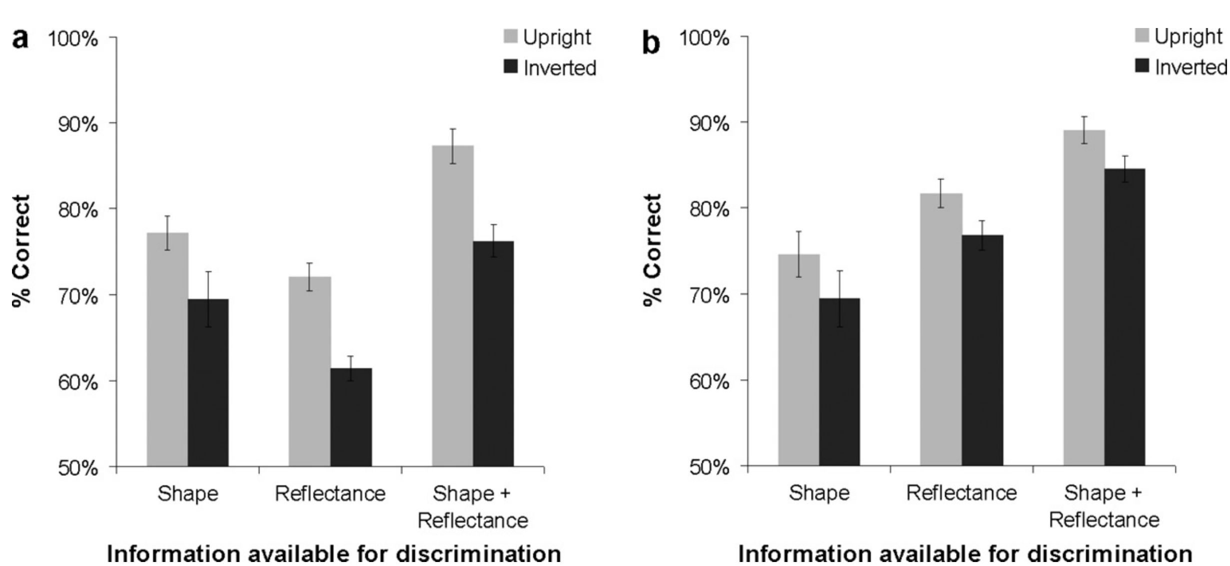

Fig. 2. Experimental results, expressed as percentage of correct responses. Error bars indicate standard errors of the mean. Light bars indicate performance with upright faces, and dark bars indicate performance with inverted faces. (a) Results of Experiment 1, with grayscale stimuli. (b) Results of Experiment 2, with full color stimuli.

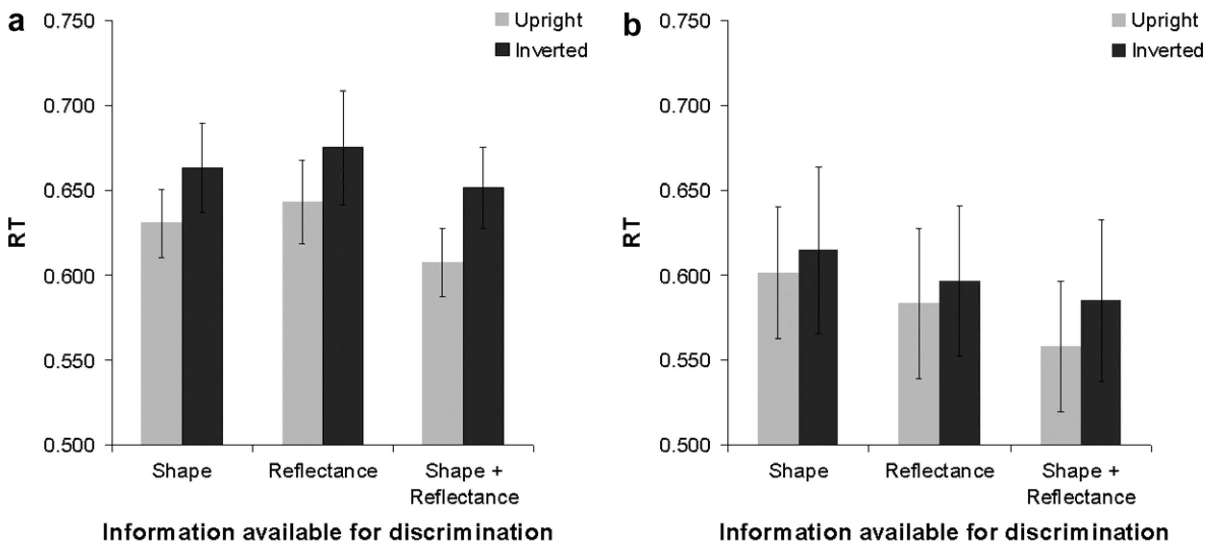

Fig. 3. Response times (RT). Error bars indicate standard errors of the mean. Light bars indicate performance with upright faces, and dark bars indicate performance with inverted faces. (a) Response times (RT) for Experiment 1, with grayscale stimuli. (b) Response times (RT) for Experiment 2, with full color stimuli.

3.2. Experiment 2—full color images not matched for Gabor-jet similarity

In Experiment 2 there was better performance with Reflectance cues than with Shape cues, a reversal of the ordering found in Experiment 1 (Fig. 2b). As in Experiment 1, there was a main effect of condition, $F_{2,28} = 34.7$, $p < .001$, and subsequent Tukey's HSD post hoc comparisons indicated that performance with each of the conditions was significantly different at the 0.05 level. There was a main effect of orientation, $F_{1,14} = 15.7$, $p < .001$, with performance worse with inverted than upright faces. Again, the interaction between condition and orientation was not significant, $F_{2,28} < 1$, indicating that orientation did not disrupt one type of information more than the other. There was no significant main effect of the sex of the face being matched, $F_{1,14} < 1$. The interaction between the sex of the face and orientation was significant, $F_{1,14} = 5.3$, $p < .05$, with greater effects of inversion for the female faces. Indeed, there appears to have been a very small effect of inversion on the male faces in Experiment 2. There was also a significant interaction between the sex of the face and condition, $F_{2,28} = 9.1$, $p < .05$, with better performance in the Reflectance condition with female faces than with male faces but better performance in the Shape and Shape + Reflectance conditions with male faces. Because there were significant effects of the sex of the face in the performance data, we present these results broken down by the sex of the face in Fig. 4. The response times (Fig. 3b) showed a similar pattern of results, with significant main effects of condition, $F_{2,28} = 16.1$, $p < .001$, and orientation, $F_{1,14} = 4.6$, $p = .05$, but not of sex, $F_{1,14} < 1$. Unlike the performance data, the response time data did not yield significant interactions.

3.3. Comparison between grayscale and full color

The difference in performance with grayscale and full color stimuli can be appreciated by comparing graphs (a) and (b) in Fig. 2. To analyze the effects of color, we ran an ANOVA with condition (Shape, Reflectance, Shape + Reflectance), orientation (upright, inverted), and sex (male, female) as within subjects factors, and color (grayscale, full color) as a between subjects factor, and percent correct responses as the dependent variable. To compare performance with grayscale and full color stimuli, our concern here was with the color factor and its interaction with other factors. There was a significant main effect of color, $F_{1,28} = 6.7$, $p < .05$, with better performance with full color than grayscale faces. There was a significant interaction between orientation and color, $F_{1,28} = 6.6$, $p < .05$, with a larger inversion effect for the grayscale than the color faces. There was also with a significant interaction between condition and color, $F_{2,56} = 16.5$, $p < .001$. Inspection of Fig. 2 suggests that this is caused primarily by better performance in the Reflectance condition with full color than with grayscale faces, and perhaps to a lesser extent by better performance in the Shape + Reflectance condition with full color. Finally, there was a three-way interaction between sex, condition, and color, $F_{2,56} = 4.2$, $p < .05$. Inspection of Fig. 4 does not immediately suggest any single reason for this three-way interaction. Running the same ANOVA with RT as the dependent variable found fewer differences of color. Though responses were faster with color faces, the main effect of color was not significant, $F_{1,28} = 1.4$, $p > .1$. While there was a larger effect of inversion with grayscale faces, the interaction between orientation and color was not quite significant, $F_{1,28} = 2.8$, $p = .1$. The only significant interaction was between condition and color, $F_{2,56} = 5.0$ $p < .05$, with faster response in the Reflectance condition with full color faces.

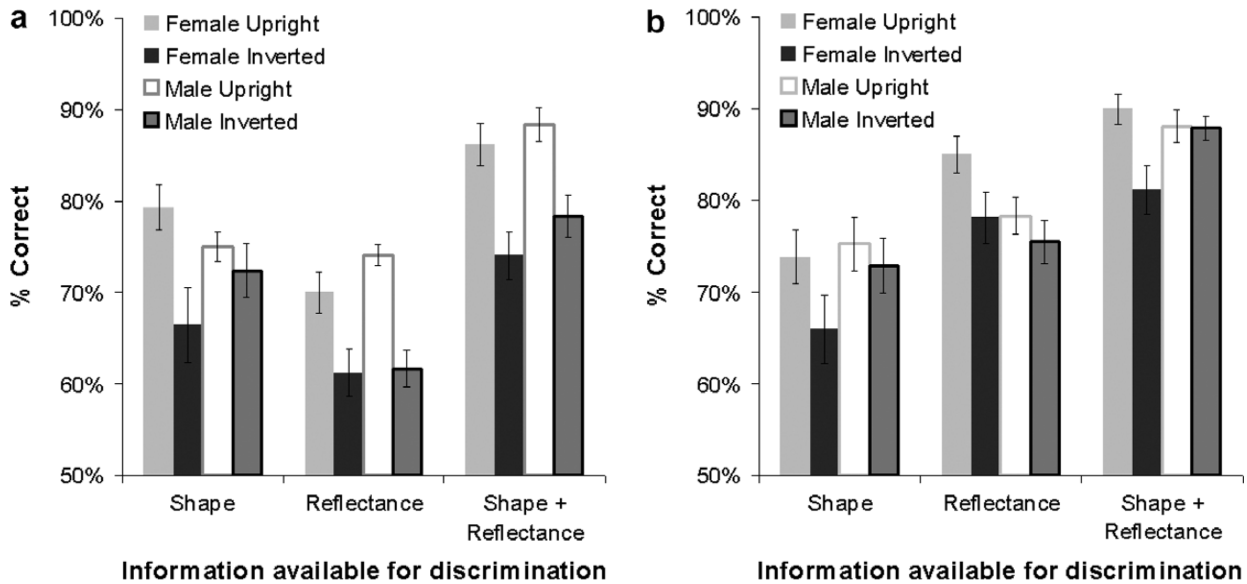

Fig. 4. Experimental results of Experiments 1 and 2, separated by sex of the face being matched, expressed as percentage of correct responses. Error bars indicate standard errors of the mean. Light bars indicate performance with upright faces, and dark bars indicate performance with inverted faces. Solid bars indicate female faces and outlined bars indicate male faces. (a) Results of Experiment 1, with grayscale stimuli. (b) Results of Experiment 2, with full color stimuli.

4. Discussion

Performance with either shape or reflectance information alone was significantly better than chance, but was also significantly worse than performance using both shape and reflectance information, suggesting that both were employed. The direct comparison between performance with shape alone and performance with reflectance alone yielded different results in the two experiments. With grayscale images for which the Gabor-jet similarity of the shape and reflectance sets were equated, subjects performed better at matching the faces using shape than reflectance information. This suggests the possibility that the visual system may use shape cues more efficiently than reflectance cues for face recognition. With full color versions of the same images (for which we can assume that the reflectance stimuli were less similar than were the shape stimuli) this relationship was reversed, and subjects performed better using reflectance than shape information.

Overall, the results are broadly consistent with the notion that shape and reflectance are both important for face recognition. These findings are in agreement with those of O'Toole et al. (1999), and show that they are tolerant to differences in the task and the procedures used to create the stimuli. The advantage of Shape over Reflectance in Experiment 1 where the two sets were equated for information content suggests that shape may be used more efficiently by the visual system for face recognition than reflectance. To test this more completely it would be helpful to vary the information content parametrically rather than simply equating it across conditions.

In both experiments, face inversion disrupted the use of shape and reflectance about equally. Because the faces in the Reflectance condition did not differ in terms of their second-order relations, our results are consistent with two recent studies (Riesenhuber et al., 2004; Yovel & Kanwisher, 2004) in providing evidence against the assertion that it is impairment in the perception of second-order relations (the spacing of features) that causes poorer recognition of inverted faces. The current study extends these findings by presenting evidence that the disruption caused by inversion is not even specific to shape in general, let alone to the spacing of features. This suggests that the processes disrupted by inversion utilize both shape and reflectance information. It is important to note that there are spatial relations in the reflectance map of a face (e.g. the dark area of the eyebrow is below the lighter area of the forehead). While inversion does not specifically disrupt shape, it is entirely possible that inversion does specifically disrupt *spatial relations*, whether those of shape or of reflectance.

The relative ordering of the utility of shape and reflectance information apparent with grayscale faces was reversed through the addition of color. This is presumably because color provides reflectance but not shape information. The effect of inversion was smaller with color than with grayscale faces. We are not aware of any other study that has investigated face inversion with color photographs, but the current result suggests that those aspects of reflectance enhanced by color may be invariant to 2D orientation. Another curious finding was the interaction between orientation and the sex of the face, which was significant or nearly significant with both color and grayscale stimuli. We are not aware of any similar findings, and have no speculation as to the cause.

Some caveats are in order. We only used Caucasian faces as stimuli, and it is possible that the utility of reflectance is not the same in all ethnic groups. Also, we do not wish to give the impression that the static, intrinsic cues that we have investigated here are the only useful information for face recognition. Dynamic cues such as distinctive emotional expressions or speech movements (Knappmeyer, Thornton, & Bulthoff, 2003; Lander & Chuang, 2005), and contextual cues such as illumination (Braje, Kersten, Tarr, & Troje, 1998; Enns & Shore, 1997; Hill & Bruce, 1996; Johnston, Hill, & Carman, 1992; McMullen, Shore, & Henderson, 2000), have been shown to play a role in face recognition. It is also likely that the relative utility of shape and reflectance differs under different kinds of illumination. The diffuse, frontal illumination that we used here reduced shading variation, which is a cue to shape, and may have enhanced pigmentation cues. Dim or less diffuse lighting would likely enhance the relative utility of shape cues to some degree.

The current study provides evidence in support of the notion that reflectance plays an important role in face recognition. For recognition of unfamiliar faces, shape and reflectance information were both found to be useful, though performance was better using shape than reflectance when information content was equated across the two conditions. However, when color cues were added, performance was better in the reflectance condition. Finally, the data provide an important constraint for explanations of the face inversion effect, in that inversion was found to disrupt the recognition of faces when they differed by reflectance as much as when they differed by shape.

Acknowledgments

We thank Nancy Kanwisher, Ken Nakayama, and Molly Potter for helpful comments on the manuscript, and Xiaomin Yue for assistance with the Gabor-jet system. This research was supported by a grant to P. Sinha from the DARPA Human ID Program administered by the Information Awareness Office, and by Grants NSF 0420794, 0426415, and 0531177 to I. Biederman.

Appendix A. 2D image warping and morphing

The software we used for morphing and warping to create our stimuli set is proprietary (MorphMan 3.0, Stoik Imaging), and so we cannot provide the precise algorithm used. Instead, we provide a general description of image morphing based on triangulation, the option we selected for rendering the images in MorphMan. Image morphing is conducted on pairs of images (the 'source' and 'target' images), and involves three steps: (1) locating corresponding feature points common to both images (2) warping one or both images into an intermediate shape or the shape of the other image (3) weighted blending of corresponding pixel values in the warped images.

The first step of locating features points on the images is typically done by hand. We used approximately 250 reference points per face to perform the morphing and warping. With the exception of the points over the eye centers, these points were placed as contour lines rather than individual points, the significance of which is described below. Lines were used to outline the jaw and hairline, the eyebrows, the eyes and irises, the nose, and the mouth.

Image warping is a geometric transformation for aligning two images via corresponding feature points. Because only a very small fraction of the total pixels in the images are labeled as feature points, a key problem in warping is determining how the pixels other than the feature points should move. A commonly used solution to this problem is called 'triangulation', because it involves partitioning the image into triangular regions by connecting neighboring feature points (Rowland & Perrett, 1995; Wolberg, 1990). There are many possible ways to tessellate the image with triangles. The points inside a triangle will move as a function of the feature points at the triangle vertices, and so it is desirable to tessellate the image such that points inside a triangle are closer to its three vertices than to the vertices of another triangle. This is commonly achieved through *Delaunay triangulation* of the set of vertices, which has the property that no triangle vertex falls in the interior of the circumcircle (the circle that passes through all three vertices) of any triangle in the triangulation. As mentioned above, we primarily used feature lines consisting of linked points, rather than individual feature points to establish correspondence between the source and target images. This almost certainly indicates that a *constrained* Delaunay triangulation was used, meaning that the triangulation was constrained such that each line segment between feature points was present as an edge in the triangulation (Bern & Eppstein, 1995). This ensures that the entire feature line of the source will align with that of the target, including pixels along the line in between the feature points used to establish the line.

During warping, the movement of the pixels within a particular triangular region is a function of the movement of the three bounding feature points. Each point P within a triangle ABC can be specified in terms of its distance from the triangle vertex A along the vectors $\boldsymbol{V}_1 = \overrightarrow{AB}$ and $\boldsymbol{V}_2 = \overrightarrow{AC}$. This can be stated as $P = \lambda_1 \boldsymbol{V}_1 + \lambda_2 \boldsymbol{V}_2$ where $\lambda_1 >= 0$, $\lambda_2 >= 0$ and $\lambda_1 + \lambda_2 <= 1$. Warping an image involves moving the feature points of the source image into the shape of the target or an intermediate shape. The values of λ_1 and λ_2 can be calculated for each pixel in the generated image using the enclosing triangle and the set of feature points in the target image. To calculate an image warp that is n% along the morph trajectory (where $n = 0$ is the source image and $n = 100$ is the target image) the first step is to calculate the location of the feature points n% from the source to the target image. Using the equation above, the entire image can be warped to this intermediate point. If a 'morph' rather than a 'warp' is desired, the pixels of the two images are then blended together with the pixel weighings determined by n. The value of each pixel in the morphed image M is given as $M = (n/100)T + (1\text{-}(n/100))S$ where T is the value of the corresponding pixel in the target image and S is the value of the corresponding pixel in the source image. This pixel value can be altered to avoid aliasing, for example by interpolation of neighboring pixels.

References

Barton, J. J. S., Keenan, J. P., & Bass, T. (2001). Discrimination of spatial relations and features in faces: effects of inversion and viewing duration. *British Journal of Psychology, 92*, 527–549.

Bern, M., & Eppstein, D. (1995). Mesh generation and optimal triangulation. In D.-Z. Du & F. Hwang (Eds.), *Computing in Euclidean Geometry* (pp. 47–123). Singapore: World Scientific Publishing Company.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology, 20*, 38–64.

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London B, 352*, 1203–1219.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Braje, W. L., Kersten, D., Tarr, M. J., & Troje, N. F. (1998). Illumination effects in face recognition. *Psychobiology, 26*(4), 371–380.

Bruce, V., Hanna, E., Dench, N., Healey, P., & Burton, M. (1992). The importance of 'mass' in line drawings of faces. *Applied Cognitive Psychology, 6*, 619–628.

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

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

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

Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., & Sagar, M. (2000). Acquiring the reflectance field of a human face. *Proceedings of SIGGRAPH 2000, 2000*, 145–156.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107–117.

Enns, J. T., & Shore, D. I. (1997). Separate influences of orientation and lighting in the inverted-face effect. *Perception & Psychophysics, 59*(1), 23–31.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information. *Perception, 29*, 159–170.

Hancock, P. J. B., Bruce, V., & Burton, M. A. (1998). A comparison of two computer-based face identification systems with human perceptions of faces. *Vision Research, 38*, 2277–2288.

Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 986–1004.

Johnston, A., Hill, H., & Carman, N. (1992). Recognising faces: effects of lighting direction, inversion, and brightness reversal. *Perception, 21*, 365–375.

Kalocsai, P., Biederman, I., & Cooper, E. E. (1994). To what extent can the recognition of unfamiliar faces be accounted for by a representation of the direct output of simple cells. *Investigative Opthalmology Visual Science, 35*, 1626.

Knappmeyer, B., Thornton, I. M., & Bulthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research, 43*, 1921–1936.

Lades, M., Vortbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P., et al. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers, 42*, 300–311.

Lander, K., & Chuang, L. (2005). Why are moving faces easier to recognize. *Visual Cognition, 12*, 429–442.

Leder, H. (1999). Matching person identity from facial line drawings. *Perception, 28*, 1171–1175.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: the role of configural information in face recognition. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology, 53A*(513–536).

LeGrand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. *Nature, 410*, 890.

Liu, C. H., Collin, C. A., Burton, A. M., & Chaurdhuri, A. (1999). Lighting direction affects recognition of untextured faces in photographic positive and negative. *Vision Research, 39*, 4003–4009.

McMullen, P. A., Shore, D. I., & Henderson, R. B. (2000). Testing a two-component model of face identification: effects of inversion, contrast reversal, and direction of lighting. *Perception, 29*, 609–619.

O'Toole, A. J., Vetter, T., & Blanz, V. (1999). Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: an application of three-dimensional morphing. *Vision Research, 39*, 3145–3155.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, Massachusetts: MIT Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Rhodes, G., Brennan, S. E., & Carey, S. (1987). Recognition and ratings of caricatures: implications for mental representations of faces. *Cognitive Psychology, 19*, 473–497.

Riesenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society of London, B—Biological Sciences, 271*(Suppl. 6), S448–S450.

Rowland, D. A., & Perrett, D. I. (1995). Manipulating facial appearance through shape and color. *IEEE Computer Graphics and Applications, 15*(5), 70–76.

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception, 35*, 749–759.

Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences, 5*(5), 211–215.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, Massachusetts: MIT Press.

Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471–491.

Vuong, Q. C., Peissig, J. J., Harrison, M. C., & Tarr, M. J. (2005). The role of surface pigmentation for recognition revealed by contrast reversal in faces and Greebles. *Vision Research, 45*, 1213–1223.

Wiskott, L., Fellous, J.-M., Kruger, N., & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 7*(4), 935–948.

Wolberg, G. (1990). *Digital image warping*. Los Alamitos, CA: IEEE Computer Society Press.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141–145.

Yovel, G., & Kanwisher, N. (2004). Face perception: domain specific, not process specific. *Neuron, 44*, 889–898.