Vision Research 55 (2012) 41–46

Contents lists available at SciVerse ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Predicting the psychophysical similarity of faces and non-face complex shapes by image-based measures

Xiaomin Yue [a,*], Irving Biederman [b], Michael C. Mangini [c], Christoph von der Malsburg [d], Ori Amir [b]

[a] Martinos Center for Biomedical Imaging, Massachusetts General Hospital, 149 13th Street, Suite 2301, Charlestown, MA 02129, United States
[b] Department of Psychology and Neuroscience Program, University of Southern California, 3641 Watt Way, Los Angeles, CA 90089, United States
[c] Psychology Department, Concordia College, 901 8th St S. Morehead, MN 56562, United States
[d] Frankfurt Institute of Advanced Studies, 60438 Frankfurt am Main, Germany

ARTICLE INFO

Article history:
Received 19 April 2011
Received in revised form 29 September 2011
Available online 5 January 2012

Keywords:
Scaling shape similarity
Face discrimination
Gabor filtering
Shape discriminability

ABSTRACT

Shape representation is accomplished by a series of cortical stages in which cells in the first stage (V1) have local receptive fields tuned to contrast at a particular scale and orientation, each well modeled as a Gabor filter. In succeeding stages, the representation becomes largely invariant to Gabor coding (Kobatake & Tanaka, 1994). Because of the non-Gabor tuning in these later stages, which must be engaged for a behavioral response (Tong, 2003; Tong et al., 1998), a V1-based measure of shape similarity based on Gabor filtering would not be expected to be highly correlated with human performance when discriminating complex shapes (faces and teeth-like blobs) that differ metrically on a two-choice, match-to-sample task. Here we show that human performance is highly correlated with Gabor-based image measures (Gabor simple and complex cells), with values often in the mid 0.90s, even without discounting the variability in the speed and accuracy of performance not associated with the similarity of the distractors. This high correlation is generally maintained through the stages of HMAX, a model that builds upon the Gabor metric and develops units for complex features and larger receptive fields. This is the first report of the psychophysical similarity of complex shapes being predictable from a biologically motivated, physical measure of similarity. As accurate as these measures were for accounting for metric variation, a simple demonstration showed that all were insensitive to viewpoint invariant (nonaccidental) differences in shape.

1. Introduction

The cortical pathway supporting object recognition is generally held to consist of a series of stages, starting with V1, where cells are tuned to a particular scale and orientation of a restricted region of the visual field. Such tuning can be well characterized by a lattice of Gabor filters (De Valois & De Valois, 1990). Cells in the last visual stage, the inferotemporal cortex (IT) of the macaque, are mostly tuned to "moderately complex features" (Kobatake & Tanaka, 1994).

Some shape variation can be characterized as *metric*, such as the degree of non-zero curvature or the length of a contour, attributes which vary continuously with orientation in depth. Other shape variation is qualitative (or *nonaccidental*), such as whether a contour is straight or curved, attributes which are invariant to orientation in depth (Biederman, 1987; Lowe, 1985). Gabor filter values represent well the metric portion. Nonaccidental tuning is characteristic of IT cells, which are also relatively invariant to position, size, precise orientation in depth, local occlusion, and direction of illumination (Kayaert, Biederman, & Vogels, 2003; Kobatake & Tanaka, 1994; Kourtzi & Kanwisher, 2001; Vogels & Biederman, 2002), all variables that would affect Gabor activations. Both humans and macaques readily exploit nonaccidental differences in shape to achieve one-shot viewpoint invariance in depth rotation (e.g., Biederman & Bar, 1999; Biederman & Gerhardstein, 1993; Logothetis et al., 1994).

The objective of the present study was to assess the extent to which the psychophysical discrimination of complex shapes, of varying metric similarity, could be predicted from a measure of Gabor Euclidean distance between the stimuli (as well as other image metrics). Poor predictability would suggest that metric aspects of shape similarity that affect psychophysical performance arise at later stages of visual processing. Conversely, high predictability would raise a question as to the degree to which later stages of visual processing play a role in defining the metric similarity of shape underlying psychophysical discrimination.

We used two types of computer-generated stimuli; novel blobs resembling teeth and faces (Fig. 1, see Section 2). Computer generation eliminated the presence of local nonaccidental features, such

* Corresponding author.
E-mail address: xiaomin@nmr.mgh.harvard.edu (X. Yue).

0042-6989/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2011.12.012

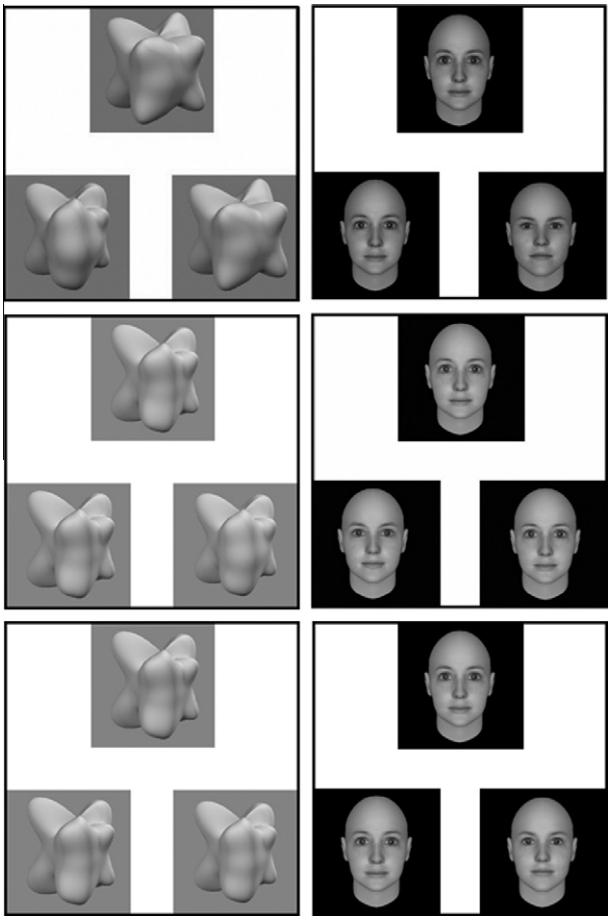

Fig. 1. Sample match-to-sample stimulus displays for blobs (left) and faces (right). The sample and matching stimuli are identical for all three blob and face displays. The only variation is with the dissimilarity of the distractor, which is at the 99th percentile (as assessed by the Gabor-jet model) of dissimilarity (high dissimilarity) in the top panel, at the 50th percentile (middle dissimilarity) in the middle panel and at the 10th percentile (low dissimilarity) in the bottom panel. High dissimilarity of the distractor to the matching stimulus would be associated with higher accuracy. From top to bottom, the correct choices for the blobs are right, left, left and for the faces, left, right, left.

as a beauty mark on a face or a scratch or stain on an object,[1] that observers tend to employ when faced with difficult discriminations (Biederman et al., 1999). The blobs were designed to allow metric variation of the contouring of smooth surfaces and thus reflect the same low level 3D shape variation as exists for faces. Given the non-Gabor coding in the later stages—which must be engaged for response selection (Sheinberg & Logothetis, 1997; Bar & Biederman, 1999; Pasupathy & Connor, 1999; Tong, 2003; Tong et al., 1998)—it might seem that predictability from a Gabor-based measure would be low.

Although prior work had established that the blobs are initially more difficult to discriminate than the faces, extensive training (>8000 trials) in discriminating the blobs renders their performance approximately equivalent to that of discriminating faces (Yue, Tjan, & Biederman, 2006). However, such training does not produce heightened sensitivity in the fusiform face area (FFA) (Kanwisher, McDermott, & Chun, 1997) to blobs nor does it noticeably affect the slope of the function relating performance to distractor similarity, although it does increase the BOLD response in the lateral occipital complex (LOC) (Grill-Spector, Kourtzi, & Kanwisher, 2001), an area sensitive to object (vs. texture) processing. Whereas face matching is dramatically impaired by contrast reversal, blob expertise training does not render the matching of blobs sensitive to contrast reversal (Nederhouser et al., 2007). The two types of stimuli thus provide a test bed for assessing the predictability by physical measures of metric variation of two major shape-based domains relevant to cognition.

For comparison measures, we studied the predictability of discrimination performance of pixel energy and the C1, S2, and C2 layers of HMAX. The HMAX model seeks to capture the coding along the ventral visual pathway with the C2 layer corresponding to IT (Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007).

2. Methods and procedure

2.1. Participants

The blob and face tasks were performed by 12 and 11 participants, all students at the University of Southern California, either for $8/h or as part of lab activities. Each participant performed 2016 trials. All procedures were approved by the USC IRB.

2.2. Tasks

To assess the effects of metric variations in shape, subjects performed a matched-to-sample task in which one of the two comparison stimuli, either computer generated faces or asymmetrical blobs, resembling teeth, was identical to the sample, which was centered above the two comparison stimuli (Fig. 1). The diagonal relationship between the sample and each of the comparison stimuli made it exceedingly difficult to pick out a local feature, e.g., a pixel, that could have distinguished distractor from matching stimulus, particularly given the limited amount of time to process the displays and that eye movements would be required to distinguish such local features.

2.3. Stimuli

Each image was 128×128 pixels. With the subject seated approximately .7 m from the display, each face or blob subtended a visual angle of 4.6°. The faces were generated by Facegen (Sinular Inversion, Vancouver, Canada) from a core image of a frontal view of a 20-year-old female Caucasian. Eight levels of two dimensions, the distance between the eyes and the mouth and the height of the cheekbones, yielded 64 faces. These dimensions were designed to vary the identity of the faces without producing differential activation of standard face categories such as pose, sex, race, age, expression, or attractiveness. The dimensions were selected to minimize the presence of nonaccidental differences among the faces such as whether the eyebrows were straight or curved. The variation of the two face dimensions, although subtle at first glance, allowed a full range of discrimination performance, from chance to 100% accuracy following a scheme for generating 2D harmonics of a circle devised by Shepard & Cermak (1973).

The blobs were parametrically defined combinations of spherical harmonics, generated by combining a sphere and the fourth harmonic of that sphere with eight different orientations of the second and third harmonics. This generated a toroidal space of 64 blobs following a scheme for generating 2D harmonics of a circle devised by Shepard and Cermak (1973). The blobs were designed to be non-face control stimuli for faces with preservation of the approximate smooth sculpting, compactness, and complexity of a face without appearing face like. Although these

[1] Close inspection of the stimuli did not reveal any artifactual local features that, if present, could have affected discrimination performance, an observation confirmed by the consistency of the ordering of accuracy across adjacent similarity values determined by the underlying stimulus generating functions. The presence of artifactual cues would have been expected to render adjacent similarity values more discriminable than expected from these functions.

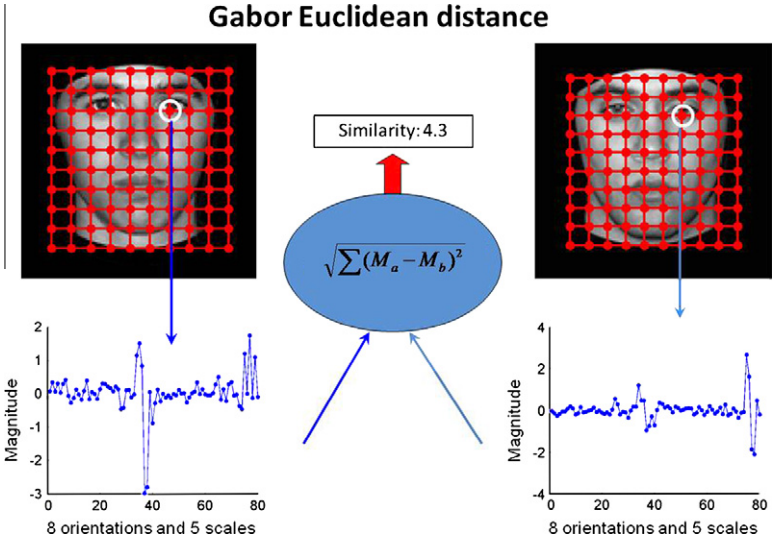

Fig. 2. Illustration of the dissimilarity calculation of the simple cell Gabor-jet model. A jet of 80 filters (5 scales × 8 orientations × 2 phases [sine and cosine]) is centered at each of the 10 × 10 nodes in the grid. *M* is the magnitude of activation of each filter. For the Gabor-jet measure, the dissimilarity of two images is the Euclidean distance between their responses to 80 Gabor filters on a 10 × 10 grid (see Section 2). Here, we illustrate this process with photos of faces of two individuals.

blobs did not have albedo variations that matched that of faces, such variations can readily be applied to their surfaces as in Nederhouser et al. (2007). The blobs are further described in Yue, Tjan, and Biederman (2006) and can be downloaded from http://geon.usc.edu/Create_blobs.m.

On each trial, three images were presented in a triangular array, with one image on top and two below (Fig. 1). The top stimulus was the sample and one of the two lower stimuli—the "matching" stimulus—was identical to the sample. A match-to-sample task has an advantage over the frequently employed same-different task in that the former eliminates decision criterion effects which can lead to significant below chance responding for stimuli that are different but highly similar in shape.

The exposure durations (300 ms for the faces, 1500 ms for the blobs) were selected so as to allow a wide range of error rates as a function of distractor dissimilarity from near chance to near error-less performance. As noted previously, the longer exposure duration for the blobs was necessary with observers who were not extensively trained on them. Such training (>8000 trials) can bring the accuracy of their matching to a level that is equivalent to that of faces without producing face-like coding of the blobs or affecting the correlation with the Gabor jet measure (Yue, Tjan, & Biederman, 2006).

2.3.1. Stimulus similarity scaling

The dissimilarity of the non-matching stimulus to the matching stimulus was varied with values scaled according to three measures of physical image similarity—Gabor-jet simple cells, pixel energy, and the C2 stage of HMAX[2] (the other three measures, Gabor-jet complex cells, the C1, S2 stages of HMAX, were also calculated). The similarity correlation matrix between those six measures are available online (http://geon.usc.edu/scaling_paper_correlation.doc), which allowed performance as a function of the similarity measures of the non-matching to the matching stimulus to be assessed.

The Gabor-jet model (Lades et al., 1993 $f(\vec{x}) = \frac{\vec{k}^2}{\delta^2} e^{\left(-\frac{\vec{k}^2\vec{x}^2}{2\delta^2}\right)} \left(e^{i\vec{k}\vec{x}} - e^{\left(\frac{\delta^2}{2}\right)}\right)$), illustrated in Fig. 2, tiles the visual field with columns (or "jets") of cells at multiple orientations and scales—corresponding to a simplified V1 simple cell hypercolumn. The implementation of the model that we adopted had five scales and eight orientations at sine and cosine phases in each jet (for a total of 80 kernels) with a jet centered at each node in a 10 × 10 grid centered over the stimulus. (All the stimuli were generated from the same central position so no repositioning of the grid was required.) A Matlab version of the model can be downloaded at http://geon.usc.edu/GWTgrid_simple.m.

For the simple cell Gabor-jet model, each jet was composed of 80 Gabor filters (or kernels) at eight equally spaced orientations (i.e., 22.5° differences in angle) × 5 scales, each centered on their jet's grid point. The coefficients of the kernels (the magnitude corresponding to an activation value for a simple cell) within each jet were then concatenated to an 8000-element (100 jets × 80 kernels) vector *G*: $[g_1, g_2, \ldots, g_{8000}]$. For any pair of pictures with corresponding jet coefficient vectors *G* and *F*, the dissimilarity of that pair was calculated by the Euclidean distance ||*G*–*F*||, the square root of the sum of the squared differences of the coefficients of the sine and cosine simple cells (corresponding to activation values). The Euclidean distances of all image pairs were then sorted in ascending order, and then grouped into 11 levels (10 levels for the Gabor complex cells and HMAX C2) with equal distances between each level. With equal distances, each level had different numbers of trials, because the similarity distribution was not uniform.

Correlations with the data were also computed for the C1, S2, and C2 stages of HMAX (Mutch & Lowe, 2008; Serre, Oliva, & Poggio, 2007; http://www.mit.edu/~jmutch/fhlib/(v8)). HMAX is a widely cited model of ventral pathway cell tuning in which first stage, S1 (similar to the Gabor simple cell model), cell activations are combined in an operation that selects the highest (maximum) activation to form complex cells, C1, across scales. Such cells thus create larger receptive fields. By combining S2 units, the C2 stage of HMAX computes units with still larger receptive fields and more complex feature combinations (such as vertices).

The model was trained on a database (http://www.vision.caltech.edu/html-files/archive.html) of photographs of faces of 18 people, with 18 exemplars of each individual. All were frontal views, with variation in background, degree of lighting, and expression. When tested on additional exemplars of those individuals that were not included in the training set, the model was correct on 76% trials. We also trained the model on the Caltech 101

[2] We also assessed the correlation of error rates with the S1 stage of HMAX. However, since that stage is strongly and linearly correlated with the Gabor-jet simple-cell model (Lades et al., 1993), .96 for blobs and .99 for faces, we elected to just present results for the latter as it enjoys precedent and has been used in prior studies (e.g., Nederhouser et al., 2007; Xu & Biederman, 2010).

Table 1
Correlations among the three measures of (a) blob similarity and (b) face similarity.

	Gabor simple	Pixels	HMAX_C2
(a)			
Gabor simple	1	0.9190	0.8980
Pixels	0.9190	1	0.9065
HMAX_C2	0.8980	0.9065	1
(b)			
Gabor simple	1	0.9928	0.9060
Pixels	0.9928	1	0.9318
HMAX_C2	0.9060	0.9318	1

database of objects. However, since the dissimilarity measures for both faces and blobs were nearly identical with the two training sets ($r > .95$) we present here only the measures produced by the model trained on faces. Two additional measures were the pixel-wise Euclidean distance and the Gabor-jet complex cell. All measures and the data are available from http://geon.usc.edu/Art_FacePairs_Euclidean_Gabor80_Pix_HMAXC2.xls for faces, and http://geon.usc.edu/BlobPairs_Euclidean_Gabor80_Pix_HMAXC2.xls for blobs.

Table 1 shows the intercorrelations among the three image measures for the present stimuli. These correlations are all moderately to strongly positive so it would be expected that if one of these measures would be a good predictor of performance, the others would be good predictors as well.

3. Results

For each measure, 11 bins provided a compromise between having a sufficiently large number of observations within each bin so that there were no equally spaced bins with missing or very few values[3] and enough bins so that a psychophysical function could be well defined (Figs. 3 and 4). Within each bin, the error rate was calculated as the number of incorrect responses divided by the total number of trials for that bin. Overall, it is readily apparent that, for both blobs and faces, Figs. 3 and 4, respectively, the correlations between error rates and the stimulus dissimilarity values are impressively high for all three measures of image dissimilarity, especially considering that there is some inherent variability in the performance of the subjects not associated with the similarity of the distractors. For blobs, the correlation of the simple cell measure was significantly higher than that for the pixel measure, $t(11) = -3.01$, $p < 0.05$, but there were no reliable differences between simple cell vs. HMAX C2, and pixel vs. HMAX C2 measures. For faces, none of the differences in correlations were significant.

4. Discussion

The high correlations of the Gabor measures with performance suggest that despite the necessity to pass through later stages of the ventral pathway, as well as response selection stages in frontal cortex, discrimination performance of metrically varying complex stimuli is well predicted by a Gabor-based measure of similarity. It also may not be necessary to posit explicit face units, such as those assumed by Jiang et al. (2006), nor norm-based encoding, such as that assumed by Leopold, Bondar, and Giese (2006), to account for psychophysical similarity mediating metric discriminability although other processes, such as adaptation, may reflect these other forms of representation.

Our interest was in the representation of metric shape, rather than in testing various schemes for achieving invariance, such as "spotlight" models of attention, so we did not test translation and scale invariance for our stimuli, which in the absence of a capacity for achieving such invariance would have been low for all measures but C2 of HMAX. That the C2 level of HMAX had approximately equivalent correlations with the Gabor simple cell measurement for faces suggests that the additional levels in HMAX (i.e., C1, S2, and C2) do not provide an enhanced representation of these metrically-varying stimuli. Instead, all their effect here might be in achieving scale and translation invariance. To the extent that they produce complex features that might facilitate classification into basic and superordinate level categories (Serre, Oliva, & Poggio, 2007), such features were not differentially present in our stimuli over the different similarity levels.

What are the likely limits to the extent to which Gabor type coding can represent the similarity of differences between shapes of visual entities? That the pixel measures showed a reasonably high correlation with error rates is likely a function of these stimuli lacking significant protuberances differing in orientation (see Fig. 5). (A frontal view of a face with an elongated part sticking out of its side would not be a face.) The pixel measure, which makes neither orientation nor scale of a shape feature explicit, would not be able to reflect that a large difference in orientation or position of a part should produce a higher dissimilarity value than a small difference in orientation or position, given that the changes in both cases had no pixels in common with the original stimulus. These effects are illustrated with the line drawings in Fig. 5 where the measure of pixel similarity has A&B equivalent to A&C whereas the latter would be much easier to discriminate. The Gabor-jet measure does, indeed, have A&B more similar than A&C. Hummel (2000) has provided a number of instances where pixel template models fail to capture such readily recognizable differences among stimuli. Where all the measures considered in the current work come up short is in their lack of sensitivity to nonaccidental differences. The Gabor-jet simple cell measure has AB equivalent to AD. Somewhat surprising, the HMAX C2 measure, whether trained on faces or objects, has the orientation difference of AB *more* dissimilar than the nonaccidental difference of AD (raw Euclidean, face training, AB = .54, AD = .43; object training, AB = .91; AD = .59) yet for humans the discrimination of A&D would likely be much easier and faster than A&B. This is not a "one-off" result. A systematic study (Amir, Biederman, & Hayworth, 2011, unpublished) compared the discriminability of nonaccidental vs. metric differences among a set of geons where the differences were equated according to the Gabor-jet model (Lades et al., 1993). The HMAX C2 stage (current version of the that model) also failed to produce greater Euclidean distances for the nonaccidental differences despite their markedly greater discriminability in terms of human performance (e.g., Biederman & Bar, 1999).

To our knowledge, this is the first time that the psychophysical similarity of complex shapes has been predicted on the basis of a biologically-based physical measure of the stimuli. This high predictability from image-based dissimilarity measures of discrimination performance for the two classes of stimuli means that it is possible to achieve an "apples and oranges" scaling so that a given distance between a pair of faces, for example, can be matched to a given distance between pairs of blobs, as illustrated in Fig. 6.

The high correlation between first stage cortical dissimilarity and psychophysical performance should not be interpreted to mean that pairs of stimuli from different classes, say faces and blobs, at equal levels of scaled dissimilarity, would yield equivalent performance levels. As noted previously, other factors, say familiarity or inherent dedicated mechanisms for processing a particular class or stimuli, could affect performance. In fact, this was likely the case with the faces and blobs in that, as noted in the Introduction, extensive training on the blobs renders their overall performance levels equivalent to that of faces, and produces a larger

[3] There were only 10 bins for the Gabor complex cells and HMAX C2 because with 11 bins the bin with the most dissimilar values had less than 10 trials, rendering those data points unreliable. The plots for all six measures can be download from http://geon.usc.edu/scaling_paper_six_plot.pdf.

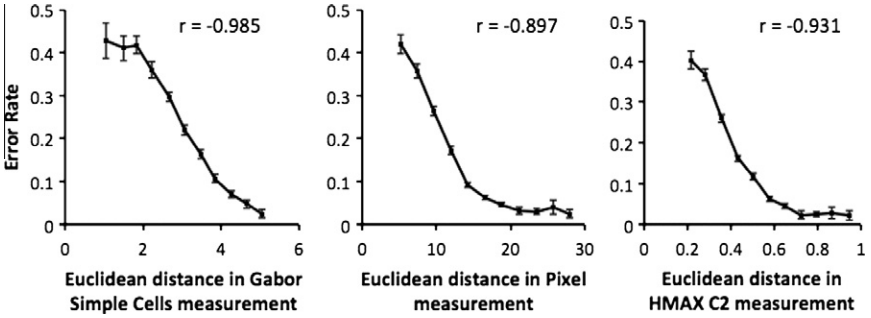

Fig. 3. Blob match-to-sample error rates as a function of the dissimilarity of the distractor to the matching stimulus for the three measures of shape dissimilarity. The error bars are the standard error of the means with between subjects variance removed. The functions for Gabor complex cell, HMAX C1, and S2 are available at the link given previously.

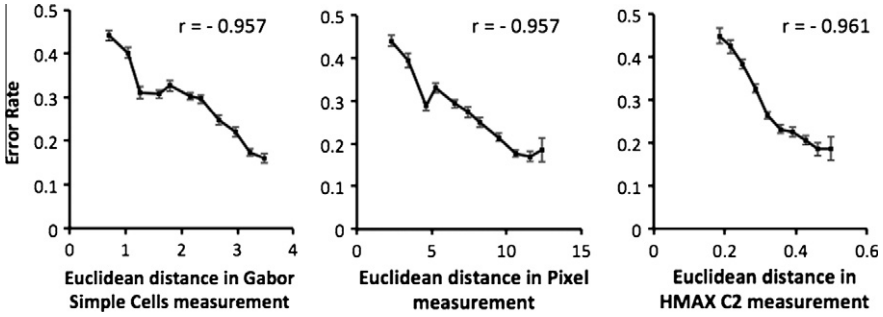

Fig. 4. Face match-to-sample error rates as a function of the dissimilarity of the distractor to the matching stimulus for the three measures of shape similarity. The error bars are the standard errors of the mean with between subjects variance removed. The functions for Gabor complex cell, HMAX C1 and S2 are available at the link given previously.

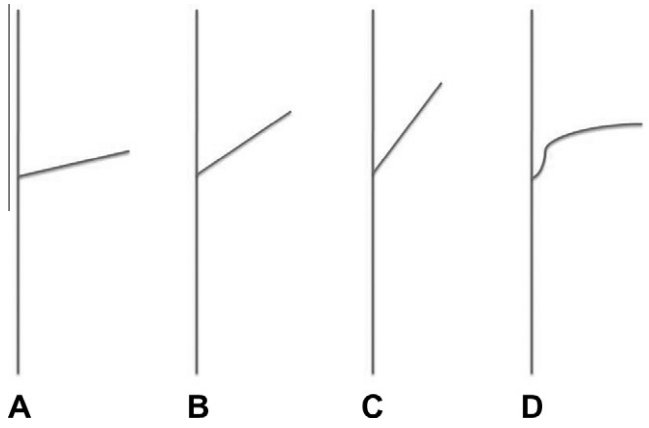

Fig. 5. Four simple stimuli illustrating the pixel measure's insensitivity to orientation differences and the insensitivity of the Gabor jet and HMAX models to nonaccidental differences. The pixel similarity measure has A&B and A&C equivalent, reflecting that measure's insensitivity to non-overlapping orientation differences, in contradiction to their discrimination difficulty. Similarly, the Gabor simple cell similarity measure has A&B equivalent to A&D, reflecting the insensitivity of that measure to nonaccidental differences. HMAX C2 scales A&B to be *more* dissimilar than A&D in clear contradiction to their likely discrimination difficulty.

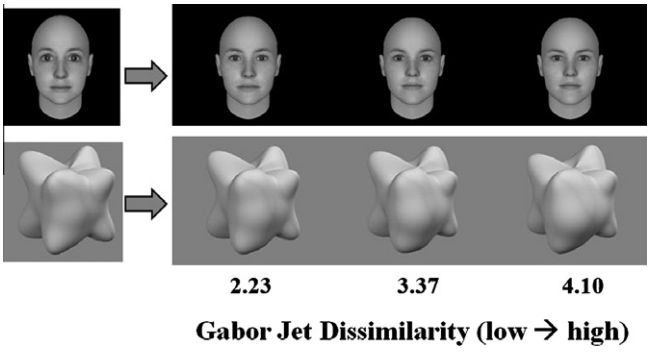

Fig. 6. Illustration of scaling equivalences for faces and blobs at different Gabor-jet simple cell dissimilarity values. The face and blob at the same horizontal location are at equal distances (dissimilarities) from the leftmost stimulus.

BOLD response in LOC (Yue, Tjan, & Biederman, 2006). However, this reduction in error rates was on the intercept, not on the slope relating the performance measures to Gabor similarity.

We hypothesize that when the Gabor dissimilarity measure is not predictive of performance for equally practiced sets of stimuli then cells sensitive to nonaccidental properties are being differentially activated by the stimuli. Such cells presumably are to be found in later stages of the ventral pathway (e.g., Kayaert, Biederman, & Vogels, 2003; Kobatake & Tanaka, 1994), Without a scaling of V1 dissimilarity, one would not know whether any observed behavioral differences in discriminating different classes of stimuli was a reflection of early vs. later stages of processing (Yue et al., 2011).

In the main, our results show that an image-based similarity measure derived from V1-like simple-cell tuning, i.e., Gabor filtering, predicts the psychophysical similarity of metric variation of complex shapes almost perfectly.

Acknowledgments

This work was supported by the National Science Foundation Grant numbers 0531177 and 0617699 to I.B. We thank Bosco Tjan for helpful discussions.

References

Amir, O., Biederman, I., & Hayworth, K. J. (2011). *Sensitivity to nonaccidental vs. metric shape differences*. Unpublished ms., University of Southern California.

Bar, M., & Biederman, I. (1999). Localizing the cortical region mediating visual awareness of object identity. *Proceedings of the National Academy of Sciences, 96*, 1790–1793.

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

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research, 39*, 2885–2899.

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

Biederman, I., Subramaniam, S., Bar, M., Kalocsai, P., & Fiser, J. (1999). Subordinate-level object classification reexamined. *Psychological Research, 62*, 131–153.

De Valois, R. L., & De Valois, K. K. (1990). *Spatial vision*. New York: Oxford.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41*, 1409–1422.

Hummel, J. E. (2000). Where view-based theories break down: The role of structure in shape perception and object recognition. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 157–185). Hillsdale, NJ: Erlbaum.

Jiang, X., Rosen, E., Zeffiro, T., VanMeter, J., Blanz, V., & Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using FMRI and behavioral techniques. *Neuron, 50*(1), 159–172.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*, 4302–4311.

Kayaert, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *Journal of Neuroscience, 23*, 3016–3027.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology, 71*, 856–867.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital cortex. *Science, 293*, 1506–1509.

Lades, M., Vorbruggen, 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.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature, 442*(7102), 572–575.

Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology, 4*, 401–414.

Lowe, D. (1985). *Perceptual organization and visual recognition*. Boston: Kluwer.

Mutch, J., & Lowe, D. G. (2008). Object class recognition and localization using sparse features with limited receptive fields. *International Journal of Computer Vision, 80*(1), 45–57.

Nederhouser, M., Yue, X., Mangini, M. C., & Biederman, I. (2007). The deleterious effect of contrast reversal on recognition is unique to faces, not objects. *Vision Research, 47*, 2134–2142.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology, 82*, 2490–2502.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, 104*, 6424–6429.

Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences, 94*, 3408–3413.

Shepard, R. N., & Cermak, G. W. (1973). Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology, 4*, 351–377.

Tong, F. (2003). Primary visual cortex and visual awareness. *Nature Reviews Neuroscience, 4*, 219–229.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21*, 753–759.

Vogels, R., & Biederman, I. (2002). Effects of illumination intensity and direction on object coding in macaque inferior temporal cortex. *Cerebral Cortex, 12*, 756–766.

Xu, X., & Biederman, I. (2010). Loci of the release from fMRI adaptation for changes in facial expression, identity, and viewpoint. *Journal of Vision, 10*, 1–13.

Yue, X., Cassidy, B. S., Devaney, K. J., Holt, D. J., & Tootell, R. B. H. (2011). Lower-level stimulus features strongly influence responses in the fusiform face area. *Cerebral Cortex, 21*, 35–47.

Yue, X., Tjan, B., & Biederman, I. (2006). What makes faces special? *Vision Research, 46*, 3802–3811.