Object feature validation using visual and haptic similarity ratings

Theresa Cooke, Sebastian Kannengiesser, Christian Wallraven, and Heinrich H. Bülthoff
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

The perceived similarity between objects may well vary according to the sensory modality/modalities in which they are experienced, an important consideration for the design of multimodal interfaces. In this study, we present a similarity-based method for comparing the perceptual importance of object properties in touch and in vision and show how the method can also be used to validate computational measures of object properties. Using either vision or touch, human subjects judged the similarity between novel, 3D objects which varied parametrically in shape and texture. Similarities were also computed using a set of state-of-the art 2D and 3D computational measures. Two resolutions of 2D and 3D object data were used for these computations in order to test for scale dependencies. Multidimensional scaling (MDS) was then performed on all similarity data, yielding maps of the stimuli in both perceptual and computational spaces, as well as the relative weight of shape and texture dimensions. For this object set, we found that visual subjects accorded more importance to shape than texture, while haptic subjects weighted them roughly evenly. Fit errors between human and computational maps were then calculated to assess each feature's perceptual validity. Shape-biased features provided good overall fits to the human visual data; however, no single feature yielded a good overall fit to the haptic data, in which we observed large individual differences. This work demonstrates how MDS techniques can be used to evaluate computational object features using the criterion of perceptual similarity. It also demonstrates a way of assessing how the perceptual validity of a feature varies as a function of parameters such as the target modality and the resolution of object data. Potential applications of this method for the design of unimodal and multimodal human-machine interfaces are discussed.

Categories and Subject Descriptors: I.4.7 [**Image Processing and Computer Vision**]: Feature Measurement—Feature representation, size and shape, texture; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—Artificial, augmented and virtual systems, evaluation/methodology; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—Haptic I/O, evaluation/methodology

General Terms: Experimentation, Human Factors, Measurement

Additional Key Words and Phrases: similarity, multidimensional scaling, perception, vision, touch, haptic, features, validation, shape, texture

1. INTRODUCTION

The design of effective and efficient multimodal displays requires an understanding of how humans make use of their different senses to build up representations of their surroundings. Models of human visual object processing have proposed that the visual system extracts object features or properties from images projected onto the retina [Marr 1982]. These fea-

e-mail: theresa.cooke@tuebingen.mpg.de

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.
© 20YY ACM 0000-0000/20YY/0000-0001 $5.00

tures are then used as the basis for representing objects in the brain [Bülthoff and Edelman 1992; Ullman 1996]. Inspired by this, a similar approach has been taken in the field of computer vision: a set of computational measures are extracted from 2D images of objects (or scenes) and used to create artificial representations of objects for automated reconstruction, recognition, or categorization tasks [Riesenhuber and Poggio 1999; Ullman et al. 2002]. Work in this field has given rise to a large number of feature extraction algorithms, including biologically-inspired filters (Gabor jets) which mimic the response of cells in visual cortex [Jones and Palmer 1987] and algorithms derived from statistical optimization procedures [Ankerst et al. 1999]. These computational measures have been evaluated in a variety of ways, e.g., based on their performance in machine vision tasks. Biological plausibility has mainly been assessed at a relatively low level (e.g., by matching receptive field properties). In this paper, we propose a new method for validating computational measures based on the *high-level, cognitive criterion of object similarity*. Similarity is thought to underlie a number of cognitive processes, including both categorization [Rosch et al. 1976] and recognition [Edelman 1999]. The similarity-based criterion we propose is as follows: a good feature is one which, for a set of parametrically-defined objects, generates similarity-based stimulus configurations akin (in one or more respects) to those derived from human similarity ratings.

Most perceptual validation of computational object features has been carried out in relation to *visual* perception. However, a feature's perceptual validity may well vary as a function of sensory modality used to interact with the objects, e.g., [Klatzky et al. 1987]. The method presented in this paper provides a solution to this problem by enabling validation to be performed relative to any sensory modality. For the haptic modality, measures computed on 3D objects are particularly interesting, e.g., [Nefs and Kappers 2003], and a large number of such measures have been proposed in the 3D graphics literature [Funkhouser et al. 2003]. However, there have been few studies which have assessed these measures relative to the haptic modality using a high-level, cognitive criterion such as similarity. Knowledge of which 3D computational measures correlate with high-level human stimulus representations derived from haptic perception would not only help in the design of more realistic artificial haptic systems (for example, [Acosta et al. 2002]) and reduce the heavy demands of haptic rendering [Salisbury et al. 2004], but could also play an important role in elucidating the computational mechanisms of the human haptic system.

Our method can be situated in the context of a framework connecting the development of artificial representational systems and advances in our understanding of human representational systems (Figure 1). Physical objects constitute the input to both types of systems, which use various sensors to measure object properties (photoreceptors, mechanoreceptors, etc.). For artificial systems, the way these properties are extracted depends on the sensor and the computational algorithm applied to the measured quantities. For humans, it is a function of sensory modality. In both human and artificial systems, the extracted properties can then be used (either directly or indirectly) to embed objects in a *representational space* or "map." With the appropriate tools, these representational spaces can be compared at either the *unimodal* or at the *multimodal* level. Comparing a map derived from human unimodal perception (e.g., from pure visual exposure to the objects) to a map derived using a computational measure (e.g., pixel-wise differences between images) allows for *unimodal validation* of computational measures. Two human unimodal maps can also be compared to identify modality-dependent differences in human object processing. The

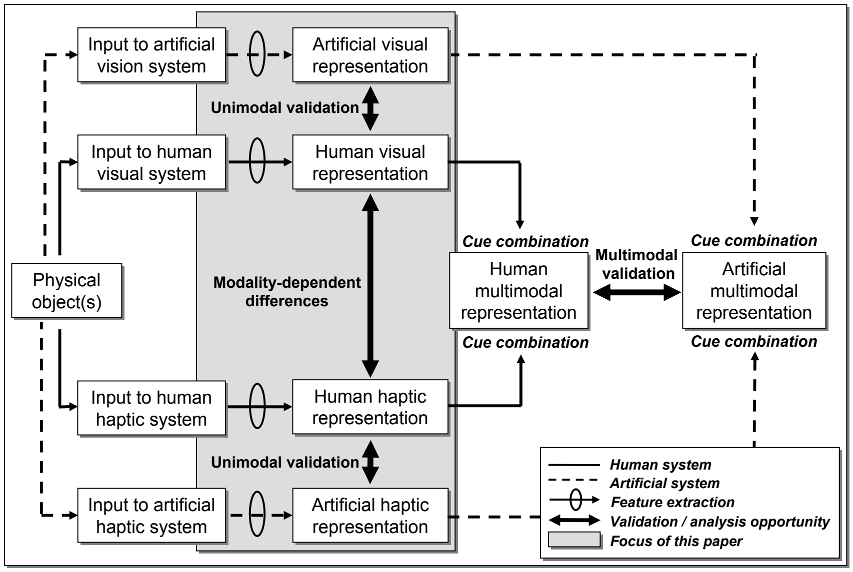

Fig. 1. A framework for studying human and artificial uni/multimodal object representations.

same approach can be applied at the *multimodal level* to test hypotheses about human cue combination and to validate approaches to artificial cue combination (e.g., in the design of visuotactile interfaces for telemedecine).

The method presented in this paper connects perceptual and artificial systems at the level of *unimodal* representations. We derive maps of our stimuli based on human visual and haptic similarity measures, and from similarity measures using a set of computational methods which we wish to perceptually validate. We first show how our method can be used to compare *human* haptic and visual stimulus maps. Then, we demonstrate how the method can be used to evaluate the perceptual validity of the computational measures by comparing the human maps against those derived from the computational measures.

2. METHODS

2.1 Stimuli

The stimuli consist of a family of novel, 3D objects (Figure 2), created in the graphics package 3D Studio Max 6.0. This software provides full control of object properties such as size, shape, and texture, allowing them to be varied in defined steps. The family begins with a family “prototype” (see Figure 2, object 1), which consists of: 1) three parts connected to a center sphere, defining the object’s macrogeometry and 2) a displacement map applied to the 3D mesh, specifying the object’s microgeometry. The other family members are generated by two manipulations. The first manipulation increases the smoothness of the object’s microgeometry (or “texture”) by decreasing the amount of mesh displacement caused by the displacement map. The second manipulation increases the smoothness of the object’s macrogeometry (or “shape”) by moving mesh vertices towards a local average, removing sharp angles in the global shape. Note that from a haptic rendering perspective, two distinct sets of force-rendering algorithms would be needed to convey these two kinds of variations: geometric-dependent rendering algorithms for shape variations and surface property-dependent rendering algorithms for texture variations [Salisbury et al. 2004].

The displacement map applied to the objects consisted of triangular-shaped elements with a base width of 3mm, a peak width of 2mm, and a maximum height of 3mm from the surface of the object; texture elements were spaced 3-5 mm apart. The scale of this pattern qualifies it as a macrotexture, the properties of which are known to be encoded by SAI mechanoreceptors [Klatzky and Lederman 2003]. The displacement manipulation simultaneously reduced element height and increased the peak width of elements; inter-element spacing remained constant. Increasing element width has been shown to decrease perceived roughness [Sathian et al. 1989]. In previous work with this stimulus set [Cooke et al. 2005; Cooke et al. 2006], subjects consistently described the objects as varying in "texture," and often referred to their "roughness" or "bumpiness."

The macrogeometrical manipulation applied to the objects averages out sharp angles in the mesh. In previous studies, subjects consistently referred to this manipulation as a change of "shape." Since the sharpest angles are located at the objects' extremities (maximum surface area of roughly 1 cm^2), the mesh relaxation affects these areas much more than the rest of the object (and thus has a more localized character than the evenly-distributed texture manipulation). It has been shown that SAI mechanoreceptors from a single finger are capable of curvature estimates for angles which fall onto the same region of the skin [Srinivasan and LaMotte 1991]. In addition to static cues, changing local curvature gradients created during object exploration can also provide macrogeometrical information [Pont et al. 1999]. Finally, kinesthetic cues to macrogeometry are provided by systematic changes in finger joint and wrist angles during object exploration.

Objects created using these variations can be plotted in a 2D space whose dimensions correspond to microgeometry and macrogeometry (Figure 2). The 3D models were printed out (Dimension 3D Printer, Stratasys, Minneapolis, USA) into hard, white, and opaque objects, measuring 9.0 +/- 0.1 cm wide, 8.3 +/- 0.2 cm high, and 3.7 +/- 0.1 cm deep and weighing about 40 g.

2.2 Visual similarity ratings

Ten subjects with normal or corrected-to-normal vision were paid 8 Euros per hour to rate the similarities between photographs of the objects presented at 75 Hz on a Sony Trinitron 21" monitor with a resolution of 1024 x 768 pixels. Photographs of the objects were displayed using the Psychtoolbox extension for MATLAB [Brainard 1997] on a Macintosh G4 computer. The photographs were taken such that the three object parts were aligned with the image plane (referred to as "frontal view"). This viewpoint was chosen in order to provide the best possible match to the viewpoint presented in the haptic task (see below). The image size was 7.6 x 7.6 degrees of visual angle (set to be approximately the same size as if the object were being held at arm's length). Subjects had never seen or touched the objects before. They were seated approximately 60 cm from the monitor in a dimly-lit room. A fixation cross appeared for 500 ms and then each of the objects appeared for 500 ms, separated by a 500 ms interstimulus interval. At the end of each trial, subjects had to rate the similarity of the objects on a scale between one (low similarity) and seven (high similarity). A set of practice trials enabled the subjects to become familiar with the task. Response time was unlimited. There were six experimental blocks of 325 randomized trials (each object was compared once with itself and once with every other object, yielding 25

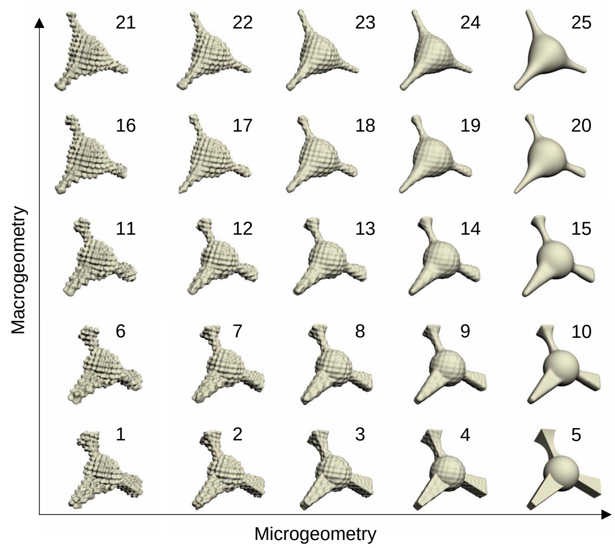

Fig. 2. Stimuli varied parametrically in terms of microgeometry (texture) and macrogeometry (shape).

+ (25·24)/2 = 325 trials.) The order of appearance of stimuli in each pair was randomized over the blocks. The total experiment lasted about two hours. At the end of the experiment, subjects were asked to write a short description of how they had judged similarity amongst the objects.

2.3 Haptic similarity ratings

Ten right-handed subjects were paid 8 Euros per hour to rate the similarities between the objects after exploring them haptically. None of the subjects had participated in the visual experiment, or had seen or touched the stimuli before. Subjects sat in front of a table, facing an opaque curtain through which they placed their right hand. They were instructed to keep their eyes closed during the experiment (subjects in a pilot study had reported that they were better able to concentrate on the task with their eyes closed). Behind the curtain, the experimenter presented two objects, one after the other. The objects were always presented in the same fixed position, face up on the table. Subjects were given up to ten seconds to trace the contour of each object, after which they rated the similarity between the objects on a scale from one (low similarity) to seven (high similarity). The contour-following procedure was chosen because it has been shown to allow for haptic extraction of a wide range of object properties, including local texture and global shape [Lederman and Klatzky 1993]. In the ten seconds provided, even untrained subjects had sufficient time to trace the object's contour twice. A set of practice trials allowed the subjects to become familiar with the task. The full experiment consisted of three blocks of 325 randomized trials spread out over five two-hour sessions on consecutive days. The order of appearance of stimuli in each pair was randomized over blocks. At the end of the experiment, subjects were asked to write a short description of how they had judged similarity among the objects.

2.4 Computational similarity measures

We implemented nine computational similarity measures: five operating on 2D photographs of the objects and four operating on the objects' 3D mesh geometry. The photographs used to compute similarity values were the same images presented in the human visual similarity experiment.

Given the basic 2D nature of visual input and the 3D nature of haptic input, we hypothesized that 2D measures could provide better fits to maps derived from visual similarities, while 3D measures could provide better fits to maps derived from haptic similarities. Human performance in visual object recognition tasks has been successfully explained by 2D view-based models [Bülthoff and Edelman 1992]. At the same time, however, the human visual system is capable of extracting and using 3D object properties from 2D images, such as in the case of shape-from-shading [Blake and Bülthoff 1991] and shape-from-texture [Todd et al. 2004]; thus we were also compared object maps based on human vision against maps generated from 3D features.

Fewer studies have examined the question of feature dimensionality for haptic object representations. A number of neuroimaging and psychophysical studies have emphasized the role of the visual cortex in the mediation of high-level tactile object representations (e.g., [Deibert et al. 1999]), while other studies have argued for more independent representations [Reed et al. 2004]. Our approach to this issue is to test whether 2D features correlate equally well with vision and haptics, or whether 3D features provide better correlations with haptics than with vision, which would support the idea of greater representational independence. In order to investigate these questions, object representations derived from human visual and haptic similarity ratings were compared against maps derived from the following 2D and 3D computational measures:

(1) *2D image subtraction* A simple pixel-wise subtraction between RGB images of two objects was performed. We took the mean absolute difference over all pixels and RGB channels as the dissimilarity between the two objects.

(2) *2D edge detection* Canny edge detection was performed on each image using MATLAB's edge detection algorithm, resulting in binary edge images (pixel value set to 1 if the pixel location is found to belong to an edge and 0 if not). The mean pixel-wise difference between two edge images was taken as the dissimilarity. Edge detection has been suggested as an important step in models of human vision [Marr 1982].

(3) *2D Gabor jets* The images were filtered with Gabor jets [Nestares et al. 1998] and the pixel-wise difference in the filter response images was computed. The Gabor jet filter has been proposed as a biologically-plausible model for receptive fields in early visual cortex [Jones and Palmer 1987] and has recently been successfully applied in models of object and motion recognition [Giese 2004]. Here, we used a variant which applies Gabor filtering in four orientations using both even and odd channels. A response image was generated for each orientation by summing and squaring the even and odd responses. Pixel-wise difference images were computed at each orientation and the final dissimilarity between two images was computed as the mean over all pixels and all orientations.

(4) *2D Visual Difference Predictor* In addition to the first three relatively straightforward 2D measures, we generated similarity data using two more elaborate measures: the Visual Difference Predictor (VDP) and the Structural Similarity (SSIM) measure. These

are industry standards for computing image differences and thus serve as a benchmarks for comparing the performance of the other 2D measures [Cadik and Slavik 2004]. The VDP [Daly 1993; Mantiuk et al. 2005] incorporates a model of low-level human visual processing, including the visual system's non-linear adaptive response to light, its contrast sensitivity function, and a masking function which models variations in sensitivity related to image content. As our computed measure, we took the total number of pixels which the VDP detected as different in the two images with a probability of at least 95%.[1]

(5) *2D Structural Similarity* Like the VDP, the SSIM [Wang et al. 2004] takes properties of the human visual system into account and computes an index of structural difference between two images after removing differences in average luminance and contrast. [2]

(6) *3D subtraction* Mean Euclidean distance between all 3D vertex coordinates of two object meshes was computed; point-by-point subtraction was possible because the meshes were in correspondence. This measure is the 3D equivalent of our 2D image subtraction measure.

(7) *3D perimeter* Object perimeter was measured along a cross-section taken parallel to the frontal view and the difference was taken for each pair of objects. Although perimeter can be measured in 2D or 3D space, we refer to it here as a 3D measure because of the special role it could play in haptic feature extraction, particularly given that in this experiment, participants explored the objects by following their contours.

(8) *3D curvature* A stable, reliable curvature estimate was obtained by fitting an implicit surface representation to the object and extracting curvatures from it [Steinke et al. 2005]. To estimate an object's "bumpiness," the absolute value of the mean curvature was averaged over the whole surface.

(9) *3D shape* A measure based on 3D shape histograms was implemented, inspired by [Ankerst et al. 1999]. 3D space is partitioned in the radial direction (into shells), which are further subdivided to create bounded sectors (the bound of the outermost shell is determined by the size of the largest object). For each object, we counted the number of vertices populating each sector, thereby creating a 3D shape histogram for that object. As a dissimilarity measure, we took the mean absolute sector-wise difference in vertex count between two object meshes.

Scale variation To demonstrate how our method can be used to quantify the effects of changing the parameters used for computing object features, we chose to vary the resolution of object data. Selecting the scale of resolution at which to compute features is an important problem in computer vision [Lindeberg 1994]; the size of stimuli relative to sensory receptors is also a fundamental issue for both visual and haptic perception [Koenderink 1984; Klatzky and Lederman 2003]. For 2D images, data at the finer scale of resolution were 600x600 pixel images (presented in the visual similarity experiment) and data at the coarser scale consisted of the same images downsampled to 38x38. For 3D object data, meshes at the finer scale consisted of the 4728 vertices constituting the original

[1]The VDP can be seen either as a single computational measure or as a combination of several measures; in the latter case, our evaluation of the VDP can be considered to be an evaluation of this particular combination of measures.
[2]A MATLAB implementation of the SSIM was downloaded from http://www.cns.nyu.edu/~lcv/ssim/

models and data at the coarser scale consisted of the same models downsampled to 296 vertices, which remained in correspondence.

2.5 MDS analysis of similarity data

Similarity data were analyzed using a multidimensional scaling (MDS) technique. Performing multidimensional scaling analysis of similarity ratings (or other kinds of proximity data) is a standard approach used in cognitive psychology to explore the psychological structure in a data set [Borg and Groenen 1997]. Human and computational similarity data were analyzed using a non-metric MDS algorithm implemented in MATLAB. Non-metric MDS uses the *ranks* of the pair-wise distances as input, as opposed to their precise values. Because of this, the relationship between the similarity data and the distances in the output configuration may be non-linear. The algorithm returns the stress value (Kruskal's stress formula 1), which is used to determine the appropriate dimensionality of the output configuration. Stress values below 0.2 are generally accepted as an indication that the dimensionality of the output space is sufficient to faithfully represent the input distance information [Cox and Cox 2001]. We also calculated the proportion of variance in the similarity data accounted for by the output configuration of a given dimensionality, which we refer to as RSQ. The optimal number of dimensions needed to represent the objects can be determined by looking for either a sharp drop in the stress plot or a plateau in the RSQ plot. Here, we used the RSQ to estimate the perceptual importance of each dimension in the output maps: the RSQ for the 1D solution was taken as the weight for the first dimension and the additional increase in RSQ for the 2D solution was taken as the weight for the second dimension. MDS also returns the coordinates of each object in the output space (though the scaling and rotation of the configuration is not determined). The non-metric MDS technique we used does not provide an interpretation of the dimensions: they must be interpreted by visual inspection of the output configuration.

2.6 Validation of computational measures

To assess the perceptual validity of the computational measures, stimulus maps derived from these measures using MDS were fit to the stimulus maps derived from human similarity ratings. Errors in these fits were used to quantify the correspondence between the computational measure and human perception. Map fitting was performed using the Procrustes function in MATLAB. This function determines a linear transformation (translation, reflection, orthogonal rotation, and symmetric scaling) of the points in a matrix Y which minimizes the sum of squared distances to points in a second matrix X, i.e., it computes

$$\min_{b,T,c} \{\|Z - X\| : Z = bYT + c\}$$

where b is a scaling factor, T is an orthogonal rotation and reflection matrix, and c is a translation component. The returned minimum value is normalized by the scale of X which makes it possible to express the fit error as a percentage value and compare it across data sets with different scales.

3. RESULTS AND DISCUSSION

3.1 Visual similarity ratings

Similarity data Mean visual similarity ratings for the twenty-five objects are shown in Figure 3 (left). Similarity between pairs is distinctly higher when both objects come from either the set 1-15 or the set 16-21 than when one object comes from the set 1-15 and one comes from the set 16-21. Within these two sets, similarity varies according to texture level (e.g., decreasing similarity between object 1 and objects 2, 3, 4, and 5).

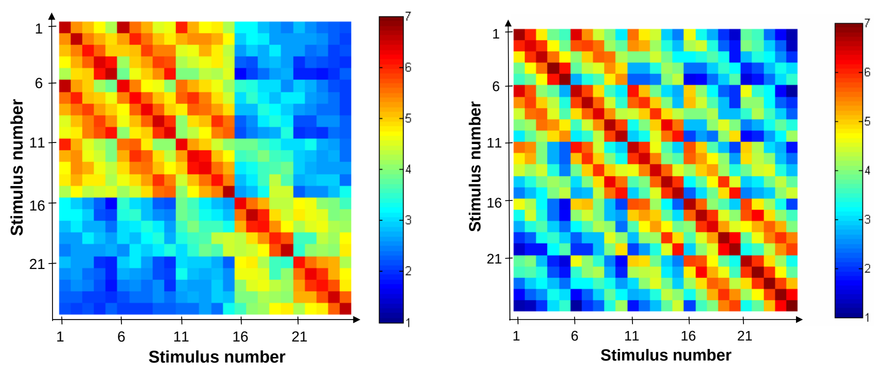

Fig. 3. Mean human visual similarity ratings (left) and mean human haptic similarity ratings (right).

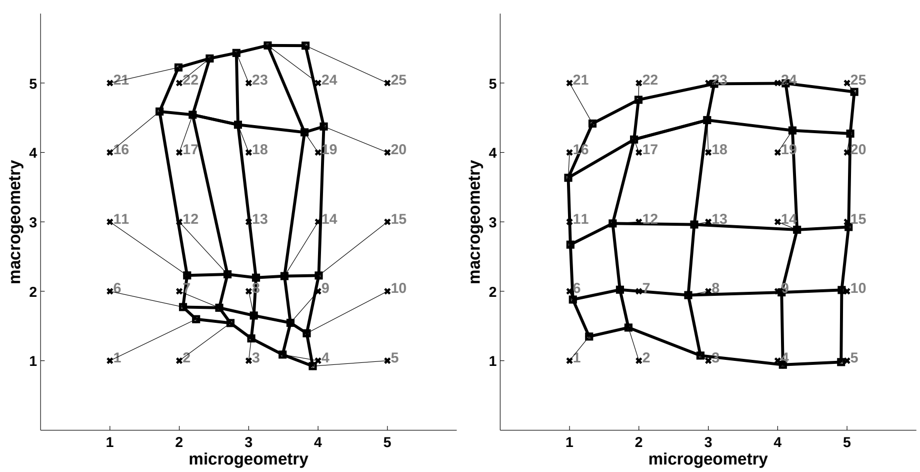

Fig. 4. Perceptual stimulus maps based on mean human visual similarity ratings (left) and mean human haptic similarity ratings (right).

MDS analysis Performing MDS allows these patterns in the similarity matrices to be more intuitively visualized as distances in a perceptual space. In order to determine the appropriate dimensionality of the output space, one needs to consider the stress value of the corresponding MDS solution. For mean visual similarity ratings, the stress for a one-dimensional solution was 0.14, indicating that one perceptual dimension is sufficient.

Dimension labels were interpreted by visual inspection of the output configuration: the map's most important dimension of variation corresponds to shape, while the second dimension corresponds to texture. Despite the high dimensionality of the visual measurement space, subjects were on average able to recover this low-dimensional variation in the stimulus set (see General Discussion).

In addition to the stress values, the dominance of shape can be seen from the RSQ weights for individual subjects (Figure 5, left). The mean shape weight across subjects was 0.85 (std. err. = 0.03), while the mean weight of the second dimension was 0.06 (std. err. = 0.01). Using a two-tailed t-test for independent samples with equal variances, the mean shape weight was found to be significantly different from the mean texture weight ($t(18)=23.6$, $p<0.01$). The greater importance of shape was also reflected in subjects' descriptions of how they judged similarity: 9 out of 10 subjects mentioned the word "shape" or global shape properties (e.g., geometric descriptions of parts), while 6 out of 10 subjects mentioned the word "texture" or texture-related properties (e.g., bumpiness).

Despite the dominance of shape, most subjects *were* able to recover the structure of the stimulus set along the texture dimension. Another interesting feature of the map is the emergence of two stimulus clusters along the shape dimension, suggesting a connection between similarity judgments and category structure (see General Discussion).

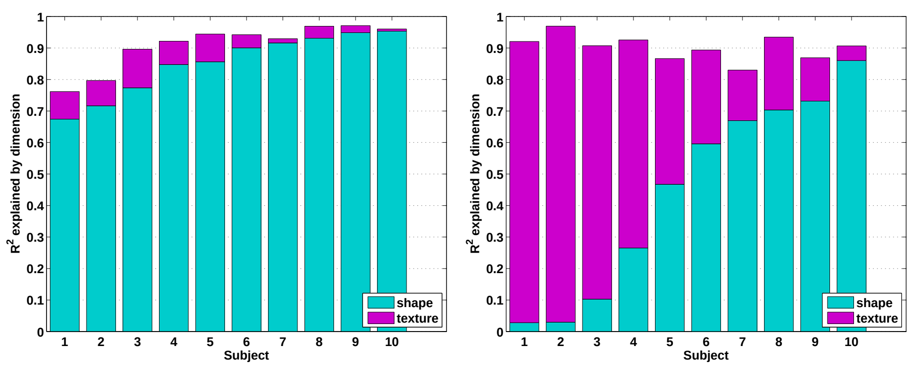

Fig. 5. Dimension weights for subjects in the visual (left) and haptic (right) similarity ratings experiments.

3.2 Haptic similarity ratings

Similarity data In contrast to the visual data, there is no sharp change in similarity visible in the matrix (3, right). Rather, similarity decreases relatively smoothly with shape change (e.g., stimulus 1 compared to 6, 11, 16, and 21). Similarity also varies smoothly as a function of texture change, even when two very different shapes are compared (e.g., object 1 compared to 21-25).

MDS analysis When subjects provided similarity ratings after touching the stimuli, MDS stress computed over mean similarity data was 0.37 for a one-dimensional solution, indicating that a single dimension is *insufficient* to explain the data. Stress dropped to 0.1 for a two-dimensional solution, indicating two dimensions are sufficient. Plotting the output

Similarity Measure	1D Stress	2D Stress
2D Subtraction	0.09	0.03
2D Edge Detection	**0.34**	**0.21**
2D Gabor Jet	0.17	0.10
2D VDP	0.12	0.03
2D SSIM	0.13	0.04
3D Subtraction	0.10	0.05
3D Shape Histogram	0.10	0.04
3D Perimeter	0	0
3D Curvature	0	0

Table I. Stress for features computed on finer scale object data for 1D and 2D MDS solutions. Values > 0.2 (bolded) indicate that the dimensionality of the output configuration is insufficient to explain similarity data.

Similarity Measure	1D Stress	2D Stress
2D Subtraction	0.09	0.03
2D Edge Detection	0.15	0.08
2D Gabor Jet	0.11	0.04
2D VDP	0.07	0.04
2D SSIM	0.10	0.03
3D Subtraction	0.10	0.04
3D Shape Histogram	0.08	0.05
3D Perimeter	0	0
3D Curvature	0	0

Table II. Stress for features computed on *coarser* scale object data for 1D and 2D MDS solutions.

stimulus configuration (Figure 4, right) enabled us to interpret these perceptual dimensions as texture and shape. Ordinal relationships in the stimulus set were recovered along both dimensions - a remarkable feat given the complexity of the haptic measurement space.

On average, shape and texture played equal roles in haptic similarity judgments. The mean shape weight across subjects was 0.45 (std. err. = 0.1) and the mean texture weight was also 0.45 (std. err. = 0.1). Using a two-tailed t-test for independent samples with equal variances, the mean shape weight was not significantly different from the mean texture weight ($t(18)=-0.1$, $p=0.94$). This agrees with the fact that *all* subjects in this experiment mentioned *both* shape-related and texture-related properties when explaining how they had made their similarity judgments. There was, however, a large amount of variation in the way individual subjects weighted shape and texture, from texture dominance, to rough equality between shape and texture, to shape dominance. This finding makes it particularly interesting to try fitting haptic and computational stimulus maps to ascertain whether the variation can be explained by specific computational mechanisms.

3.3 Computational similarity measures

In this section, we present the results of the MDS analysis on the similarities computed using the nine computational measures on the two sets of object data (fine and coarse). The data are comprised of stress values for one and two dimensional solutions (Tables I and II), plots of the two dimensional object maps (Figures 6 and 7), and relative shape/texture weights (Figure 8).

Results using higher-resolution object data For all measures except 2D edge detection,

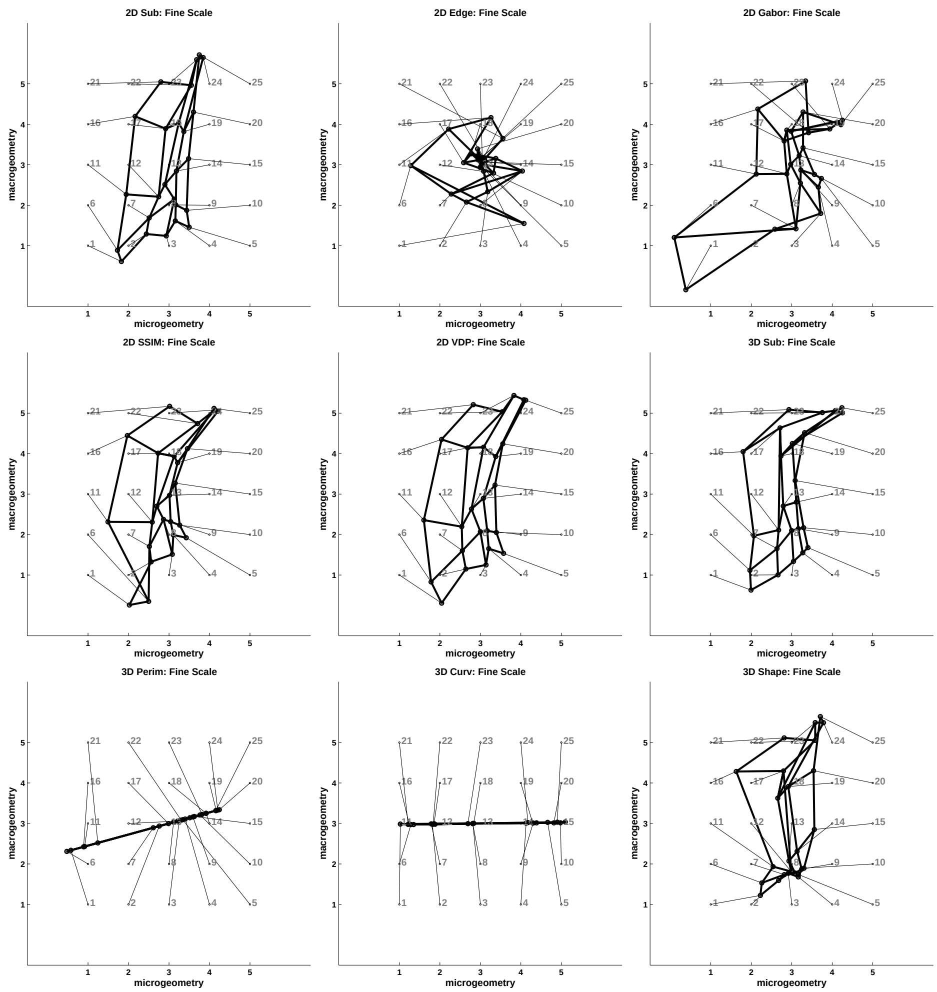

Fig. 6. Stimulus maps derived by MDS analysis of computed similarity data using *finer* resolution object data

MDS stress fell below the threshold of 0.2 for a one-dimensional solution, implying that one dimension sufficed to explain similarity data derived using these measures (Table I). For 2D subtraction, 3D subtraction, VDP, SSIM, and 3D shape histograms, this one dimension corresponded to shape. The dominance of shape over texture for these measures can also be seen from the RSQ weights (Figure 8). For the remainder of the discussion, these measures will be referred to as "shape-dominated measures." For 3D curvature and 3D perimeter, the single dimension required corresponded to texture. In the case of 2D edge detection, two dimensions were required (although texture was weighted much more strongly than shape). One-dimensional stress for the Gabor jet was 0.17 and although shape was technically sufficient to explain computed similarities using Gabor jets, the MDS map shows that the measure is indeed sensitive to texture changes, especially for the bumpiest objects (Figure 6).

MDS maps for the shape-dominated and Gabor jet measures exhibit a larger difference

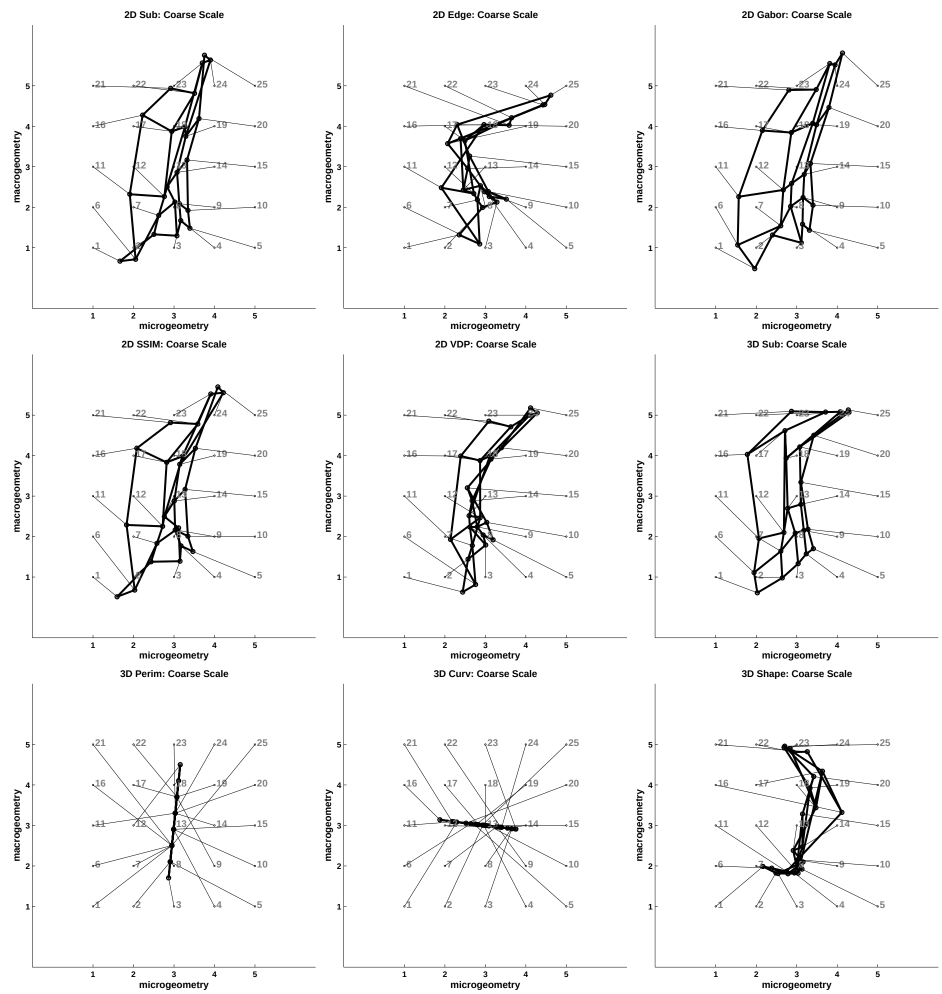

Fig. 7. Stimulus maps derived by MDS analysis of computed similarity data using *coarser* resolution object data

between the bottom three and top two rows of stimuli than, a gap which was also observed in human visual maps. Although they are shape-dominated, these measures are also sensitive to texture differences amongst the objects. However, relative to human maps, they are more sensitive to differences between the bumpiest objects and less sensitive to differences between smooth objects.

The 2D and 3D subtraction maps are quite similar to one another. This is explained by the fact that the 2D images used in this experiment capture most of the variation in the 3D objects. The shape manipulation affects sharp angles in the macrogeometry (such as tips and joints), most of which are visible from the selected 2D view. Had we taken photographs of our stimuli from a viewpoint in which some of these parts were occluded, the maps generated by 2D and 3D subtraction would have differed more. Although texture changes occur over the whole object and are therefore not limited to the frontal view, they have a smaller net effect on 3D vertex positions and pixel values. This also explains the

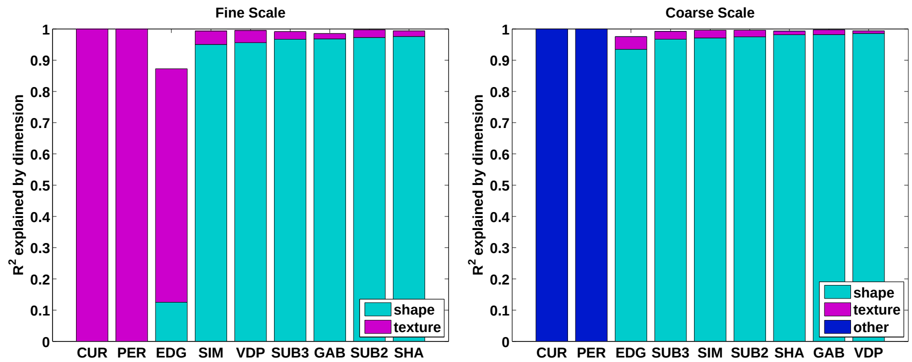

Fig. 8. Dimension weightings based on RSQ for features computed at a finer scale (left) and a coarser scale (right). (CUR: 3D curvature estimate; EDG: 2D edge detection; GAB: 2D Gabor jets; SUB3: 3D subtraction; PER: 3D perimeter; SHA: 3D shape histograms; SIM: 2D Structural Similarity; SUB2: 2D subtraction; VDP: 2D Visual Difference Predictor)

absence of strong texture-related modulation in maps based on global 2D/3D differences.

In contrast, the maps derived from perimeter and curvature show that these measures are not sensitive to changes in object shape for the finer dataset; both features are solely responsive to variations in object texture. The perimeter measure is particularly sensitive to the differences between the most highly-textured objects and the rest. The curvature measure yields a map with more regular spacing between texture levels.

Results using lower-resolution object data Using lower resolution object data had a large effect on curvature, perimeter, and edge measures, as can been seen from the maps (Figure 7). At the coarser scale, the perimeter and curvature measures are no longer able to recover texture variation in the stimuli; although their similarity data can be explained using a single dimension (Figure 8), it is not clear how this dimension can be interpreted. The edge measure responds more regularly to shape when computed on the downsampled data, though the map is still quite noisy. Lowering resolution also had a noticeable effect on the Gabor jet map, whose hypersensitivity to the highest texture level was reduced. On the other hand, lowering resolution had lesser effects on the shape-dominated measures. The VDP and shape histogram measures lost some of their ability to separate texture levels. No major effects were observed for 2D subtraction, 3D subtraction, or SSIM.

3.4 Perceptual validation of computational measures

In order to assess the perceptual validity of a computational measure, we verified how well the stimulus configuration generated from the measure compared to the configuration generated from human similarity ratings. This was done by fitting the computational maps to the individual visual/haptic human maps, using the fit error as the goodness-of-fit measure, as described in section 2.6.

The fit error enables us to make a *relative* assessment of goodness-of-fit, i.e., we can say that the fit obtained with one measure is better or worse than another; however, it does not provide an *absolute* criterion. To determine such a criterion, we reasoned that a given measure can be deemed to fit the human data well and thus “perceptually valid” to a certain extent (see General Discussion) *if the mean error in fitting the computational map to all individual maps is statistically equivalent to the mean error obtained by fitting each* *individual subject map to all other individual maps.* We refer to the procedure of fitting each individual map to all the other individual maps as "cross-fitting" the individual data and refer to the resulting error as the "cross-fitting error." To test whether a measure met our criterion, we performed a two-tailed t-test between the cross-fitting errors and the fit errors generated by each measure (5% confidence level, assuming independent samples and equal variances).

Cross-fitting results Fitting maps derived from individual visual similarity data led to a cross-fitting error of 24% with standard error of 2%. With individual haptic data, we obtained a mean cross-fitting error of 19% with standard error of 2%. Lower cross-fitting error in the haptic condition was related to the fact that maps recovered from individual haptic subject data were more regular than those recovered from individual visual subject data (7/10 haptic maps had 3 or fewer violations of ordinal relationships compared to only 2/10 visual maps). This likely stemmed from differences in experimental settings across conditions; notably, visual stimuli were presented on a computer screen in a darkened room, whereas haptic stimuli were presented by an experimenter in a naturally-lit room. In a follow-up study in which experimental conditions were equated for visual and haptic similarity judgments [Cooke et al. 2006], ordinal relationships were recovered equally well using either modality.

Fit between computational measures and human visual perception Several measures met our criterion for perceptual validity: the 2D measures of image subtraction, VDP, and SSIM, as well as the 3D measures of mesh subtraction and shape histograms provided mean fit errors which were not significantly different from mean visual cross-fitting error (all p's $>$ 0.4) when computed using finer resolution object data (Figure 9). Gabor jets provided a slightly worse fit. The 3D curvature, edge and perimeter measures provided poor fits to visual data, regardless of the scale of object data. For coarse resolution data, the loss of sensitivity to texture differences resulted in slightly worse fits for VDP and shape histograms, and a noticeable improvement in the fit for the Gabor jet measure. The fit also improved for the edge measure, which responded more regularly to shape when computed at a coarser scale. Although we have used mean fit error over all subjects to define our criterion for perceptual validity, it is also informative to consider how each measure fits individual subjects (Figures 9 and 10, right). For the visual subjects, we see a consistent pattern of fits across subjects, with shape-dominated measures consistently outperforming other measures.

Fit between computational measures and human haptic perception All of the computational measures we implemented yielded fit errors which differed significantly from the mean haptic cross-fitting error, i.e., none of the features tested met our criterion for perceptual validity relative to the haptic modality for either data set (Figures 9 and 10, left). However, good fits were found in individual cases. When curvature was computed using higher resolution data, it provided good fits to subjects 1 and 2, who were the most texture-dominated. The VDP and 2D subtraction measures provided good fits to subject 10, who was the most shape-dominated of the haptic subjects. The Gabor jet and subtraction measures also fit this subject well for the coarse data set. The worst fits were obtained for subjects for whom *both* shape and texture were important. One reason for this is that all the measures we tested are either shape or texture dominated. Thus one way of modelling our data would be to use a combination of these measures. On the other hand, a more sophisticated 3D measure which we have not included may be capable of modelling our

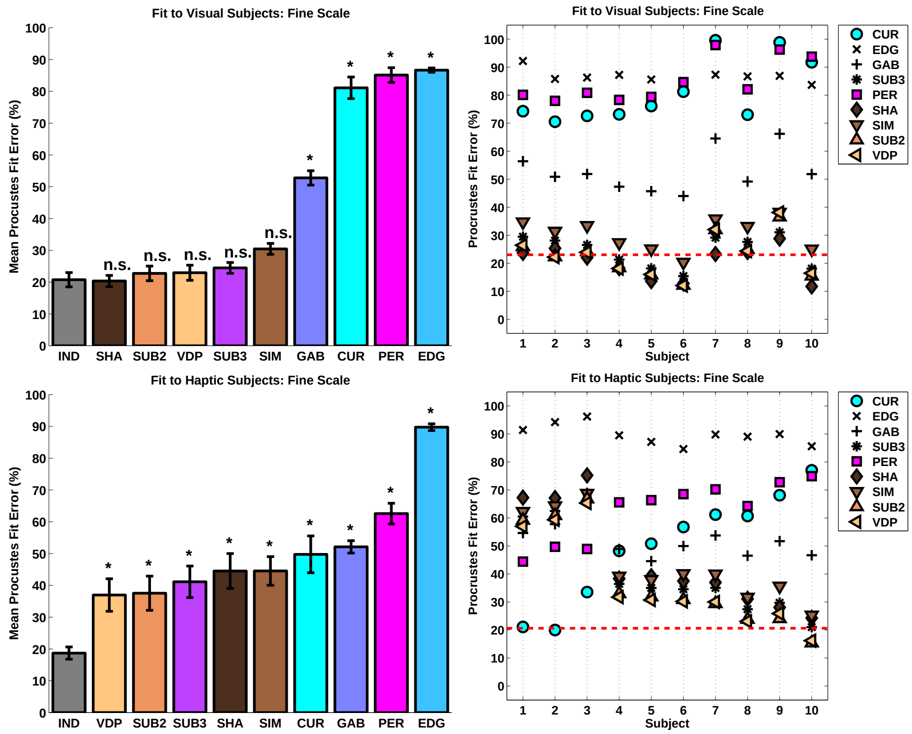

Fig. 9. Fits between computational measures and human maps when features are computed using *finer* resolution object data. Mean fit over all subjects (left) and fits to individual subjects (left). Fits relative to visual data (top) and haptic data (bottom). Error bars represent standard error. * = significant difference compared to mean cross-fitting error (IND); n.s. = not significant. In the right-hand figures, the dashed red line is drawn at one standard error away from the mean cross-fitting error. (CUR: 3D curvature estimate; EDG: 2D edge detection; GAB: 2D Gabor jets; SUB3: 3D subtraction; PER: 3D perimeter; SHA: 3D shape histograms; SIM: 2D Structural Similarity; SUB2: 2D subtraction; VDP: 2D Visual Difference Predictor)

data; further 3D features need to be tested in order to investigate this possibility.

4. GENERAL DISCUSSION

4.1 Effect of modality on perceptual similarity

Visual similarity In visual similarity judgments, shape was the dominant perceptual dimension, whereas texture variation played a lesser role. This finding agrees with the idea that shape is a key determinant of similarity relationships between objects [Edelman 1999]. It is also consistent with the notion that the extraction of global form is one of the visual system's areas of expertise [Klatzky and Lederman 2003]. Distinct clusters of stimuli based on shape appeared in the visual similarity map, hinting at the formation of shape-based categories in similarity space. This observation is interesting given debate surrounding the question of whether similarity relationships form the basis for perceptual categorization [Hahn and Ramscar 2001] and coincides with evidence for a special role of shape in the formation of category structure. For instance, young children have been shown to use shape as a basis for naming generalization, ignoring other properties such as size and tex-

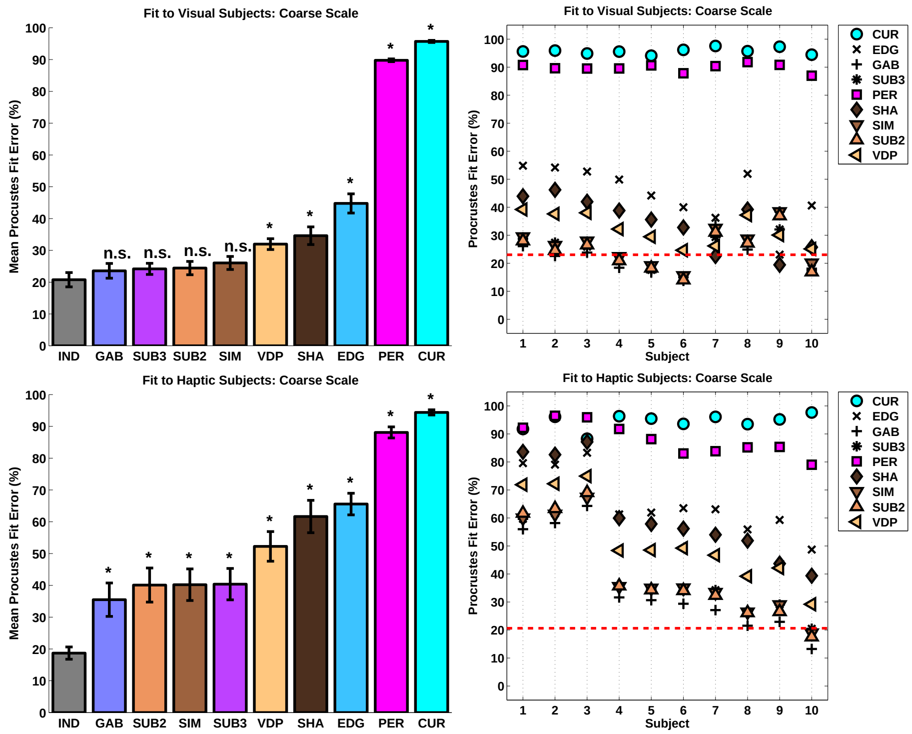

Fig. 10. Fits between computational measures and human maps when features are computed using *coarser* resolution object data. Mean fit over all subjects (right) and fits to individual subjects (left). Fits relative to visual data (top) and haptic data (bottom). Error bars represent standard error. * = significant difference compared to mean cross-fitting error (IND); n.s. = not significant. In the right-hand figures, the dashed red line is drawn at one standard error away from the mean cross-fitting error. (CUR: 3D curvature estimate; EDG: 2D edge detection; GAB: 2D Gabor jets; SUB3: 3D subtraction; PER: 3D perimeter; SHA: 3D shape histograms; SIM: 2D Structural Similarity; SUB2: 2D subtraction; VDP: 2D Visual Difference Predictor)

ture [Landau et al. 1998]. Models of visual object categorization have also been developed on the basis of shape primitives [Biederman 1987] as well as on the basis of similarity relationships amongst shape primitives [Edelman 1999]. An interesting question is whether shape also plays a special role in category formation when objects are first experienced through touch. We have recently found evidence that *both* shape and texture dimensions are capable of determining spontaneous category structure in vision as well as in touch [Cooke et al. 2006]. In addition, we found that the relative importance of shape/texture in determining category structure was the same as the relative weight in judging similarities (shape was more important than texture for vision, while for haptics, shape and texture were roughly equally important).

Haptic similarity When subjects touched the objects, they weighted the relative importance of shape and texture differently than when they saw the objects: instead of shape dominating their similarity judgments, *both* shape and texture were important. Given that local material properties are thought to be more accessible to the haptic system than global geometric properties [Klatzky and Lederman 2003], one might have expected haptic similarity ratings to be more strongly affected by texture differences. [Klatzky et al. 1987] found that haptic free sorting of wafer shapes based on similarity was performed preferentially on the basis of material properties as opposed to geometrical properties. However, in a follow-up study [Lederman et al. 1996], in which stimuli were fully 3D and shape variation was no longer limited to the edges, geometric properties played a *more* important role than material properties in haptic similarity judgments. This result indicates that the distribution of geometrical features (e.g., 2D vs. 3D shape information) has an influence on the relative weights of object properties. In our stimulus set, shape information is 3D, but most variation in shape features can be captured in the frontal 2D plane [Cooke et al. 2005]. In light of the two aforementioned studies, it could be that this “2 1/2 D” distribution of shape features contributed to the even weighting of shape and texture properties in this experiment. Experiments involving stimuli with controlled variations in the distribution of shape features would be required to test this hypothesis.

Another difference between our study and [Lederman et al. 1996] is that subjects freely explored the objects and used a variety of different hand movements, whereas subjects in this experiment were restricted to contour-following. Although contour-following allows for the extraction of both shape and texture properties, it is thought to be optimal for extracting precise shape information. We are currently investigating how controlled variation in exploratory procedure affects relative property weightings for this stimulus set.

Differences and commonalities in haptic and visual similarities We found a larger amount of individual variation in the dimension weights derived from haptic similarity data than in those derived from visual similarity data. This may have arisen due to differences in exploration time: in the visual condition, viewing time was controlled by a software program and was kept constant (500ms) for all subjects, whereas in the haptic condition, subjects were allowed to explore the object for up to 10s. Actual exploration time varied from individual to individual and was as short as 3s per object. [Lakatos and Marks 1999] found that that local and global geometrical properties had comparable effects on haptic similarity ratings for short exploration times (0.5s to 4s), but that the role of global properties increased with exploration time. This suggests that subjects who took longer to explore the objects may have been more shape-biased than those who used shorter exploration times. However, in a follow-up study [Cooke et al. 2006], in which exploration times were kept between 3 and 5s, we found the same pattern of dimension weights: visual subjects were consistently shape-dominated, while haptic subjects exhibited variable weights, with shape and texture being equally weighted on average. One remaining explanation is that since subjects have less experience with haptic similarity ratings, they tend to invoke cognitive strategies and rules more often when access is provided via touch instead of vision, which in turn leads to greater variability in dimension weights.

Despite the differences we found between visually and haptically-derived representations, there were also important commonalities. In both visual and haptic conditions, subjects were able to extract the two kinds of parametric variation which were used to create the stimuli. These two stimulus variations, which we initially referred to as changes in “macrogeometry” and “microgeometry” were consistently referred to by subjects as changes in “shape” and “texture.” The fact that subjects were able to extract systematic variation along these two dimensions is a non-trivial ability given the high-dimensionality of the visual and haptic measurement spaces. For instance, assuming gaze fixation, the visual measurement space might be approximated by the number of pixels in the images of the stimuli. The haptic measurement space might be approximated by the 3D forces exerted on the finger plus the relevant joint angles and positions, taken over the course of the contour-following procedure. Furthermore, the two stimulus manipulations may have had non-linear effects on measurements of object data. In spite of this, maps derived from mean human similarity data exhibit clear, regular responses to the two stimulus manipulations and ordinal relationships present in the stimulus set are recovered. Understanding how the visual and haptic systems deliver such close results despite large differences in the anatomical structure of receptors and pathways which convey object information to the brain is a key motivation for our work.

4.2 Similarity-based feature validation

In this paper, we have proposed a criterion for feature validity based on the fit error between maps based on the fit between computational and human similarity measures. However, as discussed above, the results of our human experiments indicate that perceptual similarity between objects can vary as a function of the modality used to experience the objects. Therefore, feature validity, when based on perceptual similarity, depends on the modality assumed to be used for perception. This underscores the importance of specifying the modality (or combination of modalities) to be used when evaluating feature validity.

Visual feature validation Several of the features we implemented met our criterion for perceptual validity relative to the visual modality (2D/3D subtraction, 3D shape histograms, VDP and SSIM). These results are in accordance with those reported in [Watson et al. 2001]. The strong performance of the VDP and SSIM, which is are industry standards for assessing image differences, is to be expected. In this sense, they can also be considered benchmarks against which the performance of the other measures can be compared. Surprisingly, the much simpler subtraction-based measures yielded comparable stimulus maps and fit errors. One apparent difference between the shape-dominated measures and the human visual data lies in their response to texture changes: the human data (Figure 4) do not exhibit the same hypersensitivity to high texture levels observed in some of the computational maps (Figures 6 and 7).

The fact that the 3D subtraction map met our criterion for perceptual validity could be taken as an indication that the human visual system reconstructs 3D geometry from 2D images; however, as pointed out earlier, 2D and 3D subtraction measures likely yielded similar results on our stimuli since most of the variation among stimuli occurs in the image plane. A stronger test of whether 3D measures are indeed perceptually valid for the visual modality would require the use of a stimulus set in which variation occurs in depth.

Perimeter and curvature measures provided poorer fits to human visual maps. This is mainly due to the insensitivity of these measures to changes in shape. This result shows that the visual system does not rely solely on curvature or perimeter estimates (at least not as we have implemented them) to judge similarities. This is not as trivial as it may seem: it is indeed possible to extract object perimeter from 2D images and, since perimeter can also be extracted in the haptic modality, it could serve as a convenient feature for sharing information between vision and touch. Curvature can also be extracted from 2D images; in fact, the visual system could use shading-related changes in pixel intensities to estimate both local curvature (texture-from-shading [Todd et al. 2004]) and global curvature (shape-from-shading [Blake and Bülthoff 1991]). Our findings do not rule out the possibility that the visual system uses these features, but they indicate that neither perimeter nor curvature (as we computed them) is sufficient to explain our human visual similarity data.

Haptic feature validation None of the measures we tested met our criterion for perceptual validity relative to the haptic modality. This is due in part to the strong shape or texture bias of the measures we implemented, which meant that good fits were not obtained for subjects with intermediate shape/texture weightings. Although identifying such a feature would help to address this problem, one would still need to account for intra-subject variability in haptic similarity judgments. For this, it may be necessary to implement an individually-adjusted combination of features.

At the individual subject level, good fits were obtained for subjects who were strongly biased either towards shape or texture. A surprising finding was that despite the fact that subjects explored the objects via a contour-following procedure, the map based on the perimeter measure did not yield good fit values for either of the two scales we tested. Possible explanations for this could be that subjects do not compute perimeter during contour following, or that it is computed but not used to judge similarity, e.g., because the estimate is not statistically reliable [Ernst and Banks 2002].

Finally, contrary to our expectations, we did not find that measures computed on 3D data provided generally better fits to haptic data than measures computed on 2D data. Studies involving objects with greater 3D variation are needed to further test whether 2D features are indeed sufficient to model haptic object representations.

4.3 Variation of object scale

In this study, we computed features using object data at two different scales and looked for differences in fits to human data caused by this variation. For our object set, the main effect of presenting coarser data was to make it more difficult for measures to recover texture variation. This had little effect on fits to human visual data, since the recovery of shape information is the critical factor in determining a good fit. For edge detection, downsampling the data led to a better recovery of shape variation and fit error was lower in the coarse than in the fine condition. Texture-dominated measures (curvature and perimeter) were strongly affected by downsampling; the measures no longer captured texture variation in the stimuli and, as a result, fits to texture-dominated haptic subjects worsened. The fact that our curvature measure was strongly affected by downsampling object data is interesting in light of findings that human haptic curvature estimation is also scale-dependent [Louw et al. 2000; Nefs and Kappers 2003]. Further studies are needed to systematically investigate how the scale of object data affects *perceptual* similarity judgments and compare this to the sensitivity of computed measures. In turn, this type of knowledge can help to design effective haptic, visual, and multimodal interfaces. More generally, we have demonstrated how our method can be used to assess the sensitivity of features to computational parameters (in this case, data resolution). We envisage that this procedure could be performed iteratively in an optimization setting, e.g., to find the resolution of object data for which feature X provides the best fit to human data.

4.4 Summary of findings and outlook

In this work, we have shown how a similarity-based approach can be used to assess the perceptual validity of features relative to a specific sensory modality. With the example of scale variation, we also showed how the method can be used to assess the sensitivity of features to changes in the way they are computed. This work is just a first step, however, towards developing a rigorous test of perceptual validity. In particular, our stimulus set only contained 25 objects, meaning that the fits between computational and perceptual maps were based on relatively few data points. Another limitation is that we limited our study to a single stimulus class. That having been said, we were able to provide a "plausibility test" for perceptual validity for a set of standard 2D and 3D features. We found a number of perceptually plausible features for the visual condition, with the critical factor being the features' ability to recover shape variation in the stimulus set. The features we tested did not provide good overall fits to the haptic data, although good fits were found for individuals who were particularly biased toward either shape or texture variation in the stimuli. A more sophisticated 3D feature or an individually-adjusted combination of features may be required to model our haptic data.

Although similarity-based methods have already been applied to compare perception in different modalities (e.g., [Garbin 1988]), our *combination* of similarity measures and *parametrically-related* stimuli differentiates our approach and allows us to compare how different computations or modalities recover high-level, topological relationships in the stimulus set. In addition to the rich qualitative information contained in the MDS maps, the method provides two important quantitative metrics: 1) weightings of the dimensions which span the output space generated by a given modality or computational measure and 2) a goodness-of-fit measure between two stimulus configurations in the output space.

We suggest that the method can be used for three distinct purposes:

(1) First, it can be used to visualize and quantify changes in human similarity-based representations of objects when different or multiple modalities are used to explore objects. The effects of varying, adding, or removing stimulus properties as well as the effects of stimulus-independent manipulations (e.g., changes in viewpoint or illumination) on the structure of human object representations can be studied. The method also makes it possible to identify object properties which are important for unimodal perception, but are not good predictors of multimodal perception.

(2) Second, the method can be used to visualize and quantify how well a computed object feature is able generate human-like stimulus maps and to compare the relative performance of different features on a given set of object data.

(3) Third, the method makes it possible to assess the sensitivity of features to changes in inputs (e.g., resolution) or algorithm parameters. The method offers a criterion which can used to optimize such parameters relative to human perception.

The next step in our work is to develop more rigorous tests of perceptual validity, which can be added as a second stage once perceptual plausibility has been established using the method presented in this paper (Figure 11). This second stage of perceptual validity testing has two important components: 1) a test of generalization to a larger number of more complex stimulus classes (left-hand boxes) and 2) a test of generalization to a larger number of points within each stimulus space (right-hand boxes). A rigorous evaluation of perceptual validity which incorporates these components should provide substantial benefit for developing efficient artificial representations of objects and also help to elucidate the computational mechanisms underlying human perception.

5. ACKNOWLEDGMENTS

The authors wish to thank the following persons for their help with the work presented here, a part of which was also presented in [Cooke et al. 2005]: Rafal Mantiuk and Karol Myszkowski for graciously providing code to implement the VDP, Florian Steinke for providing code to compute 3D curvature, Martin Breidt and Michael Renner for assistance with stimulus production, and Karin Bierig for helping to run the haptic experiments.

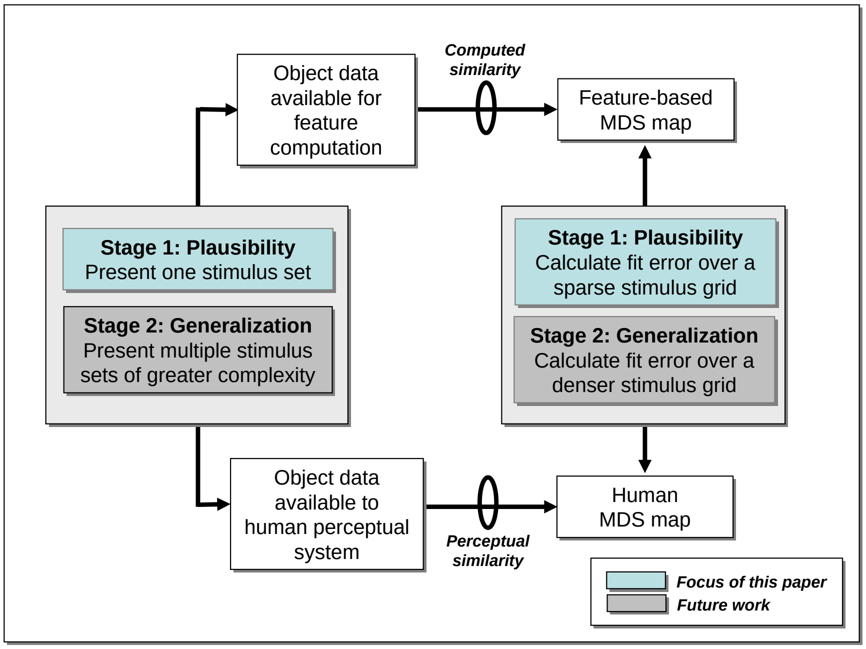

Fig. 11. A two-stage model for perceptual feature validation. The first stage involves a less demanding, but relatively straightforward plausibility test. In the second stage, perceptual validity is tested more rigorously by testing whether goodness-of-fit holds for a larger number of points in the stimulus space and to more complex stimulus classes.

REFERENCES

ACOSTA, E., TEMKIN, B., GRISWOLD, J., DEEB, S., KRUMMEL, T., HALUCK, R., AND KAVOUSSI, L. 2002. Heuristic haptic texture for surgical simulations. *Stud Health Technol Inform 85*, 14–16.

ANKERST, M., KASTENMÜLLER, G., KRIEGEL, H., AND SEIDL, T. 1999. 3D shape histograms for similarity search and classification in spatial databases. In *Advances in Spatial Databases, 6th International Symposium, SSD'99*, R. Güting, D. Papadias, and F. Lochovsky, Eds. Vol. 1651. Springer, Hong Kong, China, 207–228.

BIEDERMAN, I. 1987. Recognition-by-components: a theory of human image understanding. *Psychol Rev 94*, 2 (Apr), 115–147.

BLAKE, A. AND BÜLTHOFF, H. 1991. Shape from specularities: Computation and psychophysics. *Philosophical Transactions of the Royal Society (London) Series B 331*, 237–252.

BORG, I. AND GROENEN, P. 1997. *Modern multidimensional scaling*. Springer.

BRAINARD, D. 1997. The psychophysics toolbox. *Spat Vis 10*, 433–436.

BÜLTHOFF, H. AND EDELMAN, S. 1992. Psychophysical support for a 2-d view interpolation theory of object recognition. *Proceedings of the National Academy of Science 89*, 60–64.

CADIK, M. AND SLAVIK, P. 2004. Evaluation of two principal approaches to objective image quality assessment. In *Proceedings of the 8th International Conference on Information Visualisation*. IEEE Computer Society, Washington, DC, USA, 513–518.

COOKE, T., JÄKEL, F., WALLRAVEN, C., AND BÜELTHOFF, H. 2006. Multimodal similarity and categorization of novel, three-dimensional objects. *Neuropsychologia in press*, –.

COOKE, T., STEINKE, F., WALLRAVEN, C., AND BÜLTHOFF, H. 2005. A similarity-based approach to perceptual feature validation. In *APGV 2005 - Symposium on Applied Perception in Graphics and Visualization*. ACM SIGGRAPH.

COX, T. AND COX, M. 2001. *Multidimensional Scaling*, 2nd ed. Chapman & Hall.

DALY, S. 1993. *Digital images and human vision*. MIT Press, Watson, AB (ed.).

DEIBERT, E., KRAUT, M., KREMEN, S., AND HART, J. J. 1999. Neural pathways in tactile object recognition. *Neurology 52,* 7, 1413–1417.

EDELMAN, S. 1999. *Representation and recognition in vision*. MIT Press.

ERNST, M. AND BANKS, M. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature 415*, 429–433.

FUNKHOUSER, T., MIN, P., KAZHDAN, M., CHEN, J., HALDERMAN, A., DOBKIN, D., AND JACOBS, D. 2003. A search engine for 3d models. *ACM Trans. Graph. 22,* 1, 83–105.

GARBIN, C. 1988. Visual-haptic perceptual nonequivalence for shape information and its impact upon cross-modal performance. *J Exp Psychol Hum Percept Perform 14,* 4, 547–553.

GIESE, M. 2004. *A neural model for biological movement recognition: a neurophysiologically plausible theory*. Kluwer Academic Publishers, Norwell, MA, USA, 443–470.

HAHN, U. AND RAMSCAR, M., Eds. 2001. *Similarity and categorization*. Oxford University Press.

JONES, J. AND PALMER, L. 1987. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol 58,* 6, 1233–1258.

KLATZKY, R. AND LEDERMAN, S. 2003. *Experimental psychology*. Handbook of Psychology, vol. 4. Healy, AF and Proctor, RW and Weiner, IB (eds.), Wiley, Chapter Touch, 147–176.

KLATZKY, R., LEDERMAN, S., AND REED, C. 1987. There's more to touch than meets the eye: the salience of object attributes for haptics with and without vision. *J Exp Psychol Gen 116,* 4, 356–369.

KOENDERINK, J. 1984. The structure of images. *Biol Cybern 50,* 5, 363–370.

LAKATOS, S. AND MARKS, L. 1999. Haptic form perception: relative salience of local and global features. *Percept Psychophys 61,* 5, 895–908.

LANDAU, B., SMITH, L., AND JONES, S. 1998. Object perception and object naming in early development. *Trends Cogn Sci 2,* 1, 19–24.

LEDERMAN, S. AND KLATZKY, R. 1993. Extracting object properties through haptic exploration. *Acta Psychol 84*, 29–40.

LEDERMAN, S., SUMMERS, C., AND KLATZKY, R. 1996. Cognitive salience of haptic object properties: role of modality-encoding bias. *Perception 25,* 8, 983–998.

LINDEBERG, T. 1994. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA.

LOUW, S., KAPPERS, A., AND KOENDERINK, J. 2000. Haptic detection thresholds of Gaussian profiles over the whole range of spatial scales. *Exp Brain Res 132,* 3, 369–374.

MANTIUK, R., DALY, S., MYSZKOWSKI, K., AND SEIDEL, H. 2005. Predicting visible differences in high dynamic range images - model and its calibration. In *Human Vision and Electronic Imaging X, IS&T/SPIE's 17th Annual Symposium on Electronic Imaging*. Vol. 5666. SPIE, San Jose, USA, 204–214.

MARR, D. 1982. *Vision: a computational investigation into the human representation and processing of visual information*. W. H. Freeman, San Francisco.

NEFS, H. AND KAPPERS, AML KOENDERINK, J. 2003. Detection of amplitude modulation and frequency modulation in tactual gratings: a critical bandwidth for active touch. *Perception 32,* 10, 1259–1271.

NESTARES, O., NAVARRO, R., PORTILLA, J., AND TABERNERO, A. 1998. Efficient spatial domain implementation of a multiscale image representation based on gabor functions. *7,* 1, 166–173.

PONT, S., KAPPERS, A., AND KOENDERINK, J. 1999. Similar mechanisms underlie curvature comparison by static and dynamic touch. *Percept Psychophys 61,* 5, 874–894.

REED, C., SHOHAM, S., AND HALGREN, E. 2004. Neural substrates of tactile object recognition: an fMRI study. *Hum Brain Mapp 21,* 4, 236–246.

RIESENHUBER, M. AND POGGIO, T. 1999. Hierarchical models of object recognition in cortex. *Nat Neurosci 2,* 11, 1019–1025.

ROSCH, E., MERVIS, C., GRAY, W., JOHNSON, D., AND BOYES-BRAEM, P. 1976. Basic objects in natural categories. *Cognit Psychol 8,* 382–439.

SALISBURY, K., CONTI, F., AND BARBAGLI, F. 2004. Haptic rendering: Introductory concepts. *IEEE Computer Graphics and Applications 24,* 2, 24–32.

SATHIAN, K., GOODWIN, A., JOHN, K., AND DARIAN-SMITH, I. 1989. Perceived roughness of a grating: correlation with responses of mechanoreceptive afferents innervating the monkey's fingerpad. *J Neurosci 9,* 4, 1273–1279.

SRINIVASAN, M. AND LAMOTTE, R. 1991. *Encoding of shape in the responses of cutaneous mechanoreceptors*. Wenner-Gren International Symposium Series. MacMillan Press, Chapter 5, 59–61.

STEINKE, F., SCHÖLKOPF, B., AND BLANZ, V. 2005. Support vector machines for 3D shape processing. In *Computer Graphics Forum (Proceedings of EUROGRAPHICS 2005)*. Vol. 24, 3.

TODD, J., OOMES, A., KOENDERINK, J., AND KAPPERS, A. 2004. The perception of doubly curved surfaces from anisotropic textures. *Psychol Sci 15,* 1, 40–46.

ULLMAN, S. 1996. *High-level Vision*. MIT Press.

ULLMAN, S., VIDAL-NAQUET, M., AND SALI, E. 2002. Visual features of intermediate complexity and their use in classification. *Nat Neurosci 5,* 7, 682–687.

WANG, Z., BOVIK, A., SHEIKH, H., AND SIMONCELLI, E. 2004. Image quality assessment: From error measurement to structural similarity. *IEEE Trans. Image Processing 13*, 600–612.

WATSON, B., FRIEDMAN, A., AND MCGAFFEY, A. 2001. Measuring and predicting visual fidelity. In *SIGGRAPH*. ACM Press, 213–220.