LETTER Communicated by Paul Viola

Inferring Perceptual Saliency Fields from Viewpoint-Dependent Recognition Data

Florin Cutzu
Michael Tarr
Department of Cognitive and Linguistic Sciences, Brown University,
Providence, RI 02912, U.S.A.

We present an algorithm for computing the relative perceptual saliencies of the features of a three-dimensional object using either goodness-of-view scores measured at several viewpoints or perceptual similarities among several object views. This technique addresses the inverse, ill-posed version of the direct problem of predicting goodness-of-view scores or viewpoint similarities when the object features are known. On the basis of a linear model for the direct problem, we solve the inverse problem using the method of regularization. The critical assumption we make to regularize the solution is that perceptual salience varies slowly on the surface of the object. The salient regions derived using this assumption empirically indicate what object structures are important in human three-dimensional object perception, a domain where theories typically have been based on somewhat ad hoc features.

1 Direct and Inverse Problems in Object Recognition and Similarity Modeling

The problem of how humans mentally represent and recognize objects is often taken as a problem of finding the right features. That is, faced with highly variable viewing conditions under which a familiar object may appear in different orientations, illuminations, or configurations, vision scientists have sought features that remain stable over changes in the image. Thus, extant models of object recognition typically have begun by defining putatively stable features. On this basis they can then predict human recognition performance as a function of viewing parameters (most often viewpoint). As a specific example of what we term the *direct problem* of object recognition, consider Biederman's (1987) recognition-by-components theory. This recognition model specifies what object parts (features) are important ("geons") for recognition and derives view recognizability on that basis (Biederman & Gerhardstein, 1993).

In contrast to approaches where predefined sets of features are used as shape grammars or building blocks, several researchers have posited that the particular features used for recognition are a function of the recognition task and training experience (Edelman, 1995; Schyns, Goldstone, & Thibaut, 1998; Tarr & Bülthoff, 1995). In this article we expand on this view by addressing the inverse of the direct problem of recognition. Specifically, we ask whether it is possible to deduce the features of recognition by examining how recognition performance varies with changes in viewing parameters. More precisely, given the geometry of an object and the recognition performance (goodness-of-view scores, recognition times, error rates) for several views of that object we will develop a method for identifying the features of recognition and their salience. This approach is appealing because it relies on judgments that are both accessible to human observers and stable across observers: goodness of view (Palmer, Rosch, & Chase, 1981) and perceptual similarity (Cutzu & Edelman, 1998). In contrast, judgments about what features are used in object perception are not necessarily consciously accessible and are notoriously unstable across observers.

In spirit, our approach is closely related to recent attempts to model the perceptual similarity of objects for purposes of recognition (Cutzu & Edelman, 1998). Visual similarity can be defined for different views of the same object or for different objects from the same category (Cutzu & Tarr, 1997). Most models of similarity (see the collection of papers in Ashby, 1992) treat this as a direct problem in that they attempt to predict similarity given an assumed feature set. In contrast, our goal is to deduce the features of similarity from perceptual similarity data given that we know only the geometry of the object (and nothing about the feature set)—that is, the *inverse problem*.

2 Modeling Goodness-of-View and View Similarity

2.1 View Recognizability. Our basic hypothesis is that view recognizability/goodness-of-view depends on two factors: surface salience, which characterizes the object in a given perceptual task, and surface visibility, which is dependent on the viewpoint of the observer relative to the object.

2.1.1 Role of Surface Salience. The goodness, or recognizability, of a view depends on which of the object features appear in the image. Goodness-of-view measurement experiments have established that certain regions of the object's surface are perceptually more important (more salient) than others (Palmer et al., 1981). The reasons for variations in salience include the diagnosticity of particular features for a given discrimination task, the functional role of particular object parts, and the stability of particular features over transformations.

To express this idea quantitatively, a salience density field ρ was defined by associating a positive number $\rho(x, y)$ with each point (x, y) of the surface of a three-dimensional object. The more important perceptually the elementary surface patch located at (x, y) is, the higher its salience density. Salience density is assumed to depend on both subjective and objective factors, such as biases, experimental task, and object geometry. We emphasize that ρ does not depend on viewpoint.

We required that $\rho(x, y)$ be continuous almost everywhere and bounded, thus integrable. The salience of a region S of the object surface is, by definition, the integral of the salience density field over the region:

$$p(S) = \int_S \rho ds. \tag{2.1}$$

We assumed that the rate of change of $\rho(x, y)$ across the surface of the object is slow in comparison to the rate of change of surface shape, that is, variation of the surface normal. This very common "smoothness" condition is physically intuitive and computationally convenient but does not allow for localized surface features with sharp boundaries. The solution is to impose only piecewise smoothness on the salience field, as discussed in section 2.1.5.

2.1.2 Role of Surface Visibility. The second factor that influences goodness-of-view is the degree of visibility for object surfaces appearing in the image. This, in turn, depends on viewpoint and object geometry.

Let θ and ϕ denote, respectively, camera latitude and longitude on the viewing sphere surrounding the object. In view (θ, ϕ) the visibility of the elementary surface patch located at (x, y), denoted by $a(\theta, \phi, x, y)$, is defined as the cosine of the angle $\psi(\theta, \phi, x, y)$ between the normal $\vec{N}(x, y)$ to the patch and the viewing direction $\vec{V}(\theta, \phi)$. The visibility of an occluded patch is zero. Therefore:

$$a(\theta, \phi, x, y) = \begin{cases} \cos(\psi) & \text{if } \cos(\psi) > 0, \text{ no self-occlusion.} \\ 0 & \text{if } \cos(\psi) \leq 0 \text{ or self-occlusion.} \end{cases} \tag{2.2}$$

In practice $a(\theta, \phi, x, y)$ is determined by using a hidden surface removal algorithm.

2.1.3 Joint Salience-Visibility Model Formulation. View recognizability or goodness-of-view depends on both the salience of the visible object features and their relative degree of visibility. Assuming a linear model for this joint dependence, the goodness-of-view score r for viewpoint (θ, ϕ) can be expressed in the following form (see Figure 1):

$$\int_x \int_y \rho(x, y) a(\theta, \phi, x, y)\, dx\, dy = r(\theta, \phi). \tag{2.3}$$

Since the salience density function, $\rho(x, y)$, is assumed continuous and bounded over the surface of the body, the domain of integration depends on

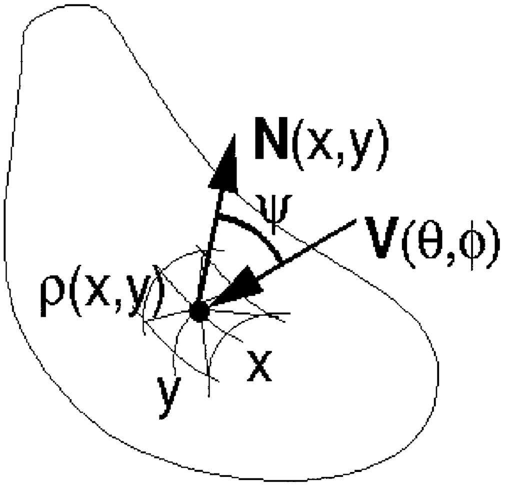

Figure 1: The normal to the surface of a three-dimensional object at the point (x, y) is denoted **N** and the salience density is $\rho(x, y)$. The viewing direction is denoted **V**.

the domain of definition of the visibility function $a(\theta, \phi, x, y)$. For a smooth object, the integral is taken over the whole surface of the body; for a piecewise smooth object (the object surface has ridges), one sums up integrals of the form in equation 2.3 taken over the domains of smoothness of the surface.

In practice, $r(\theta, \phi)$ is measured at a finite number of views (θ_i, ϕ_i), while the salience density $\rho(x, y)$, defined over the entire surface of the object, is unknown and must be calculated. Equation 2.3 represents a Fredholm linear integral equation of the first kind in ρ. The approach to the solution of this equation is presented next.

2.1.4 The Discrete Formulation. The integral equation, 2.3, has no analytic solution. Therefore, a quadrature based on the discretization of the object's surface was chosen as solution strategy. To this end, the surface of the body was approximated by a fine triangular mesh, as illustrated in Figure 2. For a sufficiently fine mesh, a point on the body corresponds uniquely to a point on the mesh surface, and therefore the unknown ρ can now be defined on the mesh.

To reduce the number of degrees of freedom of the problem, following the fundamental idea of the finite element method (Schwarz, 1988), the unknown function $\rho(x, y)$ is assumed to vary linearly within the triangular elements. Therefore, the value of the discretized function ρ is fully defined by its values at the nodes of the mesh. This piecewise linear approximation

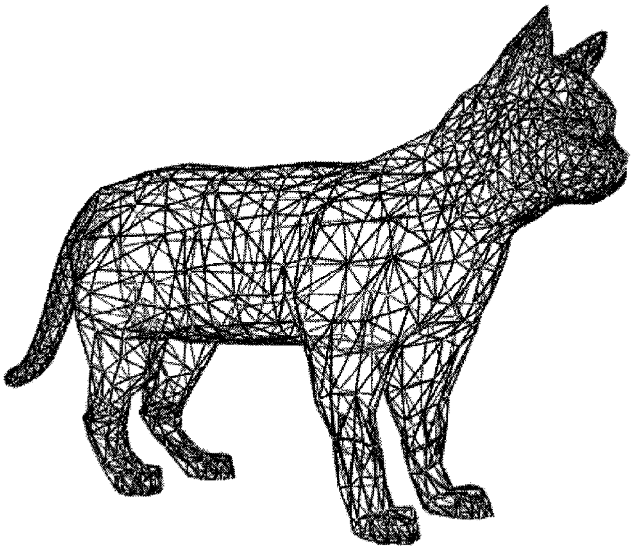

Figure 2: The three-dimensional object models used in the psychophysical and computational experiments are, structurally, high-resolution triangular meshes.

can be shown to be asymptotically convergent to the original function $\rho(x, y)$ with increasing mesh density. By this procedure, the unknown, continuous function $\rho(x, y)$ is replaced by a vector of unknown scalars, that is, the values of ρ at the nodal points.

In a typical experiment, the object is imaged from O different orientations. By writing the discretized version of equation 2.3 for each of the O orientations, one arrives at the matrix equation,

$$\mathbf{r} = \mathbf{B}\rho, \tag{2.4}$$

where:

$\mathbf{r}$ is the vector of goodness-of-view scores (or recognition times, error rates, or something else) associated with the O views: $\mathbf{r} = (r_1, \ldots, r_O)^T$.

$\mathbf{B}$ is a $O \times N$ matrix, $\mathbf{B}_{kv} = \sum_{i=1}^{m_k} S_{ki} a_{ki}^v$ is the summed visibility-area product over all m_k triangles sharing mesh node k ($k = 1, \ldots, N$), at viewpoint v ($v = 1, \ldots, O$). S_{ki} is the surface area of ith mesh triangle sharing node k and a_{ki}^v its visibility at viewpoint v.

ρ is the vector of vertex salience densities: $\rho = (\rho_1, \ldots, \rho_N)^T$.

2.1.5 Solving for Salience. Ill-posedness of the problem. Equation 2.4 predicts the recognizability (or the goodness) of a view when the saliences are known; this is the direct problem that has been typically studied in the human visual object recognition literature (see Tarr & Bülthoff, 1995; and Biederman & Gerhardstein, 1995, for a discussion of the merits of different approaches to this problem). Here we are attempting to solve the inverse problem: given the recognizability scores for O object views and given the geometry of the object, to identify the saliences of the different regions of the object's surface. In other words, we want to derive ρ from $\mathbf{r}$ and $\mathbf{B}$. There are, however, several difficulties in achieving an inverse solution.

Equation 2.3 is a Fredholm equation of the first kind, and such equations are ill conditioned (Groetsch, 1993). Intuitively the kernel a (which encodes the shape the closed object surface) performs an averaging, smoothing operation on the saliency field ρ to yield the goodness-of-view score; therefore, inverting this operation will be very sensitive to small changes in the data. Unfortunately, the discretized version, equation 2.4, inherits the ill-conditioned character of the continuous problem, the singular values of the discretized kernel gradually decaying to zero.

A second problem is that typically the number of views O that can be reasonably tested in an experiment is much smaller than the number of mesh triangles of a realistic three-dimensional model, and therefore system 2.4 is severely underdetermined. For example, the cat model has 3000 triangles, and only 20 views were tested psychophysically in our experiments.

Finally, in practice $\mathbf{r}$ is an observed quantity and thus subjected to measurement errors. Therefore, if $\hat{\mathbf{r}}$ denotes the experimental, noisy data,

$$\hat{r}_v = \sum_{k=1}^{N} \mathbf{B}_{kv}\rho_k + e_v, \tag{2.5}$$

where e_v, the error affecting view v, is a zero-mean normal random variable describing the contribution of all effects (nonlinearities, errors) not explicitly modeled by equation 2.4. We assumed that the off-diagonal elements of the covariance matrix of the errors $C_{vw} = \mathrm{Cov}(e_v, e_w)$ were negligible, with $\sigma_v = (C_{vv})^{1/2}$.

The regularization method. The difficulties listed above, characteristic for ill-posed inverse problems, are generally due to a loss of information in the transformation that must be inverted. They can be overcome by using some prior knowledge of the nature of the solution. As a result, the problem becomes well posed, with solutions unique and continuously dependent on the data. Regularization and the maximum entropy method are among the most widely used techniques for approximating solutions to inverse problems in vision.

The maximum entropy method (see, for example, Skilling, 1989), which applies when the solution can be interpreted as a probability distribution, seeks a solution compatible with the data that has maximum entropy. Maximum entropy has been successfully applied to certain image restoration problems, especially in astronomy. Because saliency is positive and can be normalized so that it represents a probability distribution, maximum entropy in principle can be applied to our problem. However, the saliency values at different points on the surface of an object do not represent independent samples from some probability distribution; in fact, there exist significant spatial relationships among them that are not captured by the maximum entropy formulation.

Regularization methods (Tikhonov & Arsenin, 1977; Engl, Hanke, & Neubauer, 1996) impose smoothness constraints on the the desired solution. These constraints employ spatial derivatives and penalize large spatial variations of the solution. The standard formulation of regularization is due to Tikhonov and Arsenin (1977) and has been widely used in computational vision (Poggio, Torre, & Koch, 1985).

Our algorithm employs a form of Tikhonov regularization. To uniquely determine a well-behaved solution, we used a smoothness constraint imposing similar saliency density values on neighboring points on the object surface.

Therefore, we are seeking an approximate solution $\hat{\rho}$ for equation 2.4 that must satisfy the following conditions:

1. Positivity: $\hat{\rho}_i \geq 0$.
2. Accuracy: $\hat{\rho}$ must minimize the χ^2 deviation from the measured data:

$$\mathcal{E}(\hat{\rho}) = \sum_{v=1}^{O} \left[\frac{\hat{r}_v - \sum_{k=1}^{N} \mathbf{B}_{kv} \hat{\rho}_k}{\sigma_v} \right]^2 . \tag{2.6}$$

Redefining $\hat{r}_v := \hat{r}_v / \sigma_v$ and $\mathbf{B}_{kv} := \mathbf{B}_{kv} / \sigma_v$:

$$\mathcal{E}(\hat{\rho}) = \|\mathbf{B}\hat{\rho} - \hat{\mathbf{r}}\|_2^2. \tag{2.7}$$

3. Smoothness: A quadratic penalty function (regularizer) was used to model the constraint, imposing similar values of salience on neighboring points on the object's surface:

$$\mathcal{S}(\hat{\rho}) = \sum_{\mathcal{N}} \frac{(\hat{\rho}(i) - \hat{\rho}(j))^2}{d(i,j)^2}, \tag{2.8}$$

where the summation ranges over the set $\mathcal{N}$ of all pairs of nodes i and j connected by a mesh edge; $d(i,j)$ is the length of the mesh edge joining nodes i and j. More general penalty functions, allowing discontinuities in the salience field, are discussed later in this section.

These three conditions convert the ill-posed inverse problem into a well-posed constrained minimization problem,

$$\min_{\hat{\rho}}\{\mathcal{E}(\hat{\rho})+\lambda\mathcal{S}(\hat{\rho})\} \text{ subject to } \hat{\rho}\geq 0, \tag{2.9}$$

where $\lambda > 0$ is the regularization parameter. It can be expressed as a quadratic programming problem,

$$\min_{\hat{\rho}}\{\hat{\rho}^T(\lambda\mathbf{M}+\mathbf{B}^T\mathbf{B})\hat{\rho}-\hat{\mathbf{r}}^T\mathbf{B}\hat{\rho}\} \text{ subject to } \hat{\rho}\geq 0, \tag{2.10}$$

where the symmetric, positive definite matrix $\mathbf{M}$ is obtained by partial differentiation of $\mathcal{S}$ with respect to $\hat{\rho}_i$. Because the Hessian matrix $(\lambda\mathbf{M}+\mathbf{B}^T\mathbf{B})$ is positive definite, the solution (if it exists) is unique.

Choice of the regularization parameter. The solution to equation 2.9 depends on the free regularization parameter λ. For small λ, the χ^2 discrepancy $\|\mathbf{B}\hat{\rho}-\hat{\mathbf{r}}\|$ is very small, but the solution has a very large norm and oscillates wildly. A larger λ has the opposite effect: it decreases $\|\hat{\rho}\|$ at the cost of increasing the χ^2 discrepancy, yielding a solution that varies slowly across the object surface but reconstructs poorly the measured data. A compromise between these two extremes is clearly desirable. According to the discrepancy principle (Morozov, 1993), the regularization parameter is chosen so that the size of the discrepancy $\|\mathbf{B}\hat{\rho}-\hat{\mathbf{r}}\|$ is the same as the error level in the data. The number of degrees of freedom of the unknown function ρ is O, and thus the expected value of the χ^2 discrepancy (see equation 2.6) is O. Therefore, the regularization parameter is to be chosen to render the χ^2 discrepancy measure in equation 2.6 equal to O. Since the discrepancy is a continuous, increasing function of λ, there exists a unique solution λ^0 satisfying the condition $\chi^2 = O$. If reliable noise estimates are unavailable, methods such as the L-curve plot (Regińska, 1996; Hansen, 1998) or generalized cross-validation (Wahba, 1990) can be used.

Discontinuities in the salience field. The quadratic regularizer in equation 2.8, however convenient computationally, leads to oversmoothing: it imposes smoothness everywhere on the object, and the penalty for large differences is too extreme. To allow for the recovery of sharply delimited features, the smoothness constraint must be switched off for large differences in salience between neighboring nodes. In other words, global smoothness needs to be replaced with piecewise smoothness.

We briefly describe two approaches to this problem. First, one may introduce the discontinuities implicitly (Geman & Reynolds, 1992) by replacing the quadratic penalty function with a concave function of the form: $\phi(u)=-(1+|u|^{\gamma})^{-1}$ where $u=(\hat{\rho}(i)-\hat{\rho}(j))/d(i,j)$. Since $\lim_{u\to\infty}\phi(u)=0$ this function allows large jumps (discontinuities) in the salience field.

Second, we can explicitly introduce discontinuities in the salience field (Geman & Geman, 1984; Marroquin, 1984). The salience field is modeled as a markov random field (MRF) (Li, 1995) defined on the nodes of the mesh. The MRF model is appropriate since we assume spatial interactions only between neighboring mesh nodes; the associated Gibbs energy includes potentials for cliques up to size two, modeling the deviation from the data and the smoothness constraint. Coupled to the the salience MRF is a second MRF, the line process, located on the edges of the mesh. The line process variables are binary, indicating the presence or absence of a discontinuity across the corresponding mesh edge. Unfortunately, the determination of the salience field and line process variables is a difficult minimization problem.

Piecewise smoothing is superior to global quadratic smoothing; however, given that our problem was severely underdetermined, the use of a quadratic regularizer (see equation 2.8) is the only practical option.

2.2 View Similarity. The formalism presented above for goodness-of-view data can be used to model perceptual similarities among the different views of an object. Our basic hypothesis was that the dissimilarity of two views of an object increases with both the extent of feature visibility change (induced by the change in object orientation) and feature salience. A general model for the dissimilarity (psychological distance) between views v and u is given by the elliptical distance:

$$d_{vu}^2 = (\mathbf{B}_{v:} - \mathbf{B}_{u:})\mathbf{G}(\mathbf{B}_{v:} - \mathbf{B}_{u:})^T, \tag{2.11}$$

where $\mathbf{G}$ is a symmetric, positive semidefinite matrix. $\mathbf{B}_{v:}$ denotes row v of matrix $\mathbf{B}$, which describes the visibilities of the mesh triangles in view v. We made the simplifying assumption that $\mathbf{G}$ is diagonal. Therefore, the above formula reduces to a weighted Euclidean distance (Carroll & Chang, 1970),

$$d_{vu}^2 = \sum_{k=1}^{N} \rho_k (\mathbf{B}_{kv} - \mathbf{B}_{ku})^2, \tag{2.12}$$

where the saliences ρ_k are positive. According to this model rotations in the image plane yield $d = 0$, since the visibility of the object's features does not change. Note that the object features that are invisible in both views do not contribute to d, since for them $\mathbf{B}_{kv} = \mathbf{B}_{ku} = 0$.

The similarity inverse problem is formally identical to the goodness-of-view inverse problem, with the difference that the saliences ρ_k must now be derived from the perceptual dissimilarities between the tested views. This problem is also ill posed and must be regularized by imposing some smoothness constraint on the solution. Assuming, as before, that the saliences ρ_k have similar values for neighboring points on the object's surface, we solved for ρ_k by applying the same regularization algorithm (section 2.1.5).

3 Applications

The algorithm described in the preceding section was applied to goodness-of-view and similarity data collected in psychophysical experiments employing three-dimensional animal models. The rationale for using animal models was that their salient features correspond, by and large, to their anatomically defined parts (Tversky & Hemenway, 1984), and thus the performance of the algorithm could readily be assessed. The experimental design and psychophysical results will be detailed in a different article; here we briefly explain the experimental methodology and summarize some of the results.

3.1 Goodness-of-View Data.

3.1.1 Experimental Design. Each test object was imaged from 20 viewpoints uniformly distributed on the viewing sphere centered on the object. The subjects (Brown University students) were shown pairs of views of the same object and were instructed to select the better view in each pair, "better" being defined as "more informative" or "more representative." All 190 possible view pairs were tested for each test object. A minimum of five subjects were used for each object. For a given test object, the data from all subjects were pooled and jointly used to derive the goodness-of-view values and variances necessary for equation 2.6. This derivation was based on Thurstone's law of comparative judgment, case IV (Torgerson, 1958). The saliences for the mesh vertices were determined by minimizing expression 2.9, and λ chosen as described in section 2.1.5. To illustrate the solution graphically, the salience values were gray-level coded, black representing minimum salience and white representing maximum salience. In other words, the salience density field was painted on the object.

3.1.2 Verification of the Algorithm. In an initial experiment, we verified that the algorithm correctly recovered the perceptually salient object features or parts. This was accomplished by instructing subjects to focus on certain predetermined object parts when judging goodness-of-view. For the deer model, in one experiment we asked subjects to focus on the right ear, and in a second experiment, the muzzle. For the hand model, we asked subjects to focus on the thumb and the index finger, and in a second experiment, the middle finger. The results, displayed in Figure 3, confirmed that the algorithm properly selected the relevant object features as defined in the instructions to the subject.

3.1.3 Results. In the actual goodness-of-view rating experiments, subjects were given no instructions about which object parts to focus on. The best views of the animal models corresponded to either the frontal or the so-called three-fourths viewpoint, which is frontolateral and represents the

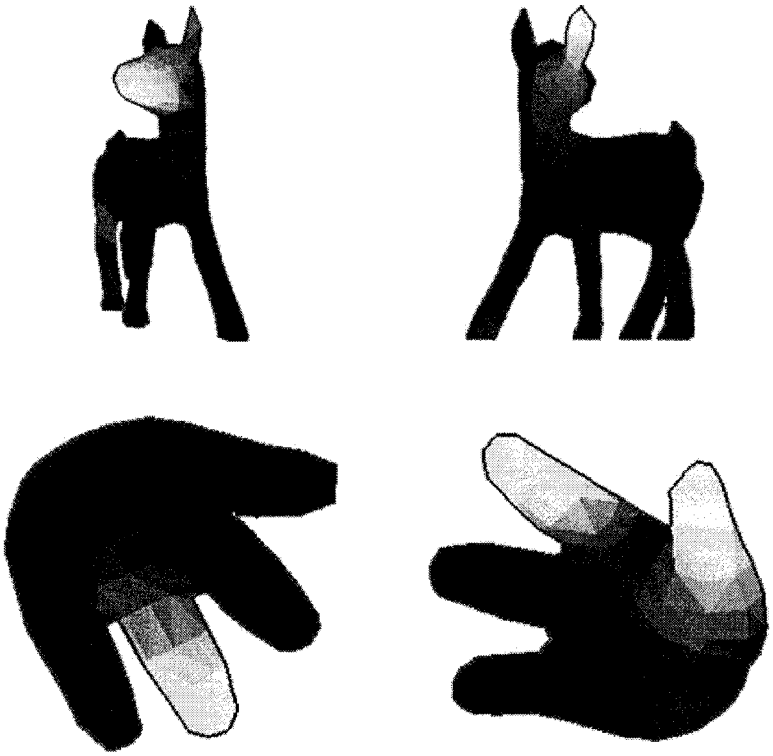

Figure 3: To verify the performance of the inversion algorithm, subjects were instructed to focus on predetermined object parts when judging goodness-of-view. The salience values computed from the averaged subject data were coded so that white represents maximum salience and black represents minimum salience. The algorithm correctly assigned the maximum salience to the predetermined features. (Top) The muzzle and the ear were correctly recovered as perceptually dominant features in two verification experiments involving the deer model. (Bottom) The index together with the thumb, and the middle finger were correctly recovered as perceptually salient features in the two experiments involving the hand model.

canonical viewpoint (Palmer et al., 1981) for objects (such as animals) having natural front, side, and back sides. The views showing the animals from the back were rated as the worst.

The saliency fields derived from the goodness-of-view data are shown on the left side in Figures 4, 5, 6, 7, and 8. The more salient regions of the object's surface correspond to the head, neck/chest, and forelimbs, which is compatible with frontal and three-fourths best views. We informally de-

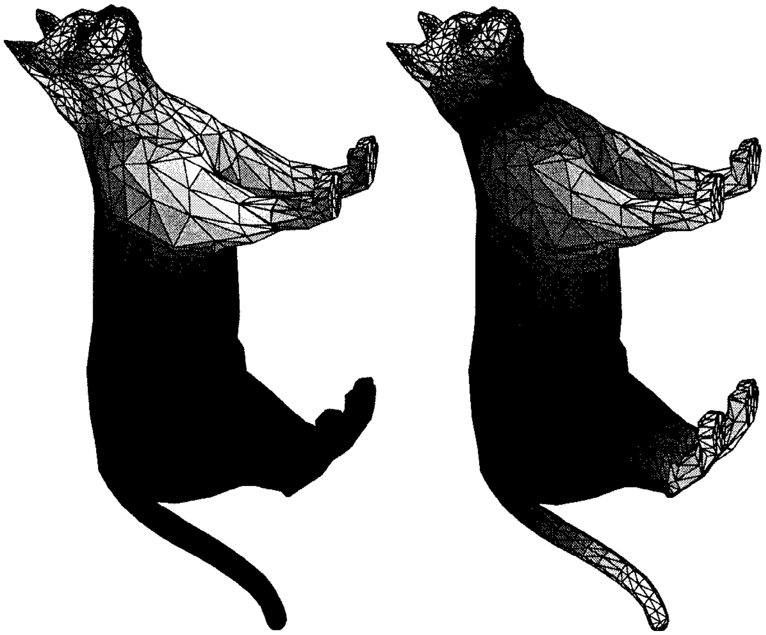

Figure 4: (Left) Salience field for the cat model derived from goodness-of-view data. (Right) Salience field for the cat model derived from view similarity data. White represents maximum salience, and black represents minimum salience.

briefed subjects to determine whether the results of the algorithm were in agreement with their subjective behavior. In general, we found good correspondence between the computational and observational measures.

3.2 View Similarity Data.

3.2.1 Experimental Design. In the viewpoint similarity experiment subjects were shown pairs of views of the same object and were asked to rate their similarity on a scale from 1 to 10. Five subjects were used for each object. As in the goodness-of-view experiment, each object was imaged from 20 viewpoints uniformly distributed on its viewing sphere. All 190 possible view pairs were rated for both test objects. The similarity ratings, averaged over subjects, were used to derive the dissimilarity values necessary for equation 2.12.

3.2.2 Results. The saliency fields derived from the viewpoint similarity data are shown on the right side in Figures 4, 5, 6, 7, and 8. Note that the salient regions correspond to all major anatomically defined body parts.

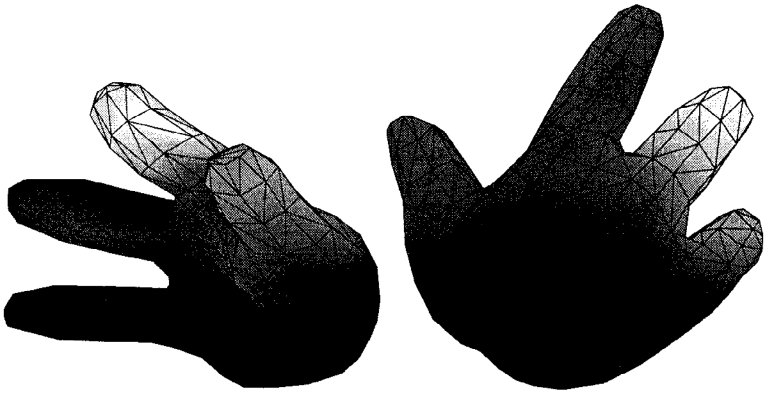

Figure 5: (Left) Salience field for the hand model derived from goodness-of-view data. (Right) Salience field for the hand model derived from view similarity data. White represents maximum salience, and black represents minimum salience.

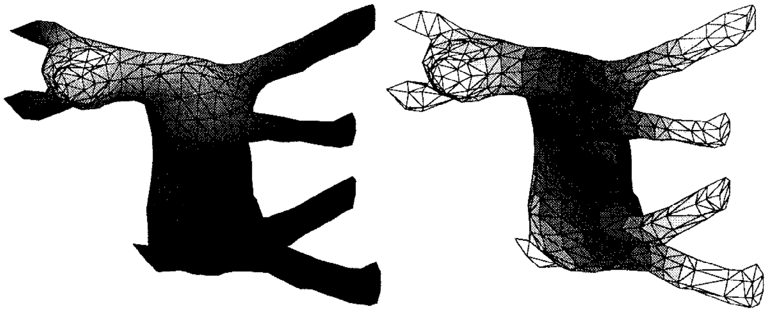

Figure 6: (Left) Salience field for the deer model derived from goodness-of-view data. (Right) Salience field derived from view similarity data. White represents maximum salience, and black represents minimum salience.

4 Discussion

We have defined the perceptual features of an object as highly salient regions on its surface. Features thus defined are viewpoint independent, since each point on the surface of the object is assumed to have an intrinsic saliency value, independent of the orientation of the object relative to the viewer. They are fundamentally different from the features proposed by Koenderink and van Doorn (1976), which are singularities of the $3D \to 2D$ imaging transformation, and therefore viewpoint dependent in an essential way.

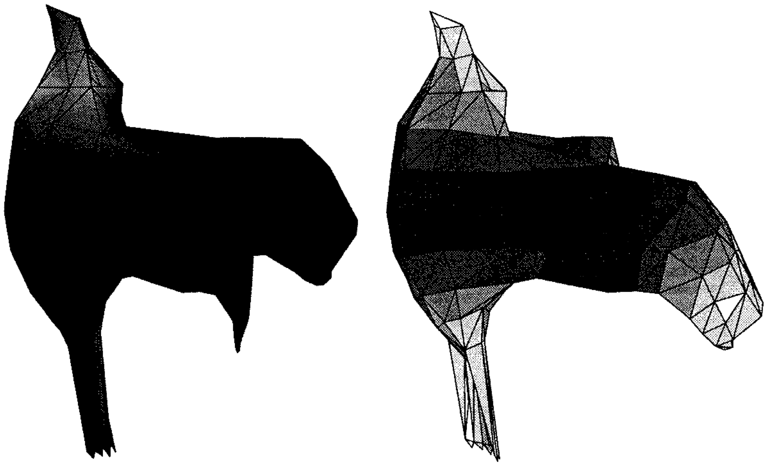

Figure 7: (Left) Salience field for the seagull model derived from goodness-of-view data. (Right) Salience field derived from view similarity data. White represents maximum salience, and black represents minimum salience.

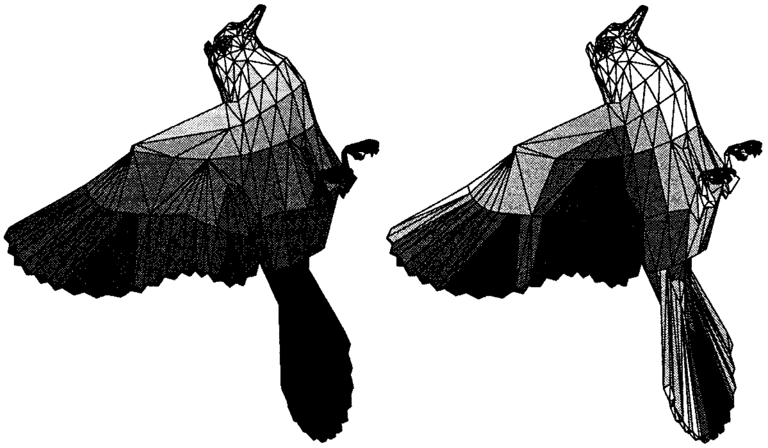

Figure 8: (Left) Salience field for the dove model derived from goodness-of-view data. (Right) Salience field derived from view similarity data. White represents maximum salience, and black represents minimum salience.

Koenderink and van Doorn's features are properties of the projection transformation and cannot be located on the object in a physical sense. On the contrary, our features are intrinsic properties of the object's surface. These definitions appear to be complementary, and possibly could capture two different aspects of object perception and representation.

A limitation of our model is its linearity: equation 2.3 does not include interactions between different surface patches. Such interactions are important in the makeup of spatially extended, holistic features such as edges and in encoding configural information for stimuli such as faces. Our model, being essentially local, accounts for such phenomena only indirectly. Take the case of a salient long edge or, say, of an eye in a face. The patches associated with these features would be deemed salient by our algorithm. However, if some component patches of the edge or of the eye are occluded, the patch configuration is disrupted; as a result, the perceived salience of the entire feature decreases to a much larger extent than predicted by the model. The problem is to a large extent one of experimental limits. Although the model could in principle be extended to include pairwise (or even triplewise) interactions, the resulting explosion in the number of unknown parameters would render any inverse solution meaningless.

Despite its simplicity, our linear salience model has yielded some interesting results. The application of the inversion algorithm to goodness-of-view and similarity data collected in psychophysical experiments has confirmed the intuition of experimental psychologists (Palmer et al., 1981; Tversky & Hemenway, 1984) that natural parts such as the limbs and head are highly salient in human object perception. It is worth pointing out, however, that this result does not imply that parts form the fundamental units of mental representations of three-dimensional objects. Rather, it is simply the case that stable clusters of features tend to co-occur repeatedly in the same relative locations within object parts. Therefore, it remains an open question as to the nature of the perceptual units used in object representation (for example, see Hayward & Tarr, 1997; Tarr et al., 1997). What we do know is that attending to such feature clusters or parts can account for the subjects' behavior in the comparative judgment experiments.

The similarity judgment task resulted in richer, better defined sets of salient features than the simpler goodness-of-view task. The salient surfaces recovered from goodness-of-view data correspond, by and large, to the surfaces visible in the three-fourths view of the object. On the other hand, the salient surfaces recovered from viewpoint similarity data correspond to all the major anatomical parts of the animal. Note that "major" does not mean large, but rather perceptually prominent: the tail of the deer model in Figure 6 is salient although negligible in size. It is tempting to hypothesize that our analysis reveals the features the subjects have used in their similarity judgments, much like multidimensional scaling reveals the perceptual dimensions of the representational space from the same types of data.

To our mind, these results provide a powerful new tool for studying human object recognition; the model presented in section 2 needs not be restricted to goodness-of-view data. For example, we are planning to use more objective-recognition performance measures, such as response times and error rates measured in naming and recognition memory experiments. Following the derivation of the salient features according to the algorithm described in this article, psychophysical experiments can be run in which the salient features thus derived are masked; a drastic decline in recognition performance for such masked stimuli would confirm that the algorithm has indeed found the features of recognition (see Biederman, 1987, for a similar methodology). Such methods will shed new light on the problem of the features of recognition in human vision. To date, little has been learned about the nature of such features; most current theories of recognition posit relatively ad hoc unverified feature sets because no methodology was available to do any better. Given our solution to the inverse problem, more principled feature sets can be derived and then tested directly (by predicting performance).

We should also note that although common objects were used in this article, there are reasons to extend the work to novel three-dimensional objects (which has become a standard in the field; see Biederman & Gerhardstein, 1993; Bülthoff & Edelman, 1992; Hayward & Tarr, 1997; Tarr, 1995; Tarr et al., 1997). Novel objects have the advantage that the features of recognition are less obvious, and less influenced by previous knowledge or biases. After exploring a large number of objects from different shape classes, it should be possible to describe some of the unifying geometrical characteristics of highly salient object features (independent of function or other learned biases). Our ultimate goal is to develop an algorithm able to predict the generic salience field when given only the geometry of a closed three-dimensional surface (of course, some of these saliences will vary with the task).

Finally, another interesting application of the algorithm is to modeling similarities between different objects from the same basic-level category (the category that is considered the default level of access, for example, “chair,” “car,” or “bird”; see Rosch et al., 1976). In other words, similarities in view space would be replaced by similarities in shape space. To see how this could work, consider a set of N objects from the same basic-level class, such as a set of three-dimensional head models obtained with a three-dimensional scanner. Assume that the correspondence between the mesh elements of the objects is known. The dissimilarity measure in equation 2.12 must be generalized by replacing the change in visibility due to viewpoint change with a measure of the physical changes (such as relative position or size) of corresponding mesh elements across the objects in the set. The rest of the derivations would remain identical, resulting in a salience map assigning different perceptual weights to the features shared by all objects in the class. One could thus derive the object features that are involved in categorization at both the basic and subordinate levels.

In summary, we have presented a new algorithm for computing perceptual saliences from behavioral data. The importance of this method lies in the fact that there are no known robust empirical methods for directly determining the features used in human object perception. In contrast, there are proved methods for measuring performance, including directly collecting goodness-of-view and similarity preferences or recording response times and error rates from human observers. The method we have presented here leverages this claim by providing a solution to the inverse problem of predicting object perception and recognition performance: What are the features that mediate such performance given that ratings can be collected or performance can be measured? Given that relatively little progress has been made regarding the features used in object perception, it seems reasonable to take this inverse approach, using our method as one tool for inferring the features of perception and recognition.

References

Ashby, F. G. (Ed.). (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.

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

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19*(6), 1162–1182.

Biederman, I., & Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1506–1514.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Natl. Acad. Sci. USA, 89*, 60–64.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of the Eckart-Young decomposition. *Psychometrika, 35*, 283–319.

Cutzu, F., & Edelman, S. (1998). Representation of object similarity in human vision: Psychophysics and a computational model. *Vision Research, 38*(15 / 16), 2229–2258.

Cutzu, F., & Tarr, M. (1997, February). The representation of three-dimensional object similarity in human vision. In *Proc. SPIE Conf. on Electronic Imaging: science and technology*, San Jose, CA.

Edelman, S. (1995). Representation of similarity in 3D object discrimination. *Neural Computation, 7*, 407–422.

Engl, H. W., Hanke, M., & Neubauer, A. (1996). *Regularization of inverse problems*. Dordrecht: Kluwer.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721–741.

Geman, D., & Reynolds, G. (1992). Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 14*(3), 367–383.

Groetsch, C. W. (1993). *Inverse problems in the mathematical sciences*. Braunschweig: Vieweg & Sohn.

Hansen, P. C. (1998). *Rank-deficient and discrete ill-posed problems: Numerical aspects of linear inversion*. Philadelphia: Society for Industrial and Applied Mathematics.

Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23*(5), 1511–1521.

Koenderink, J. J., & van Doorn, A. J. (1976). The singularities of the visual mapping. *Biological Cybernetics, 24*, 51–59.

Li, S. Z. (1995). *Markov random field modeling in computer vision*. Berlin: Springer-Verlag.

Marroquin, J. (1984). *Surface reconstruction preserving discontinuities* (A.I. Memo No. 792). Cambridge, MA: Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Morozov, V. A. (1993). *Regularization methods for ill-posed problems*. Boca Raton, FL: CRC Press.

Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 135–151). Hillsdale, NJ: Erlbaum.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature, 317*, 314–319.

Regińska, T. (1996). A regularization parameter in discrete ill-posed problems. *SIAM J. Sci. Comput., 17*(3), 740–749.

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

Schwarz, H.-R. (1988). *Finite element methods*. Orlando, FL: Academic Press.

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, in press.

Skilling, J. (Ed.). (1989). *Maximum entropy and Bayesian methods*. Dordrecht: Kluwer.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review, 2*(1), 55–82.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon-structural-descriptions or by multiple-views? *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1494–1505.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science, 8*(4), 282–289.

Tikhonov, A. N., & Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. Washington, D.C.: W. H. Winston.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113*, 169–193.

Wahba, G. (1990). *Spline models for observational data*. Philadelphia: Society for Industrial and Applied Mathematics.

Received January 5, 1998; accepted November 18, 1998.