On Similarity to Prototypes in 3D Object Representation

Sharon Duvdevani-Bar Shimon Edelman
Dept. of Applied Mathematics and Computer Science
The Weizmann Institute of Science
Rehovot 76100, Israel
[sharon,edelman]@wisdom.weizmann.ac.il

July 20, 1995

Abstract

A representational scheme under which the ranking between represented dissimilarities is isomorphic to the ranking between the corresponding shape dissimilarities can support perfect shape classification, because it preserves the clustering of shapes according to the natural kinds prevailing in the external world. We discuss the computational requirements of rank-preserving representation, and examine its plausibility within a prototype-based framework of shape vision.

Contents

1 Introduction

The greatest challenge in devising an effective representation of the visual world is the variability of the appearance of objects, both familiar and unfamiliar, encountered by a visual system. This variability, caused by factors such as changing illumination, viewpoint, and, in the case of articulated or flexible objects, by the changes in the shape of the object, goes largely unnoticed in everyday life, because of the great flexibility of the brain in dealing with the naturally occurring visual stimuli. In this report we propose that the impressive performance of the object recognition system in primate vision stems from prototype-based representation of similarities among natural shape kinds. Our approach is motivated by a series of recent experimental findings concerning shape processing in the brain, and their theoretical interpretation, as outlined below.

1.1 Psychophysical constraints on representation

Psychophysical studies conducted in the past few years led to the characterization of certain basic limitations of the visual system in generalizing shape-based recognition to novel conditions (Rock and DiVita, 1987; Tarr and Pinker, 1989; Bülthoff and Edelman, 1992; Edelman and Bülthoff, 1992; Humphrey and Khan, 1992; Edelman, 1995a; Cutzu and Edelman, 1994); see (Jolicoeur and Humphrey, 1994) for an extensive review and a discussion. Specifically, it was found that the recognition of novel views of objects tends to be slower and more prone to errors than the recognition of highly familiar views, that the effect of viewpoint increases when the objects to be distinguished are more similar to each other, and that it persists even when full 3D shape information is available to the subject through, e.g., binocular stereo cues.

Attempts to model the human recognition performance led to a certain progress in understanding the mathematics and the possible neurobiological mechanisms behind the psychophysical data. In particular, it is becoming increasingly clear that human performance in recognizing objects under changes of viewpoint can be accounted for to a significant degree by a theory based on the notion of interpolation among view-specific representations (Poggio and Edelman, 1990; Edelman and Weinshall, 1991; Bülthoff and Edelman, 1992; Edelman, 1995a). Detailed computational modeling showed that this approach to representation can be implemented by putting together simple building blocks such as the nonlinear graded-profile overlapping receptive fields[1] found in biological visual systems (Edelman, 1995a; Edelman, 1995b).

1.2 View prototypes and the representation of 3D objects for recognition

The computational basis of representation by multiple views is twofold. On the one hand, the space of all views of a 3D object can be characterized mathematically as a smooth low-dimensional manifold in the space of all views of all possible objects (Ullman and Basri, 1991). Thus, a "black box" capable of recognizing an object can be built by learning to interpolate the view-space manifold from examples, using standard methods for function approximation (Poggio and Edelman, 1990; Poggio, 1990). On the other hand, interpolation from examples can be considered as taking a linear combination of responses of "units" each encoding a particular example (i.e., a prototypical input/output relationship; Poggio and Girosi, 1990). Furthermore, the response profiles of such units need not be more complicated than

[1]Many of the lower-level areas in the primate visual system are organized retinotopically, that is, as maps which preserve to a certain degree the topography of the retina. A unit that is a part of such a retinotopic map normally responds selectively to stimulation in a well-localized part of the visual field, referred to as its receptive field (RF).

the familiar Gaussian-profile receptive fields, commonly found in biological information processing systems. Thus, the prototypical-view approach possesses both mathematical and biological appeal.

1.3 Shape prototypes: from recognition to classification

A more challenging problem arises if one wishes to represent unfamiliar objects, for which particular view prototypes are not likely to exist in the system. Because the view-space theorems such as that of (Ullman and Basri, 1991) do not apply in this case, no principled predictions regarding the structure of the manifold of views of novel objects can be made. Nevertheless, representations of a sufficient variety of objects, or prototypes, taken together, can serve as a basis for a provably faithful distributed (ensemble) representation of a completely unfamiliar shape, as noted in (Edelman, 1995c). This approach to representation relies on the concept of *shape space* (superseding that of view space), and is centered around the observation that faithful proximal representation of shape contrasts allows the reconstruction of distal shape contrasts, which, in turn, can serve as a basis for solving problems such as shape classification.

The computational principle behind this idea is the same one that underlies nonmetric multidimensional scaling:[2] for a set of points (in the present case, points in the shape space), the ranks of interpoint distances constrain the point configuration sufficiently to allow its retrieval from the rank data. If the measured (proximal) ranks approximate well the distal ranks of dissimilarities between object shapes, the relative location of each object in the shape space (in other words, the shape of the object, described in terms of the other objects) can be effectively estimated. In the present paper, we explore various theoretical and computational aspects of this idea; some results of a parallel psychophysical research effort can be found in (Edelman, 1995b; Cutzu and Edelman, 1995). In section 2 we describe representation by second-order isomorphism, first formulated by Shepard in 1968; this notion serves as a convenient starting point for developing a formal approach to representation by similarity, and, in particular, for placing constraints on the proximal-to-distal mapping. Section 3 then introduces a prototype-based scheme called Chorus and analyzes its computational properties, and, in particular, its compliance with the requirements for faithful representation. Section 4 examines the possibility of implementing the Chorus scheme using receptive fields such as those found in biological visual systems. Finally, section 5 includes a summary and a discussion of some open issues.

2 First and second-order isomorphism in representation

2.1 The two problems of representation

It is possible to distinguish between two problems about mental representation (Cummins, 1989). The first of these, the problem of representations (plural), is basically empirical; two instances of this problem are the unraveling of the representations employed by natural cognitive systems, and the design of the representational substrate for artificial cognitive modules. The second problem is called by Cummins the Problem of *Representation* (singular). Here, the central question is how, in principle, can an internal state of a system refer to anything at all in the external world.

Curiously, the Problem of Representation is rarely discussed in cognitive psychology. Arguments put forward by Marr (1982) prompted cognitive scientists to contemplate the fact that not all representations are computationally equivalent; the classical example here is the difficulty of doing arithmetics

[2] Nonmetric MDS (Shepard, 1980) is an algorithm for embedding a set of points in a metric space, while preserving as closely as possible the ranking of pairwise distances between the points, specified in advance. In psychophysical applications (see section 5.2), this ranking is derived from the subject's response data.

with Roman numerals, compared to the ease of calculation using positional notation. At the same time, it is generally taken for granted that a certain pattern of neuronal activity in someone's brain can represent the shape of a cat or the number 17, or, amazingly, that physically distinct patterns of activity in the brains of two people can stand for the same shape (that of a cat) or the same number (17).

Reviews in cognitive psychology are likely to adopt an extreme position concerning representation. For example, Palmer (1978) asks "What is representation," then proceeds with an in-depth survey of the field, the upshot of which is that "we ... do not really understand our concepts of representation." In contrast, Suppes, Pavel and Falmagne (1994) state flatly "Representation of something is an image, model, or reproduction of that thing"; this amounts to the Aristotelian idea of representation by resemblance, which happens to have been discredited so strongly as to become a rare point of consensus in the philosophy of mind (Cummins, 1989). The present paper takes a middle path between these two extremes, by abandoning the question "what is representation" in favor of "what can make it work," and by trying to bring certain recent findings in vision to bear on that latter formulation of the Problem of Representation.

Barring resemblance, or "first-order structural isomorphism" (Shepard, 1968) between the object and the entity that stands for it internally, what relationship can qualify as representation of something by something else? Shepard (1968) suggests *"second-order" isomorphism*: "... the isomorphism should be sought — not in the first-order relation between *(a)* an individual object, and *(b)* its corresponding internal representation — but in the second-order relation between *(a)* the relations among alternative external objects, and *(b)* the relations among their corresponding internal representations. Thus, although the internal representation for a square need not itself be square, it should (whatever it is) at least have a closer functional relation to the internal representation for a rectangle than to that, say, for a green flash or the taste of a persimmon," (Shepard and Chipman, 1970, p.2).

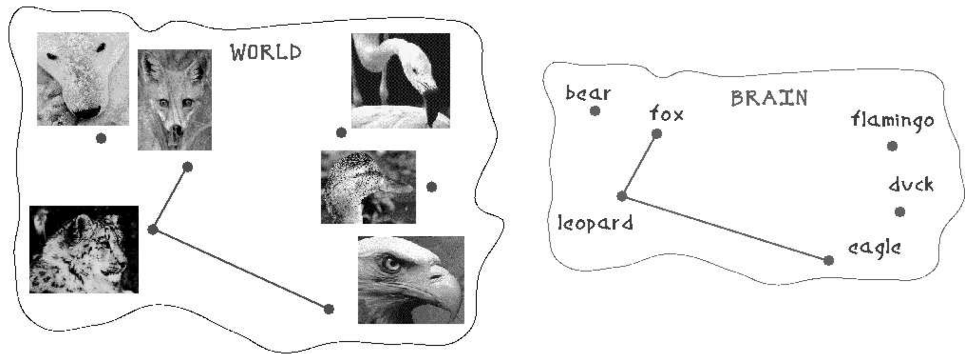

Figure 1: Clustering by natural kind, and its representation that fulfills the requirement of second-order isomorphism (in this case, isomorphism between distance ranks in the two spaces).

2.2 Conditions for faithful representation

Let us now identify mathematical conditions under which isomorphism between distal and proximal relations gives rise to representation which is, in a sense, true to the original.[3] Let $\mathcal{D}$ be the set of distal objects, and $\mathcal{P}$ — the set of internal tokens that participate in the process of representation. Let $f : D \subset \mathcal{D} \to \mathcal{X}$ and $g : P \subset \mathcal{P} \to \mathcal{Y}$ be functions defined, respectively, over sets of distal and proximal entities. According to the requirement of second-order isomorphism, $\mathcal{D}$ is isomorphic to $\mathcal{P}$

[3]Some qualifications to this formulation of veridicality, or "truth in representation," can be found in section 5.3.

under a mapping M if

$$\forall D \subset \mathcal{D}, f(D) \sim g(M(D)) \tag{1}$$

where the relation $\sim$ is that of set isomorphism, defined over $\mathcal{X}, \mathcal{Y}$. To properly constrain the choice of M, we have to be more specific in defining the functions f, g. In the context of shape perception, it is natural to consider for this purpose *similarity* (actually, two similarity functions must be defined, one for the external objects and one for the internal tokens standing for those objects). Intuitively, we would like the representation to capture the similarity relationships within and between natural kinds (Quine, 1969). More specifically, similarity is seen to be relevant both to recognition, in which case resemblance between the viewed shape and some previously seen ones is to be assessed, and to classification, where similarities between a shape and a number of shape classes are compared (Nosofsky, 1988; Goldstone, 1994). For a pair of internal representations, similarity is not directly observable, but can still be defined operationally as the degree to which an optimal stimulus for one token activates the other one (Shepard and Chipman, 1970; Edelman, 1995c).

As noted by philosophers (Nicod, 1930; Quine, 1973), similarity should be construed as a triadic relation: knowing that A is more similar to B than to C is much more informative than merely knowing that A is similar to B and B to C. Furthermore, we may assume that similarity is defined qualitatively and not quantitatively, that is, the range of f, g is $\mathcal{X} = \mathcal{Y} = \{>, <\}$. This assumption limits neither classification (because knowledge of similarity for every triplet of points belonging to a metric space defines unequivocally the clustering of the points), nor recognition (because the location of each point can be recovered from triadic similarities using nonmetric multidimensional scaling; see section 5.2). We thus obtain $f : \mathcal{D} \times \mathcal{D} \times \mathcal{D} \rightarrow \{>, <\}$, and, analogously, $g : \mathcal{P} \times \mathcal{P} \times \mathcal{P} \rightarrow \{>, <\}$. Condition 1 then becomes:

$$\forall D_1, D_2, D_3 \in \mathcal{D}, f(D_1, D_2, D_3) = g(M(D_1), M(D_2), M(D_3)) \tag{2}$$

In other words, the mapping M is required to preserve similarity ranks.

2.3 Constraints on the shape space

The notion of similarity invoked in the preceding section can be put on a firm basis if we assume that the distal shape space and the internal representation space are both endowed with metric structure, in which case (dis)similarity in each space can be derived from the appropriate distance function. Now, to demonstrate the existence of a metric for a given shape family (i.e., a subspace of the shape space), it suffices to exhibit at least one parameterization of all the shapes in the family (the possibility of different parameterizations is discussed in section 5.3). A metric on shapes can then be defined simply as the Euclidean distance in the underlying parameter space R^n.[4] An example of objects produced using such a parameterization for a family of animal-like shapes is illustrated in Figure 2.

2.4 Constraints on the proximal-to-distal mapping

2.4.1 Strict rank preservation

Given that distances can be measured between entities both in the external shape space and in its internal representation, the similarity ranking of three distinct points $\mathbf{x}, \mathbf{y}, \mathbf{z} \in R^n$ can be formalized

[4] As we have already noted above, defining a metric on internal representations is more of a technical issue that a matter of principle, because it does not rely on assumptions regarding the external world.

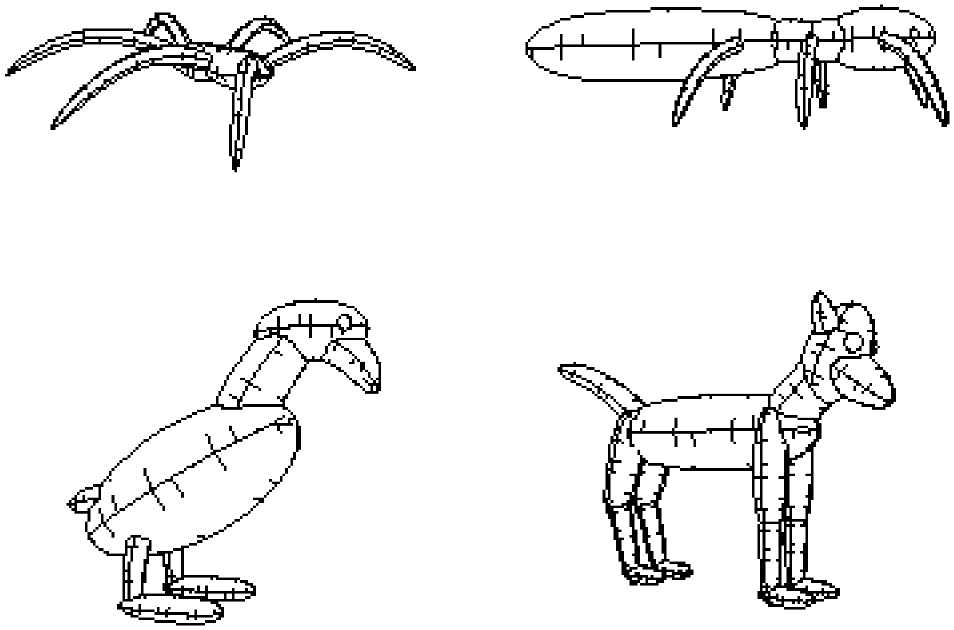

Figure 2: Four shapes jointly parameterized by a set of 57 parameters (Manolache and Edelman, 1993). The possibility of such parameterization for a given shape family is a prerequisite for its faithful representation, as argued in section 2.3.

as follows. Consider the *simple ratio* of $\mathbf{x}, \mathbf{y}, \mathbf{z}$, defined as

$$\langle \mathbf{x}\,, \mathbf{y}\,, \mathbf{z} \rangle = |\mathbf{x} - \mathbf{y}|/|\mathbf{y} - \mathbf{z}| \tag{3}$$

Obviously, a mapping $S : R^n \to R^n$ preserves distance ranks iff it preserves $\langle \mathbf{x}\,, \mathbf{y}\,, \mathbf{z} \rangle$ for any choice of points. As shown in (Reshetnyak, 1989), p.32, a bijective mapping S with this property must be a similitude (that is, a mapping of the form $S(\mathbf{x}) = \lambda\, P(\mathbf{x})$, where $\lambda > 0, \lambda \in R$, and $P : R^n \to R^n$ is an orthogonal transformation). Thus, the rank preservation requirement is quite restrictive in the class of mappings that it allows. As we shall see in the next two subsections, one can increase the size of the class of admissible mappings by relaxing the strict rank preservation requirement in two respects.

2.4.2 Local rank preservation

Locally, the rank preservation requirement is satisfied by *any* well-behaved mapping in one or two dimensions. Specifically, a mapping realized by an analytic function with a non-vanishing Jacobian in a given region is conformal there, that is, it preserves angles (Cohn, 1967). Such a function preserves similitude of small triangles; in particular, a scalene triangle formed by a triplet of points will be mapped into a triangle with the same ranking of side lengths (see Figure 1).

Conformality in higher dimensions is much more restrictive. As proved by Liouville in 1850, already for $n = 3$ there are no conformal mappings from R^n to itself besides those which are composed of finitely many inversions with respect to spheres. Such mappings, called Möbius transformations, constitute a finite-dimensional Lie group which includes the group of motions in R^n and is only slightly broader than that group (Reshetnyak, 1989).

2.4.3 Local approximate rank preservation

A considerably broader class of mappings emerges if the requirement of conformality is replaced by that of quasiconformality. Intuitively, a regular topological mapping is quasiconformal if there exists a constant q, $1 \leq q \leq \infty$, such that almost any infinitesimally small sphere is transformed into an ellipsoid for which the ratio of the largest semiaxis to the smallest one does not exceed q (Reshetnyak, 1989). Under such a mapping, the ranks of distances between points are not strictly preserved, but are guaranteed not to change too much, if q is sufficiently close to 1.

How relevant is local approximate preservation of distance ranks, offered by a quasiconformal proximal-to-distal mapping M, to the representation of real-world shapes? A part of the answer to this question emerges from a consideration of the hierarchical structure of perceived categories. Numerous studies in cognitive science reveal that in the hierarchical structure of object categories there exists a certain level, called basic level, which is the most salient according to a variety of criteria (Rosch et al., 1976; Rosch, 1978). Taking as an example the hierarchy quadruped, dog, terrier, the basic level is that of dog. Objects whose recognition implies less detailed distinctions than those required for basic-level categorization are said to belong to a superordinate level. A reasonable assumption is that faithful representation of similarities is required *within* superordinate categories (e.g., within quadruped: giraffe to camel, leopard, horse),[5] but not *between* superordinate categories. In other words, the informativeness (and, indeed, the well-posedness) of the statement that a giraffe is more similar to a banana than to a beetle is questionable.

The above considerations suggest that the locality constraint per se does not preclude basing representation upon the principle of conformality of the proximal-to-distal mapping M. The question of relevance, raised in the preceding paragraph is reduced, therefore, to a practical one: for a given point in the distal shape space, how large is the domain over which the distortion coefficient q of the mapping M is likely to be close enough to 1? Theoretical considerations and computer simulations presented in the remainder of this paper suggest that the answer to that question is, probably, "large enough."

3 Representation by distances to prototypes

3.1 Similarity to prototypes: giraffe = [0.6 camel, 0.4 leopard]

Having discussed some mathematical aspects of the idea of representation by second-order isomorphism, we now turn to an analysis of a concrete representational scheme closely related to that idea. The scheme, called the Chorus of Prototypes, has been proposed in (Edelman, 1995c), and is illustrated in Figures 3 and 4. The basic motivation behind Chorus is to represent the shape of an object by holistic similarities to related shapes, and thus to avoid the need to specify explicitly the object's geometry.

A link between the Chorus scheme and the requirements on the proximal-to-distal mapping M, derived in the preceding section, is provided by the generic connectionist classifier (see Figure 4, left). Such a classifier constitutes a convenient building block for realizing a smooth regular mapping M. The Chorus scheme is implemented by a bank of classifiers, each tuned to a particular point in the distal space (its optimal stimulus), with the response falling off gradually and monotonically with distance from the optimal point, as in the hidden-layer units in a radial basis function (RBF) approximation module (Poggio and Girosi, 1990). Intuitively, this corresponds to representing the input (say, an image of a giraffe) by similarities to a set of prototypical stored patterns (say, a camel and a leopard), simply by computing the response of a camel-recognizer and a leopard-recognizer to the input image.

[5] And, of course, within basic categories.

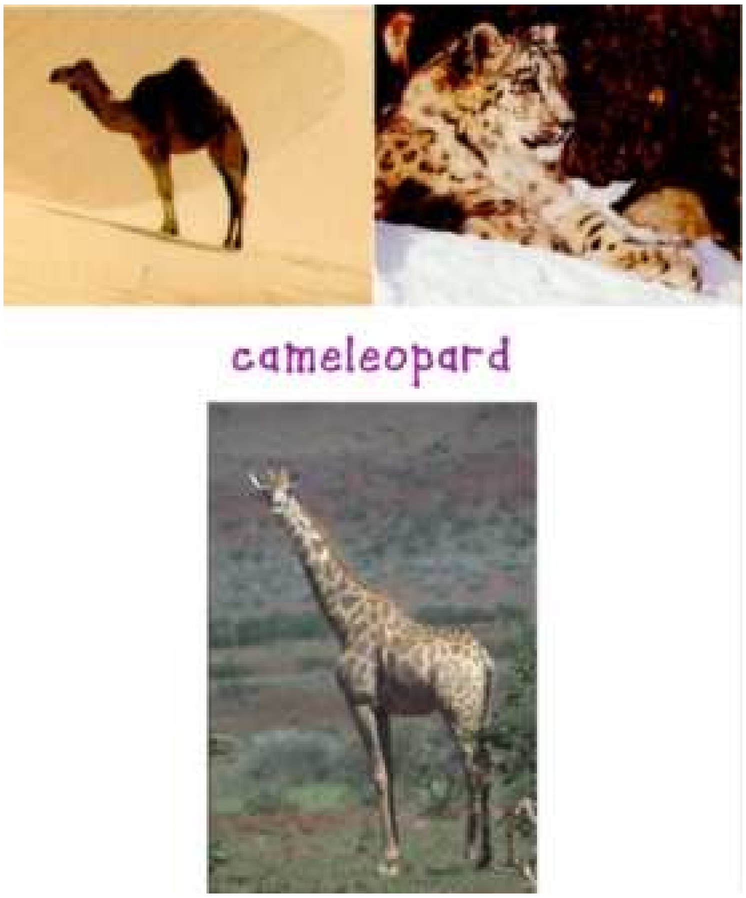

Figure 3: **Representation by similarity**: a giraffe as something of a camel and something of a leopard (see section 3).

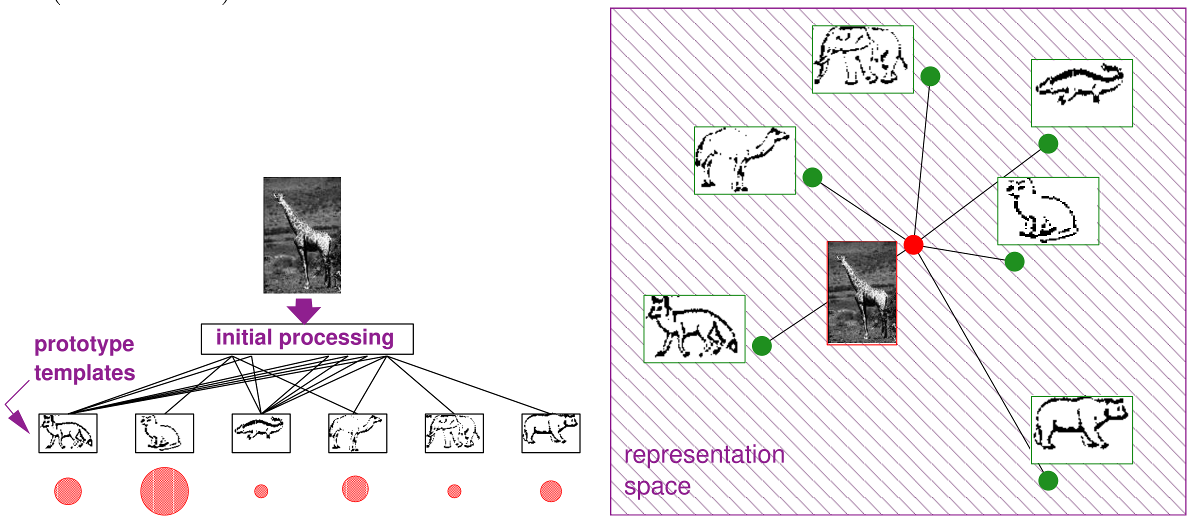

Figure 4: *Left:* A practical way of implementing representation by similarity, using a bank of object recognizers. *Right:* The outputs of recognizers correspond to distances between the stimulus and the preferred stimulus of each recognizer. Adapted from (Edelman, 1995c).

Consider the mapping $M : R^n \to R^n$ implemented by such a system. First, M is smooth, because of the graded tuning curves of the individual classifiers (typically, it is a composition of a linear mapping with a differentiable nonlinearity such as the logistic function or the Gaussian). Second, M is regular, provided that the input space is properly covered by the classifiers so as to avoid metameries (Snippe and Koenderink, 1992). Thus, according to the considerations stated in the preceding section, a bank of classifiers can serve as a substrate for locally approximately conformal representation of the distal shape space.[6]

3.2 The Chorus Transform

Let us now analyze the situation in which the number n of classifiers (i.e., prototypical patterns) is small compared to the dimensionality of any intermediate representation space that may intervene between the distal shape space and the classifier stage.[7] Let $\mathbf{p}_1, \ldots, \mathbf{p}_n$ be n prototypes and let $\mathbf{x} \in R^d$, $d \gg n$, be an input stimulus, represented by a vector in a multidimensional feature space. The Chorus Transform (CT) is defined as the mapping taking an input vector $\mathbf{x} \in R^d$ into a vector of distances between $\mathbf{x}$ and the n prototypes:

$$CT(\mathbf{x}) = \frac{1}{\sqrt{n}} \begin{pmatrix} \|\mathbf{x} - \mathbf{p}_1\| \\ \vdots \\ \|\mathbf{x} - \mathbf{p}_n\| \end{pmatrix} \tag{4}$$

We next discuss some of the useful properties of the CT mapping.

3.3 Dimensionality reduction

The dimensionality of a "default" representation of a 3D object — its image — as it confronts a visual system is equal to the number of independent measurements that the system performs on the input. If, for example, the input is a digital snapshot of an object, its dimensionality is simply the number of pixels it contains, which is likely to be high. Now, problems formulated in high-dimensional spaces tend to be computationally unmanageable — a phenomenon which came to be known as the curse of dimensionality (Bellman, 1961). In particular, learning and generalization, whether formulated in statistical terms (Stone, 1982) or in terms of training a neural network (Haussler, 1992), require too many examples if attempted in a high-dimensional space.

Dimensionality reduction has been posed in the past as a primary purpose of feature extraction (Intrator, 1992). The concept of a feature in pattern recognition usually refers to relatively simple measurements on the input. In comparison, the Chorus Transform offers a method for describing the input in terms of holistic features (here, literally, distances to whole objects). Such features have the computational advantage of being directly learnable (Edelman, 1995c). Furthermore, if the number of prototypical objects is small, CT also results in a serious reduction of dimensionality: an application of $CT : R^d \longrightarrow R^n$ maps the input into a space of dimensionality n, which is equal to the number of prototypes $\{\mathbf{p}_i\}_{i=1}^n$, $n \ll d$.

[6]It is interesting to note that better coverage of the original space is provided by a system that is not overtrained and is not forced into a winner-take-all mode (unlike the analog to symbolic transducers, a desideratum of many cognitive models).

[7]In primate vision, the million-fiber optic nerve constitutes such a high-dimensional intermediate representation, through which any information reaching the brain must pass.

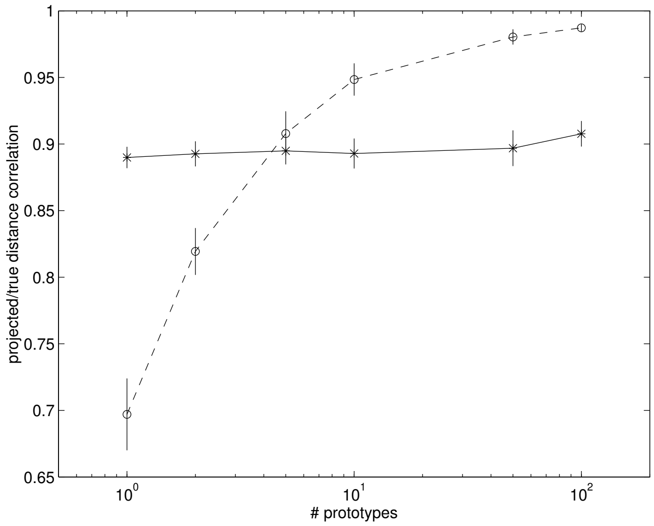

Figure 5: **Preservation of distance ranks by the Chorus Transform**. Pearson product-moment correlation (dotted line) and Spearman rank order correlation (solid line) between true and CT-transformed distances among 25 points in a space of dimensionality 1000, plotted against the number of prototypes employed by CT. Note that with as few as five prototypes good correlation (Pearson 0.908, Spearman 0.895) was attained. Error bars were calculated for 10 independent runs.

3.3.1 Viewing CT as a method for dimensionality reduction

Importantly, CT reduces dimensionality in a manner that is likely to preserve the relationships between interpoint distances in the input space. To demonstrate that, we calculated linear (Pearson) and rank order (Spearman) correlation coefficients between true distances among 25 randomly chosen points in R^{1000}, and their distances in R^n, following the application of CT, for varying number of prototypes n. The correlation coefficients were then plotted against the number of prototypes (Figure 5). Apparently, the application of CT with five prototypes (equivalent to a 200-fold dimensionality reduction with respect to the original 1000-dimensional space) yields distances correlated at about 0.9 with the distances in the original space.

3.3.2 Bounds on dimensionality reduction: the embedding theorems

To understand why it is possible to reduce dimensionality while preserving certain aspects of the metric structure of the original space, we turn to a series of recent analytical results on near-isometric embedding of metric spaces. A typical theorem from this field (Johnson and Lindenstrauss, 1984) states that, given n points $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in any Euclidean space $\mathcal{S}$ with distances $d_{ij} = d(\mathbf{v}_i, \mathbf{v}_j)$, it is possible to find n points $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ in $\mathcal{R}^t$ where $t = O(\frac{\log n}{\epsilon^2})$, such that for all i, j the distance $\delta_{ij} = d(\mathbf{w}_i, \mathbf{w}_j)$ satisfies $(1+\epsilon)\delta_{ij} > d_{ij} > \delta_{ij}$. In other words, the n points can be embedded in l_2 of $\frac{\log n}{\epsilon^2}$ dimensions with distortion $\leq (1+\epsilon)$ in the distances. Simulations described in Figure 6 were designed to visualize the results of Johnson and Lindenstrauss. According to the conditions of the theorem, $n = 100$ points were embedded in lower-dimensional spaces, for varying values of ϵ, $0 < \epsilon \ll 1$. As shown in Figure 6, left, the embedding resulted in a significant dimensionality reduction, $O(\frac{\log n}{\epsilon^2})$. The quality of the embedding (i.e., its suitability for supporting similarity-based representation) was then measured by calculating correlation coefficients between the original and the projected pairwise

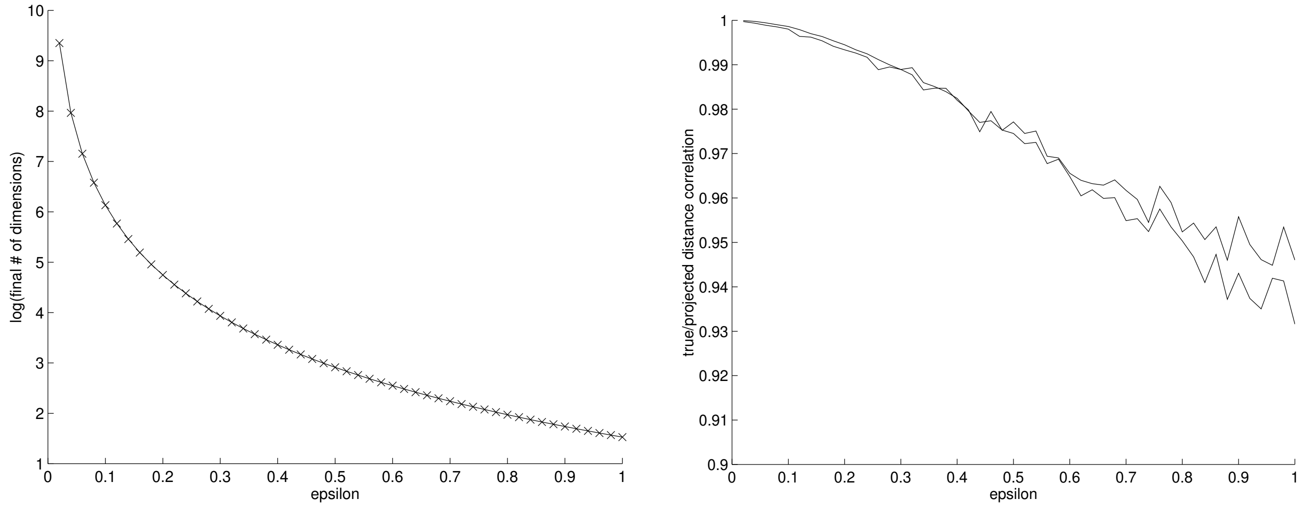

Figure 6: **Visualizing dimensionality-reducing embedding**, after (Johnson and Lindenstrauss, 1984). *Left:* The embedding of a set of $n = 100$ points results in a significant dimensionality reduction of an order $O(\frac{\log n}{\epsilon^2})$. *Right:* Pearson and Spearman (rank order) correlation coefficients for the original/projected pairwise distances among the $n = 100$ points indicate how well distances are preserved. For example, distortion of $\epsilon = 1$, allowing embedding into 5 dimensions, resulted in distance correlation of about 0.94.

distances among the $n = 100$ points. Clearly, the distances are faithfully preserved (Figure 6, right).

Note that the simulations we just described concentrated not on the isometry between original and target space under CT, but rather on the correlation between distances in the two spaces. This measure of performance of CT, dictated by the requirement for faithful representation, differs from the measure accepted in the statement of the problem of isometric embedding (where the goal is to preserve distances, not distance ranks). Indeed, if individual distances are considered in addition to the distance rank correlation, it becomes apparent that CT does distort distances (Figure 7), while preserving distance ranks to a surprising degree (Figure 8). Thus, while the existing embedding theorems provide an indirect explanation for the efficacy of CT, a new mathematical framework must be developed, if the preservation of distance ranks is to be better understood.

3.3.3 CT as an implementation of Bourgain's method

The work of Johnson and Lindenstrauss (1984) provided not only the bounds on possible dimensionality reduction, but also an algorithm for carrying it out: they showed that any orthogonal projection of the original n-dimensional space onto a random t-dimensional subspace is likely to produce a mapping with the desired embedding properties. The embedding method implemented by CT is different in that the target space is far from being a *random* subspace of the original one. To realize this, note first that a finite metric space with n points can be isometrically embedded into l_∞^n, simply by setting the coordinates of a point to the vector of its distances to all other points in the set. As shown by Bourgain (1985), a variation on this embedding method can result in logarithmic dimensionality reduction, if for each point distances to randomly picked subsets of the $n - 1$ remaining points (in CT, distances to a few prototypes) are retained instead of the distances to all the $n - 1$ other points. Using this method, an n-point metric space can be embedded into an $O(\log n)$-dimensional Euclidean space with distortion $\leq O(\log n)$ (Bourgain, 1985; Linial et al., 1994).

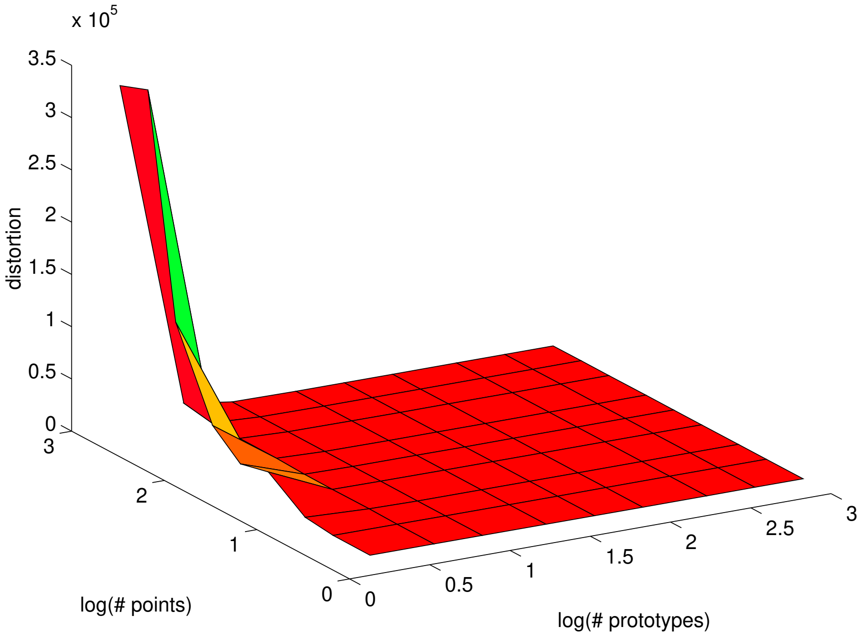

Figure 7: **Transformation of distances under the Chorus Transform**. A random set of n points in R^d, $d = 10000$, was mapped by CT into R^t, with t being the number of prototypes. The plot shows the distortion E_0 of the CT-embedding, vs. n and t. Definitions: $E = \text{argmin}\{(1+E)\delta_{ij} > d_{ij} > \delta_{ij}\}\ \ \forall i,j$, $E_0 = E/C$, where $C = \sqrt{t/n}$ is the factor prescribed by the proof of Johnson-Lindenstrauss embedding theorem; d_{ij} and δ_{ij} are, respectively, the original and the transformed distances between points i, j.

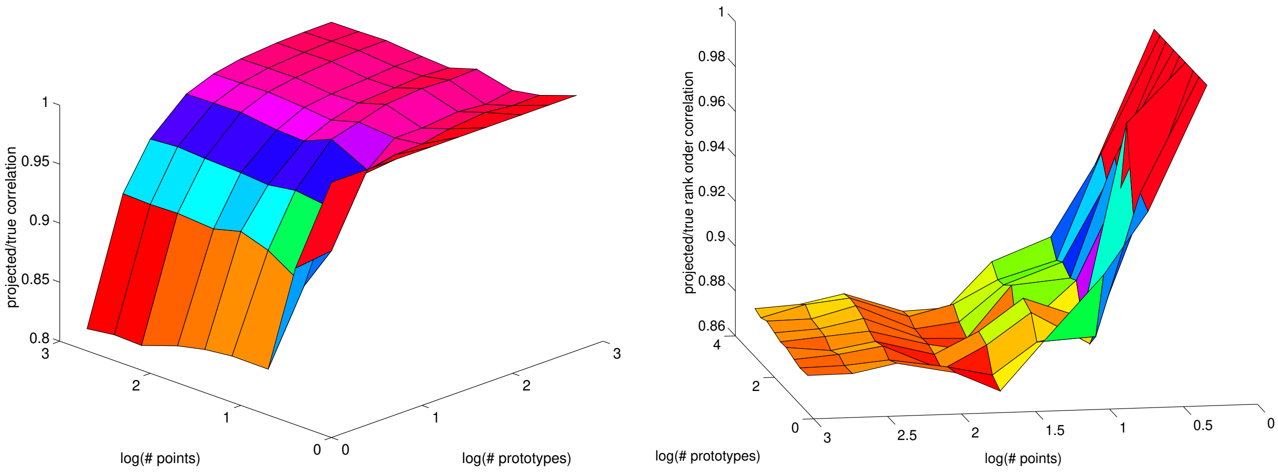

Figure 8: **Transformation of distance ranks under the Chorus Transform**. Even for relatively small values of t, the original/projected distance correlations (left) and distance rank correlations (right) are high.

3.4 CT and the Voronoi tessellation of the input space

The reliance of CT on distances to a set of points (prototypes) suggests a possible connection to the Voronoi diagram induced by the prototypes in the input space. The Voronoi diagram (Preparata and Shamos, 1985) of a set $S = \{\mathbf{p}_i\}_{i=1}^n$ is a sequence $Vor(S) = \{V(\mathbf{p}_i)\}_{i=1}^n$ of convex polyhedra covering d-space, where $V(\mathbf{p}_i) = \{\mathbf{x} \in R^d : \ d(\mathbf{x}, \mathbf{p}_i) \leq d(\mathbf{x}, \mathbf{p}_j) \ \forall j \neq i\}$, where $d(\mathbf{x}, \mathbf{p}_i)$ denotes the Euclidean distance between the points $\mathbf{x}$ and $\mathbf{p}_i$. The Voronoi diagram is a useful geometrical data structure, and is especially suitable for supporting nearest-neighbor classification (Duda and Hart, 1973).

Within the Voronoi diagram, the cell $V(\mathbf{p}_i)$ containing the i'th prototype is the set of all vectors $\mathbf{x}$ for which $\|\mathbf{x} - \mathbf{p}_i\| \leq \|\mathbf{x} - \mathbf{p}_j\|$, $\forall j \neq i$. In appendix (section B) we show that CT preserves $V(\mathbf{p}_i)$ for each i:

$$\mathbf{x} \in V(\mathbf{p}_i) \Longrightarrow CT(\mathbf{x}) \in V(CT(\mathbf{p}_i)). \tag{5}$$

3.5 Rank order of distances to prototypes

We now introduce a representation related to the Voronoi diagram, which provides more information regarding the input vector than merely the label of its nearest neighbor among the prototypes. Let $\mathbf{p}_1, \ldots, \mathbf{p}_n$ be n prototypes, and consider a representation that associates with each input $\mathbf{x}$ the rank order of its distances to prototypes (ROD). That is, an input vector $\mathbf{x}$ is represented by an ordered list of indices $ROD(\mathbf{x}) = (i_1, i_2, ..., i_n)$, which is taken to mean that among all the prototypes $\mathbf{x}$ is the most similar to $\mathbf{p}_{i_1}$, then to $\mathbf{p}_{i_2}$, and so on. Note that in a list corresponding to a prototype $\mathbf{p}_i$, the index i always serves as the first element (a prototype is most similar to itself).

The total number of distinct representations under the ROD scheme is $n!$ (the number of permutations of the n indices). To compare two representations, one may use Spearman's rank order correlation of the index lists. A short form of Spearman's rank correlation coefficient formula (which can be reliably employed when the total number of tied measurements is not large relative to n) is $r_s = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$, where d_i are the differences in the ranks of the i^{th} component of the vectors to be correlated (Mendenhall and Sincich, 1988). In that formula, only the number of mismatches in ranks and their magnitude enter the calculation. In our case, however, we must take into account the positions of the mismatches as well. For example, assume that we have the following three rank order vectors: $q_1 = (1,2,3)$, $q_2 = (2,1,3)$ and $q_3 = (1,3,2)$. The difference between q_1 and q_2 is in their first and second-largest rank order components (difference of one position), whereas q_1 and q_3 have the same largest rank order component, but differ in their second and third components (again, difference of one position). The difference vectors are: $q_1 - q_2 = (1,1,0)$, $q_1 - q_3 = (0,1,1)$, therefore, their Spearman's rank correlation coefficient will be the same, although intuition seems to require that they be different. Position of mismatched ranks can be taken into account using the following variation of Spearman's formula: $r_s = 1 - \frac{6\Sigma(n-i+1)d_i^2}{\frac{1}{2}n(n+1)(n^2-1)}$, or, $r_s = 1 - \frac{6\Sigma w_i d_i^2}{(n^2-1)}$, where $w_i = \frac{(n-i+1)}{\frac{1}{2}n(n+1)}$, $\Sigma_{i=1}^n w_i = 1$, and d_i are as before. In the above example, the modified formula yields $r_s(q_1, q_2) = 0.375$, compared to $r_s(q_1, q_3) = 0.625$. Thus, a shift in the first two components is given more weight than the same shift in two lower-order components.

To realize the connection between the ROD representation and the Voronoi diagram, note that the leading element i_1 of $ROD(\mathbf{x})$ determines the Voronoi cell $V(\mathbf{p}_{i_1})$ to which the vector $\mathbf{x}$ belongs. Each of the subsequent elements of $ROD(\mathbf{x})$ adds more information regarding the location of $\mathbf{x}$ with respect to the prototypes, resulting in a finer subdivision of the Voronoi cell $V(\mathbf{p}_{i_1})$.

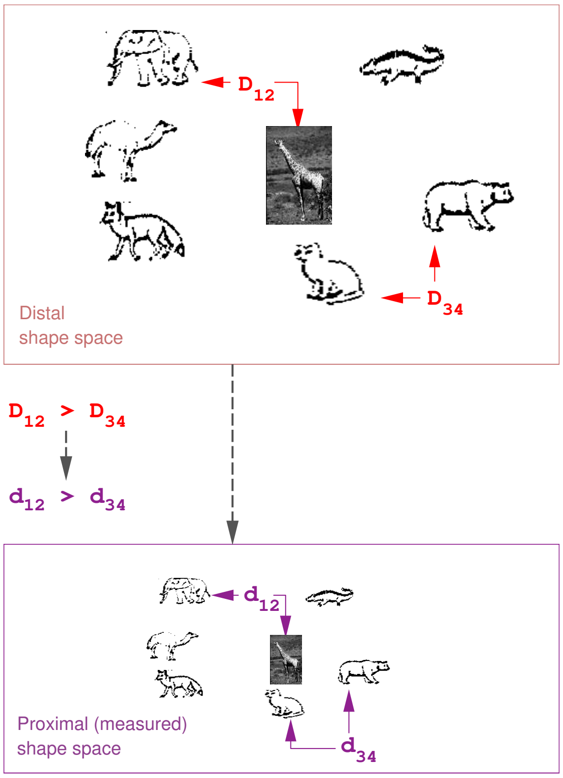

Figure 9: Faithful representation can be realized by a proximal-to-distal mapping that preserves distance ranks (see section 4).

4 Satisfaction of monotonicity requirements for different classes of objects

Any scheme relying on the idea of second-order isomorphism between the world and the pattern of representations (including Chorus) requires that represented distances (in Chorus, the distances to prototypes) covary monotonically to the dissimilarities between the corresponding entities in the world. As we have argued in the preceding sections, it would suffice for the proximal-to-distal mapping M to be locally (quasi)conformal. In visual perception, M is composed of two mappings: M_1 that maps the shape onto the sensorium (the retina of an eye, or the detector array of a camera), and M_2 that is realized by the perceptual system proper. It has been pointed out in the past that singularities in M_1 (such as accidents of projection, illumination, etc.) may pose a problem for distance-based recognition systems by mapping disparate objects into similar images (Adini et al., 1993). Recent experiments with human face images indicate that this situation can be amended to a considerable extent by the action of an appropriately devised mapping M_2 (Lando and Edelman, 1995); in some cases, simulating the action of the initial stages of primate vision results in satisfactory remapping of the representational space (Weiss and Edelman, 1995).

Given the distance rank preservation properties of the Chorus Transform, the practical value of representation by second-order isomorphism depends on the possibility to demonstrate that the stages preceding CT (that is, a composition of the imaging stage with the initial stages of vision) also preserve distance ranks, for a wider variety of objects than human faces.[8] In this section, we examine the degree to which distance ranks are preserved by imaging and by transduction with Gaussian receptive fields. We consider in turn objects that are collections of points in 3D (section 4.1) and computer-generated objects that possess extended surfaces (section 4.2).

4.1 Point objects

Consider objects that are collections of a fixed number of points in 3D (see Figure 10, left), and assume that point correspondence for any two objects, or any two views (2D projections) of objects is given. Let each object starts rotating away from some reference orientation in 3D. A natural objective distance can then be simply the angle of rotation, or, alternatively, distance computed according to some 3D point-based metric such as the Hausdorff metric (Figure 10, center), or the Euclidean metric.[9] Obviously, the distance between a rotated view of the object and the reference view need not covary monotonically with the objective distance (if only because of the non-monotonicity of projected lengths of segments connecting pairs of points). One may observe (Figure 10, right), however, that projected *ratios* of segment lengths do tend to stay close to the true ratios (Burns et al., 1993). Thus, one may hope that the set of all pairwise projected distances will carry sufficient information to allow the recovery of the configuration of the points in 3D. Apparently, Hausdorff distance between projections (2D images) of a reference and rotated point objects is approximately monotonic with distal measurements (Figure 10, middle). Moreover, distance ranks are preserved under (CT) projection from 3D to 2D (Figure 10, right).

4.2 3D shapes

To investigate the behavior of shape-space distances for objects composed of surfaces rather than points, we need a mechanism for transducing images of arbitrary shapes into a convenient vector-space form. The particular mechanism we used — receptive fields — is described below.

4.2.1 Image transduction by receptive fields

Receptive fields (RFs) are probably the most prominent and ubiquitous computational mechanism employed by biological information processing systems. A collection of graded and highly overlapping RFs can support fine spatial discrimination (Snippe and Koenderink, 1992), including hyperacuity-level resolution (Weiss et al., 1993) and discrimination among highly complex stimuli such as images of human faces (Edelman et al., 1992). The results of the simulations described in this section illustrate another aspect of the computational power of simple graded-profile RFs: changes in 3D object appearance due to gradual changes in viewing position are reflected monotonically in the high-dimensional space spanned by the activities of the RFs.

In the simulations, 3D objects rotating in space were rendered using the SGI Open Inventor graphics toolkit (see Figure 12). The images were then transduced using a set of 200 filters with radially elongated Gaussian receptive fields, randomly positioned over the image (Weiss and Edelman, 1995);

[8] Such a demonstration would substantiate Dretske's remark that sufficiently complex systems relying on representation by covariation can reflect properties of distal objects, and not only of the objects residing in the system's sensorium (Dretske, 1981).

[9] Note that the Hausdorff metric does not require the knowledge of point-to-point correspondence between the objects, but is sensitive to the omission of some of the points, such as caused, e.g., by occlusion, and to the presence of outliers, or points wrongly attributed to a given object. The recovery of 3D shape from distances among its multiple views using MDS is, therefore, one of the few methods for building 3D object representations that does not rely on point correspondence.

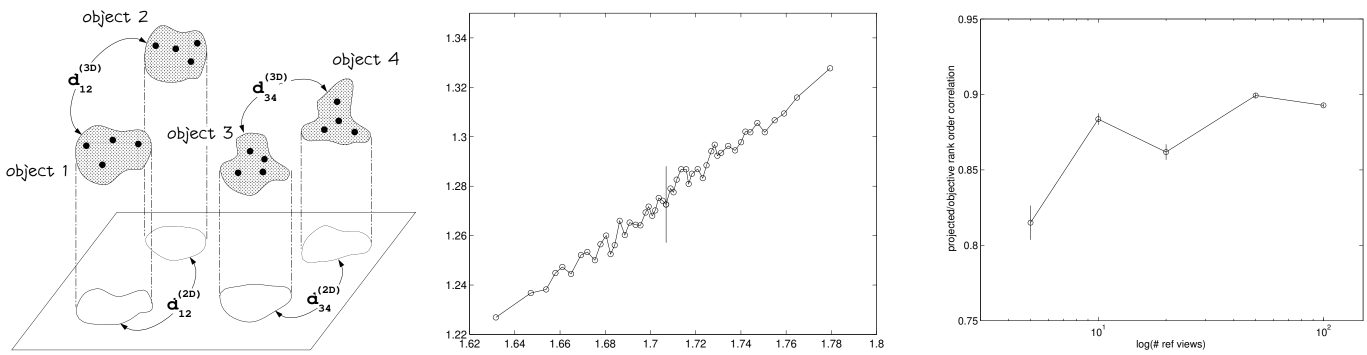

Figure 10: **Point objects in 3D; view space**. *Left:* a symbolic depiction of two pairs of objects (point clouds in 3D; points uniformly distributed between -0.5 and 0.5), and their 2D projections. In the simulations described here, the centers of mass of the objects were assumed to coincide (that is, differences in shape, not location, were considered). *Center:* behavior of different views of the same object. Hausdorff distance between projections of a reference cloud and a rotated cloud is approximately monotonic with the Hausdorff distance between the reference and the rotated clouds in 3D. Error bar represents the standard error of the mean, computed for 250 views (only the meam error bar is shown). *Right:* Preservation of similarity ranks under 2D projection, and an application of CT. Spearman rank correlation between objective and projected distances, vs. the number of reference views used by CT. The error bars are computed over 25 runs for each set of reference views.

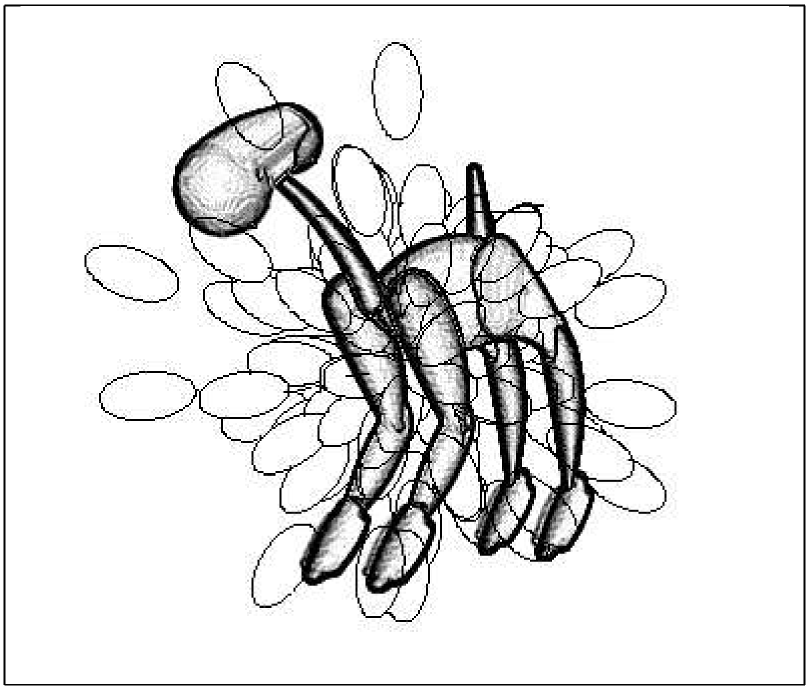

Figure 11: An image of a 3D object, overlayed by RFs used in the transduction stage (see section 4.2.1).

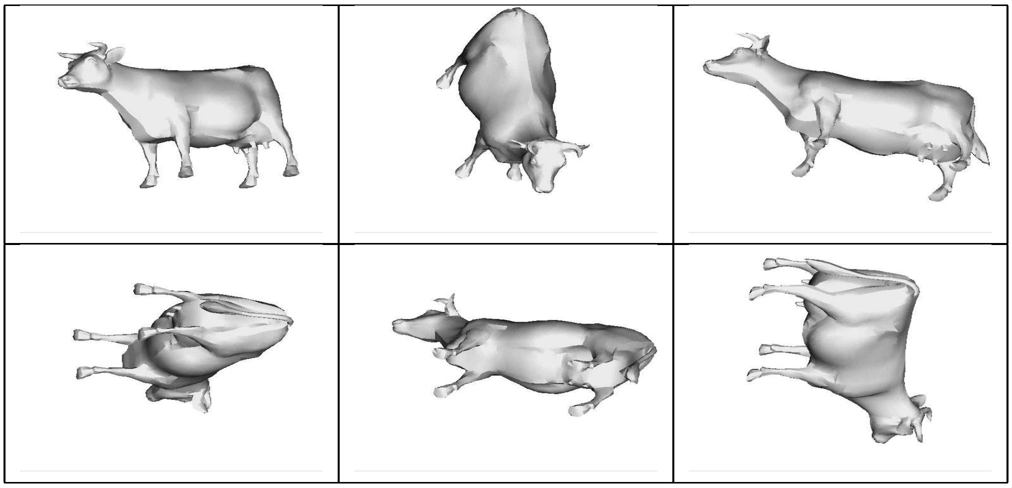

Figure 12: A series of images of a 3D object (cow), taken from different viewing positions.

see Figure 11. Thus, different object appearances corresponding to successive orientations were represented by vectors in a 200-dimensional Euclidean space, R^{200}, of the RF activities. To visualize the arrangement of views of an object in the R^{200} space, pairwise distances were calculated among vectors of RF activities corresponding to views that differed by $10°$ rotation steps around a fixed axis. Those distances were then submitted to nonmetric multidimensional scaling, or MDS (Kruskal and Wish, 1978; Sas, 1989). The successful recovery of the gradual view change by MDS (Figure 13) indicates that an RF-based representation can reflect faithfully the local relationships among different views of an object.

4.2.2 View space

Our next series of simulations were designed to determine whether RF transduction preserves the metric structure of the view space of individual 3D objects. Faithful representation of the relationships between different views of the same object requires that the RF-space distance between views covary monotonically with the objective view-space distance. We define the distance between two views of an object as the shortest-path angular rotation transforming one view into the other.[10] The simulations we describe next involved 75 object models, some of which are shown in Figure 14. Images of different views (as in Figure 12) were transduced into a 200-dimensional Euclidean space, using radially elongated Gaussian receptive fields, described in section 4.2.1. The results (see Figure 15, top) indicate that, up to a point, distance in the RF space covaries monotonically with the angular rotation distance between object views. Figure 15, bottom, shows the mean deviation from monotonicity for each object. The "elbow" in each plot corresponds to the minimal rotation for which monotonicity no longer holds for that object. On the average, the monotonicity was found to hold for rotation distance of up to $\theta = 126°$, for the 75 objects used in the experiment.
We next show the proximal-to-distal behavior of views of a 3D object, for randomly chosen angular rotation distances in the $[0, \pi)$ range, around a random axis of rotation. The results (Figure 16, left) indicate that RF-space distance generally covaries monotonically with angular rotation distance between

[10]The angular rotation distance is the angle α for which $1 + 2\cos\alpha = \mathrm{trace}(M)$, where M is the rotation matrix transforming one view into the other (see (Kanatani, 1990), p.205); other measures of distance can be used as well.

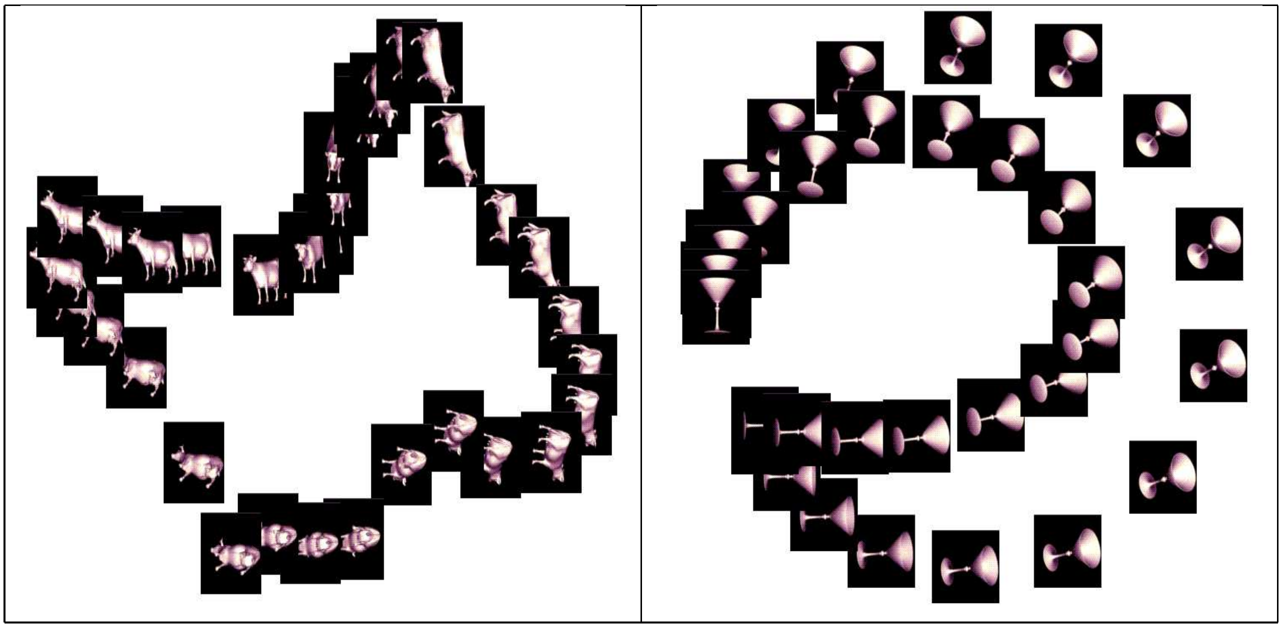

Figure 13: **The arrangement of points corresponding to successive views of 3D objects**. *Left:* views of a cow - the *Badness-of-Fit* criterion (Kruskal, 1964) is 0.08. *Right:* views of a martini - the *Badness-of-Fit* is 0.03.

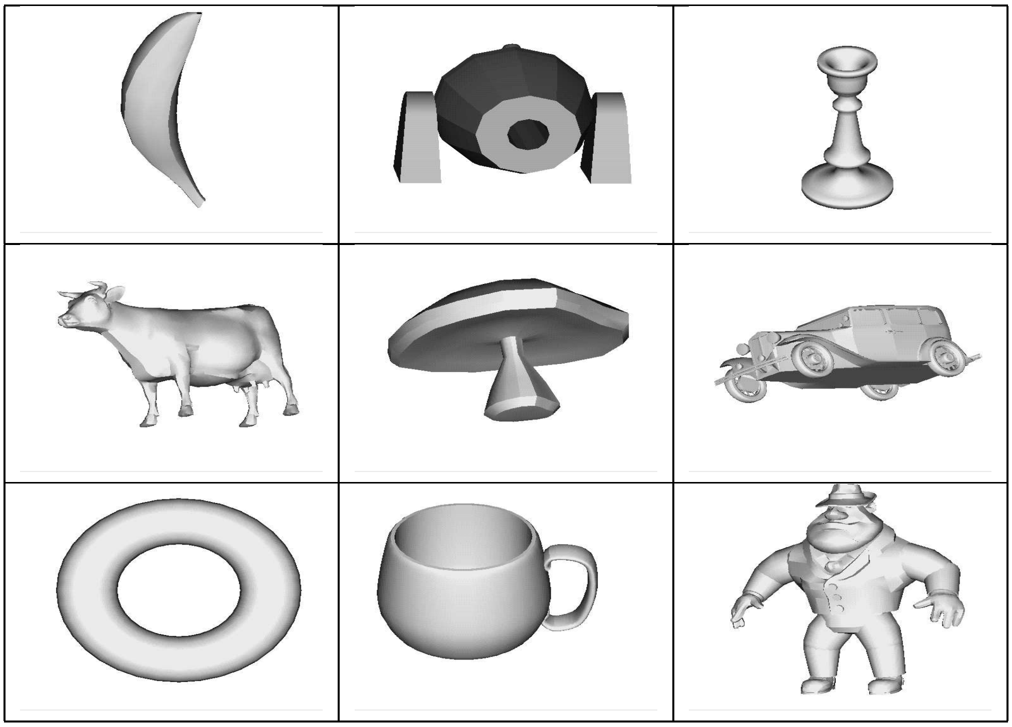

Figure 14: Images of some of the 3D object models used in our simulations. The models are available electronically, at URL ftp://avalon.vislab.navy.mil/pub/objects/.

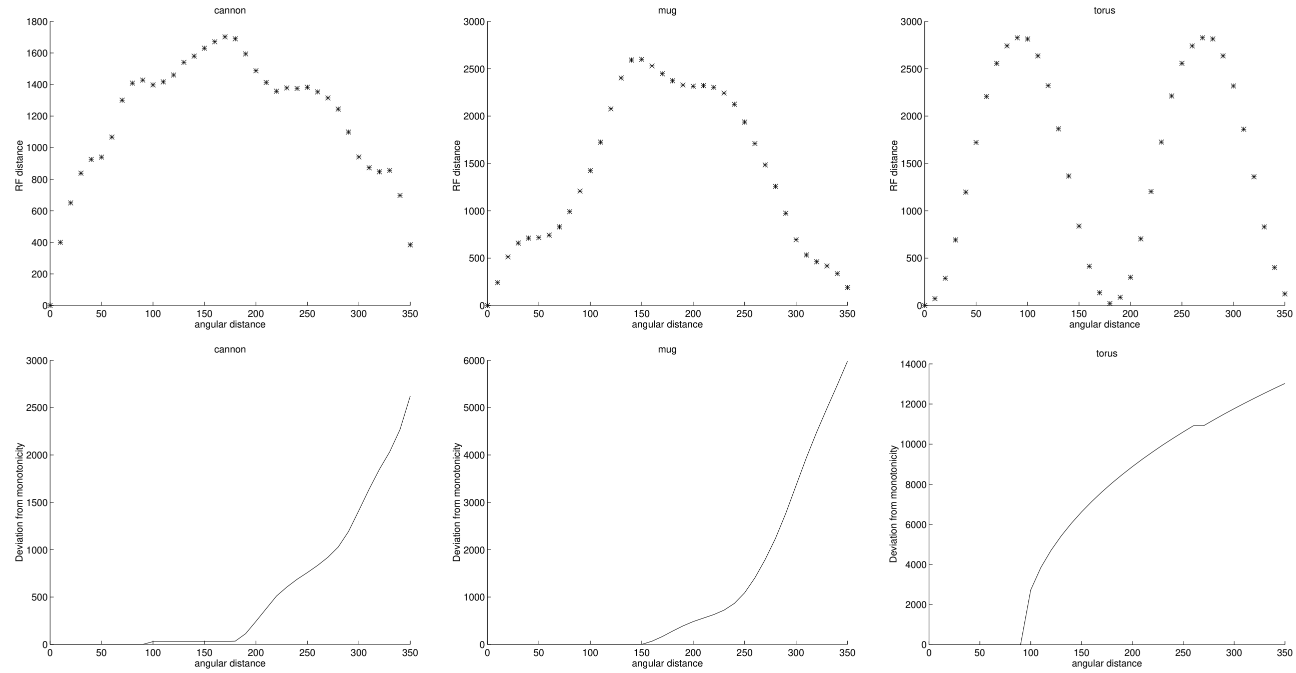

Figure 15: **Monotonic behavior and deviation from monotonicity of 3D object views**. *Top:* Objective and RF-space projective distances for three of the 3D object shown in Figure 14 (left: cannon, middle: mug, right: torus). *Bottom:* The deviation from monotonicity for each of these objects. Note that the torus object yields a fluctuating objective/projective distance curve, resulting in a larger total deviation from monotonicity.

object views. The preservation of distance ranks under 3D-object rotation, RF-space transduction, and an application of CT is shown in Figure 16, right. Spearman rank order correlation between rotation-space distances (from a reference to a test view) and the corresponding RF-space distances was calculated for a varying number of reference views (used by CT). As expected from a basic claim on isometric embedding of finite metric spaces[11], Spearman rank correlation is improved with the number of reference points employed by CT.

The results of this exploration of the view space can be summarized as follows. First, the RF-space image of an object rotating in 3D changes in a smooth and gradual manner (Figure 13). Second, the RF-space distance relative to a reference orientation grows monotonically, up to a certain orientation, which is usually tens of degrees away from the reference one, but can be closer in certain degenerate cases. The most degenerate case in this respect is that of rotational symmetry with respect to a point: a spheroidal object appears the same from all viewpoints. Finally, distance ranks are preserved under view-space rotation, RF-space transduction and CT application.

[11] Any finite set of n points can be isometrically embedded into l_∞^n, simply by setting the coordinates of a point to the vector of its distances to all other points in the set (Linial et al., 1994).

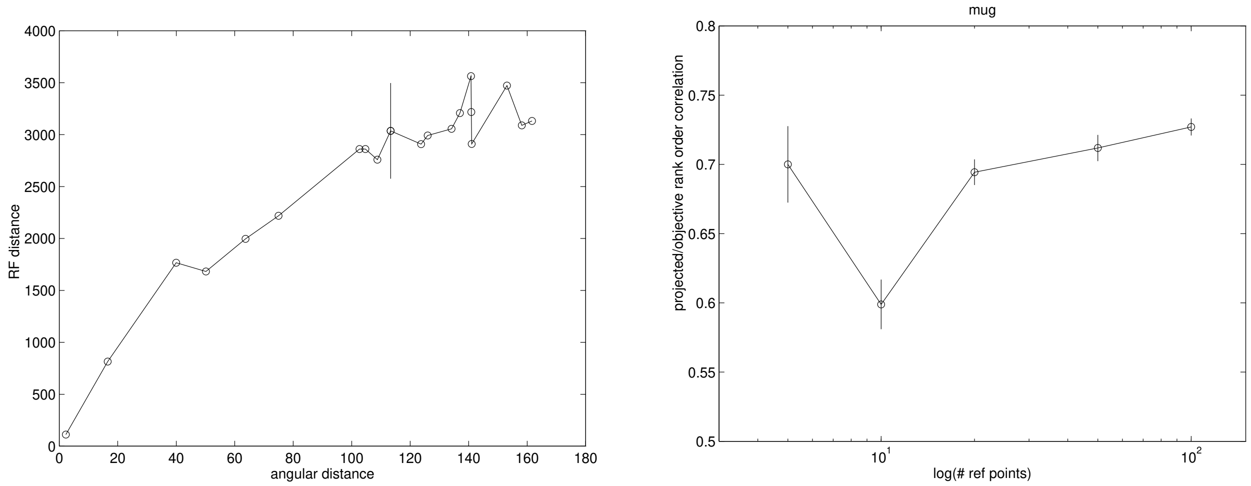

Figure 16: **3D objects; view space**. *Left:* RF-space distance between a canonical and test images of a mug, plotted against objective (rotation-space) distance between the corresponding views. The error bar denotes the mean standard error of the mean, calculated for 10 random views. *Right:* Preservation of similarity ranks under rotation, RF-space transduction, and CT application. The plot shows Spearman correlation between rotation-space distances (from a canonical to a test view) and the corresponding RF-space distances. CT used a varying number of reference views (shown on the abscissa). Error bars are calculated over 25 runs with randomly chosen reference points.

Figure 17: Seven different animal-like objects arranged on a line in the R^{70} parameter space; see (Cutzu and Edelman, 1995), for details. The intermediate shapes were constructed from the two extreme one, A and B, using parameter-space interpolation: $M = \alpha A + (1 - \alpha)B$, $0 \leq \alpha \leq 1$, for equally spaced values of α.

4.2.3 Shape space

We next consider the question of preservation of distance ranks by RF transduction under deformation in shape space, rather than rotation in view space. To explore shape space, we employed parameterized animal-like objects, courtesy of F. Cutzu (Cutzu and Edelman, 1995). Each object was encoded by a vector of 70 parameters (a point in R^{70}); the Euclidean distance between two such points is a natural measure of dissimilarity between the two corresponding shapes. Seven objects were employed, for which the representation vectors were arranged on a line in the R^{70} parameter space (see Figure 17). RF-space distances between the shapes were computed and submitted to the MDS procedure. Each object appeared at one of four viewpoints, corresponding to the vertices of a regular tetrahedron centered on the object.

In the configuration recovered by MDS, corresponding views of the different objects clustered together, as can be seen in Figure 18 (cf. Lando and Edelman, 1995). It may be noted that RF-space distances, calculated between images of six test shapes and a reference one, covaried monotonically with the parameter-space distances among the corresponding shapes (Figure 19). The comparison, carried out for each of the four different views separately, is justified by the clustering-by-view observation.

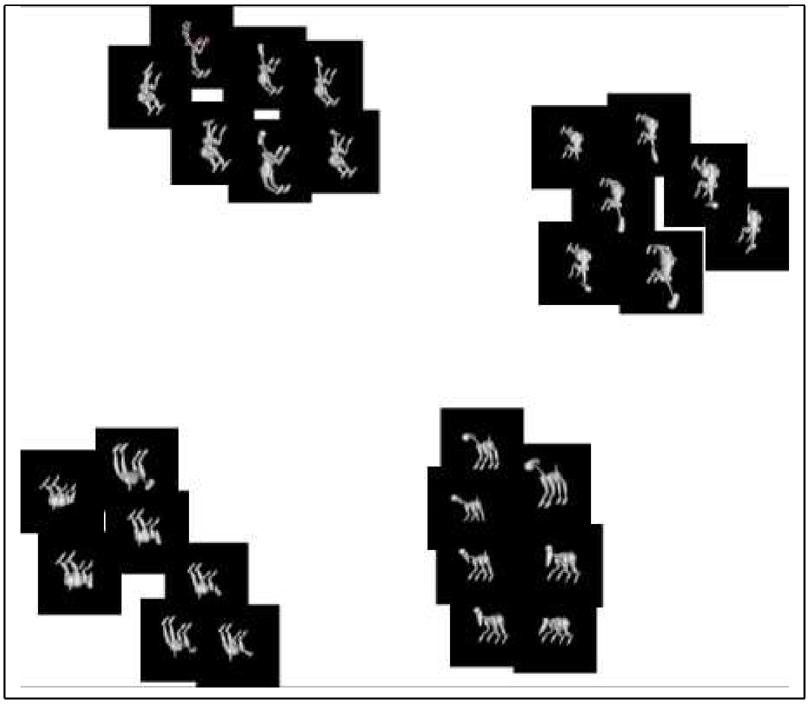

Figure 18: An MDS layout of 4 different views of the 7 animal-like objects from Figure 17. Note that corresponding views cluster together. This implies that distances measured among corresponding views are smaller than distances among arbitrarily chosen views.

To generalize these monotonicity results, the following series of simulations have been devised. In the simulations, animal shapes were deformed randomly in a small hyper-cube in the parameter space (rather then being deformed along a line). Pairwise distances among the different shapes were calculated both in RF and parameter space. The observed proximal-to-distal behaviour (Figure 20, left) indicates that even for arbitrary shape deformation, distances among RF-transduced images covary monotonically with objective shape distances. Spearman correlation calculated for distances between deformed shapes and their corresponding RF-space distances (for a varying number of reference views used by CT, as in Figure 16) demonstrates the rank preservation under 3D-object deformation, RF-space transduction, and CT application. As noted in Section 4.2.2, regarding Figure 16, Spearman rank correlation is improved with the number of reference points employed by CT, as expected.

The implications of the above findings for representing 3D shapes by RF-space similarities to prototypes are discussed below.

5 Discussion

5.1 Summary

In this report we have attempted (1) to characterize mathematically the intuitive notion of representation that is formally faithful to the universe of entities it stands for, and (2) to evaluate computationally the properties of a particular representational scheme, the Chorus of Prototypes. We found that representation fulfilling a condition of second-order isomorphism to the shape space (in the sense of Shepard,

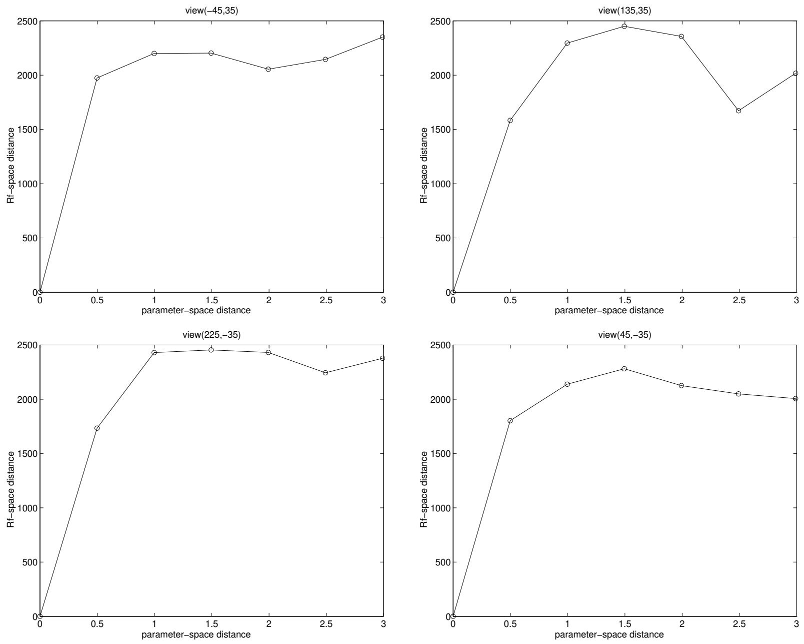

Figure 19: **Monotonicity of proximal (RF-space) measurements with respect to distal measurements made in the parameter space**. For 4 different views, the RF-space distances between 6 test objects and a reference object are plotted against the Euclidean distance of the corresponding points in the parameter space. The separate treatment of each view is justified, because of the clustering by view, apparent in Figure 18.

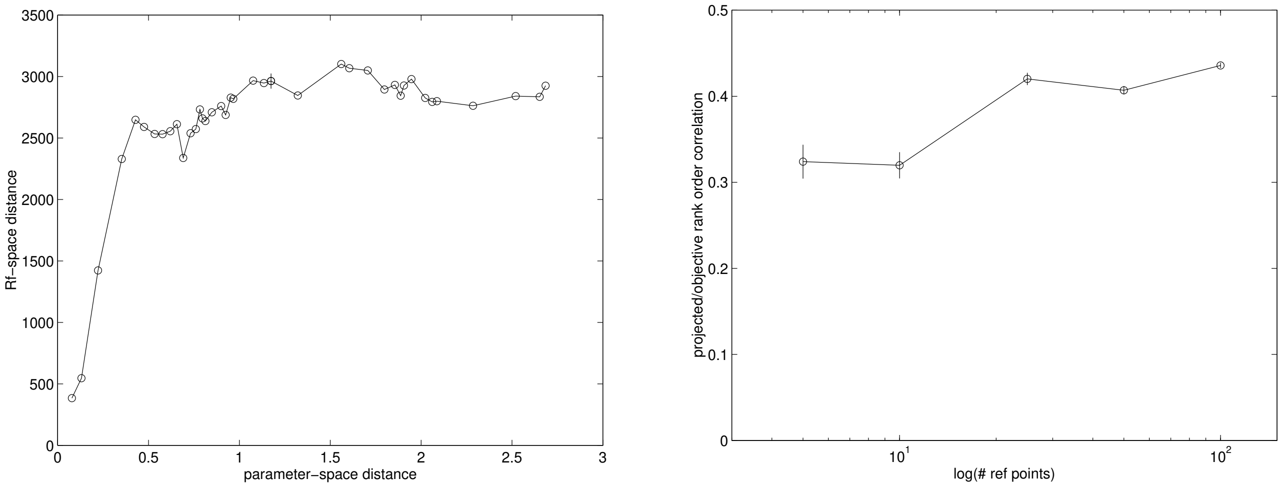

Figure 20: **3D objects; shape space**. *Left:* RF-space distance between pairs of 3D objects (parameterized animal-like shapes, courtesy of F. Cutzu), plotted against objective distance between the corresponding points in the parameter space. Error bar denotes the mean standard error of the mean, calculated for 25 randomly chosen pairs. *Right:* Preservation of similarity ranks under deformation, RF-space transduction, and CT application. Spearman correlation is plotted against the number of reference points used by CT. Error bars are calculated over 25 runs with randomly chosen reference points.

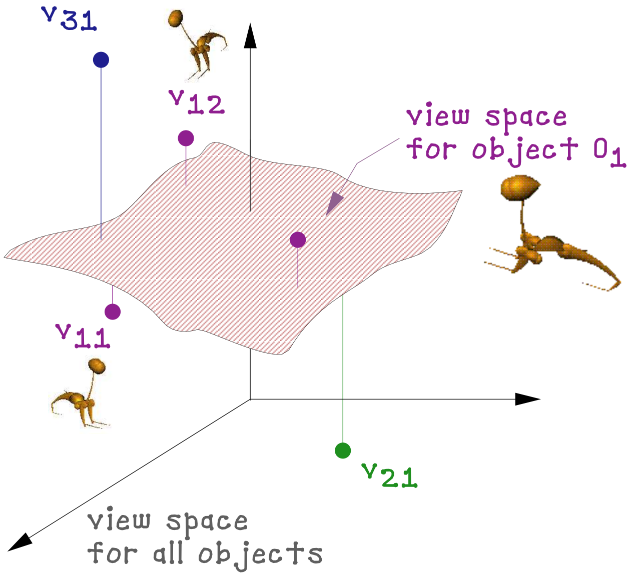

Figure 21: Distance between an input view and the *view space* spanned by a number of "snapshots" of a particular object ($\mathbf{v}_{1i}$). This distance can serve as the basis for object recognition (Ullman and Basri, 1991).

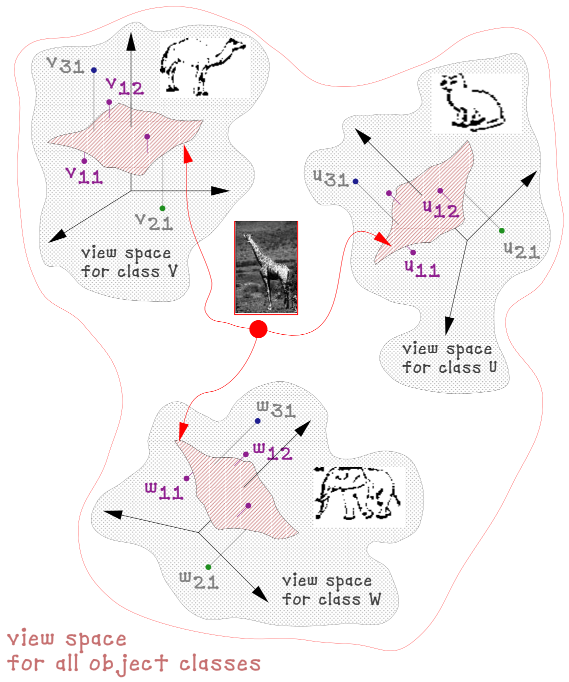

Figure 22: Distance between an input view and a number of objects in a *shape space*. Note that the distances here are to be computed to the view space of each reference object, and not to the individual views (unless view-to-view distance covaries monotonically with view-to-viewspace distance, in which case the former is an acceptable approximation of the latter).

1968) can be considered faithful, provided that the mapping between a parameterization of the (distal) shape space and the internal (proximal) representation space preserves distance ranks. It turned out that a proximal-to-distal mapping realized by a bank of typical connectionist classifiers, working in parallel as in the Chorus scheme, is likely to satisfy the requirements for distance rank preservation over appropriately limited regions of the distal space. Furthermore, results of simulations conducted with a variety of shapes indicate that the extent of a region of rank preservation is typically large enough to satisfy practical needs of classification and recognition. Thus, a shape-space analog of the multiple-view representation scheme (see Figures 21 and 22) appears to be both mathematically appealing and computationally feasible. In the remainder of this section, we cite psychophysical support for the idea of parametrically faithful representation, discuss a number of open issues, and list directions for future research.

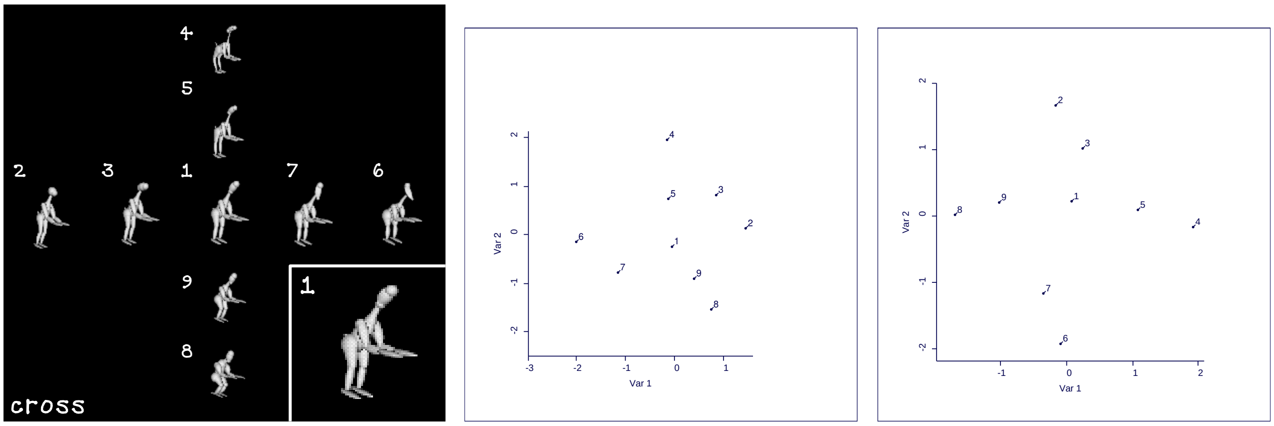

Figure 23: *Left:* the cruciform parameter-space configuration used for generating the stimuli in one of the experiments described in (Cutzu and Edelman, 1995). The inset shows an enlargement of one of the objects. *Middle:* the 2D MDS solution for subject VAL (badness-of-fit 0.04, distance correlation 0.99). The labeling of the individual points is the same as in the diagram on the left. *Right:* the 2D MDS solution for all subjects (badness-of-fit 0.14, distance correlation 0.92). Note the good recovery of the parameter-space configuration from the subject data.

5.2 Faithful parametric representation in psychophysics

A number of recent works have explored the veridicality of parametric shape representation using MDS (Edelman, 1995b; Cutzu and Edelman, 1995). In each of a series of experiments, which involved pairwise similarity judgment, delayed match to sample, and long-term recognition memory, subjects were confronted with several classes of computer-rendered 3D animal-like shapes, arranged in a complex pattern in a common high-dimensional parameter space. Response time and error rate data were combined into a measure of subjective shape similarity, and the resulting proximity matrix was submitted to nonmetric MDS. In the resulting solution (see Figure 23), the relative geometrical arrangement of the points corresponding to the different objects invariably reflected the complex low-dimensional structure in parameter space that defined the relationships between the stimuli classes, as predicted by the idea of second-order isomorphism in representation.

5.3 The possibility of different parameterizations

In discussing representation, the idea of faithfulness to "the original" can be easily confused with extreme Platonism, if it is cast in terms of the recovery of a unique true parameterization of shapes. It should, therefore, be clarified that veridicality does not require or entail uniqueness of parameterization of the shape space. Consider the set of all parameterizations $\mathcal{P}$ related to a given one p_0 by some conformal mapping T. Obviously, $\mathcal{P}$ is an equivalence class; moreover, because the composition $T \circ M$ is conformal if M is, veridical representation of some $p \in \mathcal{P}$ is equivalent to the representation of any other $p \in \mathcal{P}$.

Now, a conformal mapping M will give rise to a proper (i.e., second-order isomorphic) representation of object clustering under all parameterizations belonging to *some* class $\mathcal{P}_x$. The nature of that class will depend on the nature of the mapping (which can emphasize some distances among objects at the expense of others, with or without altering the distance ranks). A system that is a product of natural selection is expected to have evolved a mapping better suitable for the representation of those aspects of its habitat most important for its survival and behavior. Thus, while faithful representation is possible, it is also possible that two perceptual systems implementing different mappings will have incompatible (or even conflicting) pictures of the world. Two examples illustrating this point are, on the one hand, the apparent inability of pigeons to learn the distinction between scrambled and intact cartoon characters (Cerella, 1987), and, on the other hand, the striking similarity between patterns of internal representation of shape, recovered using multidimensional scaling from chimpanzee and human behavioral data (Tomonaga and Matsuzawa, 1992).

5.4 Future work

We conclude this work by listing a number of directions in which it can be extended.

Characterization of embeddings that preserve distance ranks. Mathematical investigations of mappings with bounded distortion (Reshetnyak, 1989) and of nearly isometric dimensionality-reducing embeddings (Linial et al., 1994) are all formulated in terms of distance preservation. It would be interesting to see whether useful results can be derived under the constraint of distance *rank* preservation, which is directly relevant to the issue of faithful representation.

Choice of prototypes in Chorus. The constraint of distance rank preservation can be applied not only to the characterization of general properties of perceptual mappings, but also to the derivation of parameter values for specific representational schemes that rely on similarity to prototypes. In particular, the problem of optimal choice of prototype locations in Chorus can be given a computational formalization, and a solution can then be sought by standard optimization techniques.

Psychophysics of shape representation. Psychophysical investigations constitute an important source of support for theories of representation. One of our central aims is, therefore, to construct a computational model replicating the psychophysical findings of (Cutzu and Edelman, 1995). A working model of this kind would elucidate the mechanisms that enable human subjects to represent faithfully the complicated geometrical arrangement of 3D shapes in parameter space, leading to a better understanding of an existing solution to the Problem of Representation.

Acknowledgments

We thank Ad Aertsen, Ronen Basri, Florin Cutzu, Nathan Intrator, Nathan Linial, Gideon Schechtman, and Shimon Ullman, for many useful discussions.

References

Adini, Y., Moses, Y., and Ullman, S. (1993). Face recognition: the problem of compensating for changes in illumination direction. CS-TR 93-21, Weizmann Institute of Science.

Bellman, R. E. (1961). *Adaptive Control Processes*. Princeton University Press, Princeton, NJ.

Bourgain, J. (1985). On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52:46–52.

Bülthoff, H. 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.

Burns, J., Weiss, R., and Riseman, E. (1993). View variation of point-set and line segment features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:51–68.

Cerella, J. (1987). Pigeons and perceptrons. *Pattern Recognition*, 19:431–438.

Cohn, H. (1967). *Conformal mappings on Riemann surfaces*. McGraw-Hill, New York.

Cummins, R. (1989). *Meaning and mental representation*. MIT Press, Cambridge, MA.

Cutzu, F. and Edelman, S. (1994). Canonical views in object representation and recognition. *Vision Research*, 34:3037–3056.

Cutzu, F. and Edelman, S. (1995). Explorations of shape space. CS-TR 95-01, Weizmann Institute of Science.

Dretske, F. (1981). *Knowledge and the flow of information*. MIT Press, Cambridge, MA.

Duda, R. O. and Hart, P. E. (1973). *Pattern classification and scene analysis*. Wiley, New York.

Edelman, S. (1995a). Class similarity and viewpoint invariance in the recognition of 3D objects. *Biological Cybernetics*, 72:207–220.

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

Edelman, S. (1995c). Representation, Similarity, and the Chorus of Prototypes. *Minds and Machines*, 5:45–68.

Edelman, S. and Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of 3D objects. *Vision Research*, 32:2385–2400.

Edelman, S., Reisfeld, D., and Yeshurun, Y. (1992). Learning to recognize faces from examples. In Sandini, G., editor, *Proc. 2nd European Conf. on Computer Vision, Lecture Notes in Computer Science*, volume 588, pages 787–791. Springer Verlag.

Edelman, S. and Weinshall, D. (1991). A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics*, 64:209–219.

Goebel, K. and Kirk, W. A. (1990). *Topics in metric fixed point theory*. Number 28 in Cambridge studies in advanced mathematics. Cambridge Univ. Press.

Goldstone, R. L. (1994). The role of similarity in categorization: providing a groundwork. *Cognition*, 52:125–157.

Haussler, D. (1992). Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, 100:78–150.

Humphrey, G. K. and Khan, S. C. (1992). Recognizing novel views of three-dimensional objects. *Can. J. Psychol.*, 46:170–190.

Intrator, N. (1992). Feature extraction using an unsupervised neural network. *Neural Computation*, 4:98–107.

Johnson, W. B. and Lindenstrauss, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, 26:189–206.

Jolicoeur, P. and Humphrey, G. K. (1994). Perception of rotated two-dimensional and three-dimensional objects and visual shapes. In Walsh, V. and Kulikowski, J., editors, *Perceptual constancies*, chapter 10. Cambridge University Press, Cambridge, UK. in press.

Kanatani, K. (1990). *Group-theoretical methods in image understanding*. Springer, Berlin.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27.

Kruskal, J. B. and Wish, M. (1978). *Multidimensional Scaling*. Sage Piblications, Beverly Hills, CA.

Lando, M. and Edelman, S. (1995). Generalization from a single view in face recognition. CS-TR 95-02, Weizmann Institute of Science.

Linial, N., London, E., and Rabinovich, Y. (1994). The geometry of graphs and some of its algorithmic applications. *FOCS*, 35:577–591.

Manolache, F. and Edelman, S. (1993). Generation of natural-looking 3D shapes by simulated evolution. CS-TR 93-13, Weizmann Institute of Science.

Marr, D. (1982). *Vision*. W. H. Freeman, San Francisco, CA.

Mendenhall, W. and Sincich, T. (1988). *Statistics for the engineering and computer sciences*. Macmillan, London.

Nicod, J. (1930). *Foundations of Geometry and Induction*. Routledge & Kegan Paul.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14:700–708.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In Rosch, E. and Lloyd, B. B., editors, *Cognition and Categorization*, pages 259–303. Erlbaum, Hillsdale, NJ.

Poggio, T. (1990). A theory of how the brain might work. *Cold Spring Harbor Symposia on Quantitative Biology*, LV:899–910.

Poggio, T. and Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343:263–266.

Poggio, T. and Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247:978–982.

Preparata, F. P. and Shamos, M. I. (1985). *Computational Geometry.* Springer Verlag, New York.

Quine, W. V. O. (1969). Natural kinds. In *Ontological relativity and other essays*, pages 114–138. Columbia University Press, New York, NY.

Quine, W. V. O. (1973). *The roots of reference.* Open Court, La Salle, IL.

Reshetnyak, Y. G. (1989). *Space mappings with bounded distortion*, volume 73 of *Translations of mathematical monographs.* Amer. Math. Soc., Providence, RI.

Rock, I. and DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, 19:280–293.

Rosch, E. (1978). Principles of categorization. In Rosch, E. and Lloyd, B., editors, *Cognition and Categorization*, pages 27–48. Erlbaum, Hillsdale, NJ.

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

Sas (1989). *SAS/STAT User's Guide, Version 6.* SAS Institute Inc., Cary, NC.

Shepard, R. N. (1968). Cognitive psychology: A review of the book by U. Neisser. *Amer. J. Psychol.*, 81:285–289.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210:390–397.

Shepard, R. N. and Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1:1–17.

Snippe, H. P. and Koenderink, J. J. (1992). Discrimination thresholds for channel-coded systems. *Biological Cybernetics*, 66:543–551.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of statistics*, 10:1040–1053.

Suppes, P., Pavel, M., and Falmagne, J. (1994). Representations and models in psychology. *Ann. Rev. Psychol.*, 45:517–544.

Tarr, M. and Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21:233–282.

Tomonaga, M. and Matsuzawa, T. (1992). Perception of complex geometric figures in chimpanzees (*pan troglodytes*) and humans (*homo sapiens*): analyses of visual similarity on the basis of choice reaction time. *J. Comparative Psychol.*, 106:43–52.

Ullman, S. and Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:992–1005.

Weiss, Y. and Edelman, S. (1995). Representation of similarity as a goal of early visual processing. *Network*, 6:19–41.

Weiss, Y., Edelman, S., and Fahle, M. (1993). Models of perceptual learning in vernier hyperacuity. *Neural Computation*, 5:695–718.

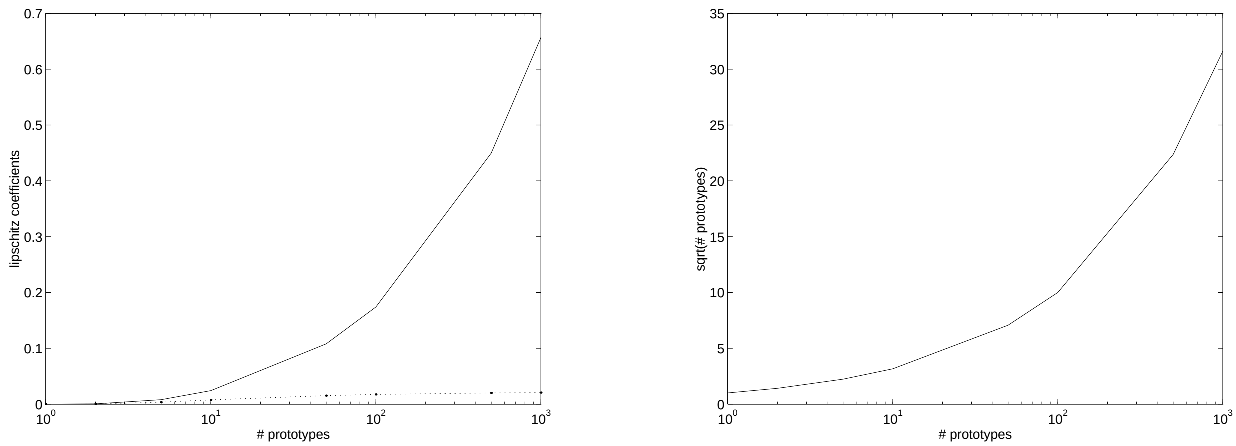

Figure 24: Determining the Lipschitz coefficient of CT (see section A in the appendix). *Left*: An approximation to the Lipschitz coefficient L, for which $\|CT(\mathbf{x}) - CT(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|$ was calculated for 100 points in R^{1000} for the non-normalized version of CT (without the factor of $\sqrt{n}$; solid line) and for the normalized CT (dotted line). Compare the behavior of the resulting Lipschitz coefficient to that of the normalization factor $f(n) = \sqrt{n}$, where n is the number of prototypes (right).

A Non-expansion of CT

The first application of the CT mapping yields the projection of each prototype onto an axis in R^n. Thus, high-dimensional input vectors are projected by CT onto the space spanned by the n prototypes, and the i'th component of the projected vector equals its distance to the i'th prototype. The non-normalized version of CT (in which distances to prototypes are not multiplied by $\frac{1}{\sqrt{n}}$) is a Lipschitz mapping.[12] In its normalized version (eq. 4), CT is nonexpansive (Goebel and Kirk, 1990), that is, $\|CT(\mathbf{x}) - CT(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|$ (a proof appears below). The following simulations clarify these statements (Figure 24). For 100 randomly chosen points in R^{1000}, we estimated the Lipschitz coefficients for the non-normalized version of CT and for its normalized version, by computing $\|CT\|_{Lip} = \max_{x \neq y} \frac{\|CT(\mathbf{x}) - CT(\mathbf{y})\|}{\|\mathbf{x}-\mathbf{y}\|}$. The behavior of the Lipschitz coefficient of the non-normalized CT (Figure 24, left, solid line) is the same as that of $\sqrt{n}$ (Figure 24, right), where n in the number of prototypes. Indeed, when normalized by a factor of $\sqrt{n}$, the Lipschitz coefficient L of CT indicates nonexpansion: $0 < L \leq 1$ (Figure 24, left, dotted line).

We now show that CT is nonexpansive, that is,

$$\rho_d(CT(\mathbf{x}), CT(\mathbf{y})) \leq \rho_n(\mathbf{x}, \mathbf{y}), \tag{6}$$

where $\rho_{d,n}$ are the metrics induced by the Euclidean norm in R^d and R^n, respectively.

Let $\{\mathbf{p}_i\}_{i=1}^n \in R^d$ be the set of prototypes and let $\mathbf{x} \in R^d$ be any input vector. Note that the first application of CT is $CT : R^d \longrightarrow R^n$, whereas all subsequent applications are $CT : R^n \longrightarrow R^n$. We establish the nonexpansion property;

[12]For a space M with a distance function (metric) δ, a mapping $T : M \to M$ is said to be *Lipschitz* if there exists $L \geq 0$ such that for all $x, y \in M$, $\delta(Tx, Ty) \leq L\delta(x, y)$ (Goebel and Kirk, 1990).

$$
\begin{aligned}
&\|CT(\mathbf{x}) - CT(\mathbf{y})\| = \\
&= \left\| \frac{1}{\sqrt{n}} \begin{pmatrix} \|\mathbf{x}-\mathbf{p}_1\| \\ \vdots \\ \|\mathbf{x}-\mathbf{p}_n\| \end{pmatrix} - \frac{1}{\sqrt{n}} \begin{pmatrix} \|\mathbf{y}-\mathbf{p}_1\| \\ \vdots \\ \|\mathbf{y}-\mathbf{p}_n\| \end{pmatrix} \right\| = \\
&= \left\| \frac{1}{\sqrt{n}} \begin{pmatrix} \|\mathbf{x}-\mathbf{p}_1\| - \|\mathbf{y}-\mathbf{p}_1\| \\ \vdots \\ \|\mathbf{x}-\mathbf{p}_n\| - \|\mathbf{y}-\mathbf{p}_n\| \end{pmatrix} \right\| \doteq \left\| \frac{1}{\sqrt{n}} \mathbf{V} \right\|.
\end{aligned} \tag{7}
$$

Without loss of generality, consider the i'th component of the vector V. Because $\rho(\mathbf{a},\mathbf{b}) \doteq \|\mathbf{a}-\mathbf{b}\|$ is a metric, it satisfies the *triangle inequality*, that is, $\rho(\mathbf{x},\mathbf{p}_i) \leq \rho(\mathbf{x},\mathbf{y}) + \rho(\mathbf{y},\mathbf{p}_i)$. Therefore,

$$\mathbf{V}_i = \|\mathbf{x}-\mathbf{p}_i\| - \|\mathbf{y}-\mathbf{p}_i\| \leq \|\mathbf{x}-\mathbf{y}\| \tag{8}$$

This is true for all i, therefore,

$$\|CT(\mathbf{x}) - CT(\mathbf{y})\| = \left\| \frac{1}{\sqrt{n}} \mathbf{V} \right\| = \sqrt{\sum_{i=1}^{n} (\frac{1}{\sqrt{n}} V_i)^2} \leq \sqrt{n \frac{1}{n} \|\mathbf{x}-\mathbf{y}\|^2} = \|\mathbf{x}-\mathbf{y}\|, \tag{9}$$

as claimed.

The same proof holds for all subsequent applications of the Chorus transform, thus establishing the nonexpansion property of CT. This property will be made use of in the next section, to demonstrate that CT preserves the Voronoi diagram of a set of points in the input space.

B Preservation of Voronoi structure by CT

We show that

$$\mathbf{x} \in V(\mathbf{p}_i) \Longrightarrow CT(\mathbf{x}) \in V(CT(\mathbf{p}_i)) \tag{10}$$

Consider $\mathbf{x} \in V(\mathbf{p}_i) \iff \|\mathbf{x}-\mathbf{p}_i\| \leq \|\mathbf{x}-\mathbf{p}_j\|,\ \forall j \neq i$. We have to demonstrate that $CT(\mathbf{x}) \in V(CT(\mathbf{p}_i))$, that is, $\|CT(\mathbf{x}) - CT(\mathbf{p}_i)\| \leq \|CT(\mathbf{x}) - CT(\mathbf{p}_j)\|,\ \forall j \neq i$. Suppose, by contradiction, that there exists some $j_0 \neq i$ for which

$$\|CT(\mathbf{x}) - CT(\mathbf{p}_{j_o})\| < \|CT(\mathbf{x}) - CT(\mathbf{p}_i)\|$$

Because CT is nonexpansive, we have, in addition,

$$\|CT(\mathbf{x}) - CT(\mathbf{p}_i)\| \leq \|\mathbf{x}-\mathbf{p}_i\|$$

Hence,

$$\|CT(\mathbf{x}) - CT(\mathbf{p}_{j_0})\| < \|\mathbf{x}-\mathbf{p}_i\| \tag{11}$$

Bearing in mind that this should be true for any $\mathbf{x} \in V(\mathbf{p}_i)$, and in particular for $\mathbf{x} = \mathbf{p}_i$, eq. (11) can be written as

$$\|CT(\mathbf{p}_i) - CT(\mathbf{p}_{j_0})\| < 0$$

which is a contradiction because $\|\cdot\|$ is a metric.