PROCEEDINGS REPRINT

SPIE—The International Society for Optical Engineering

Reprinted from

Geometric Methods in Computer Vision

25–26 July 1991
San Diego, California

SPIE PROCEEDINGS SERIES

Volume 1570

Mathematical Theories of Shape: Do they model perception?

David Mumford

Harvard University, Dept. of Mathematics and Division of Applied Science
Cambridge, MA 02138

ABSTRACT

The mathematics of shape has a long history in the fields of differential geometry and of topology. But does this theory of shape address the central problem of vision: finding the best data structure plus algorithm for storing a shape and later recognizing the same and similar shapes. Several criteria may be used to evaluate this: does the data structure capture our intuitive idea of 'similarity'? does it allow reconstruction of typical shapes to compare with new input? One direction in which mathematics and vision have converged is toward multiscale analyses of visual signals and shapes. In other respects, however, the recognition process in animals shows features that still defy mathematical modeling.

1. INTRODUCTION

There is a long history of mathematical research in the fields of topology and differential geometry with many deep and beautiful results dealing, in some sense, with shape. However, the term "shape" has never been picked up as a technical term in either subject, although both subjects started out as attempts to explore different aspects of the intuitions of shape. In the context of computer vision or animal vision, a theory of shape has quite specific goals which are unrelated to those considered by Poincaré and Gauss, the founders of topology and differential geometry. The basic task of a vision system is to be able to *recognize an object when seen again in new surroundings, and to recognize similar objects which belong to the same category.* This poses two difficult and subtle problems not dealt with in traditional mathematics.

The first of these problems is what *data structure* is appropriate for storing the "shape" of the perceived object? From an engineering perspective, there is a major need for data compression. If we were to record every view of the world that we see with each saccade of our eyes, over a lifetime we would have to store* something like 10,000 terabytes. From a psychological perspective, many experiments have tried to probe the nature of our memory of shapes. Clearly we do not store, pixel by pixel, everything we saw, but how do we code whatever it is that we store?

The second, closely related, problem is how we allow for the variations in shape when an object is recognized. We may list the sources of variation as follows:

i) noise in the imaging system,
ii) different perspectives between object and imaging system,
iii) partial occlusions, of object by itself (if not convex) and by other objects,
iv) jointed movements of parts of an object (e.g. limbs of a body),
v) more complex movements of "soft" objects (e.g. clothing),
vi) variations within a category (e.g. 2 hammers, 2 trees, 2 faces).

The interesting point here is that people have a clear intuition of what is meant by two shapes being "similar" which seems to play a role in this ability to recognize objects in spite of such variations. The type of variation which is expected with different objects itself varies, and somehow the range of "allowable" or "reasonable" variations is also stored. The ability to perceive similarity of shape is clearly affected by the type of data structure, because similarity of the data structure may or may not reflect similarity of the shape itself.

*I am assuming 4 saccades/second and that each of the million axons in the optic nerve transmits a byte of data with each saccade.

 0-8194-0698-8/91/$4.00

In this paper, I want to discuss briefly four aspects of the problem of shape: 1) Can simple linear filters solve the problem of shape recognition? 2) Is the curvature of the boundary of a shape the best data structure? 3) There are many mathematical ways to define a numerical measure of the similarity of 2 shapes: do any of these approximate the human idea of similarity? And 4) is the human idea of similarity well modeled by a metric in the first place?

2. FILTERS, GOOD AND BAD

The most natural way to describe a shape S in $\mathbf{R}^2$ is to measure it in various ways and consider the vector $\vec{f}(S)$ of these measurements as a feature vector in some Euclidean space $\mathbf{R}^n$. And, the simplest way to measure S is to use a linear filter, i.e. let

$$f_i(S) = \int_S \phi_i(x,y)\,dxdy$$

for some test functions ϕ_i on $\mathbf{R}^2$. Two sets of test functions ϕ_i have been often used: polynomials and exponentials. The first associates to a set S its *moments*

$$m_{ij}(S) = \int_S x^i y^j\,dxdy,$$

and the second associates to S its *Fourier transform*

$$\widehat{\chi_S}(a,b) = \int_S e^{2\pi i(ax+by)}dxdy.$$

The problem with moments is that they are extremely sensitive to the outlying parts of S and don't give a robust way of reconstructing S. Thus, the higher order moments are all dominated by the points in S furthest from the origin and reconstructing the rest of S is very unstable. With Fourier coefficients, the inverse Fourier transform allows one to stably reconstruct S, but only if *all* the high frequencies are used, in which case there is no data compression. Omitting these, you get ringing of edges and aliasing.

However, the use of linear filters has taken on an entirely new lease on life with the development of the theory of wavelets. At this stage there is a major battle being waged over which is the best scheme for adapting these ideas to real life images, with DARPA and the FBI (which is anxious to compress its fingerprint files) as cheerleaders. However, certain clear ideas have been demonstrated. The first goes back to the original discovery by Grossman and Morlet[18]. This is to start with many more filters than needed, i.e. to *oversample* and then choose only the strongest responses. To be specific, if we want to describe a $1D$ signal $f(x)$, we start with any nice function $\phi(x)$ such that

$$\int \phi(x)dx = 0.$$

The idea is that $\phi(ax+b)$ is the wavelet ϕ scaled by $1/a$ and translated by $-b/a$. We seek to write $f(x)$ as a superposition of such wavelets. In fact, if we define

$$\hat{f}(a,b) = \int \phi(ax+b)\cdot f(x)dx, \text{ all } a>0, b\in\mathbf{R},$$

then it is easy to check that

$$f(x) = \iint \phi(ax+b)\cdot \hat{f}(a,b)\cdot dadb.$$

At first this seems silly because it uses a function of 2 variables to reconstruct one of only 1 variable. But what is remarkable[7] is that numerically $f(x)$ is well approximated by replacing this double integral by a small finite sum involving only the largest values of $\hat{f}$. Two-dimensional versions of this have been explored by Mallat[15] and others and apply to the description of shape.

The second approach is due to Meyer and Daubechies[6,17] and is based on the remarkable fact that there exist ϕ such that

$$\phi_{ij}(x) = \phi(2^i x + j)$$

are an orthonormal basis of $L^2(\mathbf{R})$. For such a ϕ, any f is a unique superposition of the ϕ_{ij}'s. Meyer, Coifman and Wickerhauser have forged a powerful tool for data compression by combining these two approaches[5]. They start with a large set of wavelets $\phi_\alpha(x)$, which are highly redundant, i.e. have many linear dependencies, but from which one can extract, in many ways, subsets which form an orthonormal basis. The idea is to choose, for each $f(x)$, an orthonormal subset which most rapidly captures the salient features of $f(x)$.

All of these methods offer a powerful method of data representation because they incorporate arguably the most efficient method of data compression and very rapid and stable reconstruction. One might argue that reconstruction is not relevant to perception because it is not needed in pure recognition algorithms. However, both psychology and neuropsychology suggest the opposite. Mental imagery seems to play a major role in human cognitive processes[14] and in many analyses of recognition tasks[4]. Neurophysiology shows that the number and size of top-dawn pathways in the brain are just as great as those of bottom-up pathways[9,20]. This strongly suggests that top-down reconstruction of memory traces, in more concrete form for comparison with present stimuli, plays a vital role in recognition.

The whole idea of wavelets is to analyze signals and shapes in a multi-scale way: in terms of a crude first approximation to its largest features, plus refinements on a smaller scale, and refinements to the refinements on a yet smaller scale, etc. This idea in computer vision goes back to Rosenfeld, Uhr, Burt and others[23] and seems to be adopted in essentially all animal vision systems. The center-surround organization of retinal ganglion cells seems to be a wavelet data compression scheme for making optimal use of the optic nerve. The variable receptive fields of cells throughout the visual system strongly suggests a multi-scale processing scheme.

3. PROS AND CONS OF CURVATURE DESCRIPTIONS

The central concept in the field of differential geometry is, without a doubt, curvature. A plane curve C has a curvature at each point P, and a surface S in 3-space has two "principle curvatures" at each point P (given by the minimum and maximum curvature of plane curves which are slices of the surface S with planes meeting the surface perpendicularly at P). Since these are invariant under rotation and translation, they seem to offer to recognition algorithms a golden route to compensating for differing viewpoints, etc. And, in fact, it is an easy theorem that a plane curve C is determined, up to rotation and translation, by its curvature $\kappa(s)$, as a function of arc length on C (a similar result holds for surfaces).

But, as with moments, $\kappa(s)$ is a highly unstable way to reconstruct C. For instance, suppose C surrounds a shape with a narrow neck and let $P, Q \in C$ be near each other but on opposite sides of the neck. Then altering $\kappa(s)$ by a small amount between P and Q can move P and Q far apart and destroy the "neck" altogether (see figure 1).

This points up a basic fact about shape: a plane shape $S \subset \mathbf{R}^2$ has a 1-dimensional side given by features of its boundary $C = \partial S$; and a 2-dimensional side given by its interior. No successful theory of shape description can ignore one or the other. An essential supplement to the curvature description of C is the description of its interior by the medial axis transform (related to the theory of evolutes in differential geometry). In this construction, one looks at S as the union of the curves

$$C_a = \{P \in S | dist(P, C) = a\}$$

plus the axis itself

$$A(S) = \{P \in S | \exists \text{ a circle inside } S \text{ touching } C \text{ twice}\}$$

plus its endpoints

$$A_0(S) = \{P \in S | P = \text{ center of an osculating circle to } C \text{ contained in } S\}.$$

These endpoints $A_0(S)$ give a natural way of approximating C by a polygon. For each point of $A_0(S)$, take the point where the osculating circle touches C, a point which is a local maximum of the curvature of C, and join these into a polygon. This procedure produced Attneave's drawing of a cat (figure 2), which argued for the theory that most shapes can be recognized from these polygonal approximations. This description becomes much richer for surfaces in $\mathbf{R}^3$, where the curvature of the surface defines many types of special points and curves on the surface, and the medial axis transform defines many special surfaces and axes in the interior. A beautiful description of these rich phenomena can be found in Koenderink's book[13]. A major problem with this is that unless one can reconstruct depth very accurately, it is not clear which of these mathematical constructs are useful for intensity based vision systems[†].

The most exciting mathematics recently on the description of shape via curvature has been the intensive study of a geometric heat equation for shape. This can be defined intuitively as follows: at all $P \in C$, let $\vec{n}_P$ be the normal vector to C pointing into S, and let $\kappa(P)$ be the curvature of C at P. Then the curve C_a just defined, for $a = \epsilon$, is the locus of points $P + \epsilon \cdot \vec{n}_P$. We may vary this procedure and define a new family of curves C'_a, where C'_ϵ is the locus of points $P + \epsilon\kappa(P) \cdot \vec{n}_P$, i.e. always move points on C'_a normal to C'_a for a distance proportional to the curvature of C'_a (see figure 3 for an example of such a family).

Written out, if C is given by $y = f(x)$ and C'_a by $y = f(x, a)$, we find:

$$\frac{\partial y}{\partial a} = \frac{\partial^2 y/\partial x^2}{\left[1 + (\partial y/\partial x)^2\right]}.$$

This is exciting because it adapts multi-scale ideas to shapes. Although the curves C_a get sharper and sharper corners, the curves C'_a get smoother and smoother, ultimately approaching a perfect circle (which shrinks away to nothing)[10,11]. Applied to a shape like a coastline, the capes and bays dissolve one by one until only the course shape remains. Running it backwards, you can read off the features in terms of the appearance of points of inflection and associated concavities.

The analogous equation for shapes in $\mathbf{R}^3$ (or $\mathbf{R}^n$) uses mean curvature, the sum of the principal curvatures, and if S is given as a graph $y = f(x_1, \cdots, x_{n-1}, a)$, the geometric heat equation works out to be:

$$\frac{\partial y}{\partial a} = \frac{\Delta y}{(1 + ||\nabla y||^2)}.$$

Unfortunately, it develops singularities, i.e. a dumbbell will squeeze down in its neck to a singularity, and a complete theory of this equation is not known at this time[8,21]. A synthesis of the medial axis and the heat equation has been given by Kimia and Zucker[12]. They define a 2-dimensional family of shapes $C_{a,b}$ where infinitesimally:

$$C_{\epsilon,\eta} = \text{ locus of points } P + (\epsilon\kappa(P) + \eta) \cdot \vec{n}_P.$$

To date, these theories have been used to give qualitative descriptions of shape, but have not solved the data compression or reconstruction problem the way wavelets have.

4. METRICS ON THE SPACE OF SHAPES

The direct attack on modeling perception is to seek a metric on the space of shapes which abstracts our intuitions about which shapes are 'similar'. What is confusing here are the number of choices for such a metric. Here are 6 (!) very different approaches to defining a metric:

i. *Hausdorff metric*: Given $S_1, S_2 \subset \mathbf{R}^2$, let

$$d_H(S_1, S_2) = \sup_{x_1 \in S_1}\left[\inf_{x_2 \in S_2} ||x_1 - x_2||\right] + \sup_{x_2 \in S_2}\left[\inf_{x_1 \in S_1} ||x_1 - x_2||\right]$$

(i.e. d_H = sum of the distance of the furthest point in S_1 from S_2 plus the furthest point in S_2 from S_1).

ii. *Template metric*:

$$d_T(S_1, S_2) = \text{area}(S_1 - S_2) + \text{area}(S_2 - S_1)$$

† If laser range data is available, it is another story altogether: cf. article by Gordon, this volume.

iii. *Transport metric*:

$$d_M(S_1, S_2) = \inf_\rho \int_{S_1}\int_{S_2} ||x_1 - x_2|| \cdot d\rho(x_1, x_2)$$

where ρ is a probability measure on $S_1 \times S_2$ such that

$$\int_{S_1}\int_{U_2} d\rho(x_1, x_2) = \frac{\text{area}(U_2)}{\text{area}(S_2)}$$

$$\int_{U_1}\int_{S_2} d\rho(x_1, x_2) = \frac{\text{area}(U_1)}{\text{area}(S_1)}.$$

Roughly speaking, this means fill S_1 uniformly with 'stuff' and find the shortest paths along which to move this 'stuff' so that it now fills S_2. This is a popular metric in operations research and is readily calculated by linear programming[22].

iv. *Optimal diffeomorphism*:

$$d_O(S_1, S_2) = \inf_\phi \left[\int_{S_1} ||J\phi||^2 + \int_{S_2} ||J(\phi^{-1})||^2\right]$$

where $\phi : S_1 \longrightarrow S_2$ is a 1-1, onto differentiable map with differentiable inverse ϕ^{-1} and J means the matrix of 1^{st} derivatives. The above integral is called "energy" in differential geometry. Variants of this use the stress of ϕ from elasticity theory, or norms on the second derivatives of ϕ.

v. *Maps with tears*:

$$d_A(S_1, S_2) = \inf_{R \subset S_1 \times S_2} \text{Area}(R)$$

where R is a surface and $p_1 : R \longrightarrow S_1$, $p_2 : R \longrightarrow S_2$ are 1-1 over most of S_1 and S_2, i.e. R is almost everywhere the graph of an invertible map, but has exceptional "tears[‡]" in places (various ways of making this precise are possible).

vi. *Graph matching*: We don't try to make this precise except to say that to do so, we first construct a graph Γ_i attached to S_i, where the nodes of Γ_i represent its parts on multiple scales and the edges of Γ_i represent adjacency or inclusion of parts, as in Marr's 3D-model[16]. The vertices and nodes may also carry labels used in interpreting 'best' in:

$$d_G(S_1, S_2) = \text{measure of the best partial match between } \Gamma_1 \text{ and } \Gamma_2.$$

Running through this list, here is a commentary:

i. This is an L^∞-type metric (like $||f||_\infty = \sup_x |f(x)|$) and, as such, it is very sensitive to any outlier points in S_1 and S_2, i.e. small outlying details cannot be overlooked. Moreover, as basic a function of S as Area(S) is not continuous in this metric. However, it is the basis of the morphological theory of shape. Its natural domain is the space of *compact* shapes in $\mathbf{R}^2$, i.e. it extends to this space, where it has the technical property of being complete.

ii. This is, in many ways, the opposite of d_H: it is a L^1-type metric (like $||f||_1 = \int ||f(x)||dx$), which is looking at a shape through its "median" points. It is totally insensitive to outliers, e.g. if a blob grows a large but thin appendage, d_T essentially ignores it. In this metric, the basic construction of morphology, the map from S to

$$S + D_\epsilon = \{P \in \mathbf{R}^2 | \text{dist}(S, P) \leq \epsilon\}$$

is not continuous. The natural domain for d_T is the space of measurable subsets of $\mathbf{R}^2$ modulo subsets of measure zero.

[‡]Tears in the sense of tearing paper, not shedding tears.

iii. In many ways, d_M fixes the problems of d_H and d_T ('M' stands for the mathematician Monge, who considered this metric). It incorporates distances in $\mathbf{R}^2$, but weights equally all points of the shape. Its natural domain is the space of probability measures on $\mathbf{R}^2$: i.e. associate to S the measure ρ_S given by:

$$\rho_S(U) = \frac{\text{area}(S \cap U)}{\text{area}(S)}.$$

The distance $d_M(\rho_1, \rho_2)$ is defined in the obvious way. This extension of d_M allows one to include very natural generalizations of shape to a) purely 1D shapes in $\mathbf{R}^2$, with length used instead of area, and to b) fuzzy shapes, á la Zadeh, where membership near the boundary is not a 0/1 decision.

iv. + v.: Measures like d_O have been proposed by the Brown group working on X-rays of hands[1,2]: they compute optimal diffeomorphisms between a stimulus hand and an ideal hand template. Its problem is that if a small break cuts a shape into two, the original and the broken shape are infinitely far apart because they are not even topologically the same. This goes against the multi-scale philosophy because whether an apparent break actually divides a shape in two or not may be only visible at a very fine scale. d_A is proposed as a way to make d_O more robust. Our group at Harvard has used d_A to construct stereo matching, because most stereo pairs include quite a few non-matching pixels, i.e. points visible to one eye but not the other, and this means that the matching must have tears. The metric d_A is very close to d_M, but the theory of this has not been worked out.

vi. This type of metric has been proposed by many workers, but it suffers from a huge problem: small changes in a shape can push the size of various parts across thresholds, resulting in a major reorganization of the tree of parts. This makes Γ very brittle. To make this approach really work, something like a theory of graph rotation is needed. This would give a catalog of these sudden rearrangements and a way of computing multiple "parses" for borderline shapes, i.e. shapes whose best parse is near some threshold.

5. "SIMILARITY" ISN'T A METRIC ANYWAY

I believed a metric, like those in section 4, was the right way to understand the intuition of "similar" until I participated in a psychological experiment producing actual measurements of similarity[19]. Our results, plus a major dose of the startling experiments of Amos Tversky[24] made me realize how this description captures only a small part of the significance of a similarity judgment.

Our experiment started with a small set of shapes: in the first of these we used 15 polygons (see figure 4) and in the second we used the 34 letters of the Gujarati alphabet. We used two sets of subjects: pigeons and Harvard students, and aimed, for each, at measuring the perception of similarity between each pair of shapes. The pigeons were trained to peck first at one shape, then at another, etc., until all shapes had been their target several times. For each run with a fixed correct shape, their errors for each of the other false shapes were recorded, resulting in a matrix of confusability measures: the $(i, j)^{\text{th}}$ entry represents the errors in which the i^{th} shape was misidentified as the j^{th} shape. For people, again each shape was taken in turn as the target, and the reaction time was recorded for the subject to accept or reject this and the other shapes when presented by slide. Again, we get a confusability matrix in which long reaction times indicate difficulty in discrimination, hence high similarity, and fast times the opposite.

The first thing that struck us was how *asymmetric* these matrices were. If "similarity" was a metric, then $d(S_1, S_2)$ should equal $d(S_2, S_1)$ and our matrix should be roughly symmetric. It wasn't! Bootstrap estimates indicated an average correlation of 0.8 of the anti-symmetric part of the matrix with itself, so this does not seem to be an artifact. Turning to Tversky, we found that there tend to be consistent asymmetries in human verbal judgments of similarity too: e.g. his subjects said that the number 99 was very similar to the number 100, but balked at describing 100 as very similar to 99! They said (this was in cold war days) that 'North Korea' was more similar to 'China' than vice versa. What seems to be happening here is that if you say 'A' is similar to 'B', you mean that B is some kind of prototype in a category which includes A. Thus the stimulus input A being analyzed is treated differently from the memory benchmark B. In the context of shape recognition, this sort of asymmetry can be interpreted like this: suppose the shape B to be recognized has a prominent strong feature (e.g. a sharp pointy top) so it is a good prototype of some class of shapes. Then the mind configures itself to look immediately for this feature, and if the stimulus A is without this feature, it is rejected very fast. But if the nondescript A is the shape to be recognized, there is no easy test for 'A-ness' and the mind doesn't focus on any single feature. Then rejecting the stimulus B as an instance of A takes longer.

Moreover, a second problem showed up that we first conjectured and then verified in follow-up experiments: the similarity of A and B depends strongly on the context in which this discrimination is being made, i.e. the whole set of shapes in the experiment. Suppose for example that A,B,C and D must be discriminated. If C and D are very different from both A and B, A and B appear relatively similar and are more often confused; but if A and C are close and B and D are close, then A and B are relatively easy to discriminate. In the same vein, Tversky asked people whether Austria was most similar to Sweden, Norway or Hungary: 60% chose Hungary. But asked whether Austria was most similar to Sweden, Poland or Hungary, 49% chose Sweden (n.b. at the time of this experiment, Poland and Hungary were linked in people's minds as being Eastern Bloc countries).

There is yet another difficulty in modeling animal shape discrimination, their confusion between mirror image shapes. To a computer, the whole idea of getting left and right confused seems very silly, and one has to go out of one's way to *force* a data structure to make mirror image shapes similar. My interpretation of this phenomenon is that it is due to the symmetry of the body and the consequent fact that the brain develops in two nearly identical hemispheres. Because of this, the brain tends to form *two* memory traces of an object, one in each hemisphere, one the reflection of the other. But the two hemispheres are integrated in a complex way, exchanging their data continuously, and this produces the mirror image confusion. In any case, this is one aspect of animal recognition that we don't want to emulate.

We carried out an analysis of the animal data using metrics of the template type d_T, the graph matching type d_G and also via back-propagation neural nets. The pigeon data was best modeled by d_T, with a moderately good correlation of 0.69 while the human data was best modeled by d_G, with a reasonable correlation of 0.8, but only after including a crude fix to artificially add mirror-image confusions.

I'd like to conclude by making some conjectures about what I believe is the best way both to model animal recognition and to construct a truly robust machine recognition algorithm. My belief is that the neural network model of Grossberg and Carpenter, which they call 'adaptive resonance theory'[3], is moving in the right direction. Rather than choosing between feature vectors and templates, the idea is that *both* must be used in an iterative architecture involving closely coupled bottom-up and top-down stages. The following scenario describes my extension of their ideas[4,20]: A new scene or shape, after some pre-processing, first activates a set of features bottom-up. These stimulate various higher-level categories of objects, and, in a top-down channel, templates of prototypes of these objects are produced. The lower-level tries to match these to the scene and this triggers new features describing the 'residuals', the mismatched features. Meanwhile, the higher area also stores data on the range of allowable variations for each class of objects, and, on receiving these residuals, modifies the template reconstructions: the template is better thought of as a 'flexible template'. Eventually, a set of the higher level categories, with suitable parameters describing this specific exemplar, produces a mental image like template which matches the scene or shape up to 'noise', and no strong residuals remain. In this architecture, 'similarity' is totally customized to the type of object being recognized: for each category of object, the degree of similarity with a stimulus depends on how much variation is built into the template, and how strong are the features of the ultimate residual. Something like d_M or d_A seem to be the most promising general purpose tool for modeling this type of flexible template, but the features which are passed bottom-up must also be incorporated into measure of disparity being minimized in either metric (e.g. in d_M, not just the distance of x from y but the difference of various measured quantities at x and y should be used in the integral).

6. REFERENCES

1. Y.Amit, A non-linear variational problem for image matching, preprint, *Brown Dept. of Applied Math.*, 1991.

2. Y.Amit, U.Grenander and M.Piccioni, Structural image restoration through deformable templates, to appear in *JASA*, 1991.

3. G.Carpenter and S.Grossberg, A massively parallel architecture for a self-organizing neural pattern recognition machine, *Comp. Vision, Graphics and Image Proc.*, 37, pp.54-115, 1987.

4. P.Cavanagh, What's up in top-down processing, *Proc. 13th ECVP*, to appear 1991.

5. R.Coifman, Y.Meyer and V.Wickerhauser, Wavelet analysis and signal processing, *preprint, Yale Dept. of Math.*, 1990.

6. I.Daubechies, Orthonormal bases of compactly supported wavelets, *Comm. Pure and Appl. Math.*, 41, pp.909-996, 1988.

7. I.Daubechies, The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Inf. Theory*, 36, pp. 961-1005, 1990.

8. L.Evans and J.Spruck, Motion of level sets by mean curvature, I, *J. Diff. Geom.*, 33, pp.635-682, 1991.

9. D.Felleman and D.Van Essen, Distributed hierarchical processing in primate cortex, to appear in *Cerebral cortex*, 1991.

10. M.Gage and R.Hamilton, The heat equation shrinking convex plane curves, *J. Diff. Geom.*, 23, pp.69-96, 1986.

11. M.Grayson, The heat equation shrinks embedded plane curves to round points, *J. Diff. Geom.*, 26, pp.285-314, 1987.

12. B.Kimia, A.Tannenbaum and S.Zucker, On the evolution of curves via a function of curvature, I, *preprint, McGill Research Center for Intelligent Machines*, 1990.

13. J.Koenderink, *Solid Shape*, MIT press, Cambridge, 1990.

14. S.Kosslyn, *Image and Mind*, Harvard Univ. Press, Cambridge, 1980.

15. S.Mallat, A theory of multiresolution signal decomposition: the wavelet representation, *IEEE Trans. PAMI*, 11, pp. 674-693, 1989.

16. D.Marr, *Vision*, Freeman and Co., New York, 1982.

17. Y.Meyer, *Ondelettes et Operateurs*, Hermann, Paris, 1990.

18. J.Morlet, G.Arens, I.Fourgeau and D.Giard, Wave propagation and sampling theory, *Geophysics*, 47, pp.203-236, 1982.

19. D.Mumford, R.Herrnstein, S.Kosslyn and W.Vaughan, Analysis and synthesis of human and avian categorizations of 15 simple polygons, *preprint, Harvard Univ. Dept. of Psych.*, 1989.

20. D.Mumford, On the computational architecture of the neo-cortex, to appear in *Biological Cybernetics*, 1991.

21. S.Osher and J.Sethian, Fronts propagating with curvature dependent speed, *J. Comput. Physics*, 79, pp.12-49, 1988.

22. S.Rachev, The Monge-Kantorovich mass transference problem and its stochastic applications, *Theory of Prob. and Applic.*, 29, pp.647-676, 1985.

23. A.Rosenfeld, ed. *Multiresolution image processing and analysis*, Springer-Verlag, Berlin, 1984.

24. A.Tversky, Features of similarity, *Psychological Review*, 84, pp.327-352, 1977.

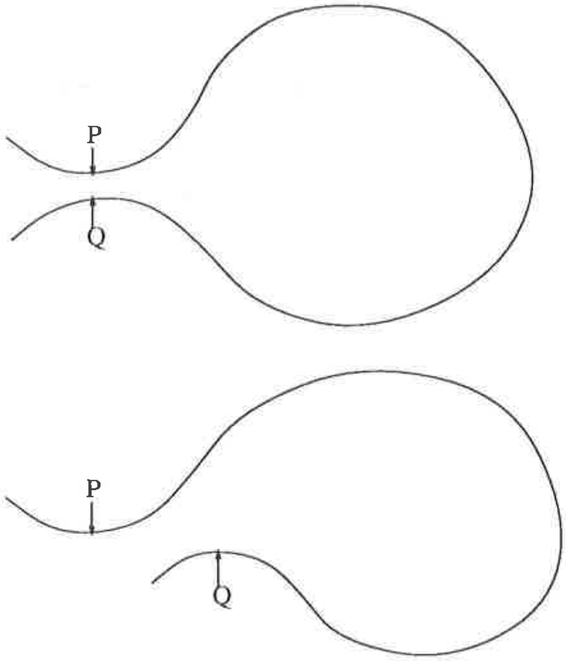

FIGURE 1

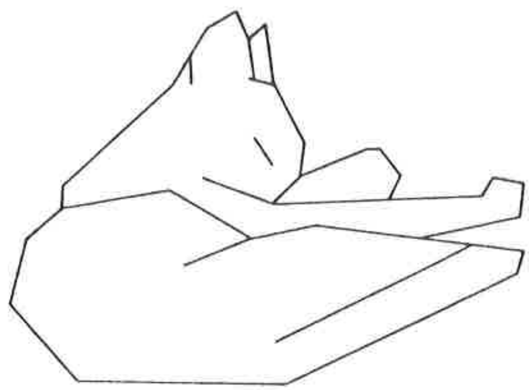

FIGURE 2

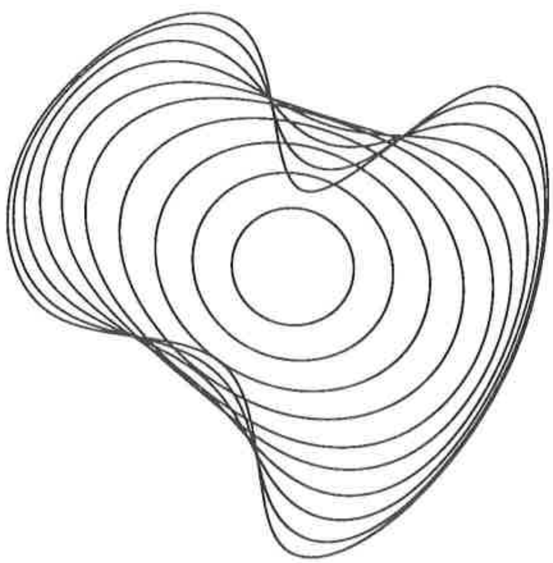

FIGURE 3

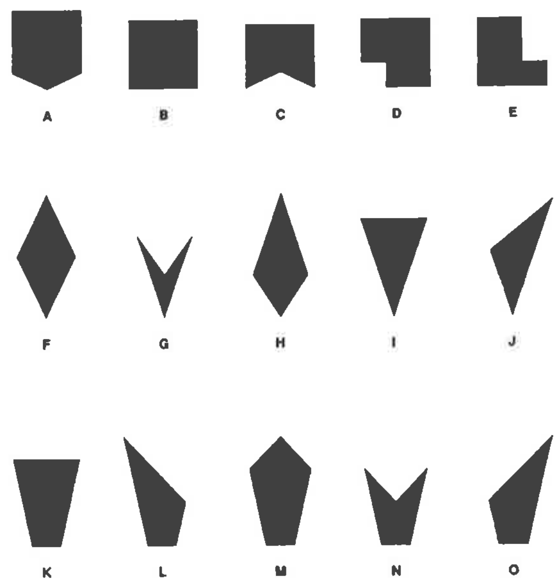

FIGURE 4