COMPUTER SOCIETY PRESS REPRINT

THE PROBLEM OF ROBUST SHAPE DESCRIPTORS

David Mumford

Reprinted from PROCEEDINGS OF THE IEEE FIRST INTERNATIONAL CONFERENCE ON COMPUTER VISION, London, England, June 8-11, 1987

The Computer Society of the IEEE
1730 Massachusetts Avenue NW
Washington, DC 20036-1903

Washington • Los Alamitos • Brussels

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

IEEE

IEEE COMPUTER SOCIETY PRESS

THE PROBLEM OF ROBUST SHAPE DESCRIPTORS

David Mumford

Department of Mathematics and Division of Applied Science,
Harvard University

INTRODUCTION

A fundamental problem in both Computer Vision and Artificial Intelligence is the transition from events in the real world which are always described by continuous data, to symbolic descriptions of the sort that computers can readily manipulate. There is a kind of mismatch between these two aspects of reality, between the signals of electrical engineers and the terms of symbolic logic. Outside our minds lies what William James termed the "blooming, buzzing world" when he discussed the perceptions of a newborn infant. Inside a computer, although perhaps not inside our minds, lies a crisp, sanitized, predictable environment. To deal with this mismatch, we must take certain precautions.[1]

Thus, suppose we need to measure the length of an object along a specific side. We take a ruler, line up one end, and estimate the nearest marking to the other end. Obviously we haven't found a the exact length; at best we have found an interval within which the length falls. Thus if we say the length is 13.53" we surely mean between 13.525" and 13.535". Actually this too cannot be right, for suppose the length were 13.5350001": then, unless we had very great accuracy, we might estimate it as 13.53" instead of the closer number 13.54". We had better allow an **overlap** between the intervals we use for our estimates or the borderline cases will call for infinite accuracy on our part. The best strategy is to say that a measurement of 13.53" means the length is between 13.5225" and 13.5375" (or something like this), so that if we are able to estimate accurately up to +/-.0025", we can be sure of finding some assertion which allows for this degree of error. Even so, there is possibility for error: no one can guarantee that they always have at least such and such an accuracy. Even being very conservative, there will always be instances when you or your tools slip, break, or are subject to unforeseen influences and, without your realizing it, your measurement is further off. To deal with this contingency, the program which subsequently manipulates the measurement had better be robust enough to handle occasional errors, i.e. to 'know' that 13.53" +/-.0075" might mean 13.52" or even 13.51".

The purpose of this paper is to apply the same principles to the description of **shape**. Shapes, like lengths, also vary smoothly: a triangle can be slowly smoothed and rounded by a process that changes it without any sudden shifts into a circle. The main difference is that, intuitively as well as mathematically, shapes vary not in a finite number of independent ways, but in infinitely many ways. We may say that the space of shapes is infinite-dimensional. In fact, as we shall see, it is not merely infinite-dimensional, it is also **very complicated.** Miller & Johnson-Laird, in their monumental study of the semantic primitives of perception, were right on the mark when they referred to "the extremely complex attribute called the shape of a perceptual object".[2]

Our first order of business is going to be making the words "the space of shapes" precise. What this means is that one must say first exactly what a "shape" is and one must say secondly when two shapes are similar. The best way to say when they are similar is to define a "distance" between two shapes. Once such a definition is given, the totality of all our shapes forms what is called mathematically a **metric space**, i.e. we have constructed a set S whose points are in 1-1 correspondence with shapes, and within S we have defined how far apart any 2 points are. (A weaker but sometimes more satisfactory way is to merely define on S the structure of topological space.)

If we want next to **describe** shapes, the second order of business is to cover the space of shapes by open sets U(i), determining not merely the U(i)'s but their overlaps as well. It is then a reasonable request to ask for a signal processing front-end that observes shapes in the real world, with all the uncertainty this entails, to output statements like "the blue shape in front of me belongs to the class of shapes U(i)". Otherwise said, the processor has assigned shape descriptor i to a particular real shape. Knowing the shape is in U(i) means we know that the shape is roughly such-and-such, and it means knowing that U(i)-shapes might sometimes be classified as U(j)-shapes for a known set of 'nearby' shape-descriptors j. Without an explicit description of nearby j's, I believe a theory of shape-descriptors can never be used successfully in a high-level recognition or categorization program.

It seems common-sense that this is what a theory of shape-descriptors should deliver. However, to my knowledge, no existing theory does deliver this, and it does not seem simple to find one that does (I will describe a solution to much easier 1D shape description problem below). Consider, for example, Marr's theory of 3D-models. He proposes an elegant and intuitive tree-like data structure with nodes corresponding

CH2465-3/87/0000/0602$01.00 © 1987 IEEE

to salient parts, and links corresponding to "part-of" relations. He gives many examples of objects and corresponding trees. But suppose you start with an object O with descriptor M. If you vary O gradually, altering proportions, size and relations of parts, slant and tilt of its surface, at some point the appropriate descriptor will change - to M´, M´´,.... What are these ´nearby´ descriptors? I suspect the answer to this is rather complicated, and also contingent on how you fill in the ´details´ to make his theory really precise.

To take another example, consider the description of shapes by moments. This theory does, in a weak sense, tell you what nearby descriptors are: they are ones whose moments are near those of the given shape. I say ´weak sense´ because the theory does not specify how to weight the higher moments vs. the lower moments. This is crucial since there are infinitely many moments, and even in practice there are a huge number of higher moments available and a variation in one of the 100th moments is certainly not as important as a similar variation in one of the 1st moments. The major weakness of this approach to shape description, however, is whether closeness of moments in any sense reflects accurately closeness of the shapes. We shall see below that for one definition of closeness of shapes, it does not. A common rule-of-thumb is that higher order moments are too unstable to be useful.[3]

In the first part of this paper, I will give several examples of important ways to define the distance between shapes. In the second half of the paper, I will focus particularly on the case of 1-dimensional shapes, and give a describe a particular covering by open sets U(i), showing how it relates to a Marr-like hierarchical descriptor.

METRICS ON THE SPACE OF ALL SHAPES

What do we mean by the space of **all** shapes? We certainly want to include circles and triangles as special cases, so the smallest reasonable class of shapes seems to be those whose boundary is made up of a finite number of differentiable arcs, meeting at a finite set of corners. In the 1D case, the analogous choice is to look at the subsets of the line made up of a finite set of disjoint intervals. On the other hand, other shapes come with fine texture on their boundary, e.g. the outline of a tree or furry animal. Perhaps it is more reasonable to consider closed sets in the plane with arbitrary simple closed curves as boundaries. And maybe these shapes should be allowed to have infinitely many holes in them: think of the shape given by froth or a section of the lungs. In 1D, a general closed set is gotten by starting with an interval and deleting a possibly infinite set of open subintervals, getting smaller and smaller.

Which is the right choice really depends on the application, and is linked also to the question of when two shapes are close. For example, consider the family of shapes obtained by deleting a circle of radius a from the unit circle: $a^2 <= x^2+y^2 <= 1$. As a goes to zero, do these shapes approach the unit circle without a hole? If the hole is considered to be ´visible´ or ´salient´ even when tiny, then it doesn´t; but if the hole, when tiny, is considered to disappear, then it does. We can make this precise by recalling the concept of the **Hausdorff metric**. This is a function which gives one notion of distance between 2 shapes S and T. For this definition, S and T may be subsets of the line, the plane or any Euclidean space, but must be bounded closed sets. First define the distance of a point P from a set S by:

$$d(\{P\},S) = \mathrm{MIN} \{ |P - Q| \mid \text{points } Q \text{ in } S\}$$

Then define

$$d(S,T) = \mathrm{MAX} \{ d(P,T) \mid \text{points } P \text{ in } S \} + \mathrm{MAX} \{ d(S,Q) \mid \text{points } Q \text{ in } T \}.$$

This is the Hausdorff distance. It satisfies the axioms for a metric:

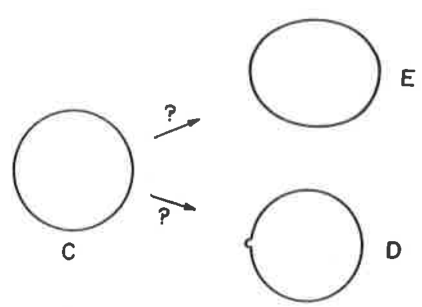
i) $d(S,T) = 0$ if and only if $S = T$,
ii) $d(S,T) = d(T,S)$,
iii) $d(R,T) <= d(R,S) + d(S,T)$.

In this distance, however, holes can disappear. If holes are salient, one probably wants a definition invariant under figure-ground reversal, and then a natural choice is:

$$d'(S,T) = d(S,T) + d(CS,CT),$$

where CS (resp. CT) is the complement of S (resp. T). With metric d´, a shape with a long thin crack forinstance is **not** near the same shape without the crack; but with metric d, it is near.

There are other ´stronger´ definitions of nearness which would be appropriate in other circumstances. For example, which is closer to a circle C: an ellipse E which moves inside and outside C a distance e, or a slightly smaller circle D with one ´tooth´ extending abruptly beyond C, but still within e/2 of C[4]:



In some situations, E in isolation might seem identical to C whereas D would be clearly different. The natural way to modify the metric to make D farther from C is to add a term involving the maximum difference between the **slope** of the boundary contour of C (resp. D) at a point P and the slope of the contour of D (resp.

C) at the point Q nearest to P, giving a new metric $d^{(1)}$.

Other choices of metric include the simple template distance:

$$d_t(S,T) = area(S-T) + area(T-S),$$

and the Kantorowitz distance. To define this, imagine S and T as being defined by unit quantities of pigment spread evenly on each set. Then define the distance between S and T as the total distance pigment must be moved to change S into T.[5]

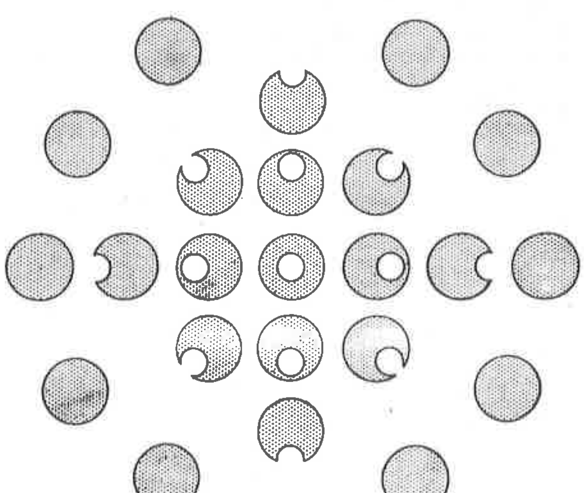
The point is that there is a sequence of 'weaker' and 'stronger' distance measures between pairs of shapes, the stronger requiring that more features of the shapes be similar before they are close. Correspondingly, for these distances to make sense, the shapes being compared must be more and more restricted. Thus the Hausdorff distance d is a well-defined distance measure on the space of all bounded closed sets in the line or the plane: we will call these spaces, in this metric, SH1 and SH2. But for $d^{(1)}$ to be defined, the boundary of the shape must be a differentiable arc, thus excluding all fractal or textured contours - so $d^{(1)}$ is a distance on a much smaller space of shapes.

An important point to understand about these spaces is that **length** is not a continuous function on SH1, nor is **area** a continuous function on SH2 (although **diameter**, i.e. the maximum separation of two points on the shape, is continuous on both). This is seen immediately from the example:

X_n = Union of 2^n intervals of length $1/(4^n)$ centered at the points $i/(2^n)$, $1 <= i <= 2^n$,
X = unit interval [0,1].

Then $X_n \longrightarrow X$ in the Hausdorff metric, but $length(X_n) \longrightarrow 0$! For the same reason, the moments are not continuous on SH1 or SH2.

The space of shapes also has interesting topology as is seen in the following construction. We will define a family of 2D shapes, i.e. sets in the plane, one for each point on a 2-dimensional sphere. It looks like this:



Each shape is given by starting from the unit circle $x^2+y^2 <= 1$ and deleting a circle $(x-a)^2 + (y-b)^2 <= 0.25$. To the north pole we associate the ring shape gotten when the deleted circle is also centered at the origin: a=b=0. To the northern hemisphere we associate the shapes where the deleted circle is still wholly within the unit circle: $a^2+b^2 <= 0.25$. To the southern hemisphere we associate the shapes gotten by deleting circles overlapping but partly outside the unit circle: $0.25 <= a^2+b^2 <= 2.25$. To the south pole itself we associate the unit circle without any deletion. It seems likely that with a suitable metric on the space of bounded 2D shapes, this 2-sphere of shapes is topologically 'non-trivial' in the sense that it cannot be collapsed to a point.

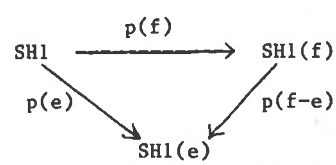
DILATION ON THE SPACE OF ONE-DIMENSIONAL SHAPES

I want to focus in the rest of this paper on the simplest case: the space SH1 of all bounded closed subsets of the line with the Hausdorff metric d. It might seem that SH1 is an impossible space to describe. This is not so. The process of dilation from the theory of shape morphology gives a nice way of visualizing SH1. For any positive e and shape X, define

$$X(e) = X \oplus [-e/2,+e/2] = \{ x + y \mid x \text{ in } X, -e/2 <= y <= e/2 \}$$

SH1(e)= set of shapes X which are unions of intervals of length at least e.

Then we define a map $p(e) : X \longrightarrow X(e)$ by $p(e)(X) = X(e)$. If $e <= f$, then the map p(f) factors via p(e), i.e. in the diagram:



the map p(f) is the composition p(f-e)op(e). The following theorem is not hard to prove:

THEOREM: The topological space SH1 is homeomorphic to the inverse limit[6] of the spaces SH1(e):

$$SH1 \xrightarrow{\approx} \varprojlim_e SH1(e).$$

This is an abstract formulation of the idea that any shape can be described hierarchically by a sequence of finer and finer "up to size e" descriptions. But even better, if we put a bound on the shape, e.g. X contained in [0,1] - call these shapes $SH1_{[0,1]}(e)$ - then $SH1_{[0,1]}(e)$ is a **finite-dimensional** space. This is clear because any X in $SH1_{[0,1]}(e)$ is made up of less than 1/e intervals, and describing any interval by its two endpoints, we have

$$\dim SH1_{[0,1]}(e) <= 2*[1/e].$$

What do the spaces $SH1_{[0,1]}(e)$ look like? These are good finite approximations of SH1 itself, so we may expect that their structure gives some insight into SH1. These spaces are quite interesting and intricate - they are not 'flat' spaces (i.e. manifolds with boundary), but they have singular points and their dimension is different near different points.

A SYSTEM OF 1D SHAPE DESCRIPTORS

The purpose of this section is to describe an explicit set of 1D shape descriptors based on a hierarchical structure of overlapping open sets covering the space of compact subsets of the line. The 1D case can be solved by a simple variant of quad-trees, but nonetheless one which I think can be used as the basis of much more robust scheme for describing sets of intervals. We first present the corresponding solution to the simpler problem of describing **points** by a hierarchical structure of overlapping open sets.

Suppose a point in the interval [0,1] with coordinate x is fixed. At first glance, with a crude yardstick, we simply say x lies more or less to the left, more or less in the middle or more or less to the right. We might mean by this:

x is to the left if	x in [0,1/2),
x is in the middle if	x in (1/4,3/4)
x is to the right if	x in (1/2,1].

A typical point to the left is 1/4, in the middle is 1/2, to the right is 3/4. Taking a second glance and measuring with a more accurate yardstick, we start with the typical values 1/4, 1/2 or 3/4. Then make a correction of -1/8, 0 or +1/8 and note whether x is actually a little to the left, in the middle or to the right. This process can continue infinitely or as far as your measuring sticks allow. The result is a non-deterministic assignment of an infinite string of letters L,M or R to the point x, except that the string should not end in all L's or all R's.

Making this precise is straightforward:

x is represented by an initial string
$s = a_1a_2...a_n$ of letters L,M and R
if and only if
x in $(s - e_n, s + e_n)$ (call this interval U(s))

where
$$s = 1/2 + \underline{a_1}/4 + \underline{a_2}/8 + ... + \underline{a_n}/(2^{n+1}),$$
$$e_n = 1/(2^{n+1}),$$
$$\underline{L} = -1,\ \underline{M} = 0,\ \underline{R} = 1.$$

Some numbers have unique representations in this system (e.g. 0.5 = MMMM.....), some have a finite number (e.g. 0.375 = MLMMM...or LRMMM....) and some have an infinite number (e.g. 1/3). The non-uniqueness is forced on us by the necessity for **overlapping** open sets. The overlapping may be described as follows: let s and t be strings of L,M,R of length n, then

$$U(s^*) \cap U(t^*) \neq \emptyset$$
if and only if
s may be transformed into t by a finite sequence of steps:

RL ⟷ *MR*
LR ⟷ *ML*

plus one step affecting the last symbol only of the form:

*R ⟷ *M or
*L ⟷ *M.

This is easy to prove.

Now consider the more complicated problem of describing 1D shapes X. We assume that by a preliminary normalization of X using translation and scaling, X is a closed subset of the interval [0,1]. Let
$$X(e) = X \oplus [-e/2,+e/2].$$
Our hierarchical description of X is based on first examining X(1/4), then X(1/8), then X(1/16), etc. The full infinite description determines X itself, but partial descriptions carried out to order n-1 will define an open set in $SH1(1/2^n)$, the space of shapes $X(1/2^n)$.

We first describe the top level categories for Y = X(1/4). Let I be a subset of {-1,0,1,2,3,4,5} containing -1 and 5, and let

U(I) = the set of all Y=X(1/4) such that
i) for all i in I, i/4 is not in Y,
ii) for all other i's, Y meets the open interval (i/4 - 1/16, i/4 + 1/16).

It is immediate that these U(I)'s cover SH1(1/4). Now consider each I as a union of 'components', where is a component is just a set of consecutive integers {i,i+1,...,j} in I, with i-1 and j+1 not in I. If {i,..,j} is such a set of consecutive integers in I, and Y is in U(I), then the whole closed interval [i/4,j/4] is disjoint from Y: this is because Y is made up of intervals of length at least 1/4, so there is no room for such an interval between consecutive points i/4. Thus the first component of I defines an interval to the left of Y, the last defines an interval to the right of I and the other components define gaps in the closed set Y. Moreover, if Y is in U(I), then any gap in Y of length at least 3/8 would have to be marked by an i in I with i/4 in the gap. So we divide up the I's and Y's into qualitatively different categories as follows:

i) I's like {-1,0,5} or {-1,3,4,5} whose only components are the strings starting at -1 and the one ending at 5; Y's in the corresponding U(I)'s are said to have **small gaps**.
ii) I's with one middle component {2} or {2,3}; Y's in the corresponding U(I)'s are said to have **gap left.**
iii) I's with one middle component {3} or {2,3,5}; Y's in the corresponding U(I)'s are said to have **gap middle.**
iv) I's with one middle component {4} or {4,5}; Y's in the corresponding U(I)'s are said to have **gap right.**
v) I's with two middle components {2} and {4}: Y's in the corresponding U(I)'s are said to have **gap left and right.**

The same covering can readily be extended to finer levels. It should be noted beforehand that as one looks at an arbitrary closed set X at finer and finer scales, one of two things can happen: either X has a small number of clear and distinct 'features' at each level, or else at some scale, X becomes 'textured', i.e. many gaps of roughly the same size appear in an interval which has been unbroken by gaps over a long series of coarser scales. Let me give examples:

Ex.1: NO TEXTURE - A Cantor set or a finite approximation thereof: one big gap (1/8,7/8) is taken out of [0,1] leaving [0,1/8] [7/8,1]. Then from each subinterval smaller gaps (1/64,7/64) and (57/64,63/64) are removed and so on recursively.

Ex.2: TEXTURE: [0,1] appears whole until the scale of 1/128 is reached. Then 20 random subintervals (i/128,(i+1)/128) are removed.

In the former case, a recursive tree-like data structure with a small number of qualitative choices at each node will describe the shape; in the latter case, there will be some nodes where many bits are needed to describe the texture fully and where, for most purposes, a statistical summary would be preferred.

Here are the details. Let $N = 2^n$, let e = 1/N, and let I be any subset of {-1, 0, 1, 2,..., N-1,N,N+1} containing -1 and N+1. Let

U(I,e) = the set of all Y = X(e) such that
i) for all i in I, i*e is not in Y,
ii) for all other i's, Y meets the open interval (i*e - e/4, i*e + e/4).

As before, the U(I,e)'s cover SHl(e). Moreover, each sequence I can be broken up into 'components' {i,i+1,...,j}, and for each such component the closed interval [i*e,j*e] is disjoint from Y. Thus the set of components of I mirrors the set of components of the complement CY of Y, except that small gaps in Y of length less than 3*e/2 may sometimes be ignored.

This covering {U(I,e)} of SHl(e) **refines** the analogous covering {U(I',2*e)} of SHl(2*e). Thus if I' is a subset of { -1, 0,.., N/2, N/2+1 } and p : SHl(e) ⟶ SHl(2*e) is the map taking Y to Y $\oplus$ [-e/2,+e/2], then p^{-1}(U(I',2*e)) is covered by the sets U(I) where I runs over the subsets of {-1,0,...,N,N+1} such that:

for each component {i',i'+1,...,j'} of I', I has one of the sequences {i,i+1,...,j} as a component, where

i = 2*i', 2*i'-1 or 2*i'-2,
j = 2*j', 2*j'+1 or 2*j'+2.

We can now put together a full description of a closed set X by an infinite tree, whose structure down to level n corresponds to the choice of an open set $U(I_n,1/2^n)$ containing $X(1/2^n)$. At level n, we imagine $X(1/2^n)$ divided up by the gaps identified in I_n and make one node for each resulting piece. Between levels, we link a piece Z of $X(1/2^n)$ to the piece of $X(1/(2^{n-1}))$ containing Z $\oplus$ $[-1/(2^{n+1}),+1/(2^{n+1})]$. At each node, we must record 2 things. The first are the rough locations of the left and right endpoints of the corresponding piece Z of $X(1/2^n)$, i.e. the integers i and j in I_n such that $i/2^n$ and $j/2^n$ bracket Z. Given the rough locations of the higher node Z is linked to, we have seen that there are three choices for i and three for j. In fact, going down the tree, we read off from which of the three choices L,M or R is made at each level, precisely a ternary expansion of the endpoints of part of X in the system explained at the beginning of the section. The other thing that must be recorded at each node are the rough locations of new gaps at which Z is broken up at the next level. These are simply given by giving the corresponding intervals {i,..,j} of I_n. For each new gap, the node defined by Z splits off an extra child node.

We omit a formal description (i) of these trees and the open sets in SHl defined by the top n levels of such a tree, (ii) of when two such open sets overlap and (iii) of the way the trees are transformed when passing from one open set to an overlapping one. Full details will appear elsewhere.

---

[1] The discussion which follows has been strongly influenced by the penetrating paper "One the expression of synchronized motion in mathematical logic", by H.Madjid and J.Myers (preprint, Div.of Applied Sci., Harvard Univ., 1986).

[2] G.Miller & P.Johnson-Laird, Language and Perception, Harvard Univ. Press, 1976, p.46.

[3] See also comments in A.Latto, D.Mumford and J.Shah, The Representation of Shape, Proc. of IEEE Workshop on Computer Vision, Annapolis, 1984.

[4] This example was discussed by T.Pavlidis in a lecture at Harvard in 1984.

[5] The idea of using this interesting metric is due to P.Diaconnis.

[6] i.e. (i) the maps p(e) are continuous, (ii) for any family of shapes X_e such that $p(f-e)(X_e) = X_f$, there is a shape X with $X_e = p(e)(X)$, and (iii) we can guarantee $d(X,Y) < e$ by requiring $d(X(e_1),Y(e_1)) < e_2$ (e.g. $e_1 = e_2 = e$).