Shape Description and Classification using the Interrelationship of Structures at Multiple Scales *

Gregory Dudek

McGill Research Centre for Intelligent Machines, McGill University, Montreal, Quebec, Canada H3A 2A7

Abstract. This paper deals with the classification of objects described by planar curves in an image. Invariance to deformation is an important aspect of shape representation and two representations are described with different degrees of such invariance. One of these is a new statistical method for shape description exhibiting a large degree of such invariance.

Using scale-space to describe shape statistically allows for a texture-like form of object classification. The scale-space used is one based on curvature-tuned smoothing (CTS). This allows a curve to be represented as a set of descriptors at various scales. The spatial correlation of these descriptors produces a statistical description of a contour that has similarities to a large-scale texture measure. The texture being measured is, in fact, the combination of substructures that define the object's shape.

Keywords: shape, shape description, classification, object recognition, recognition, regularization, tuned smoothing, active contours, statistical shape, scale-space, texture, differential geometry.

1 Introduction

For the purposes of object recognition, an object's shape is its most important characteristic. Computational approaches to shape-based recognition have largely focused on shape matching based on shape similarity as a template-like matching process combined with a limited amount of deformation (notwithstanding several exceptions noted below). Vision-based object recognition amounts to a process of finding the exemplar shape from a library of models whose contours best match the input shape according to some distance measure. This approach fails to describe the alternative types of shape-based recognition that is performed by humans. Consider the recognition of 2-dimensional objects such as the silhouettes of clouds or plants: such objects are eminently recognizable from their silhouettes but are often highly dissimilar in any template-like sense.

* The author gratefully acknowledges the financial support of the Natural Sciences and Engineering Research Council and the comments of M. Langer, J. K. Tsotsos and S. W. Zucker.

1.1 Representational Constraint

Despite its intuitiveness, the concept of what it means for objects to have similar shapes is surprisingly hard to define. This may be, in part, because there are multiple mechanisms that contribute to the concept of shape [8, 24].

Computationally, there are several classes of shape-description techniques which can be organized along a continuum or taxonomy according to their degree of *representational shape constraint*; that is, the degree of spatial freedom they permit in individual parts (or sub-parts) of an object without change to the representation (or the deformation invariance properties). Template-like representations are the most constraining allowing almost no deformation in an object's shape [3], metric representations with parameterized deformation are somewhat less constraining [20, 14], representations based on feature topology are less constraining still [1], and finally statistical shape description, a method described below, captures shape properties with extremely little positional constraint on the individual sub-shapes or features.

Two matching methods along this continuum are presented based on the same input primitives. One method is a minimum-deformation matching method, the other is a new method for shape description and representation based on statistical properties of an object's shape. The complex relationship between spatial scale and object structure has become apparent in attempts to describe object shape computationally [25, 4, 15, 12]. The statistical shape-recognition method exploits the multi-scale aspect of object shape by describing objects in terms of the interrelationship between different shape features at a single location of an object contour. This leads to an object similarity measure that associates objects having similar structural properties even when they are dissimilar in a template matching or part-by-part sense. The notion of statistical shape properties and the relationship between different scales has some similarities to a texture measure [16, 18]. A key difference from conventional microtexture descriptors [23, 9] is that the primitive features here are large-scale shape primitives.

2 Curvature-scale-space Description

Shape primitives can be extracted using a variational method called curvature-tuned smoothing [5, 6]. This description has its basis in curvature measurements [2, 13], and tolerates sparse data or noise [19, 22]. The multi-scale nature of the representation allows multiple alternative descriptions for portions of a curve to be retained. It produces a description of a curve where a single region may be described in terms of one or more arcs of different curvatures (of one or more sizes), and hence makes the information at different spatial scales explicit. The term scale is used to refer to the size or spatial extent of a processing operation or feature.

The curve representation is produced by repeatedly minimizing the following energy functional with respect to a piecewise C^2 solution $\mathbf{u}(t) = (x(t), y(t))$:

$$E(\mathbf{u}(t), c) = \int_{t_a}^{t_b} ||\mathbf{u}(t) - \mathbf{d}(t)||^2 + \alpha p(\mathbf{u}(t)) + \lambda(c)(\kappa_u(t) - c)^2 \, dt, \qquad (1)$$

where t is the arc length, $\mathbf{d}(t) = (x(t), y(t))$ is a list of initial data points estimating the input curve, $p(x, y)$ is a potential function derived from the input image (i.e. a measure of edge strength), $\kappa_u(t)$ is the curvature of $\mathbf{u}(t)$, c is the *curvature tuning*, α is a constant, and λ is the stabilizing constant selected as a function of c. This solution is determined for various values of c, denoted by c_i. The first two terms constrain the solution to be consistent with an initial input description and with image support for the curve position. The third term expresses an *a priori* bias for a solution with a specific curvature given by c.

In practice, the discrete form of this equation is used:

$$\sum_{t \in \text{data}} ||\mathbf{u}_i(t) - \mathbf{d}(t)||^2 + \alpha p(\mathbf{u}_i(t)) + \lambda(c)(1 - l_i(t))(\kappa_u(t) - c)^2, \qquad (2)$$

where $l_i(t)$ is an independent Boolean discontinuity function (line process) at each scale. Discontinuities are progressively inserted at each scale to satisfy a smoothness criterion.

For each value of the tuning parameter, a slightly different solution curve $\mathbf{u}(t)$ is produced that reflects structure. This combines smoothing of the input data akin to that of active contours models (i.e. *snakes* [11]), with model fitting at multiple scales although the process can also be used directly on a parameterized input curve (i.e. with $\alpha = 0$) [6].

The use of multiple alternative stabilizers for curvature-tuned smoothing leads to selecting not only various structures at different curvatures, but also structures with different spatial extents. Low curvature segments are components of circles with large radii. Conversely, the segments selected when the curvature tuning is large must also have large curvatures. As a result, differently tuned stabilizers lead to different sets of discontinuities that decompose the curve into different segments.

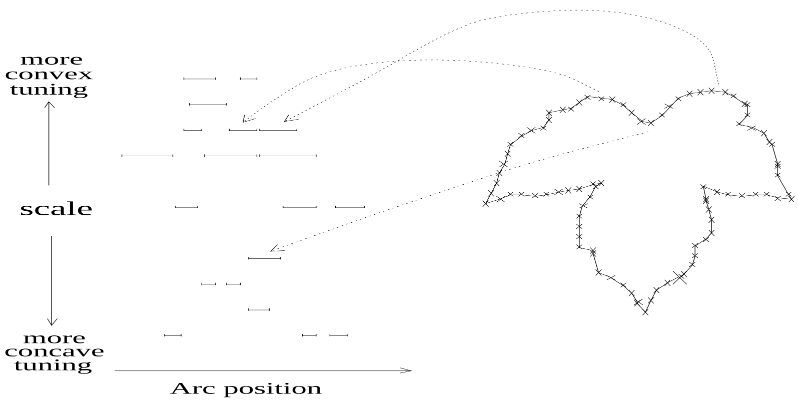

Fig. 1. Poison sumac leaf and scale-space.

The description of the poison sumac leaf (object s1) extracted using curvature-tuned smoothing. Segments corresponding to certain features on the leaf illustrated.

2.1 Abstraction into Segments

From the set of arc-like segments produced by the minimization operations it is possible to extract a small subset of the segments with high smoothness as a simplified description [7, 6]. These are the segments that best match the input data since their low energy implies that they had to deform least to suit the data (such a description is shown in Fig. 1). The segments themselves are sections of *approximately* uniform curvature, yet together they capture most of a curve's structure. The structure of each segment is so simple that it is unnecessary to retain all the internal point locations. As a coarse description the curve segments can be encoded only by their initial and final positions (t_j^I and t_j^F) and the curvature tuning c used to extract them. This encoding will be referred to as the *segment descriptor* for a segment j:

$$s_j = (t_j^I, t_j^F, c_j). \tag{3}$$

The set of segment descriptors for an object o constitutes its description $\mathcal{S}(o)$:

$$\mathcal{S}(o) = \bigcup_j s_j. \tag{4}$$

.

3 Matching with Deformation

Dynamic programming is one of the techniques used to match curves based on a sequence of extracted primitives such as those described above [10]. By constructing a matching function that ensures that matched curves have the same *sequence* of (multi-scale) primitives, matching is made insensitive to local deformations in a curve. For two segments s_1 and s_2 the mismatch is measured as $\langle s_1, s_2 \rangle_s = w_1 |\log c_1 - \log c_2| + |l_1 - l_2|$, where w_1 is a constant and l_1 and l_2 are the segment lengths ($|t_j^I - t_j^F|$). Note that logarithmic weighting is applied to the curvature components to impose a preference for coarse-scale information [25].

Curve matching can be formulated as a dynamic programming problem in terms of matching an increasingly long subsequence of segments from one curve to a series of segments from the other. Invariance to the initial position on either curve can be achieved by doubling the series of tokens and looking only for a substring of half the total length [6]. This has been demonstrated using an algorithm that constructs an incrementally expanded table of costs such that for two curves composed of segments, entry $C(i, j)$ in the cost table reflects the match the first i segments from one curve makes with the first j segments from the other. The process of matching one contour with another is then a process of executing the dynamic program for an observed data set against the set of models.

This procedure has been shown to be appropriate for matching curves that are noisy versions of one another or that have undergone a limited amount of deformation [6]. For pairs of curves that have significant structural variations with respect to one another, there will be substantial mismatch error. For many natural processes structural variations may be present at a global level while sub-parts and local structures are similar. It has been suggested that one way in which this can occur is when local generative processes at different scales are combined in a pseudo-random or non-rigid manner [17, 25]. In such cases the alternative approach described below may be appropriate for shape recognition.

4 Statistical Measurement

Conventional approaches to curve recognition using local characteristics, such as the one described above, are based on determining the position of features on a curve and then using the position or spatial topology of these features for recognition. The approach described here as *scale-space statistics* is an alternative to using the relative locations of features on a curve for object recognition or classification.

At a given scale, the ease with which a curve can be described as having a given curvature c can be considered as a one-dimensional signal similar to a goodness-of-fit and will be denoted by

$$\phi(t, c) \in 0, 1 \tag{5}$$

that varies along a curve. A simple form of $\phi(t, c)$ is a binary function that indicates whether any segment descriptor having curvature c spans point t:

$$\phi(t, c) = \begin{cases} 1 \text{ iff } \exists\ s_j = (t_j^I, t_j^F, c)\ \in\ \mathcal{S}(o) \text{ and } t_j^I \le t \le t_j^F, \\ 0 \text{ otherwise }. \end{cases} \tag{6}$$

By observing the mean value $\overline{\phi(c)}$ of this function, we can describe "how much" of a contour can be well-approximated at the given curvature.

The similarity between the one-dimensional functions $\phi(\cdot, c)$ for different values of c indicates the inter-relationship between the different-scale substructures that make up the curve at each point. As noted above and in the texture literature, specific statistical interrelationships are characteristic of many shapes including a variety of natural forms. Common examples include the trunks of trees, typified by a large-scale cylindrical curve combined with fine-scale bark patterns, geological formations, or the way the bumps and ridges on the leaves of a tree are combined. Note also that many objects are recognizable even though the sequence of sub-curves that compose them may be highly variable (Fig. 2).

The cross-correlation matrix $\mathcal{C}$ has elements defined by

$$C_{ij} = \int \frac{(\phi(t, c_i) - \overline{\phi(c_i)})(\phi(t, c_j) - \overline{\phi(c_j)})}{\sigma_{\phi(c_i)}\ \sigma_{\phi(c_j)}} dt \tag{7}$$

between this value at one curvature and the value of this function at another curvature. It provides a measure of what types of substructure in curvature space occur within a structure at another scale. This corresponds to taking a slice of

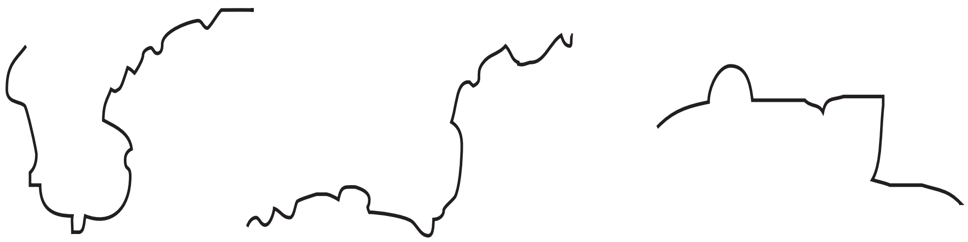

Fig. 2. Statistically similar objects. The first two coastal curves are similar in a structural or statistical sense, yet they cannot be globally deformed into one another easily; the third is different.

Fig. 3. Sample input curves. Left to right, top to bottom: r1 (raspberry), m1 (maple), a9 (unknown), r2 (raspberry).

the scale-space for a fixed position and measuring the statistical likelihood of features at one scale given the presence (or absence) of features at another scale.

Together, the vector $\overline{\phi}$ and the matrix $\mathcal{C}$ provide a statistical description of a curve which is similar to a texture measure for an intensity pattern. Whereas texture is often measured by decomposing a signal into different components such as bandpass channels [23, 21], the statistical shape measure presented here relates texture to the goodness-of-fit of shape operators at different curvature-based scales.

For appropriate classes of shapes, these statistical scale-space measures can be used directly for shape matching. The simplest such shape measure for two shapes o_1 and o_2 being compared is

$$\mathcal{M}(o_1, o_2) = \frac{\mathcal{C}_1 \cdot \mathcal{C}_2}{||\mathcal{C}_1||||\mathcal{C}_2||}, \tag{8}$$

where $\cdot$ denotes the dot or inner product. Shapes with identical scale-space statistics thus match with value 1, while unrelated shapes have a match score of zero.

Since $\mathcal{C}_1$ and $\mathcal{C}_2$ have uniform diagonals caused by autocorrelation, $\mathcal{M}$ has a positive offset. Cross-talk between the responses at different scales leads to a consistent positive bias for near diagonal elements as well. This off-diagonal coupling across scales, however, cannot readily be estimated *a priori* and depends

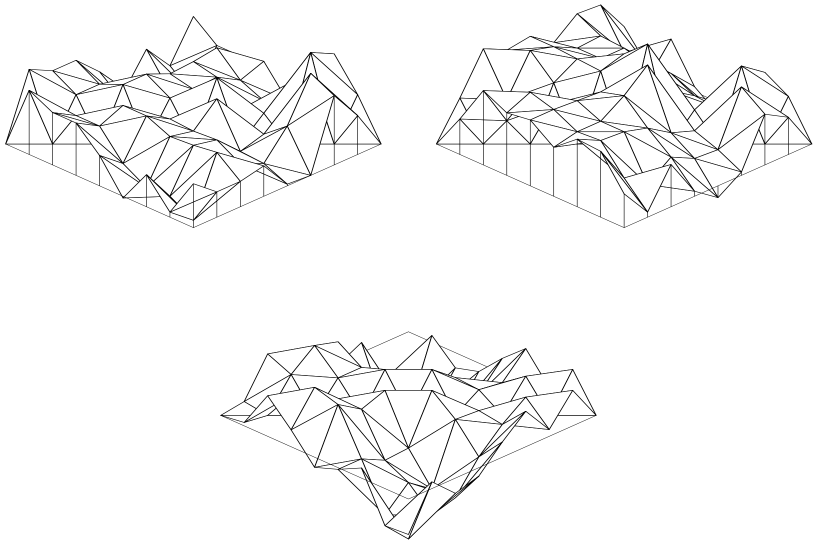

Fig. 4. Scale-space correlation surfaces. Three scale-space correlation surfaces for three different curves from the previous figure (leaf silhouettes). Curvature tuning (or scale) varies along each axis and the amplitude at any point reflects the correlation between ϕ signals for the two curvatures. The top two surfaces are from two different leaves of the same type (examples r1 and r2 at different orientations in depth). The lower surface is from a different type of leaf (example a9); note its qualitatively different profile.

on non-linear discontinuity effects in the original solutions. Hence, an additional heuristic is of utility: elements (correlations) of $\mathcal{C}$ that are well off the diagonal, corresponding to correlations between signals well separated in scale, can be more heavily weighted. This is further grounded in the observation that, in general, structures at different scales are independent except where non-accidental processes lead this to be otherwise; hence such structural correlations are especially salient [26, 17].

Hence, we have a refined measurement of the form:

$$\mathcal{M}^w(o_1, o_2) = \frac{\mathcal{C}_1 \odot W \cdot \mathcal{C}_2 \odot W}{||\mathcal{C}_1 \odot W|| \ * \ ||\mathcal{C}_2 \odot W||}, \tag{9}$$

where $\odot$ denotes the Hadamard product $(C(i,j) * W(i,j))$ and W is a weighting matrix of the form:

$$W(i,j) = 1 - e^{-|i-j|}. \tag{10}$$

In this way, an improved signal-to-noise ratio for the matching task is obtained.

5 Results

The results of matching particular contours (e.g. object *m1* of Fig. 3) to several others are tabulated below, each to two significant figures (the first letter indicates the leaf species, the numerical suffix indicates the example; m and r species are intuitively similar):

Curve	$\mathcal{M}^w(m1,\cdot)$	$\mathcal{M}^w(t1,\cdot)$	$\mathcal{M}^w(a9,\cdot)$
m1	1.0	0.65	0.29
m2	0.79	0.72	0.44
r1	0.73	0.87	0.24
r2	0.71	0.78	0.17
s1	0.65	0.54	0.21
t1	0.65	1.0	0.086
a9	0.29	0.086	1.0

Note that the m1 and r1 contours and their deformed versions are rated similar to one another while other contours have much lower scores.

The statistical representations $\mathcal{C}$ and the matching function $\mathcal{M}^w$ describe the relationship between structures of different types without regard for the precise spatial arrangement of the structures. For example a large bump may equivalently contain several concavities without regard for the positions of the concavities with respect to one another. This form of view invariance has both advantages and shortcomings. A disadvantage of this coarse abstraction of a curve's shape is that it is insensitive to a large variety of possible variations in the object, in particular those that are obtained by reordering the major sections of the shape. On the other hand, this abstraction permits measurement of the the similarity between different shapes that have the same cross-scale structure because they are composed of the same building-block parts, but in different numbers or arrangements. For example, various natural forms such as cloud types are typified by the combination and co-occurrence of particular forms at multiple scales, for example lobes with serrations, whereas the specific spatial arrangement of the forms is highly variable. In essence, this simple shape measurement is best suited to classes of objects where a small number of interacting generative processes are responsible for each object, and each of these processes can be typified as creating subshapes at a particular scale but with random or hard-to-typify spatial arrangements. This characterization appears to be appropriate for many types of natural form such as rocks, leaves, microscopic particles, and clouds.

6 Conclusion

The use of a collection of curvature-based minimizing operators, which are termed collectively *curvature-tuned smoothing*, has been previously developed to address several difficulties with existing approaches to smoothing, interpolation, segmentation, and curve description. Using this representation as input, techniques for describing and recognizing objects via the sequencing of descriptors along the curve and via the correlation statistics of the descriptors in this space have been outlined.

The statistical method provides a notion of recognition based on structural regularities in shape rather than direct point-to-point similarity. As such, it allows objects to be recognized or deemed alike even when they have no identical sub-contours. Because the primitive elements in this description (bumps and valleys) are perceptually and functionally salient, the shape-similarity space can be described in intuitive or generative terms (for example it can be related to processes that produce bumps and valleys). Statistical shape description can also be formulated in terms of alternative primitive shapes if this is appropriate to specialized domains. This particular class of similarity appears well suited to the recognition of certain classes of biological and geological forms where the structural characteristics are common to the class, but individual members vary in terms of their particular layout.

The two matching techniques presented illustrate very different positions along a proposed continuum for the classification of shape matching methods.

1. Ansari, N. and Delp, E. J. (1990). Partial shape recognition: A landmark-based approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(5), pp. 470–483.
2. Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61, pp. 183–193.
3. Cass, T. A. (1988). A robust parallel implementation of 2-d model based recognition. In: *Proc. of the Conf on Computer Vision and Pattern Recognition*, Ann Arbor, MI., pp. 879–884.
4. Crowley, J. L. and Parker, A. C. (1984). A representation for shape based on peaks and ridges in the difference of low-pass transform. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1(2), pp. 156–170.
5. Dudek, G. and Tsotsos, J. K. (1989). Using curvature information in the decomposition and representation of planar curves. *NATO Advanced Study Institute of Robotics and Active Vision*, Maratea, Italy.
6. Dudek, G. and Tsotsos, J. K. (1991). Shape representation and recognition from curvature. In: *Proceedings of the 1991 Conference on Computer Vision and Pattern Recognition*, Maui, Hawaii. IEEE Press, pp. 35–41.
7. Dudek, G. L. (1990). Shape representation from curvature. PhD Thesis, Dept. of Computer Science, University of Toronto, Toronto, Canada.
8. Fischler, M. A. and Bolles, R. C. (1983). Perceptual organization and the curve partitioning problem. In: *Proc. of the International Joint Conf. on Artificial Intel.*, pp. 1014–1018, Karlsruhe, Germany.
9. Fogel, I. and Sagi, D. (1989). Gabor filters as texture discriminators. *Biological Cybernetics*, 61, pp. 103–113.
10. Gorman, J. W., Mitchell, O. R., and Kuhl, F. P. (1988). Partial shape recognition using dynamic programming. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 10(2), pp. 257–266.
11. Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4), pp. 321–331.

12. Kimia, B. B., Tannenbaum, A., and Zucker, S. W. (1990). Toward a computational theory of shape: An overview. In: *Proceedings of the First European Conference on Computer Vision*, Antibes, France.
13. Koenderink, J. J. and van Doorn, A. J. (1980). Photometric invariants related to solid shape. *Optica Acta*, 27(7), pp. 981–996.
14. Milios, E. (1988). Recovering shape deformation by an extended circular image representation. In: *Proceedings of the 2nd Interlnational Conf. on Computer Vision*, Tarpon Springs, FL., IEEE Press, pp. 20–29.
15. Mokhtarian, F. and Mackworth, A. (1986). Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(1), pp. 34–43.
16. Pentland, A. P. (1984). Fractal-based description of natural scenes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6(6), pp. 661–674.
17. Pentland, A. P. (1985). Perceptual organization and the representation of natural form. technical note 357, SRI International.
18. Pentland, A. P. (1987). Perceptual organization and the representation of natural form. In: Fischler, M. A., Firschein (eds.), *Readings in Computer Vision (also in SRI TR-357 1985)*, Morgan Kaufman Publishers, Los Altos, California, pp. 680–698.
19. Rektorys, K. (1980). *Variational Methods in Mathematics, Science and Engineering*. Reidel, Dordrecht, Holland.
20. Solina, F. (1987). Shape recovery and segmentation with deformable part models. PhD Thesis, Dept. of Computer and Information Science, Univ. Pennsylvania.
21. Subirana-Vilanova, J. B. (1991). On contour texture. In: *Proc. computer vision and pattern recognition 1991*, Maui, HA. IEEE Computer Society, pp. 753–754.
22. Terzopoulos, D. (1986). Regularization of inverse visual problems involving discontinuities. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(4), pp. 413–424.
23. Turner, M. R. (1986). Texture discrimination by gabor functions. *Biological Cybernetics*, 55, pp. 71–82.
24. Warrington, E. K. and Taylor, A. M. (1978). Two categorical stages of object recognition. *Perception*, 7, pp. 695–705.
25. Witkin, A. P. (1983). Scale-space filtering. In: *Proc. 3rd Internat. Joint Conf on Artificial Intelligence*, volume 2, Karlsruhe, West Germany.
26. Witkin, A. P. and Tenenbaum, J. M. (1983). On the role of structure in vision. In: Rosenfeld, J., Beck. B. and Hope. B. (eds.), *Human and Machine Vision*. Academic Press.

This article was processed using LaTeX with SHAPE.sty modified from LMAMULT