SCALE-SPACE FILTERING

Andrew P. Witkin

Fairchild Laboratory for Artificial Intelligence Research

ABSTRACT—The extrema in a signal and its first few derivatives provide a useful general-purpose qualitative description for many kinds of signals. A fundamental problem in computing such descriptions is scale: a derivative must be taken over some neighborhood, but there is seldom a principled basis for choosing its size. Scale-space filtering is a method that describes signals qualitatively, managing the ambiguity of scale in an organized and natural way. The signal is first expanded by convolution with gaussian masks over a continuum of sizes. This "scale-space" image is then collapsed, using its qualitative structure, into a tree providing a concise but complete qualitative description covering all scales of observation. The description is further refined by applying a stability criterion, to identify events that persist of large changes in scale.

1. Introduction

Hardly any sophisticated signal understanding task can be performed using the raw numerical signal values directly; some description of the signal must first be obtained. An initial description ought to be as compact as possible, and its elements should correspond as closely as possible to meaningful objects or events in the signal-forming process. Frequently, local extrema in the signal and its derivatives—and intervals bounded by extrema—are particularly appropriate descriptive primitives: although local and closely tied to the signal data, these events often have direct semantic interpretations, e.g. as edges in images. A description that characterizes a signal by its extrema and those of its first few derivatives is a *qualitative* description of exactly the kind we were taught to use in elementary calculus to "sketch" a function.

A great deal of effort has been expended to obtain this kind of primitive qualitative description (for overviews of this literature, see [1,2,3].) and the problem has proved extremely difficult. The problem of *scale* has emerged consistently as a fundamental source of difficulty, because the events we perceive and find meaningful vary enormously in size and extent. The problem is not so much to eliminate fine-scale noise, as to separate events at different scales arising from distinct physical processes.[4] It is possible to introduce a *parameter of scale* by smoothing the signal with a mask of variable size, but with the introduction of scale-dependence comes ambiguity: every setting of the scale parameter yields a different description; new extremal points may appear, and existing ones may move or disappear. How can we decide which if any of this continuum of descriptions is "right"?

There is rarely a sound basis for setting the scale parameter. In fact, it has become apparent that for many tasks no one scale of description is categorically correct: the physical processes that generate signals such as images act at a variety of scales, none intrinsically more interesting or important than another. Thus the ambiguity introduced by scale is inherent and inescapable, so the goal of scale-dependent description cannot be to eliminate this ambiguity, but rather to manage it effectively, and reduce it where possible.

This line of thinking has led to considerable interest in multi-scale descriptions [5,2,6,7]. However, merely computing descriptions at multiple scales does not solve the problem; if anything, it exacerbates it by increasing the volume of data. Some means must be found to organize or simplify the description, by relating one scale to another. Some work has been done in this area aimed at obtaining "edge pyramids" (e.g. [8]), but no clear-cut criteria for constructing them have been put forward. Marr [4] suggested that zero-crossings that coincide over several scales are "physically significant," but this idea was neither justified nor tested.

How, then, can descriptions at different scales be related to each other in an organized, natural, and compact way? Our solution, which we call *scale-space filtering,* begins by continuously varying the scale parameter, sweeping out a surface that we call the *scale-space image.* In this representation, it is possible to track extrema as they move continuously with scale changes, and to identify the singular points at which new extrema appear. The scale-space image is then collapsed into a tree, providing a concise but complete qualitative description of the signal over all scales of observation.[1]

2. The Scale-Space Image

Descriptions that depend on scale can be computed in many ways. As a primitive scale-parameterization, the gaussian convolution is attractive for a number of its properties, amounting to "well-behavedness": the gaussian is symmetric and strictly decreasing about the mean, and therefore the weighting assigned to signal values decreases smoothly with distance. The gaussian convolution behaves well near the limits of the scale parameter, σ, approaching the un-smoothed signal for small σ, and approaching the signal's mean for large σ. The gaussian is also readily differentiated and integrated.

The gaussian is not the only convolution kernel that meets these criteria. However, a more specific motivation for our choice is a property of the gaussian convolution's

---

[1] A complementary approach to the "natural" scale problem has been developed by Hoffman [9].

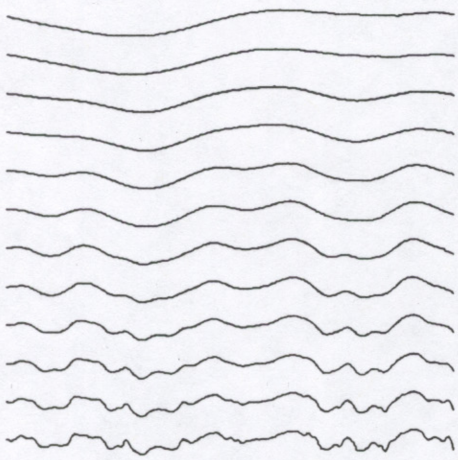

Figure 1. A sequence of gaussian smoothings of a waveform, with σ decreasing from top to bottom. Each graph is a constant-σ profile from the scale-space image.

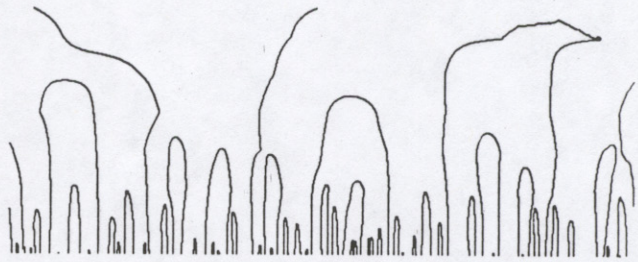

Figure 2. Contours of $F_{xx} = 0$ in a scale-space image. The x-axis is horizontal; the coarsest scale is on top. To simulate the effect of a continuous scale-change on the qualitative description, hold a straight-edge (or better still, a slit) horizontally. The intersections of the edge with the zero-contours are the extremal points at some single value of σ. Moving the edge up or down increases or decreases σ.

zero-crossings (and those of its derivatives): as σ decreases, additional zeroes may appear, but existing ones cannot in general disappear; moreover, of convolution kernels satisfying "well behavedness" criteria (roughly those enumerated above,) the gaussian is the *only* one guaranteed to satisfy this condition [12]. The usefulness of this property will be explained in the following sections.

The gaussian convolution of a signal $f(x)$ depends both on x, the signal's independent variable, and on σ, the gaussian's standard deviation. The convolution is given by

$$F(x,\sigma) = f(x) * g(x,\sigma) = \int_{-\infty}^{\infty} f(u)\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-u)^2}{2\sigma^2}}\,du, \tag{1}$$

where "$*$" denotes convolution with respect to x . This function defines a surface on the (x,σ)-plane, where each profile of constant σ is a gaussian-smoothed version of $f(x)$, the amount of smoothing increasing with σ. We will call the (x,σ)-plane ***scale space*** , and the function, F, defined in (1), the ***scale-space image*** of f.[2] Fig. 1 graphs a sequence of gaussian smoothings with increasing σ. These are constant-σ profiles from the scale-space image.

At any value of σ, the extrema in the nth derivative of the smoothed signal are given by the zero-crossings in the $(n+1)$th derivative, computed using the relation

$$\frac{\partial^n F}{\partial x^n} = f * \frac{\partial^n g}{\partial x^n},$$

where the derivatives of the gaussian are readily obtained. Although the methods presented here apply to zeros in any derivative, we will restrict our attention to those in the second. These are extrema of slope, i.e. inflection points. In terms of the scale-space image, the inflections at *all* values of σ are the points that satisfy

$$F_{xx} = 0, F_{xxx} \neq 0, \tag{2}$$

using subscript notation to indicate partial differentiation.[3]

3. Coarse-to-fine Tracking

The contours of $F_{xx} = 0$ mark the appearance and motion of inflection points in the smoothed signal, and provide the raw material for a qualitative description over all scales, in terms of inflection points. Next, we will apply two simplifying assumptions to these contours: (1) the ***identity*** assumption, that extrema observed at different scales, but lying on a common zero-contour in scale space, arise from a single underlying event, and (2) the ***localization*** assumption, that the true location of an event giving rise to a zero-contour is the contour's x location as $\sigma \to 0$.

Referring to fig. 2, notice that the zero contours form arches, closed above, but open below. The restriction that zero-crossings may never disappear with with decreasing σ (see section 2) means that the contours may *never* be closed below. Note that at the apexes of the arches, $F_{xxx} = 0$, so by eq. (2), these points do not belong to the contour. Each arch consists of a pair of contours, crossing zero with opposite sign.

The ***localization assumption*** is motivated by the observation that linear smoothing has two effects: qualitative simplification—the removal of fine-scale features—and spatial distortion—dislocation, broadening and flattening of the features that survive. The latter undesirable effect may be overcome, by tracking coarse extrema to their fine-scale locations. Thus, a coarse scale may be used to ***identify*** extrema, and a fine scale, to ***localize*** them. Each zero-contour therefore reduces to an (x,σ) pair, specifying its fine-scale location on the x-axis, and the coarsest scale at which the contour appears.

A coarse-to-fine tracking description is compared to the

[2] It is actually convenient to treat $\log\sigma$ as the scale parameter, uniform expansion or contraction of the signal in the x-direction will cause a translation of the scale-space image along the $\log\sigma$ axis.

[3] Note that the second condition in (2) excludes zero-crossings that are parallel to the x-axis, because these are not zero-crossings in the convolved signal.

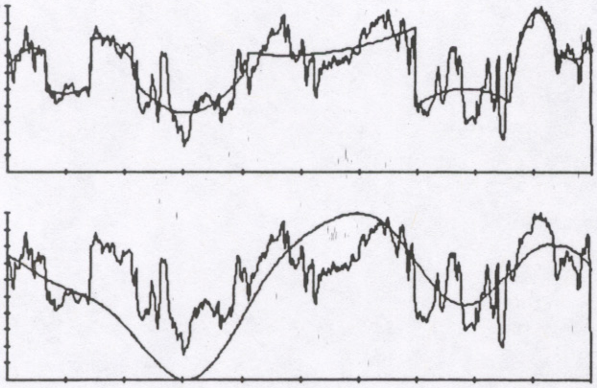

Figure 3. Above is shown a signal with a coarse-to-fine tracking approximation superimposed. The approximation was produced by independent parabolic fits between the localized inflections. Below is shown the corresponding (qualitatively isomorphic) gaussian smoothing.

corresponding linear smoothing in Fig. 3.[4]

4. The Interval Tree

While coarse-to-fine tracking solves the problem of localizing large-scale events, it does not solve the multi-scale integration problem, because the description still depends on the choice of the continuous global scale parameter, σ, just as simple linear filtering does. In this section, we reduce the scale-space image to a simple tree, concisely but completely describing the qualitative structure of the signal over all scales of observation.

This simplification rests on a basic property of the scale-space image: as σ is varied, extremal points in the smoothed signal appear and disappear at singular points (the tops of the arches in fig. 2.) Passing through such a point with decreasing σ, a pair of extrema of opposite sign appear in the smoothed signal. At these points, and only these points, the *un*distinguished interval (i.e. an interval bounded by extremal points but containing none) in which the singularity occurs splits into three subintervals. In general, each undistinguished interval, observed in scale space, is bounded on each side by the zero contours that define it, bounded above by the singular point at which it merges into an enclosing interval, and bounded below by the singular point at which it divides into sub-intervals.

Consequently, to each interval, I, corresponds a node in a (generally ternary-branching) tree, whose parent node denotes the larger interval from which I emerged, and whose offspring represent the smaller intervals into which I subdivides. Each interval also defines a rectangle in scale-space, denoting its location and extent on the signal (as defined by coarse-to-fine tracking) and its location and extent on the scale dimension. Collectively, these rectangles tesselate the (x, σ)-plane. See fig. 4 for an illustration of the tree.

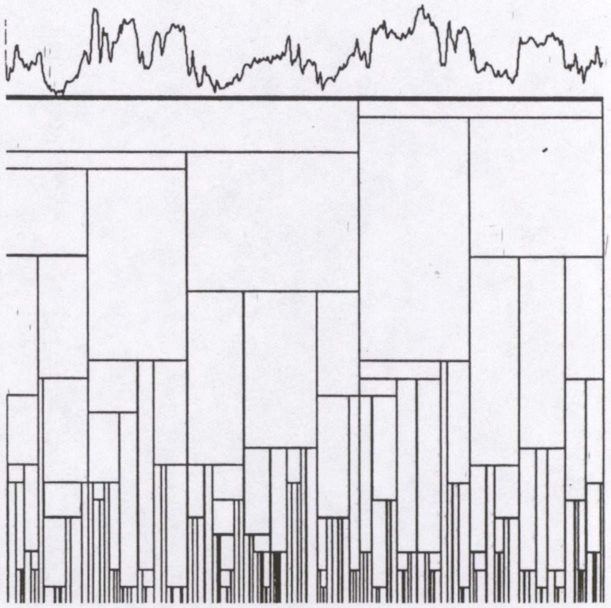

Figure 4. A signal with its interval tree, represented as a rectangular tesselation of scale-space. Each rectangle is a node, indicating an interval on the signal, and the scale interval over which the signal interval exists.

This *interval tree* may be viewed in two ways: as describing the signal simultaneously at all scales, or as generating a family of single-scale descriptions, each defined by a subset of nodes in the tree that cover the x-axis. On the second interpretation, one may move through the family of descriptions in orderly, local, discrete steps, either by choosing to subdivide an interval into its offspring, or to merge a triple of intervals into their parent.[5]

We found that it is in general possible, by moving interactively through the tree and observing the resulting "sketch" of the signal, to closely match observers' spontaneously perceived descriptions. Thus the interval tree, though tightly constrained, seems flexible enough to capture human perceptual intuitions. Somewhat surprisingly, we found that the tree, rather than being too constraining, is not constrained enough. That is, the perceptually salient descriptions can in general be duplicated within the tree's constraints, but the tree also generates many descriptions that plainly have no perceptual counterpart. This observation led us to develop a *stability* criterion for further pruning or ordering the states of the tree, which is described in the next section.

5. Stability

Recall that to each interval in the tree corresponds a rectangle in scale space. The x boundaries locate the interval on the signal. The σ boundaries define the scale range over which the interval exists, its *stability* over scale changes. We have observed empirically a marked correspondence between the stability of an interval and its perceptual salience: those intervals that survive over a broad range of scales

[4] In this and all illustrations, approximations were drawn by fitting parabolic arcs independently to the signal data on each interval marked by the description. This procedure is crude, particularly because continuity is not enforced across inflections. Bear in mind that this procedure has been used only to display the qualitative description.

[5] For previous uses of hierarchic signal descriptions see e.g. [10,11,2].

Figure 5. Several signals, with their maximum-stability descriptions. These are "top-level" descriptions, generated automatically and without thresholds. You should compare the descriptions to your own first-glance "top-level" percepts. (the noisy sine and square waves are synthetic signals.)

tend to leap out at the eye, while the most ephemeral are not perceived at all. To capture this relation, we have devised several versions of a stability criterion, one of which picks a "top-level" description by descending the tree until a local maximum in stability is found. Another iteratively removes nodes from the tree, splicing out nodes that are less stable than any of their parents and offspring. Both of these radically improve correspondence between the interval tree's descriptions and perceptual features (see fig. 5.)

6. Summary

Scale-space filtering is a method that describes signals qualitatively, in terms of extrema in the signal or its derivatives, in a manner that deals effectively with the problem of scale—precisely localizing large-scale events, and effectively managing the ambiguity of descriptions at multiple scales, without introducing arbitrary thresholds or free parameters. The one-dimensional signal is first expanded into a two-dimensional *scale-space image*, by convolution with gaussians over a continuum of sizes. This continuous surface is then collapsed into a discrete structure, using the connectivity of extremal points tracked through scale-space, and the singular points at which new extrema appear. The resulting tree representation is a a concise but complete qualitative description of the signal over all scales of observation. The tree is further constrained using a maximum-stability criterion to favor events that persist over large changes in scale.

We are currently developing applications of scale-space filtering to several signal matching and interpretation problems, and investigating its ability to explain perceptual grouping phenomena. The method is also being extended to apply to two-dimensional images: the scale-space image of a 2-D signal occupies a volume, containing zero-crossing surfaces.[6]

REFERENCES

[1] A. Rosenfeld and A. C. Kak Digital Picture Processing. *Academic Press, New York, New York,* 1976.

[2] D. H. Ballard and C. M. Brown Computer Vision. *Prentice Hall, Englewood Cliffs, New Jersey,* 1982.

[3] Pavlidis, T. *Structural Pattern Recognition.* Springer, 1977

[4] D. Marr Vision, *W. H. Freeman, San Fransisco,* 1982.

[5] Rosenfeld, A. and Thurston, M. "Edge and curve detection for visual scene analysis." *IEEE Transactions on computers, Vol C-20 pp. 562-569.* (May 1971).

[6] Marr, D and Poggio, T. "A computational theory of human stereo vision." *Proc. R. Soc. Lond.,* B. 204 (1979) pp. 301–328

[7] D. Marr and E. C. Hildreth Theory of Edge Detection. *M.I.T. Artificial Intelligence Memo Number 518, Cambridge, Massachusetts,* April 1979.

[8] Hong, T. H., Shneier, M. and Rosenfeld, A. Border Extraction using linked edge pyramids. *TR-1080, Computer Vision Laboratory, U. Maryland,* July 1981

[9] Hoffman, D. Representing Shapes For Visual Recognition *Ph.D. Thesis, MIT, forthcoming.*

[10] Erich, R. and Foith, J., "Representation of random waveforms by relational trees," *IEE Trans. Computers, Vol C-26, pp. 725-736,* (July 1976).

[11] Blumenthal A., Davis, L., and Rosenfeld, R., "Detecting natural 'plateaus' in one-dimensional patterns." *IEEE Transactions on computers,* (Feb. 1977)

[12] J. Babaud, R. Duda, and A. Witkin, *in preparation.*

[6] **Acknowledgments**—I thank my colleagues at FLAIR, particularly Richard Duda and Peter Hart, as well as J. Babaud of Schlumberger, ltd., for their help and encouragement.