From Encyclopedia of Mathematics

Scale-Space Theory

Theory of multi-scale representation of sensory data developed by the image processing and computer vision communities. The purpose is to represent signals at multiple scales in such a way that fine scale structures are successively suppressed, and a scale parameter t is associated with each level in the multi-scale representation.

For a given signal $f \colon R^N \to R$, a linear scale-space representation is a family of derived signals $L \colon R^N \times R \to R$, defined by $L(\cdot;\ 0) = f(\cdot)$ and

$$L(\cdot;\ t) = h(\cdot;\ t) * f(\cdot) \tag{1}$$

for some family $h \colon R^N \times R \to R$ of convolution kernels [1, 2]. An essential requirement on a the scale-space family L is that the representation at a coarse scale constitutes a simplification of the representations at finer scales. Several different ways of formalizing this requirement about non-creation of new structures with increasing scales show that the Gaussian kernel

$$g(x;\ t) = \frac{1}{(2\pi t)^{N/2}} \exp\left(-\frac{x_1^2 + \ldots + x_N^2}{2t} \right) \tag{2}$$

constitutes a canonical choice for generating a scale-space representation [3, 4, 5, 6]. Equivalently, the scale-space family satisfies the diffusion equation

$$\partial_t L = \frac{1}{2}\,\nabla^2 L. \tag{3}$$

The motivation for generating a scale-space representation of a given data set originates from the basic fact into account that real-world objects are composed of different structures at different scales and may appear in different ways depending on the scale of observation. For example, the concept of a "tree" is appropriate at the scale of meters, while concepts such as leaves and molecules are more appropriate at finer scales. For a machine vision system analysing an unknown scene, there is no way to know what scales are appropriate for describing the data. Thus, the only reasonable approach is to consider descriptions at all scales simultaneously [1, 2].

From the scale-space representation, we can at any level of scale define scale-space derivatives by

$$L_{x^\alpha}(x;\ t) = \partial_{x^\alpha}\left(g(x;\ t) * f(x)\right) \tag{4}$$

where $\alpha = (\alpha_1, \ldots, \alpha_D)^T$ and $\partial_{x^\alpha} L = L_{x_1^{\alpha_1} \ldots x_D^{\alpha_D}}$ constitute multi-index notation for the derivative operator ∂_{x^α}. Such Gaussian derivative operators provide a compact way to characterize the local image structure around a certain image

point at any scale. Specifically, the output from scale-space derivatives can be combined into multi-scale differential invariants, to serve as feature detectors.

More generally, a scale-space representation with its Gaussian derivative operators can serve as basis for expressing a large number of early visual operations, including feature detection, stereo matching, computation of motion descriptors and the computation of cues to surface shape [3, 4]. Neurophysiological studies have shown that there are receptive field profiles in the mammalian retina and visual cortex, which can be well modelled by the scale-space framework [7].

Pyramid representation [8] is a predecessor to scale-space representation, constructed by simultaneously smoothing and subsampling a given signal. In this way, computationally highly efficient algorithms can be obtained. A problem noted with pyramid representations, however, is that it is usually algorithmically hard to relate structures at different scales, due to the discrete nature of the scale levels. In a scale-space representation, the existence of a continuous scale parameter makes it conceptually much easier to express this deep structure [2]. For features defined as zero-crossings of differential invariants, the implicit function theorem directly defines trajectories across scales, and at those scales where bifurcations occur, the local behaviour can be modelled by singularity theory [3, 5].

Extensions of linear scale-space theory concern the formulation of non-linear scale-space concepts more committed to specific purposes [9]. There are strong relations between scale-space theory and wavelet theory, although these two notions of multi-scale representation have been developed from slightly different premises.

References:
[1] A. P. Witkin "Scale-space filtering", *Proc. 8th Int. Joint Conf. Art. Intell.*, 1019–1022, Karlsruhe, West Germany, Aug. 1983.

[2] J. J. Koenderink, "The structure of images", *Biological Cybernetics*, 50:363–370, 1984.

[3] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer, 1994.

[4] L. M. J. Florack, *Image Structure*, Kluwer, 1997.

[5] J. Sporring *et al* (eds) *Gaussian Scale-Space Theory*, Kluwer, 1997.

[6] B. M ter Haar Romeny *et al* (eds) *Proc. First Int. Conf. Scale-Space'97*, Springer LNCS, vol. 1252, 1997.

[7] R. A. Young, "The Gaussian derivative model for spatial vision: Retinal mechanisms", *Spatial Vision*, 2:273–293, 1987.

[8] P. J. Burt and E. H. Adelson, "The Laplacian Pyramid as a Compact Image Code", *IEEE Trans. Communications*, 9:4, 532–540, 1983.

[9] B. M ter Haar Romeny (ed) *Geometry-Driven Diffusion in Computer Vision*, Kluwer, 1994.