Real-time scale selection in hybrid multi-scale representations[*]

Tony Lindeberg and Lars Bretzner

Computational Vision and Active Perception Laboratory (CVAP)
Department of Numerical Analysis and Computer Science
KTH (Royal Institute of Technology)
SE-100 44 Stockholm, Sweden

Technical report ISRN KTH/NA/P–03/07–SE, June 2003.

Abstract

Local scale information extracted from visual data in a bottom-up manner constitutes an important cue for a large number of visual tasks. This article presents a framework for how the computation of such scale descriptors can be performed in real time on a standard computer.

The proposed scale selection framework is expressed within a novel type of multi-scale representation, referred to as hybrid multi-scale representation, which aims at integrating and providing variable trade-offs between the relative advantages of pyramids and scale-space representation, in terms of computational efficiency and computational accuracy. Starting from binomial scale-space kernels of different widths, we describe a family pyramid representations, in which the regular pyramid concept and the regular scale-space representation constitute limiting cases. In particular, the steepness of the pyramid as well as the sampling density in the scale direction can be varied.

It is shown how the definition of γ-normalized derivative operators underlying the automatic scale selection mechanism can be transferred from a regular scale-space to a hybrid pyramid, and two alternative definitions are studied in detail, referred to as variance normalization and l_p-normalization. The computational accuracy of these two schemes is evaluated, and it is shown how the choice of sub-sampling rate provides a trade-off between the computational efficiency and the accuracy of the scale descriptors. Experimental evaluations are presented for both synthetic and real data. In a simplified form, this scale selection mechanism has been running for two years, in a real-time vision system.

[*] The support from the Swedish Research Council, the Royal Swedish Academy of Sciences and the Knut and Alice Wallenberg Foundation is gratefully acknowledged. We would also like to thank Pascal Grostabussiat and Jani Niemenmaa for their contributions to earlier stages of this project.

Contents

1 Introduction

Recent works have shown how the notion of automatic scale selection constitutes an essential complement to traditional scale-space representation. While a scale-space representation provides a well-founded framework to represent image structures at different scales, the scale-space representation by itself contains no explicit information about what scales are relevant for further processing. By using automatic scale selecting as a pre-processing stage to early visual operations, hypotheses can be generated about interesting scales and image structures for further analysis. Moreover, visual operations can be normalized with respect to size variations.

For addressing the problem of choosing interesting scale levels from image data, a number of different approaches have been developed in the literature (see the review in section 2). If one aims at real-time performance, however, a common problem of most present approaches for automatic scale selection, is computational efficiency. Since scale selection is performed by either minimizing or maximizing feature measures over scales, the algorithms involve explicit search over scales. The purpose of this article is to show how these problems can be remedied for a class of scale selection methods based on normalized derivatives, and how real-time performance can be obtained on a standard PC by implementation in terms of an oversampled pyramid representation referred to as a hybrid multi-scale representation. It will also be shown how hybrid pyramid representations allows different trade-offs to be reached between computational efficiency, computational accuracy and algorithmic simplicity.

2 Related work

An early approach to scale selection focused on the detection of blob-like image features and scale levels were selected from local maxima over scales of a normalized measure of blob strength (Lindeberg 1993*a*). Later, this idea was generalized to a wide class of differential image features, by selecting scale levels from local maxima over scales of differential invariants expressed in terms of normalized derivatives (Lindeberg 1993*b*, Lindeberg 1994). This principle has been applied to various problems relating to the detection of image features (Lindeberg 1998*b*, Lindeberg 1998*a*, Chomat et al. 2000, Almansa & Lindeberg 2000, Pedersen & Nielsen 2000, Nielsen & Lillholm 2001, Kadir & Brady 2001). In particular, and motivated by the observation that single-scale ridge detection may be highly sensitive to the choice of scale level, special emphasis has been on the detection of ridges for medical image analysis (Pizer et al. 1994, Eberly et al. 1994, Koller et al. 1995, Lorenz et al. 1997, Sato et al. 1998, Staal et al. 1999, Frangi et al. 1999, Majer 2001); for related applications to brain activation analysis, see (Worsley et al. 1996b, Coulon et al. 1997, Lindeberg et al. 1999). Moreover, for the purpose of obtaining zoom invariant image features for further processing, scale selection mechanisms have proven highly useful for interest point detection (Mikolajczyk & Schmid 2001, Mikolajczyk & Schmid 2002) with applications to object recognition (Lowe 1999, Lowe 2000, Hall et al. 2000, Schmid 2001, Sidenbladh & Black 2001) and tracking (Bretzner & Lindeberg 1998, Laptev & Lindeberg 2001, Bretzner et al. 2002). Other approaches for scale selection have been presented from the behaviour of entropy measures over scales (Jägersand 1995, Sporring & Weickert 1999, Hadjidemetriou et al. 2002), the minimization of error measures over scales (Lindeberg 1995, Elder & Zucker 1996, Niessen & Maas 1996, Yacoob & Davis 1997, Lindeberg 1998*c*, Pedersen & Nielsen 2001), probabilistic methods (Marimont & Rubner 1998) or by making explicit use of depth information (Olson 2000). A scale selection mechanism for mean shift analysis has been developed by (Comaniciu et al. 2001).

The algorithms that will presented bear close relations to previous work by (Crowley & Parker 1984) for detecting peaks and ridges in a bandpass pyramid, as well as previous works performing scale selection in a regular scale-space representation (Lindeberg 1994, Lindeberg 1998*b*) without spatial subsampling, although reformulated to be expressed in a hybrid pyramid representation (Lindeberg 1995, Grostabussiat 1997, Niemenmaa 2001). Parallel developments of real-time algorithms for automatic scale selection are being made by (Crowley 2002) and (Lowe 2002).

3 Background: Scale-space and pyramid representations

Since we will build upon pyramid and scale-space representation, we shall first briefly review basic notions concerning these concepts.[1]

3.1 Scale-space representation

Given any continuous D-dimensional signal $f\colon \mathbb{R}^D \to \mathbb{R}$, its scale-space representation $L\colon \mathbb{R}^D \times \mathbb{R}_+ \to \mathbb{R}$ is defined as the result of convolving f with Gaussian kernels

$$g(x;\ t) = \frac{1}{(2\pi t^2)^{D/2}}\, e^{-(x_1^2+\dots+x_D^2)/(2t)} \tag{1}$$

of different widths t. In other words, for $t = 0$ the scale-space representation L is defined by $L(\cdot;\ 0) = f$, and for $t > 0$ by

$$L(x;\ t) = \int_{\xi \in \mathbb{R}^D} g(\xi;\ t)\, f(x - \xi)\, d\xi. \tag{2}$$

Equivalently, L can be defined as the solution to the (linear) diffusion equation

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

with initial condition $L(x;\ 0) = f(x)\ \forall x \in \mathbb{R}^D$. From this representation, scale-space derivatives are defined by

$$L_{x^\alpha}(\cdot;\ t) = L_{x_1^{\alpha_1} x_2^{\alpha_2} \dots x_D^{\alpha_D}}(\cdot;\ t) = \partial_{x^\alpha} L(\cdot;\ t) = g_{x^\alpha}(\cdot;\ t) \tag{4}$$

where the multi-index notation $\alpha = (\alpha_1, \dots, \alpha_D)$ denotes the order of differentiation. Several results have been presented concerning uniqueness properties of this representation as a visual front-end, see (Iijima 1962, Witkin 1983, Koenderink 1984, Babaud et al. 1986, Yuille & Poggio 1986, Koenderink & van Doorn 1992, Lindeberg 1994, ter Haar Romeny 1994, Pauwels et al. 1995, Florack 1997, Sporring et al. 1996, ter Haar Romeny et al. 1997, Weickert 1998, Nielsen et al. 1999, Kerckhove 2001).

[1] For a more extensive background, see e.g. chapters 2–4 in (Lindeberg 1994).

Scale-space for discrete signals. For a discrete signal $f\colon \mathbb{Z}^D \to \mathbb{R}$, the canonical way of defining an analogous scale-space representation $L\colon \mathbb{Z}^D \times \mathbb{R}_+ \to \mathbb{R}$ is by solving a semi-discretized version of the diffusion equation, in which the continuous scale parameter is left untouched and the Laplacian operator is replaced by discrete second-order difference approximation (Lindeberg 1990). In the one-dimensional case, this corresponds to convolution with the discrete analogue of the Gaussian kernel $T\colon \mathbb{Z} \times \mathbb{R}_+ \to \mathbb{R}$, *i.e.*,

$$L(x;\ t) = \sum_{n \in \mathbb{Z}} T(n;\ t)\, f(x-n) \tag{5}$$

where

$$T(n\ \ t) = e^{-t} I_n(t) \tag{6}$$

and I_n are the modified Bessel functions of integer order. In terms of differential equations, this discrete scale-space satisfies the semi-discretized diffusion equation

$$\partial_t L = \frac{1}{2}\, \delta_{xx} L, \tag{7}$$

where δ_{xx} denotes the second-order difference operator with coefficients $(1, -2, 1)$. In two dimensions, the corresponding discrete scale-space is given by the solution to the semi-discrete diffusion equation

$$\partial_t L = \frac{1}{2} \nabla^2_{\lambda} L = \frac{1}{2} \left((1-\lambda) \nabla^2_5 L + \lambda \nabla^2_{\times^2} L \right) \tag{8}$$

where ∇^2_5 and $\nabla^2_{\times^2}$ are five-point and cross-point approximations to the Laplacian operator and $\lambda \in [0, 1]$ is a free parameter. With $\lambda = 0$, this two-dimensional discrete scale-space corresponds to the Cartesian product of the one-dimensional scale-space according to (5) and (7), while $\lambda = \frac{1}{3}$ gives the two-dimensional discrete scale-space with the highest degree of rotational symmetry (Lindeberg 1994).

$$\nabla^2_5 = \begin{pmatrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{pmatrix} \qquad \nabla^2_{\times^2} = \begin{pmatrix} \frac{1}{2} & & \frac{1}{2} \\ & -2 & \\ \frac{1}{2} & & \frac{1}{2} \end{pmatrix} \qquad \nabla^2_{1/3} = \begin{pmatrix} \frac{1}{6} & \frac{4}{6} & \frac{1}{6} \\ \frac{4}{6} & -\frac{20}{6} & \frac{4}{6} \\ \frac{1}{6} & \frac{4}{6} & \frac{1}{6} \end{pmatrix}$$

Figure 1: Computational molecules corresponding to; (a) the five-point operator ∇^2_5, (b) the cross-operator $\nabla^2_{\times^2}$, and (c) the linear combination $\nabla^2_{\lambda} = (1-\lambda)\nabla^2_{\times^2} + \lambda \nabla^2_{\times^2}$ when $\lambda = \frac{1}{3}$.

3.2 Pyramid representation

In a pyramid, the smoothing operation is combined with a subsampling step. For simplicity, let us first assume that the smoothing filter is separable, and that the number of filter coefficients along one dimension is odd. Then, it is sufficient to study the following one-dimensional *reduction operator*, which with $L^{(0)} = f$ states how to compute a coarser-scale representation $L^{(i+1)}$ at level $k+1$ from the representation $L^{(i)}$ at the current scale level k, given a set of filter coefficients $c\colon \mathbb{Z} \to \mathbb{R}$:

$$L^{(i+1)} = \textsc{Reduce}(L^{(i)}) \tag{9}$$

$$L^{(i+1)}(x) = \sum_{n=-N}^{N} c(n)\, L^{(i)}(2x-n). \tag{10}$$

Common conditions on the filter coefficients include:

- positivity: $c(n) \geq 0$,
- unimodality: $c(|n|) \geq c(|n|+1)$,
- symmetry: $c(-n) = c(n)$, and
- normalization: $\sum_{n=-N}^{N} c(n) = 1$.

Another common condition is that all pixels should contribute equally to the next level. With a subsampling factor equal to two, it can be shown that this condition implies that the sum of the filter coefficients with odd indices should be equal to the sum of the filter coefficients with even indices, or equivalently, that the kernel (1/2, 1/2) should occur as one factor in the smoothing kernel. For $N = 1$, the only kernel that satisfies all these conditions is the *binomial three-kernel*.

$$(\frac{1}{4}, \quad \frac{1}{2}, \quad \frac{1}{4}) \tag{11}$$

A negative property of this filter, however, is that when combined repeatedly with a subsampling operation, the equivalent convolution kernel, corresponding to the combined effect of iterated smoothing and subsampling, tends to a triangular function (see the left column in figure 4 for a few illustrations). For $N = 2$, the same conditions imply that the kernel has to be of the form

$$(\frac{1}{4} - \frac{a}{2}, \quad \frac{1}{4}, \quad a, \quad \frac{1}{4}, \quad \frac{1}{4} - \frac{a}{2}). \tag{12}$$

(Burt & Adelson 1983) proposed to determine a such that the equivalent smoothing kernel should be as similar to a Gaussian as possible, and suggested $a \approx 0.4$. For further descriptions about pyramids, see (Burt 1981, Crowley 1981, Burt & Adelson 1983, Rosenfeld 1984, Crowley & Parker 1984, Crowley & Stern 1984, Meer et al. 1987, Chehikian & Crowley 1991, Jähne 1995) and the references therein.

3.3 Connections between scale-space and pyramid representations

There is a close connection between pyramid filters and the diffusion interpretation of scale-space representation. If (7) is discretized in the scale direction using Eulers forward method with scale step Δt, we obtain a smoothing kernel of the form

$$\left(\frac{\Delta t}{2}, \quad 1 - \Delta t, \quad \frac{\Delta t}{2}\right). \tag{13}$$

The limit case for this kernel to be a scale-space kernel, $\Delta t = \frac{1}{2}$, corresponds to the commonly used binomial kernel (11). If iterated twice, we obtain a kernel of the form (12) with $a = 3/8$, the so-called *binomial five-kernel*

$$\left(\frac{1}{4}, \quad \frac{1}{2}, \quad \frac{1}{4}\right)^2 = \left(\frac{1}{16}, \quad \frac{4}{16}, \quad \frac{6}{16}, \quad \frac{4}{16}, \quad \frac{1}{16}\right). \tag{14}$$

Moreover, if we consider the limit case of composing K such kernels in cascade, all having the same scale step $\Delta t = t/K$, and let K tend to infinity, then as limit case we obtain the discrete analogue of the Gaussian kernel (6)

$$\lim_{K \to \infty} \left(\frac{t}{2K}, \quad 1 - \frac{t}{K}, \quad \frac{t}{2K}\right)^K = T(\cdot;\, t). \tag{15}$$

which satisfies similar scale-space axioms for discrete signals as the Gaussian kernel does for continuous signals (Lindeberg 1990, Lindeberg 1994).

3.4 Separable and non-separable smoothing operations

In view of the abovementioned theory, there are two main ways of computing a pyramid representation in two dimensions. If the one-dimensional binomial diffusion filter is applied as a separable filter in two dimensions, we obtain a two-dimensional primitive diffusion smoothing filter with the coefficients

$$\begin{pmatrix} \frac{\Delta t^2}{4} & \frac{\Delta t(1-\Delta t)}{2} & \frac{\Delta t^2}{4} \\ \frac{\Delta t(1-\Delta t)}{2} & (1-\Delta t)^2 & \frac{\Delta t(1-\Delta t)}{2} \\ \frac{\Delta t^2}{4} & \frac{\Delta t(1-\Delta t)}{2} & \frac{\Delta t^2}{4} \end{pmatrix} \tag{16}$$

On the other hand, discretizing the two-dimensional semi-discrete diffusion equation (8) using Eulers method in the scale direction gives a two-dimensional primitive diffusion filter of the form

$$\begin{pmatrix} \frac{\lambda\,\Delta t}{4} & \frac{(1-\lambda)\,\Delta t}{2} & \frac{\lambda\,\Delta t}{4} \\ \frac{(1-\lambda)\,\Delta t}{2} & 1-(2-\lambda)\Delta t & \frac{(1-\lambda)\,\Delta t}{2} \\ \frac{\lambda\,\Delta t}{4} & \frac{(1-\lambda)\,\Delta t}{2} & \frac{\lambda\,\Delta t}{4} \end{pmatrix} \tag{17}$$

(which is separable only if $\Delta t = \lambda$). In situations where a good approximation of rotational symmetry is critical, we can expect kernels generated from (17) with $\lambda = \frac{1}{3}$ to give better performance compared to the grid effects that may be introduced by only using separable smoothing of one-dimensional kernels according to (16). Corresponding arguments can be carried out in higher dimensions.

4 Hybrid pyramid representations

While pyramids and scale-space representation have both been developed from the intuitive idea of representing a given data set at multiple scales in such a way that the resulting representation can be used as input to a large number of visual processes, these concepts have their relative advantages and disadvantages.

A pyramid representation is highly efficient in the sense that it leads to a rapidly decreasing image size, while a scale-space representation successively becomes more redundant as the scale parameter increases. The highly discretized nature of a pyramid can, however, lead to algorithmic problems at coarse scales, while in scale-space representation the task of operating on the data will be successively simplified at coarser scales.

When processing data at a coarse scale in a scale-space representation, it thus seems natural that a certain amount of subsampling can be performed without affecting the performance too seriously. On the other hand, one could also consider decomposing the smoothing operation in a pyramid into a set of smoothing stages, so as to obtain a denser sampling along the scale direction. In this way, we obtain an *oversampled pyramid*, characterized by the fact that not every smoothing step is followed by a subsampling operation.

The goal of this section is to present a general class of multi-scale representations, which comprises both regular pyramids, oversampled pyramids and scale-space representation as special cases.

4.1 Reduction operators: The separable case

To formalize these notions, let us decompose the reduction operator in (9) into a smoothing operation and a subsampling stage. Moreover, let us assume that the smoothing operation can be decomposed into several smoothing steps:

$$\begin{array}{rcl} \textsc{ReduceCycle} & := & \textsc{SubSample} \\ & & \textsc{Smooth}^{+} \end{array} \tag{18}$$

where the notation $\textsc{Op}^{+}$ means that several operators of the form $\textsc{Op}$ may occur. The subsampling operation is here defined by

$$S = \textsc{SubSample}(L) \tag{19}$$

$$S(x) = L(2x). \tag{20}$$

and each smoothing step according to

$$S = \textsc{Smooth}(L) \tag{21}$$

$$S(x) = \sum_{n=-N}^{N} c(n)\, L(x-n). \tag{22}$$

For simplicity of presentation, we shall usually assume that the smoothing operation corresponds to diffusion smoothing repeated K times

$$\textsc{Smooth}(L) = \textsc{DeltaSmooth}(L;\ \Delta t, K) = [\textsc{DeltaSmooth}(L;\ \Delta t, 1)]^{K} \tag{23}$$

where in one dimension the $\textsc{DeltaSmooth}(L;\ \Delta t, 1)$ operator corresponds to convolution with a binomial diffusion filter of the form (13)

$$D = \textsc{DeltaSmooth}(L;\ \Delta t, 1) \tag{24}$$

$$D(x) = \frac{\Delta t}{2}\, L(x-1) + (1-\Delta t)\, L(x) + \frac{\Delta t}{2}\, L(x+1) \tag{25}$$

and in two dimensions we either apply this diffusion filter as a separable filter along each dimension (16) or use a genuine two-dimensional diffusion filter (17). Thus, for one-dimensional filtering, the binomial three-kernel (11) corresponds to

$$\textsc{Bin3Kernel}(L) = \textsc{DeltaSmooth}(L;\ \tfrac{1}{2}, 1) \tag{26}$$

and the binomial five-kernel (14) to

$$\textsc{Bin5Kernel}(L) = \textsc{DeltaSmooth}(L;\ \tfrac{1}{2}, 2) \tag{27}$$

Using this notation, we can now define different types of oversampled pyramid representations as illustrated in figure 2 and figure 3. To index the levels in such a hybrid representations, we shall henceforth use the index $l \in [1 \ldots L]$ for the subsampling levels and the index $j \in [1 \ldots J]$ within each subsampling level (see figure 3).

Bin3ReduceCycle := SubSample
Bin3Kernel

Bin3Reduce6Cycle := SubSample
Bin3Kernel
Bin3Kernel
Bin3Kernel
Bin3Kernel
Bin3Kernel
Bin3Kernel

Bin5ReduceCycle := SubSample
Bin5Kernel

Bin5Reduce3Cycle := SubSample
Bin5Kernel
Bin5Kernel
Bin5Kernel

Figure 2: Examples of regular and oversampled pyramids as generated using the notation for hybrid multi-scale representations defined in (18)–(27).

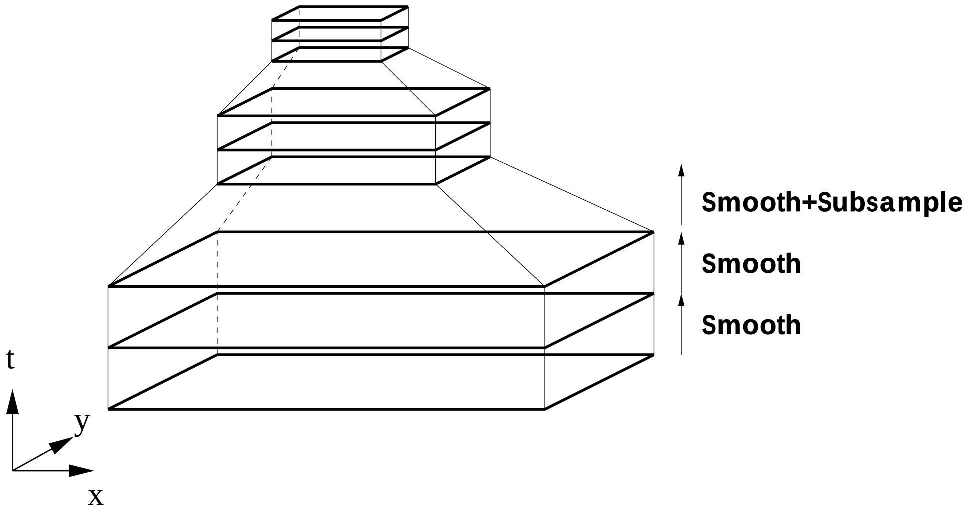


Figure 3: A hybrid pyramid with $I = 3$ levels for each resolution.

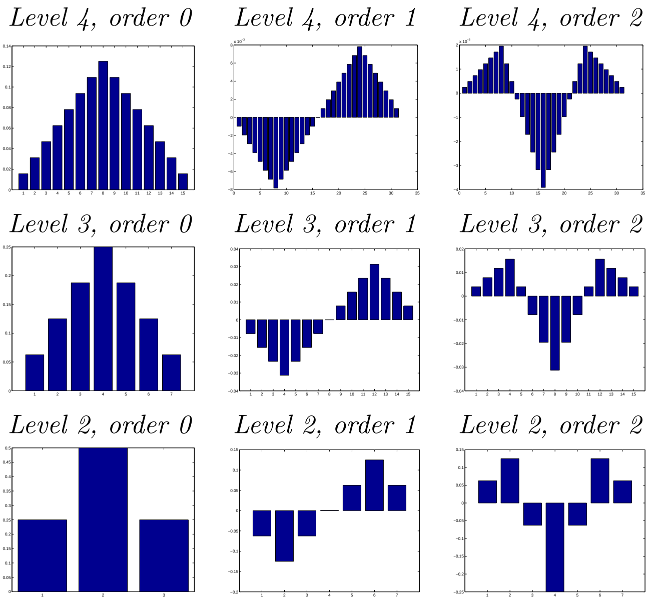


Figure 4: Examples of equivalent convolution kernels and equivalent derivative approximation kernels for the Bin3Pyramid derived from the Bin3ReduceCycle in figure 2. (The values of the scale parameters for these kernels are given in table 1.)

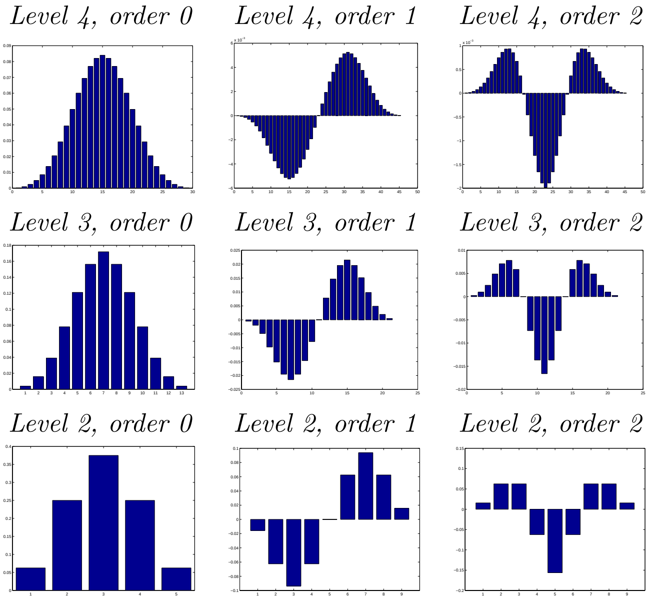


Figure 5: Examples of equivalent convolution kernels and equivalent derivative approximation kernels for the Bin5Pyramid derived from the Bin5ReduceCycle in figure 2. (The values of the scale parameters for these kernels are given in table 1.)

4.2 Equivalent convolution and derivative approximation kernels

Since the representation at each level is constructed from a set of repeated smoothing and subsampling operations, which are all linear operations, the composed operation can equivalently be modeled as the result of applying one kernel $C^{(i,j)}$, termed *equivalent convolution kernel*, to the original image, followed by a pure subsampling step. If we define a dual operator[2] to the ReduceCycle operator according to

$$\begin{array}{rcl} \textsc{ExpandCycle} & := & \textsc{Smooth}^{+} \\ & & \textsc{Enlarge} \end{array}$$

where the Enlarge operation enlarges any D-dimensional image by a factor of 2 along each dimension

$$E = \textsc{Enlarge}(L) \tag{28}$$

$$E(x) = \begin{cases} s^D L(x/s) & \text{if all indices in } x \text{ are even} \\ 0 & \text{if any index in } x \text{ is odd} \end{cases} \tag{29}$$

the equivalent convolution kernel corresponding to level (i, j) can be obtained by expanding a discrete delta function $\delta^{(i,j)}$ at the given pyramid level (i, j) down to the original image

$$C^{(i,j)} = \textsc{ExpandAll}(\delta^{(i,j)}) \tag{30}$$

Thus, ExpandAll denotes the ExpandCycle operators corresponding to the set of all the ReduceCycle operators used for reaching this level. Similarly derivative approximations are computed by taking the grid spacing h at the current into explicit account

$$\partial_{x^r} \approx \mathcal{D}_{x^r} = \frac{1}{h^{|r|}} \delta_{x^r}, \tag{31}$$

at any level with resolution h in the pyramid, the corresponding *equivalent derivative approximation kernel* is given by

$$C^{(i,j)}_{x^r} = \textsc{ExpandAll}(\delta^{(i,j)}_{x^r}) \tag{32}$$

where higher dimensional difference approximations $\delta_{x^r} = \delta_{{x_1}^{r_1}} \delta_{{x_2}^{r_2}} .. \delta_{{x_D}^{r_D}}$ are expressed in terms of the one-dimensional rth order difference operator according to

$$\delta_{x^r} = \begin{cases} (\delta_{xx})^{r/2} & \text{if } r \text{ is even} \\ \delta_x \, \delta_{x^{r-1}} & \text{if } r \text{ is odd} \end{cases} \tag{33}$$

and δ_x and δ_{xx} denote the first-order symmetric difference operators with computational molecules

$$\delta_x = (-\frac{1}{2}, \, 0, \, \frac{1}{2}) \tag{34}$$

and

$$\delta_{xx} = (1, \, -2, \, 1). \tag{35}$$

Figures 4–6 show examples of equivalent derivative approximations of orders $r = 1$ and $r = 2$ computed in this way for the pure and oversampled pyramid representations defined in figure 2.

[2] The interpretation of this operator is that the same weights $c(n)$ are used for propagating grey-level values from a coarser resolution at level i to a finer resolution at level $i - 1$ as were used when constructing the coarser-scale representation $L^{(i)}$ from the the finer-scale representation $L^{(i-1)}$. The factor 2^D is needed to preserve the L_1-norm (the mass) of the signal.

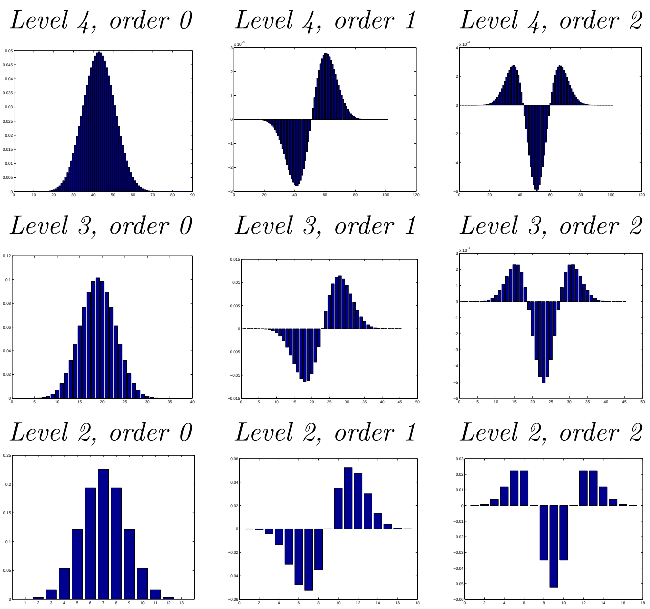


Figure 6: Examples of equivalent convolution kernels and equivalent derivative approximation kernels for the Bin3(6)Pyramid and Bin5(3)Pyramid pyramids derived from the Bin3Reduce6Cycle and Bin5Reduce3Cycle reduction cycles in figure 2. (The values of the scale parameters for these kernels are given in table 1.)

4.3 Measuring the scale parameter

In a multi-scale representation, it is natural to measure the scale parameter in terms of the variance of the convolution kernel used for reaching any scale level. Thus, in an oversampled pyramid representation, it is natural to define the scale parameter in terms of *the covariance matrix of the equivalent convolution kernel*:

$$t_{(i,j)} = (\det V(C^{(i,j)}))^{1/D} = (\det V(\textsc{Expand}(\delta^{(i,j)})))^{1/D} \qquad (36)$$

where $V(C)$ represents the spatial covariance matrix of a kernel C and D is the dimension of the signal.[3] A major motivation underlying this choice is that for non-negative distributions, variances and covariance matrices obey an additive property under convolutions

$$V(C_1 * C_2) = V(C_1) + V(C_2). \qquad (37)$$

This definition is also natural from the viewpoint of the diffusion formulation of the scale-space representation. For pyramid generation kernels of the form $(\frac{\Delta t}{2}, 1 - \Delta t, \frac{\Delta t}{2})$, the scale step Δt exactly corresponds to the variance of the kernel, and if we compose a set of such kernels in cascade, the concatenation of scale steps perfectly corresponds to the addition of variances of the primitive convolution kernels. Moreover, if using such filters for separable filtering along each dimension, if follows that the scale parameters will be added for each dimension. Thus, at coarser levels of resolution with grid spacing $h \in \mathbb{Z}_+$, the operator $\textsc{DeltaSmooth}(L;\ \Delta t, K)$ in (23)

[3] For the isotropic scale-spaces and pyramids we consider in this paper, the covariance matrix will always be proportional to the unit matrix $V(C^{(i,j)}) = t_{(i,j)}\, I$. A more general treatment of non-uniform scale-spaces with non-diagonal covariance matrices is given in (Lindeberg 2001).

corresponds to scale values at levels k and $k+1$ related according to

$$t^{(i,j+1)} - t^{(i,j)} = K\, h^2\, \Delta t. \tag{38}$$

Table 1 shows scale levels computed in this way for the examples of pure and oversampled pyramids defined in figure 7.

Bin3Pyramid
0.0
0.5
2.5
10.5
42.5
170.5
682.5

Bin5Pyramid
0.0
1.0
5.0
21.0
85.0
341.0
1365.0

Bin3(6)Pyramid			
0.0	0.5	...	2.5
3.0	5.0	...	13.0
15.0	23.0	...	55.0
63.0	95.0	...	223.0
255.0	383.0	...	895.0
1023.0	1535.0	...	3583.0

Bin5(3)Pyramid		
0.0	1.0	2.0
3.0	7.0	11.0
15.0	31.0	47.0
63.0	127.0	191.0
255.0	511.0	767.0
1023.0	2047.0	3071.0

Table 1: Scale values for the examples of pure and oversampled pyramids defined in figure 2 (here, without adding an initial pre-smoothing stage as will be described later in figure 7).

4.4 Measuring the subsampling rate

Given that an oversampled pyramid is defined in terms of a set of Smooth and *subsample* operations as exemplified in figure 2, we are interested in describing how the grid spacing h depends on the scale parameter t. Conceptually, it is natural to let the maximum allowed grid spacing h_{max} at any level be proportional to the scale parameter measured in units of the standard deviation $\sigma = \sqrt{t}$ of the equivalent convolution kernel. We may thus define a subsampling factor ρ from the relation

$$h_{max} = \rho\, \sigma = \rho \sqrt{t} \tag{39}$$

and for reasons of computational efficiency define the actual grid spacing as the maximum power of two that does not exceed this upper bound

$$h(t, \rho) = \begin{cases} \max_{h' = 2^{i-1}\,:\, i \in \mathbb{Z}_+ \setminus \{0\}} h' : h' < h_{max}(t, \rho) & \text{if } h_{max} \geq 1 \\ 1 & \text{otherwise} \end{cases} \tag{40}$$

Thus, a subsampling factor of $\rho = 0$ corresponds to preserving the original resolution at all levels of scales, while increasing values of ρ correspond to higher decimation of the number of samples with increasing scale.

In this context, the relation $h < \rho\sqrt{t}$ holds at the first pyramid level if and only of $t_{start}(\rho) \geq \frac{1}{\rho^2}$ If we aim at self-similarity over scales, it is in many situations natural to assume that the input image has been pre-smoothed by this amount. Moreover, physical imaging devices always imply a certain amount of pre-smoothing, given by

$$\begin{array}{rcl}\textsc{Bin3Pyramid} & := & [\textsc{Bin3ReduceCycle}]^{+} \\ & & \textsc{DeltaSmooth}(\cdot;\ \frac{1}{6}, 1)\end{array}$$

$$\begin{array}{rcl}\textsc{Bin5Pyramid} & := & [\textsc{Bin5ReduceCycle}]^{+} \\ & & \textsc{DeltaSmooth}(\cdot;\ \frac{1}{6}, 2)\end{array}$$

$$\begin{array}{rcl}\textsc{Bin3(6)Pyramid} & := & [\textsc{Bin3Reduce6Cycle}]^{+} \\ & & \textsc{DeltaSmooth}(\cdot;\ \frac{1}{2}, 2)\end{array}$$

$$\begin{array}{rcl}\textsc{Bin5(3)Pyramid} & := & [\textsc{Bin5Reduce3Cycle}]^{+} \\ & & \textsc{DeltaSmooth}(\cdot;\ \frac{1}{2}, 2)\end{array}$$

Figure 7: Pure and oversampled pyramids after adding an initial pre-smoothing stage (with pre-smoothing amount t_{start} according to (43)) to the four sample hybrid multi-scale representations defined in figure 2.

the spatial extent of the sensor. Hence, when performing synthetic experiments, or in the absence of knowledge about the physical imaging device for real-world images, we will often add this amount of smoothing prior to the computation of the actual pyramid.

For a pyramid with a reduction cycle of the form (18), it is straightforward to compute the subsampling factor ρ as well as the initial amount of pre-smoothing t_{start} from the assumption that equality in (40) should hold at the first level after each sub-sampling stage. If the total amount of smoothing in the composed Smooth^{+} stage between two sub-sampling stages in (18) corresponds to a variance of $h^2 \Delta t_{cycle}$,, where for hybrid pyramids generated according to (23) and (24) we have

$$h^2 \Delta t_{cycle} = h^2\, J\, K\, \Delta t, \tag{41}$$

then from the resolutions $h = 1$, $h = 2$, $h = 4$ and $h = 8$ etc, we can form the system of equations

$$\begin{cases} 1 = \rho \sqrt{t_{start}} \\ 2 = \rho \sqrt{t_{start} + \Delta t_{cycle}} \\ 4 = \rho \sqrt{t_{start} + \Delta t_{cycle} + 2^2 \Delta t_{cycle}} \\ 8 = \rho \sqrt{t_{start} + \Delta t_{cycle} + 2^2 \Delta t_{cycle} + 4^2 \Delta t_{cycle}} \\ \vdots \end{cases} \tag{42}$$

and solve for ρ and Δt_{cycle}, which gives

$$\rho = \sqrt{\frac{3}{\Delta t_{cycle}}}, \quad t_{start} = \frac{\Delta t_{cycle}}{3} \tag{43}$$

Table 2 shows values of ρ and t_{start} computed in this way for the pyramids in figure 7, as well as a Bin3(12)Pyramid and a Bin5(6)Pyramid object.

For a more general pyramid corresponding to the application of the binomial DeltaSmooth(L; $\Delta t, K$) operator J times at each pyramid level, we have $\Delta t_{cycle} = J\, K\, \Delta t$. Thus, for a general Bin3(M)Pyramid consisting of J_3 layers of binomial three-kernels (11) at each subsampling level (corresponding to $\Delta t = 1/2$ and $K = 1$ in (23) and (24)), we have

$$\rho_{Bin3(M)Pyramid} = \sqrt{\frac{6}{J_3}} \tag{44}$$

while for a general Bin5(M)Pyramid generated from J_5 layers of binomial five-kernels (14) at each subsampling level (corresponding to $\Delta t = 1/2$ and $K = 2$ in (23) and (24)), the corresponding subsampling rate is:

$$\rho_{Bin5(M)Pyramid} = \sqrt{\frac{3}{J_5}} \tag{45}$$

Pyramid	$t^{(i,j+1)} - t^{(i,j)}$	Levels	ρ	t_{start}	d_{mean}
Bin3Pyramid	$h^2/2$	1	$\sqrt{6}$	1/6	2
Bin5Pyramid	h^2	1	$\sqrt{3}$	1/3	2
Bin3(6)Pyramid	$h^2/2$	6	1	1	1/3
Bin5(3)Pyramid	h^2	3	1	1	2/3
Bin3(12)Pyramid	$h^2/2$	12	$1/\sqrt{2}$	2	1/6
Bin5(6)Pyramid	h^2	6	$1/\sqrt{2}$	2	1/3

Table 2: The subsampling rate ρ computed for the four sample pyramids with initial pre-smoothing stages in figure 2 extended with initial pre-smoothing stages according to figure 7.

4.5 Measuring the sampling density in the scale direction

A major aim of a multi-scale representation is to capture the behaviour of image structures over scale. To measure how stable image structures are over scales as well as for quantifying how densely we sample the multi-scale representation in the scale direction, a natural unit to use is effective scale. For the scale-space representation of a continuous signal, it can be shown that effective scale is given by $\tau(t) = A \log t + B$ for some $A \in \mathbb{R}_+$ and $B \in \mathbb{R}$, while for discrete signals a well-founded way of defining this entity can be expressed in terms of the expected number of local extrema as function of scale (Lindeberg 1994).

Thus, for a reduction cycle of the form (18), with the Smooth$^+$ operation corresponding to J steps of the DeltaSmooth(L; $\Delta t, K$) operator at each subsampling level i, we can define the average subsampling density as

$$d_{mean} = \frac{\tau(t^{(i+1,1)}) - \tau(t^{(i,1)})}{J} \tag{46}$$

where we for simplicity of presentation shall approximate $\tau(t)$ as

$$\tau(t) \approx \log_2(t) \tag{47}$$

For pyramids generated according to (40), (42) and (43), the first scale level after any subsampling step will be given by

$$t^{(i,1)} = t_{start} + \Delta t_{cycle} + 2^2 \Delta t_{cycle} + 2^4 \Delta t_{cycle} + \cdots + 2^{2(i-2)} \Delta t_{cycle} = 4^{i-1} \Delta t_{cycle} \tag{48}$$

which means that the explicit value for the subsampling density is

$$d_{mean} = \frac{2}{J} \tag{49}$$

Table 2 lists these values for the four sample pyramids in figure 2 extended with an initial pre-processing stage according to figure 7. Notably, the BIN3(6)PYRAMID and the BIN5(3)PYRAMID pyramid have the same ρ-values, while they differ in d_{mean}.

5 Scale selection in hybrid multi-scale representation

Our next goal is to express a scale selection mechanism within a hybrid pyramid representation. In previous works, it has been shown that a powerful principle for automatic scale selection consists of selecting interesting scale levels from the scales at which (possibly non-linear) combinations of *γ-normalized derivatives*

$$\partial_{\xi_i} = t^{\gamma/2}\,\partial_{x_i}, \tag{50}$$

assume local maxima over scales (see section 2). Intuitively, this corresponds to selecting scale levels at which the normalized feature response is locally strongest.

General scale invariance property. A basic property of this scale selection method is as follows: If $\mathcal{D}(L)$ is a homogeneous differential expression, and if a local maximum of a signal f is detected at scale t_{locmax}, then under a rescaling of f by a factor s, this local maximum over scale is transferred to the scale level $s^2 t_{locmax}$.

Interpretation in terms of L_p-norms. With respect to the computation of derivatives of the scale-space representation, it can be shown that γ-normalization corresponds to normalizing the corresponding γ-normalized Gaussian derivative operators $g_{\xi^m}(\cdot;\ t) = t^{m\gamma/2} g_{x^m}(\cdot;\ t)$ to constant L_p-norms

$$\|g_{\xi^m}(\cdot;\ t)\|_p = \left(\int_{x\in\mathbb{R}^D} |g_{\xi^m}(\cdot;\ t)|^p dx\right)^{1/p} \tag{51}$$

over scales, where the parameter p in the L_p-norm is related to the parameter γ in the γ-normalized derivative concept according to

$$p = \frac{1}{1 + \frac{m}{D}(1-\gamma)}, \tag{52}$$

where m is the order of differentiation and D denotes the dimension of the signal. Specifically, $\gamma = 1$ corresponds to $p = 1$ and thus to L_1-normalization of all the Gaussian derivative kernels. For orders up to four, the kernel norms are in the one-dimensional case with with $p = 1$ given by

$$N_1 = \int_{-\infty}^{\infty} |g_{\xi}(u;\ t)|\, du = \sqrt{\frac{2}{\pi}} \approx 0.797885, \tag{53}$$

$$N_2 = \int_{-\infty}^{\infty} |g_{\xi^2}(u;\ t)|\, du = \sqrt{\frac{8}{\pi\, e}} \approx 0.967883, \tag{54}$$

$$N_3 = \int_{-\infty}^{\infty} |g_{\xi^3}(u;\ t)| \, du = \sqrt{\frac{2}{\pi}} \left(1 + \frac{4}{e^{3/2}}\right) \approx 1.51003, \tag{55}$$

$$N_4 = \int_{-\infty}^{\infty} |g_{\xi^4}(u;\ t)| \, du$$
$$= \frac{4\sqrt{3}}{e^{3/2+\sqrt{3/2}}\,\sqrt{\pi}} (\sqrt{3-\sqrt{6}}\, e^{\sqrt{6}} + \sqrt{3+\sqrt{6}}) \approx 2.8006. \tag{56}$$

while for other (non-integer) p-values, it is straightforward to compute the corresponding integrals numerically. Then, due to separability, the L_p-norm of a partial Gaussian derivative operator in higher dimensions is given by the product of L_p-norms of one-dimensional Gaussian derivative kernel along each dimension; in two dimensions we have

$$\|\partial_{x^m y^n} g(x, y;\ t)\|_p = \|\partial_{x^m} g(x;\ t)\|_p \|\partial_{y^m} g(y;\ t)\|_p \tag{57}$$

where $g(x, y;\ t)$ denotes the two-dimensional Gaussian kernel and $g(x;\ t)$ and $g(y;\ t)$ denote one-dimensional Gaussian kernels along the x- and y-directions.

5.1 Defining normalized derivatives with spatial subsampling

For transferring this notion of γ-normalized derivatives from a scale-space representation to a hybrid pyramid, our next goal is to define normalization parameters α_r such that normalized derivative approximations can be written:

$$\mathcal{D}_{x^r,norm} = \alpha_r \, \mathcal{D}_{x^r}. \tag{58}$$

Here, two approaches will be considered and evaluated:

- *variance-based normalization:* multiplying the equivalent derivative approximation kernel (32) at any level in the pyramid by the variance (36) of the equivalent convolution kernel at the corresponding level

$$\alpha_{r,var} = \left(t^{(i,j)}\right)^{\gamma|r|/2} = \left(\det(V(C^{(i,j)}))^{1/D}\right)^{\gamma|r|/2} \tag{59}$$

- *l_p-normalization:* requiring the l_p-norm of the normalized equivalent derivative approximation kernel to be equal to the L_p-norm of the corresponding Gaussian derivative operator $\partial_{\xi^r} g(x;\ t)$

$$\alpha_{r,l_p} \|C_{x^r}^{(i,j)}\|_p = \|\partial_{\xi^r} g(x;\ t)\|_p \tag{60}$$

Experiments: Scale-space signatures for Gaussian blobs. For a rotationally symmetric Gaussian blob with variance t_0 in two dimensions $f(x, y) = g(x, y;\ t_0)$ it can be shown that the evolution over scales of the γ-normalized Laplacian response at the center of the blob is in the case when $\gamma = 1$ given by

$$(\nabla^2_{norm} L)(0, 0;\ t) = t\,(\partial_{xx} + \partial_{yy}) L(0, 0;\ t) = \frac{t}{\pi(t_0 + t)^2} \tag{61}$$

and there is a unique maximum over scales in $-(\nabla^2_{norm} L)(0, 0;\ t)$ at $t = t_0$.

Figure 8 shows a few examples of such scale-space signatures computed for Gaussian blobs of different sizes, using a separable Bin3(6)Pyramid with an initial presmoothing stage. As can be seen from these graphs, l_p-normalization (stars) gives a closer approximation of the continuous behaviour (the solid curve) than variance-based normalization (crosses). Moreover, for variance-based normalization there are a number of "kinks" in the graph at the scales where subsamplings occur. In these respects, l_p-normalization has clear advantages compared to variance-based normalization.

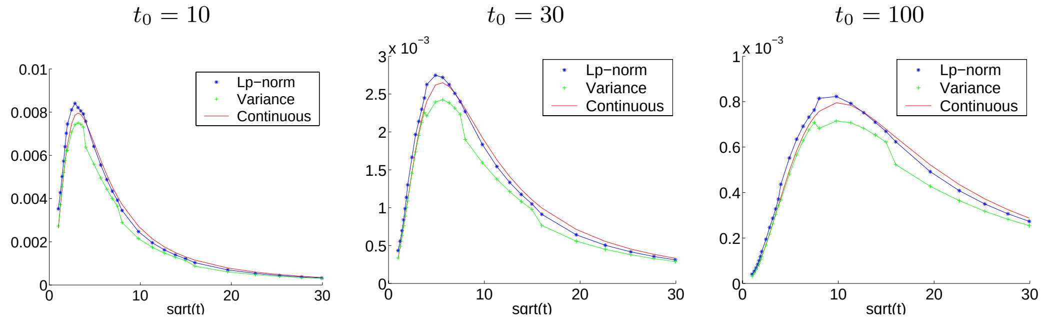

$t_0 = 10$ $t_0 = 30$ $t_0 = 100$

Figure 8: Scale-space signatures of the normalized Laplacian response for rotationally symmetric Gaussian blobs with variances $t_0 = 10$, $t_0 = 30$ and $t_0 = 100$, respectively, computed using a separable Bin3(6)Pyramid in two dimensions using l_p-normalization (stars) and variance-based normalization (crosses). For reference, the corresponding continuous behaviour is shown as well (solid curve).

5.2 Detecting scale-space maxima

A method for complementary scale selection and detection of interest points consists of simultaneously maximizing differential entities over both space and scale. If $\mathcal{D}_{space}L$ denotes the differential entity used for spatial selection and if $\mathcal{D}_{scale,norm}L$ is the γ-normalized differential entity used for scale selection, such *interest points with automatic scale selection* can be characterized by

$$\begin{cases} \nabla(\mathcal{D}_{space}L) = 0 \\ \mathcal{H}(\mathcal{D}_{space}L) \text{ negative definite} \\ \partial_t(\mathcal{D}_{scale,norm}L) = 0 \\ \partial_{tt}(\mathcal{D}_{scale,norm}L) \leq 0 \end{cases} \tag{62}$$

where $\mathcal{H}(\mathcal{D}_{space}L)$ denotes the Hessian of $\mathcal{D}_{space}L$. In the special case when $\mathcal{D}_{space}L = \mathcal{D}_{scale,norm}L$ such points are referred to as *scale-space maxima* of $\mathcal{D}_{scale,norm}L$. Our next goal is to investigate how the performance of a blob detector with automatic scale selection depends on the choice of normalization method as well as the subsampling rate ρ in the pyramid.

To quantify the difference between these two normalization approaches, 1000 Gaussian images were generated containing one blob each with random variance between $t_0 = 10$ and $t_0 = 100$ and at a random position within a central 128×128 window in the image. Local maxima in the normalized Laplacian response were detected as described in appendix A.1, and the scale-space maximum having the strongest response was selected. Then, quadratic interpolation over scales was performed as described in appendix A.2 to estimate the scale $\hat{t}$ of the peak in the scale-space signature as well as its position $(\hat{x}, \hat{y})$ with higher accuracy. The relative error in the estimate was computed

$$\varepsilon_n = \log_2 \left(\frac{\hat{t}_n}{t_{0,n}} \right) \tag{63}$$

and the performance was measured in terms of the following descriptors

$$\varepsilon_{mean} = \frac{1}{N} \sum_{n=1}^{N} \varepsilon_n, \qquad \varepsilon_{spread} = \sqrt{\frac{\sum_{n=1}^{N} \varepsilon_n^2}{N}} \tag{64}$$

where N is the number of blobs. These error measures were then transformed into relative error factors measured in dimension length $\sigma = \sqrt{t}$ according to

$$r_{mean} = \sqrt{2^{\varepsilon_{mean}}}, \qquad r_{spread} = \sqrt{2^{\varepsilon_{spread}}} \tag{65}$$

where the ideal case corresponds to $r_{mean} = 1$ and $r_{spread} = 1$. In addition, the absolute error in the estimated position $(\hat{x}, \hat{y})$ was measured as $\delta = \sqrt{(\hat{x} - x_0)^2 + (\hat{y} - y_0)^2}$ and a relative error measure in relation to the scale level $\sigma_0 = \sqrt{t_0}$ was defined as $\delta_{rel} = \delta/\sigma_0$. This procedure was repeated for different types of separable two-dimensional pyramids as shown in tables 3–4.

Pyramid type	l_p-normalization		variance-based	
	r_{mean}	r_{spread}	r_{mean}	r_{spread}
Bin3Pyramid	0.65	1.61	0.62	1.70
Bin5Pyramid	0.78	1.34	0.77	1.36
Bin3(6)Pyramid	0.93	1.11	0.93	1.15
Bin5(3)Pyramid	0.93	1.12	0.92	1.15
Bin3(12)Pyramid	0.96	1.08	0.95	1.13
Bin5(6)Pyramid	0.94	1.10	0.94	1.13

Table 3: Performance of the scale selection method when performing simultaneous spatial and scale selection based on scale-space maxima of the normalized Laplacian response using different types of hybrid multi-scale representations and either l_p-normalization or variance-based normalization.

As can be seen from the results, there is a substantial variation in the accuracy of the estimate local maximum over scales depending on the type of pyramid — the oversampled Bin3(6)Pyramid and the Bin5(3)Pyramid perform significantly better than the regular Bin3Pyramid and the Bin5Pyramid, and further improvement is obtained if we increase the amount of oversampling by using a Bin3(12)Pyramid or a Bin5(6)Pyramid. In all of these cases, l_p-normalization leads to better performance measures than variance-based normalization. For this reason, we will henceforth prefer l_p-normalization.

Concerning the spatial localization error, we can see how the error decreases as we increase the degree of oversampling in the hybrid pyramid, by decreasing ρ and h_{max}. For the Bin3(6)Pyramid, the Bin5(3)Pyramid, the Bin3(12)Pyramid and the Bin5(6)Pyramid, the average error in all cases corresponds to a fraction of a pixel, and true sub-pixel accuracy is obtained for these synthetic data.

Pyramid type	l_p-normalization		variance-based	
	δ	δ_{rel}	δ	δ_{rel}
Bin3Pyramid	1.86	0.32	1.76	0.29
Bin5Pyramid	1.21	0.21	1.21	0.21
Bin3(6)Pyramid	0.18	0.03	0.05	0.01
Bin5(3)Pyramid	0.19	0.03	0.07	0.01
Bin3(12)Pyramid	0.05	0.01	0.03	0.00
Bin5(6)Pyramid	0.05	0.01	0.02	0.00

Table 4: Measures of the spatial localization error when performing simultaneous spatial and scale selection based on scale-space maxima of the normalized Laplacian response using different types of hybrid multi-scale representations and either l_p-normalization or variance-based normalization.

5.3 Post-processing the scale-space maxima from a hybrid pyramid

While the previous results show that scale-space maxima can be detected in a hybrid pyramid using conceptually very clean operations, there is a minor complication with the previous approach. From the quantitative measure r_{mean} shown in table 3, it can be seen that there is a certain bias in the scale selection procedure that leads to an average underestimate of the scale estimate by 4 to 7 % for the sample types of oversampled hybrid pyramid representations that have been evaluated here.

When analysing the image data in more detail, it can be observed that a major reason for this scale bias is due to the detection of local maxima when translational invariance has been violated by the subsampling step. If the position of the original blob is far away from the closest grid point at the scale levels around the scale level t_0 at which it would be detected without spatial subsampling, the magnitude of the normalized Laplacian at the available grid points at the desired scale level $t_k \approx t_0$ may be significantly smaller than they would have been without spatial subsampling. As a result of this, the values of the normalized Laplacian at lower scale levels may be higher (since the grid sampling there is denser), which in turn means that a lower scale level is selected than in the ideal case without spatial subsampling (see figure 9).

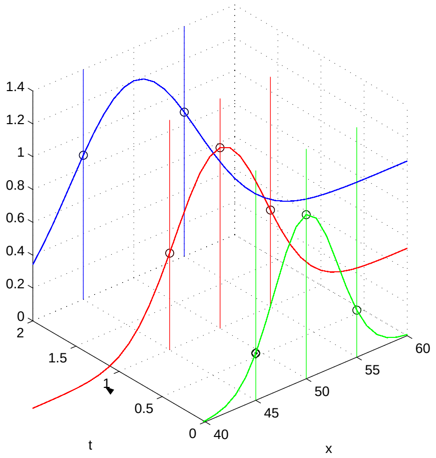


Figure 9: Illustration of how detection of local maxima in combination with sub-sampling may lead to a bias towards smaller scale values.

To reduce this problem, an additional post-processing stage is applied: If a scale-space maximum is detected at a scale level where the next coarser scale level is at lower resolution, then a computation of image values at (one level of) finer resolution is initiated in a spatial 3×3 neighbourhood around the scale space maximum at this pyramid level. If the magnitude of the normalized differential entity is greater at this scale, then the scale-space maximum is translated to this nearest coarser scale level. Moreover, a tri-quadratic interpolation is performed in a $3 \times 3 \times 3$ neighbourhood in space and scale to estimate the position and the scale of the scale-space maximum with sub-pixel accuracy.

Table 5 shows the results obtained by adding these two post-processing stages to the previously methodology. As can be seen from a comparison with table 3, for the Bin5(3)Pyramid and the Bin5(6)Pyramid the average bias in the scale estimate is reduced by basically one order of magnitude, from 6–7 % to 0.4–0.6 %. Moreover, the measure r_{spread} of the spread in the scale values is reduced from 10–12 % to 1–3 %.

Pyramid type	l_p-normalization		variance-normalization	
	r_{mean}	r_{spread}	r_{mean}	r_{spread}
Bin5Pyramid	1.196	1.250	1.182	1.239
Bin5(3)Pyramid	1.006	1.032	0.999	1.180
Bin5(6)Pyramid	0.996	1.019	0.999	1.082

Table 5: Performance of the scale selection method when adding extended coarser scale level search and tri-quadratic interpolation to the previously developed method for performing simultaneous spatial and scale selection based on scale-space maxima of the normalized Laplacian response (see table 3). The numerical values show the mean r_{mean} and the spread r_{spread} of the relative error according to (63) for 1000 Gaussian blobs with random variances between $t_0 = 10$ and $t_0 = 100$.

6 Trade-off: Computational efficiency vs. accuracy

From the experiments on blob detection with automatic scale selection, we have seen how decreasing the value of ρ improves the accuracy of the results. On the other hand, increasing ρ improves the computational efficiency, since fewer grid points are computed. Thus, the hybrid pyramid concept allows us to obtain different trade-offs between computational efficiency vs. accuracy by varying ρ.

To quantify this trade-off, we started out by measuring the computational efficiency in the following way: For a given image size of 384*288 pixels, a threshold on the magnitude of the blob response was determined such that around 500 blobs would be detected between $t_{min} = 4$ and $t_{max} = 2000$ in a Bin5(6)Pyramid. Keeping this threshold fixed, blobs were then detected using the Bin5Pyramid, Bin5(2)Pyramid, ... Bin5(5)Pyramid. A similar experiment was performed using a lower threshold on the blob response, determined in such a way that about 1000 blobs would be obtained in the Bin5(6)Pyramid. Table 6 shows the computation time for detecting scale-space extrema in this way, with and without using the additional localization stage described in section 5.3. To allow for comparison, a denser estimation of the scale and localization errors for Gaussian blob detection was also performed for the same types of pyramids and using the methodology described in section 5.2 — see table 7.

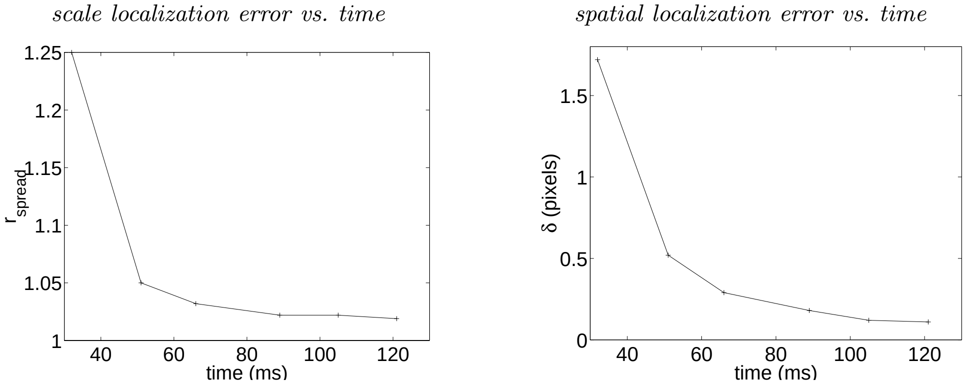
Pyramid type	ρ	500 blobs		1000 blobs	
		det	det+loc	det	det+loc
Bin5Pyramid	1.73	16	32	17	45
Bin5(2)Pyramid	1.22	23	51	25	79
Bin5(3)Pyramid	1.00	39	66	43	97
Bin5(4)Pyramid	0.87	55	89	63	127
Bin5(5)Pyramid	0.77	72	105	81	153
Bin5(6)Pyramid	0.71	88	121	101	173

Table 6: Computation times (in ms) for blob detection in different hybrid pyramids with and without the additional post-processing stage for scale localization. The timings have been performed on a 2.4 GHz DELL PC with a Pentium 4 processor. No hardware specific software libraries have been used and no extensive code optimization has been performed.

Pyramid type	ρ	δ (pixels)	r_{spread}
Bin5Pyramid	1.73	1.72	1.250
Bin5(2)Pyramid	1.22	0.52	1.050
Bin5(3)Pyramid	1.00	0.29	1.032
Bin5(4)Pyramid	0.87	0.18	1.022
Bin5(5)Pyramid	0.77	0.12	1.022
Bin5(6)Pyramid	0.71	0.11	1.019

Table 7: The spatial and scale localization errors for different subsampling factors ρ using l_p-normalization. The experiments were performed on 1000 Gaussian blobs with random position and random variances between 10 and 100.

If we regard these measures as representative indicators of the computational effort and the computational accuracy in the scale estimates, we thus obtain the following trade-off curves for how ρ affects r_{spread} and the computation time:


scale localization error vs. time

spatial localization error vs. time

Figure 10: Trade-offs between the localization error (vertical axis) and the computation time (horizontal axis) for hybrid pyramids with different values of ρ: (left) scale localization error, (right) spatial localization error.

7 Stability of the scale descriptors

In addition to the abovementioned quantitative experiments on synthetic data with ground truth, it is of particular interest to investigate the stability of the scale descriptors on real-world images. To investigate this, we performed the following experiment: An image sequence was taken for a set of uniformly spaced distances to an object. In each image, blob detection was performed by detecting scale-space extrema of the normalized Laplacian response in a Bin5(6)-pyramid using l_p-normalization. Five scale-space maxima were selected manually in the first frame, and these features were matched over time as illustrated in figure 11.

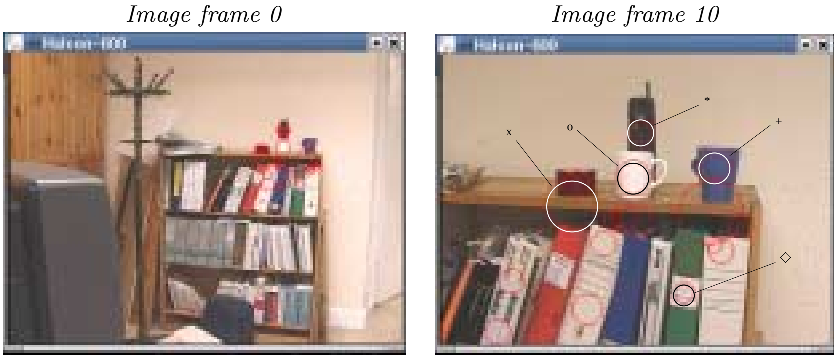


Figure 11: Two out of eleven images in an image sequence used for testing the stability of the scale descriptors over time. In each image, a set of detected image features is indicated, out of which a subset has been matched over time and been used for measuring variations in scale levels over time. In the last image, five scale-space maxima used for scale measurements have been marked by corresponding symbols used in figure 12.

For each one of these five features, a straight line of the form $\frac{1}{\sqrt{t}} = A\tau + B$ was fit to the data (with τ denoting time), and the time to collision was estimated by extrapolating the line to $\tau \to \infty$ (see figure 12). Here, the mean value of the five different estimates of the time to collision was 14.89 time units and the standard deviation 0.30 time units. Considering that these estimates are based on measurements at single points in scale-space, the results show how scale descriptors computed from a hybrid multi-scale representation can be stable enough to be used as a visual cue in its own right.

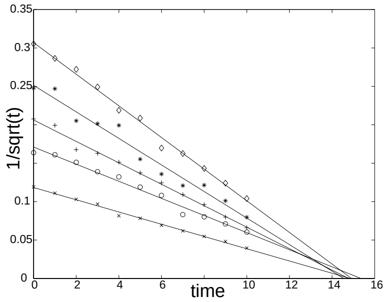


Figure 12: Graph showing the variation over time of $1/\sqrt{t}$ for five image features matched over time as a camera approaches an object with uniform velocity.

8 Summary and discussion

We have presented a general framework for defining subsampled multi-scale representations in such a way that the theory comprises both traditional pyramid representations and discrete scale-space as limiting cases. Regular pyramids arise as a special case when we have only one scale level between any pair of successive subsampling stages (*i.e.* a reduction cycle with $J = 1$), while a regular discrete scale-space representation is obtained as the limiting case if we let the scale increment Δt in the diffusion smoothing operator tend to zero, while keeping the product of $J\Delta t$ constant and equal to the maximum scale level t_{max} that needs to be accessed. Since this family of multi-scale representations provides a way to express different trade-offs between the relative advantages of pyramids and scale-space representation, we refer to it as hybrid multi-scale representations.

Then, we presented a theory for how scale selection mechanisms based on the maximization over scales of γ-normalized derivatives can be expressed within this family of subsampled multi-scale representations. Two ways of defining normalized derivatives in the presence of spatial subsampling have been studied, and it has been shown that the approach referred to as l_p-normalization performs significantly better than the possibly more straightforward approach of variance-based normalization. Specifically, we have quantified how the steepness of a hybrid representation, parameterized by the subsampling rate ρ, allows us to obtain different trade-offs between computational accuracy as enabled by dense sampling and computational efficiency as promoted by sparse sampling. While we have here focused on separable pyramids, the extension to non-separable pyramids is straightforward, and allows for more accurate approximation of rotational symmetry in higher-dimensional scale-spaces.

We have also shown how the scale descriptors computed from a hybrid multi-scale representation are stable enough to be used as a cue in its own right. Combined with a multi-scale tracking and recognition method described elsewhere (Laptev & Lindeberg 2001), an integrated real-time computer vision based on a simplified hybrid pyramid has been presented in (Bretzner et al. 2002).

A Appendix

A.1 Detection of scale-space maxima in hybrid pyramids

For a regular discrete scale-space representation without spatial subsampling, it is straightforward to detect scale-space maxima by mere comparisons with the nearest neighbours. For example, in the two-dimensional case, one compares any point $(x, y;\ t_k)$ with its 26-neighbours $(x+i, y+j;\ t_{k+l}$ for all $x, y, l \in \{-1, 0, +1\}$. The use of different levels of resolution in a subsampled multi-scale representation, however, complicates the situation somewhat, and in this section we will describe the discrete implementation that has been used for all the experiments reported in this article.

The discrete detection of scale-space maxima is based on comparisons with spatial nearest neighbours at:

- the representation at the current scale level $f^{(k)}$
- the representation at the nearest lower scale level $f^{(k-1)}$
- the representation at the nearest higher scale level $f^{(k+1)}$

For simplicity, we describe the implementation of each of these comparisons in the one-dimensional case only. The generalization to higher dimensions is straightforward, by extending the comparison to comparisons over the Cartesian product of the subsets indicated below for each dimension.

Comparisons at the current scale level: In the one-dimensional case, any image point x is compared to its nearest neighbours $x + 1$ and $x - 1$. Specifically, the point x is rejected as a discrete scale-space maximum if either $f^{(k)}(x) < f^{(k)}(x - 1)$ or $f^{(k)}(x) < f^{(k)}(x + 1)$.

Comparisons with the nearest lower level: If the representation at the nearest lower scale level $f^{(k-1)}$ has the same resolution as the current scale level, then comparisons are made using the same subset of image points as for the current level. Thus, the point x is rejected as a scale-space maximum if either $f^{(k)}(x) < f^{(k-1)}(x - 1)$, $f^{(k)}(x) < f^{(k-1)}(x)$ or $f^{(k)}(x) < f^{(k-1)}(x + 1)$.

If the representation at the nearest finer scale has a higher resolution than the current scale level, then the point x at the current level is projected to a new point x_{lower} at the lower level of resolution. (Due to the convention that representations at coarser scales will always be subsets of representations at finer scales, the projection will always be on an actual grid point.) Then, comparisons are made relative to the nearest neighbours of x_{lower}. Thus, the point x is rejected as a scale-space maximum if either $f^{(k)}(x) < f^{(k-1)}(x_{lower} - 1)$, $f^{(k)}(x) < f^{(k-1)}(x_{lower})$ or $f^{(k)}(x) < f^{(k-1)}(x_{lower} + 1)$.

Comparisons with the nearest higher level: If the representation at the nearest higher scale level $f^{(k+1)}$ has the same resolution as the current scale level, then comparisons are made using the same subset of image points as for the current level. Thus, the point x is rejected as a scale-space maximum if either $f^{(k)}(x) < f^{(k+1)}(x - 1)$, $f^{(k)}(x) < f^{(k+1)}(x)$ or $f^{(k)}(x) < f^{(k+1)}(x + 1)$.

If the representation at the nearest finer scale has a higher resolution than the current scale level, then two cases can be distinguished: If we use a subsampling factor of two and project the point x to its nearest higher level, it can either be the case that the projection falls on a grid point or in the middle between two grid points:

- If the projection to the nearest higher level falls on a grid point x_{higher}, then comparisons to the nearest higher level are made only relative to this grid point. Thus, the point x is rejected as a scale-space maximum if $f^{(k)}(x) < f^{(k+1)}(x_{higher})$.

- If the projection to the nearest higher level falls just between two grid point, $x_{higher1}$ and $x_{higher2}$, then the comparisons at the nearest higher level are made relative to both these grid points. Thus, the point x is rejected as a scale-space maximum if either $f^{(k)}(x) < f^{(k+1)}(x_{higher1})$ or $f^{(k)}(x) < f^{(k+1)}(x_{higher2})$.

Accepting scale-space maxima. Finally, the point x is accepted as a scale-space maximum at level k if it is not rejected by any of the abovementioned comparisons.

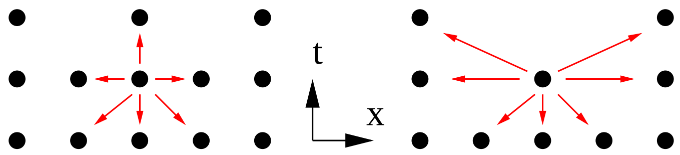


Figure 13: Examples of neighbourhood comparisons situations when detecting scale-space maxima.

A.2 Sub-pixel estimation of local extrema

In several algorithms presented in this paper, the following straightforward procedure is used for estimating the position of a local maximum in a discrete signal by sub-pixel accuracy. Assume that a discrete local maximum has been found at index n in a discrete vector $(f_1, f_2, \dots, f_{n-1}, f_n, f_{n+1}, \dots, f_N)$ in other words $f_n \geq f_{n-1}$ and $f_n \geq f_{n+1}$. Moreover, assume that the discrete vector is associated with attribute values $(t_1, t_2, \dots, t_{n-1}, t_n, t_{n+1}, \dots, t_N)$ where $t_i < t_{i+1}$. Then, we can interpolate the data (t_{n-1}, f_{n-1}), (t_n, f_n) and (t_{n+1}, f_{n+1}) by a quadratic polynomial. Centering these values around (t_n, f_n)

$$\left\{ \begin{array}{l} g_{n-1} = f_{n-1} - f_n \\ u_{n-1} = t_{n-1} - t_n \end{array} \right. \qquad \left\{ \begin{array}{l} g_{n+1} = f_{n+1} - f_n \\ u_{n+1} = t_{n+1} - t_n \end{array} \right. \tag{66}$$

and interpolating the pairs (u_{n-1}, g_{n-1}), $(0, 0)$ and (u_{n+1}, g_{n+1}) by

$$g(u) = A\frac{u^2}{2} + Bu + C \tag{67}$$

results in $C = 0$ and

$$A = \frac{2}{u_{n-1}\, u_{n+1} (u_{n-1} - u_{n+1})} \left(u_{n+1}\, g_{n-1} - u_{n-1}\, g_{n+1}\right) \tag{68}$$

$$B = \frac{1}{u_{n-1}\, u_{n+1} (u_{n-1} - u_{n+1})} \left(u_{n-1}^2\, g_{n+1} - u_{n+1}\, g_{n-1}\right) \tag{69}$$

Thus, we obtain the following estimate of the position

$$t_{max} = t_n - \frac{B}{A} \tag{70}$$

and the value

$$f_{max} = f_n - \frac{B^2}{2A} \tag{71}$$

of the local maximum. Of course, a similar approach applies to a local minimum

References

Almansa, A. & Lindeberg, T. (2000), 'Fingerprint enhancement by shape adaptation of scale-space operators with automatic scale-selection', *IEEE Transactions on Image Processing* **9**(12), 2027–2042.

Babaud, J., Witkin, A. P., Baudin, M. & Duda, R. O. (1986), 'Uniqueness of the Gaussian kernel for scale-space filtering', *IEEE Trans. Pattern Analysis and Machine Intell.* **8**(1), 26–33.

Bretzner, L., Laptev, I. & Lindeberg, T. (2002), Hand-gesture recognition using multi-scale colour features, hierarchical features and particle filtering, *in* 'Proc. Face and Gesture', Washington D.C., USA, pp. 63–74.

Bretzner, L. & Lindeberg, T. (1998), 'Feature tracking with automatic selection of spatial scales', *Computer Vision and Image Understanding* **71**(3), 385–392.

Burt, P. J. (1981), 'Fast filter transforms for image processing', *Computer Vision, Graphics, and Image Processing* **16**, 20–51.

Burt, P. J. & Adelson, E. H. (1983), 'The Laplacian pyramid as a compact image code', *IEEE Trans. Communications* **9:4**, 532–540.

Chehikian, A. & Crowley, J. L. (1991), Fast computation of optimal semi-octave pyramids, *in* 'Proc. 7th Scandinavian Conf. on Image Analysis', Aalborg, Denmark, pp. 18–27.

Chomat, O., de Verdiere, V., Hall, D. & Crowley, J. (2000), Local scale selection for Gaussian based description techniques, *in* 'Proc. Sixth European Conference on Computer Vision', Vol. 1842 of *Lecture Notes in Computer Science*, Springer Verlag, Berlin, Dublin, Ireland, pp. 117–133.

Comaniciu, D., Ramesh, V. & Meer, P. (2001), The variable bandwidth mean shift and data-driven scale selection, *in* 'Proc. 8th Int. Conf. on Computer Vision', Vancouver, Canada, pp. 438–445.

Coulon, O., Bloch, I., Frouin, V. & Mangin, J.-F. (1997), Multi-scale measures in linear scale-space for characterizing cerebral functional activations in 3D PET difference images, *in* B. M. ter Haar Romeny, L. M. J. Florack, J. J. Koenderink & M. A. Viergever, eds, 'Scale-Space Theory in Computer Vision: Proc. First Int. Conf. Scale-Space'97', Springer Verlag, New York, Utrecht, The Netherlands, pp. 188–199.

Crowley, J. L. (1981), A Representation for Visual Information, PhD thesis, Carnegie-Mellon University, Robotics Institute, Pittsburgh, Pennsylvania.

Crowley, J. L. & 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 Intell.* **6**(2), 156–170.

Crowley, J. L. & Stern, R. M. (1984), 'Fast computation of the Difference of Low Pass Transform', *IEEE Trans. Pattern Analysis and Machine Intell.* **6**, 212–222.

Crowley, J. L. (2002), Personal communication.

Eberly, D., Gardner, R., Morse, B., Pizer, S. & Scharlach, C. (1994), 'Ridges for image analysis', *J. of Mathematical Imaging and Vision* **4**(4), 353–373.

Elder, J. H. & Zucker, S. W. (1996), Local scale control for edge detection and blur estimation, *in* 'Proc. 4th European Conf. on Computer Vision', Vol. 1064 of *Lecture Notes in Computer Science*, Springer Verlag, Berlin, Cambridge, UK, pp. 57–69.

Florack, L. M. J. (1997), *Image Structure*, Series in Mathematical Imaging and Vision, Kluwer Academic Publishers, Dordrecht, Netherlands.

Frangi, A. F., Niessen, W. J., Hoogeveen, R. M., van Walsum, T. & Viergever, M. A. (1999), Quantitation of vessel morphology from 3d MRI, *in* 'MICCAI', pp. 358–367.

Grostabussiat, P. (1997), On hybrid multi-scale representations, Licentiate thesis, KTH, Stockholm, Sweden.

Hadjidemetriou, E., Grossberg, M. D. & Nayar, S. K. (2002), Resolution selection using generalized entropies of multiresolution histograms, *in* 'Proc. 7th European Conference on Computer Vision', Vol. 2350 of *Lecture Notes in Computer Science*, Springer Verlag, Berlin, Copenhagen, Denmark, pp. I:220–235.

Hall, D., de Verdiere, V. & Crowley, J. (2000), Object recognition using coloured receptive fields, *in* 'Proc. Sixth European Conference on Computer Vision', Vol. 1842 of *Lecture Notes in Computer Science*, Springer Verlag, Berlin, Dublin, Ireland, pp. 164–177.

Iijima, T. (1962), Observation theory of two-dimensional visual patterns, Technical report, Papers of Technical Group on Automata and Automatic Control, IECE, Japan.

Jägersand, M. (1995), Saliency maps and attention selection in scale and spatial coordinates: An information theoretic approach, *in* 'Proc. 5th Int. Conf. on Computer Vision', Cambridge, MA, pp. 195–202.

Jähne, B. (1995), *Digital Image Processing*, Springer Verlag, New York.

Kadir, T. & Brady, M. (2001), 'Saliency, scale and image description', *Int. J. of Computer Vision* **45**(2), 83–105.

Kerckhove, M., ed. (2001), *Scale-Space and Morphology: Proc. Scale-Space'01*, Lecture Notes in Computer Science, Springer Verlag, New York, Vancouver, Canada.

Koenderink, J. J. (1984), 'The structure of images', *Biological Cybernetics* **50**, 363–370.

Koenderink, J. J. & van Doorn, A. J. (1992), 'Generic neighborhood operators', *IEEE Trans. Pattern Analysis and Machine Intell.* **14**(6), 597–605.

Koller, T. M., Gerig, G., Szèkely, G. & Dettwiler, D. (1995), Multiscale detection of curvilinear structures in 2-D and 3-D image data, *in* 'Proc. 5th Int. Conf. on Computer Vision', Cambridge, MA, pp. 864–869.

Laptev, I. & Lindeberg, T. (2001), Tracking of multi-state hand models using particle filtering and a hierarchy of multi-scale image features, *in* M. Kerckhove, ed., 'Proc. Scale-Space'01', Vol. 2106 of *Lecture Notes in Computer Science*, Springer-Verlag, Vancouver, Canada, pp. 63–74.

Lindeberg, T. (1990), 'Scale-space for discrete signals', *IEEE Trans. Pattern Analysis and Machine Intell.* **12**(3), 234–254.

Lindeberg, T. (1993*a*), 'Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention', *Int. J. of Computer Vision* **11**(3), 283–318.

Lindeberg, T. (1993*b*), On scale selection for differential operators, *in* K. H. K. A. Høgdra, B. Braathen, ed., 'Proc. 8th Scandinavian Conf. on Image Analysis', Norwegian Society for Image Processing and Pattern Recognition, Tromsø, Norway, pp. 857–866.

Lindeberg, T. (1994), *Scale-Space Theory in Computer Vision*, The Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers, Dordrecht, Netherlands.

Lindeberg, T. (1995), Direct estimation of affine deformations of brightness patterns using visual front-end operators with automatic scale selection, *in* 'Proc. 5th Int. Conf. on Computer Vision', Cambridge, MA, pp. 134–141.

Lindeberg, T. (1995), unpublished manuscript on scale selection in hybrid multi-scale representations.

Lindeberg, T. (1998*a*), 'Edge detection and ridge detection with automatic scale selection', *Int. J. of Computer Vision* **30**(2), 117–154.

Lindeberg, T. (1998*b*), 'Feature detection with automatic scale selection', *Int. J. of Computer Vision* **30**(2), 77–116.

Lindeberg, T. (1998*c*), 'A scale selection principle for estimating image deformations', *Image and Vision Computing* **16**(14), 961–977.

Lindeberg, T. (2001), Linear spatio-temporal scale-space, report, ISRN KTH/NA/P--01/22--SE, Dept. of Numerical Analysis and Computing Science, KTH, Stockholm, Sweden.

Lindeberg, T., Lidberg, P. & Roland, P. (1999), 'Analysis of brain activation patterns using a 3-D scale-space primal sketch', *Human Brain Mapping* **7**(3), 166–194.

Lorenz, C., Carlsen, I.-C., Buzug, T. M., Fassnacht, C. & Weese, J. (1997), Multi-scale line segmentation with automatic estimation of width contrast and tangential direction in 2d and 3d medical images, *in* J. Troccaz, E. Grimson & R. Mosges, eds, 'CVRMed-MRCAS'97', Vol. 1205, Grenoble, France, pp. 233–242.

Lowe, D. (1999), Object recognition from local scale-invariant features, *in* 'Proc. 7th Int. Conf. on Computer Vision', Corfu, Greece, pp. 1150–1157.

Lowe, D. G. (2000), Towards a computational model for object recognition in IT cortex, *in* 'Biologically Motivated Computer Vision', Vol. 1811 of *Lecture Notes in Computer Science*, pp. 20–31.

Lowe, D. (2002), Personal communication.

Majer, P. (2001), The influence of the γ-parameter on feature detection with automatic scale selection, *in* M. Kerckhove, ed., 'Proc. Scale-Space'01', Lecture Notes in Computer Science, Springer Verlag, New York, Vancouver, Canada, pp. 245–254.

Marimont, D. & Rubner, Y. (1998), A probabilistic framework for edge detection and scale selection, *in* 'Proc. 6th Int. Conf. on Computer Vision', Bombay, India, pp. 207–214.

Meer, P., Baugher, E. S. & Rosenfeld, A. (1987), 'Frequency domain analysis and synthesis of image pyramid generating kernels', *IEEE Trans. Pattern Analysis and Machine Intell.* **9**, 512–522.

Mikolajczyk, K. & Schmid, C. (2001), Indexing based on scale invariant interest points, *in* 'Proc. 8th Int. Conf. on Computer Vision', Vancouver, Canada, pp. I:525–531.

Mikolajczyk, K. & Schmid, C. (2002), An affine invariant interest point detector, *in* 'Proc. 7th European Conference on Computer Vision', Vol. 2350 of *Lecture Notes in Computer Science*, Springer Verlag, Berlin, Copenhagen, Denmark, pp. I:128–142.

Niemenmaa, J. (2001), Feature detection in images with the pyramid representation, MSc thesis, KTH, Stockholm, Sweden.

Nielsen, M., Johansen, P., Olsen, O. F. & Weickert, J., eds (1999), *Scale-Space Theories in Computer Vision: Proc. Second Int. Conf. Scale-Space'99*, Lecture Notes in Computer Science, Springer Verlag, New York, Corfu, Greece.

Nielsen, M. & Lillholm, M. (2001), What do features tell about images, *in* M. Kerckhove, ed., 'Proc. Scale-Space'01', Lecture Notes in Computer Science, Springer Verlag, New York, Vancouver, Canada, pp. 39–50.

Niessen, W. & Maas, R. (1996), Optic flow and stereo, *in* J. Sporring, M. Nielsen, L. Florack & P. Johansen, eds, 'Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory', Kluwer Academic Publishers, Copenhagen, Denmark.

Olson, C. F. (2000), 'Adaptive-scale filtering and feature detection using range data', *IEEE Trans. Pattern Analysis and Machine Intell.* **22**(9), 983–991.

Pauwels, E. J., Fiddelaers, P., Moons, T. & van Gool, L. J. (1995), 'An extended class of scale-invariant and recursive scale-space filters', *IEEE Trans. Pattern Analysis and Machine Intell.* **17**(7), 691–701.

Pedersen, K. S. & Nielsen, M. (2000), 'The hausdorff dimension and scale-space normalisation of natural images', *J. of Mathematical Imaging and Vision* **11**(2), 266 – 277.

Pedersen, K. S. & Nielsen, M. (2001), Computing optic flow by scale-space integration of normal flow, *in* 'Proc. Scale-Space'01', Vol. 2106 of *Lecture Notes in Computer Science*, Springer Verlag, New York, pp. 14–.

Pizer, S. M., Burbeck, C. A., Coggins, J. M., Fritsch, D. S. & Morse, B. S. (1994), 'Object shape before boundary shape: Scale-space medial axis', *J. of Mathematical Imaging and Vision* **4**, 303–313.

Rosenfeld, A. (1984), *Multiresolution Image Processing and Analysis*, Vol. 12 of *Springer Series in Information Sciences*, Springer Verlag, New York.

Sato, Y., Nakajima, S., Shiraga, N., Atsumi, H., Yoshida, S., Koller, T., Gerig, G. & Kikinis, R. (1998), '3d multi-scale line filter for segmentation and visualization of curvilinear structures in medical images', *Medical Image Analysis* **2**(2), 143–168.

Schmid, C. (2001), Constructing models for content-based image retrieval, *in* 'Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition', Kauai Marriott, Hawaii.

Sidenbladh, H. & Black, M. J. (2001), Learning the statistics of people in images and video, *in* 'Proc. International Conference on Computer Vision', Vol. 2, Vancouver, Canada, pp. 709–716.

Sporring, J., Nielsen, M., Florack, L. & Johansen, P., eds (1996), *Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory*, Series in Mathematical Imaging and Vision, Kluwer Academic Publishers, Copenhagen, Denmark.

Sporring, J. & Weickert, J. A. (1999), 'Information measures in scale-spaces', *IEEE Trans. Information Theory* **45**(3), 1051–1058.

Staal, J., Kalitzin, S., ter Haar Romeny, B. & Viergever, M. (1999), Detection of critical structures in scale-space, *in* '2nd International Conference on Scale-Space Theories in Computer Vision', Vol. 1682 of *Lecture Notes in Computer Science*, Springer Verlag, New York, Corfu, Greece, pp. 105–116.

ter Haar Romeny, B., ed. (1994), *Geometry-Driven Diffusion in Computer Vision*, Series in Mathematical Imaging and Vision, Kluwer Academic Publishers, Dordrecht, Netherlands.

ter Haar Romeny, B., Florack, L., Koenderink, J. J. & Viergever, M., eds (1997), *Scale-Space Theory in Computer Vision: Proc. First Int. Conf. Scale-Space'97*, Lecture Notes in Computer Science, Springer Verlag, New York, Utrecht, Netherlands.

Weickert, J. (1998), *Anisotropic Diffusion in Image Processing*, Teubner-Verlag, Stuttgart, Germany.

Witkin, A. P. (1983), Scale-space filtering, *in* 'Proc. 8th Int. Joint Conf. Art. Intell.', Karlsruhe, Germany, pp. 1019–1022.

Worsley, K. J., Marret, S., Neelin, P. & Evans, A. C. (1996b), 'Searching scale space for activation in PET images', *Human Brain Mapping* **4**, 74–90.

Yacoob, Y. & Davis, L. S. (1997), Estimating image motion using temporal multi-scale models of flow and acceleration, *in* M. Shah & R. Jain, eds, 'Motion-Based Recognition', Kluwer Academic Publishers.

Yuille, A. L. & Poggio, T. A. (1986), 'Scaling theorems for zero-crossings', *IEEE Trans. Pattern Analysis and Machine Intell.* **8**, 15–25.