

In: K. Ikeuchi (ed.) *Computer Vision: A Reference Guide*, Springer, pages 701–713, 2014, DOI: 10.1007/978-0-387-31439-6.

Scale Selection

TONY LINDBERG, SCHOOL OF COMPUTER SCIENCE AND COMMUNICATION,
KTH ROYAL INSTITUTE OF TECHNOLOGY, STOCKHOLM, SWEDEN

Synonyms

- Automatic scale selection
- Scale invariant image features and image descriptors

Related Concepts

- Scale-space
- Feature detection
- Scale invariance
- Interest point detection
- Blob detection
- Corner detection
- Edge detection
- Ridge detection
- Frequency estimation
- Feature tracking
- Image-based matching and recognition
- Object recognition

Definition

The notion of scale selection refers to methods for estimating characteristic scales in image data and for automatically determining locally appropriate scales in a scale-space representation, so as to adapt subsequent processing to the local image structure and compute scale invariant image features and image descriptors.

An essential aspect of the approach is that it allows for a bottom-up determination of inherent scales of features and objects without first recognizing them or delimiting alternatively segmenting them from their surrounding.

Scale selection methods have also been developed from other viewpoints of performing noise suppression and exploring top-down information.

Background

The concept of *scale* is essential when computing features and descriptors from image data. Real-world objects may contain different types of structures at different scales and may therefore appear in different ways depending on the scale

of observation. When observing objects by a camera or an eye, there is an additional scale problem due to perspective effects, implying that distant objects will appear smaller than nearby objects. A vision system intended to operate autonomously on image data acquired from a complex environment must therefore be able to handle and be robust to such scale variations.

For a vision system that observes an unknown scene, there is usually no way to *a priori* know what scales are appropriate for extracting the relevant information. Hence, a multi-scale representation of the image data is essential, whereby the original signal is embedded into a *one-parameter family* of signals using scale as the parameter. Given an N -dimensional signal $f: \mathbb{R}^N \rightarrow \mathbb{R}$ and with the notation $x = (x_1, \dots, x_N) \in \mathbb{R}^N$, the *scale-space representation* [28,8,11] of f is defined by the convolution operation

$$L(x; t) = \int_{\xi \in \mathbb{R}^N} f(x - \xi) g(\xi; t) d\xi \quad (1)$$

where $g: \mathbb{R}^N \times \mathbb{R}_+ \rightarrow \mathbb{R}$ denotes the *Gaussian kernel*

$$g(x; t) = \frac{1}{(2\pi t)^{N/2}} e^{-|x|^2/2t} \quad (2)$$

and the variance $t = \sigma^2$ of this kernel is referred to as the *scale parameter*. Based on this representation, *Gaussian derivatives*, or *scale-space derivatives*, at any scale t can then be computed by differentiating the scale-space representation or equivalently by convolving the original image with Gaussian derivative kernels:

$$L_{x^\alpha}(\cdot; t) = \partial_{x^\alpha} L(\cdot; t) = (\partial_{x^\alpha} g(\cdot; t)) * f(\cdot). \quad (3)$$

(with multi-index notation $\alpha = (\alpha_1, \dots, \alpha_N)$ for $\partial_{x^\alpha} = \partial_{x_1}^{\alpha_1} \dots \partial_{x_N}^{\alpha_N}$). Such Gaussian derivatives can be used as a basis for expressing a large number of visual modules including feature detection, feature classification, image matching, motion, shape cues, and image-based recognition [15].

Theory

The notion of scale selection complements traditional *scale-space theory* by providing explicit mechanisms for generating hypotheses about interesting scales.

Scale selection from γ -normalized derivatives: A particularly useful methodology for computing estimates of *characteristic scales* is by detecting *local extrema over scales* of differential expressions in terms of *γ -normalized derivatives* [13,12] defined by

$$\partial_\xi = t^{\gamma/2} \partial_x. \quad (4)$$

A general and very useful property of this construction is that if two signals f and f' are related by a scaling transformation

$$f'(x') = f(x) \quad \text{with} \quad x' = s x, \quad (5)$$

and if there is a local extremum over scales at $(x_0; t_0)$ in a differential expression $\mathcal{D}_{\gamma-norm}L$ defined as a homogeneous polynomial of Gaussian derivatives computed from the scale-space representation L of the original signal f , then there will be a corresponding local extremum over scales at $(x'_0; t'_0) = (s x_0; s^2 t_0)$ in the corresponding differential expression $\mathcal{D}_{\gamma-norm}L'$ computed from the scale-space representation L' of the rescaled signal f' [13, section 4.1].

This scaling result holds for all homogeneous polynomial differential expressions, including rotationally invariant differential invariants, and implies that local extrema over scales of γ -normalized derivatives are preserved under scaling transformations. Thereby, such local extrema over scales provide a theoretically well-founded way to automatically adapt the scale levels to local scale variations.

Specifically, scale-normalized scale-space derivatives of order $|\alpha| = \alpha_1 + \dots + \alpha_N$ at corresponding points will then be related according to

$$L'_{\xi'\alpha}(x'; t') = s^{|\alpha|(\gamma-1)} L_{\xi\alpha}(x; t) \quad (6)$$

which means that $\gamma = 1$ implies perfect scale-invariance in the sense that the γ -normalized derivatives at corresponding points will be equal. If $\gamma \neq 1$, the difference in magnitude can on the other hand be easily compensated for using the scale values of corresponding scale-adaptive image features.

These results imply that detection of image features and computation of image descriptors at scale levels equal to or proportional to the scales at which there are local extrema over scales constitutes a very general methodology for obtaining *scale-invariant image features* and *scale-invariant image descriptors*.

Indeed, it can also be axiomatically shown that the notion of γ -normalized derivatives arises by necessity, given the condition that local extrema over scales of scale-normalized derivatives should be preserved under scaling transformations [13, appendix A.1].

Relation to frequency estimation: There is a conceptual similarity between this principle and local frequency estimation from peaks in the Fourier transform. For a one-dimensional sine wave

$$f(x) = \sin(\omega x) \quad (7)$$

it can be shown [13, section 3] that there will be a peak in the magnitude of the m :th order γ -normalized derivative at a scale

$$\sigma_{max} = \frac{\sqrt{\gamma m}}{2\pi} \lambda \quad (8)$$

proportional to the wavelength $\lambda = 2\pi/|\omega|$ of the signal. Two conceptual differences compared to Fourier-based frequency estimation, however, are that (i) no window size is needed for computing the Fourier transform, and (ii) this approach applies also to non-linear differential expressions.

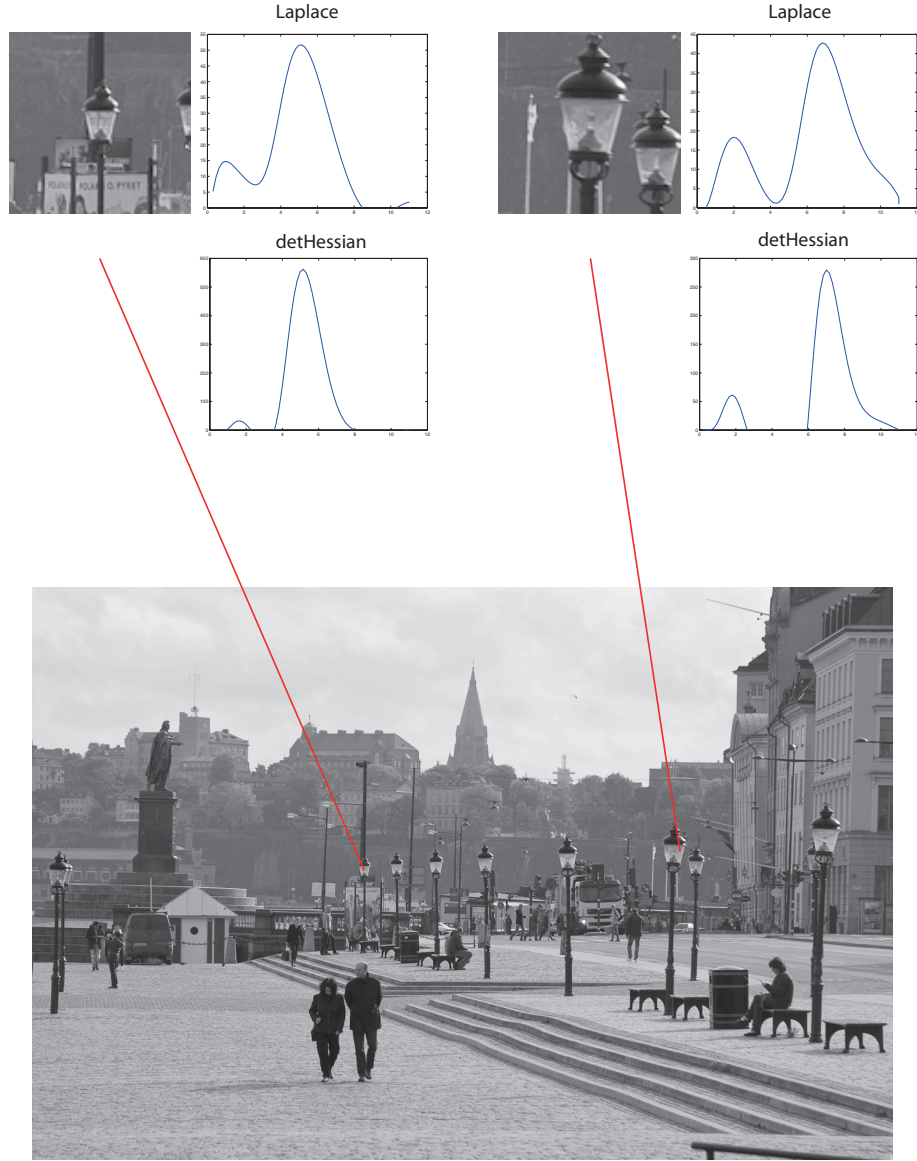


Fig. 1. *Scale-space signatures* accumulated for image structures of different size in the image domain. The upper part of the illustration shows windows around two details from the image at the bottom, with corresponding scale-space signatures of the scale-normalized Laplacian $\nabla_{norm}^2 L$ and the scale-normalized determinant of the Hessian $\det \mathcal{H}_{norm} L$ accumulated at the central point in each window. As can be seen from the graphs, the local extrema over scales are assumed at coarser scales for the larger size object than for the smaller size object. Specifically, the ratio between the scale values at which the local extrema are assumed provides an estimate of the relative amount of scaling, when measured in dimension [length]. (In the graphs, the horizontal axis represents *effective scale* [11, pages 180–182] approximated by $\tau \approx \log_2(1+t)$.) (Reprinted from [18] with permission.)

Relations to image statistics: It can be shown [13, section 9.1] that the notion of γ -normalized derivatives corresponds to normalizing the m :th order N -dimensional Gaussian derivatives to constant L_p -norms over scale with

$$p = \frac{1}{1 + \frac{m}{N}(1 - \gamma)} \quad (9)$$

where the perfectly scale invariant case $\gamma = 1$ corresponds to L_1 -normalization for all orders m .

It can also be shown [13, section 9.2] that the γ -normalized derivatives are *neutral* with respect to *self-similar* power spectra of the form

$$S_f(\omega) = |\omega|^{-N-2m(1-\gamma)}. \quad (10)$$

Natural images often show a qualitative behaviour similar to this [5].

Scale-space signatures: Figure 1 illustrates the basic idea, by showing so-called *scale-space signatures* accumulated in the two-dimensional case¹ for two generally applicable differential entities for scale selection; the *scale-normalized Laplacian* [11, section 13.3] [13, section 5] (with $\gamma = 1$)

$$\nabla^2 L_{norm} = t(L_{xx} + L_{yy}) \quad (11)$$

and the *scale-normalized determinant of the Hessian* [11,13] (also with $\gamma = 1$)

$$\det \mathcal{H}_{norm} L = t^2 (L_{xx}L_{yy} - L_{xy}^2). \quad (12)$$

In the scene in figure 1, there are strong perspective scaling effects due to differences in depth between similar objects in the world. These scale variations are reflected in the scale-space signatures in the respect that the local extrema over scales are assumed at finer scales for distant objects and at coarser scales for nearby objects. If one computes the ratio between the scale values in terms of a scale parameter $\sigma = \sqrt{t}$ of dimension [length], then the ratio between the scale values is in very good agreement with the ratio between the sizes of the objects in the image domain as measured by a ruler. This property illustrates one of the scale-invariant properties of the scale selection mechanism.

General framework for defining scale-invariant image descriptors. By computing an image descriptor at a scale proportional to the detection scale \hat{t} of a scale-invariant image feature or by normalizing an image patch by a corresponding scaling factor $\hat{\sigma} = \sqrt{\hat{t}}$, provides two very general scale normalization mechanisms that can be used for defining much wider classes of *scale-invariant image descriptors* [13,16,17,19] (see the “Applications” section below for two specific examples regarding image-based recognition). The scale-invariant properties of

¹ In the specific 2-D case, the simplifying notation $(x, y) \in \mathbb{R}^2$ is used instead of $x = (x_1, x_2) \in \mathbb{R}_2$ implying that $L_{x_1 x_1} = L_{xx}$, $L_{x_1 x_2} = L_{xy}$, $L_{x_2 x_2} = L_{yy}$, etc.

these descriptors originate from the general scale-invariant property of local extrema over scales of differential expressions in terms of γ -normalized derivatives.

Figure 2 illustrates how scale normalization can be performed in this way by rescaling the local image patches around the two details in figure 1 using the scale values $\hat{\sigma} = \sqrt{\hat{t}}$ at which the Laplacian $\nabla_{norm}^2 L$ and the determinant of the Hessian, respectively, assumed their strongest local extrema over scales. In this sense, scale normalization from the detection scales \hat{t} constitutes a general mechanism for establishing a *common scale-invariant reference frame* with regard to scaling transformations.

It should be noted, however, that multiple extrema over scales may in general be found in the scale-space signature, as can be seen in figure 1 and figure 2, where two significant local extrema over scales are obtained in each scale-space signature, with the coarser-scale response corresponding to the lamp as a whole and the finer-scale response corresponding to the light bulb inside. Because of this inherent multi-scale nature of real-world objects, a vision system intended to interpret images from a natural environment must in general be able to handle *multiple scale hypotheses over scales*.

Scale-space extrema. The notion of scale selection from scale-normalized derivatives can be complemented by *spatial selection* by detecting points in scale-space that assume *local extrema with respect to both space x and scale t* . Such points are referred to as *scale-space extrema*. Specifically, detection of scale-space extrema of rotationally invariant differential operators provides a general, effective and robust methodology for detecting *interest points with built-in scale selection* [11,13,16,17,19]. Thus, given a scale-normalized differential expression $\mathcal{D}_{\gamma-norm} L$, one simultaneously obtains spatial positions \hat{x} and scale estimates \hat{t} according to

$$(\hat{x}; \hat{t}) = \underset{(x; t)}{\operatorname{argmaxminlocal}} (\mathcal{D}_{\gamma-norm} L)(x; t) \quad (13)$$

Figure 3 shows the result of detecting the 50 strongest scale-space extrema of the scale-normalized Laplacian $\nabla_{norm}^2 L$ and the scale-normalized determinant of the Hessian $\det \mathcal{H}_{norm} L$ from an image that contains two objects of different size. Each scale-space extremum has been illustrated by a circle with the radius proportional to the detection scale $\hat{\sigma} = \sqrt{\hat{t}}$. In figure 4, each feature has been visualized by a sphere in the 3-D scale-space volume of the 2-D image, with the radius of the sphere increasing with the detection scale. As can be seen from this illustration, the notion of scale-space extrema can effectively reveal interest points and characteristic scales of those (see the “Applications” section below for more details about scale-invariant interest points detectors). Specifically, the differences in the radii of the circles in the 2-D illustration and in the heights over the image plane in the 3-D graphics reveal the scale differences between corresponding image features from the two objects.

The differential operators $\nabla_{norm}^2 L$ and $\det \mathcal{H}_{norm} L$ do in general both produce strong responses at the centers of blob-like structures that are either brighter or darker than their surrounding, provided that the these differential entities are

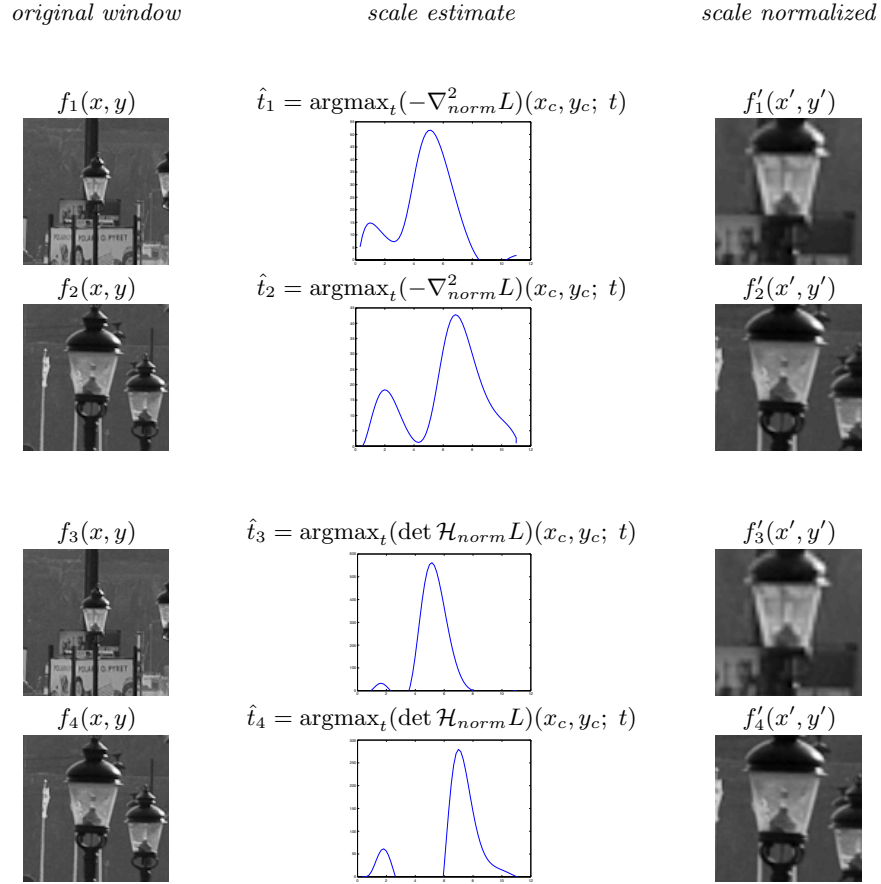


Fig. 2. In this illustration, local windows around two of the lamps in figure 1 (shown in the left column) have been rescaled by scaling factors $\hat{\sigma} = \sqrt{\hat{t}}$ obtained from the dominant response over scales of the Laplacian $\nabla_{norm}^2 L$ and the determinant of the Hessian $\det \mathcal{H}_{norm} L$ (shown in the middle column) to compute a *scale-normalized window* (shown in the right column) around each detail. In this way, scale selection can be used for defining a *scale-normalized reference frame* for subsequent computation of *scale-invariant image descriptors* (Reprinted from [18] with permission.).

computed at scale levels that roughly match the size of the corresponding image structures. For this reason, they constitute very useful differential entities for *blob detection*.

Discrete implementation: Detection of scale-space extrema from an N -dimensional discrete image can be performed by nearest-neighbour comparisons in the $N + 1$ -dimensional scale-space volume. For a 2-D image, this implies that nearest-

neighbour comparisons are performed by local comparisons with the 26 neighbours in a $3 \times 3 \times 3$ neighbourhood over space and scale [11] [13, footnote 16] [22]. Scale estimates and position estimates of higher accuracy can then be obtained by fitting a parabola to the data around any scale-space extremum [20,22].

Discrete analogues of γ -normalized derivatives can be obtained either by (i) *variance-based normalization* which implies that the discrete derivative approximations $\delta_{x^\alpha} L$ are multiplied by an appropriate power of the scale parameter

$$L_{\xi^\alpha}(\cdot; t) = L_{\xi_1^{\alpha_1} \dots \xi_N^{\alpha_N}}(\cdot; t) = (t^{\gamma/2})^{|\alpha|} (\delta_{x^\alpha} L)(\cdot; t) \quad (14)$$

or by using the notion of (ii) *l_p -normalization* [13, appendix A.4.2]

$$L_{\xi^\alpha}(\cdot; t) = L_{\xi_1^{\alpha_1} \dots \xi_N^{\alpha_N}}(\cdot; t) = C_\alpha (\delta_{x^\alpha} L)(\cdot; t) \quad (15)$$

where the discrete normalization constants C_α are determined such that the l_p -norms of the scale-normalized discrete derivative approximation kernels $\delta_{x^\alpha} T$ [11, chapter 5] are to be equal to the L_p -norms of the corresponding γ -normalized Gaussian derivative kernels $\partial_{x^\alpha} g$

$$C_\alpha \left(\sum_{n \in \mathbb{Z}^N} |(\delta_{x^\alpha} T)(n; t)|^p \right)^{1/p} = (t^{\gamma/2})^{|\alpha|} \left(\int_{x \in \mathbb{R}^N} |(\partial_{x^\alpha} g)(x; t)|^p dx \right)^{1/p}. \quad (16)$$

Experiments in [20] show that the notion of l_p -normalization gives more accurate scale estimates in situations where discretization effects become important.

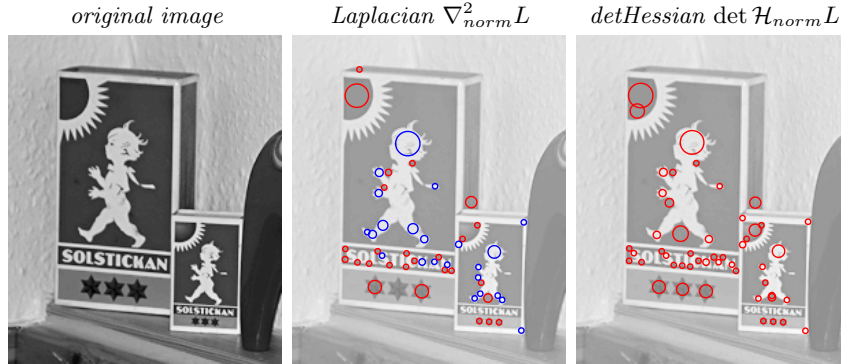
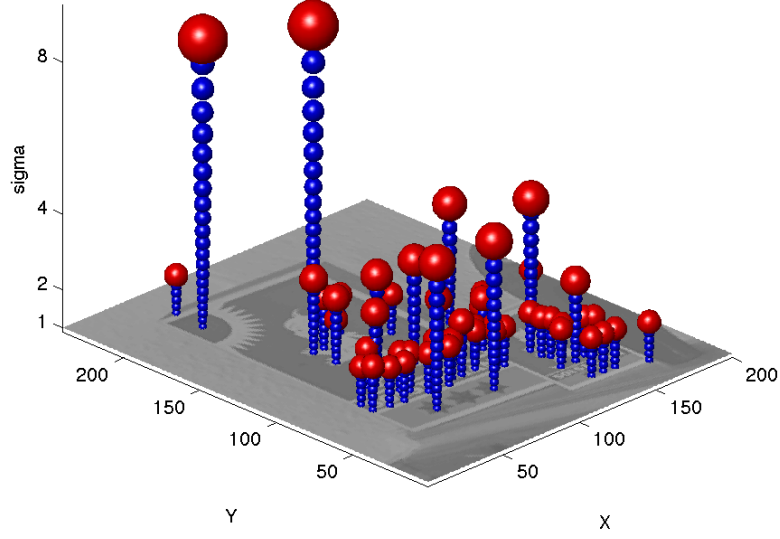


Fig. 3. 2-D illustration of the 50 strongest *scale-space extrema* of the Laplacian $\nabla_{norm}^2 L$ and the determinant of the Hessian $\det \mathcal{H}_{norm} L$ computed from an image with two similar objects of different physical size. Each feature is illustrated by a circle centered at the position (\hat{x}, \hat{y}) of the scale-space extremum and with the radius proportional to the detection scale $\hat{\sigma} = \sqrt{\hat{t}}$. Red circles represent scale-space maxima, while blue circles represent scale-space minima.

scale-space extrema of the Laplacian $\nabla_{norm}^2 L$



scale-space extrema of the determinant of the Hessian $\det \mathcal{H}_{norm} L$

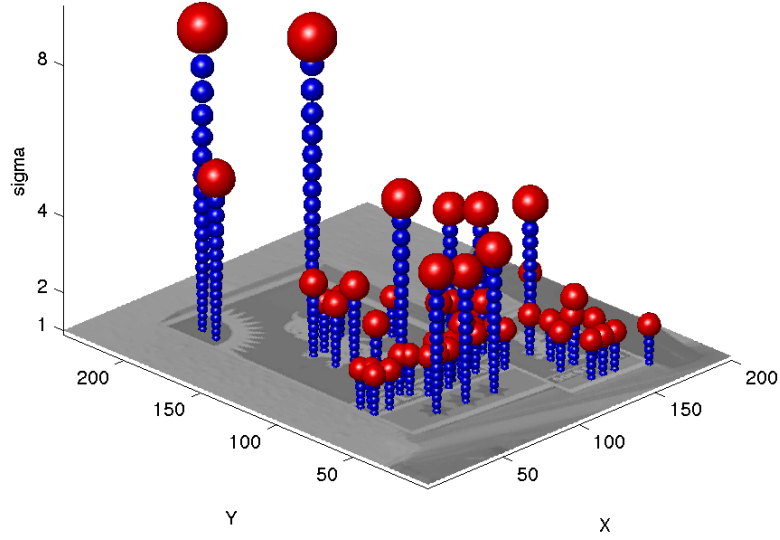


Fig. 4. 3-D illustration of the 50 strongest *scale-space extrema* of the Laplacian $\nabla_{norm}^2 L$ and the determinant of the Hessian $\det \mathcal{H}_{norm} L$ computed from the image in figure 4. Here, each feature is illustrated by a red sphere centered at the position $(\hat{x}_0, \hat{y}_0; \hat{t}_0)$ of the scale-space extremum and with the radius increasing with the detection scale \hat{t}_0 . The blue spheres have been inserted to simplify visual interpretation.

A particularly convenient way of implementing scale-space smoothing in this context is by convolution with the *discrete analogue of the Gaussian kernel* [11, pages 84–87]

$$T(n; t) = e^{-t} I_n(t) \quad (17)$$

which implies that semi-group property of the Gaussian scale-space holds exactly also for the discrete scale-space kernels $T(\cdot; t_1) * T(\cdot; t_2) = T(\cdot; t_1 + t_2)$ and the *cascade smoothing property*

$$L(\cdot; t_2) = T(\cdot; t_2 - t_1) * L(\cdot; t_1) \quad (18)$$

for $t_2 \geq t_1 \geq 0$ implies that one can perform a set of incremental convolutions with kernels of smaller support instead of computing each scale level from the original signal f independently.

The notion of scale selection from scale-normalized derivatives can also be transferred to a *pyramid* representation to allow for real-time implementation on standard processors [20,3,22].

Alternative approaches to scale selection: A number of other mechanisms for scale selection have also been developed based on ideas of:

- detecting peaks over scales in weighted entropy measures [7] or Lyapunov functionals [27],
- minimizing normalized error measures over scales in order to compute more accurate localization estimates for coarser-scale corner features [13, section 7.2] or for coarse-to-fine matching of highly noisy image data [14],
- determining minimum reliable scales for feature detection according to an *a priori* determined noise suppression model [4],
- determining optimal stopping times in non-linear diffusion-based image restoration methods using similarity measurements relative to the original data [24],
- performing image segmentation from the scales at which a supervised classifier delivers class labels with the highest posterior [21,10].

Relations between the different approaches to scale selection: The different approaches to scale selection may have quite different properties depending on the types of data they are applied to. For noise free data, an adaptive noise suppression scheme optimized for suppressing high-frequency noise can be expected to not smooth the data at all, thus implying the selection of a zero scale, whereas scale selection based on local extrema over scales will always select a scale level reflecting a characteristic length in the image data.

Provided that the characteristic lengths of the relevant image features are greater than the typical characteristic lengths in the noise, scale selection based on scale-normalized derivatives will therefore lead to scale-invariant image features. Smoothing approaches that are optimized for suppressing superimposed high-frequency noise will on the other hand lead to an amount of smoothing that is primarily determined by the noise level, and therefore not necessarily corresponding to scale-invariant image descriptors. In this respect, these two types of scale determination approaches can lead to fundamentally different results.

If the task is to detect fine-scale details with amplitude and/or characteristic scales comparable to the noise, it does, however, not seem unlikely that the two types of approaches could possibly benefit from each other.

Applications

Interest point detectors with built-in scale selection: Below, four different interest point detectors with automatic scale selection will be presented. A more general set of scale-invariant interest point detectors defined according to a similar methodology can be found in [16,17] with an in-depth theoretical analysis of their scale selection properties in [19].

Blob detection: Based on the notion of scale-space extrema, straightforward methods for blob detection can be obtained by detecting scale-space extrema of either (i) the scale-normalized Laplacian $\nabla_{norm}^2 L = t(L_{xx} + L_{yy})$ or (ii) the scale-normalized determinant of the Hessian $\det \mathcal{H}_{norm} L = t^2(L_{xx}L_{yy} - L_{xy}^2)$ [11,13]. Specifically, using the *Laplacian* operator one can detect:

- bright blobs from negative scale-space minima of $\nabla_{norm}^2 L$, and
- dark blobs from positive scale-space maxima of $\nabla_{norm}^2 L$.

Using *the determinant of the Hessian* one can on the other hand detect:

- bright blobs from positive scale-space maxima of $\det \mathcal{H}_{norm} L$ that satisfy $\nabla^2 L < 0$,
- dark blobs from positive scale-space maxima of $\det \mathcal{H}_{norm} L$ that satisfy $\nabla^2 L > 0$, and
- saddle-like image features from negative scale-space minima of $\det \mathcal{H}_{norm} L$.

These two blob detection approaches do both satisfy the basic scale selection property that if the scale adaptive blob detector is applied to a two-dimensional Gaussian blob with scale value t_0 , *i.e.* $f(x, y) = g(x, y; t_0)$, then the select scale \hat{t} will be equal to the scale of the blob in the input data, *i.e.* $\hat{t} = t_0$.

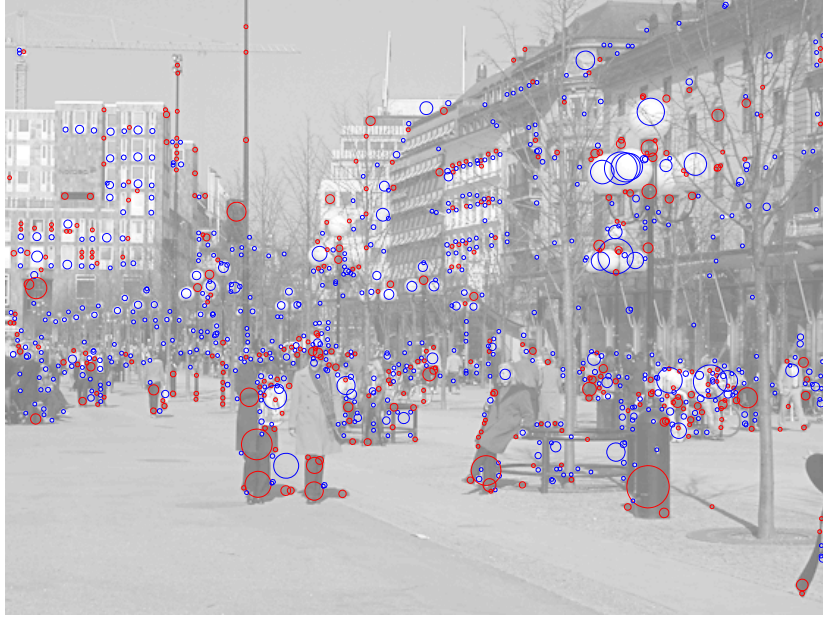
In comparison, the image features obtained from the determinant of the Hessian blob detector do often have better repeatability properties under affine image deformations than Laplacian image features [13,16,17,19].

Figure 5 shows the result of applying these interest point detectors to a grey-level image. Please, note how the variations in the detection scales of the blob responses reflect the perspective scaling effects in the scene.

Corner detection: A straightforward method for scale-invariant corner detection can be obtained by detecting positive scale-space maxima and negative scale-space minima of the scale-normalized *rescaled level curve curvature* measure

$$\tilde{\kappa}(L) = t^{2\gamma} |\nabla L|^2 \kappa(L) = t^{2\gamma} (L_x^2 L_{yy} + L_y^2 L_{xx} - 2L_x L_y L_{xy}) \quad (19)$$

scale-space extrema of the Laplacian $\nabla_{norm}^2 L$



scale-space extrema of the determinant of the Hessian $\det \mathcal{H}_{norm} L$



Fig. 5. *Scale-invariant interest points* obtained from the 1000 strongest scale-space extrema of the Laplacian $\nabla_{norm}^2 L$ and the determinant of the Hessian $\det \mathcal{H}_{norm} L$ with the size of each circle reflecting the detection scale of the corresponding feature. Red circles represent local maxima of the operator response, while blue circles indicate local minima. (Adapted from [16].)

scale-space extrema of the rescaled level curve curvature $\tilde{\kappa}(L)$



the Harris-Laplace operator



Fig. 6. *Scale-invariant interest points* obtained from the 1000 strongest scale-space extrema of the rescaled level curve curvature $\tilde{\kappa}(L)$ and the Harris-Laplace operator. The size of each circle reflects the detection scale of the corresponding feature. For the rescaled level curve curvature operator $\tilde{\kappa}(L)$, the colour of the circles show the sign of the curvature; red circles represent a local maxima of the operator response, while blue circles indicate local minima. (Adapted from [16].)

where $\kappa(L)$ denotes the curvature of the level curves of the Gaussian smoothed image at any scale and $\gamma = 7/8$ turns out to be a good choice [13, section 6] [16]; see figure 6(a) for an illustration.

The *Harris-Laplace* operator [23] is structurally different in the respect that it uses different entities for spatial selection (the Harris measure) and scale selection ($\nabla_{norm}^2 L$); see figure 6(b).

Edge detection: With regard to edge detection, the evolution properties over scales of the *scale-normalized gradient magnitude*

$$|\nabla L|_{norm} = t^{\gamma/2} \sqrt{L_x^2 + L_y^2} \quad (20)$$

can be shown to reveal local characteristics of the type of edge [12, section 4]. Specifically, by choosing $\gamma = 1/2$ a local maximum over scales will be assumed at a scale corresponding to the diffuseness of a one-dimensional diffuse step edge

$$\Phi(x; t_0) = \int_{u=-\infty}^x g(u; t_0) du \quad (21)$$

and may then provide cues to *e.g.* focus blur, shadow edges or rounded edges.

Ridge and valley detection: Let e_p and e_q denote the eigendirections of the Hessian matrix $\mathcal{H}L$ such that the mixed second-order derivative in this coordinate frame is zero $L_{pq} = 0$ and denote the eigenvalues of the Hessian matrix by L_{pp} and L_{qq} . These eigenvalues are also referred to as *principal curvatures* and these directions are assumed to be ordered such that $L_{pp} < L_{qq}$.

Then, a differential geometric definition of the *ridges* in the image at any scale can be expressed as the set of points that satisfy [12, section 5.2]

$$L_p = 0, \quad L_{pp} \leq 0, \quad |L_{pp}| \geq |L_{qq}|. \quad (22)$$

Similarly, the *valleys* at any scale can be defined from [15]

$$L_q = 0, \quad L_{qq} \geq 0, \quad |L_{qq}| \geq |L_{pp}|. \quad (23)$$

With $R_{\gamma-norm}$ denoting a scale-normalized measure of *ridge strength* (or valley strength) defined from the principal curvatures L_{pp} and L_{qq} , one can also express ridge and valley detection methods with automatic scale selection by detecting *scale-space ridges* using the definition

$$L_p = 0, \quad L_{pp} \leq 0, \quad \partial_t(R_{\gamma-norm}) = 0, \quad \partial_{tt}(R_{\gamma-norm}) \leq 0 \quad (24)$$

and *scale-space valleys* according to

$$L_q = 0, \quad L_{qq} \geq 0, \quad \partial_t(R_{\gamma-norm}) = 0, \quad \partial_{tt}(R_{\gamma-norm}) \leq 0. \quad (25)$$

Specifically, it can be shown that for natural measures of ridge or valley strength, the choice $\gamma = 3/4$ implies that the selected scale will reflect the width of a Gaussian ridge (or valley) [12]. For generalizations to 3-D images, see [26,6,9].

Feature tracking: By adapting the scales for feature detection by a local scale selection mechanism, the resulting image features will be robust to scale changes, which means that they can be matched over substantial size variations [2]. Indeed, the variations over time in the characteristic scale estimates obtained during feature tracking can, if appropriately implemented, be robust enough for computing estimates of *time-to-collision* [20,25].

Image-based matching and recognition: The SIFT descriptor [22] comprises a bottom-up keypoint detection stage with scale-space extrema detection in a differences-of-Gaussians (DoG) pyramid. The scale-invariant properties of the SIFT descriptor can be explained as follows:

From the way that the DoG operator is implemented in the pyramid in [22], it follows that the normalisation will be similar to the scale-normalized Laplacian. Using the fact that the scale-space representation satisfies the diffusion equation, it follows that the Laplacian operator can be approximated from the difference between two levels in the scale-space representation

$$\frac{1}{2}\nabla^2 L(x, y; t) = \partial_t L(x, y; t) \approx \frac{L(x, y; t + \Delta t) - L(x, y; t)}{\Delta t} = \frac{DOG(x, y; t, \Delta t)}{\Delta t} \quad (26)$$

i.e., from the difference of two Gaussian smoothed images.

With the scale levels distributed such that the ratio between successive scale levels is k when measured in terms of $\sigma = \sqrt{t}$, (*i.e.*, $\sigma_{i+1} = k \sigma_i$ and $t_{i+1} = k^2 t_i$ which implies that $\Delta t_i = (k^2 - 1) t_i$), it follows that [16]

$$\begin{aligned} DOG(x, y; t) &= L(x, y; k^2 t) - L(x, y; t) \\ &\approx (k^2 - 1) t (\partial_t L(x, y; t)) = (k^2 - 1) t \frac{1}{2} \nabla^2 L(x, y; t) \\ &= \frac{(k^2 - 1)}{2} t \nabla^2 L(x, y; t) = \frac{(k^2 - 1)}{2} \nabla_{norm}^2 L(x, y; t) \end{aligned} \quad (27)$$

Hence, with self-similar sampling of the scale levels, the pyramid implemented DoG interest point operator can be interpreted as an approximation of the scale adapted Laplacian operator in equation (11).

In the SURF descriptor [1], local feature detection is performed by detecting local extrema over space and scale of an approximation of the the determinant of the Hessian operator in terms of Haar wavelets, with the filters normalized to constant l_1 - or Frobenius norm over scales. According to equation (9), the γ -normalized derivative concept corresponds to normalisation of the Gaussian derivative operators to unit L_p -norm over scales. Furthermore, it was shown in [20] that normalizing the filter responses to constant l_p -norm over scales gives better accuracy in a practical implementation than normalisation of the discrete filters by multiplication with the scale parameter raised to a power of $m\gamma/2$, where m denotes the order of differentiation. Hence, the initial feature detection step in the SURF descriptor can be seen as an approximation of the scale-normalized determinant of the Hessian operator in equation (12).

The scale invariant property of the actual image descriptors in the SIFT and SURF descriptors do in turn follow from the scale-invariant properties of the initial feature detection step, in line with the general framework for computing scale-invariant image descriptors from scale estimates obtained from local extrema over scales of scale-normalized differential expressions, as described in the “Theory” section above.

In these ways, the notion of scale selection constitutes a general mechanism for computing scale-invariant image descriptors for image-based matching and recognition.

Recommended Readings

- [1] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool. Speeded up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.
- [2] L. Bretzner and T. Lindeberg. Feature tracking with automatic selection of spatial scales. *Computer Vision and Image Understanding*, 71(3):385–392, Sep. 1998.
- [3] J. Crowley and O. Riff. Fast computation of scale normalised receptive fields. In L. Griffin and M. Lillholm, editors, *Proc. Scale-Space Methods in Computer Vision (Scale-Space’03)*, volume 2695 of *Lecture Notes in Computer Science*, pages 584–598, Isle of Skye, Scotland, Jun. 2003. Springer.
- [4] J.H. Elder and S.W. Zucker. Local scale control for edge detection and blur estimation. *IEEE Trans. Pattern Analysis and Machine Intell.*, 20(7):699–716, July 1998.
- [5] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Optical Society of America*, 4:2379–2394, 1987.
- [6] A. F. Frangi, Niessen W. J., R. M. Hoogeveen, T. van Walsum, and M. A. Viergever. Model-based quantitation of 3D magnetic resonance angiographic images. *IEEE Trans. on Medical Imaging*, 18(10):946–956, 2000.
- [7] T. Kadir and M. Brady. Saliency, scale and image description. *Int. J. of Computer Vision*, 45(2):83–105, 2001.
- [8] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [9] K. Krissian, G. Malandain, N. Ayache, R. Vaillant, and Y. Troussset. Model-based detection of tubular structures in 3D images. *Computer Vision and Image Understanding*, 80(2):130–171, 2000.
- [10] Y. Li, D. M. J. Tax, and M. Loog. Supervised scale-invariant segmentation (and detection). In *Proc. Int. Conf. Scale Space and Variational Methods in Computer Vision (SSVM 2011)*, volume 6667 of *Lecture Notes in Computer Science*, pages 350–361, Ein Gedi, Israel, 2012. Springer.
- [11] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Springer, 1994.
- [12] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *Int. J. of Computer Vision*, 30(2):117–154, 1998.
- [13] T. Lindeberg. Feature detection with automatic scale selection. *Int. J. of Computer Vision*, 30(2):77–116, 1998.

- [14] T. Lindeberg. A scale selection principle for estimating image deformations. *Image and Vision Computing*, 16(14):961–977, 1998.
- [15] T. Lindeberg. Scale-space. In Benjamin Wah, editor, *Encyclopedia of Computer Science and Engineering*, pages 2495–2504. John Wiley and Sons, Hoboken, New Jersey, 2008.
- [16] T. Lindeberg. Generalized scale-space interest points: Scale-space primal sketch for differential descriptors. 2012. (under revision for *Int. Journal of Computer Vision*).
- [17] T. Lindeberg. Image matching using generalized scale-space interest points. In *Proc. Int. Conf. Scale-Space and Variational Methods for Computer Vision (SSVM 2013)*, volume 7893 of *Lecture Notes in Computer Science*, pages 355–367. Springer, 2013.
- [18] T. Lindeberg. Invariance of visual operations at the level of receptive fields. *PLOS ONE*, 8(7):e66990, 2013.
- [19] T. Lindeberg. Scale selection properties of generalized scale-space interest point detectors. *J. of Mathematical Imaging and Vision*, 46(2):177–210, 2013.
- [20] T. Lindeberg and L. Bretzner. Real-time scale selection in hybrid multi-scale representations. In L. Griffin and M. Lillholm, editors, *Proc. Scale-Space Methods in Computer Vision (Scale-Space’03)*, volume 2695 of *Lecture Notes in Computer Science*, pages 148–163, Isle of Skye, Scotland, 2003. Springer.
- [21] M. Loog, Yan Li, and David Tax. Maximum membership scale selection. In *Multiple Classifier Systems*, volume 5519 of *Lecture Notes in Computer Science*, pages 468–477. Springer, 2009.
- [22] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. of Computer Vision*, 60(2):91–110, 2004.
- [23] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *Int. J. of Computer Vision*, 60(1):63–86, 2004.
- [24] P. Mrázek and M. Navara. Selection of optimal stopping time for nonlinear diffusion filtering. *Int. J. of Computer Vision*, 52(2–3):189–203, 2003.
- [25] A. Negre, C. Braillon, J. L. Crowley, and C. Laugier. Real-time time-to-collision from variation of intrinsic scale. *Experimental Robotics*, 39:75–84, 2008.
- [26] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis. 3D multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. *Medical Image Analysis*, 2(2):143–168, 1998.
- [27] J. Sporring, C. J. Colios, and P. E. Trahanias. Generalized scale selection. In *Proc. Int. Conf. on Image Processing (ICIP’00)*, pages 920–923, Vancouver, Canada, 2000.
- [28] A. P. Witkin. Scale-space filtering. In *Proc. 8th Int. Joint Conf. Art. Intell.*, pages 1019–1022, Karlsruhe, Germany, Aug. 1983.