Segmentation of Natural Images Using Scale-Space Representation with Multi-Scale Edge Supervised Hierarchical Linking

Technical Report

Oscar Divorra Escoda and Pierre Vandergheynst
Signal Processing Laboratory (LTS)
Swiss Federal Institute of Techonlogy in Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
WWW home page: http://ltspc4.epfl.ch
{oscar.divorra,pierre.vandergheynst}@epfl.ch

Abstract— **In general purpose computer vision systems, non-supervised image analysis is mandatory in order to achieve an automatic operation. In this paper a different approach to image segmentation for natural scenes is presented. Scale-Space representation is used to extract the structure from meaningful objects in the image. A hierarchical decomposition of the image is performed from the iso-intensity paths. The Scale-Space stack is generated using isotropic diffusion on the basis of linear Scale-Space theory. From that, the independence of the algorithm from the image content and particular characteristics is ensured. In the framework of this work, it is also introduced the use of additional information to improve the robustness in the structure extraction. In addition to the set of several diffused versions of the image, a representation of edges through scale is included as a feature in order to supervise the generation of the hierarchical tree that represents the image.**

Keywords—**Scale-Space, isotropic diffusion, unsupervised segmentation, edge detection, Wavelets, image structure, uncommitted visual front-end.**

I. Introduction

In the future, smart systems [1] relying on visual information will need from general purpose non-adapted visual front ends. These will have to adapt to the most diverse situations and scenes being its working principle independent of particular features of the image. The idea of an uncommitted visual front end in computer vision is very closely related to the concept of artificial intelligence. Such a system should be able to identify meaningful objects in the scene independently of its nature. It is for sure that, despite a complete uncommitment is desired, a smart system will be supposed to have some previous knowledge about the object that will be identified. Anyway, before such a high level of image understanding, a lower level of analysis is necessary to extract from the scene suitable information about its general characteristics.

In the literature appear many image and sequences analysis algorithms that are more or less based on low level features. An elaborated and complete statistical based example can be found in [2], [3]. This approach has demonstrated to perform well in several applications, but is not less expensive computationally. Anyway, it has also the evidence of keeping some dependence on human supervision, since a large number of parameters and factors need to be tuned. Thus, a general statement of the present problem can be explained as: The necessity of some methodology to extract and treat the most important information and features of an image or sequence with absolute or almost absolute independence of human intervention.

One of the most relevant informations contained in an image is structure. It gives information about how the different regions in a scene are organized. Structure gives a good low level basis for a primary ordering of information. It allows to perform an early classification according to scale and consistent regions. In fact, studies [4], [5] stablish links between analysis of image structure and the HVS (human visual system). In particular, there are evidences about the possibility that the brain uses some kind of scale-space like analysis to perform preliminary processing on the images before any semantical classification is realized. It follows from this fact that in case the HVS really uses this principle, it is a very natural direction of evolution for computer-vision systems.

The use of scale-space representation to perform early image analysis provides a tool very suitable to extract the image structure. Furthermore, it allows a non-committed configuration in a particular case of scale-space. The idea of multi-scale analysis to perform unsupervised low level image segmentation was already introduced by *Burt et al.* [6] and further developed by *Ziliani and Jensen* [7] to a multi-scale and multi-feature algorithm. They used the idea of hierarchical representation of the input data for segmentation purposes.

The concept of the necessity of analysing image structure for a suitable image understanding was introduced by *Koenderink* [8]. From this, many works have appeared on several directions. On linear scale-space, *Lindeberg* [9] has come to be a reference. From the scale-space theory, and from the fact of evolution through scale of different image phenomena, different approaches trying to extract structure from the changes suffered by extrema can be found in the literature, see for example *Lifshitz* [10] and *Henkel* [11], [12].

In this paper, an approach to a possible non-committed visual front-end is presented. On the basis of the work performed by *Vincken* [13] for medical images, a new scope is proposed. On the idea of a possible general computer vision system several experiences are realized using the image structure. A promising future is foreseen for a large spectrum of applications where the supervision of a computer system would be desired, and the main input is visual information. Additionally to the low-pass

representation through scale, the use of multi-scale band-pass information is used as well. Taking the HVS as example additional information is introduced in the data set analysis for the generation of the hierarchical tree that will hold the image structure. As it will be seen later, this comes from the fact that HVS besides of extracting objects structure, makes great emphasis on edges. This will improve the precision on the detection of consistent segments according to meaningful objects.

This paper is structured as follows. In section II a general introduction to Scale-Spaces can be found. In section III *Vincken's*[13] hierarchical segmentation algorithm is described. Section IV presents two possible basis for edge representation through scale and the best fit for our application is discussed. Section V will show how the information from edges can be included in the retrieval of image structure. Finally several results are presented and discussed in section VI followed by the conclusions in section VII.

II. Scale-Space

The structure of images has a close relation with multi-scale representation. One of the most clear examples of multi-scale (or multi-resolution) data representation is Scale-Space. Such a representation is composed by the stack of successive versions of the original data set at coarser scales. It is assumed then, the bigger the scale, the less information referred to local characteristics of the input data will appear. Anyway, we also impose that general information applying to large scales will last through scale. Taking that into account, it is reasonable to think that local and high resolution scale information can be related to general and low resolution information. This will enable us to extract image structure.

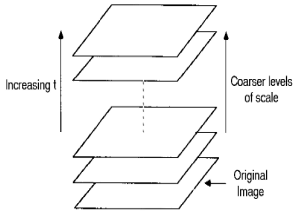


Fig. 1. Scale-Space stack.

A. Scale-Space Flavors

Scale-spaces can be generated on the basis of many different principles. It is just necessary to be able to obtain in some way a description of the image structures through scale. According to the application, it will be possible to derive the scale stack from different scale operators. In the literature, different approaches can be found. General comparisons are available in [13], [4]. A rough classification might be:

Linear Scale-Space is a one parameter family of images derived from the linear diffusion equation. Koenderink [8] derived the unique linear kernel that satisfies such condition: the Gaussian. Further details are given in sec. II-B, since it is the base for the present work.

Non-Linear Scale-Spaces relax the constraint of uncommitment in the processing of visual information, but keeping the main properties of a scale-space. Some of them are:

1. Luminance conserving scale-spaces where examples are *Gradient dependent diffusion* [14] and *Tensor Dependent Diffusion* [15].
2. Geometric flows where the evolution through scale of curves and surfaces are considered as a function of their geometry [16].
3. Morphological scale-spaces are the ones coming from the successive erosion or dilation of an image with a structuring element of increasing size [17].

The choice of one or other principle to obtain the derived set of images at different scales is a matter of the particular application. Depending on the previous knowledge about the characteristics of the images to analyse, one can be selected. That will allow to take advantage of some special feature and will allow to preserve some image particularity.
In the case where there is no previous knowledge of the kind of scene, it is not possible to foresee which will be the most advantageous scale-space. In that case, the best is to stay on the basis of the uncommitted visual front-end [18] where properties like:

- linearity,
- spatial shift invariance,
- isotropy,
- scale invariance,

will be kept. Such a set of properties is satisfied by the Linear Scale-Space. It will then be taken as the basis of this work, since no dependence on the input data is desired and a maximum of flexibility is preferable.

B. Linear Scale-Space

Assumptions made by *Lindeberg* [19] are based on the idea of using successive convolutions to generate the scale-space. *Koendering* first reailzed [8] about which should be the basis for the structure of images analysis. Under several constraints, he defined the diffusion equation, given by (1), as the generator of its scale-space.

$$\frac{\partial I\left(\vec{x},t\right)}{\partial t} = \Delta I\left(\vec{x},t\right), \tag{1}$$

where I stands for the luminance of the image which depends on $\vec{x}$, position ($\vec{x} = (x, y)$), and t, scale.

From (1) and from the constraint of using convolution to generate the subsequent scale levels one finds that the unique kernel that satisfies both is the Gaussian.

The Gaussian is the Green function of the diffusion equation and for an infinite domain it is given by:

$$I(\vec{x},t) = \int_D G(\vec{x}-\vec{x}',t)\cdot I(\vec{x}',0), \tag{2}$$

where G stands for the Gaussian function :

$$G(\vec{x},t) = \frac{1}{2\pi t} e^{\left(-\frac{x^2+y^2}{4t}\right)}\Big|_{\vec{x}=(x,y)}. \tag{3}$$

Notice the linear relation between the scale parameter t and the variance of the Gaussian kernel $\sigma^2 = 2t$.

From this simple formulation, it follows that the problem of a basis generation for an uncommitted front-end turns into the simple successive blurring of the image to analyse. A linear scale-space representation will be then defined by: The succession of an infinitely dense set of images derived from the original one through convolution by a Gaussian kernel where the continuous scale parameter t variate monotonically ascending. In other words: A continuous three-dimensional blurring representation of a continuous image where the third dimension is defined by the scale parameter, which is monotonic and ascending.

The fact that in this case scale-space is being generated by blurring, confers to this representation some interesting properties which are:

- Causality: coarser scales can only be caused by what happened at finer scales.
- Maximum principle: at any scale change, the maximal luminance at coarser scale is always lower then the maximum intensity at the finer scale, the minimum is always larger.
- No new extrema at larger scales: this holds **only** for one-dimensional signals [19].
- Physics of luminance diffusion: the decay of the luminance with scale is equal to the divergence of a flow.

Once the Gaussian kernel is established as the unique scale-space operator to change scale, there is an important additional result. One of the most useful results in linear scale-space theory is that the spatial derivatives of the Gaussian are as well solutions of the diffusion equation, and together with the zeroth order Gaussian they form a complete family of differential operators [9]. It can be seen that:

$$\frac{\partial}{\partial x}(I * G) = I * \frac{\partial G}{\partial x}, \tag{4}$$

which is trivial according to the linearity of the differential operator and the convolution. (In II-B $*$ stands for convolution).

From this, it was seen that derivatives are given at certain scale. Multiscale differential analysis can thus be performed.

B.1 Discrete Approximation: Stack Generation

The main problem with Scale-Space applications is that the theory holds on the continuous domain. In order to be able to use such analysis approach, a discrete approximation of the principle is mandatory.

Lindeberg [9] postulates that the Linear Scale-Space should be generated by convolution with a one-parameter family of kernels, i.e. $L(\vec{x}, 0) = I(\vec{x})$ and for $t > 0$.

$$L(\vec{x}, t) = \sum_{\vec{x}' \in \mathcal{Z}^{\mathcal{N}}} G(\vec{x}', t) I(\vec{x} - \vec{x}'). \tag{5}$$

where functions are defined to by discrete spatially and continuous in the scale dimension.

He defines the Scale-Space Family of Kernels $G : \mathbb{Z}^{\mathbb{N}} \times \mathbb{R}_+ \to \mathbb{R}$ as satisfying:

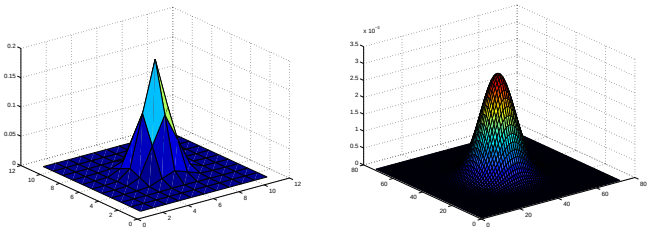
- $G(\cdot, 0) = \delta(\cdot)$,
- *the semi-group property* $G(\cdot, s) * G(\cdot, t) = G(\cdot, s + t)$,
- *the symmetry properties* and
- *the continuity requirement* $||G(\cdot, t) - \delta(\cdot)||_1 \to 0$ *when* $t \downarrow 0$.

and a Scale-Space representation:

Definition 1: Let $I : \mathbb{Z}^{\mathbb{N}} \to \mathbb{R}$ be a discrete signal and let $G : \mathbb{Z}^{\mathbb{N}} \times \mathbb{R}_+ \to \mathbb{R}$ be a scale-space family of kernels. Then, the one-parameter family of signals $L : \mathbb{Z}^{\mathbb{N}} \times \mathbb{R} \to \mathbb{R}$ given by Eq. 5 is said to be the scale-space representation of I generated by G.

If we want to approximate the discrete Scale-Space kernel, by a discrete Gaussian kernel, we face two main problems.

1. The Gaussian is defined over an infinite domain, which means that for a practical implementation a truncated version will be necessary.
2. Discrete Gaussian approximations differ much from the continuous Gaussian function at lower scales.



(a) Discrete Gaussian kernel to generate the 1 octave distance image.

(b) Discrete Gaussian kernel to generate the 7.25 octave distance image.

Fig. 2. The problem of the approximation of the continuum with discrete.

From eq. 6 we find the expression of the discrete Gaussian,

$$g(n, m, \sigma) = \frac{1}{2\pi\sigma^2} e^{(-\frac{n^2+m^2}{2\sigma^2})}, \tag{6}$$

where $n, m \in \mathbb{Z}$. The same reasoning can be done in the Fourier domain, where it follows from applying the DFT to eq. 6:

$$G(k, l, \sigma) = e^{(-((\frac{k}{N})^2 + (\frac{l}{M})^2) 2\pi^2\sigma^2)}, \tag{7}$$

where N and M specifies the Fourier transform dimensions and $k, l \in \mathbb{Z}$ represent the discrete frequencies in the Fourier domain.

As said before, the kernel needs to be truncated since the Gaussian function has an infinite support. Defining the Gaussian matrix as a NxN matrix, then $N = 2K\sigma$ (or equivalently truncating the Gaussian at radius $r = K\sigma$), where K defines the accuracy factor used in the approximation. $K = 3$ is considered as being an acceptable approximation.

The choice of domain to generate the stack is a matter of different preferences. In what refers to spatial convolution, pixels affected by boundary effects can be controlled exactly. Approximation error will be bigger at fine scales. Convolution is computationally very expensive, although in this case, since the Gaussian kernel is separable some fast computing can be implemented. In the Fourier domain, the generation is less expensive. The boundary problem gets somehow arranged by the implicit periodization, although literature [10], [13] qualifies such approach as including undesirable artifacts at large scales, because

it modifies the original structure of the image. The main problem in the Fourier domain is that the Gaussian kernel is worse approximated at bigger scales. A way to solve that is to increase the resolution of the Fourier transform. Performing the Fourier transform using much more samples than the image size showed to improve significantly the quality of the segmentation. Furthermore, it seemed to perform similarly when using different kinds of padding around the original image (see section II-B.2) if the Fourier transform was big enough. In our test, stack is being generated in the Fourier domain with a transform of size significantly bigger (about 3 or 4 times) than the original image in order to have enough resolution for the Gaussian kernel.

Besides the approximation of the kernel in what concerns to space dimensions, the scale parameter has to be taken as well into account. In a practical application it has to be sampled too. Sampling the scale dimension will determine (according to the resolution) the possibility of following the structure. So, sampling very coarsely the scale parameter will lead to wrong and undesired effects in segmentation.

To obtain a uniform sampling in the scale direction, and relate it linearly with a parameter δ_τ, the std. deviation σ_n of the Gaussian must have the form:

$$\sigma_n = \varepsilon e^{(\tau_0 + n \cdot \delta_\tau)} \tag{8}$$

where ε is a parameter which specifies the initial scale for $n = 0$, τ_0 is just a possible offset of the first level in Scale-Space, and δ_τ specifies the scale sampling.

In this work, sampling of the scale parameter has been taken up to 4 layers per octave $\delta_\tau = \frac{\ln 2}{4}$ in order to reach a compromise between resources needs and accuracy.

Discrete Gaussian kernels with discrete scale parameter lack the property of generating one level of the scale-space from the level below with an iterative. The property of *semi-group* is failing. That is one of the reasons that impulsed *Lindeberg* to work directly on a kernel approximating the discrete version of the diffusion equation [19]. When the stack is being generated with the Gaussian kernel, all the levels will have to be the product of convolving by a kernel of the appropriate scale with the original image.

B.2 Finite Image Limitations: Border Extensions

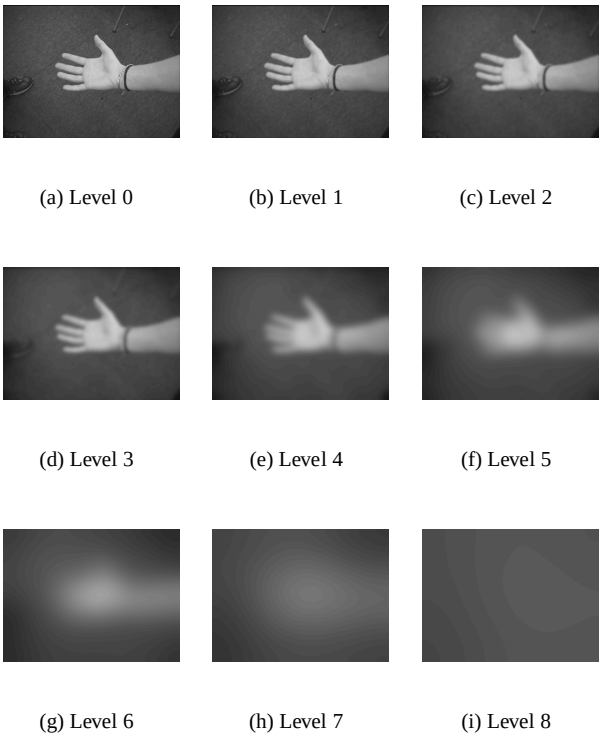
One of the most important problems on scale-space generation is the fact that theory was conceived in an infinite space. This problem, is especially noticed when reaching the calculation of large scale levels. There Gaussian kernels become really big compared to the image and border effects stand for an important handicap. The most relevant solutions to the problem are quite logical and some of them quite used in other domains of image processing. Those are:

- *Zero Padding* [10] which has been shown to be not very suitable when looking for iso-intensity paths, since it affects the structure concerning to low pass components. It may serve, anyway, when just the retrieval of extrema is desired.
- *Mean Image Value Padding* has its explanation on the fact that scale-space generation is affected by the borders when the kernel is very big. If it is considered that at big scales the image tends to its mean, then it should be fair padding around the image with the mean. When the influence of extern pixels to the image will be relevant, the image will already be converging to the average, and so it will not be very affected.
- *Periodization* would keep also the mean level at large scales but will introduce additional structure with a huge slope around the image.
- *Mirroring* solves the problem of the slope around the image, but still adds structure to the scale space.
- *Extrapolation* can be seen as a solution in the case where no slope is desired, although depending on the extrapolation order, in general it will affect the image mean value.

The last four points are further explained and compared in [13].

In our case, the Mean Image Value Padding was used since it does not affect the structure in terms of DC component and it does not add new structures around the image.

Summarizing, the choice of border extensions has to be done correctly according to the way the scale-space is going to be extracted. In our case, since the iso-intensity paths are going to be followed, we should not affect the DC component.



(a) Level 0 (b) Level 1 (c) Level 2

(d) Level 3 (e) Level 4 (f) Level 5

(g) Level 6 (h) Level 7 (i) Level 8

Fig. 3. Scale-Space representation by Gaussian blurring of image (a) with 1 sample per octave in scale.

III. Scale-Space Segmentation

The basis of scale-space segmentation is the extraction of the hierarchical structure of the image. In figure 4 the description corresponding to the general algorithmcan be seen. First, the Scale-Space representation is generated. Right after, the structure analysis is realized building up the tree-like hierarchy (figure 5). From this, a set of segments is obtained. Those correspond to all the pixels hanging down the selected roots from the hierarchy. In the end just a morphological filtering on the encountered regions masks is performed to erase little spots or

regions corresponding to mistakes occurred during the phase of structure analysis. In this work, the multi-resolution segmentation algorithm by *Vincken* [13] is taken as a starting point.

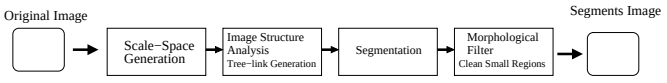


Fig. 4. Segmentation Scheme.

A. Space Generation

Scale-space generation is explained in detail in section II-B.1. We show the constraints and limitations due to discrete approximations.

B. Linking up through space

The algorithm for the construction of the structure, on a simple approach [13], is based on the tracking of the iso-intensity paths through scale. Other algorithms where proposed relying on extrema [12], [11], [20], [10], [21], but we considered to be more consistent and generic to search for the iso-intensity paths. This is because image pixels can not be fully described by maxima and minima.

The algorithm sets up the structure establishing relations between pixels of consecutive levels. On the finest scale (the original image) all the pixels are related to the pixel from the first blurred image on the scale direction. At this level, not all pixels will receive a link from a pixel from the level below. This is because due to blurring, the image contains less information, and so a pixel from the upper level (bigger scale), will be related to a bigger number than one pixel from the level below (finer scale). Pixels from the finer scale level will represent the details lost by blurring in the upper level. This linking up is performed between all the scale levels. Figure 5 shows a simple schema of the idea. Levels are linked in a tree like structure. These links converge through scale according to the reduction of information imposed by the low-pass filtering.

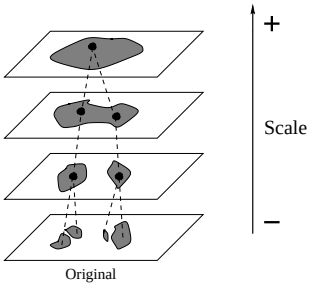


Fig. 5. Hierarchical analysis of the image structure linking pixels through levels.

The basic problem that arises is the search of the parent pixels at a larger scale. *Vincken* proposes as linkage criteria the gray level difference between two different pixels of different neighbor levels. Those pixels having the smallest difference from a limited spatial neighborhood will be linked. That means that taking a valid pixel from a determined level (a pixel who has at least one link from the level below), a search on a circular area around that point will be performed. This search area is proportional to the inner scale (see figure 6).

In addition to the base criteria of gray level difference, some others where added in order to help the convergence [13]. Those rely on different features like for example volume of pixels hanging from the selected parent pixel. This would influence in the way that a pixel having many children is very likely to have more. Another feature would be the average gray level of the hanging pixels. Such a characteristic is quite advantageous when segmenting regions with a uniform gray level, like for example medical images. Factors are represented by:

$$\mathcal{C}_I = 1 - \frac{|I_p - I_c|}{\Delta I_{max}}, \qquad \mathcal{C}_G = \frac{SG_p}{SG_{max}}, \tag{9}$$

where $\mathcal{C}_I$ is the driving feature that relies on the intensity of pixels parent (I_p) and children (I_c). $\mathcal{C}_G$ represents the accessory feature that favorize big segments, SG_p represents the number of pixels associated to a parent pixel, and SG_{max} the maximum value associated to a parent pixel.

$$\mathcal{C}_M = 1 - \frac{|M_p - M_c|}{\Delta I_{max}}, \tag{10}$$

where $\mathcal{C}_M$ is the feature associated to the mean gray value of segments.

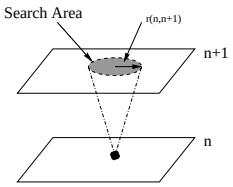


Fig. 6. Search area for a parent pixel.

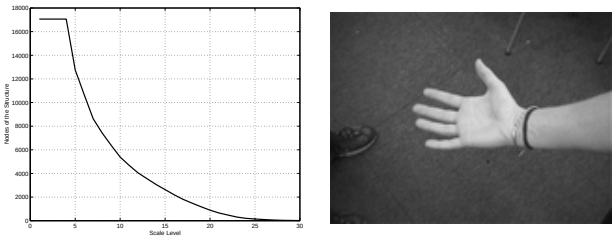
In figure 6, $r_{n,n+1} = k \cdot \sigma_{n,n+1}$ and

$$\sigma^2_{n,n+1} = \sigma^2_{n+1} - \sigma^2_n, \tag{11}$$

where n indicates the scale level.

In figure 7, the convergence rate can be seen (in fact, the number of nodes in every level) through scale. It shows how iso-intensity paths converge to few points at larger scales. The use of the additional features proposed by *Vincken* will accelerate the convergence rate, although in some cases they can contribute to break the structure of the scale-space based on isotropic diffusion since they do not really take into account the scale-space theory.

To improve the linking phase of the algorithm, the most important and decisive stage in terms of final performance, an approach based on a maximum likelihood linking path retrieval was proposed by *Vincken* in [13], [22]. Instead of selecting the parent for a given pixel in each linking level, all the possible paths are kept. The most probable path will be selected in the segments reconstruction stage where from the whole set of possible ones. Although it improves the quality of final segments, the increase of necessary memory resources turns into a real limitation.



(a) Number of nodes (b) Rosa's HAND

Fig. 7. Number of nodes (parent pixels) in every level for the analysis of the image structure relying only on the gray level through scale and the image used to generate the values.

In this method, only two values are used to estimate the iso-intensity paths and local links. Thus, the linking procedure lacks from robustness, mainly at fine scales where noise can be present. It gets easily lost in the search of the strongest link between to pixels from two levels and it is quite common that little segmented regions (mainly in the single parent linking method) emerge to the top of the scale-space stack. Those little miss-segmented regions are regions of pixels such that their hierarchical tree has not found the right path till the largest scale level. There can appear very clearly errors in the linking procedure on the border of big segments. The assumption that those little segments are incorrectly linked is evident if the selected scale level is big. According to the inner scale of a level, regions sizes should be (although very roughly) around the generating kernel size.
The full criteria for linking is:

$$\mathcal{A} = \mathcal{D} \cdot \frac{\sum_{i=1}^{N} w_i \cdot \mathcal{C}_i}{\sum_{i=1}^{N} w_i}, \qquad with\, \mathcal{C}_i \in [0,1], \tag{12}$$

where w_i are possible configurable weights and $\mathcal{D}$ is a weight depending on distance (equation III-B). In here w_i will be fixed and further exploitation of features relaying in scale-space will be studied. We have used:

$$\mathcal{D} = \begin{cases} 1 & \text{if } d_{c,p} \leq 0.5\sigma_p \\ \frac{D(d_{c,p})}{D(0.5\sigma_p)} & \text{if } d_{c,p} > 0.5\sigma_p \end{cases} \tag{13}$$

where

$$D(d_{p,c}) = e^{\left(-\frac{d_{c,p}^2}{2\cdot(\sigma_p^2 - \sigma_c^2)}\right)}. \tag{14}$$

In section V in addition to the low pass information structure obtained from Gaussian blurring, band pass information structure is also taken into account for the hierarchical analysis. Anyway, some other technique of parent search should be studied. It is not enough to take into account which is the most similar pixel in a given region on the upper scale level. A more precise estimation of the direction of variation or gradient of scale-space could be quite useful when extracting the structure. The principle proposed by *Vincken* (called in his work single parent linking) could be considered as a particular case where only a two tap filter is used to look for the minimum gray level variation.

C. Reconstructing Segments

Once the image structure has been estimated, the obtention of segments is evident. To carry out the segmentation it is necessary to select the scale of analysis. From this, all the nodes at that scale level will define a segment each. The segments will be all the pixels connected through the hierarchical tree to the upper selected node.

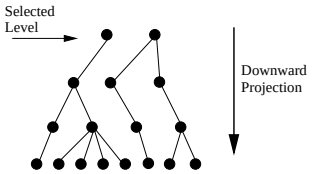


Fig. 8. Scanning of the image structure to obtain the segments

Selection of the upper nodes that define the final number of segments, can be done in different ways. The most simple is the selection of scale level, and from there take all the segments emerging from the hierarchical tree. Other possible approaches appeared in the literature. Those try mainly to look for the nodes that should be root instead of being further linked up. In order to do that, segmentation algorithms apply different seeding rules based on several different features of the scale stack. In [13], [22], [23], [24] some thresholding is applied on the distance measurement between pixels for the linking procedure. When the closest pixel to link is above some similarity distance, it is not linked and it will define a new segment. This can be done selecting the threshold heuristicaly or on the basis of some statistics. The idea of seeding rule retrieval on the basis of a statistical model can be found in the segmentation algorithm based on the pyramid of *Ziliani* [7].

Since scale-space analysis is intended to be used as visual front-end, the use of simple classification of features to select roots on hierarchical trees turns into a lack in segmentation quality. It should be more appropriate the use a feedback from the high level analysis stage in a complete visual system.

The only root node classification in this work will be the selection of scale level. All the nodes in a certain level will be considered as root nodes. It is out of the scope of this paper to attempt to perform an abstract understanding of the image, but investigate the possibilities of scale-space representation for natural images analysis.

D. Cleaning up regions

The problem of little regions miss-segmented due to linking errors introduces quite a high number of little segments of few pixels. It is clear that they do not belong to the selected scale level. A way to remove them is to delete regions smaller than certain area proportional to scale and re-assign those pixels to the big neighbor segments on the basis of some criteria, like average gray level, or big existing regions can be grown using geodesy with some morphological operators.

IV. Edges Through Scale

Edge detection through scale can rely on the application of Gaussian derivatives. According to section II-B, Gaussian derivatives are also a solution of the diffusion equation (1). Thus all the statements that hold for the scale-space generated by Gaussian blurring will also hold for the scale-space generated by one of the Gaussian derivatives. We will find, thus, the hierarchical structure of image edges in this space.

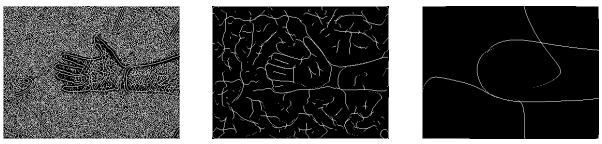
In this work, two different approaches have been studied, those are the use of the first derivative (spatial gradient) and the second derivative (the Laplacian of the Gaussian).

The First Gaussian derivative corresponds, in practical terms, to the spatial gradient of the scale-space levels extracted in sec. II-B.1. All the existent ridges in the gradient module (15) at all levels are extracted. Those define where are the local maximas of image variations, and consequently the location of edges. There is no selection of the most important edges, since those will persist through scale. For the ridges extraction, a morphological procedure using a directional dilation with reconstruction is used [1]. Anyway, any other approach could be taken to extract ridges [9], [18].

The module of the first derivative is represented by:

$$|\vec{\nabla} I(\vec{r},t)| = \sqrt{\left(\frac{\partial I(\vec{r},t)}{\partial x}\right)^2 + \left(\frac{\partial I(\vec{r},t)}{\partial y}\right)^2}, \qquad (15)$$

where t represents the scale.



(a) Level 1 (b) Level 4 (c) Level 7

Fig. 9. Edge representation through scale using the gradient module. 1 sample per 3 octaves (first sample on the first octave)

The Second Gaussian derivative is the Laplacian of the Gaussian. An equivalent scale-space of the second derivative is computed on its basis. It is given explicitly by:

$$\nabla \mathcal{G}(x,y) = -\frac{1}{\pi\sigma^4} \cdot \left[1 - \frac{x^2+y^2}{2\sigma^2}\right] e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (16)$$

Instead of using directly the second derivative, we approximate it. We used the Difference of Gaussians ($\mathcal{DOG}$), which has been recently proved to be the response of the receptive field in the cats retina [25]. To detect edges in scale, the difference between two consecutive levels of the scale-space is computed (figure 12) and then, a zero-crossing detection is performed.

$$\mathcal{DOG}(x) = A_1 e^{\left(-\frac{x-\mu}{2\sigma_1}^2\right)} - A_2 e^{\left(-\frac{x-\mu}{2\sigma_2}^2\right)}, \qquad (17)$$

where $\sigma_1 > \sigma_2$.

For example with two Gaussians A and B of scale 6 and 5 octaves respectively, the $\mathcal{DOG}$ will be the equivalent to the second derivative of a Gaussian of scale 5.5 octaves.

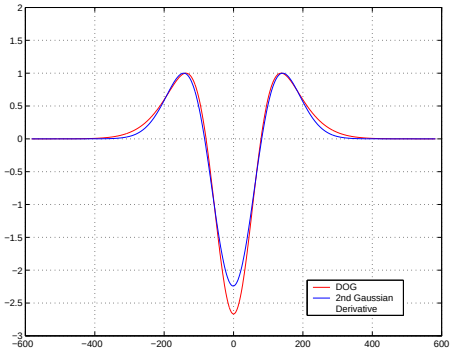
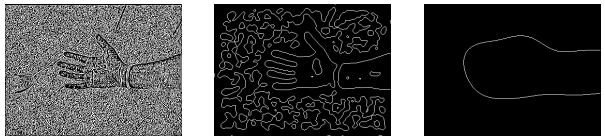


Fig. 10. Comparison between the Laplacian of a Gaussian and the DOG approximation



(a) Level 1 (b) Level 4 (c) Level 7

Fig. 11. Edge representation through scale using DOG. 1 sample per 3 octaves (first sample on the first octave)

In the images from Fig. 11 and 9, we can see how edges of the most important structure are kept through scale.

The use of edges representation through scale on the basis of the second derivative of a Gaussian, in fact, is nothing else than a wavelet representation of the image. In this particular case the the use of a second derivative of a Gaussian is known as the *Mexican Hat* wavelet [26]. This is another analogy with the HVS [5]. There are evidences of certain similarity between some parts of HVS analysis and wavelet analysis. Wavelet representation of images, allows to work and to represent in a structured fashion band-pass information of signals in general.

A possible use of the inherent structure represented by wavelets could lead to an estimation of image structure through them instead of using the low pass information. In any manner, here, edges through scale will be used as an accessory to the Gaussian blurred data in order to get rid of incorrect linking.

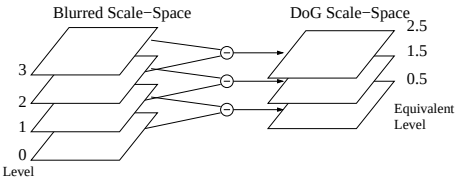


Fig. 12. Generation of the approximation of the Laplacian of a Gaussian from the blurred scale-space.

V. Supervising Segmentation with Edges

Two sources of image structure have been exposed. One extracts information from the low-pass approximation of the image in a multi-resolution way. The other extracts structure from image through the multi-resolution representation provided by

wavelets. One possibility would be to use either one or the other to extract the structure. Another would be the cooperation between both principles in order to profit from both to build a more precise structure to describe the scene under analysis.

In this work the use of both scale-spaces to generate the hierarchical tree has been studied. As a basis, we took the algorithm described in section III-B. The use of the second scale-space is introduced in the procedure of linking through scale. The multiscale edge representation extracted from the second derivative of the Gaussian (see section IV) is used to supervise the linking procedure.

During the linking up procedure, a search to find parents is performed. All the pixels that where already linked from a level below will be linked up to a parent pixel. This linking procedure [13], [22], [23], [24] from level to level does not take into account orientation of structure in itself. It looks for the nearest gray level pixel in a circular area. This is performed independently of the shape of the region where both pixels (child and parent) belong. This uncontrolled link search turns into the possibility that pixels can be linked outside the region they represent. Although locally it is true that the most similar pixels in the upper level are very likely to be the best parents for the child pixel, when search windows are large, children pixels can find sometimes better fits for their gray level some distance away from the supposed ideal pixel. In this situation, when paths evolve through scale, this small mistake turns into a divergence of a whole branch.

Figure 13 shows the algorithm proposed to reduce the divergence of paths during linking. When looking for the relation

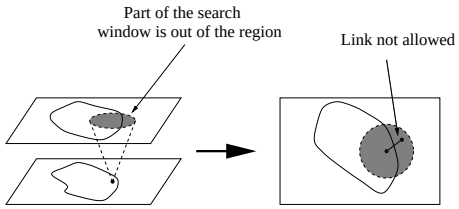


Fig. 13. Wrong linkage problem.

between two pixels, we test if they belong to the same region or blob at that scale. This means that when looking for linkage, all those links that cross an edge of the second derivative representation at the same scale level will not be taken into account. It follows that the area of search for a parent pixel is modified. Only that area that is included into the blob of the child pixel is taken into account in the search window.

In figure VI we show the convergence of pixels through the scale when using the edge supervision. If compared with figure 7 where there is no use of edges in the linking procedure we can see no difference in the speed of convergence. That means the application of edges does not affect the structure negatively. It does not contribute to split segmented regions more but to redirect links into the appropriate blobs, avoiding inter-blob linking of pixels, as it should.

VI. Segmentation Experiments

In this section a set of results and several segmentation experiments are presented. All are realized on natural image sets. Ac-

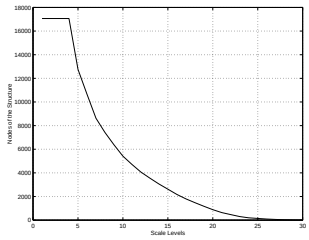


Fig. 14. Number of nodes (parent pixels) in every level for the analysis of the image structure relying on the gray level through scale and the edges representation through scale

cording to the results obtained, the influence of different aspects of the segmentation algorithm will be discussed in the following sections.

A. Parameter Influence

Segmentation are performed relying mainly on the gray level difference between the child pixel and the supposed parent pixels. As it appears in section III-B and in [1], [13], some additional features are used in order to increase stability and force little lost segments to join big segments. Anyway, an excess on those additional components will break the tracking of the real scale-space since such features do not rely on a scale-space basis.

(a) Level of segmentation: $\sigma_n = 50$ pixels

(b) Meaningful object from the segmentation in (a).

Fig. 15. Segmentation of Rosa's hand relaying only on the gray level component through the scale and filtered morphologicaly.

Figure 15 shows the result of applying the algorithm from [13] with single parent linking trying to follow the iso-intensity paths with no additional feature. It is easy to see a great amount of little segments due to wrong linking. Although the target object is segmented, part of the background gets merged to this.

In figure 16 some components of the two secondary features described in [13] are slightly introduced. With respect to the gray level difference a smaller weight is used for those additional components. The figure shows how segments are more consistent and the target object is reached.

In figure 17 a much stronger weight is used for the feature that takes into account the gray level mean of the segments. Although a finer segmentation is achieved in this case, it affects the structure and the tracking up of the scale-space defined by the diffusion equation. Such a case of linking procedure could be considered as a modification of the original scale-space in order to take profit from some special characteristics of a particular application. Those special features could be the uniform

(a) Segmentation of the image Rosa's hand. Level of segmentation: $\sigma_n = 50$ pixels

(b) Meaningful object from the segmentation in (a).

Fig. 16. Segmentation of Rosa's hand using the three components with weights $W_{\mathcal{C}_i} = 1.0$, $W_{\mathcal{C}_m} = 0.4$, $W_{\mathcal{C}_g} = 0.4$ and filtered with the morphological filter.

gray level in the regions of interests, quite a common situation in medical image analysis.

Fig. 17. Segmentation of the image Rosa's hand, using two of the three components: $\mathcal{C}_i$ and $\mathcal{C}_m$, with 1.0 and 1000 as weight values respectively.

The use of the parameter relying on the size can be considered as a help (with small weights) to avoid lost and small segments. Anyway, although the number of small segments will reduce, they will not necessarily correctly merge. The other additional feature, the one that relies on the mean gray level, can be considered as a variation of the scale-space basis (towards a kind of non-linear one) in order to take profit of a special features of the image uniformity.

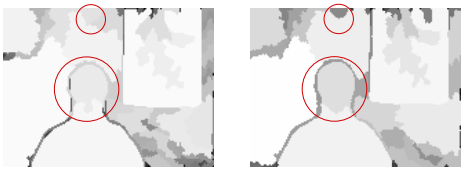
On one hand additional parameters should not be used in an early stage of the segmentation if not introduced by a posterior sketch on the basis of some preliminar analysis. On the other hand, a better extraction of the link structure would help much more, i.e. the use of edges through scale-space (section VI-B) or to track the iso-intensity paths taking into account the direction of minimum variation of the gray level paths through scale, instead of taking just a circular window to look for the parent pixel.

B. Edge Supervision Influence

Edge detection is intended to avoid incorrect linking between different regions separated by an edge. In Fig. 18 we see the effect of the use of edges. Both segmentations are computed using the same parameters, and are segmented on the basis of the same scale level. The only difference is in the use of edges to supervise the correct linking.

An improvement is clearly seen. In the images the most relevant details are signaled where the use of edges are more influent. In figure (b) we see how part of the head is merged to the body, and next to the picture on the wall, there is a little box, which does not appear on the segmentation without edges. In

(a) Image Sergi.



(b) None using edges.

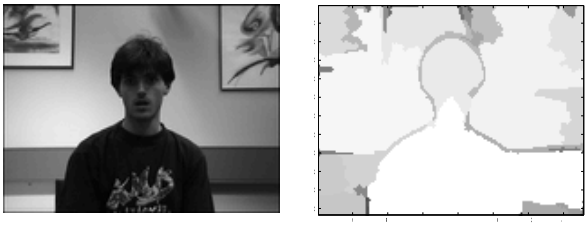
(c) Using edges.

Fig. 18. Comparison of the effect of edge detection on the segmentation. Segmentation of the image Sergi. Level of segmentation: $\sigma_n = 25$ pixels.

(c), since we use the edges at each scale, we keep from linking through them, and we success in avoiding the incorrect linking of the head, improving the definition of the contours. We can observe that here the region that defines the box on the wall is kept, and not wrongly merged.

C. The Contrast Problem

In this segmentation technique, we have been working with the information available on the intensity image only. The main and unique feature is the gray level, with the additional feature of the edge detection. The problem that arises here is that neighbor regions, can have the same or similar intensity. This brings the problem of two neighbor objects badly contrasted can merge. It is clear that sooner or later regions merge in the structure, but bad contrast between regions make them merge before they should.



(a) (b)

Fig. 19. Segmentation (a) of the image Oscar (b) with edge supervision using the Laplacian version and filtered with the morphological filter. Level of segmentation $\sigma_n = 35$

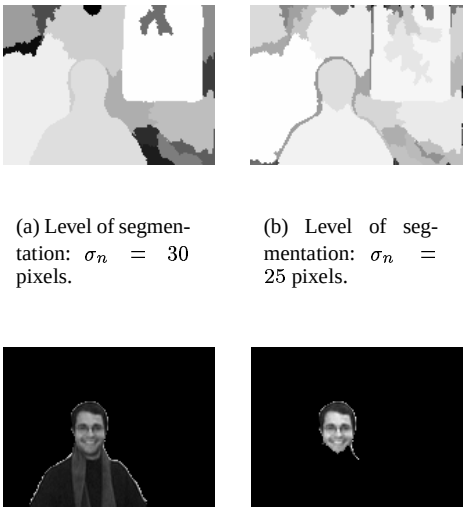
This is the example that can be seen in Fig. 19. We can observe how the algorithm success in segmenting the body, and the head of the subject. But, a part of the wall merges with the body due to the relative low contrast. This is because when building

the structure, the most suitable link connects both regions and the edge estimation can do nothing since it is also affected by the low contrast and the edge at the corresponding scale is also not found. In order to be able to split them and perform a correct analysis, the intervention of some higher level of image understanding might be needed. In fact, it could act as a feed back in order to change the uncommitted front end diffusion basis to some appropriate anisotropic diffusion. In fact, human vision has the same problem.

D. Scale Selection

Image structure gives a hierarchical description of the scene through scale. As it is explained in section III, in order to obtain the segments a scale level is selected. This selection contributes to set the roots of the hierarchical trees that will represent the whole segments. In the underlying idea of the present segmentation principle, this selection of roots would be carried by the high abstraction level layer that would interpret the structures obtained from the the analysis using the scale-space.

As it is seen in figure 20 the scale parameter plays a fundamental rol. It is evident that from image structure can be extracted much useful information to generate segments. In addition, the scale selection helps on the definition of the desired size of segments.



(a) Level of segmentation: $\sigma_n = 30$ pixels.

(b) Level of segmentation: $\sigma_n = 25$ pixels.

(c) Meaningful object from (a).

(d) Meaningful object from (b)

Fig. 20. Obtention of meaningful objects using the Scale-Space segmentation.

When selecting the scale parameter to set the root segments, an implicit selection of the approximated size of segments is performed. From figure 20 we see that for different settings of the root level different segments are obtained. In (a),(c) a scale of $\sigma_n = 30$ pixels is selected. The segments obtained correspond to big meaningful objects according to their structure. We can see in (a) segments clearly defined. A segment represents clearly the subject, another represents the picture on the wall,etc. The wall does not looks like a single segment since scale is still not enough high. In (b),(d), we see how the variation of the scale parameter provides a change on the segment selection. The reduction on the imposed scale produces the splitting of several meaningful segments into smaller segments corresponding to a more detailed partition of the image. One of the most relevant effects is how the head of the subject and the body are splitted. Here a meaningful object is splitted to obtain further smaller divisions meaningful as well. This reflects clearly the applied concept of structure analysis. Selecting the scale level results in chosing the size of the meaningful segmented objects.

VII. Conclusions

In scale-space segmentation, we have seen, that we can not obtain all the possible segments in an image at the same time. We have to chose between bigger or smaller details. Then, according to the choice, we will obtain the segments of objects that are approximately belonging to that scale. Depending on the application, this can be a problem, or an advantage. In the present one, this can help in discriminate implicitly small details. In this way, the task reserved for the following step in the system (segments analysis for image understanding) will be easier.

The technique proposed was shown to be very sensitive to the quality of the generated scale-space. Much care must be taken in its generation, avoiding as much as possible coarse approximations. One of the most important factors in the quality of scale-space was the frequency resolution of the Fourier transform. The other limitations are the spatial resolution (which limits also the use of finer scale sampling rates), the boundary problems (which are more or less solved with the image mirroring or the image mean padding), and the scale sampling rate. Furthermore, the direction of variation of iso-intensity paths should be taken into account.

The inclusion of edge detection is the band-pass part of the image analysis in this procedure. As the human eye, we want to integrate in the same procedure low-pass and band-pass processing of the image. Using the low-pass as a basis to extract the structure and the band-pass as correcting feature, improved the performance.

The tests have shown that the segmentation technique has some problems when two objects or regions not very well contrasted are neighbors. If the segmentation scale is not very well chosen, then these regions can merge. Another problem can be when the optimal scale to begin the segmentation is between two scale samples, and is not available.

Very complicated scenes, can have segmentation problems. If they are very complicated, then it means that there are many details. Small details are more difficult to segment because they have less scale levels to converge. In addition, it is possible that some other band-pass analysis should be necessary for their correct analysis, or some non-linear scale-space analysis.

Finally, it has to be emphasized that the present technique is intended to be a low level image analysis tool. Further higher level understanding analysis layers are supposed to be included in a complete computer vision application. The use of a scale-space framework opens a door to further adaptive analysis. There is always the possibility of variating what kind of diffusion is applied in order to adapt it to the scene.

Scale-Space can be considered as a promising technique for image analysis and segmentation.

ACKNOWLEDGMENTS

We would like to acknowledge the contributions of all the people at LTS with whom we had very interesting discussions. We would like especially to thank the fruitful conversations with Francesco Ziliani, Andrea Cavallaro, Rosa Maria Figueras i Ventura and Sergi Alquezar Alquezar.

The work presented in this paper was carried out in the framework of the M. Sc. Thesis of Oscar Divorra Escoda.

REFERENCES

[1] Divorra Escoda O., "Motion detection & segmentation for audio-visual source separation," M.S. thesis, UPC Barcelona (Catalonia) and EPFL Lausanne (Switzerland), August 2000.

[2] Castagno R., *Video Segmentation Based on Multiple Features for Interactive and Automatic Multimedia Applications*, Ph.D. thesis, EPFL, Lausanne, 1998.

[3] Kunt M. Castagno R.; Ebrahimi T., "Video segmentation based on multipe features for interactive multimedia applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 5, September 1998.

[4] ter Haar Romeny B. M., "Introduction to scale-space theory: Multiscale geometric image analysis," Tech. Rep., Utrecht University, 1996.

[5] Marr D., *Vision*, Freeman Publishers, 1982.

[6] Burt P.; Hong T. H.; Rosenfeld A., "Segmentation and estimation of image region properties through cooperative hierarchical computation," *IEEE Transaction on Systems, Man, and Cybernetics*, 1981.

[7] Ziliani F.; Jensen B., "Usupervised segmentation using modified pyramidal linking approach," in *In Proceedings of ICIP*, Chicago, October 1998.

[8] Koenderink J. J., "The structure of images," *Biological Cybernetics*, vol. 50, pp. 363–370, 1984.

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

[10] Lifshitz L. M.; Pizer S. M., "A multi-resolution hierarchical approach to image segmentation based on intensity extrema," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 6, June 1990.

[11] Henkel R. D., "Segmentation with synchronising neural oscillators," Tech. Rep., Zentrum fr Kognitionswissenschaften, Universitt auf Bremen, 1994.

[12] Henkel R. D., "Segmentation in scale-space," in *In Proceedings of the 6th International Conference on Computer Analysis of Images and Pattern (CAIP)*, Prague, 1995.

[13] Vincken K., *Probabilistic Multi-Scale Image Segmentation by the Hyperstack*, Ph.D. thesis, Utrecht University, 1995.

[14] Perona P.; Malik J., "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions Patern Analysis and Machine Intelligence*, 1990.

[15] Weikert J., "Scale-space properties of nonlinear diffusion filtering with a diffusion tensor," Tech. Rep., Laboratory of Technomathematics, University of Kaiserslautern, 1994.

[16] Osher S.; Sethian S., "Fronts propagating with curvature dependent speed: Algorithms based on the hamilton-jacobi formalism," *Computational Physics*, 1988.

[17] Jackway P. T.; Deriche M., "Scale-space properties of multiscale morphological dilation-erosion," *IEEE Transactions on Patterb Analysis and Machine Intelligence*, vol. 18, 1996.

[18] Lindeberg T., "Scale-space: A framework for handling image structures at multiple scales," in *In Proc. CERN School of Computing*, The Netherlands, September 1996.

[19] Lindeberg T., "Scale-space for discrete signals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, March 1990.

[20] Florack L. M.; ter Haar Romeny B. M.; Koenderink J. J.; Viergever M. A., "Linear scale-space," *Journal of Mathematical Imaging and Vision*, vol. 4, 1994.

[21] Lindeberg T.; Eklundh J. O., "Scale detection and region extraction from a scale-space primal sketch," in *Third International Conference on Computer Vision*, 1990.

[22] Vincken K. L.; Koster A. S. E.; Viergever M. A., "Probabilistic multiscale image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, February 1997.

[23] Niessen W. J.; Vincken K. L.;Viergever M. A., "Comparison of multiscale representations for a linking-based image segmentation model," in *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*, June 1996, vol. 21-22, pp. 263–272.

[24] Vincken K. L.; Niessen W. J.;Viergever M. A., "Blurring strategies for image segmentation using a multiscale linking model," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings CVPR '96*, June 1996, vol. 18-20, pp. 21–26.

[25] Cai D.; Deangelis G. C.;Freeman R. D., "Spatiotemporal receptive field organization in lateral geniculate nucleus of cats and kittens," *Journal of Neurophysiology*, vol. 78, no. 2, August 1997.

[26] Mallat S., *A Wavelet Tour of Signal Processing*, Academic Press, 1998.