

Multiresolution Segmentation of Natural Images: From Linear to Non-Linear Scale-Space Representations

Ana Petrovic *Student Member, IEEE*, Oscar Divorra Escoda, *Student Member, IEEE*
and Pierre Vanderghyest, *Member, IEEE*

EDICS #'s : 2-SEGM

The authors are with the Signal Processing Institute, Swiss Federal Institute of Technology, Lausanne, Switzerland.
Email :{ana.petrovic, oscar.divorra, pierre.vanderghyest}@epfl.ch

Abstract

In this paper, we introduce a framework that merges classical ideas borrowed from scale-space and multi-resolution segmentation with non-linear partial differential equations. A non-linear scale-space stack is constructed by means of an appropriate diffusion equation. This stack is analyzed and a tree of coherent segments is constructed based on relationships between different scale layers. Pruning this tree proves to be a very efficient tool for unsupervised segmentation of different classes of images (e.g. natural, medical ...). This technique is light on the computational point of view and can be extended to non-scalar data in a straightforward manner.

Index Terms

Unsupervised Segmentation, Visual front-end, Multi-resolution, Linear Scale-Space, Non-Linear Scale-Space, PDE, Diffusion, Edge Detection, Image Structure, Hierarchical Trees.

I. INTRODUCTION

Segmentation is a classical problem in computer vision that can roughly be described as the process of partitioning an image into a set of non-overlapping homogeneous regions Ω_i :

$$\bigcup_{i=1}^n \Omega_i = \Omega, \quad \Omega_i \cap \Omega_j = \emptyset \text{ if } i \neq j,$$

where Ω is the image domain.

Many criterions might be given for defining such regions (grayscale levels or color homogeneity, texture coherence ...), each of which leads to different solutions. Most of the time, feature vectors are extracted by means of classical image processing techniques (histograms, Gabor filtering [1], wavelet coefficients analysis [2], ...) and clustering or classification algorithms are then used for obtaining set representatives. Another class of approaches consists in tracking the border of objects through edge detection techniques either based on filtering or active contours [3].

Although there exist a wide variety of methods more or less suited for particular applications, a very strong interest has recently emerged for a general framework based on partial differential equations (PDEs). One of the earliest formulations is due to Mumford and Shah [4]. In this work, the optimal image partition is defined as the solution of a variational problem that tends to simplify the original image, while preserving the borders of original regions. The idea of minimizing a given energy functional later on found many encouraging applications in image enhancement and denoising. A parallel between general non-linear diffusion PDEs and Scale-Space theory [5], [6], [7] can easily be made since the latter is generated by applying the linear heat flow equation to the image. Based on this, Alvarez et al. [8] established a rigorous

connection between both theories and paved the way to studying and using a wide variety of generalized, non-linear, scale-spaces. Multi-resolution stands as a fairly reasonable paradigm for image segmentation and has been used under several instances (see [9], [2], [10] for example). It allows for a progressive refinement of the partition and is very well adapted to processing inspired by the Human Visual System (HVS). Evidences have been found that the HVS performs some structure analysis on the incoming visual data [11], [6]. The structure of images has a close relation with multi-scale representation [11]. One of the clearest examples of multi-scale (or multi-resolution) data representation is Scale-Space [5]. Such a representation is composed by the stack of successive versions of the original data set at coarser scales. It is assumed that, the bigger the scale, the less information referred to local characteristics of the input data will appear. 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 structures.

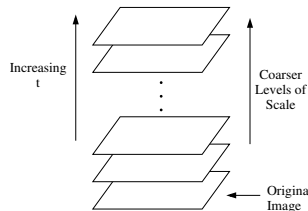


Fig. 1. Scale-Space stack.

Scale-spaces can be generated on the basis of many different principles. It is just necessary to be able to obtain a description of 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 and general comparisons are available in [12], [13], [14], [8]. Scale-spaces may be classified in two main groups: *Linear Scale-Space* and *Non-Linear Scale-Spaces*. The former is derived from the linear heat equation. *Koenderink* [6] was one of the first to point out its relation with structure analysis due to its multi-resolution nature. The main property of linear scale-space is the diffusion independence with respect to the signal. *Non-Linear Scale-Spaces* relax the constraint of uncommitment in the processing of visual information, keeping the main properties of a scale-space. They may have properties like *luminance conservation*, e.g. *gradient dependent diffusion* [15] or *tensor dependent diffusion* [16]. They may rely on *geometry* to influence the evolution of surfaces or curves [17], or consider morphology based on the erosion or dilation of images [18], etc. The choice of one or other principle to obtain the derived set of images at different scales is a matter of the particular application.

The outline of the paper is as follows. Section II reviews basic properties of linear scale-space and introduces a segmentation algorithm based on building a hierarchical structure on the scale-space stack. We show that carefully monitoring edges through scale gives more precise results. Section III extends previous results to the case of non-linear scale-spaces. The basic idea is to embed the edge monitoring step inside the stack generation and track non-linear flow information through scale. Section IV presents several results using our segmentation algorithm as well as some comparisons with other known techniques. An extension of the approach to color images is proposed in section V. Finally, in Section VI we expose our conclusions.

II. LINEAR SCALE-SPACE SEGMENTATION

When no a priori assumptions about the signal to analyze are available, the best is to stay on the basis of an uncommitted visual front-end [6], [7] where properties like linearity, spatial shift invariance, isotropy and scale invariance, are kept. Such a set of properties is satisfied by the Linear Scale-Space [7], [19].

A. Linear Scale-Space

A linear scale-space representation is constructed by embedding an image $I(\vec{x})$, $\vec{x} = (x, y)$, into a one parameter family of images $I(\vec{x}, t)$ that satisfy the diffusion equation [6]:

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

with initial condition $I(\vec{x}, t = 0) = I(\vec{x})$ and the scale parameter (t) being monotonically ascending. It has been shown that $I(\vec{x}, t)$ can be obtained as well by means of convolution. Since the unique kernel satisfying (1) is the Gaussian function [5], [6], [7], then,

$$I(\vec{x}, t) = \int_D G(\vec{x} - \vec{x}', t) \cdot I(\vec{x}', 0) d\vec{x}', \quad s.t. \quad G(\vec{x}, t) = \frac{1}{4\pi t} e^{\left(-\frac{\|\vec{x}\|^2}{4t}\right)}. \quad (2)$$

Hence, the problem of generating a signal representation for an uncommitted visual front-end turns into a simple set of successive blurrings.

An important additional result depicted from Equation (1) is that derivatives of $I(\vec{x}, t)$ also satisfy the heat flow equation, and so any Gaussian derivative generates also a scale-space [19], [20]:

$$\frac{\partial}{\partial \vec{x}} (I * G) = I * \frac{\partial G}{\partial \vec{x}}, \quad (3)$$

where $*$ stands for convolution. This is of special interest for multi-scale differential analysis, e.g. analysis of edges at different scales [21], [11], [22].

A detail to consider is that scale-space theory holds in the continuous domain [7]. Its use requires a careful discrete approximation that ensures a maximum fidelity, as discussed in [23].

B. Segmentation Process: Linking Up Through Scale

The segmentation algorithm is based on the analysis of the image structure from the scale-space stack. In this work we consider as starting point the approach introduced by *Vincken* [12], [24], [25] based on the tracking through scale of the iso-intensity paths generated by the diffusion eqs. (1) and (2). Other approaches relying on extrema [26], [27], [19], [28], [29] are proposed in the literature. However we considered it more general not to be limited to extrema since they do not fully describe all image pixels.

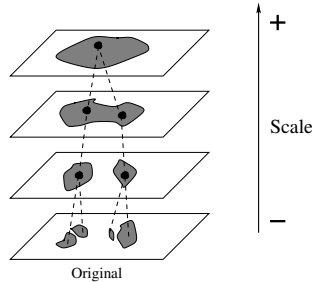


Fig. 2. Hierarchical analysis of the image structure linking pixels through levels.

The algorithm sets up a tree-like structure establishing relations between pixels of consecutive levels (Fig. 2). Given a level n , only pixels having a link from level $n - 1$ are related to a parent pixel at the next level $n + 1$. Children pixels are connected to a parent pixel by means of a search area (Fig. 3) centered in each one of them with radius proportional to the level scale $r_{n,n+1} \propto (\sigma_{n+1}^2 - \sigma_n^2)$ s.t. $t = \frac{\sigma_n^2}{2}$ in Equation (2) [12]. All pixels located within this area are potential parent pixels. The selection is mainly driven by the minimum intensity gradient (see [30] for additional criteria).

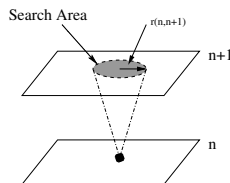


Fig. 3. Search area for a parent pixel.

The tree-like structure comes from the convergence of links when scale increases. This grouping of the iso-intensity paths is due to the progressive blurring of I . At the top of the tree (biggest scale), only a node (parent pixel) will remain.

A segment is defined as the image region that is fully linked to a node of the hierarchical tree. The selection of a set of root nodes will determine the final set of segments. This selection can be done in several ways, the simplest is to choose all the nodes belonging to a given scale level. To obtain more

precise results, a finer analysis of the tree is required. Possible approaches that look for more complex seeding rules can be found in [12], [10].

In a complete visual system, scale-space analysis is intended to be used as a visual front-end. Thus, it seems reasonable to think that node selection should be driven by the feedback from a high abstraction analysis layer, which should interpret structures obtained from the scale-space.

C. Supervising Segmentation with Multi-Scale Edges

The linking procedure [12], [25], [9], [24] from level to level simply looks for the nearest most suitable pixel in the circular search area. Linking is performed independently of the region shape where child and parent pixels belong. This uncontrolled linking allows pixels to be connected outside the appropriate region. In a very reduced neighborhood, wrong linking is supposed to rarely occur. Mistakes are possible when the search window is large and children pixels may find better fits for their gray level some distance away from the ideal parent. These small mistakes may turn into the divergence of a whole branch of the tree.

1) *Edges Through Scale:* According to Sec. II-A, multi-scale edge analysis can profit from the scale-space structure of diffusions generated by means of Gaussian derivatives. Spatial zero crossing detection, based on a scale-space generated by the second derivative of a Gaussian will supply the set of edges related to that scale. These will be the region or blob contours of the non-derivative scale-space that satisfies (1) (Gaussian filtering). For the sake of simplicity the Laplacian of a Gaussian may be approximated by the Difference of Gaussians (*DOG*).

This multi-scale second derivative representation, is nothing else than a wavelet representation of the image using the *Mexican Hat* wavelet [22] and is considered to be also an analogy with the HVS [11], [22].



Fig. 4. Edge representation through scale using DOG (levels 1,4 and 7) and zero crossing detection. 1 sample per 3 octaves (first sample on the first octave)

2) *Use of Edges in the Linking Process:* To reduce the divergence of the estimated iso-intensity paths, we propose the use of edge information to supervise the linking process. The search area is modified such

that parent and child pixels must belong to the same region or blob (Fig. 5). Thus, only parent candidates within the intersection of the circular search area and the projection of the child region are considered. All those link candidates that cross an edge at the scale of analysis (i.e. those not respecting the signal structure) will be dismissed.

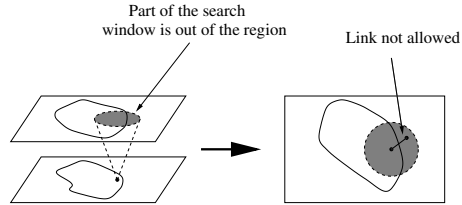


Fig. 5. Wrong linkage problem.

D. Segmentation Results

All experiments have been performed on natural images. In the following the influence of two main factors are illustrated and discussed.

1) *Edge Supervision Influence:* Edge detection is introduced in order to avoid incorrect linking between different regions. The effect of this correction is presented in Fig. 6. Both segmentations are computed using the same parameters and the same scale level. The difference is the edge based supervision on the right image in addition to the simpler affinity criteria used in [24] (left image). An improvement can be clearly seen on the marked details of the head and a little box on the wall. On the left figure we see how part of the head is merged to the body and that the little box is not detected. On the right figure, the wrong linking of the head is avoided, improving the definition of the contours. Furthermore, the box on the wall is kept and not wrongly merged.



Fig. 6. Comparison of the effect of edge detection on the segmentation. Segmentation of the original image Sergi (top). Level of segmentation: $\sigma_n = 25$ pixels (left: not using edges, right: using edges).

2) *Scale Selection:* The selection of a stopping scale plays a fundamental role since it defines the roots of the hierarchical tree that correspond to the segmented objects. The stopping scale implicitly defines

the level of details in the segmentation result (Fig. 7) and roughly the size of the objects. Both results from the figure (left and right) correspond to the same scale-space and tree analysis, only the stopping scale changes.

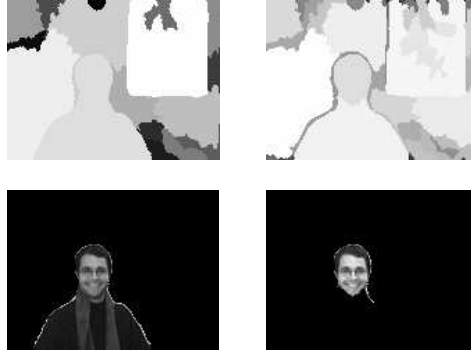


Fig. 7. Meaningful objects (down) obtained using the Scale-Space segmentation (regions up). Level of segmentation are left: $\sigma_n = 30$ pixels, right: $\sigma_n = 25$ pixels.

III. NON-LINEAR SCALE-SPACE: A FRAMEWORK FOR HANDLING IMAGE STRUCTURE

As shown in the previous section with the help of numerical simulations, locally supervising image edges improves segmentation results. This is easily explained in the settings of our algorithm by the fact that we don't link pixels that belong to *different* structures through scales. A simple way of achieving the same task in an unsupervised manner would be to use a nonlinear scale-space in which coherent structures are preserved by the flow. According to the HVS, edges are very important primitives in natural images. We emphasize that they should be conserved in order to avoid wrong linking and this paves the way to using non-linear diffusion as a natural scale-space candidate.

A. Non-Linear Scale-Spaces

In what follows, we will give an overview of three non-linear PDEs models.

1) *Anisotropic Diffusion Equation:* In section II-A we gave a description of the linear scale-space. An important improvement of this classical linear multi-scale analysis, with a more accurate edge detection, was proposed by Perona and Malik [15]. The main idea was to introduce a part of the edge detection step in a partial differential equation model, encouraging smoothing within homogeneous regions in preference to smoothing across the boundaries. Blurring would then take place separately in each region, letting region boundaries remain sharp. They proposed to replace the heat equation (1) by the following nonlinear diffusion equation:

$$\frac{\partial I}{\partial t} = \operatorname{div} [g(\|\nabla I\|) \nabla I], \quad (4)$$

where $\|\nabla I\|$ is the gradient magnitude of the image I and $g(\|\nabla I\|)$ is an edge stopping function chosen to satisfy $g(0) = 1$, $g(x) \geq 0$ and $g(x) \rightarrow 0$ when $x \rightarrow \infty$. The choice of g corresponds to a sort of thresholding which can be compared to the thresholding of $\|\nabla I\|$ in edge detection methods (e.g. [31]). This has a cost: the model contains a "contrast threshold" which can only be fixed manually. But the experimental results obtained by Perona and Malik are perceptually impressive.

Let us denote by (ξ, η) the local orthogonal coordinate system (Fig. 8), where the η -axis indicates the direction parallel to the gradient and the ξ -axis is orthogonal to the gradient direction:

$$\hat{\xi} = \frac{(-I_y, I_x)}{\|\nabla I\|}, \quad \hat{\eta} = \frac{(I_x, I_y)}{\|\nabla I\|}. \quad (5)$$

Now, we can rewrite the anisotropic diffusion (4) with respect to this new coordinates:

$$\frac{\partial I}{\partial t} = g(\|\nabla I\|) I_{\xi\xi} + (\|\nabla I\| g(\|\nabla I\|))' I_{\eta\eta}. \quad (6)$$

This new form shows that (4) does not diffuse only in the direction orthogonal to the gradient, but also in the direction of the gradient. If the diffusion coefficient in the orthogonal direction to the edge is negative, the flow behaves like the reverse heat equation [14] and there is no unique solution and no stability.

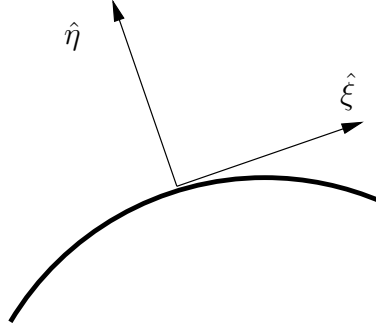


Fig. 8. Local coordinate system

2) *Mean Curvature Motion:* Defining g as $g(s) = \frac{1}{s}$ in (4) and multiplying the whole expression by $\|\nabla I\|$ we get the mean curvature equation:

$$\frac{\partial I}{\partial t} = \|\nabla I\| \operatorname{div} \left(\frac{\nabla I}{\|\nabla I\|} \right). \quad (7)$$

This equation with respect to local coordinates (5) becomes:

$$\frac{\partial I}{\partial t} = I_{\xi\xi}. \quad (8)$$

This means that mean curvature motion diffuses only in the direction orthogonal to the gradient. Therefore, in the ideal case, the exact location and sharpness of the edge is preserved while the picture is smoothed on both sides of the edge.

Another possible way to rewrite (7) is:

$$\frac{\partial I}{\partial t} = k \|\nabla I\|, \quad (9)$$

where k is the curvature of the level sets of I . This equation shows that the level sets C of the solution move in the normal direction \vec{N} with a speed proportional to their curvature,

$$\frac{\partial C}{\partial t} = k \vec{N}. \quad (10)$$

Directional diffusion is then equivalent to smoothing each one of the level sets according to the geometric heat flow. Fig. 9 shows a curve evolution under the Euclidean geometric heat flow. We can see that the curve eventually shrinks to a point. On Fig. 10 the image is evolving under mean curvature motion. Note



Fig. 9. Steps of the Euclidean geometric heat flow.

again how the level lines tend to shrink to a point.

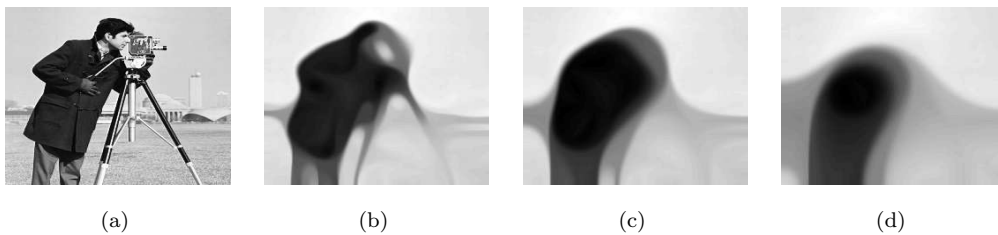


Fig. 10. Mean curvature flow at different scale levels S , $dt=0.025$. (a) $S = 2$; (b) $S = 4096$; (c) $S = 16384$; (d) $S = 65536$.

3) *TV Flow*: The idea of minimizing the total variation norm:

$$TV[I] = \int_{\Omega} \|\nabla I\| dx, \quad (11)$$

for image processing purposes was first introduced by Osher and Rudin [32]. The main property of this functional is that it can be defined for functions that have discontinuities, leading to the correct edge representation. Thus, the TV norm does not penalize discontinuities in I and allows one to recover edges

of the original image. The minimization of the Total Variation (11) can be achieved by a gradient descent method which yields the following evolution equation:

$$\frac{\partial I}{\partial t} = \operatorname{div} \left(\frac{\nabla I}{\|\nabla I\|} \right) = k. \quad (12)$$

We may rewrite (12) with respect to local coordinates:

$$\frac{\partial I}{\partial t} = \frac{1}{\|\nabla I\|} I_{\xi\xi}. \quad (13)$$

Written in this way, the method appears as a diffusion in the direction orthogonal to the gradient, tuned by the magnitude of the gradient. The main advantage of this evolution is edge enhancement, longer edge conservation and intra-region smoothing, as can be seen in Fig. 11.

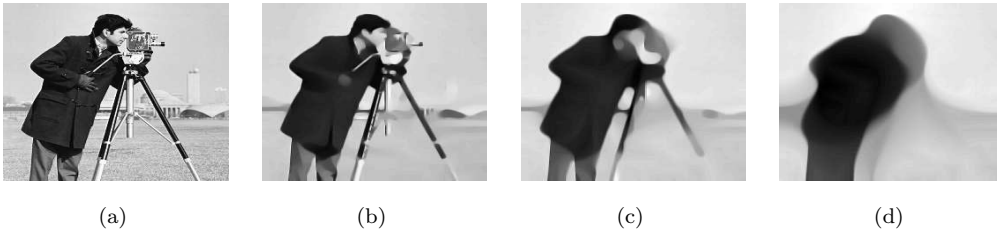


Fig. 11. TV flow at different scale levels S , $dt=0.025$. (a) $S = 2$; (b) $S = 4096$; (c) $S = 16384$; (d) $S = 65536$.

B. TV flow: a Good Framework for Segmentation

Numerous non-linear PDEs exist, each of them having their own properties. We could look at this multi-scale image representation as a set of similar images that contains more information than the original image alone. This decomposition could be a good framework for many high-level image processing tasks. However, more work is required in order to understand what kind of information is present in the decomposition, and how a high level process could take advantage of it.

Akin to the section II-B, it seems natural that the segmentation process takes advantage of an image decomposition in which only the most important components are preserved through scale. Due to the importance of preserving edges, we have presented three different non-linear flows. We showed that the Perona and Malik equation (4) may give an unstable solution, inconvenient for further examination. The only difference between the mean curvature flow (7) and TV flow (12) is the scaling parameter $\|\nabla I\|$ of the Equation (7). Therefore, in both cases diffusion goes in the direction orthogonal to the gradient but TV diffusion is inhibited in regions where the gradient is high. This turns into a longer edge preservation by TV flow than by the mean curvature flow (see Fig. 12 and Figs. 10 and 11). It is obvious that TV

flow preserves the image coherent structures much better than the curvature flow, and this leads us to the conclusion that it is more appropriate for image segmentation tasks.

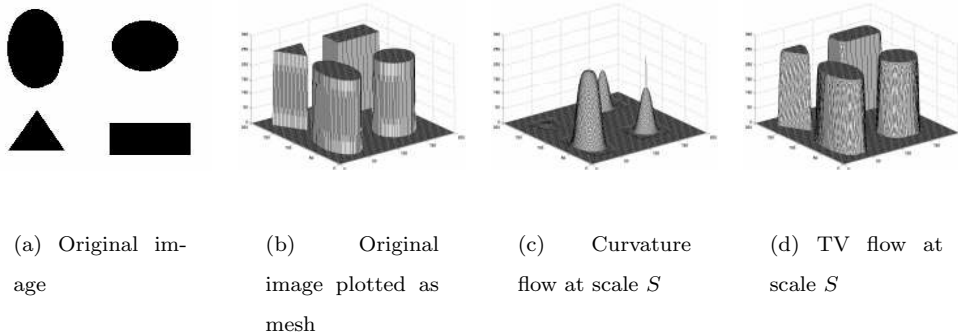


Fig. 12. Comparison of TV and mean curvature flow.

C. Image Representation

Following to multi-scale TV flow representation of the original image, the next step is to find an appropriate image representation at each scale. This image representation should be well adapted to the segmentation problem, that is, a decomposition of the image into homogeneous regions separated by boundaries or "edges". If we define isolevel sets with level λ of an image I as the sets of pixels satisfying:

$$\chi_\lambda(I) \equiv \{x, I(x) = \lambda\}, \quad (14)$$

then edges are borders of isolevel sets called *level lines*. As stressed by Mathematical Morphology (Matheron [33], Serra [34]), the main shape information is contained in the image level sets. Moreover, isolevel sets are contrast invariant in the sense

$$\chi_{g(\lambda)}(g(I)) = \chi_\lambda(I), \quad (15)$$

for any continuous increasing contrast change g .

Finally, Kronkord [35] showed that if the I is a continuous function, then the isolevel sets χ_λ are nested, making an inclusion tree. This tree relies on the geometric properties of the image without affecting contrast information. In this way, isolevel sets provide a complete, contrast invariant and parameter free representation of an image and can be a good starting point for clever non local grouping algorithms. Therefore, we choose our basic objects or regions at each scale as connected components of isolevel sets.

D. Segmentation Algorithm

A hierarchical decomposition of the non-linear Scale-Space can be seen as a tree where nodes correspond to picture regions (isolevel sets) and where links between nodes indicate a set inclusion (Fig. 2). At each scale we perform a partition of the image into isolevel sets and try to link them with the "closest" one in the next scale. Regions from two subsequent scales can be linked only if they overlap and if the intensity difference is the smallest among all the overlapping regions. In this way, the segmentation is obtained by grouping regions of the initial image that are connected to the same region at some stopping scale S . In what follows, we set up the terminology and the construction procedure for the proposed hierarchical algorithm.

Let Ω be an open subset of \mathbb{R}^2 and I an original grayscale image treated as a function defined on Ω . From I we construct a sequence of images $I^0 = I, I^1, I^2, \dots, I^S$ according to the numerical implementation of the TV flow using an explicit Euler method in time:

$$\begin{cases} I^k = I^{k-1} + \Delta t \frac{I_{xx}I_y^2 - 2I_{xy}I_xI_y + I_{yy}I_x^2}{(I_x^2 + I_y^2)^{3/2}}, & k \geq 1 \\ I^k = I, & k = 0. \end{cases} \quad (16)$$

Gradients are approximated by the central difference in space. We use the terminology I^k to denote the image at scale k .

The parameter S represents the stopping scale of the hierarchical tree. The choice of this parameter controls mostly the presence of details in the segmentation result. The best scale to stop the evolution of the image I is mainly determined by the number of meaningful objects that one wishes to extract. Increasing the parameter S reduces the final number of segmented objects. In practice, it is possible to tune this parameter according to the image content.

Once the sequence $\{I^k\}_{k=0}^S$ has been obtained, we can define on each of the image domains $\{\Omega^k\}_{k=0}^S$ a set of partitions corresponding to the connected component of isolevel sets. Therefore, at each scale k we have partitions $\{\Omega_0^k, \Omega_1^k, \dots, \Omega_{P(k)}^k\}$ such that

$$\bigcup_{i=0}^{P(k)} \Omega_i^k = \Omega^k, \quad \Omega_i^k \cap \Omega_j^k = \{0\}, \quad \forall i \in \{0, \dots, P(k)\} \quad I^k(\Omega_i^k) = \text{const},$$

where $P(k)$ defines the total number of regions at scale k .

To construct the hierarchical tree (Fig. 13), we have to define a linking operator L to connect two regions of neighboring scales. As TV flow preserves coherent structures, the gray level of an object just slightly changes on consecutive levels. Hence, the linking criteria will be the minimal gray level distance

among all overlapping regions. This rule defines the following class C_i^k of admissible linking regions for Ω_i^k :

$$C_i^k = \{\Omega_j^{k+1} \mid 0 \leq j \leq P(k+1), \Omega_j^{k+1} \cap \Omega_i^k \neq \{0\}\}.$$

We can then define the linking operator as

$$L(\Omega_i^k) = \operatorname{argmin}_{\Omega_j^{k+1} \in C_i^k} [abs [I^{k+1}(\Omega_j^{k+1}) - I^k(\Omega_i^k)]] .$$

In order to give a more formal definition of the segmentation, we first define the global operator $\Lambda(\Omega_i^{k1}, k2)$, $k1 \leq k2$. The result of this operator will be the region at level $k2$ that is linked with Ω_i^{k1} in the tree. Taking into account the construction of the hierarchical tree, the operator Λ is injective. Notice also that recursive application of operator L from scale $k1$ to $k2$ leads to Λ (Fig. 13).

$$\begin{aligned} \Lambda(\Omega_i^{k1}, k2) &= \Lambda(L(\Omega_i^{k1}), k2), & k1 < k2 \\ \Lambda(\Omega_i^{k1}, k2) &= \Omega_i^{k1}, & k1 = k2 \end{aligned}$$

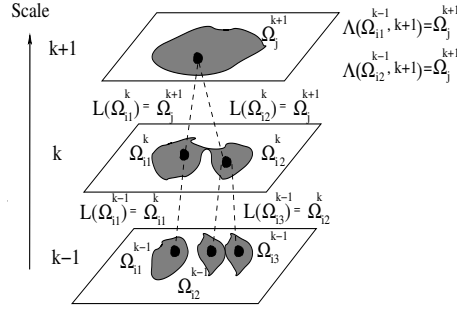


Fig. 13. Hierarchical tree and linking operators L and Λ .

If at scale S we decide to stop the hierarchical tree, the segmentation can be induced according to the following simple rule: all connected regions of the initial scale that are connected to the same region on the stopping scale S represent one object. Denote by O_i^S , $0 \leq i \leq P(S)$ the induced segmented objects associated respectively to the regions Ω_i^S , $0 \leq i \leq P(S)$. Then segmented objects may be written as:

$$O_i^S = \left\{ \bigcup_{j=0}^{P(0)} \Omega_j^0 \mid \Lambda(\Omega_j^0, S) = \Omega_i^S \right\}.$$

If $O_i^S = \{0\}$ the object is empty and the real number of objects is:

$$| \{O_i^S \mid 0 \leq i \leq P(S), O_i^S \neq \{0\}\} |$$

This is the final stage and the segmented objects are extracted. The previous procedure can be given in the iterative form:

Initialization $k = 0$

- 1) Calculate an image I_k at scale k using equation (16).
- 2) Split the image at scale k into connected components of isolevel sets $\{\Omega_0^k, \Omega_1^k, \dots, \Omega_{P(k)}^k\}$.
- 3) If $k > 0$, perform linking between regions of two subsequent scales $k - 1$ and k . $\forall i \in \{0, \dots, P(k)\}$ set up connection between Ω_i^k and $L(\Omega_i^k)$.
- 4) $k = k + 1$.
- 5) If $k \neq S$
 - go to step 1
 - else
 - calculate O_i^S , $0 \leq i \leq P(S)$

Using this segmentation algorithm, for any given data set, it is always possible to construct a hierarchical tree and induce the segmentation. The presented algorithm combines region growing method and variational-based approach for image segmentation. Indeed, if we look carefully, the growing of regions is defined by a linking strategy and this is encapsulated in a variational framework (minimization of the TV norm). Hence, segmentation is achieved by grouping regions of the original image according to the minimization of the TV norm. Moreover, a *multiscale analysis* is performed and the segmentation results are given according to the stopping scale S . For different values of S , different segmentation results are obtained, corresponding to different levels of resolution. This problem of multiscale segmentation was raised by Morel and Solimni in their work on Variational Methods in Image Segmentation [36]. They have stated that most segmentation methods can be interpreted as attempts to solve some variational formulation and that is naturally associated with multiscale theory.

IV. EXPERIMENTAL RESULTS

A. Performance of our Multi-Scale Segmentation Algorithm with Improved Edge Preservation

Fig. 14(a) shows an original bird image taken from literature [37] (see Fig.5 (c)). The next three columns present the results based on the main segmentation steps. Each row contains the images obtained at the same scale level. The first column (Figs. 14 (b), (e), (h)) contains images of the TV flow corresponding to different scales. The second column (Figs. 14 (c), (f), (i)) shows the segmentation results corresponding to the same scales as the first column. To clearly show the effect of the choice of scale in the segmentation, the last column (Figs. 14 (d), (g), (j)) presents the contours of the objects of each segmented image from

the second column. For different scale parameter, we can observe the obvious variation in segmentation results. On Figs. 14(c) and (d), the background is recognized as one object but the small structures of the bird are kept. At further scales (Figs. 14(f) and (g)), the structure of the bird is simplified, and the main objects are the bird's head, the bird's body and the small bush on the right bottom corner. On Figs. 14(i) and (j), the contour of the bird is detected as the only object. This segmentation stack is generated only once and gives possibility for different segmentation problems to choose the appropriate scale parameter. The results are given without any pre-processing or post-processing. The only implicit parameter in this model is the time step Δt in the TV flow discretization (16). Experimentally, we discovered that the choice of this parameter does not considerably affect the results and in all examples the value of Δt is fixed to 0.025.

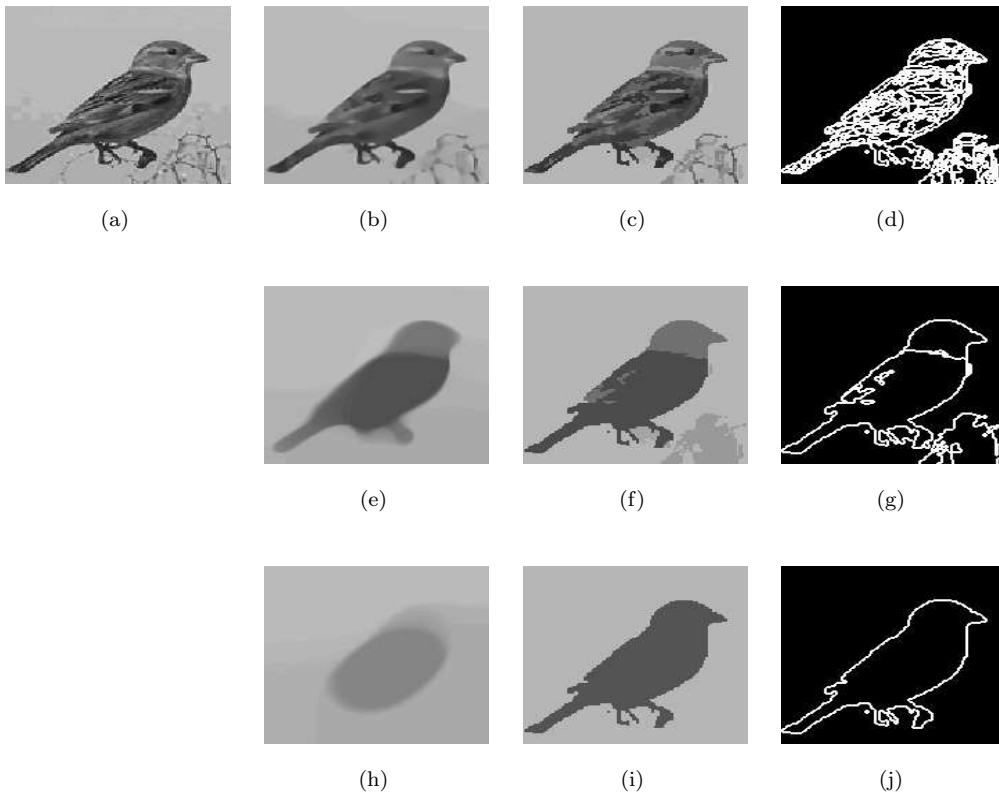


Fig. 14. Scale-space and segmentation stack. (a) Original bird image; (b),(e),(h) samples of scale-space corresponding to scale parameters 1536, 16384, 65536; (c),(f),(i) the segmentation results for image(a) corresponding to (b),(e),(h) respectively; (d),(g),(j) the borders of detected regions corresponding to (c),(f),(i) respectively.

The next example demonstrates how our model deals with noisy images (Fig. 15). Gaussian noise was added to the original bird picture (Fig. 14 (a)) such that the resulting PSNR is 24 dB. Although the

initial image is noisy, at the final step of the segmentation process (Fig.15(d)) the result is similar to the noiseless case (Fig.14(i)). The smoothing effect of the TV flow on the degraded image, makes the algorithm robust to some amount of noise.



Fig. 15. Segmentation results for a noisy image, obtained applying our algorithm with linking up to different levels S of scale-space. (a) Noisy bird image; (b) $S = 24$; (c) $S = 6144$; (d) $S = 65536$.

On Fig. 16, we performed an experiment on an MR image of the brain. It has to be noted that nothing was assumed concerning the nature of the image, the algorithm was run blindly and a meaningful segment corresponding to brain white matter was selected with the help of a physician. This results illustrates the potential of our technique when applied to diverse data.

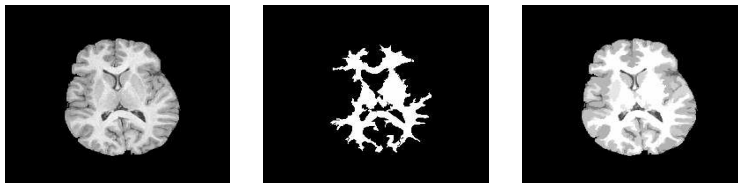


Fig. 16. Segmentation of a part the brain. Original MR image (left). Segmented region (middle). Superposition of the segment onto the original image (right).

B. Comparisons with Other Known Segmentation Techniques

In the following, we give some comparison between our method and some other known segmentation techniques. We first present three classical segmentation tools: thresholding, maximum likelihood clustering and region merging of gradient watershed regions. The results of those techniques are obtained by using the software package Kuim [38]. We introduce also two active contour methods and we finally present the results obtained by our approach.

Perhaps the simplest segmentation method is based on image thresholding [39]. This class of segmentation algorithms depends on a threshold selection. Figs. 17(b) and (e) show the segmentation results using the automatic threshold selection based on edge strengths [38]. It is evident that some additional

processing must be done in order to obtain a cleaner result. The technique captures the person and the picture in the top right corner, but is not able to cope with the nonuniform luminance of the background. This method is parameter independent, it is very fast and easy to implement but is not appropriately adjusted to capture the multi-scale structure of objects.

Another class of segmentation algorithms considers statistical image classification [39]. Maximum Likelihood Segmentation is a well-developed method from statistical decision theory and has been applied on Figs. 17(c) and (f). The drawback is the need to define a priori the number of output classes. Another problem that arises is that almost all types of data have a natural variability that cause class distribution to be mixed (i.e. structure is not necessarily respected). This produces artifacts on the background of the image. The method is based on grouping pixels according to the intensity distribution information and therefore it is not able to group nonuniform regions within one object (the head can not be detected as one object).

The next segmentation algorithm belongs to the classes of region merging techniques [39] and watershed segmentations [39]. This method performs image segmentation by merging adjacent gradient watershed regions. An iterative merging procedure is based on region "similarity", or difference in region means. Input parameters control how many internal iterations are performed and how similar regions must be in order to be merged. Segmentation results are presented on Figs. 17(d) and (g), and like the previous techniques, this method also suffers from over-segmentation problems with no control of the scale of the results.

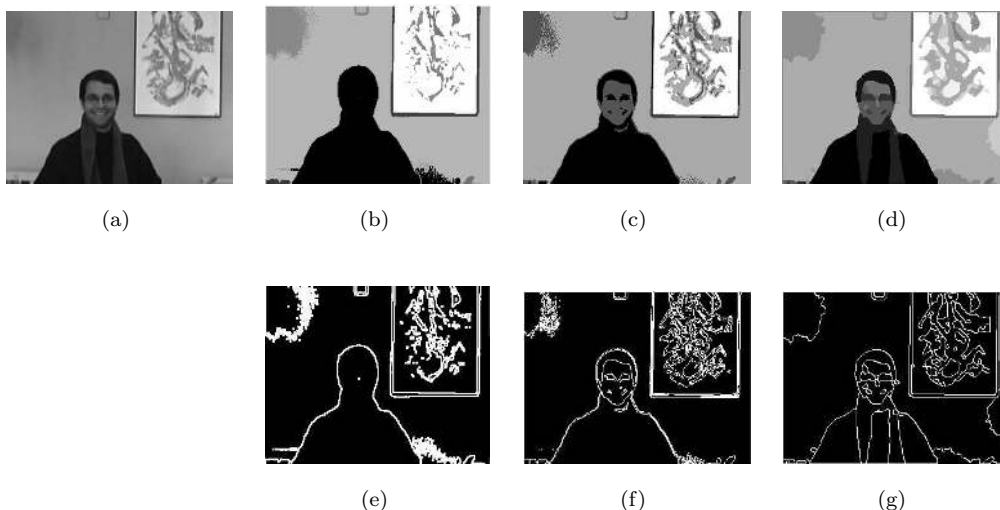


Fig. 17. Different segmentation techniques. (a) Original image. Top row: segmented images. Bottom row: region's borders. Intensity Thresholding: (b),(e) ; Maximum Likelihood Segmentation: (c),(f) ; Gradient Watershed Regions: (d),(g).

An extensive work was done on boundary detection via “snakes” or active contours initially proposed by Kass and al. [3], [40]. To be independent on the curve parametrization, a novel geodesic formulation for active contours was proposed by Caselles, Kimmel and Sapiro [41]. The approach is based on active contours evolving in time according to intrinsic curve properties and image measures concerning the object boundaries. This technique has a strong mathematical formulation regarding existence, uniqueness, stability, and consistence of the solution. When it is implemented using the level-set based numerical algorithm [17], this model allows automatic changes of topology and simultaneous detection of several objects without previous knowledge of their exact number. The weakness of this model is its dependence on the topology of the initial curve as shown in Figs. 18 (b) and (c), where different results are obtained using different initial conditions. Another possible problem of this approach is the need to estimate the parameters that determines the behavior of active contours.

Some of these weaknesses were overcome in the work of Vese and Chan [42]. Their model has some advantages: edge detection with or without the use of gradient, detection of interior contours, automatic change of topology, robustness with respect to noise, representation of complex topologies. However, in the piecewise constant case, the number of segments in the image has to be known in advance. In the piecewise-smooth case, only two level set functions can represent any number of segments of a given image. But then, the global minimizer is not unique in general, and the numerical results may depend on the initial curves and only a local minimum is computed. Finally, we show in Fig. 18 (d) the segmentation result using the four-phase model.

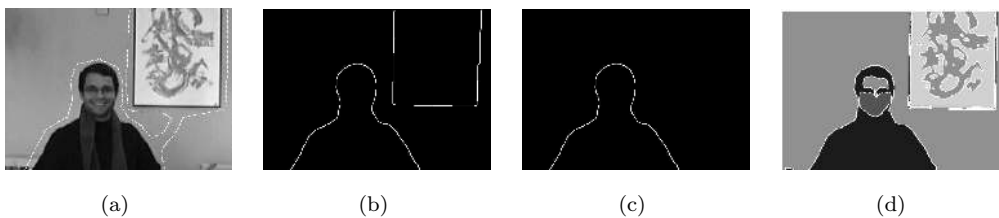


Fig. 18. Segmentation results obtained applying active contours. (a) Initial contour for geodesic active contour method; (b) Segmentation result using initial contour of image (a); (c) Segmentation result using different initial contours; (d) Segmentation results using 4-phase piecewise constant model based on the Mumford shah functional.

The results of our multiscale segmentation scheme are presented in Fig. 19. We apply the segmentation algorithm up to different scale levels S (Figs.19(a), (b), (c) and (d)). On Fig. 19 (a), we obtain many segments due to the non uniformity of objects but as we continue linking through the scale, those small objects are merged into a bigger one. Obviously the further we go through scales, the less details appear

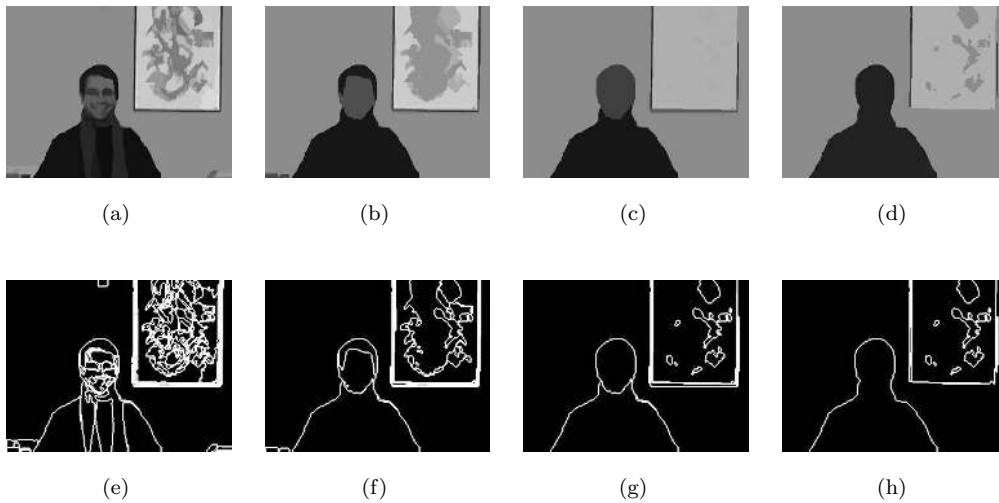


Fig. 19. Segmentation results obtained applying our algorithm up to different scale levels S . Top row: segmented images. Bottom row: region's borders. (a) $S = 3072$; (b) $S = 24576$; (c) $S = 49152$; (d) $S = 131072$.

on the segmentation results. For example, we can see that if we stop our algorithm at scale $S = 3072$, the hair, several face regions and the scarf are still recognized as independent objects (Fig. 19 (a)), but at $S = 24576$ the scarf is already merged with the body and represents one single object, the face turns into a segment and the painting appears as a clear segment on the picture background (Fig. 19 (b)). Eventually, if we perform linking up to $S = 131072$, the hair, the face and the body are merged into an object that represents the person (Fig. 19 (c)). We can conclude that sufficiently good results have been obtained for the extraction of “meaningful” objects at different scales.

V. EXTENSION TO COLOR IMAGES

The presented algorithm is very easily extended to color images. The first step is to implement an appropriate flow according to the main requirement of preserving edges during the blurring process. Our choice is the Beltrami flow for color images that satisfies the previous requirement. For details on the Beltrami flow, the interested reader is referred to [43].

The second step is to partition images into regions with respect to an appropriate segmentation criterion. This process should correspond to a partition of the image into isolevel sets. The basic idea is to group neighboring pixels which are “close enough” in the sense of color difference. As far as color difference is concerned, a very important issue is the choice of the metric as well as the choice of color space suitable for that metric. The most frequently used metric is simply the Euclidean distance in the space

$$\Delta^2(C_1, C_2) = (\Delta T_1)^2 + (\Delta T_2)^2 + (\Delta T_3)^2,$$

where ΔT_i , $i=1,2,3$ are the color component differences between the colors C_1 and C_2 . However, in order to have a subjective justification, equal increments of each component ΔT_i should correspond to the same perceptual difference of colors C_1 and C_2 . Hence, the equation $\Delta^2(C_1, C_2) = 1$ should perceptually represent a sphere with radius equal to the least noticeable color difference with center in (T_1, T_2, T_3) . Experimental results showed that the RGB color system is highly nonuniform, e.g. our visual system is less sensitive to the blue channel. Therefore, we use the perceptually uniform color system CIELUV (L^* , u^* , v^*) [44] for which the color difference is given as follows:

$$\Delta^2(C_1, C_2) = (\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2.$$

Once images from the scale-space stack are partitioned, the next step is to build the hierarchical tree by linking through the scale-space. We follow the principle described for gray images in Section III-D with the simple modification that for color images we consider color differences instead of gray level differences. Linking is also performed at each two subsequent levels, starting from the original image. Since the original image is getting more and more simplified through scale, the number of regions will decrease at each level. Consequently, the further we go on with the linking process the more regions of the original image are linked to the same region of the top level. Edges which are preserved at higher levels play an important role in correct tracking. They will prevent incorrect linking, preserving regions that naturally belong to some object from diverging. Fig. 20 shows results of the segmentation algorithm for a color image. Clearly there remains some open problems and parameters, like the appropriate choice of color space or color diffusion scheme.



Fig. 20. Original image and segmentation results for different choices of the highest level (this is a color image).

VI. CONCLUSIONS

Building on the ideas of [12], [29], [45], [46], we have introduced an unsupervised segmentation technique for a visual front-end based on studying the multi-scale structure of the image. This is accomplished by first generating a scale-space stack and then building a hierarchical tree of coherent components through scale. In this paper, we focused on scale-spaces induced by Partial Differential Equations. We showed that results based on the classical heat flow, or Gaussian scale-space, can be improved by monitoring edges through scale. Turning to non-linear diffusion equations, such as the TV flow, we showed that the associated scale-spaces are better suited because they *naturally preserve* the meaningful edge information while performing intra region simplification. Extension to non-scalar data is very natural and the simple case of multichannel (for example color) images has been covered. Ultimately, our technique might be seen as a way of tracking the information flow through scale. We validated the proposed model by various numerical results and comparisons to other known segmentation tools.

ACKNOWLEDGMENTS

Ana Petrovic would like to acknowledge the support of the Swiss National Science Foundation through grant number 2100-061647.00. Oscar Divorra Escoda is supported by grant 6044.1 KTS from the Swiss Federal Office for Education and Technology. The authors would also like to thank Xavier Bresson for fruitful discussions and his contribution to the results of active contour segmentation.

REFERENCES

- [1] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition*, vol. 12, pp. 1167–1186, 1991.
- [2] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 1549–1560, 1995.
- [3] M. Kass, A. Witkin, and D. Terzopoulos, "Active contour models," in *Proceedings of the First International Conference on Computer Vision*, London, GB, June 1987, pp. 259–268.
- [4] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Comm. Pure and Appl. Math.*, vol. 42, pp. 577–684, 1989.
- [5] A. Witkin, "Scale-space filtering," in *Proceedings of the Eight International Joint Conference on Artificial Intelligence*, Karlsruhe (Germany), 1983, pp. 1019–1022.
- [6] J. J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, pp. 363–370, 1984.
- [7] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.
- [8] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel, "Axioms and fundamental equations of image processing," *Archive Rat. Mech. Anal.*, vol. 123, no. 3, pp. 199–257, 1993.

- [9] J. W. Niessen, K. L. Vincken, and M. A. Viergever, "Comparison of multiscale representations for a linking-based image segmentation model," in *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*, vol. 21-22, June 1996, pp. 263–272.
- [10] F. Ziliani and B. Jensen, "Unsupervised segmentation using modified pyramidal linking approach," in *In Proceedings of ICIP*, Chicago, October 1998.
- [11] D. Marr, *Vision*. Freeman Publishers, 1982.
- [12] K. Vincken, "Probabilistic multi-scale image segmentation by the hyperstack," Ph.D. dissertation, Utrecht University, 1995.
- [13] B. M. ter Haar Romeny, "Introduction to scale-space theory: Multiscale geometric image analysis," Utrecht University, Tech. Rep., 1996.
- [14] F. Guichard, L. Moisan, and J.-M. Morel, "A review of p.d.e. models in image processing and image analysis," *Journal de Physique IV*, vol. 12, pp. 137–154, 2002.
- [15] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions Pattern Analysis and Machine Intelligence*, 1990.
- [16] J. Weikert, "Scale-space properties of nonlinear diffusion filtering with a diffusion tensor," Laboratory of Technomathematics, University of Kaiserslautern, Tech. Rep., 1994.
- [17] S. O. S and S. Sethian, "Fronts propagating with curvature dependent speed: Algorithms based on the hamilton-jacobi formalism," *Computational Physics*, 1988.
- [18] P. T. Jackway and M. Deriche, "Scale-space properties of multiscale morphological dilation-erosion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, 1996.
- [19] L. M. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and A. M. Viergever, "Linear scale-space," *Journal of Mathematical Imaging and Vision*, vol. 4, no. 4, pp. 325–251, 1994.
- [20] T. Lindeberg, "Scale-space: A framework for handling image structures at multiple scales," in *In Proc. CERN School of Computing*, The Netherlands, September 1996.
- [21] —, "Edge detection and ridge detection with automatic scale selection," *Int. J. of Computer Vision*, vol. 30, no. 2, 1998.
- [22] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1998.
- [23] T. Lindeberg, "Scale-space for discrete signals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, March 1990.
- [24] K. L. Vincken, W. J. Niessen, and M. A. Viergever, "Blurring strategies for image segmentation using a multiscale linking model," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings CVPR '96*, vol. 18-20, June 1996, pp. 21–26.
- [25] K. L. Vincken, A. S. E. Koster, and M. A. Viergever, "Probabilistic multiscale image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, February 1997.
- [26] R. D. Henkel, "Segmentation in scale-space," in *In Proceedings of the 6th International Conference on Computer Analysis of Images and Pattern (CAIP)*, Prague, 1995.
- [27] —, "Segmentation with synchronising neural oscillators," Zentrum für Kognitionswissenschaften, Universität auf Bremen, Tech. Rep., 1994.
- [28] L. M. Lifshitz and S. M. Pizer, "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.
- [29] T. Lindeberg and J. O. Eklundh, "Scale detection and region extraction from a scale-space primal sketch," in *Third International Conference on Computer Vision*, 1990.

- [30] A. S. E. Koster, K. L. Vincken, and M. A. Viergever, "Heuristic linking models in multi-scale image segmentation," *Computer Vision and Image Understanding*, vol. 65, no. 3, 1997.
- [31] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [32] L. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *Proc. IEEE Internat. Conf. Imag. Proc.* IEEE Press, Piscataway, NJ, 1994, pp. 31–35.
- [33] G. Matheron, *Random Sets and Integral Geometry*. John Wiley, 1975.
- [34] J. Serra, *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
- [35] A. S. Kronrod, "On functions of two variables," *Uspehi Math. Sci.*, vol. 5, no. 35, pp. 24–134, 1950.
- [36] J. M. Morel and S. Solimini, *Variational Methods in Image Processing*. Birkhauser, 1994.
- [37] C. S. Tong, Y. Zhang, and N. Zheng, "Variational-based image segmentation and its multiscale realizations," *International Conference on Inverse Problems.*, Jan 9-12 2002.
- [38] J. Gauch, "Kuim," *image processing software, available online at <http://www.ittc.ku.edu/jgauch/research/kuim/html/>*.
- [39] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 2nd ed. Brooks/Cole Publishing Company, 1998.
- [40] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Computer Vision.*, vol. 1, pp. 321–331, 1988.
- [41] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Computer Vision.*, vol. 22(1), pp. 61–79, 1997.
- [42] L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the mumford shah model," *Int. J. Computer Vision.*, vol. 50(3), pp. 271–293, 2002.
- [43] N. Sochen, R. Kimmel, and R. Malladi, "A general framework for low level vision," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 310–318, 1998.
- [44] D. H. Brainard, *The Science of Color*, 2nd ed. Washington D.C.: Optical Society of America, 2003, pp. 191–216.
- [45] A. Kuijper and L. Florack, "Hierarchical pre-segmentation without prior knowledge," in *Proceedings of the 8th International Conference on Computer Vision 2001*, Vancouver, Canada, July 9–12 2001, pp. 487–493.
- [46] O. Divorra Escoda, A. Petrovic, and P. Vanderghyest, "Segmentation of natural images using scale-space representations: a linear and a non-linear approach," in *EUSIPCO*, Toulouse, France, September 2002.