Multiscale Image Segmentation Using Active Contours

Laura Gui, Xavier Bresson and Jean-Philippe Thiran
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Institute
CH-1015 Lausanne, Switzerland
{Laura.Gui, Xavier.Bresson, JP.Thiran}@epfl.ch

Technical Report TR_ITS_ 2005.012

Abstract

We propose a new approach for image segmentation at different scales of observation, based on a multiscale image decomposition and on the active contour segmentation model. The proposed method consists of two steps. Firstly, a representation of a given image at multiple scales is derived, by means of a smoothing method which minimizes the weighted total variation norm of the image. This method allows the longtime preservation of edges and contrast with increasing scale, facilitating the detection of underlying structures. Secondly, image structures are extracted at each scale, using a level set formulation of active contours, minimizing the Mumford-Shah functional. Promising results of the proposed segmentation approach on natural images are reported.

1. Introduction

Scale is a fundamental concept in computer vision. Lindeberg emphasized this point in [8], observing that real-world objects are perceived differently depending on the scale of observation. Most natural images contain objects at different scales and these multiscale objects are often recursive, i.e. they contain substructures, which contain further substructures etc. In the context of a computer vision system, a high level image understanding process could dictate the scale of the details to be analyzed. Consequently, it is desirable for the lower level segmentation process to provide the detected structures corresponding to different scales. This is the kind of segmentation process that we target in our work.

In order to extract such multiscale objects from images, an essential concept is the *scale space* of an image, whose bases were set by the pioneering works of Iijima [5], Marr [10], Witkin [19] and Koenderink [6]. The main principle of scale spaces is to decrease the amount of information in an image, by smoothing/simplifying the objects within it gradually, resulting in a hierarchy of fine to coarse replicas of the original image. Mathematically speaking, scale spaces are hierarchical representations at a continuum of scales, embedding the original image $u_0 : \mathbb{R}^N \to \mathbb{R}$ into a family $u : \mathbb{R}^N \times [0, \infty) \to \mathbb{R}$ of gradually more simplified images.

This multiscale hierarchy can be used to extract/identify multiscale objects with a given segmentation method. The segmentation of multiscale objects was explored in many different works, such as [9], [7], [4], [18], [3], [11], [14] and [13]. Closest to our approach is the method developed by Petrovic and Vandergheynst in [14]. They derive a scale space of an image using the total variation flow and then segment homogeneous image regions at each scale by applying a region growing algorithm, which minimizes the piecewise-constant approximation of the Mumford-Shah functional.

In our approach, we keep the same two-step framework for multiscale segmentation: scale-space derivation, followed by structure extraction at each scale. However, we investigate new methods for each of the two steps. More precisely, we derive our scale-space using a new variational formulation, based on the minimization of the weighted total variation norm of the image [16], [1]. This method yields better results than the total variation flow used in [14], in that it better preserves the image geometry (edge positions) and contrast. For structure extraction, we use the multiphase level set framework developed by Vese and Chan in [17], which has shown very good results in the segmentation of real world images.

2. The Weighted Total Variation Scale-Space

The weighted total variation (TV) scale-space [16], [1] is obtained by an enhancement of the well-known total variation flow developed by Rudin, Osher and Fatemi [15]. The Rudin-Osher-Fatemi model is a powerful variational technique for image denoising, remarkable for the longtime preservation of image edges with increasing scales of denoising. It is based on the minimization of the following energy:

$$E_{TV}(u,\lambda) = \int_\Omega |\nabla u|\,dx + \lambda \int_\Omega (u-u_0)^2\,dx, \qquad (1)$$

where $\Omega \subset \mathbb{R}^N$ designates the signal (image) domain, u_0 is the given image, u is its regularized approximation and λ is a positive parameter, dictating the *scale of observation* of the solution.

Weighted TV flow with an L^1-norm data fidelity term [1] is obtained by the minimization of the modified energy

$$E_{gTV}(u,\lambda) = \int_\Omega g\,|\nabla u|\,dx + \lambda \int_\Omega |u-u_0|\,dx, \qquad (2)$$

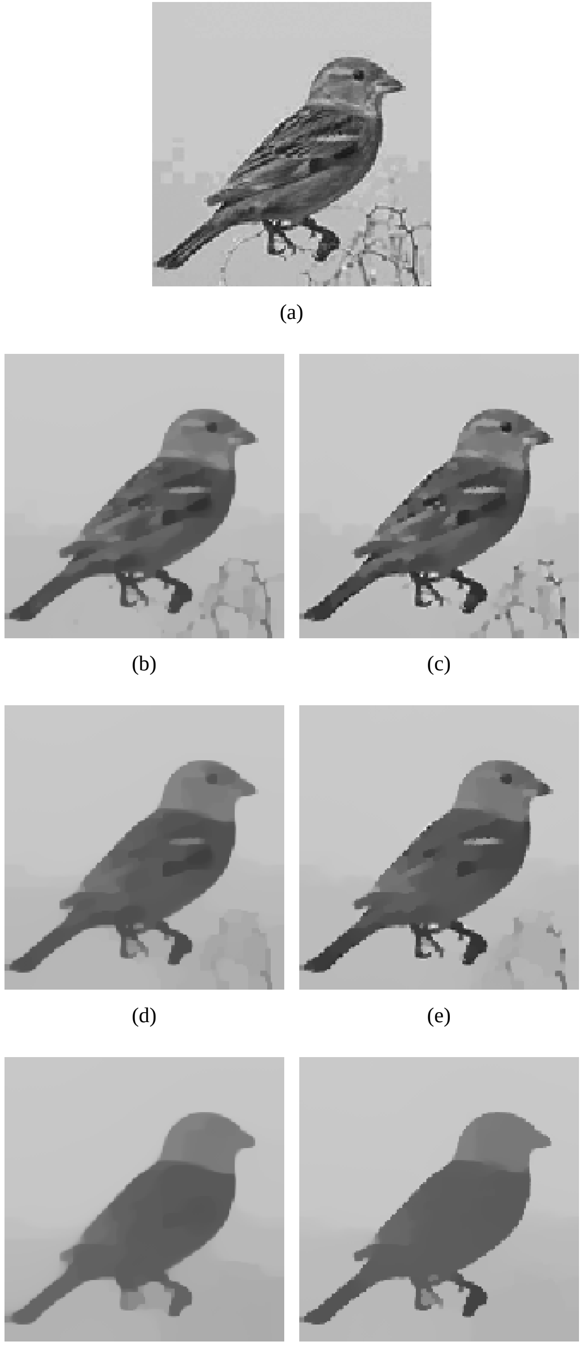
where $g = g(|\nabla u_0 * G_\sigma|)$ is an edge detecting function, dependent on the gradient magnitude of the gaussian smoothed original image (G_σ is the Gaussian kernel), non-increasing, with $g(0) = 1$, $g(s) \geq 0$ and $\lim_{s\to\infty} g(s) = 0$. For instance $g(s) = 1/(1+\beta\,s^2)$, with $\beta > 0$ an arbitrary parameter. The minimization of E_{gTV} in (2), using the Euler-Lagrange equations and the gradient descent method, yields the following evolution equation for u:

$$u_t = \nabla \cdot \left(g \frac{\nabla u}{|\nabla u|} \right) + \lambda \frac{u-u_0}{|u-u_0|}. \qquad (3)$$

As compared to the TV flow, the introduction of the weighting function g further inhibits diffusion at edge locations, contributing to edge preservation through scales. Moreover, the replacement of the L^2-norm with the L^1-norm as a fidelity measure helps to preserve the contrast of the image, as shown in [2] and confirmed by our experimental results. Figure 1 illustrates the advantages of weighted TV flow with respect to TV flow: observe the quality of the two final images 1(f) and 1(g), where the bird's eye is no longer discernible.

3. Active Contours Driven by the Mumford-Shah Functional

In [12], Mumford and Shah defined the image segmentation problem as follows. Given an observed image $u_0 : \Omega \to \mathbb{R}$, find a decomposition of the image domain Ω into connected components Ω_i and an optimal piecewise smooth approximation u of u_0, so that u varies smoothly within each Ω_i and rapidly/discontinuously across the boundaries of Ω_i. For 2D images, this translates into the following minimization problem: find (u, C) that minimize the energy func-



(a) (b) (c) (d) (e) (f) (g)

Figure 1. TV flow versus weighted TV flow. (a) Original image u_0; (b), (d), (f) TV flow, scales $\lambda = 0.04, 0.02, 0.008$; (c), (e), (g) weighted TV flow, scales $\lambda = 0.4, 0.2, 0.08$. Weighted TV flow better preserves edges and contrast of the original image through scales.

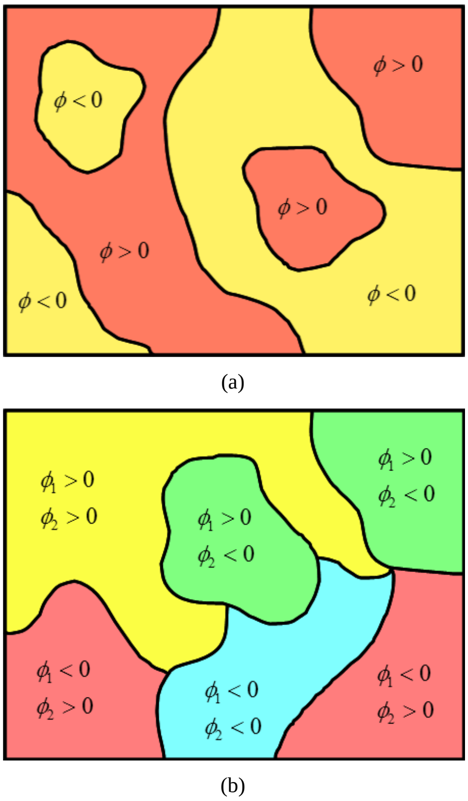


Figure 2. a) Image partition representable using one single level set function ϕ. b) Image partition containing triple junctions, representable using two level set functions ϕ_1 and ϕ_2.

tional

$$E_{MS}(u, C) = \int_{\Omega} (u - u_0)^2 dxdy + \mu \int_{\Omega \setminus C} |\nabla u|^2 dxdy + \nu \mathcal{H}^{N-1}(C), \tag{4}$$

where C is the set of discontinuities/boundaries of the regions Ω_i, $\mathcal{H}^{N-1}(C)$ is the $(N-1)$-dimensional Hausdorff measure, representing the length of C and $\mu, \nu > 0$ are weighting constants for the three competing terms. The first term of the energy (4) imposes fidelity of the approximation u to u_0, the second term dictates smoothness of u outside of C and the third term requires smoothness of the contour C.

Vese and Chan [17] proposed to solve this problem using a multiphase level set approach. They use one or more level set functions, whose zero-level sets represent the set of curves C. With one level set function ϕ (which implies two phases $\{\phi > 0\}, \{\phi < 0\}$), one can detect multiple objects whose boundaries do not intersect, i.e. they can be represented by the zero level set of a single function: $C = \{(x,y) | \ \phi(x,y) = 0\}$ (see Figure 2(a)). Two level set functions ϕ_1 and ϕ_2 suffice to represent more complicated scenes, including triple junctions. The possible combinations $\phi_1 \lesseqgtr 0$, $\phi_2 \lesseqgtr 0$ generate four different phases, which we can use to "color"/delimit different adjacent regions in a partition (see Figure 2(b)). In the following, we present this four-phase segmentation model, which will be used to segment smooth regions in the multiscale image representation.

In this model, segmentation is achieved by evolving the level set functions ϕ_1 and ϕ_2, whose zero level sets represent the contours outlining object borders. At the same time, the piecewise smooth approximation u of the original image u_0 is computed. The link between ϕ_1, ϕ_2 and u is expressed via the auxiliary functions u^{++}, u^{+-}, u^{-+} and u^{--}, representing the four different phases, such that:

$$u(x,y) = \begin{cases} u^{++}(x,y) & \text{if } \phi_1(x,y) > 0 \text{ and } \phi_2(x,y) > 0, \\ u^{+-}(x,y) & \text{if } \phi_1(x,y) > 0 \text{ and } \phi_2(x,y) < 0, \\ u^{-+}(x,y) & \text{if } \phi_1(x,y) < 0 \text{ and } \phi_2(x,y) > 0, \\ u^{--}(x,y) & \text{if } \phi_1(x,y) < 0 \text{ and } \phi_2(x,y) < 0. \end{cases} \tag{5}$$

For brevity of the description, we denote u^{++}, u^{+-}, u^{-+} and u^{--} by u^{ij}, where $i = +, -$ and $j = +, -$. In this way, the minimization of the Mumford-Shah functional (4) can be translated to the problem of finding u, ϕ_1 and ϕ_2 which minimize

$$\begin{aligned} F(u, \phi_1, \phi_2) = & \sum_{i,j=+,-} \int_{\Omega} |u^{ij} - u_0|^2 H(i\phi_1) H(j\phi_2) dxdy \\ & + \mu \sum_{i,j=+,-} \int_{\Omega} |\nabla u^{ij}|^2 H(i\phi_1) H(j\phi_2) dxdy \\ & + \nu \int_{\Omega} |\nabla H(\phi_1)| + \nu \int_{\Omega} |\nabla H(\phi_2)|. \end{aligned} \tag{6}$$

The minimization of this energy with respect to the functions u^{++}, u^{+-}, u^{-+} and u^{--} leads to the following Euler-Lagrange equations:

$$\begin{cases} u^{ij} - u_0 = \mu \triangle u^{ij} & \text{in } \{i\phi_1 > 0, j\phi_2 > 0\}, \\ \dfrac{\partial u^{ij}}{\partial \vec{n}} = 0 & \text{on } \{\phi_1 = 0, j\phi_2 \geq 0\} \\ & \cup \ \ \{i\phi_1 \geq 0, \phi_2 = 0\}, \end{cases} \tag{7}$$

for $i = +, -$ and $j = +, -$ and where $\partial / \partial \vec{n}$ represents the partial derivative in the normal direction $\vec{n}$ at the corresponding boundary.

Minimizing the energy $F(u, \phi_1, \phi_2)$ with respect to ϕ_1 and ϕ_2 leads to the Euler-Lagrange equations for ϕ_1 and ϕ_2, which are integrated in a temporal scheme, yielding:

$$\begin{aligned}\frac{\partial \phi_m}{\partial t} = \ & \delta_\varepsilon(\phi_m)\Bigg(\nu \nabla \cdot \left(\frac{\nabla \phi_m}{|\nabla \phi_m|}\right) \\ & + \sum_{i,j=+,-} (-i)^{(2-m)}(-j)^{(2-n)}|u^{ij} - u_0|^2 H(j\phi_n) \\ & + \mu \sum_{i,j=+,-} (-i)^{(2-m)}(-j)^{(2-n)}|\nabla u^{ij}|^2 H(j\phi_n)\Bigg),\end{aligned} \tag{8}$$

for $m, n = 1, 2$ and $m \neq n$.

For segmentation, the level set functions are initialized as signed distance functions to some initial contours (circles on a grid, in our case). Then the discretized versions of the evolution equations (7) and (8) (see [17] for details of the numerical schemes) are applied successively, until the steady state is reached (the values of the evolving functions no longer modify from one iteration to the next). The final segmenting contours are given by the zero level sets of ϕ_1 and ϕ_2, while the approximation to u_0 is given by:

$$u = \sum_{i,j=+,-} u^{ij} H(i\phi_1)H(j\phi_2). \tag{9}$$

4. Multiscale Segmentation Using Active Contours

Our goal is to design an algorithm which can capture the structures present in an image, at different scales of observation of the image. Such an algorithm could be used in different ways by a computer vision system. A higher level image understanding process could indicate the desired scale of observation and the multiscale segmentation algorithm would output the detected objects *for the required scale*. Alternatively, the segmentation process could be given just the input image and be required to extract the meaningful objects *at each scale* (from a collection of scales). Then the high level process could chose to analyze the scene at a particular scale/level of detail (for instance based on the number of objects found by the segmentation at each scale).

To this end, our algorithm needs to use a simplifying/smoothing process that would generate a multiscale representation of the original image. Secondly, it should extract the structures (i.e. homogenous regions and their boundaries) from the image decomposition at each scale.

For the first task we choose the weighted total variation flow presented above, because of its edge and contrast preserving properties, favorable for the segmentation phase which follows. For structure extraction, we adopt the four-phase piecewise smooth version of the active contour model described above. The four phase model allows the segmentation of fairly complex image structures, such as triple junctions, and has offered very promising results in the literature [17].

The main problem of active contour models is their sensitivity to the initial conditions. Since they are relying on gradient descent for energy minimization, they are susceptible to fall into local minima, depending on their initial conditions. In order to solve this problem, we start the segmentation process at the coarsest/highest scale of observation of the image. Here, the image is simplified the most, since its details have been filtered out by the smoothing process. Therefore, its energy landscape contains less local minima and the active contours are be more likely to capture the global minimum. At the coarsest scale, we use the "seed initialization"(circles placed on a regular grid), which has shown good results in [17]. Once the coarsest scale has been segmented, the main structures in the given image will have been detected. We can then use the resulting active contours as initial condition for the image at the next finer scale, since their position will already be close to the global minimum.

The detection of new emerging structures in the finer scales is performed naturally within the active contour framework proposed in [17]. This is because the segmentation model is able to capture *interior structures*, i.e. structures appearing inside other structures which have already been outlined by the active contours. The reason for this is that the approximations to the Dirac and the Heaviside functions ($\delta_\epsilon(z)$ and $H_\epsilon(z)$) (which are used in the evolution equations 7, 8) have *infinite support*. Therefore, the evolution of the level set functions is not restricted to a narrow band around the zero level set, but takes place within the whole image domain, allowing the appearance of new contours.

In the following, we summarize our approach in an algorithmic setting. Let u_0 be the original image, ϕ_1 and ϕ_2 the level set functions modeling the active contours and u^{++}, u^{+-}, u^{-+}, u^{--} the auxiliary functions designating the four phases of the piecewise smooth Mumford-Shah model, denoted concisely by u^{ij}, where $i = +, -$ and $j = +, -$. The proposed algorithm consists of the following steps:

Stage 1
Derive the multiscale image representation using weighted TV flow:

- generate stack of scaled images $u(\lambda_1), u(\lambda_2), \ldots u(\lambda_m)$ by running to steady state the discretized version of equation (3) on the original image u_0, for decreasing weights of the fidelity term $\lambda_1 > \lambda_2 > \ldots > \lambda_m$ ($u(\lambda_1)$ - finest scale image, $u(\lambda_m)$ - coarsest scale image).

Stage 2
Extract structures from each scale using active contours:

- initialize level set functions (seed initialization): $\phi_1(0) = \phi_1^0$, $\phi_2(0) = \phi_2^0$;

- **for** $\lambda = \lambda_m$ **downto** λ_1
  segment $u(\lambda)$
  1. set time $n = 0$;
  2. initialize $u^{ij}(0) = u(\lambda)$ for $i, j = +, -$;
  3. compute $u^{ij}(n + 1)$ using $u^{ij}(n)$ and $\phi_1(n)$, $\phi_2(n)$ for $i, j = +, -$, according to the evolution equations (7);
  4. compute $\phi_1(n+1)$, $\phi_2(n+1)$ using $\phi_1(n)$, $\phi_2(n)$ and $u^{ij}(n+1)$, $i, j = +, -$, according to the evolution equations (8);
  5. $n = n + 1$;
  6. **if** the evolving functions ϕ_1, ϕ_2 and u^{ij}, with $i, j = +, -$ have not reached their steady state, **then** continue evolution: **go to step 3** **else** finished segmentation of $u(\lambda)$:
     – output final segmenting contours for scale λ as the zero level sets of $\phi_1(n)$, $\phi_2(n)$ and the segmented objects, as the connected components of the four phases $u^{ij}(n)$, with $i, j = +, -$;
     – initialize level set functions for the next scale as the final level set functions at this scale: $\phi_1(0) = \phi_1(n)$, $\phi_2(0) = \phi_2(n)$;
- **end for.**

5. Experimental results

We have applied our algorithm on two natural images containing structures at different scales. The images are presented in Figure 3(a) and 3(b) and will be denoted as 'Sergi' and 'Cameraman', respectively. Figures 4 and 5 illustrate the result of the first stage of the algorithm - the multiscale representations of the two images. In order to demonstrate our algorithm, we use four samples from the scale space of each image, as can be seen in Figures 4 and 5. These samples correspond to different scales of observation, given by different weights λ of the fidelity term in the weighted TV energy (2). Since the disappearance of details in the images throughout the scale space appears to be exponentially connected to the λ parameter, we have chosen an exponential law of variation for λ, in the form: $\lambda = \lambda_0 \alpha^m$. Then λ_0 and α were set to some arbitrary values, the same for both images ($\lambda_0 = 10$, $\alpha = 0.95$) and m was varied in order to obtain samples of the scale space. Therefore, in the following we will refer to m as the scale parameter. As can be seen in Figures 4 and 5, the first stage of the algorithm has achieved its purpose, providing us with a multiscale representation of the two images. The details of the images are gradually filtered out and finally only the main structures remain: the man and the painting ('Sergi' image) and the cameraman, the field and the sky ('Cameraman' image).

(a) Original image 'Sergi' (b) Original image 'Cameraman'

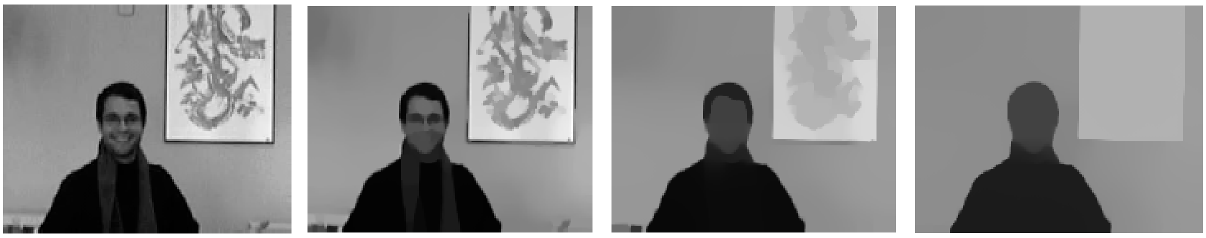
Figure 3. Original images considered for multiscale segmentation.

Figures 6 and 7 show the results of the second stage of the algorithm. For both images, the parameters used for the active contour model are $\mu = 30$, $\nu = 100$. The final segmenting contours for each scale are depicted with green and magenta, superposed on the corresponding images at each scale. They represent the zero level sets of the functions ϕ_1, ϕ_2, at the steady state of their evolution. At the coarsest level of scale, the fine details of both images have been eliminated by the weighted TV flow. Therefore, the subsequent active contour segmentation has outlined only the main objects still visible in the images: the man and the painting ('Sergi' image) and the cameraman, the field and the sky ('Cameraman' image). As we go down to finer scales, more details appear in the observed images and are outlined by the segmentation process. The obtained results prove the effectiveness of the proposed algorithm for the intended purpose: image segmentation at multiple scales of observation.

6. Conclusion and Future Work

We have introduced an algorithm for image segmentation at multiple scales. To this end, we have combined an image regularization technique with a variational segmentation technique. In a first stage, we derive a multiscale representation of the given image, using the weighted TV flow, which favors the long-time preservation of image edges and contrast with increasing scale. In the second stage, we successively segment each image in the scale-space, using active contours based on the Mumford-Shah functional. We start from the coarsest scale and use the resulting contours at one scale as initial contours in the segmentation of the next scale.

We have obtained promising results in the segmentation of natural images containing multiscale structure. We are currently investigating the merging of the two phases of our algorithm into a single variational model, by incorporating scale directly into the Mumford-Shah segmentation frame-



(a) Scale $m = 0$ (b) Scale $m = 58$ (c) Scale $m = 92$ (d) Scale $m = 108$

Figure 4. Stage 1 for 'Sergi' image: weighted total variation scale space.

(a) Scale $m = 0$ (b) Scale $m = 76$ (c) Scale $m = 92$ (d) Scale $m = 113$

Figure 5. Stage 1 for 'Cameraman' image: weighted total variation scale space.

work. We conclude that the proposed algorithm can contribute in the automation of the image understanding task, providing an outline of the significant structures in images, which correspond to different observation scales.

Acknowledgement

The authors would like to thank Ana Petrovic for her advice and ideas supporting this work.

References

[1] X. Bresson, S. Esedoglu, P. Vandergheynst, J. Thiran, and S. Osher. Global minimizers of the active contour/snake model. *UCLA CAM Report 05-04*, 2005.

[2] T. Chan and S. Esedoglu. Aspects of total variation regularized L^1 function approximation. *UCLA CAM Report 04-07*, 2004.

[3] J. Froment. A compact and multiscale image model based on level sets. In *Proceedings 2nd Int. Conf. Scale Space '99, in Lecture Notes in Computer Science*, volume 1692, pages 152–163, 1999.

[4] R. Henkel. Segmentation in scale space. In *Proc. of 6th Int. Conf. on Computer Analysis of Images and Pattern, CAIP '95*, Prague, 1995.

[5] S. Ishikawa, J. Weickert, and A. Imiya. Linear scale-space has first been proposed in Japan. *Mathematical Imaging and Vision*, 10:237–252, 1999.

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

[7] G. Koepfler, C. Lopez, and J. Morel. A multiscale algorithm for image segmentation by a variational method. *SIAM J. Numerical Analysis*, 31:282–299, 1994.

[8] T. Lindeberg. *Scale-Space Theory in Computer Vision*. The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, 1994.

[9] T. Lindeberg and J. Eklundth. Scale detection and region extraction from a scale-space primal sketch. In *3rd Int Conf Computer Vision*, pages 416–426, 1990.

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

[11] P. Monasse and F. Guichard. Scale-space from a level lines tree. *J. of Visual Communication and Image Representation*, 11:224–236, 2000.

[12] D. Mumford and J.Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications in Pure and Applied Mathematics*, 42:577–685, 1989.

[13] A. Petrovic, O. Escoda, and P. Vandergheynst. Multiresolution segmentation of natural images: From linear to nonlinear scale-space representations. *IEEE Transactions on Image Processing*, 13(8), August 2004.

[14] A. Petrovic and P. Vandergheynst. Multiscale variational approach to simultaneous image regularization and segmentation. In *Proceedings of ISPA*, Rome, Italy, September 2003.

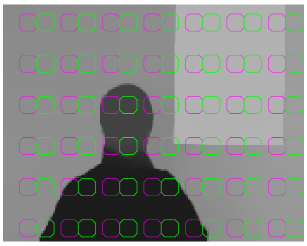
[15] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[16] D. Strong and T. Chan. Spatially and scale adaptive total variation based regularization and anisotropic diffusion in image processing. *UCLA CAM Report 96-46*, 2004.

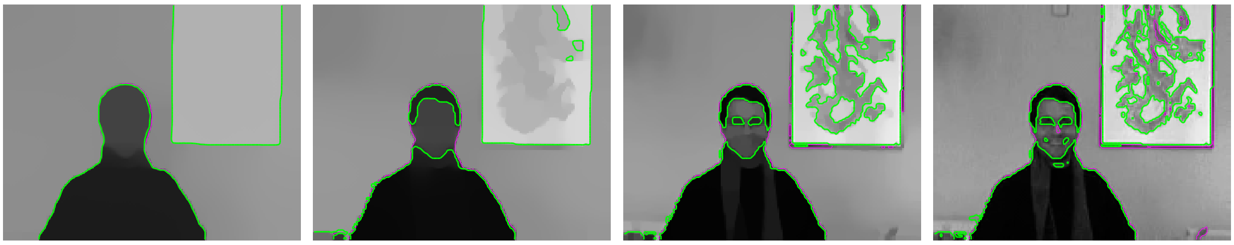
[17] L. Vese and T. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.

[18] K. Vincken, A. Koster, and M. Viergever. Probabilistic multiscale image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):109–120, February 1997.

[19] A. Witkin. Scale-space filtering. In *Proceedings of the Eight International Joint Conference on Artificial Intelingence*, pages 1019–1022, Karlsruhe, Germany, 1983.

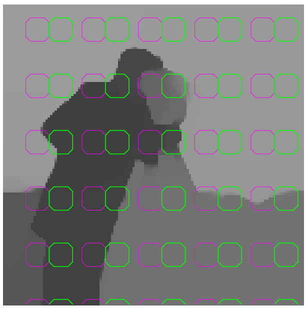


(a) Initial active contours on initial scale $m = 108$

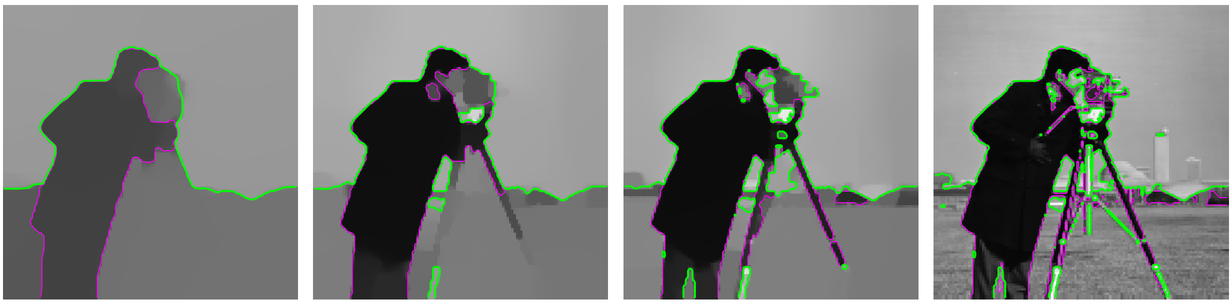


(b) Scale $m = 108$ (c) Scale $m = 92$ (d) Scale $m = 58$ (e) Scale $m = 0$

Figure 6. Stage 2 for 'Sergi' image: segmentation of the weighted TV scale space, using active contours.



(a) Initial active contours on initial scale $m = 113$



(b) Scale $m = 113$ (c) Scale $m = 92$ (d) Scale $m = 76$ (e) Scale $m = 0$

Figure 7. Stage 2 for 'Cameraman' image: segmentation of the weighted TV scale space.