Multiscale variational approach to simultaneous image regularization and segmentation

Ana Petrovic and Pierre Vandergheynst
Signal Processing Institute (STI)
Swiss Federal Institute of Technology in Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
{Ana.Petrovic, Pierre.Vandergheynst}@epfl.ch

Abstract

The problem of finding a piecewise smooth approximation of an original image while preserving edges is closely related to the image segmentation problem. Indeed, if we know the cartoon version of the original image, it is easier to perform an image partition into homogeneous regions. From this point of view, we propose to simultaneously perform regularization and segmentation.

By working in the space of functions of bounded total variation, we allow discontinuities in the result of the minimisation. Therefore we perform regularization by using a Total Variation (TV) minimisation based on the gradient descent method. Each gradient step involves solving a discrete approximation of the corresponding partial differential equation that results in a smooth estimate of the original image without blurring across edges. This non-uniform smoothing allows an explicit region growing segmentation scheme that minimises one form of the Mumford-Shah functional. Therefore, the major novelity of our approach consists in competing two processes, one for boundary detection and the other for intraregion smoothing.

1. Introduction

This paper addresses the classical problem of image segmentation. This problem can be roughly described as the process of partitioning an image into a set of non-overlaping semantically 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 colour homogeneity, texture coherence ...), each of them leading to different solutions. In all cases the areas Ω_k should correspond to "meaningful" structures (e.g. real world objects).

Structure detection is inherently a multiscale problem. In vision, a unique scale does not exist. If you move closer to a scene all objects become bigger and we are able to distinguish substructures within bigger structures. On the contrary if you move further away, all objects become smaller and we can notice only some structures without details. Our goal is to identify all image structures regardless of the scale at which they occur and that involves using scale as a parameter. Therefore the issue of scale selection and structure detection cannot be treated separately and as a result we have to operate at multiple scales. Multiresolution methods attempt to obtain a global view of an image by examining it at various resolution levels. Thus, segmentation methods that operate on multiple images will improve final segmentation.

One of the clearest examples of multi-scale data representation is scale-space [16], [15], [11]. In scale-space theory, the observed image $I(x)$ is embedded into a continuous family $I(x,t)$ of gradually smoother versions of it, where t correspond to the scale (i.e. artificial "time"). The original image corresponds to the scale $t = 0$ and increasing the scale should simplify the image without creating spurious structures. If we choose TV evolution [6], that encourages smoothing within homogeneous regions in preference to smoothing across the boundaries, the blurring will take place separately in each region letting region boundaries remain sharp. As a result, scale-space produces a hierarchy of simplified, sketchy versions of the original image approaching the piecewise constant approximation of the original image. This allows extraction of the structural decomposition at each scale and easy examination of the change in structure due to a change in the scale. Hence, a multiscale segmentation algorithm is still required to extract the image structures from the scale-space decomposition. The structure is contained implicitly within the scale-space, representing a decomposition at different level of details. Tracking the structures through the scales naturally results in a multiscale region growing algorithm that merges all the structures of the finest level into bigger structures at coarser level of resolution (Figure 4). That is, starting at the finest levels of resolution, and iteratively passing from one level to the next, the region growing algorithm is performed in such a way that the output of the process at each level is used as an input for the next level. The region growing algorithm is based on minimising a variant of the Mumford-Shah functional [9].

The multiscale segmentation algorithm that we introduce is motivated by the work of Mumford and Shah [9], [8]. They studied a general functional from the point of view of approximation theory and defined segmentation problem as the computation of an optimal approximation of a given image by piecewise smooth functions. In addition, our approach has an intellectual connection with some other important works in the field: total variational approach [12], [6]; shock filters of Osher and Rudin [10]; scale-spaces [16], [15], [11]; region merging procedures [5]. Interested readers are referred to Morel and Solimini's book on Variational Methods in Image Segmentation [7] for an exhaustive survey of variational models in image processing.

The rest of this paper is organised as follows. Section 2 gives the main motivation of our work and its connection to the Mumford-Shah global segmentation model. Section 3 reviews multiscale representations of images and presents the specific form used in this paper. Section 4 discusses the extraction of image structures from the multiscale representation, in particular the multiscale region growing algorithm that minimises a variant of the Mumford-Shah energy functional. Section 5 gives experimental results for real images. Section 6 presents concluding remarks.

2. Review of the segmentation model

S.Geman and D.Geman [3] were the first to introduce a global model for image segmentation using Markov random fields. A continuous version of Gemans's model was presented by Mumford and Shah using a variational approach. This model defines the segmentation problem as a joint smoothing/edge detection problem: given an image $g(x)$, one seeks simultaneously a "piecewise smoothed image" $u(x)$ with a set B of discontinuities that represent the "edges" of g. Then one can define the segmentation energy by:

$$E(u,B) = \mu^2 \int_R (u-g)^2 dx + \int_{R\setminus B} |\nabla u|^2 dx + \nu |B|, \quad (1)$$

where:

- R is the image domain,
- g denotes the observed data (i.e. the gray level function),
- B is the union of segment boundaries,
- u denotes the piecewise smooth approximation of g with discontinuities along B,
- $|B|$ is the total length of boundaries B,
- μ and ν are the weights.

This functional simultaneously measures the quality of the boundaries B and of the approximation function u : the minimum of this functional is close to the original image, piecewise smooth and with few discontinuities. Thus, minimising this functional requires estimating two processes, the continuous segmented field u, and a binary edge process B. Even though this model is very simple, it has been extremely important in acquiring a deep understanding of the general segmentation problem.

The disadvantage of this functional is that it lacks practical means to solve u and B. To overcome these practical difficulties associated with minimisation of the original Mumford-Shah segmentation energy, we propose to cast the problem into two coupled processes, one for non-uniform smoothing, and an other for region detection. The intraregion smoothing of the image is determined by the minimisation of the Total Variation. This can be achieved by a gradient descent which yields to a nonlinear diffusion equation. Discretisation of such non-linear equation results in a multiresolution representation of the image. Then to recover the actual object boundaries we apply a multiscale region growing algorithm that operates on this non-linear scale-space.

Thus motivated by the general Mumford-Shah functional we propose to find an equilibrium between two competing processes rather than a search for a global minimum of a single functional with many local minima. It is still an open question whether the proposed system has a unique solution but empirical results strongly suggests this to be the case.

3. Multiscale representation

Lindeberg [15] observed that objects in the real-world appear in different ways depending on the scale of observation. He gives as a simple example the concept of a branch of tree which makes sense only at a scale from a few centimeters to at most a few meters, while it is meaningless to discuss the tree concept at the nanometer or kilometer level. Besides this multi-scale properties of real-world objects, it is necessary to cope with the complexity of unknown scenes and noise. Therefore in order to detect true physical objects, while avoiding overly complex schemes, we need some simpler multiscale image model and this representation should be based on the geometry of the image. Also, according to the HVS, edges are very important primitives in natural images. One way to select true edges is to encourage smoothing within homogeneous regions in preference to smoothing across the boundaries. The blurring would then take place separately in each region, with no interaction among regions. In what follows, we will show that TV flow preserves the image coherent structures through the scales by letting the region boundaries remain sharp[1].

3.1. TV flow

The idea of minimising the total variation (roughly speaking the L^1 norm of the gradient instead of its L^2 norm) for image processing purposes was first introduced by Osher and Rudin [12].

$$TV[I] = \int_{\Omega} |\nabla I| \, dx. \tag{2}$$

Here Ω denotes the continuous signal domain, ∇I the gradient, and dx the area element of Ω. The main properties of this functional is that it can be defined for functions that have discontinuities, leading to a correct representation of edges in the image. Thus, the TV norm does not penalize discontinuities in I and allows us to recover edges of the original image. The variational problem using the total variation (TV) norm with constraints was originally introduced by Rudin and Osher [12] for the blurring and denoising problems. The minimisation of the Total Variation can be achieved by a gradient descent method which yields to the following evolution equation:

$$\frac{\partial I}{\partial t} = \text{div}\left(\frac{\nabla I}{|\nabla I|}\right). \tag{3}$$

The main problem in minimising the TV norm comes from the non differentiable argument $|\nabla I|^{-1}$. The problem is usually solved, in the discretization of $|\nabla I|$, by replacing with $\sqrt{|\nabla I|^2 + \epsilon},\ \epsilon > 0$ small. We rewrite equation (3) as:

$$\frac{\partial I}{\partial t} = \frac{1}{|\nabla I|} I_{\xi\xi}, \tag{4}$$

where $I_{\xi\xi}$ is the second order partial derivatives in the direction orthogonal to the gradient. Written in this way, the method appears as a diffusion in the direction orthogonal to the gradient, tuned by the magnitude of the gradient and the main advantage is that edges are very well preserved. As t increases, we approach a blurred version of our image, and the effect of the evolution is edge enhancement and intraregion smoothing as can be seen in Figure 1. As we can see, the bigger the scale, the less information referred to the local characteristics of the input data will appear, but the general information such as edges will remain through scales. 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 obtain an image segmentation scheme.

(a) Level 2

(b) Level 4096

(c) Level 16384

(d) Level 65536

Figure 1. TV flow at different scale levels.

4. Multiscale energy-minimising segmentation

We saw that discretization of TV diffusion equation produces a non-linear scale space. Assuming those images to be an optimal, piecewise smooth and relatively flat approximations of the original image, the next step in the development of the model is to adopt a criterion for region detection. The choice of an object detection technique depends on the properties of multiscale representation. The scale parameter t can be interpreted as a measure of details in the image: starting at the finest scale level and iteratively passing from one level to the next coarser level, all small structures are merged into bigger ones according to the multiscale property of the objects. In this way, the ***causality property*** is satisfied: if $t > t'$, then the regions boundaries for t are contained in those obtained for t' and the regions of the segmentation associated to t are the unions of some of the regions obtained for t'. This property ensures that the segmentation can be performed by a region growing algorithm in such a way that the output of the process at each level is used as an input for the next level. Therefore the region growing algorithm is applied in a multiscale framework. In other terms, it will not only compute one segmen-

[1] Other anisotropic diffusion techniques share this property and could be used here.

tation, but a hierarchy of segmentations from fine to coarse scales t.

4.1. Region Growing

In this section we present the main ideas of Koepfler, Lopez and Morel's region growing algorithm [5] based on the minimisation of the simplest case of Mumford-Shah functional. In this case one can prove [9] that for a given partition Ω of the image domain R, the piecewise constant function which minimises the energy $E(u, B)$ in equation (5), is the functional equal to the mean of the original data image in each region of partition Ω. We can define the discrete version of the energy functional by:

$$E(\Omega) = \sum_{i=1}^{N} \sum_{s \in \Omega_i} (\overline{g}_{\Omega_i} - g(s))^2 + \lambda |\Omega|, \qquad (5)$$

where

- g is the image to be segmented,
- N is the number of regions,
- $\overline{g}_{\Omega_i}$ is the mean of g in the region Ω_i,
- $g(s)$ is the gray level intensity of the pixel at position s,
- $|\Omega|$ is the total length of boundaries.

The first term is the sum of the squared distances between the original image and its piecewise constant approximation. The second term penalises excessive splitting of the image. The recursive region merging algorithm looks for a decay of global energy. Two adjacent regions will be merged if this operation reduces the global energy. Therefore, the decision to proceed to a merging of two regions Ω_i and Ω_j depends on the sign of

$$\Delta E = E_{\Omega_i \cup \Omega_j} - E_{\overline{\Omega_i \cup \Omega_j}}, \qquad (6)$$

where:

- $E_{\Omega_i \cup \Omega_j}$ is the energy of the segmentation when Ω_i and Ω_j are merged,
- $E_{\overline{\Omega_i \cup \Omega_j}}$ is the energy of the segmentation when Ω_i and Ω_j are not merged.

If the merging criterion $\Delta E < 0$ is satisfied we remove a common boundary $\partial(\Omega_i, \Omega_j)$ of two neighbouring regions Ω_i and Ω_j. By repeating this step until no merging is possible, we finally get a 2-normal segmentation [7] where no further elimination decreases the energy.

4.2. Multiscale Segmentation Algorithm

In what follows, we set up the terminology and the construction procedure for the proposed hierarchical algorithm. We are dealing with the digital image data. Let Ω be an open subset on $\mathbb{R}^2$ and g a grey-scale image treated as a function defined on Ω. The image domain Ω is a collection of pixels on a discrete grid. The ideas of section 2 are applied in order to give the precise algorithm scheme.

Let g be an original image from which we construct a sequence of images $u^0 = g, u^1, u^2, ..., u^S$ according to the numerical implementation of the TV flow. Therefore, we build a sequence $u^k,\ 0 \le k \le S$ such that:

$$\Delta u^{k-1} = \frac{u_{xx}^{k-1}(u_y^{k-1})^2 - 2u_{xy}^{k-1}u_x^{k-1}u_y^{k-1} + u_{yy}^{k-1}(u_x^{k-1})^2}{((u_x^{k-1})^2 + (u_y^{k-1})^2)^{3/2}},$$

$$\begin{cases} u^k = u^{k-1} + \Delta t \Delta u^{k-1}, & k \ge 1, \\ u^k = g, & k = 0. \end{cases} \qquad (7)$$

We have applied the explicit Euler method in time and Δt indicates the timestep. The gradients could be approximated by central differences in space. We use the terminology "image at scale k" to denote the image u^k, therefore the scale represents the index of the image. This scale factor can be also interpreted as a time factor or the level of the hierarchy. The parameter S represents the stopping scale, when the result of segmentation gives just one region.

Now, when we have constructed the sequence $\{u^k\}_{k=0}^{S}$, we can define partitions of each of the image domains $\{\Omega^k\}_{k=0}^{S}$ respectively by the region growing algorithm based on minimisation of the simplest case of Mumford-Shah functional:

$$E(\Omega^k|\Omega^{k-1}) = \sum_{i=1}^{P(k)} \sum_{s \in \Omega^k} (\overline{u}_{\Omega^k}^k - u^{k-1}(s))^2 + k\left|\Omega^k\right|, \quad (8)$$

where:

- u^k is the image to be segmented,
- $P(k)$ is the number of regions,
- $\overline{u}_{\Omega^k}^k$ is the mean of u^k in the region Ω^k,
- $u^{k-1}(s)$ is the gray level intensity of the image from the previous scale at the position s,
- $\left|\Omega^k\right|$ is the total length of boundaries.

By using $E(\Omega^k|\Omega^{k-1})$ we means that we initialise the region growing algorithm at scale k with an initial partition Ω^{k-1} that is the partitioning results at the finer scale $k-1$. This is done in order to detect the structure change among the scales.

The weight attributed to the region term $\left|\Omega^k\right|$ should obviously vary with the scale k, because there are much more regions at small scales than at large scales. We chose to vary this weight in a way that compensates for the diminishing number of regions, and in this paper, we simply set a linear weight.

The result of the multiscale region growing procedure is the partitioning of all rectangles $\{\Omega^k,\ 0 \le k \le S\}$ into finite sets of non overlapping regions constructed in such a way that the output of the process at each level is used as an input for the next level. The causality property of the scale-space ensures that the partition is unique and can be done at each scale. Therefore at each scale k we have partitions $\{\Omega_0^k, \Omega_1^k, ..., \Omega_{P(k)}^k\}$ such that:

$$\bigcup_{i=0}^{P(k)} \Omega_i^k = \Omega^k, \quad \Omega_i^k \bigcap_{i \neq j} \Omega_j^k = \{0\},$$

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

The previous procedure can be implemented in an iterative form:

Initialisation: $k = 0$

1. Calculate an image u^k at scale k using equation (7).
2. Split the image at scale k into regions $\{\Omega_i^k,\ 0 \le i \le P(k)\}$ by minimising the energy in equation 8.

$$\Omega^k = \text{argmin}_{\Omega^k} \left[E\left(\Omega^k|\Omega^{k-1}\right)\right]$$

3. $k = k + 1$.
4. If $P(k) > 1$ go to step 1.

5. Experimental results

Figure 2 shows the TV flow at different scale level S. A hierarchy of segmentations from fine to coarse scales is computed by applying the region growing segmentation algorithm through scales. Segmentation results are presented in Figures 3 and 4. Figure 3 represents partitioning results Ω^S at scale S, while Figure 4 gives the boundaries of those partitions. Notice that smaller objects are merged into a bigger one through the scales. Obviously, the further we go through scales, the less details we observe on segmented images. On Figure 3, we can see that if we stop our algorithm at scale $S = 1953$, the hair, the face and the scarf are still recognized as an objects, but at $S = 6424$ the scarf is already merged with the body and they represent one single object. Eventually, if we perform region merging up to $S = 10775$, the hair and the face are merged into one object that represent the head. We can conclude that sufficiently good results have been obtained for the extraction of meaningful objects.

(a) Original image: $S = 0$.

(b) Scale level: $S = 1953$

(c) Scale level: $S = 6424$

(d) Scale level: $S = 10775$

Figure 2. Scale-Space: TV flow.

(a) Level of segmentation: $S = 1953$

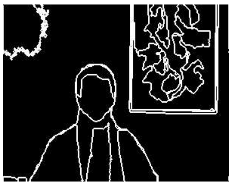
(b) Level of segmentation: $S = 6424$

(c) Level of segmentation: $S = 10775$

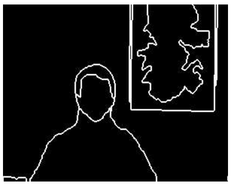
Figure 3. Segmentation results obtained applying region growing strategy up to different level S of scale-space.

6. Conclusions

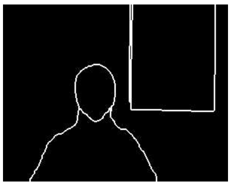
We have introduced a multiscale algorithm for an unsupervised image segmentation. This segmentation method is



(a) Level of segmentation: $S = 1953$



(b) Level of segmentation: $S = 6424$



(c) Level of segmentation: $S = 10775$

Figure 4. The borders of the objects that are extracted at scale S.

based on two coupled processes: simultaneous intraregion smoothing and boundary detection. The result of such interactions produce two hierarchies, one of segmented images (Figure 3) and an other one with more simplified approximations of the original image (Figure 4). This corresponds to the main idea of the Mumford-Shah general model of image segmentation.

A non-linear scale-space stack is constructed by minimisation of the total variation norm that gives a family of approximated, sketchy versions of the original image. This scale-space represents a natural way of merging segmented structures through scales. During evolution, substructures are smoothed out into bigger structures according to the presence of details at different scales. Therefore this is a good framework for multiscale segmentation by region growing. By approaching more simplified versions of the original image, regions are merged through scales. In particular, region growing is based on minimisation a variant of the Mumford-Shah functional.

This technique is light on the computational point of view and presented results demonstrate the efficiency of the method in segmenting real images.

References

[1] Y. Chen, B. Vemuri, and L. Wang. Image denoising and segmentation via nonlinear diffusion. *Computers and Mathematics with Applications*, 39(5/6):131–149, 2000.

[2] O. Escoda, A. Petrovic, and P. Vandergheynst. Segmentation of natural images using scale-space representations: a linear and a non-linear approach. In *EUSIPCO*, 2002.

[3] S. Geman and D.Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans*, 1984.

[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] G. Koepfler, C. Lopez, and J. Morel. A multiscale algorithm for image segmentation by variational method. *SIAM J. Numer. Anal.*, 1994.

[6] A. Marquina and S. Osher. Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal. *SIAM J. on Scientific Computing*, 22(2):387–405, 2000.

[7] J. Morel and S. Solimini. *Variational Methods in Image Processing*. Birkhauser, 1994.

[8] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1985.

[9] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and variational problems. *Comm. Pure Appl. Math.*, 1989.

[10] S. Osher and L. Rudin. Feature-oriented image enhancement using shock filters. *SIAM J. on Numerical Analysis*, 1990.

[11] P. Perona and J. Malik. Scale-space and edge detections using anisotropic diffusion. *IEEE Trans. PAMI*, 12(7):629–639, 1990.

[12] L. Rudin and S. Osher. Total variation based image restoration with free local constraints. In *Proc. IEEE Internat. Conf. Imag. Proc.*, pages 31–35. IEEE Press, Piscataway, NJ, 1994.

[13] J. Shah. Segmentation by nonlinear diffusion. In *IEEE Conf. on Vision and Pattern Recognition*, 1991.

[14] J. Shah. Segmentation by nonlinear diffusion ii. In *IEEE Conf. on Vision and Pattern Recognition*, 1992.

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

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