An Adaptive Total Variation Model for Image Segmentation

Ana Petrovic and Pierre Vandergheynst,

Technical Report: TR_ITS 2005.30

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

Abstract

In our previous work [1], tracking the iso-level sets through total variation scale-space proved to be a very efficient tool for unsupervised segmentation. Stepping on these results, we propose a new segmentation approach in a unified total variation framework. The main idea is to use the total variation energy at each scale to drive the region merging process. We show that this total variation formulation, which was originally proposed for restoration and enhancement, is also well suited for segmentation. In addition, this energy functional can be derived from a Bayesian principle using a Markov random field prior. We demonstrate the effectiveness of our method on gray scale, noisy, color and texture images.

Index Terms

Unsupervised Segmentation, Total Variation Diffusion, Total Variation Regularization, Spatially Adaptive Segmentation, Region Merging, Energy Minimization, Bayesian model, Multi-resolution.

I. Introduction

During the last decades a lot of work has been done in integrating partial differential equations in image processing for various tasks such as restoration, enhancement, segmentation, reconstruction ([2], [3], [4], [5], [6], etc.). One of the earliest formulations of anisotropic diffusion is due to Perona and Malik [7]. In this work, the edge detection step is introduced in a diffusion equation in order to simplify the original image, while preserving the borders of original regions. One of the conceptually limit cases of the Perona-Malik diffusion is total variation (TV) flow [8]. This diffusion method requires no additional parameters, it is well-posed and tends to piecewise constant approximations of the original image that are naturally suited to the segmentation problem. The effectiveness of TV scale-space for the segmentation is shown in [1], where segmentation results are induced by tracking the iso level sets through the scale-space stack. The difficulty of the method comes from simultaneous coupling of two processes: image simplification and region linking. To overcome this limitation, here we propose to unify those processes by applying a region-merging procedure that minimizes the corresponding energy of TV evolution. Thereby, driven by total variation minimization, pairs of regions are recursively merged progressively resulting in coarser segmentation.

Since we can associate with this diffusion a multiscale energy proposed and developed by Rudin, Osher and Fatemi (ROF) [9], there exist a direct connection with total variation restoration. The ROF restoration model is one of the most successful tools for image restoration and enhancement and, as we show, it can be equally used for explicit segmentation. Total variation based techniques take advantages of no particular bias toward a discontinuous or smooth solution. The interpretation of TV restoration as an approximation of the original image that has minimal total variation, leads to the design of a segmentation scheme that looks for a piecewise constant approximation of the observed data. Therefore, this technique is achieving both noise removal and exact edge location through the use of the total variation equation. Another property of total variation methods is linear contrast reduction. A natural approach to overcome this drawback is to control the amount of regularization according to intensity changes. The main idea is to smooth more in the regions of low intensity change and less in regions of high intensity change. Following this spatial adaptivity, which has proven useful in the image restoration setting [10], [11], [12], [13], we propose a spatial adaptive segmentation scheme.

The image segmentation problem can be also approached using statistical Bayesian methods [14],[15], [16] . In this case one seeks the solution which most closely matches the probabilistic behavior of the original data. So, segmentation can be reformulated as a maximum a posteriori (MAP) estimation problem which allows the introduction of a prior distribution. In order to account for the interaction among neighboring pixels, one usually adopts a Markov random field (MRF) image model. In MRF model, we use the simple potential function which penalizes the absolute differences among neighboring pixels in the estimate. This particular case of potential function does not penalize discontinuities nor smooth functions, thus behaving as total variation. Therefore, our approach can alternatively be derived as MAP estimate using MRF prior.

The outline of the report is the following. In Section II we set the framework by describing or recalling the concepts of Segmentation Energy, Region Merging and Multiscale Representation. Section III surveys TV diffusion and TV regularization, their most important features and application in image segmentation. In Section IV we derive our approach from a Bayesian statistical framework. Section V details the segmentation process and its important properties. The next Section is devoted to experimental evaluation of our algorithm. We apply the method to gray scale, noisy, color and texture images. Finally, in Section VII, we expose our conclusions.

II. Segmentation Energy, Region Merging, Multiscale Representation

This section briefly reviews the segmentation problem as an energy minimization problem considering a general class of energies that contain two terms: an approximation term and a regularization term. Then, region merging is described as an optimization strategy in order to find the minimum of this energy. Finally, the whole framework is extended in a multiscale fashion.

A. Segmentation Energy

Segmentation can be described as the process of partitioning an image into a finite set of non overlapping regions, each of which should be uniform with respect to some characteristic such as intensity level, texture, etc.

The natural properties that a genuine segmentation algorithm should satisfy are:

- region boundaries should correspond to object edges,
- boundaries should also be simple, not ragged and smooth enough.

Achieving these properties leads to a general variational formulation proposed by Terzopoulos [17] . According to this paper, most vision problems can be seen as energy minimization problems that can be globally written as:

$$E(u) = E_A(u) + \lambda E_R(u), \tag{1}$$

where u is a segmented image, E_A is an approximation term, E_R is a regularization term and λ is a regularization parameter. The first penalizes distance between the original image and the recovered model and the second measures unsmoothness of the model. Depending on the theoretical framework, many segmentation energies are proposed: variational [18], Bayesian [14], minimum encoding [19]. In this report, the idea is to express the modified Rudin, Osher, and Fatemi Total Variation denoising model [9] in terms of segmentation energy. Then, using a statistical approach, this segmentation energy is presented as maximum a posteriori (MAP) estimation of the image given the observed data. In the case of one dimensional problems, we can use dynamical programming [20] to find the global minimum fast and efficiently. For higher dimensions, it is possible to apply a Monte Carlo [14] algorithm known as simulated annealing or the graduate non-convexity method [21]. If we want to avoid overly complex optimization algorithms, one simplified solution is to apply region merging, which is the approach we propose in the following.

B. Optimization strategy: region merging

Finding the minimum of functional (1) over the set of all partitions of an image is notoriously intractable. Therefore, to minimize this function we use a greedy region merging heuristics, which results in a stepwise optimization approach to the global optimization problem. We start with the initial set of regions and update these regions iteratively to converge to a stable solution. The segmented image can be represented as an undirected weighted graph G(V,E), where vertices V represent regions and edges E represent boundaries between regions (i.e. only neighboring regions are connected by an edge). The edge weight $W_{i,j}$ between adjacent nodes Vi and Vj measures a distance between those two nodes.

Then, two adjacent nodes should be merged if this operation causes the energy to decrease, i.e. the weight between nodes is the smallest value among all weights. In this settings, the region merging algorithm decreases the global energy and can be defined by the following steps:

1) Construction of graph G with initial segmentation: each pixel is a separate region (vertex), with the four direct neighbors connected by edges of corresponding weights.
2) Remove the edge $e_{ij} = \{Vi, Vj\}$ from G that has the smallest weight $W_{i,j}$ and merge the nodes Vi and Vj into one Vij.
3) Update all edges and weights spanning from the node Vij.
4) If the number of nodes is greater than one, repeat steps 2 to 4. Otherwise stop.

This graph based formulation of the algorithm, allows efficient implementation using an adjacency list.

C. *Multi-scale extension*

One of the most important parameter in computer vision is *scale* since coherent structure in an image are only observable on a certain range of scales. Therefore, to each scale should correspond a different segmentation energy and our goal is to define this multiscale energy. Moreover, multi-scale implementation usually results in improvements in computational speed and robustness.

One possible way to define energy at different scales, proposed by Koepfler et al. [22], is to consider the regularization parameter λ in (1) as a scale parameter. Thus, increasing λ we obtain a hierarchy of segmentation from fine to coarse. The principle of increasing this regularization term results in imposing a smooth solution that correspond to coarse segmentation. In [22], the authors used the piecewise-constant Mumford-Shah model as the merging criterion. Here, we introduce a different multiscale segmentation energy within a variational context.

Our methodology is motivated by a nonlinear diffusion technique called TV flow [8] which preserves only the most important components through scale, producing segmentation like results. This multiscale-data representation allows us to define the segmentation energy that correspond to each scale. In the next section we give the most important properties of this diffusion process that explain the overall behavior of our segmentation algorithm.

III. Segmentation framework : Total variation minimization

Total Variation (TV) minimization was introduced by Rudin, Osher and Fatemi [9] for use in image restoration problems. The main idea is to minimize the total variation in the image, subject to constrains involving goodness of fit to the original data. The constrains are imposed using Lagrange multipliers, so the problem can be formulated as an unconstrained or Tikhonov problem [23]:

$$\min_u \quad \frac{1}{2}\int_\Omega (u-u_0)^2\,dx + \lambda \int_\Omega |\nabla u|\,dx. \tag{2}$$

The main advantage of TV regularization over the other regularization techniques, is that there is no particular bias toward a discontinuous or smooth solution, in other words it does not penalize edges. The solution of (2) satisfies the following Euler-Lagrange equations:

$$\lambda \nabla \left(\frac{\nabla u}{|\nabla u|}\right) - (u-u_0) = 0, \tag{3}$$

that can be rewritten as

$$u = u_0 + \lambda \nabla \left(\frac{\nabla u}{|\nabla u|}\right). \tag{4}$$

One could also consider regularization without the fitting constrains:

$$\min_u \quad \int_\Omega |\nabla u|\,dx. \tag{5}$$

Then, the corresponding Euler-Lagrange equation can be solved by gradient descent method, i.e. by marching the following PDE to the steady state:

$$\frac{\partial u}{\partial t} = \nabla \left(\frac{\nabla u}{|\nabla u|}\right), \quad u(t=0) = u_0. \tag{6}$$

The parabolic counterpart to TV regularization appears as a special case of anisotropic diffusion called TV flow [8]. The basic idea here is that the image is diffused more where edges are not present (i.e. denominator $|\nabla u|$ is small) and less where edges are present (i.e. denominator $|\nabla u|$ is big). The main property of this evolution is that it tends to preserve edges producing piecewise constant, segmentation-like results, as can be seen in Fig. 1. On that figure one can observe that the TV scale-space is a family of segmentations of the initial image, with larger values of the scale parameter t corresponding to segmentations at coarser scale. The amount of energy at every scale naturally defines a segmentation energy associated with TV evolution. Therefore, to induce segmentation, it is necessary to define this energy along the evolution.

Notice, however, that there exists a direct connection between TV regularization and TV diffusion. If one time-marches equation (6) for n time steps $t_n = n\Delta t$, then the obtained image is approximately

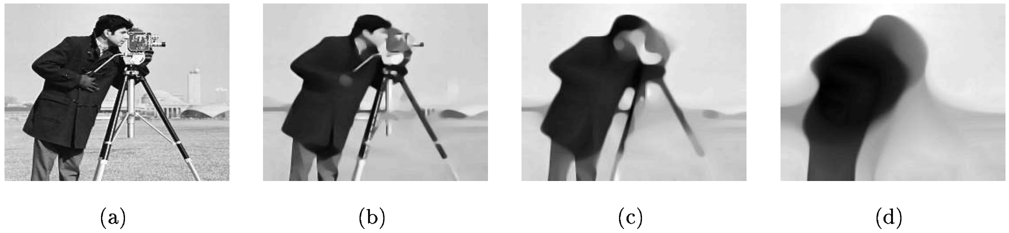


Fig. 1. Total Variation Diffusion at different scale levels t, dt=0.025. (a) $t = 2$, (b) $t = 4096$, (c) $t = 16384$, (d) $t = 65536$.

equivalent to the solution of problem (2) with $t_n = \lambda$ small. So roughly speaking, the diffused image at time t corresponds to a regularized solution with regularization parameter t. This formulation has two main interest:

- If we interpret time as a scale, than the corresponding regularization parameter could be also considered as a scale.
- It makes the TV regularization functional appear explicitly as the right one to be associated with TV diffusion.

This bring us to the conclusion that total variation minimization can be an effective algorithm for multiscale segmentation. The implementation of such an algorithm naturally results in a recursive region merging strategy explained in II-B. Since the behavior of this algorithm corresponds to TV flow, the next section briefly reviews the main properties of this diffusion.

A. Effects of TV diffusion

For a better understanding of how TV flow influences an image there is a need for an exact analytical solution. This is possible to derive in 1D discrete space [24], [25] but starting from 2D topology it is no longer possible to obtain an exact solution to the problem. Nevertheless, one can find explicit solutions for some important particular cases that yield good approximations for the general 2D case. In the following, we give the results developed in [26] for two constant regions in R^2. If we consider the simplest case in 2D with only two constant regions (Fig. 2), defined as

$$u = \begin{cases} u_1, & \vec{x} \in \Omega_1 \\ u_2, & \vec{x} \in \Omega_2, \end{cases} \tag{7}$$

it is possible to calculate the evolution speed of pixels. Pixels within the regions with the same value

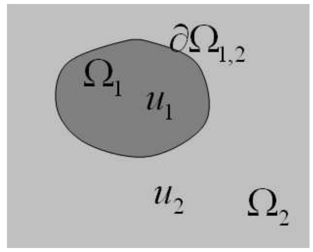


Fig. 2. Piecewise constant image of two regions.

evolve linearly with the same speed. Indeed define:

$$\delta_i = \begin{cases} +\frac{|\partial\Omega_{i,j}|}{|\Omega_i|}, & u_i < u_j \\ -\frac{|\partial\Omega_{i,j}|}{|\Omega_i|}, & u_i > u_j, \end{cases} \tag{8}$$

where $|\Omega_i|$ is the area of Ω_i, $|\partial\Omega_i|$ is its boundary length and $|\partial\Omega_{i,j}| \equiv |\partial\Omega_i \cap \partial\Omega_j|$. Then $u_i(t)$ is given by

$$u_i(t) = u_i(0) + \delta_i t. \tag{9}$$

Considering this equation, there exists a finite time $T > 0$ such that for all $t \geq T$ the image becomes constant

$$u(t) = \frac{u_1|\Omega_1| + u_2|\Omega_2|}{|\Omega_1| + |\Omega_2|}, \; \forall t \geq T, \tag{10}$$

i.e. evolving to the mean value of the whole image. These results show the following:

1) TV diffusion tends to preserve the exact edge locations. That brings us to conclude that the analogue segmentation procedure would respect object boundaries.
2) If we define the local scale of the object as the ratio of the area of an object to its boundary length, than the intensity change is inversely proportional to this local scale and is independent of the original intensity. So we can conclude that in the segmentation algorithm, properties of the objects define the merging speed.
3) The steady state is the mean value of the original image that results in recognizing the whole image as one object at the final segmentation step.

These conclusions, although developed for the precise and simple cases, give very important insights for better understanding the effects of TV diffusion (i.e. TV segmentation).

B. Piecewise Constant Image Model

The next important example is shown in Fig. 3, in which we demonstrate the behavior of TV model for two different statistical cases: where the regions' distributions have the same mean but different variances and different means with different variances. Fig. 3(a) consists of 4 regions whose intensities are generated from four Gaussian distributions with identical means: $N(160, 70^2)$, $N(160, 20^2)$, $N(160, 130^2)$ and $N(160, 70^2)$. We measure the mean value and standard deviation of regions through TV evolution that are shown in Fig. 3(b,c) respectively. As expected, the mean value remains constant through scale while the standard deviation of each region tends with different speed but very fast to zero. In this case, we need to measure the second order statistic (variance) of each region in order to induce the correct segmentation. In the Fig.3(d), regions are generated randomly from four Gaussian distributions with different means and variances: $N(90, 70^2)$, $N(30, 20^2)$, $N(240, 130^2)$ and $N(120, 70^2)$. Here, the regions change their mean values with constant speed toward the steady state solution (i.e. mean value of the entire image). More specifically, along evolution we observe a sequence of "break points" that corresponds to times at which regions merge and continue evolving with changed speed according to equation (8). As in the previous case, the variances very quickly go to zero since the diffusion process first changes the pixels values that decrease the total variation the most, i.e. the pixels that lie in the tails of the distributions. This makes the effects of noise removing more obvious and explains why the corresponding segmentation algorithm is not sensitive to noise. In this case to obtain a precise segmentation we need to measure the first order statistic (mean), while the variance does not provide useful information. In summary, we can conclude that first order statistic is a valid measure for region description if they have different mean values. Therefore, we adopt the piecewise constant model for the approximating function u, except in the special cases where first order statistic is not discriminant, which we treat separately.

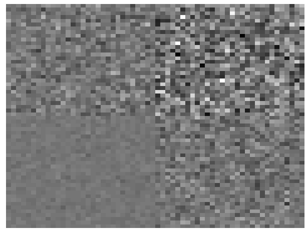
IV. From Maximum A posteriori Probability Estimation to Segmentation energy

In this section, we give an alternative derivation of segmentation energy using a stochastic image model. We follow a Bayesian approach, in which image segmentation can be regarded as a maximum a posteri probability (MAP) estimation problem.

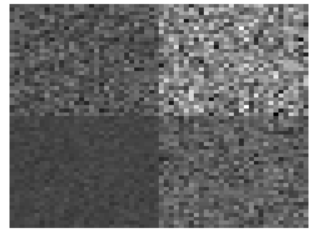
It is assumed that the entire image domain is composed of a regions $\{R_i\}$, generated by a probability distribution $p_i(U, u_i)$, where u_i are the parameters of the distribution. Here, for the sake of simplicity, we consider Gaussian distribution

$$p_i(U(x,y)|u_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{(U(x,y)-\mu_i)^2}{2\sigma_i^2}\right\}, \tag{11}$$

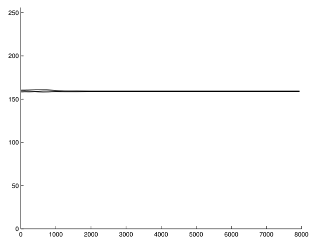
where u_i represents the mean μ_i and variance σ_i^2 of the region R_i, $u_i = (\mu_i, \sigma_i^2)$. The observed picture values are supposedly drawn from this distribution, that is, we assume that image data inside region R_i are independent and identically distributed (i.i.d) with distribution $p_i(U(x,y)|\mu_i, \sigma_i^2)$. Thus our problem



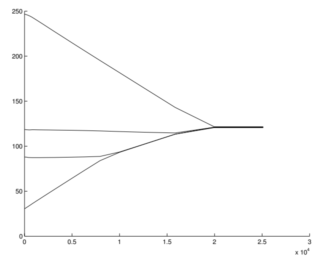
(a) Original image: $t = 0$.



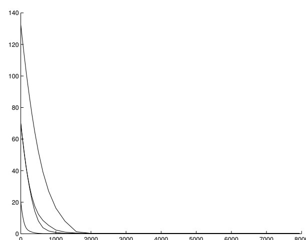
(d) Original image: $t = 0$.



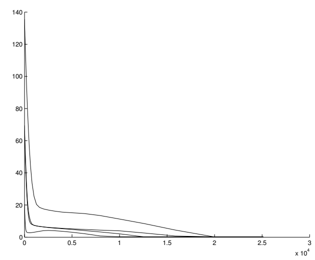
(b) Change of mean value through scales.



(e) Change of mean value through scales.



(c) Change of standard deviation through scales.



(f) Change of standard deviation through scales.

Fig. 3. Two possible region's distributions: (a) four Gaussian distributions with same mean but different variances, (d) four Gaussian distributions with different mean and different variances.

is to find the MAP estimate of the best model for a given image:

$$\begin{aligned}
\hat{u} &= \arg\max_u p(u|u_0) \\
&= \arg\max_u \{p(u_0|u)p(u)/p(u_0)\} \\
&= \arg\max_u \{\ln p(u_0|u) + \ln p(u) - \ln p(u_0)\} \\
&= \arg\max_u \{\ln p(u_0|u) + \ln p(u)\} \\
&= \arg\max_u \{E_A(u) + E_R(u)\}.
\end{aligned} \tag{12}$$

Therefore, the first term is the log likelihood function of the parameters given the data:

$$\begin{aligned}\ln L(u|u_0) &= \ln\left\{p(u_0|u)\right\} = \ln\{\textstyle\prod_{(x,y)} p(u_0(x,y)|u)\Big\} \\ &= \ln\left\{\textstyle\prod_{R_i}\prod_{(x,y)\in R_i} p_i(u_0(x,y)|u_i)\right\} \\ &= \ln\left\{\textstyle\prod_{R_i}\prod_{(x,y)\in R_i} \frac{1}{\sqrt{2\pi\sigma_i}}\exp\left\{-\frac{(u_0(x,y)-\mu_i)^2}{2\sigma_i^2}\right\}\right\} \\ &= \textstyle\sum_{R_i}\sum_{(x,y)\in R_i}\left(-\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma_i^2 - \frac{(u_0(x,y)-\mu_i)^2}{2\sigma_i^2}\right).\end{aligned} \tag{13}$$

In the previous section we adopted the piecewise constant image model that imposes zero variance over the entire image. Than, in our stochastic model the variance of each region σ_i can be approximated by a small unique constant σ and u_i becomes μ_i. Then the first term of the segmentation energy, i.e. the data fidelity term, is:

$$E_A(u) = \frac{1}{2\sigma^2}\sum_{R_i}\sum_{(x,y)\in R_i}(u_0(x,y)-u_i)^2. \tag{14}$$

In order to define the second term in the MAP model, the prior distribution of u, a common choice is a Markov random Field (MRF) model. Techniques based on MRF models have proven to be very useful in pixel labeling problems such as segmentation because they take into account spatial dependencies of neighboring pixels. Let $S = \{s = (x,y)\}$ be a two-dimensional lattice. A clique c is a subset of S in which each site is neighbor of all remaining sites (we will restrict to cliques of 4-point neighborhood) and let C be the set of all cliques. Now, if we define u as a MRF on S, its probability function is a Gibbs distribution:

$$p(u) = \frac{1}{Z}exp\left[\sum_{c\in C} V_c(u)\right], \tag{15}$$

where Z is a normalizing constant and $V_c(u)$ is the potential function defined on the cliques. In particular we will take each absolute difference of neighboring pixels as a potential function:

$$V_c(u) = \begin{cases} |u(s)-u(s')|, & s,s'\in c \\ 0, & s,s'\notin c. \end{cases}$$

Compared to the quadratic potential function, the major difference is the transition from the L^2 norm to L^1 norm which allows the sharp reconstruction of edges. So, this is a natural choice since it penalizes the complexity of the model but does not discourage abrupt discontinuities. Then, the log of the prior distribution of the MRF is:

$$E_R(u) = \ln p(u) = \frac{1}{Z}\sum_{\{s,s'\}\in c}|u(s)-u(s')|. \tag{16}$$

If we denote the total variation in the image as

$$TV(u) \equiv \int_\Omega |\nabla u|\,dx, \tag{17}$$

we can rewrite the regularization part using a discrete approximation of TV as:

$$E_R(u) = \frac{1}{Z} TV(u). \tag{18}$$

Thus our problem, ignoring the constant σ and Z, is to minimize the functional:

$$E(u) = E_A(u) + E_R(u) = \frac{1}{2} \sum_{x \in \Omega} (u - u_0(x))^2 + \lambda TV(u), \tag{19}$$

that turns out to be the same as the Tikhonov regularization functional. In other words, the segmentation problem can be formulated as total variation regularization one. The first part is recognized by many authors as the optimal way to calculate the distance between regions. First, it was proposed by Ward [27] as a total approximation error that is introduced by merging the two regions. Then, this dissimilarity measure is used by Beaulieu and Golberg in their hierarchical picture segmentation [28], it is also the first term of the Mumford and Shah functional [18], etc. The second term is mostly used in image reconstruction [29] and restoration techniques [9].

V. Spatially adaptive segmentation algorithm

In this section we give a complete description of the algorithm. We saw that the main idea is to interpret total variation regularization as a segmentation problem. The main advantage of the TV functional is that it penalizes neither discontinuities (i.e. edges) nor smooth parts of the image. This results in preserving the exact location of edges which is a highly desirable property in image segmentation. The preservation of small scale details is determined by the scale parameter that defines a trade-off between smoothing/regularizing and data fidelity term. In the case when this ratio is constant through the image, the smaller-scaled features can be lost as the effects of stronger TV regularization adapted to bigger-scaled objects. To be able to capture different scaled objects, it is necessary to adapt the amount of regularization at an image location in a spatially adaptive way. Spatial adaptivity has been extensively studied in image restoration and enhancement literature [10], [11], [12], [13]. Within a piecewise constant model, it is desirable to put relatively more weight on the regularization part while preserving small scale details imply less regularization, smaller scale parameter. To achieve this spatial adaptivity, the weighting factor is chosen to be an edge detecting function:

$$\omega = \frac{1}{\varepsilon + |\nabla \hat{u}|^2}, \tag{20}$$

where $\hat{u}$ is an estimated segmented image u in a previous segmentation step and ε is an arbitrary positive constant. The parameter ε allows a more direct local control of the adaptivity. Larger values of ε

correspond to less variation in the weighting factor and more conservative scheme. In our experiments, for images with range between 0 and 1, we used a value of $\varepsilon = 0.1$. The corresponding weighted TV norm can be written

$$TV_\omega(u) = \int_\Omega \omega(x)\, |\nabla u|\, dx. \tag{21}$$

The complete segmentation functional becomes:

$$E(u) = \frac{1}{2}\|u - u_0\|^2 + \lambda TV_\omega(u). \tag{22}$$

We applied the adaptive segmentation scheme for natural images where gradients correspond to object boundaries and the standard scheme when gradients are not necessary proportional to the likelihood of there being an edge, i.e. noisy, texture images. The functional is minimized using the standard recursive region merging algorithm explained in II-B. If we follow the variational formulation for the region growing methods [22], [30], [15], [31], [32], the merging criteria looks for an adjacent region that reduces an energy the most. Since the total variation energy functional contains a regularization parameter, the regions that reduce the energy with the smallest scale parameter will be merged first. Therefore, the weight $W_{i,j}$ between neighboring nodes V_i and V_j is defined as the scale parameter λ^* such that the change of energy introduced by merging these two nodes is equal to zero:

$$\Delta E(i,j) = E(V_i \cup V_j) - E\left(\overline{V_i \cup V_j}\right) = \Delta E_A(i,j) + \lambda^* \Delta E_R(i,j) = 0, \tag{23}$$

$$W_{i,j} = \lambda^* = -\frac{\Delta E_A(i,j)}{\Delta E_R(i,j)}. \tag{24}$$

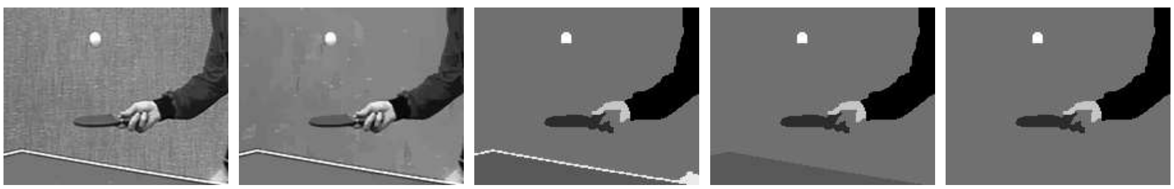
The algorithm consequently merges regions from finer to coarser scale, simplifying the image structure. This method of continually merging regions through scale, produces a hierarchy of segmentation results, detecting objects at different scales. The finest scale keeps the small and detailed features while the coarse scale corresponds to a simple topology.

VI. Simulations

In this section we illustrate our segmentation algorithm on multiple images: gray-scale, noisy, colour. It will be shown that the algorithm is applicable to images with multiple non overlapping regions characterized by different means and is not restricted to images with only two different means (objects). Moreover, up to some extension, it is possible to handle images that contain objects with the same mean but different variances and we demonstrate how to deal with this type of problems. The attractive properties of our model is that it is parameter free and not sensitive to initial condition as most of the segmentation algorithms are, where it is necessary to chose good seeds or initial curves. Another advantage of the model is that we do not need to know in advance the number of regions in order to capture them all.

A. Multiresolution approach: gray level images

Fig. 4 shows a hierarchy of segmentation results from fine to coarse scales. Multiple disjoint regions are captured up to different stage, so the multiscale nature of the objects is evident: some objects are visible only at bigger scale while other are represented only at smaller scale. The segmentation clearly delineated the main objects as: the hand, the table, the ball. Notice also that because of the high value of gradient along the fingers, they are not merged with the hand.



(a) (b) (c) (d) (e)

Fig. 4. Segmentation results obtained applying recursive region growing strategy. Number of regions: a)9999, b)999, c)7, d)6, e)5.

Next, we demonstrate on Fig. 5 how the TV scale-space evolution simplifies image structure. Comparing the Figs.4 and 5 it is clear that segmentation and diffusion processes approximate each other. This confirms our main idea to formulate segmentation technique to be dual to nonlinear diffusion filtering in terms of the energy functional.

(a) (b) (c) (d) (e)

Fig. 5. TV Scale-Space. Scale level: a) $t = 0$, b)$t = 454$, c)$t = 2212$, d)$t = 6813$, e)$t = 14452$.

B. Noisy images

We now illustrate the performance of our approach applied on a noisy image. The previous image (Fig. 4(a)), normalized between 0 and 1, is corrupted with white Gaussian noise whose standard deviation is 0.14. In this case, when an image is corrupted by noise, the adaptive edge detecting function does not correspond to objects boundaries any more. Therefore, we evolve a non-adaptive segmentation algorithm on noisy image and obtained results are shown in Fig. 6. Just as in the noise-free case, the algorithm very accurately locates the objects. Here, we point out that this denoising property of the segmentation is in accordance with the best known ROF [9] model that was originally introduced for image denoising and reconstruction. In particular, the regularization term in the functional disfavors oscillations and encourage the elimination of noise. So, we can conclude that total variation segmentation demonstrates the advantages of using L^1 norm for regularization term in denoising applications.

(a) (b) (c) (d) (e)

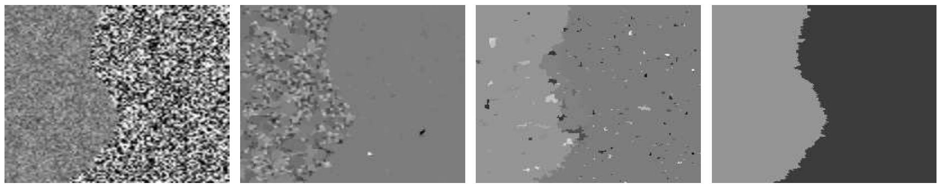
Fig. 6. Segmentation applied on noisy image. Number of regions: a)9999, b)999, c)7, d)5, e)3.

C. Images with adjacent regions with identical means but different variances

Recall that in the derivation of the model we made the assumption that the variance is constant over the entire image, i.e. regions are characterized only by their mean value. Although this model is sufficient for segmenting the whole range of natural images, it can not discriminate neighboring regions that are characterized only by different variances. To overcome this limitation we now consider different variances for each region σ_i^2. Then, the natural logarithm of the likelihood function becomes:

$$\begin{aligned}\ln\left(L(u|u_0)\right) &= ln\left(\prod_{R_i}\prod_{(x,y)\in R_i}\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left\{-\frac{(u_0(x,y)-\mu_i)^2}{2\sigma_i^2}\right\}\right)\\ &= \sum_{R_i}\sum_{(x,y)\in R_i}\left(-\frac{1}{2}\ln 2\pi-\frac{1}{2}\ln\sigma_i^2-\frac{(u_0(x,y)-\mu_i)^2}{2\sigma_i^2}\right)\\ &= -\frac{n}{2}\ln 2\pi-\frac{1}{2}\sum_{R_i}|R_i|\ln\sigma_i^2-\frac{n}{2}.\end{aligned}$$

This defines the first term of the segmentation energy, the approximation part, while the regularization term remains the total variation. Then, we proceed with the same region merging algorithm to minimize this modified energy. To examine the correctness of this model, we chose an image taken from the literature [15], composed of two regions generated randomly from two Gaussian distributions with identical means: $N(128, 10^2)$ and $N(128, 35^2)$, separated by a S-shaped curve. The segmentation results are presented in Fig. 7.



(a) (b) (c) (d)

Fig. 7. Segmentation results obtained applying region growing strategy up to different number of regions: a)31622, b)5011, c)251, d)2.

D. Vector-Valued Images

Our segmentation algorithm can be very easily generalized for vector-valued images, where each pixel is a vector belonging to $R^M, M \geq 1$. The components of the vector could correspond to red, green, and blue intensity values in color images, or to the feature channels gathered from analyzing a texture image, etc. We view the segmentation problem in terms of solving the nonlinear optimization problem :

$$E(u) = \frac{1}{2}\int_\Omega ||u - u_0||^2 dx + \lambda \int_\Omega \omega(x)||\nabla u|| dx.$$

For scalar valued functions, the norm $||.||$ is simply the absolute value and for vector-valued images the norm can be interpreted as the l^2 norm:

$$||u_i - u_j|| = \sqrt{\sum_{k=1}^{M}(u_{i,k} - u_{j,k})^2}.$$

Just as for scalar images, we use the same principle to merge two neighboring regions described in section V. Two color image experiments are shown in Fig. 8.

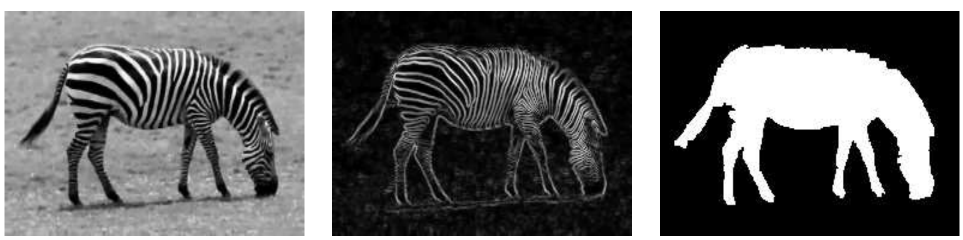
E. Texture Segmentation

A natural approach to texture segmentation is to first represent texture by feature descriptors and then to apply a vector-valued segmentation scheme. It is clear that the quality of the segmentation will depend on the extraction of good features for texture discrimination. However, the problem of optimal feature

(a) (b) (c) (d)

Fig. 8. Segmentation of a color images. a)Original image of a hand. b)Segmented image of a hand (6 regions). c)Original image of a woman. d)Segmented image of a woman (9 regions).

extraction is beyond the main interest of the presented work. Our goal here is to show the feasibility of texture segmentation using the total variation segmentation scheme. In order to do this, we use a simple set of features consisting of the original image and the magnitude of its gradient (Fig. 9 (a)(b)). The first channel contains a complete textural information and the second channel captures the local orientation of the texture elements. Then, our algorithm proceeds as in the previous case by coupling these two feature channels and simultaneously merging regions (Fig. 9(c)). Although, this example uses



(a) (b) (c)

Fig. 9. Feature channels: a) original image, b) magnitude of the image gradient. c) Segmented image.

only two texture features, it yields very good results for most cases of two texture discrimination. For more complicated texture images we consider the texture features generated by Gabor filters [33], [34], [35]. A particular Gabor elementary function can be used as the mother wavelet to generate a whole family of Gabor wavelets. We use a particular class of 2D Gabor wavelets proposed in [34]. Using this 2D Gabor-wavelet transform, images are decomposed into several channel outputs. For the image presented on Fig. 11(a) which is composed of five textures, we can see the magnitude response of the Gabor filter bank on Fig. VI-E.

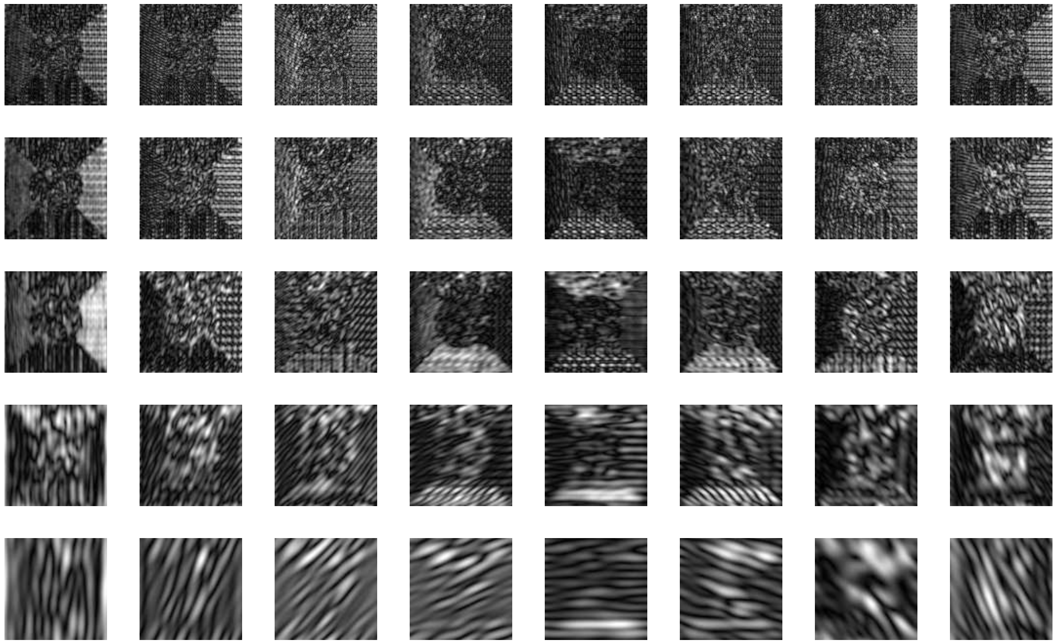
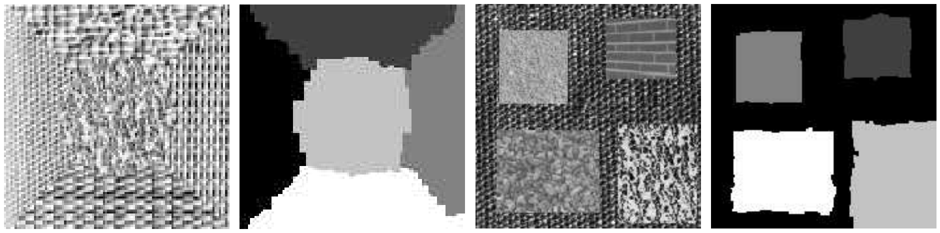


Fig. 10. Gabor Space.

Since the five texture images consist of dominant patterns (i.e. textures are localized frequency components in the form of Gabor elementary functions), it is evident that multi-channel Gabor filtering represents an effective way to capture textural information. However, all filtered channels do not result in good texture discrimination. It is assumed that the strength of the responses in regions of local coherent texture is large and spatially homogeneous which results in high variance. So, the variance of the filtered images is used to perform a systematic filter selection. We use only a subset of filtered images that have variances greater than the average variance calculated over entire Gabor Space. Then, the original texture image is replaced by a multi-valued image containing the selected subset of filtered images. Now, we can directly apply the previously defined vector-valued segmentation scheme to preform texture segmentation. Our results are presented on Fig. 11.

VII. Conclusions

In this report, the total variation approach to unsupervised image segmentation is introduced. We have shown that the TV regularization can be regarded as both a geometry-driven diffusion scheme, as well as a segmentation one. We have used the standard Total Variational functional to determine a merging



(a) (b) (c) (d)

Fig. 11. Segmentation of texture images. Original images: a), c). Segmented images: b) 5 regions, d) 5 regions.

predicate to group regions. One of the main benefits of this segmentation process is that it combines the most attractive properties of the Total Variation and region merging such as edge preservation, robustness to noise, parameter reduction and low algorithm complexity. Furthermore, we achieved a greater scale flexibility by using a spatially varying regularization term. Ultimately, we interpreted our approach in the probabilistic settings, as the log-likelihood of the observed image data given a Markov random Field model. Finally, we validated the effectiveness of our algorithm by segmenting gray level, color and texture images.

Acknowledgments

Ana Petrovic would like to acknowledge the support of the Swiss National Science Foundation through grant number 200020-105177/1. The authors would also like to thank Xavier Bresson for fruitful discussions.

References

[1] A. Petrovic, O. Divorra, and P. Vandergheynst, "Multiresolution segmentation of natural images: From liner to nonlinear scale-space representation," *IEEE Trans. on Image Processing*, vol. 13, no. 8, August 2004.

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

[3] B. M. ter Haar Romeny, Ed., *Geometry-Driven Diffusion in Computer Vision*. Kluwer Academic Publishers, 1994.

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

[5] I. Pollak, A. S. Willsky, and H. Krim, "Image segmentation and edge enhancement with stabilized inverse diffusion equations," *IEEE Trans. on signal Processing*, vol. 9, no. 2, pp. 256–266, February 2000.

[6] 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.

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

[8] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazon, "Minimizing total variation flow," *Differential and Integral Equations*, vol. 14, no. 3, pp. 321–360, March 2001.

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

[10] M. R. Banham and A. K. Katsaggelos, "Spatially adaptive wavelet-based multiscale image restoration," *IEEE Trans. on Image Processing*, April 1996.

[11] J. Biedmond and R. Lagendijk, *Iterative Identification and Restoration of Images*, Dordrecht, Ed. Kluwer Academic Publishers, 1991.

[12] D. M. Strong and T. F. Chan, "Spatially and scale adaptive total variation based regularization and anisotropic diffusion in image processing," UCLA, Tech. Rep., 1997.

[13] P. Blomgren, "Total variation methods for restoration of vector valued images," Ph.D. dissertation, UCLA, 1998.

[14] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE PAMI*, vol. 6, no. 6, pp. 721–741, 1984.

[15] S. C. Zhu and A. L. Yuille, "Region competition and its analysis: A unified theory for image segmentation," Harvard Robotics Laboratory, Tech. Rep. 95-7, 1995.

[16] C. Bouman and M. Shapiro, "A multiscale random field model for bayesian image segmentation," *IEEE Trans. on Image Processing*, vol. 3, no. 2, pp. 162–177, March 1994.

[17] D. Terzopoulos, "Regularization of inverse problems involving discontinuities," *IEEE PAMI*, vol. 8, no. 4, 1986.

[18] 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.

[19] Y. G. Leclerc, "Constructing simple stable descriptions for image partitioning," *International Journal of Computer Vision*, vol. 3, no. 1, pp. 73–102, 1989.

[20] A. A. Amini, T. E. Weymouth, and R. Jain, "Using dynamic programming for solving variational problems in vision," *IEEE Trans. on PAMI*, vol. 12, no. 9, pp. 855–867, September 1990.

[21] A. Blake and A. Zisserman, *Visual Reconstruction.* Cambridge: MIT Press, 1987.

[22] G. Koepfler, C. Lopez, and J. Morel, "A multiscale algorithm for image segmentation by variational method," *SIAM J. Numer. Anal.*, vol. 31, no. 1, 1994.

[23] C. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Integral Equations of the First Kind.* Boston, MA: Pitman, 1984.

[24] T. Brox, M. Welk, G. Steidl, and J. Weickert, *Equivalence results for TV diffusion and TV regularization.* Berlin: Spinger, June 2003, vol. 2695, ch. Lecture Notes in Computer Science, pp. 86–100.

[25] I. Pollak, A. S. Willsky, and Y. Huang, "Nonlinear evolution equations as fast and exact solvers of estimation problems," *IEEE Trans. on signal Processing*, vol. 53, no. 2, pp. 484–498, February 2005.

[26] D. M. Strong and T. F. Chan, "Exact solutions to total variation regularization problems," UCLA Math Department, CAM Report 96-41, October 1996.

[27] J. H. Ward, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, March 1963.

[28] J.-M. Beaulieu and M. Goldberg, "Hierarchy in picture segmentation: A stepwise optimization approach," *IEEE PAMI*, vol. 11, no. 2, pp. 150–163, February 1989.

[29] C. Bouman and K. Sauer, "An edge-preserving method for image reconstruction from integral projections," in *Info. Sci. and Syst.*, Baltimore, March 1991, pp. 383–387.

[30] C. Ballester, V. Caselles, and M. Gonzalez, "Affine invariant segmentation by variational method," *SIAM J. Applied Math.*, vol. 56, pp. 294–325, 1996.

[31] P. Felzenszwalb and D. Huttenlocher, "Image segmentation using local variation," in *In. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 1998, pp. 98–104.

[32] N. Redding, D.Crisp, D. Thang, and G. Newsam, "An efficient algorithm for mumford-shah segmentation and its application to sar imagery," in *In Proc. of the Digital Image Computing Techniques and applications*, Perth, Austarlia, 1999, pp. 35–41.

[33] D. Gabor, "Theory of communication," *J. Inst. Elec. Eng.*, vol. 93(III), pp. 429–457, 1946.

[34] T. S.Lee, "Image representation using 2d gabor wavelets," *IEEE Trans. on PAMI*, vol. 18, no. 10, pp. 1–13, october 1996.

[35] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition*, vol. 12, pp. 1167–1186, 1991.