UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Spatially and Scale Adaptive Total Variation Based Regularization and Anisotropic Diffusion in Image Processing

David M. Strong
Tony F. Chan

November 1996

CAM Report 96-46

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90024-1555

Spatially and Scale Adaptive Total Variation Based Regularization and Anisotropic Diffusion in Image Processing *

David M. Strong † Tony F. Chan ‡

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90095-1555

Abstract

In image processing, it is often desirable to remove noise, smooth or sharpen image features, or to otherwise enhance the image. Total Variation (TV) based regularization is a model case of geometry-driven diffusion for image processing. In our papers [14] and [15], we analyze the precise effects of TV based regularization by analytically finding exact solutions to the TV regularization problem. In this paper, we used these results to develop adaptive TV regularization schemes, as well as a scheme for automatic scale recognition. We develop our results by considering the unconstrained (Tikhonov) formulation of the TV regularization problem, in which a regularization parameter must be chosen to determine the balance between the goodness of fit to the original (e.g. measured) data and the amount of regularization to be done to the image. We first discuss how to choose the regularization parameter based on the local features of objects in the image by using the theory presented in [15]. We then discuss how to choose the regularization parameter based on the noise level in the image, and we finally discuss how to develop adaptive regularization schemes which are driven by both features and noise level in the image. We discuss several heuristics and give two schemes for doing adaptive regularization based on our results, which use *a priori* information to manually define a spatially varying regularization parameter. Using the same theory, we also give a simple scheme for local scale recognition. Our results are equally applicable to the constrained formulation of the TV regularization problem, as well as to other geometry-driven diffusion schemes, as these other schemes can be thought of as variations or modifications of the model TV regularization scheme.

*This work is supported by the ONR under contract ONR N00014-96-1-0277, and by the NSF under grant NSF DMS-9626755.

†`dstrong@math.ucla.edu`, `http://www.math.ucla.edu/~dstrong`

‡`chan@math.ucla.edu`, `http://www.math.ucla.edu/~chan`

1 Introduction and Motivation

A typical problem in image processing is to find an approximation u to true image given the noisy measured data z. In this paper, we will treat the non-blurred case (although our results can be extended to the case where blur is present), so that the corresponding equation is

$$u + \epsilon = z,$$

where ϵ represents the error or degradation—the "noise"—in the measured data. Often our task is to remove the noise or otherwise enhance the image by choosing u to be near the measured image z, while minimizing a functional $R(u)$, which is chosen to control unwanted characteristics of u, such as the irregularity or oscillatory behavior of u. One formulation of this problem is the *unconstrained* problem

$$\min_u \ \tfrac{1}{2}\|u - z\|_\Omega^2 + \alpha\, R(u), \tag{1}$$

where α is a positive parameter that determines the trade-off between goodness of fit to the measured data, and the amount of regularization done to the measured image z. Another common formulation of this problem is the *noise-constrained* problem

$$\min_u R(u) \quad \text{subject to} \quad \|u - z\|_\Omega^2 = \sigma^2, \tag{2}$$

where the error (noise) level σ is assumed to be known. For both problems, $\|u - z\|_\Omega^2 = \int_\Omega [u(\vec{x}) - z(\vec{x})]^2\, d\vec{x}$. We note that solving (1) is equivalent to solving (2) when $\alpha = \frac{1}{\lambda}$, where λ is the Lagrange multiplier found in solving (2).

More recently, PDE-driven approaches to image processing have been introduced and studied. In [11], it is proposed to time-march the following anisotropic *diffusion* equation

$$u_t = \nabla \cdot (g(|\nabla u|)\nabla u), \tag{3}$$

using $u(\vec{x}, t = 0) = z(\vec{x})$, the measured image, as the initial condition. The basic idea is that if $g(|\nabla u|)$ is decreasing in $|\nabla u|$, then the image is diffused more where no edges are present and less where edges are present. One of the main benefits of this diffusion process is excellent edge preservation in the image, with removal of noise.

TV regularization can be regarded as both a image regularization scheme, as well as a geometry-driven diffusion scheme. This dual nature of TV regularization is discussed in Section 2. In Section 3, we summarize the results of [15], in which the precise effects of TV regularization are mathematically established by analytically finding exact solutions to the TV regularization problem. In Section 4, we discuss how to choose the regularization parameter α based on the images features, and we discuss the relationship between the scale and the regularization parameter. We also discuss choice of regularization parameter based on the noise in the image. In Section 5, we give two formulations of an adaptive regularization scheme, as well as an automatic scale recognition scheme, motivated by the results of Sections 3 and 4.

2 Total Variation Based Regularization

Often the functional $R(u)$ in (1) or (2) is taken to have the form $R(u) = \|Qu\|^2$, where Q is a *linear* operator. For example, $Q = I$, $Q = \Delta$ (the Laplacian operator) and $Q = \nabla$ (the gradient operator) are often used (cf. [7]). Unfortunately, linear operators are limited in their effectiveness, particularly when there are edges present in the image, and *non-linear* operators are often required. The following non-linear total variation (TV) regularization functional has recently been proposed in [13] as a choice of $R(u)$ for use in image regularization:

$$TV(u) \equiv \int_\Omega |\nabla u(\vec{x})| \, d\vec{x} \,. \tag{4}$$

The functional $TV(u)$ simply measures the total variation of u.[1] With $R(u) = TV(u)$, equations (1) and (2) then become, respectively:

$$\min_u \; \tfrac{1}{2}\|u - z\|_\Omega^2 + \alpha\, TV(u), \tag{5}$$

and

$$\min_u TV(u) \quad \text{subject to} \quad \|u - z\|_\Omega^2 = \sigma^2. \tag{6}$$

The main advantage of TV regularization is that it does not penalize discontinuities (i.e. edges) in u. In fact we have shown in [15] that TV regularization preserves edges *exactly.* On the other hand, TV has the advantage that it does not penalize *smooth* images. (This is not quite true when a smooth image, which has been contaminated with noise—leaving it no longer smooth—is TV regularized.) with noise, resulting in a rough image.) The TV-based image restoration approach looks for an approximation to the observed image which has minimal *total variation.* There is no particular bias towards a discontinuous or smooth solution. The measured data z, as well as the regularization parameter α (when solving (5)) or the noise level σ^2 (when solving (6)), will determine the sharpness or smoothness of the restored image. See [15] for a more detailed explanation of the non-biased (i.e. toward sharp or toward smooth edges) nature of the TV functional (4).

In [6] and [9], it is shown that a unique solution to the TV regularization problems (5) and (6) exists under certain conditions. This paper will not discuss numerical methods for solving (5) or (6). For some suggested methods see [3], [4], [13] or [17] .

2.1 TV Regularization as a Special Case of Anisotropic Diffision

Total Variation regularization can be viewed as a special case of the more general class of geometry-driven diffusion schemes previously mentioned. To solve (5), we differentiate (with respect to u) the functional to be minimized in (5), and set the result equal to 0 to get

$$\alpha \nabla \cdot (\frac{\nabla u}{|\nabla u|}) - (u - z) = 0. \tag{7}$$

[1] This definition of $TV(u)$ is only valid for differentiable functions u. We must interpret $\nabla u(\vec{x})$ as the distribution derivative in order for (4) to be valid for non-differentiable functions.

In [13], a time-marching scheme is proposed to solve (1) by marching the following PDE to steady state:

$$u_t = \alpha \nabla \cdot (\frac{\nabla u}{|\nabla u|}) - (u - z) \qquad \text{with } u(\vec{x}, t = 0) = z(\vec{x}). \tag{8}$$

As with other geometry-driven diffusion schemes, the first term on the right-hand-side of (8) can be viewed as a nonlinear diffusion operator with diffusion coefficient $\frac{1}{|\nabla u|}$ which diffuses less where the gradient is large (e.g. near edges). One could also diffuse without the fitting constraint by taking

$$u_t = \nabla \cdot (\frac{\nabla u}{|\nabla u|}) \qquad \text{with } u(\vec{x}, t = 0) = z(\vec{x}). \tag{9}$$

In this form we see that TV regularization is a special case of geometry-driven diffusion. To see the direct connection between TV regularization and TV diffusion, notice that we can re-write (7) as

$$u = z + \alpha \nabla \cdot (\frac{\nabla u}{|\nabla u|}) \tag{10}$$

so that solving (5) is equivalent to doing a single *implicit* step of time-marching using (9) with step size of α. If we time-march *explicitly* using (9) for n time steps with time increment $\triangle t$, then the resulting image will be approximately equal to the solution of (5) if we have $n \triangle t = \alpha$.

One can consider a general function $f(|\nabla u|)$ in the integrand in (4). For example, if $f(|\nabla u|) = |\nabla u|^n$, then for $n > 1$ smooth edges are preferred, resulting in loss of sharp discontinuities in the image. On the other hand, if $n < 1$ then sharp edges are preferred, which can lead to unwanted edges, for example staircasing (and moreoever, the minimization problem is no longer even convex). The only unbiased choice would be $f(|\nabla u|) = |\nabla u|$, as is the case for TV regularization. See [18] for further discussion of choosing $f(|\nabla u|)$.

We can also generalize from (9) back to (3), where $g(|\nabla u|)$ is chosen as a decreasing function of $|\nabla u|$. Several choices of $g(|\nabla u|)$ have been given in looking at geometry-driven diffusion. The simplest, and the most unbiased, choice of a non-negative function $g(|\nabla u|)$ decreasing in its non-negative argument is to choose $g(|\nabla u|) = \frac{1}{|\nabla u|}$, which is inherent and automatic when doing TV regularization. This corresponds to choosing $f(|\nabla u|) = |\nabla u|$ in the integrand of (4), as just discussed.

We can consequently view TV regularization as a model or canonical case of geometry-driven diffusion, of which other geometry-driven diffusion schemes are a variation or modification. Because it can be analyzed in its variational form (5), it turns out that the effects of TV regularization can be understood more precisely than the effects of other geometry-driven schemes. This understanding is in fact developed in [15] by finding exact solutions to the TV regularization problem by solving the variational formulation of the problem. Additionally, because TV regularization can be viewed as a model geometry-driven diffusion scheme, the understanding and schemes we develop for TV regularization can be extended to other geometry-driven diffusion schemes.

2.2 Need for Better Understanding of TV Regularization

There is a need for a more precise understanding of how TV regularization (and more generally, geometry-driven diffusion) affects an image. Consider the problem of removing noise from the one-dimensional image shown in Figure 1. We must find u, the minimizer of (5), which will depend on z and our choice of α. For α too small, not enough smoothing is done to remove the noise, while for α too large, more noise is removed but we lose too much of the detail. As α becomes larger, increasingly larger scaled features are lost due to the condition of trying to lower the total variation of the image. In this sense, TV regularization is a type of multi-scale regularization, with the scale determined by α. Larger values of α when solving the unconstrained problem (5) correspond to higher estimates σ^2 of the noise level when solving (6), and to longer time-marching when solving (9). The key is to find an appropriate value of α (whether constant or varying), either based on the amount of detail we wish to preserve in the image, or based on the amount of noise (whether known *a priori* or estimated given the noisy image), or based on both factors.

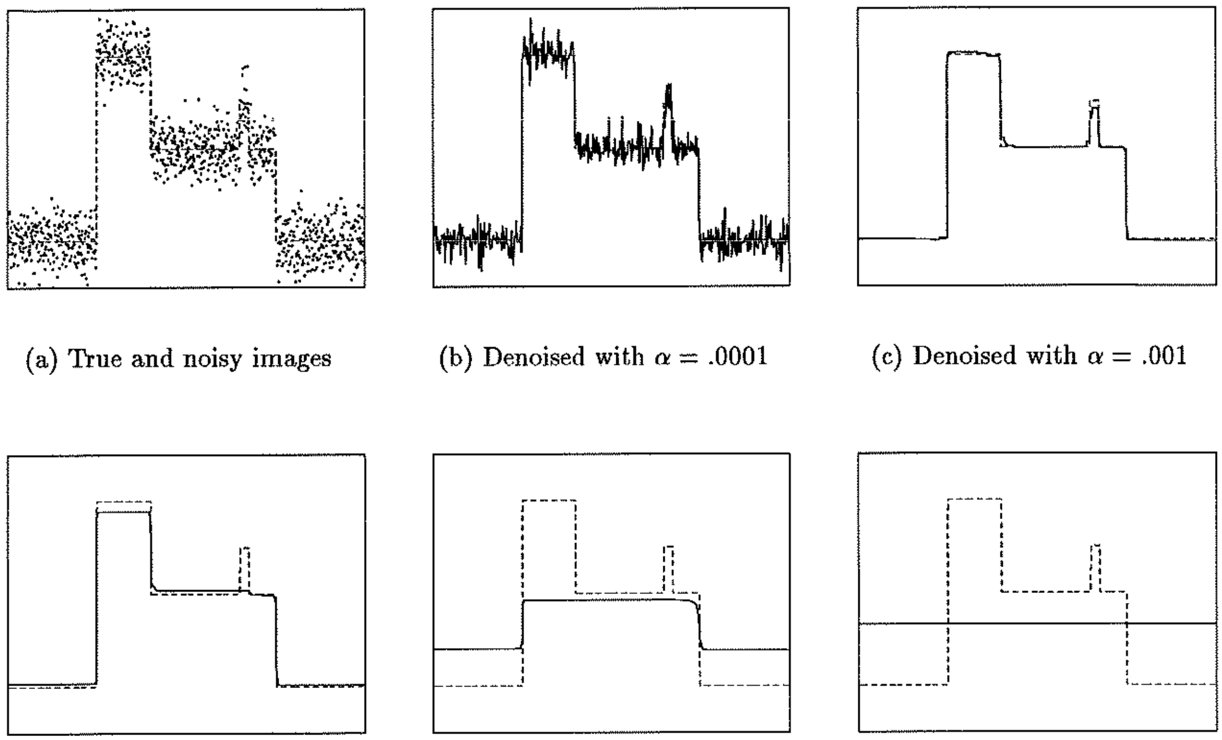
In order to determine the "optimal" value of α, it is necessary to establish a relationship between the regularization parameter α and the amount of smoothing which occurs. In [5], it is suggested that the effectiveness of TV based noise removal will depend on the "mass" (essentially the area and intensity level of the object) of the image relative to the total variation of the image. It would be very useful to define a more precise relationship between the regularization parameter α and the amount of smoothing which occurs to an object of a given scale. With a more precise relationship defined, it becomes possible to make a more intelligent and useful choice of α, whether α is constant or $\alpha = \alpha(\vec{x})$ is spatially varying. This we did in [15], and we summarize our results in Section 3.

2.2.1 Spatially Adaptive Image Restoration

Finally, image restoration and enhancement can be more effective if done in a spatially adaptive way. There is generally a trade-off between noise removal and detail preservation, and images are typically comprised of multiple objects of different spatial scales. In removing noise from an image, the details and/or contrast are often reduced and can be lost completely. A natural approach to partially alleviate this problem is to use spatial adaptivity. In general, less smoothing is desired for features of smaller scale, while more smoothing is appropriate for features of larger scale. Figure 1 demonstrates the effects of varying amounts of *non-adaptive* regularization.

Spatial adaptivity has been studied extensively in image restoration literature; see for example [1, 2]. Multi-scale TV regularization has been proposed in [12]; however, often it can be useful to do adaptive regularization with more direct local control of the adaptivity, and with a better understanding of how the image is being affected. In the context of TV restoration, one basic way of achieving spatial adaptivity would be by modifying equation (5) to

$$\min_u \quad \tfrac{1}{2}\|u - z\|_\Omega^2 + \int_\Omega \alpha(\vec{x})|\nabla u(\vec{x})|\, d\vec{x}. \tag{11}$$



(a) True and noisy images

(b) Denoised with $\alpha = .0001$

(c) Denoised with $\alpha = .001$

(d) Denoised with $\alpha = .01$

(e) Denoised with $\alpha = .1$

(f) Denoised with $\alpha = 1$

Figure 1: A comparison of using different values of α in applying TV regularization to a one-dimensional image. The dotted line is the noisy image, the dashed line is the true image, and the solid line is the regularized image. An unintelligent choice of α can result in too little noise removal, or too much detail loss, as more detail is lost with increasing values of α.

Similarly, we could solve a modified form of the constrained problem (6),

$$\min_u \int_\Omega \alpha(\vec{x}) |\nabla u(\vec{x})| \, d\vec{x} \quad \text{subject to} \quad \|u - z\|_\Omega^2 = \sigma^2. \tag{12}$$

Both (11) and (12) are equivalent to time-marching to steady state the adaptive version of (8)

$$u_t = \nabla \cdot (\alpha(\vec{x}) \frac{\nabla u}{|\nabla u|}) - (u - z) \qquad \text{with} \;\; u(\vec{x}, t = 0) = z(\vec{x}). \tag{13}$$

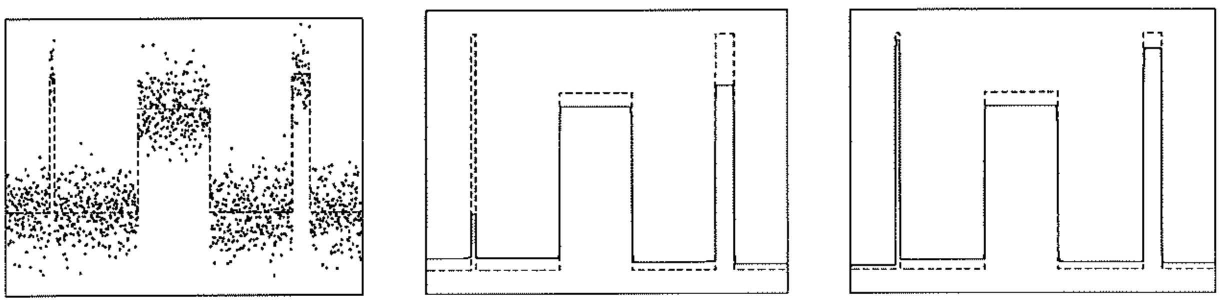
We could similarly modify (3) and (9):

$$u_t = \nabla \cdot (\alpha(\vec{x}) \, g(|\nabla u|) \nabla u) \qquad \text{with} \;\; u(\vec{x}, t = 0) = z(\vec{x}). \tag{14}$$

$$u_t = \nabla \cdot (\alpha(\vec{x}) \frac{\nabla u}{|\nabla u|}) \qquad \text{with} \;\; u(\vec{x}, t = 0) = z(\vec{x}). \tag{15}$$

For (14) and (15), instead of spatially varying $\alpha(\vec{x})$, we could instead choose to spatially vary the time step $\Delta t = \Delta t(\vec{x})$ when time-marching.

In each of these problems, $\alpha(\vec{x})$ can be chosen to spatially vary based on the local features in the image as well as on the information to be extracted from the image. In order to realize effective choices for $\alpha(\vec{x})$ in (11) - (15), we need to study the proper adaptive selection of α. The theory developed in this paper provides a simple foundation upon which to appropriately base our choice of $\alpha(\vec{x})$ in doing adaptive regularization. See Figure 2 for a simple example of adaptive regularization where the choice of α was based on the results subsequently developed in this paper.



(a) True and noisy images

(b) True and denoised images, *without* adaptivity

(c) True and denoised images, *with* adaptivity

Figure 2: A comparison of non-adaptive and adaptive TV regularization. Adaptive regularization gives better results. The dotted line is the noisy image, the dashed line is the true image, and the solid line is the denoised image.

3 Exact Solutions of TV Regularization Problems

In this section we give a summary of the results of [15], in which we analytically find the exact solutions to the TV regularization problem (5). These results, which were first explored in [14], give us an exact relationship between the regularization parameter α and the effect of TV regularization on an object or image feature of any given scale. The results can then be used to solve the adaptive TV regularization problems (11) - (15). All mathematical proofs, as well as more in-depth intuitive explanations of the results, can be found in [15].

We first give results for piecewise constant images, and then for simple smooth images. These exact results are given for simple cases, but can be extended to more complex images.

3.1 Piecewise Constant Images

We first give analytic results for piecewise constant images. Figure 3 illustrates the effects of TV regularization applied to a one-dimensional piecewise constant image.

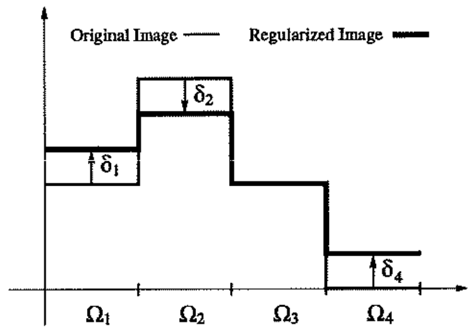


Figure 3: One-dimensional piecewise constant image, before and after TV regularization.

All one-dimensional piecewise constant images are comprised of three types of regions, which we label as *extrema*, *steps* and *boundaries*. In [15], we proved that

$$
\begin{array}{rcllll}
\delta_i & = & \frac{2}{|\Omega_i|}\alpha & & & \text{“Extremum” Region } (\Omega_2) \\
\delta_i & = & \frac{0}{|\Omega_i|}\alpha & = & 0 & \text{“Step” Region } (\Omega_3) \\
\delta_i & = & \frac{1}{|\Omega_i|}\alpha & & & \text{“Boundary” Region } (\Omega_1, \Omega_4)
\end{array}
$$

These results are valid whether the image before regularization is noise-free (as Figure 3 illustrates), or *noisy.*

We next give results for piecewise constant two- or three-dimensional images. Figure 4 illustrates the effects of TV regularization on a two- or three-dimensional piecewise constant image.

$$
\begin{array}{rcllll}
\delta_i & = & \frac{|\partial\Omega_{i,i+1}|+|\partial\Omega_{i,i-1}|}{|\Omega_i|}\alpha & = & \frac{|\partial\Omega_i|}{|\Omega_i|}\alpha & \text{“Extremum” Region } (\Omega_1, \Omega_2) \\
\delta_i & = & \frac{|\partial\Omega_{i,i+1}|-|\partial\Omega_{i,i-1}|}{|\Omega_i|}\alpha & & & \text{“Step” Region } (\Omega_3) \\
\delta_i & = & \frac{|\partial\Omega_{i,i-1}|}{|\Omega_i|}\alpha & & & \text{“Boundary” Region } (\Omega_4)
\end{array}
$$

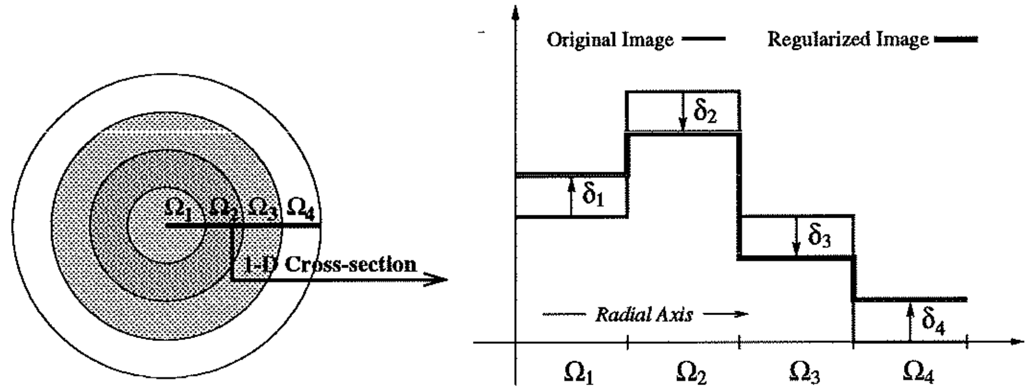
We found these analytic results for two- and three-dimensional images for the noise-free case, although numerical results show the behavior in the noisy case is essentially as predicted by these formulae. Also, these exact results are for *radially symmetric* piecewise constant images. One can have essentially the same results even in non-symmetric images, if appropriately modified (e.g. minmod [13]) approximation schemes are used.

Perhaps the most important case is in the so-called extremum regions, where

$$\delta = \frac{|\partial\Omega|}{|\Omega|}\alpha. \tag{16}$$

If we define the scale, s, of the object as the ratio of the area (volume in 3-D) of an object to its boundary length (surface area in 3-D).

$$s = \frac{|\Omega|}{|\partial\Omega|} \tag{17}$$



(a) A Two-dimensional image, or the two-dimensional cross section of a three-dimensional image

(b) The one-dimensional cross-section of (a), before and after TV regularization

Figure 4: A two- or three-dimensional image, before and after TV regularization.

then (16) can be rewritten as

$$\delta = \frac{\alpha}{s} \tag{18}$$

so that the change δ in image intensity is inversely proportional to the scale in the image. This explains why (smaller-scaled) noise is removed while (generally larger-scaled) image features are relatively unaffected when doing TV regularization. Conversely, if we can locally measure the change in intensity level due to TV regularization, then we can find the scale, by re-writing (18) as

$$s = \frac{\alpha}{\delta}. \tag{19}$$

3.2 Smooth Radially Symmetric Objects

We next give results from [15] for images with smooth edges (i.e. without discontinuities). Figure 5 illustrates the effects of TV regularization on a simple one-dimensional smooth image.

One approach in developing results in the smooth case is to treat it as the limit of the piecewise constant case. Assuming the regularized image will be as show in Figure 5, we find that

$$\delta_i = -m|\Omega_i| - \sqrt{(m|\Omega_i|)^2 + 2m\alpha} \tag{20}$$

and

$$\begin{aligned} \tilde{x}_1 &= x_1 - \frac{\delta_1}{m}, \\ \tilde{x}_2 &= x_2 + \frac{\delta_2}{m}. \end{aligned} \tag{21}$$

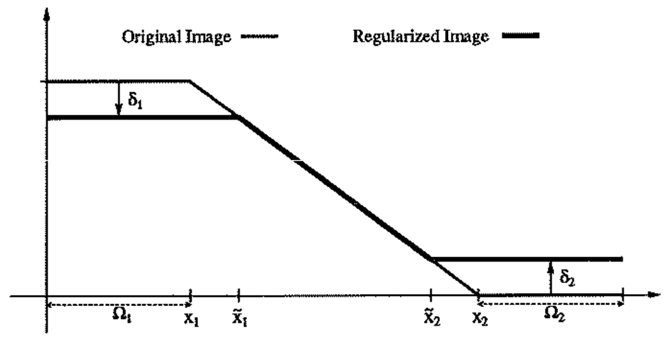


Figure 5: A smooth one-dimensional image, before and after TV regularization.

These results can be extended to more complicated one-dimensional images.

We next give results for two and three dimensions. Figure 6 illustrates the effects of TV regularization on a simple two- or three-dimensional smooth image. We find that

$$\delta(r) = \begin{cases} \frac{1}{r}\alpha & \text{in 2-D} \\ \frac{2}{r}\alpha & \text{in 3-D} \end{cases}$$

There is no explicit formula for finding the values of $\tilde{r}_1, \tilde{r}_2, \delta_1$ and δ_2, which must be found

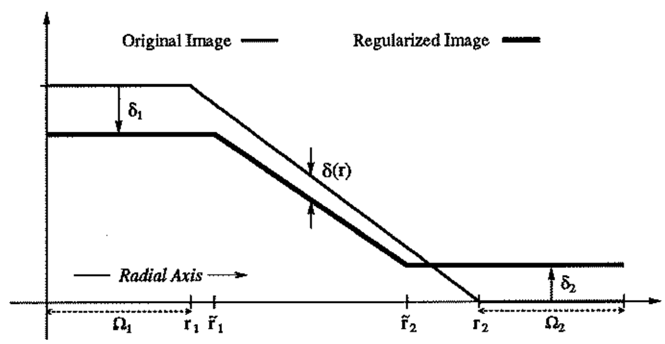


Figure 6: A one-dimensional cross-section of a two- or three-dimensional image, before and after TV regularization.

numerically, as discussed in [15]. Their values, however, are similar to those found for the one-dimensional case, as given in (20) and (21).

These results for two and three dimensions are again for radially symmetric images. The results are approximately the same for non-symmetric images, when an appropriate modified approximation scheme is used. For the preceeding smooth cases, the results are exact for noise-free images. The behavior in the noisy case is essentially as predicted by these formulae. See examples given in [15].

Finally, since any radially symmetric object is made up of piecewise constant regions and smooth regions, the results given in this section could be extended to any radially symmetric object, although the more complicated the image, the more complicated it becomes to extend the results. As described in this section, the piecewise constant regions would remain piecewise constant, sharp edges (i.e. discontinuities) would remain sharp, and smooth edges or changes in intensity would retain essentially the same shapes, with miminal change in edge position.

4 Heuristics and Applications of Exact Solutions of Regularization Problems

The results given in Section 3 can be useful in developing adaptive or multi-scale image regularization schemes, as well as in doing non-adaptive regularization more intelligently. When solving the modified unconstrained problem (11), the relationship between scale and intensity change due to regularization can be used in a very straightforward way to define a spatially varying value of $\alpha(\vec{x})$. When solving the modified constrained problem (12), we can use the regularization-scale relationship to determine more precisely the relative effect of the regularization on different image features, and we can consequently determine in a locally adaptive way the desired balance between fitting error and total variation of u. Also, when time-marching using (15), we can use our results to determine how to locally weight the diffusion of the image.

It is not our purpose to give an exhaustive discussion of how best to incorporate the results given in Section 3 into adaptive TV regularization schemes. We will, however, give some basic ideas which can be used as building blocks for adaptive regularization (denoising) and scale recognition schemes. In Section 5, we will also give two specific schemes which are built from these basic ideas.

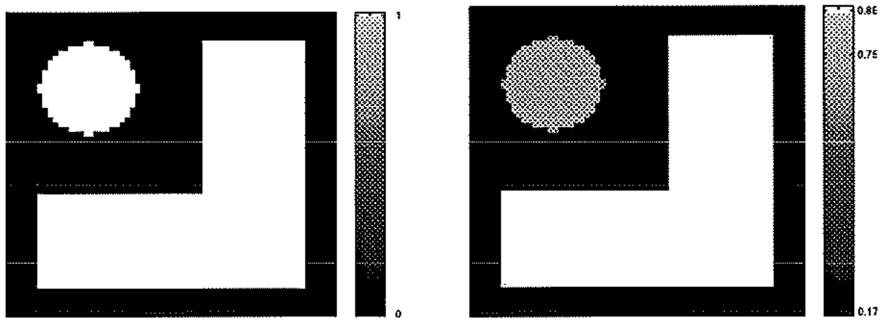
In this section, we first discuss how one might choose $\alpha^{scale} = \alpha(\vec{x})$ based on the scale of the features in the image, where the choice of α^{scale} is based on the results of Section 3. We then discuss how one could choose a $\alpha = \alpha^{noise}$ based solely on the noise level. We then give a few heuristics for choosing $\alpha(\vec{x})$ as a function of both the noise level and the scale of local features of the image.

4.1 Choosing the Regularization Parameter Based on Scale

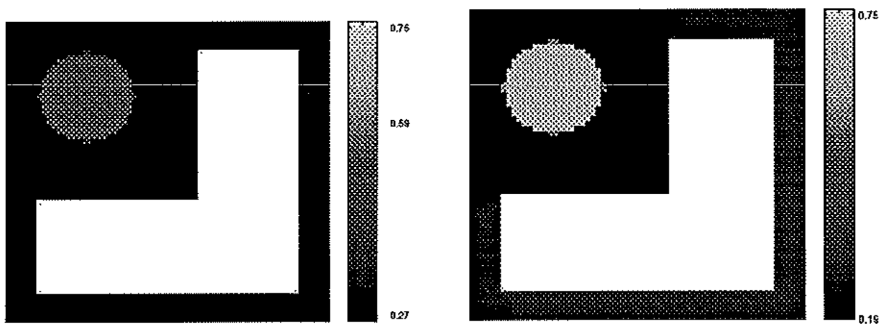
We first discuss how one might adaptively choose $\alpha(\vec{x})$ according to the image features, based on the results of Section 3. Applying TV regularization to a noisy image generally results in the removal of noise, with some reduction in contrast in the image due to changing levels of image intensity. We briefly discuss how to do adaptive regularization by controlling the amount of intensity change experienced by objects of any given scale due to TV regularization.

We define α_f, $0 \leq f \leq 1$, as the α which causes an image feature of constant intensity to be changed by the fractional amount f; that is, so that $\delta = (1-f)I$, where I is the original intensity. Then we can use (16) to find that

$$\alpha_f = \frac{(1-f)|\Omega|}{|\partial\Omega|}. \tag{22}$$



(a) True image (b) Regularized, $\alpha = \frac{3}{160}$



(c) Regularized, $\alpha = \frac{39}{1280}$ (d) Regularized, $\alpha = \frac{3}{160}, \frac{39}{1280}$

Figure 7: A noise-free image (a) is regularized, shown in (b) and (c), using two different constant values of α, and then in (d) it is regularized using a spatially varying value of $\alpha(\vec{x})$, where $\alpha(\vec{x}) = \frac{3}{160}$ in the region of the circle, and $\alpha(\vec{x}) = \frac{39}{1280}$ in the region of the L-shaped polygon. We are able to control the change in intensity level by spatially varying the value of $\alpha(\vec{x})$. Compare with Figure 8.

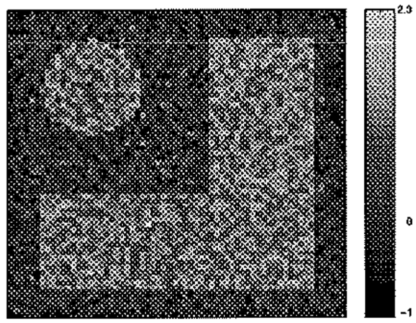
As an example, we consider an image which is comprised of a circle and an L-shaped polygon, as shown in Figure 7. We will first consider the noise-free image in order to show the effect the regularization has on the image itself, then we will add noise and see the similar effects of the regularization. We show the results of TV regularization when α is chosen according to a specific scale and controlled amount of change in intensity. Two values of α are used. We use definition (22) to find the appropriate choice for α_f, using $f = .75$. We first choose $\alpha_{.75}$ based on the scale of the circle, which has radius $r = \frac{3}{20}$, so that

$$\alpha = \frac{.25\,\pi\,(\frac{3}{20})^2}{2\pi(\frac{3}{20})} = \frac{3}{160}$$

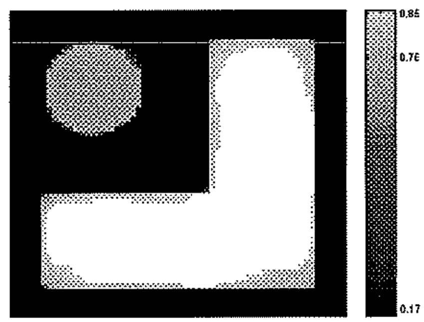
is chosen to preserve approximately 75% of the orginal level of intensity of the circle, as shown in Figure 7(b). In 7(c), $\alpha = \frac{39}{1280}$ is chosen to preserve approximately 75% of the original level of intensity of the L-shaped polygon, which has width $\frac{3}{10}$ and length $\frac{8}{10}$. In

7(d), α is chosen to be spatially varying in order to preserve 75% intensity level of each object, by choosing $\alpha = \frac{3}{160}$ in the region of the circle, and $\alpha = \frac{39}{1280}$ in the region of the L-shaped polygon. In 7(d) the reduction in intensity in the circle is slightly more than 0.25 (it should theoretically be exactly 0.25) due to the discrepency introduced by discretizing the continuous problem, as well due to the minmod approximation scheme, which was used to better preserve edges.

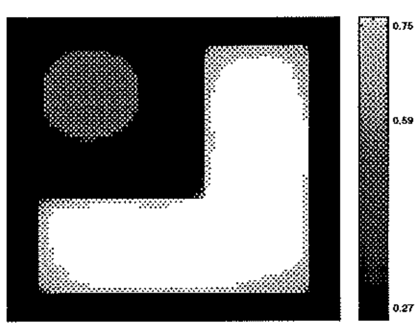
In Figure 8, we take the same image with the circle and the L-shaped polygon, add noise to it, and apply TV regularization to the noisy image, choosing $\alpha(\vec{x})$ in the same way as done for each of the images in Figure 7. The results for the noisy case are nearly identical to the results for the noise-free case given in Figure 7. The level of similarity between the noise-free and noisy regularization results will depend on the amount of noise present, the values of the regularization parameter α, and the true image itself.



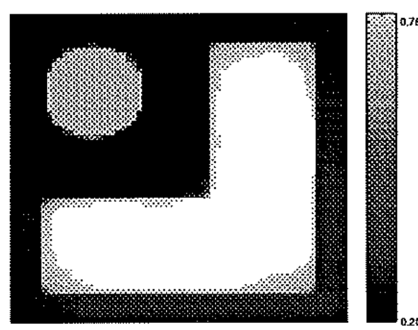
(a) Noisy image



(b) Regularized, $\alpha = \frac{3}{160}$



(c) Regularized, $\alpha = \frac{39}{1280}$



(d) Regularized, $\alpha = \frac{3}{160}, \frac{39}{1280}$

Figure 8: A noisy image (a) is regularized, shown in (b) and (c), using two different constant values of α, and then in (d) it is regularized using a spatially varying value of $\alpha(\vec{x})$, where $\alpha(\vec{x}) = \frac{3}{160}$ in the region of the circle, and $\alpha(\vec{x}) = \frac{39}{1280}$ in the region of the L-shaped polygon. We are able to control the change in intensity level by spatially varying the value of $\alpha(\vec{x})$. Compare with Figure 7.

4.2 Choosing the Regularization Parameter Based on Noise in the Image

An alternative way to choose α is based on the noise in the image. To some extent, this happens implicity and automatically when solving the constrained formulation of the problem, (6) or (12). In the unconstrained case, (5) or (11), we may try to choose α so that the solution to the unconstrained regularization problem (5) is identical to the solution of the constrained problem (6). One reason for solving the unconstrained problem rather than the constrained problem is that it is computationally less expensive. For general regularization schemes, there are other methods for choosing the regularization parameter if the noise level is not known or if it cannot be accurately estimated (see [8]). In this section we briefly discuss how one might choose α based on the noise in the image, using the results of Section 3. This approach requires no computational work to find α^{noise} if the noise level (and type of noise) is known, and it requires only the work to estimate the type and level of noise if it is unknown.

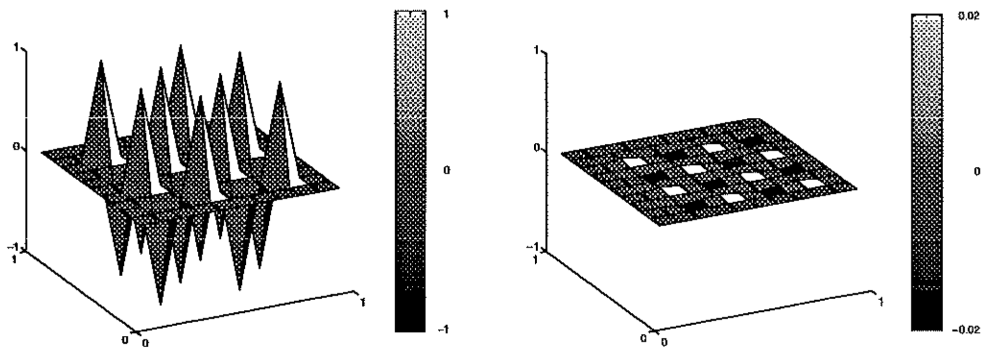
We examine the effects of TV regularization on the noise itself. The larger α is, the more noise will be removed, but with an increasing amount of detail being lost. So the task is to find the *smallest* value of α^{noise} necessary to remove essentially all of the noise. Take for example the image in Figure 9. We have constructed a small image of mutually disjoint individual noise pixels, each with magnitude one. It is constructed so that each noise element is not adjacent to any other noise element, and (less importantly) so that they alternate between positive and negative values, as shown. With this "checkerboard" assumption, the approach is to find the value of α that will completely smooth out each noise element. Since each noise element is a single $\frac{1}{n}$ x $\frac{1}{n}$ pixel of intensity level ± 1, we can use (22) to find that the theoretical value of α needed for TV regularization to annihilate a single noise element is

$$\alpha_0 = \frac{(1-f)I|\partial\Omega|}{|\Omega|} = \frac{(1-0)\cdot 1 \cdot \frac{1}{n^2}}{\frac{4}{n}} = \frac{1}{4n}.$$

We see in Figure 9 that this is essentially observed. The tiny amount of noise that still remains after regularization can be attributed to the discrepency introduced when discretizing the continuous problem. But for practical purposes, essentially all of the noise is removed.

The "checkerboard" noise in the previous example is generally not encountered in practice, but is a useful and demonstrative example. A more generic example is given in Figure 10. Here the noise is Gaussian (mean 0) with standard deviation of $\frac{1}{3}$, so that we can then assume that most of the noise has magnitude no greater than 1 (three standard deviations). If we choose $\alpha = \frac{1}{4n}$ as before, we see the results are again quite good. This time the noise was not quite completely smoothed out, primarily because there were adjacent pixels with noise values of the same sign, which in effect created local objects of scale or width greater than one pixel. Larger scaled objects require a larger value of α to be annihilated, as described by (18).

In choosing α based on the noise in the image, we could always choose α sufficiently large to completely smooth out the noise; in fact, for α sufficently large, the smoothed image



(a) Before regularization (b) After regularization

Figure 9: Isolated noise: essentially all of the noise, where each noise pixel has magnitude of 1, is annihilated due to TV regularization, using the theoretically predicted value of $\alpha = \frac{1}{4n}$.

would actually be a constant-valued image whose value is simply the mean of the original image z. Since larger values of α would also cause more detail in the image to be removed, we would consequently want to choose the smallest α that will still remove the desired amount (e.g. all) of the noise. As an example of TV regularization where α is chosen to be the (theoretically) smallest value necessary to remove all of the noise, we add Gaussian noise (mean 0, standard deviation $\frac{1}{3}$) to the same image from Figures 7 and 8, and apply TV regularization using $\alpha = \frac{1}{4n}$, as before. We threshhold the image after regularization to require that all image pixel values are non-negative. The results are seen in Figure 11.

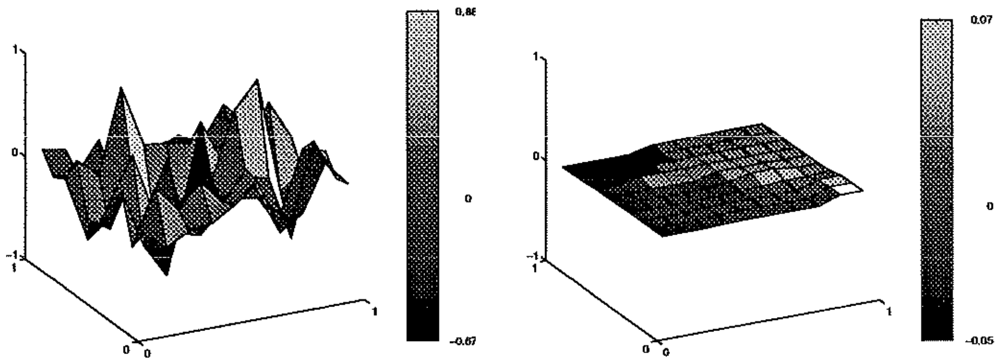
4.3 Choosing the Regularization Parameter Based on Both Noise and Scale

There is genenerally a balance between noise reduction (or other image enhancement) and detail preservation. There is no general rule to determine the optimal balance. It will depend on several factors, including the type of image being denoised and the purpose for regularization. This should vary locally for best results. Therefore, it becomes necessary to choose the amount of regularization to be done based both on the image features and on the noise level.

We outline one method in which we choose α based on both noise and local scale. We determine the locally varying α^{scale} which is based solely on the scale of the image, such as outlined in Section 4.1. We also find a value of α^{noise} which we would choose based solely on the noise level, such as done in Section 4.2. Then the spatially varying regularization parameter is a function of these two indepently found values of α,

$$\alpha(\vec{x}) = f(\alpha^{scale}(\vec{x}), \alpha^{noise}).$$

Of course α^{noise} could be spatially varying as well.



(a) Before regularization (b) After regularization

Figure 10: Gaussian noise: essentially all of the noise (mean 0, standard deviation $\frac{1}{3}$) is annihilated due to TV regularization, using the estimated value of $\alpha = \frac{1}{4n}$.

An example of defining an explicit relationship between $\alpha(\vec{x})$, α^{scale} and α^{noise} would be the convervative choice

$$\alpha(\vec{x}) = min(\alpha^{scale}(\vec{x}), \alpha^{noise}).$$

For this example, we could choose α^{scale} by using the relationship between α, scale and change in image intensity (18). We could similarly choose to set a bound on the change in intensity level, which we could accomplish by using (22) (This is a very nice feature of TV regularization, that we have this kind of local control over the smoothing.) Other ideas could include some sort of weighted average of $\alpha^{scale}(\vec{x})$ and α^{noise} in determining $\alpha(\vec{x})$.

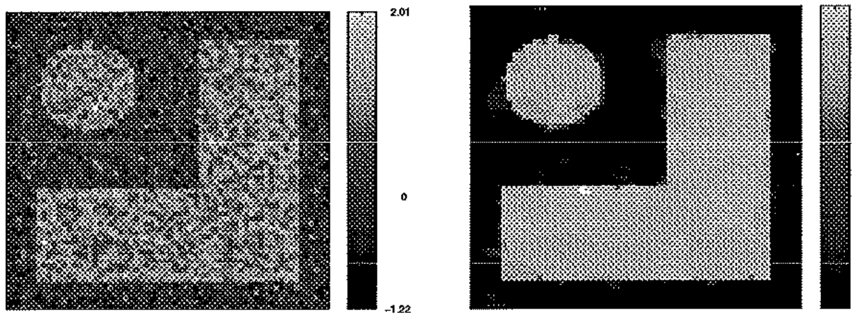
An implicit way of basing our choice of $\alpha(\vec{x})$ on α^{scale} and α^{noise} would be to solve the noise-constrained problem (12), while choosing $\alpha(\vec{x})$ based solely on the scale of the image features. We could do this by using the relationship between scale and change in image intensity (18). Choosing $\alpha(\vec{x})$ in this way allows us to locally control the amount of regularization done, based on both the noise level and the image features themselves.

5 Schemes for Adaptive Regularization and Automatic Scale Recognition

We now give two versions of a scheme for doing scale-based spatially adaptive regularization, and then a scheme for doing automatic scale recognition.

5.1 Adaptive Regularization

We give two essentially equivalent versions of an adaptive regularization scheme, and give examples of using it to denoise an image. Both versions are natural applications of the ideas



(a) Noisy image (b) After regularization

Figure 11: TV regularization, using $\alpha = \frac{1}{4n} = \frac{1}{256}$, was applied to (a) to obtain (b). Essentially all of the Gaussian noise (mean 0, standard deviation $\frac{1}{3}$) was removed, while not denoising excessively.

presented in Section 4, as they locally vary the amount of regularization or diffusion to be done, depending on the scale in the image, in a very controlled way.

The first version involves solving the adaptive noise-constrained problem (11). The idea is to locally vary the amount of smoothing to be done, while still satisfying the condition $\|u - z\|_{\Omega}^2 = \sigma^2$. We choose $\alpha(\vec{x})$ to be inversely proportional to the change in intensity level an object would experience as a result of smoothing. Since the change in intensity level is inversely proportional to the scale of the object (see (18)), we are actually choosing $\alpha(\vec{x})$ to be directly proportional to its scale. Results of choosing $\alpha(\vec{x})$ in this manner are given in Figure 12, along with the results of non-adaptive regularization for comparison.

The second version of this adaptive regularization scheme is done by applying TV based diffusion, using (9). We choose $\alpha(\vec{x})$ in exactly the same way as done for the first version of this scheme, as desribed in the previous paragraph. In order to demonstrate the similarity—indeed, the near equivalence—of TV regularization and TV diffusion, we choose the amount of time to march our diffusion equation in the following way. In computing the image in Figure 12(c) we find the Lagrange multiplier λ, and we choose the number of time steps n to time-march such that $n\,\Delta t = \frac{1}{\lambda}$. We do the same thing in the adaptive case, where we again find the Lagrange multiplier in computing the image (d). The resulting images are given in (e) and (f), which demonstrate the approximate equivalence of the two ways of viewing TV. In (d) and (f), we see the effects of adaptive regularization and diffusion particularly in the regions around the smallest rectangle and the smallest square. Here a higher level of noise is allowed to remain in order to better preserve the true image feature. There is always, of course, a balance between noise removal and detail preservation. In considering the basic unconstrained formulation of TV regularizaton, we remind the reader that solving (5) is equivalent to doing a single, *implicit* step of time-marching with (9), using the original image z as the initial condition, and using α as the step size. This fact helps explain the close correspondence of images (c) and (d) with (e) and (f), respectively. Refer to equation (10).

5.2 Automatic Scale Recognition

We next give a simple scheme for automatic scale recognition, and briefly discuss heuristics and subsequent uses of it. This scheme is another natural application of the ideas presented in Section 4, as we will determine the scale in the image by examining the change in the image due to TV regularization.

The scheme is actually quite simple. We first give a version of the scheme that can be applied to a noise-free image, and then discuss how to extend it to noisy images. We use the non-adaptive, unconstrained formulation of TV regularization (5). We apply TV regularization to an image using a sufficiently small value of α, so as not to lose any image features. We then examine on a pixel-by-pixel basis the change in intensity level between the image before regularization u, and the image after regularization z. Using the formula (19) and knowing the value of α that was used in doing the regularization, we can easily determine the scale $s = s(i,j)$ at each position in the image. The equation that describes the scale at each position i,j then is

$$s(i,j) = \frac{\alpha}{\delta(i,j)}.$$

We give an example of applying this scheme to the image shown in Figure 13.[2] For this example we suppose that we are interested in isolating a certain group of image features, say in this case the bacteria, which are on average up to 10 pixels long and 5 pixels wide. The scale of a bacteria of these dimensions would be

$$s = \frac{|\Omega|}{|\partial\Omega|} = \frac{(\frac{10}{n})(\frac{5}{n})}{2(\frac{10}{n}+\frac{5}{n})} = \frac{5}{3n} \tag{23}$$

After computing $s(i,j)$ throughout the image, we determine where the bateria is by comparing the scale values $s(i,j)$ just computed to the estimated scale value (23). Of course since there are varying sizes of bacteria, we must define an interval for scale values that we will accept as being those of bacteria. We arbitrarily define this interval to be

$$\frac{1}{2n} \leq s \leq \frac{5}{3n}. \tag{24}$$

In Figure 13(b), we see the image which shows the locations of only those image features whose scale, as computed by our algorithm, is within the interval of acceptance (24). Of course, if the interval of acceptance (24) is wider, there is more chance that all of the bacteria will be identified, but there is also more chance that features other than bacteria will be included. Similarly, smaller intervals of acceptance screen out the unwanted features, but with the added risk of losing some of the wanted features, in this case the bacteria. There is a (generally arbitrary) balance between the two extremes.

For noisy images this scheme would need to be modified to be effective, as much of the change from the original image to the image after regularization is due to the denoising

[2]The image is of a polished stainless steel coupon, immersed in water, with prominent features being the relatively large registration mark, and relatively small bacteria attached to the coupon. The image is taken from an image provided by the Center for Biofilm Engineering, Montana State University-Bozeman.

that takes place. This more complex problem is still under investigation, and we give no numerical results at this time. Still, the basic idea described in the scheme for the noise-free case still applies.

In the example we just considered, our purpose was to isolate features of a certain scale. Other uses for the information gathered about the image scale $s(i, j)$ would include using the information to automatically determine how to define $\alpha(\vec{x})$ for doing adaptive regularization or diffusion, particularly in the case of denoising.

6 Summary of Results

We have given some exact solutions to simple cases of the TV regularization problem to describe how TV regularization will affect images. These results are used to develop ideas for choosing a spatially varying regularization parameter based on scale, noise and/or both scale and noise. These ideas are then used to develop two schemes for doing adaptive TV regularization, as well as a scheme for automatic scale recognition. We develop our results by considering the unconstrained formulation of the minimization problem. Our results are also demonstrated in solving the constrained formulation of the problem. Moreover our results can be extended to any geometry-driven diffusion scheme, as other schemes can be thought of as variations or modifications of the model TV regularization scheme.

Future work that is motivated by our work includes: (1) developing additional adaptive (or more intelligent non-adaptive) TV regularization schemes based on our results, and (2) further developing the connection between TV diffusion and TV regularization, as well as TV diffusion and general geometry-driven diffusion schemes, in order to more fully extend our understanding of TV regularization to other geometry-driven diffusion schemes.

References

[1] M. R. Banham and A. K. Katsaggelos. "Spatially adaptive wavelet-based multiscale image restoration." Technical report, Dept. of E.E. and C.S., Northwestern University, 1995. IEEE Transactions on Image Processing, April 1996.

[2] J. Biedmond and R. Lagendijk. "Iterative Identification and Restoration of Images." Kluwer Academic Publishers, Dordrecht, 1991.

[3] Tony F. Chan, Gene H. Golub, and Pep Mulet. "A Nonlinear Primal-Dual Method for Total Variation-Based Image Restoration." UCLA Math Department CAM Report 95-43, September 1995

[4] Tony F. Chan, Pep Mulet. Tony F. Chan and Pep Mulet. "Iterative Methods for Total Variation Image Restoration." UCLA Math Department CAM Report 96-38, October 1996. To appear in "Iterative Methods in Scientific Computing", Proceedings of Winter School on Iterative Methods, the Chinese Universtiy of Hong Kong, December 1995, Spring Verlag, Singapore.

[5] D. Dobson and F. Santosa. "Recovery of Blocky Images from Noisy and Blurred Data." SIAM Journal of Applied Mathematics, to appear.

[6] D. Dobson and C.R. Vogel. "Convergence of An Iterative Method for Total Variation Denoising." SIAM Journal of Numerical Analysis, to appear.

[7] R.C. Gonzalez and R.E. Woods. *Digital Image Processing.* Addison Wesley, 1993.

[8] M. Hanke and P.C. Hansen. "Regularization Methods for Large Scale Problems." Survey of Mathematics in Industry, Volume 3, pp. 253-315, 1993.

[9] P.L. Lions, S. Osher, and L. Rudin. "Denoising and Deblurring Algorithms with Constrained Nonlinear PDE's." SIAM Journal of Numerical Analysis, submitted.

[10] J.M. Morel and S. Solimini. *Variational Methods for Image Segmentation.* Birkhauser, Boston, 1995.

[11] P. Perona and J. Malik. "Scale Space and Edge Detection using Anisotropic Diffusion." IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 12, pp. 629-639, 1990.

[12] L. Rudin, "MTV-multiscale total variation principle for a PDE-based solution to non-smooth ill-posed problem." Technical Report, Cognitech, Inc. Talk presented at the Workshop on Mathematical Methods in Computer Vision, University of Minnesota, Sept. 11-15, 1995.

[13] L. Rudin, S. Osher, and E. Fatemi. "Nonlinear Total Variation Based Noise Removal Algorithms." Physica D, Vol. 60, pp. 259-268, 1992.

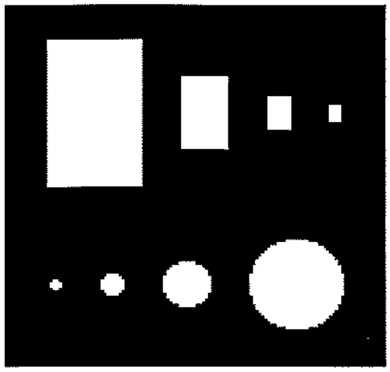
[14] D. Strong, T. Chan. "Relation of Regularization Parameter and Scale in Total Variation Based Image Denoising." UCLA Math Department CAM Report 96-7, February 1996.

[15] D. Strong, T. Chan. "Exact Solutions to Total Variation Regularization Problems." UCLA Math Department CAM Report 96-41, October 1996.

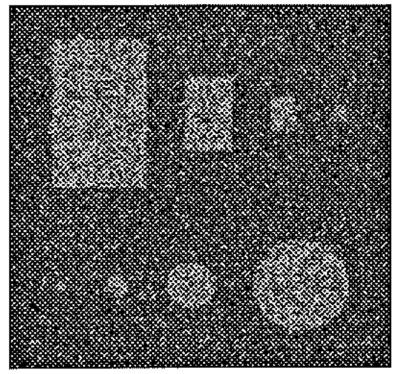
[16] B.M. ter Haar Romeny. *Geometry-driven Diffusion in Computer Vision.* Kluwer Academic Publishers, Dordrecht, 1994.

[17] C.R. Vogel and M.E. Oman. "Iterative Methods for Total Variation Denoising." SIAM Journal of Scientific Computing, Vol. 17, pp. 227-238, 1996.

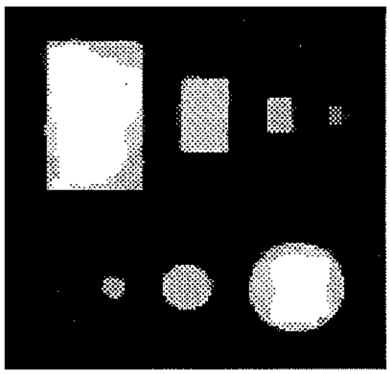
[18] Yu-Li You, M. Kaveh, Wen-Yuan Xu, A. Tannenbaum. "Analysis and Design of Anisotropic Diffusion." IEEE Transactions on Image Processing. To appear.



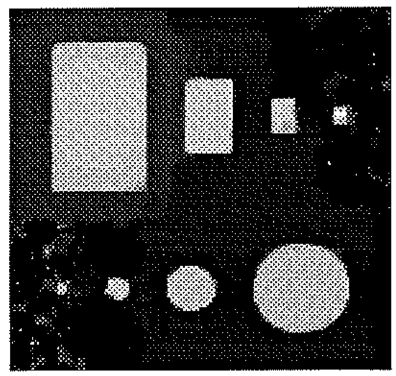
(a) True image



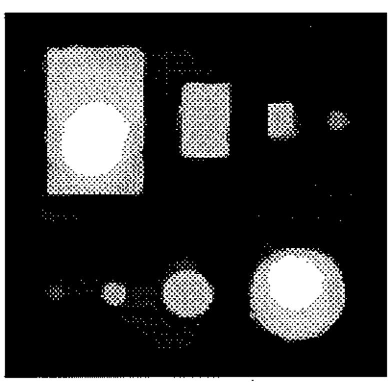
(b) Noisy image, SNR = 1



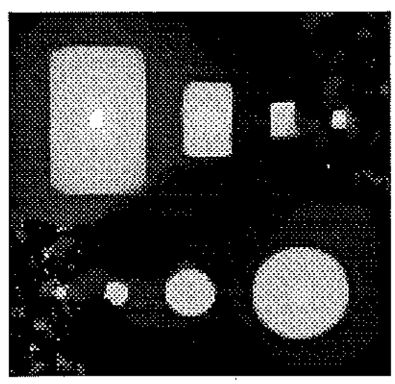
(c) *Non-adaptive* TV regularization, with noise constraint



(d) *Adaptive* TV regularization, with noise constraint

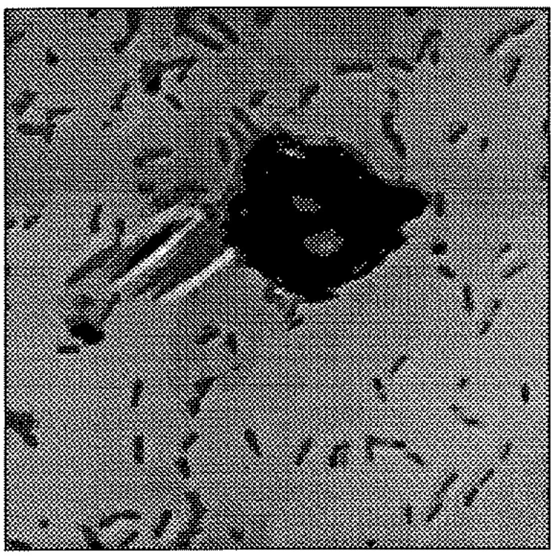


(e) *Non-adaptive* TV diffusion

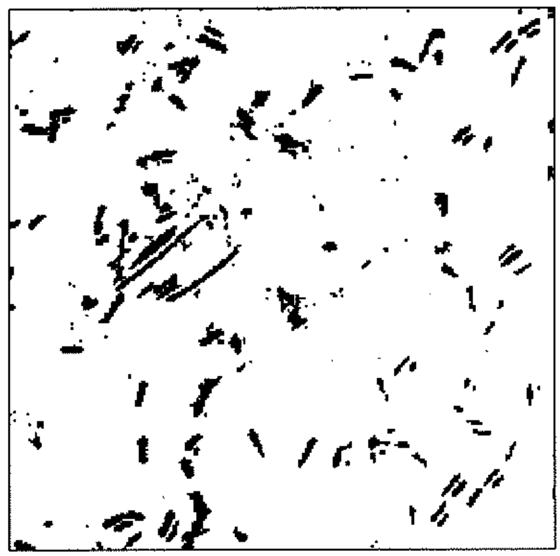


(f) *Adaptive* TV diffusion

Figure 12: TV regularization and TV diffusion, both uniform and spatially adaptive, are done to denoise (b). In (e) and (f), the amount of time-marching done was chosen such that $n\,\Delta t = \frac{1}{\lambda}$, where λ was found in computing each of the images (c) and (d). Note the approximate equivalence of (c) and (e), as well as of (d) and (f).



(a) Image in which to determine scales



(b) Features from image (a) with specified scale are shown

Figure 13: The image in (a) is regularized, and the scale at each pixel location is determined. In (b) we highlight only the image features which are the scale of interest, as precribed by the user.