Chapter 6

# Conclusions and Future Research

## 6.1 Contributions of the Thesis.

IN this thesis we have presented a new approach to edge enhancement and segmentation which is robust to very high levels of noise and blurring. Our approach is based on a new class of evolution equations for the processing of imagery and signals which we have termed stabilized inverse diffusion equations or SIDEs. These evolutions, which have discontinuous right-hand sides, have conceptual and mathematical links to other evolution-based and variational methods in signal and image processing, such as the Perona-Malik equation, shock filters, total variation minimization, region merging methods, the Geman-Reynolds functional, and the Mumford-Shah functional. SIDEs, however, have their own unique qualitative characteristics and properties which were described in Chapter 3 and led to improved performance. Specifically, a number of stability properties were derived, such as well-posedness of the 1-D equation, the maximum principle, and—for almost all input images—the invariance of the resulting segmentation with respect to small changes in the input image. SIDEs were further characterized as gradient descent equations for certain energy functionals. Several classes of Lyapunov functionals were found for these equations.

SIDEs' superior performance was demonstrated in Chapter 3 through a multitude of examples, both in 1-D and in 2-D. Their robustness—in contrast with other methods—was shown in segmenting signals with heavy-tailed noise and images with high levels of blurring and speckle noise, such as ultrasound and SAR images.

Chapter 4 provided an extensive performance analysis and a probabilistic interpretation of a variant of SIDEs. We showed that it produced maximum likelihood solutions to certain binary classification/edge detection problems in 1-D. We conjectured that similar results hold in 2-D. Extensive performance analysis for the 1-D problem of finding one change in mean was done, producing bounds on the probabilities of errors. For this problem, it was also proved that the algorithm is optimal with respect to a certain $H_\infty$-like criterion. The experiments in both 1-D and 2-D showed that the algorithm performs extremely well, and is robust to large-amplitude noise and blurring.

We also presented a fast implementation and compared it with the dynamic programming solution of the 1-D problem. Dynamic programming was faster in certain instances, but took up more memory and did not generalize to 2-D. We explored the links between our non-linear diffusion equation and the Mumford-Shah variational method of image segmentation, and showed that a certain particular case of these is a linear programming problem.

Thus, the contribution of Chapter 4 was two-fold. First, we presented a fast and robust 1-D edge detection and 2-D image segmentation method. Second, we established a link between deterministic methods for image restoration and segmentation (based on non-linear diffusions and variational formulations) and a probabilistic framework. This leads to a deeper understanding of these methods: both of their performance, and of how to use them in a variety of situations (e.g., in Section 4.4 this meant pre-processing the data by forming the log-likelihood ratios). As we will argue in the next section, we have no doubt that these lines of investigation can and should be pursued further.

Finally, in Section 5 we demonstrated that our framework can be easily adapted to non-scalar-valued images. Specifically, we used the result from Chapter 3 which showed that a scalar-valued SIDE is the steepest descent equation for a certain energy functional. We then generalized SIDEs by deriving the steepest descent equations for similar energy functionals in vector-valued and circle-valued cases. We showed that many properties of the scalar-valued SIDEs applied, and pointed out several important differences. We successfully applied the resulting evolution to the segmentation of color, texture, and orientation images.

## ■ 6.2 Future Research.

In this section, we list several interesting problems motivated by this thesis.

### ■ 6.2.1 Feature Extraction.

Once one admits the concept of sliding surfaces for signal or image evolution, the question immediately arises as to whether one can design sliding surfaces other than those used in this work. In particular, the sliding surfaces reflecting the enforced equality of neighboring points or pixels, directly correspond to piecewise-constant approximations of signals, and the resulting SIDE evolution in essence produces an adapted sequence of staircase approximations to a signal or an image. An open question is whether it is also possible to produce a sequence of other approximations (e.g., linear splines) to signals and images by appropriately defining the sliding surfaces. The challenge is to design these surfaces in such a way that the resulting SIDE evolution is both fast and robust.

### ■ 6.2.2 Choosing the Force Function.

We mentioned in Chapter 3 that choosing a SIDE force function best suited for a particular application is an open research question. We also partially addressed this issue in Chapter 4, by describing and analyzing the problems for which $F(v) = \text{sgn}(v)$ was the optimal choice. This choice was optimal in the sense that it resulted in the maximum likelihood solutions to change detection problems for certain probabilistic

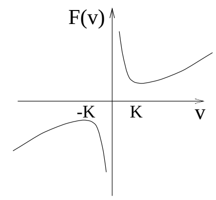


**Figure 6.1.** A force function which results in faster removal of outliers.

models of the input signal. It is unclear how to choose force functions for other signal models; an interesting question is for what models this can be done so as to guarantee that the solution produced by the SIDE is the maximum likelihood solution.

Another question is that of robustness properties of the SIDEs corresponding to different shapes of the force function. Intuitively, if the goal is segmentation in the presence of outliers, then it is appropriate to diffuse quickly both in the areas of very large gradient (corresponding to the outliers), and in the areas of very small gradient (corresponding to small-amplitude noise). Ideally, the minimum diffusion speed would be at the locations with intermediate values of the gradient, corresponding to edges. We are then led to the form of a force function depicted in Figure 6.1, which is the inverse of a Perona-Malik force function, and has a unique minimum at some location $K$. If the parameters of the probabilistic model of the input signal (such as the standard deviation of the noise, the magnitude and frequency of the outliers) are fixed, then it is natural to expect that some value of $K$ is, in some sense, optimal. Uncovering this relationship between the model and the corresponding value of the parameter $K$ is an interesting research topic.

## ■ 6.2.3 Choosing the Stopping Rule.

Another interesting question is how to stop a SIDE so as to obtain the best segmentation. When the force function is $F(v) = \text{sgn}(v)$ and the number of regions is an exponential random variable, the answer to this question is given by Proposition 4.6 of Chapter 4. We have been unable, however, to answer this question for the general form of the SIDE force function $F$. In our examples, we assumed that the desired number of regions was known, and stopped the evolution when that number was attained. In some applications where the number of objects of interest is not known a priori, this may not be a realistic assumption.

As we remarked in Chapter 4, one way to address the question of when to stop is by looking at the size of the energy of the region being removed, or at the time required to remove it. If a region has large energy, or if a long time is required to remove it, it is more likely to be significant, and less likely to be due to noise.

## 6.2.4 Image Restoration.

As mentioned in Chapter 3, while SIDEs may do an excellent job of segmentation and location of edges, they do not directly provide good estimates of the values within regions or between edges. This limitation can be addressed, for example, by using a SIDE for segmentation and then using optimal linear estimation or filtering within each segmented region in order to get both accurate edge estimates and denoising within each so-identified region.

## 6.2.5 PDE Formulation.

The problem of re-casting SIDEs as PDEs similar to (3.27) of Chapter 3 is of great theoretical interest, because PDEs have been the unifying language of most of the researchers in the field of non-linear scale spaces and geometry-driven diffusion. A PDE formulation is also important because it will allow one to obtain certain results concerning the evolution. For example, the analysis of the invariance of the equation to various coordinate transformations of the image domain is clearly more convenient to carry out in the continuous-space setting.

## 6.2.6 Further Probabilistic Analysis.

An important topic for future investigations is whether results similar to those of Chapter 4 can be obtained for signals whose samples come from more than two probability distributions. In addition, we would like to generalize these results to the situation when the samples of the observed process are not conditionally independent. This would make possible the design of probabilistic texture segmentation algorithms. We also would like to explore the case of unknown probability densities for the binary problem, which means devising ways of learning the log-likelihood ratios from the data. Proving the conjectures of Section 4.7 also has high priority.

The most ambitious question is perhaps analyzing the performance in 2-D, which amounts to analyzing random fields with abrupt changes. In order to understand the sensitivity of a segmentation algorithm to noise, we need to understand how the edges produced by the algorithm will change if the input image is perturbed by a spatial random field.

## 6.2.7 Prior Knowledge.

Another avenue of future research is understanding how to incorporate prior knowledge, which is a very important question not only for segmentation, but for any other image processing task. For example, we saw in Chapter 4 that SIDEs can be modified to make use of a prior distribution on the number of desired regions in the segmentation. It remains to be seen, however, how to incorporate other knowledge, for instance, information about sizes and shapes of the regions.

# Appendix A

# Proof of Lemma on Sliding (Chapter 3)

To simplify notation, we replace $n_i$ with $i$ in (3.11) and re-write the system in terms of $v_i = u_{i+1} - u_i$:

$$\begin{aligned} \dot{v}_i &= \frac{1}{m_{i+1}}(F(v_{i+1}) - F(v_i)) - \frac{1}{m_i}(F(v_i) - F(v_{i-1})) \\ i &= 1, \ldots, p-1. \end{aligned} \tag{A.1}$$

We need to prove that if $(i_1, \ldots, i_{p-1})$ is any permutation of $(1, \ldots, p-1)$, then, as $\mathbf{v}$ approaches $S = \cap_{k=1}^{m} S_{i_k} \backslash (\cup_{k=m+1}^{p-1} S_{i_k})$, $\lim(\dot{v}_{i_q} \text{sign}(v_{i_q})) \leq 0$ for all integers $q$ between 1 and $m$, and for at least one such $q$ the inequality is strict (i.e., the trajectories enter $S$ transversally). Note that for every point $s \in S$ and every quadrant $Q$, we only need to find one sequence of $\mathbf{v}$'s approaching $s$ from $Q$ and satisfying these inequalities. This is because the solutions vary continuously inside each quadrant.

Fix $v_{i_{m+1}}, \ldots, v_{i_{p-1}}$ at non-zero values, let

$$\varepsilon = \frac{1}{2} \min_{m+1 \leq j \leq p-1} |v_{i_j}|,$$

let initially $\delta = \varepsilon$, set $|v_{i_1}| = \ldots = |v_{i_m}| = \delta$, and drive $\mathbf{v}$ towards $S$ by letting $\delta$ go to zero. Take an arbitrary index $q$ between 1 and $m$. By our construction, $v_{i_q}$ is approaching zero, and either $v_{i_q} = \delta > 0$ or $v_{i_q} = -\delta < 0$. If $v_{i_q} = \delta$, then, by construction, $v_{i_q} \leq |v_{i_{q\pm1}}|$, implying $F(v_{i_q}) \geq F(v_{i_{q\pm1}})$, which makes the RHS of (A.1) for $i = i_q$ non-positive: $\lim_{\delta \to 0} \dot{v}_{i_q} \leq 0$. If $m < p-1$, then there is a $j$ between 1 and $m$ such that at least one of the two neighbors of $v_{i_j}$ is in the set $\{v_{i_{m+1}}, \ldots, v_{i_{p-1}}\}$, and whose absolute value is therefore staying above $\varepsilon$: $|v_{i_{j+1}}| > \varepsilon$ or $|v_{i_{j-1}}| > \varepsilon$. Without loss of generality, suppose it is the left neighbor: $|v_{i_{j-1}}| > \varepsilon$. If $m = p-1$, define $j = 1$. If our arbitrary $q$ happens to be equal to this $j$, then

$$F(v_{i_q}) - F(v_{i_{q-1}}) = F(v_{i_j}) - F(v_{i_{j-1}}) > F(v_{i_j}) - F(\varepsilon),$$

and hence (A.1) for $i = i_j$ has a strictly negative limit: $\lim_{\delta \to 0} \dot{v}_{i_j} < 0$. Similar reasoning for the case $v_{i_q} = -\delta$ leads to $\lim_{\delta \to 0} \dot{v}_{i_q} \geq 0$, and, if it happens that $q = j$, then $\lim_{\delta \to 0} \dot{v}_{i_q} > 0$. ■

Appendix B

# Proofs for Chapter 4.

## B.1 Proof of Proposition 4.1.

Let $\mathbf{h} = \mathbf{h}^*_{\leq\nu}(\mathbf{u}^0)$, and suppose that $\mathbf{h}$ has an upward edge at location $i$: $h_i = 0$, and $h_{i+1} = 1$. Suppose further that there is no upward $\alpha$-crossing at $i$. We only treat the case when the signal values are not equal to the threshold: $u^0_i \neq \alpha$, $u^0_{i+1} \neq \alpha$. The degenerate cases when one or more signal values are equal to the threshold are handled similarly.

Consider two possibilities.

**Case 1.** $u^0_i > \alpha$ (Figure B.1, (b)).

Changing $h_i$ from 0 to 1 will either maintain the same number of edges (if $i > 1$ and $h_{i-1} = 0$), or reduce the number of edges by one (if $i = 1$), or reduce the number of edges by two (if $i > 1$ and $h_{i-1} = 1$). It will also increase $\phi(\mathbf{u}^0, \mathbf{h})$ (by the amount $u^0_i - \alpha$), which contradicts the assumption that $\mathbf{h}$ is the best hypothesis with $\nu$ or fewer edges. Thus, $u^0_i$ has to be

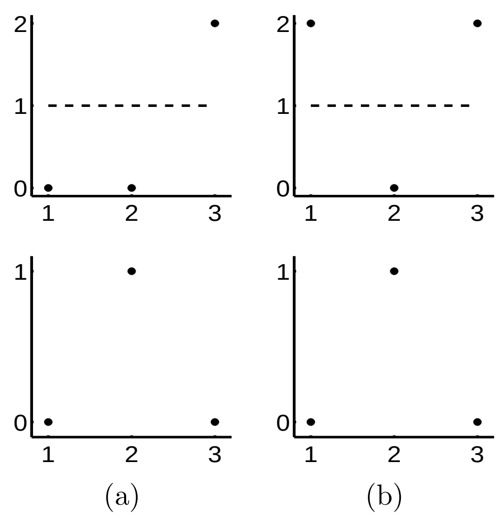


**Figure B.1.** Samples of a signal (top plots) and impossible edge configurations of optimal hypotheses (bottom plots).

smaller than $\alpha$.

**Case 2.** $u_i^0 < \alpha$ and $u_{i+1}^0 < \alpha$ (Figure B.1, (a)).
Changing $h_{i+1}$ from 1 to 0 will either maintain the same number of edges (if $i < N-1$ and $h_{i+2} = 1$), or reduce the number of edges by one (if $i = N-1$), or reduce the number of edges by two (if $i < N-1$ and $h_{i+2} = 0$). It will also increase $\phi(\mathbf{u}^0, \mathbf{h})$ (by the amount $\alpha - u_{i+1}^0$), again violating the assumption that $\mathbf{h}$ is the best hypothesis with $\nu$ or fewer edges. Combining this with Case 1, we see that $\mathbf{u}$ must have an upward $\alpha$-crossing at $i$.

It is similarly shown that every downward edge of $\mathbf{h}$ occurs at a downward $\alpha$-crossing of $\mathbf{u}^0$. ■

## ■ B.2 Proof of Proposition 4.2: SIDE as a Maximizer of a Statistic.

We note that this proof relies on Lemma 4.1, proved in Subsection 4.3.1.

We first show that at any time $t$ between 0 and $t_f$, the rate of reduction of $\phi(\mathbf{u}(t), \mathbf{h})$ is the largest for $\mathbf{h} = \mathbf{h}^*(\mathbf{u}(t_f))$, among all hypotheses $\mathbf{h}$ with $\nu_\alpha(t_f)$ or fewer edges. To simplify notation, define $\nu = \nu_\alpha(t_f)$, $g_{j+1} = N$, and let $g_1, \dots, g_j$ be the edge locations (from left to right) of an arbitrary hypothesis $\mathbf{h}$ with $\nu$ or fewer edges (i.e., $j \le \nu$). Without loss of generality, assume that the leftmost edge is upward, so that $h_{g_1} = 0$. Then

$$\begin{aligned} \dot{\phi}(\mathbf{u}(t), \mathbf{h}) &= \mathbf{h}^T \dot{\mathbf{u}}(t) && \text{(B.1)} \\ &= \sum_{i \text{ is odd}} \sum_{n=g_i+1}^{g_{i+1}} \dot{u}_n(t) && \text{(B.2)} \\ &= \sum_{i \text{ is odd}} \left\{ \sum_{n=g_i+1}^{N} \dot{u}_n(t) - \sum_{n=g_{i+1}+1}^{N} \dot{u}_n(t) \right\}, && \text{(B.3)} \end{aligned}$$

Note that, for any integer $s$ such that $1 \le s \le N-1$ and $u_{s+1}(t) - u_s(t) \ne 0$, we have, by summing up Equations (4.1) from $n = s+1$ to $n = N$,

$$\sum_{s+1}^{N} \dot{u}_n(t) = -\mathrm{sgn}(u_{s+1}(t) - u_s(t)) = \pm 1. \tag{B.4}$$

If $u_{s+1}(t) = u_s(t)$, let $p$ be the smallest index to the left of $s$ such that $u_{p+1}(t) = u_s(t)$, and let $q$ be the largest index to the right of $s$ such that $u_q(t) = u_s(t)$. If $q \le N-1$

and $p \geq 1$, then, according to Equations (4.1) and (4.3), we have:

$$\begin{aligned}\sum_{n=s+1}^{N} \dot{u}_n(t) &= \sum_{n=s+1}^{q} \dot{u}_n(t) + \sum_{n=q+1}^{N} \dot{u}_n(t) \\ &= (q-s)\dot{u}_s - \operatorname{sgn}(u_{q+1}(t) - u_q(t)) \\ &= (q-s)\frac{1}{q-p}(\operatorname{sgn}(u_{q+1}(t) - u_q(t)) - \operatorname{sgn}(u_{p+1}(t) - u_p(t))) \\ &\quad -\operatorname{sgn}(u_{q+1}(t) - u_q(t)) \\ &= -\frac{1}{q-p}\operatorname{sgn}(u_{q+1}(t) - u_q(t)) \\ &\quad -\frac{1}{q-p}\operatorname{sgn}(u_{p+1}(t) - u_p(t)) \geq -1, \end{aligned} \tag{B.5}$$

since $q - p \geq 2$. The same inequality is obtained if $q = N$ or $p = 0$. Combining (B.5) and (B.4), we see that $\sum_{n=s+1}^{N} \dot{u}_n(t) \geq -1$ for any $s$. Therefore, the minimal possible value for (B.3) is $(-1)$ times the number of sums of the form $\sum_{n=s+1}^{N} \dot{u}_n(t)$ in that expression. This is $-j$, which is, by assumption, not smaller than $-\nu$:

$$\dot{\phi}(\mathbf{u}(t), \mathbf{h}) \geq -j \geq -\nu.$$

Now note that in this double inequality, both equalities are achieved for all time $t$, $0 \leq t \leq t_f$, when $\mathbf{h} = \mathbf{h}^*(\mathbf{u}(t_f))$. Indeed, since in this case $j = \nu$, and—as easily seen from the definition of $\phi$—$g_1, \ldots, g_\nu$ are $\alpha$-crossings of $\mathbf{u}(t_f)$ (and, therefore—by Lemma 4.1—also of $\mathbf{u}(t)$ for $0 \leq t \leq t_f$), we have that for $t \in [0, t_f]$,

$$-\operatorname{sgn}(u_{g_i+1}(t) - u_{g_i}(t)) = \begin{cases} -1 & \text{if } i \text{ is odd} \\ 1 & \text{if } i \text{ is even.} \end{cases}$$

Inserting this into (B.4), and then back into (B.3), we see that in this case, (B.3) is equal to $-\nu$. By definition of $\mathbf{h}^*(\mathbf{u}(t_f))$, $\phi(\mathbf{u}(t_f), \mathbf{h})$ is the largest for $\mathbf{h} = \mathbf{h}^*(\mathbf{u}(t_f))$. On the other hand, we just showed that the amount of the reduction of $\phi(\mathbf{u}(t), \mathbf{h})$ during the evolution was the greatest for $\mathbf{h} = \mathbf{h}^*(\mathbf{u}(t_f))$, over all possible hypotheses with $\nu$ or fewer edges. Therefore, $\phi(\mathbf{u}(0), \mathbf{h})$ must also have been the largest for $\mathbf{h} = \mathbf{h}^*(\mathbf{u}(t_f))$, over the same set of hypotheses, which is the statement of the Proposition. ■

The statement of the same proposition in [51] is as follows.

**Proposition B.1.** *Fix the initial condition* $\mathbf{u}^0$ *of the SIDE (4.1), and let* $\mathbf{u}(t)$ *be the corresponding solution. Suppose that a statistic* $\phi$ *satisfies two conditions:*

**1)** $\frac{d}{dt}\left\{\phi(\mathbf{u}(t), \mathbf{h}) - \mathbf{h}^T\mathbf{u}(t))\right\} = 0$;

**2)** *there exists* $\alpha \in \mathbb{R}$ *such that,* $\forall t \geq 0$*, the optimal hypothesis* $\mathbf{h}^*(\mathbf{u}(t))$ *is generated by the set of all* $\alpha$*-crossings of* $\mathbf{u}(t)$*.*

*Let $\nu_\alpha(t)$ be the number of $\alpha$-crossings of $\mathbf{u}(t)$. Then*

$$\mathbf{h}^*_{\le \nu_\alpha(t)}(\mathbf{u}^0) = \mathbf{h}^*(\mathbf{u}(t)). \qquad \blacksquare$$

We now show that the two statements are equivalent.

**Proposition B.2.** *Propositions 4.2 and B.1 are equivalent, in the following sense.*
*i) If $\phi(\mathbf{u}, \mathbf{h})$ is as in Proposition 4.2, it satisfies the two conditions of Proposition B.1.*
*ii) Suppose that $\phi'(\mathbf{u}, \mathbf{h})$ satisfies the two conditions of Proposition B.1 for any initial data $\mathbf{u}^0 \in I\!R^N$ (where the constant $\alpha$ may depend on $\mathbf{u}^0$), and suppose that $\phi(\mathbf{u}, \mathbf{h})$ is as in Proposition 4.2. Then, for all $\mathbf{u} \in I\!R^N$ and for all $\mathbf{h} \in \{0,1\}^N \backslash \{\mathbf{0}, \mathbf{1}\}$ (where $\mathbf{1} = (1, \ldots, 1)^T \in I\!R^N$), $\phi(\mathbf{u}, \mathbf{h})$ and $\phi'(\mathbf{u}, \mathbf{h})$ can only differ by a function of $\sum_{i=1}^N u_i$:*

$$\phi'(\mathbf{u}, \mathbf{h}) - \phi(\mathbf{u}, \mathbf{h}) = f\left(\sum_{i=1}^N u_i\right),$$

*for some function $f : I\!R \to I\!R$, and thus the optimal hypotheses with respect to $\phi$ and $\phi'$ are the same.*

**Lemma B.1.** *Let $\mathbf{u}(t)$ be the solution of the SIDE (4.1). Suppose that a function $\psi : I\!R^N \to I\!R$ satisfies*

$$\frac{d}{dt}(\psi(\mathbf{u}(t))) = 0, \tag{B.6}$$

*for any initial data $\mathbf{u}^0$. Then $\psi$ only depends on the sum of the entries of its argument—i.e., there is a function $f : I\!R \to I\!R$ such that*

$$\psi(\mathbf{u}) = f\left(\sum_{i=1}^N u_i\right).$$

**Proof.** We first show that the partial derivatives of $\psi$ with respect to all the variables are equal to each other, using the identity

$$\frac{d}{dt}(\psi(\mathbf{u}(t))) = \sum_{i=1}^N \frac{\partial \psi}{\partial u_i} \dot{u}_i. \tag{B.7}$$

Take an initial condition for which $u_1^0 < u_2^0 < \ldots < u_{N-1}^0 < u_N^0$. It then follows from (4.1) that $\dot{u}_1 = 1$, $\dot{u}_N = -1$, and $\dot{u}_i = 0$ for $2 \le i \le N-1$. Substituting these into (B.7) and using (B.6), we get:

$$\frac{\partial \psi}{\partial u_1} = \frac{\partial \psi}{\partial u_N}. \tag{B.8}$$

Now take an initial condition for which $u_1^0 < u_2^0 < \ldots < u_{N-1}^0$ and $u_N^0 < u_{N-1}^0$. Then $\dot{u}_1 = \dot{u}_N = 1$, $\dot{u}_{N-1} = -2$, and $\dot{u}_i = 0$ for $2 \le i \le N-2$, and therefore

$$\frac{\partial \psi}{\partial u_1} - 2\frac{\partial \psi}{\partial u_{N-1}} + \frac{\partial \psi}{\partial u_N} = 0,$$

which, combined with (B.8), gives:

$$\frac{\partial \psi}{\partial u_1} = \frac{\partial \psi}{\partial u_{N-1}}.$$

Proceeding inductively in a similar fashion, we obtain:

$$\frac{\partial \psi}{\partial u_1} = \frac{\partial \psi}{\partial u_2} = \ldots = \frac{\partial \psi}{\partial u_N}. \tag{B.9}$$

Let

$$\begin{aligned} w_i &= u_{i+1} - u_i \text{ for } i = 1, \ldots, N-1, \\ w_N &= \sum_{i=1}^{N} u_i. \end{aligned}$$

Then it is easily verified by direct substitution that

$$\begin{aligned} u_k &= \frac{1}{N}\left(\sum_{i=1}^{N-1} i w_i + w_N\right) - \sum_{i=k}^{N-1} w_i, \text{ for } k = 1, \ldots, N-1 \\ u_N &= \frac{1}{N}\left(\sum_{i=1}^{N-1} i w_i + w_N\right), \end{aligned}$$

and therefore

$$\frac{\partial u_k}{\partial w_i} = \begin{cases} \frac{i}{N}, & i < k \\ \frac{i}{N} - 1, & k \le i < N-1 \\ \frac{1}{N}, & i = N. \end{cases}$$

This means that, if, $i \neq N$,

$$\begin{aligned} \frac{\partial}{\partial w_i}\{\psi(\mathbf{u}(\mathbf{w}))\} &= \sum_{k=1}^{N} \frac{\partial \psi}{\partial u_k}\frac{\partial u_k}{\partial w_i} \\ &= \frac{\partial \psi}{\partial u_k}\left[\sum_{k=1}^{i}(\frac{i}{N} - 1) + \sum_{k=i+1}^{N} \frac{i}{N}\right] \\ &= \frac{\partial \psi}{\partial u_k}\left[\frac{i^2}{N} - i + (N-i)\frac{1}{N}\right] = 0, \end{aligned}$$

where we used the fact (B.9) that $\frac{\partial\psi}{\partial u_k}$ is the same for all $k$. If $i = N$, then

$$\begin{aligned}\frac{\partial}{\partial w_i}\{\psi(\mathbf{u}(\mathbf{w}))\} &= \sum_{k=1}^{N}\frac{\partial\psi}{\partial u_k}\frac{\partial u_k}{\partial w_N} \\ &= \frac{\partial\psi}{\partial u_k}\left[\sum_{k=1}^{N}\frac{1}{N}\right] \\ &= \frac{\partial\psi}{\partial u_k}.\end{aligned}$$

Thus, $\psi$ does not depend on $w_1, \ldots, w_{N-1}$, only on $w_N = \sum_{i=1}^{N} u_i$. ■

**Proof of Proposition B.2.**

(i) Is straightforward.

(ii) According to Lemma B.1 above, if $\phi'(\mathbf{u}, \mathbf{h})$ satisfies Condition 1 of Proposition B.1, then

$$\phi'(\mathbf{u}, \mathbf{h}) - \mathbf{h}^T\mathbf{u}$$

is a function of $\mathbf{h}$ and $\sum_{i=1}^{N} u_i$ only. It also follows from the same Lemma that $\alpha$ of Proposition B.1 may only depend on $\sum_{i=1}^{N} u_i$. Therefore, there is a function $\psi$ such that

$$\phi'(\mathbf{u}, \mathbf{h}) = \mathbf{h}^T(\mathbf{u} - \mathbf{a}) + \psi(\mathbf{h}, \sum_{i=1}^{N} u_i).$$

Suppose that $\psi$ depends on $\mathbf{h}$ for $\mathbf{h} \neq \mathbf{0}, \mathbf{1}$. We presently show that this would lead to violating Condition 2 of Proposition B.1. Take $\mathbf{h}_1, \mathbf{h}_2 \in \{0,1\}^N \backslash \{\mathbf{0}, \mathbf{1}\}$, and $S \in I\!R$, such that

$$\psi(\mathbf{h}_1, S) > \psi(\mathbf{h}_2, S). \tag{B.10}$$

Denote the number of samples where $\mathbf{h}_1$ and $\mathbf{h}_2$ are different by $p$:

$$p = \|\mathbf{h}_2 - \mathbf{h}_1\|_1 \geq 1.$$

Let $k$ be the number of zeros in $\mathbf{h}_2$:

$$k = N - \|\mathbf{h}_2\|_1,$$

and let

$$\varepsilon = \frac{\psi(\mathbf{h}_1, S) - \psi(\mathbf{h}_2, S)}{2p} > 0.$$

**Case 1.** *There are $i$ and $j$ such that $h_{1,i} = h_{2,i} = 1$ and $h_{1,j} = h_{2,j} = 0$.*

Let $u_m = \alpha - \varepsilon$ if $h_{2,m} = 0$ and $m \neq j$, and let $u_m = \alpha + \varepsilon$ if $h_{2,m} = 1$ and $m \neq i$. If

$S - N\alpha - (N-2k)\varepsilon \geq 0$, let $u_i = \alpha + \varepsilon + S - N\alpha - (N-2k)\varepsilon$, and let $u_j = \alpha - \varepsilon$. Otherwise, let $u_i = \alpha + \varepsilon$ and $u_j = \alpha - \varepsilon + S - N\alpha - (N-2k)\varepsilon$. Then

$$\begin{aligned} \phi'(\mathbf{u}, \mathbf{h}_1) - \phi'(\mathbf{u}, \mathbf{h}_2) &= -p\varepsilon + \psi(\mathbf{h}_1, S) - \psi(\mathbf{h}_2, S) \\ &= -p\varepsilon + 2p\varepsilon = p\varepsilon > 0. \end{aligned} \tag{B.11}$$

On the other hand, the edges of $\mathbf{h}_2$ coincide with the $\alpha$-crossings of $\mathbf{u}$: $u_m > \alpha$ whenever $h_{2,m} = 1$ and $u_m < \alpha$ whenever $h_{2,m} = 0$. Therefore, if $\phi'$ is to satisfy Condition 2 of Proposition B.1, $\mathbf{h}_2$ has to be the optimal hypothesis for $\mathbf{u}$, which contradicts (B.11). We also note that $\sum_{m=1}^{N} u_m = S$ by construction.

**Case 2.** *There is no $i$ for which $h_{1,i} = h_{2,i} = 1$, and there is no $j$ for which $h_{1,j} = h_{2,j} = 0$, i.e., $h_{1,i} = 1 - h_{2,i}$ for $i = 1, \ldots, N$.*

If $2 \leq k \leq N-2$, let $i$ and $j$ be such that $h_{2,i} = h_{2,j} = 1$. Let $\mathbf{h}_3$ be obtained from $\mathbf{h}_2$ by changing the $i$-th entry from zero to one and the $j$-th entry from one to zero. Then either $\psi(\mathbf{h}_1, S) \neq \psi(\mathbf{h}_3, S)$, or $\psi(\mathbf{h}_2, S) \neq \psi(\mathbf{h}_3, S)$ (or both), and both pairs $(\mathbf{h}_1, \mathbf{h}_3)$ and $(\mathbf{h}_2, \mathbf{h}_3)$ fall under Case 1 considered above.

If $k = 1$, let $i$ be the index for which $h_{2,i} = 0$. Without loss of generality, assume that $i \neq 1$ and $i \neq N$. Form $\mathbf{h}'_1$ and $\mathbf{h}'_2$ as follows:

$$\begin{aligned} h'_{2,m} &= h_{2,m} \text{ for } m = 1, \ldots, N-1 \\ h'_{2,N} &= 0 \\ h'_{1,m} &= h_{1,m} \text{ for } m = 2, \ldots, N \\ h'_{1,1} &= 1. \end{aligned}$$

In other words,

$$\begin{aligned} h_1 &= (0, 0, \ldots, 0, 1, 0, \ldots, 0, 0)^T, \\ h'_1 &= (1, 0, \ldots, 0, 1, 0, \ldots, 0, 0)^T, \\ h_2 &= (1, 1, \ldots, 1, 0, 1, \ldots, 1, 1)^T, \\ h'_2 &= (1, 1, \ldots, 1, 0, 1, \ldots, 1, 0)^T. \end{aligned}$$

Then one of the following holds:

$$\begin{aligned} \psi(\mathbf{h}_1, S) &\neq \psi(\mathbf{h}'_1, S), \\ \psi(\mathbf{h}_2, S) &\neq \psi(\mathbf{h}'_2, S), \\ \psi(\mathbf{h}'_1, S) &\neq \psi(\mathbf{h}'_2, S), \end{aligned}$$

and all three pairs $(\mathbf{h}_1, \mathbf{h}'_1)$, $(\mathbf{h}_2, \mathbf{h}'_2)$, and $(\mathbf{h}'_1, \mathbf{h}'_2)$ fall under Case 1 considered above. The remaining cases are handled similarly to Cases 1 and 2. The conclusion is that, in each case, (B.10) leads to a violation of Condition 2 of Proposition B.1. This means that the inequality (B.10) cannot be true, and so $\psi(\mathbf{h}, S)$ is independent of $\mathbf{h}$, for $\mathbf{h} \in \{0,1\}^N \backslash \{\mathbf{0}, \mathbf{1}\}$. ∎

## ■ B.3 Proof of Lemma 4.3.

Consider the earliest time instant $t_1$ during the evolution of the SIDE when one of these $\alpha$-crossings $\{g_1, \ldots, g_\nu\}$ disappears. Let us denote the region being removed at $t_1$ by $(i, j)$. Suppose that, at time $t_1^-$, there are at least $\nu + 2$ $\alpha$-crossings remaining. We will presently show that the only $\alpha$-crossings among $\{g_1, \ldots, g_\nu\}$ that can be disappearing at $t_1$ are either $g_1$ or $g_\nu$, but not both.

**Case 1.** *The signal* $\mathbf{u}(t_1^-)$ *has at least* $\nu + 3$ *$\alpha$-crossings;* $i - 1$ *and* $j$ *are consecutive elements of the set* $\{g_1, \ldots, g_\nu\}$.

Since $\mathbf{u}(t_1^-)$ has at least $\nu + 3$ $\alpha$-crossings whereas the set $\{g_1, \ldots, g_\nu\}$ has only $\nu$ elements, there must be at least one pair of consecutive $\alpha$-crossings of $\mathbf{u}(t_1^-)$, say $i' - 1$ and $j'$, such that

**(i)** $i' - 1, j' \notin \{g_1, \ldots, g_\nu\}$, and

**(ii)** either the $\alpha$-crossing of $\mathbf{u}(t_1^-)$ immediately to the left of $i' - 1$ or the $\alpha$-crossing of $\mathbf{u}(t_1^-)$ immediately to the right of $j'$ is an element of $\{g_1, \ldots, g_\nu\}$.

Since the region $(i, j)$ disappears at time $t_1$ while the region $(i', j')$ stays, the energy (4.5) of the region $(i, j)$ must be smaller: $E_{ij} < E_{i'j'}$, or equivalently,

$$\left|\sum_{n=i}^{j}(u_n - \alpha)\right| < \left|\sum_{n=i'}^{j'}(u_n - \alpha)\right|.$$

Therefore, we can increase $\phi(\mathbf{u}^0, \mathbf{h})$ by changing $\mathbf{h}$ as follows: remove the edges $i - 1$ and $j$ and add edges at $i' - 1$ and $j'$. This contradicts our assumption that $\mathbf{h}$ is the best hypothesis with $\nu$ or fewer edges. The conclusion is that two consecutive $\alpha$-crossings from $\{g_1, \ldots, g_\nu\}$ cannot be erased.

**Case 2.** $i - 1 \notin \{g_1, \ldots, g_\nu\}$, *and* $j \in \{g_1, \ldots, g_\nu\}$.

Suppose that there are $\alpha$-crossings of $\mathbf{u}(t_1^-)$ to the left of $i - 1$. Let $i' - 1$ be the $\alpha$-crossing of $\mathbf{u}(t_1^-)$ immediately to the left of $i - 1$. Since the region $(i, j)$ is removed, we have: $E_{ij} < E_{i',i-1}$, which means that changing $\mathbf{h}$ by moving the edge from $j$ to $i' - 1$ will increase $\phi(\mathbf{u}^0, \mathbf{h})$, and so $\mathbf{h} \neq \mathbf{h}^*_{\leq\nu}(\mathbf{u}^0)$. Thus, it must be that either $i = 1$ or $i - 1$ is the leftmost $\alpha$-crossing of $\mathbf{u}(t_1^-)$, and therefore $j = g_1$.

**Case 3.** $i - 1 \in \{g_1, \ldots, g_\nu\}$, *and* $j \notin \{g_1, \ldots, g_\nu\}$.

This case is handled similarly to Case 2.

Cases 1, 2, and 3 combined imply that, at time $t_1$, only one of the $\alpha$-crossings $\{g_1, \ldots, g_\nu\}$ can disappear: either $g_1$ or $g_\nu$. Without loss of generality, suppose that it is $g_1$. We showed that it can only disappear at the time $t_1^-$ when region $(i, g_1)$ gets erased, where $i = 1$ or $i - 1$ is the leftmost $\alpha$-crossing of $\mathbf{u}(t_1^-)$. In what follows, we consider only the case $i = 1$; the other case is handled similarly.

We now show that no other $\alpha$-crossing from the set $\{g_1, \ldots, g_\nu\}$ can disappear before $t^+$. We consider the earliest time instant $t_2 > t_1$ when one of the remaining $\alpha$-crossings $\{g_2, \ldots, g_\nu\}$ disappears. Let us again denote the region being removed by

$(i, j)$. Suppose that, at time $t_2^-$, there are at least $\nu + 1$ $\alpha$-crossings remaining.
**Case 4.** *The indices $i-1$ and $j$ are consecutive elements of the set $\{g_2, \dots, g_\nu\}$.*
**Case 5.** *$i-1 \notin \{g_2, \dots, g_\nu\}$, and $j \in \{g_2, \dots, g_\nu\}$.*
**Case 6.** *$i-1 \in \{g_2, \dots, g_\nu\}$, and $j \notin \{g_2, \dots, g_\nu\}$.*
Cases 4, 5, and 6 are handled similarly to Cases 1, 2, and 3, respectively, with the result that only $(i, g_2)$ or $(g_\nu, j)$ can be removed, where $i-1$ is either 0 or the leftmost $\alpha$-crossing of $\mathbf{u}(t_2^-)$, and $j$ is either $N$ or the rightmost $\alpha$-crossing of $\mathbf{u}(t_2^-)$.
We now show how to handle the case when $(1, g_2)$ is removed at $t_2$; all other cases are handled using similar techniques.

Without loss of generality, we suppose that $g_1$ is an upward edge of $\mathbf{h}$. Then

$$\sum_{n=1}^{g_1}(u_n - \alpha) < 0, \tag{B.12}$$

as otherwise we would be able to improve $\mathbf{h}$ by removing the edge $g_1$. Also,

$$\sum_{n=g_1+1}^{g_2}(u_n - \alpha) > 0, \tag{B.13}$$

as otherwise removing the edges $g_1$ and $g_2$ would improve $\mathbf{h}$. Since the region $(1, g_1)$ disappeared at time $t_1$ while the $\alpha$-crossing at $g_2$ stayed, we have

$$\begin{aligned} E_{1,g_1} &< E_{g_1+1,g_2}, \\ \text{i.e.,} \left|\sum_{n=1}^{g_1}(u_n - \alpha)\right| &< \frac{1}{2}\left|\sum_{n=g_1+1}^{g_2}(u_n - \alpha)\right|. \end{aligned} \tag{B.14}$$

We combine (B.12), (B.13), and (B.14) to get

$$\begin{aligned} E_{1,g_2} &= \left|\sum_{n=1}^{g_2}(u_n - \alpha)\right| \\ &= \left|\sum_{n=g_1+1}^{g_2}(u_n - \alpha)\right| - \left|\sum_{n=1}^{g_1}(u_n - \alpha)\right| \\ &> \frac{1}{2}\left|\sum_{n=g_1+1}^{g_2}(u_n - \alpha)\right| \\ &= E_{g_1+1,g_2}. \end{aligned} \tag{B.15}$$

Choose the region $(i', j')$ of $\mathbf{u}(t_2^-)$ analogously to Case 1. Then, since $(1, g_2)$ gets removed at $t_2$ and $(i', j')$ does not,

$$E_{i'j'} > E_{1,g_2}.$$

Together with (B.15), this gives

$$E_{i'j'} > E_{g_1+1,g_2}.$$

Thus, $\mathbf{h}$ can be improved by replacing the edges $g_1$ and $g_2$ with $i'$ and $j'$, which is a contradiction. ■

## ■ B.4 Completion of the Proof of Proposition 4.6.

We need to show that the following situation is impossible:

- $\mathbf{h}_\psi = \mathbf{h}^*_{\leq\bar{\nu}}(\mathbf{u}^0) \neq \mathbf{h}^*_{\leq\bar{\nu}-1}(\mathbf{u}^0)$, and
- $\rho_{i^*,j^*} = 2$, i.e., the SIDE's solution goes from $\bar{\nu}+1$ zero-crossings directly to $\bar{\nu}-1$.

We denote the edges of $\mathbf{h}_1 = \mathbf{h}^*_{\leq\bar{\nu}+1}(\mathbf{u}^0)$ by $g_1, \ldots, g_{\bar{\nu}+1}$, and show that each of the three possibilities allowed by Proposition 4.4 in the situation above, leads to a contradiction.
**Case (ii)**. $\mathbf{h}^*_{\leq\bar{\nu}}(\mathbf{u}^0)$ *has edges at the locations* $g_2, \ldots, g_{\bar{\nu}+1}$.
Then

$$\begin{aligned} \eta(\mathbf{h}^*_{\leq\bar{\nu}}(\mathbf{u}^0)) - \eta(\mathbf{h}_1) &= E_{1,g_1} - \lambda, \\ \eta(\mathbf{h}_2) - \eta(\mathbf{h}^*_{\leq\bar{\nu}}(\mathbf{u}^0)) &= 2E^* - E_{1,g_1} - \lambda, \end{aligned}$$

and so, for $\mathbf{h}^*_{\leq\bar{\nu}}(\mathbf{u}^0)$ to be better (with respect to $\eta$) than both $\mathbf{h}_1$ and $\mathbf{h}_2$, we need to have:

$$\begin{aligned} E_{1,g_1} &< \lambda, \text{ and} \\ 2E^* - E_{1,g_1} &> \lambda, \end{aligned}$$

from which it follows that

$$\begin{aligned} 2E^* - E_{1,g_1} &> E_{1,g_1} \Rightarrow \\ E^* &> E_{1,g_1}. \end{aligned}$$

The latter inequality contradicts the definition of $E^*$ as the smallest energy of any region of $\mathbf{u}(t)$.
**Case (iii)**. $\mathbf{h}^*_{\leq\bar{\nu}}(\mathbf{u}^0)$ *has edges at the locations* $g_1, \ldots, g_{\bar{\nu}}$.
This case is handled similarly to Case 2.
**Case (iv)**. $\mathbf{h}^*_{\leq\bar{\nu}}(\mathbf{u}^0)$ *has edges at the locations* $\{g_1, \ldots, g_{\bar{\nu}+1}\}\backslash\{i^*, j^*\}$, *as well as one edge at some other location* $g_1'$. *This situation requires considering several sub-cases—see Case 6 of Section B.3. As they are similar, we only treat the one where the region* $(1, g_1')$ *was removed before time* $t$.
Then

$$\begin{aligned} \eta(\mathbf{h}^*_{\leq\bar{\nu}}(\mathbf{u}^0)) - \eta(\mathbf{h}_1) &= 2E^* - E_{1,g_1'} - \lambda, \\ \eta(\mathbf{h}_2) - \eta(\mathbf{h}^*_{\leq\bar{\nu}}(\mathbf{u}^0)) &= E_{1,g_1'} - \lambda. \end{aligned}$$

If $\mathbf{h}_\psi = \mathbf{h}^*_{\le\bar{\nu}}(\mathbf{u}^0)$, we then must have

$$\begin{aligned} E_{1,g_1'} &> \lambda, \text{ and} \\ 2E^* - E_{1,g_1'} &< \lambda, \end{aligned}$$

from which it follows that

$$\begin{aligned} 2E^* - E_{1,g_1'} &< E_{1,g_1'} \Rightarrow \\ E^* &< E_{1,g_1'}. \end{aligned} \tag{B.16}$$

But the region $(1, g_1')$ disappeared before $(i^*, j^*)$, and therefore

$$E^* \ge E_{1,g_1'},$$

which contradicts (B.16). This concludes the proof. ∎

## ■ B.5 Equivalence of the SIDE (4.1) to a Linear Programming Problem.

We need to show that a solution to the linear program (4.24-4.27) exists and consists of ones and zeros. There are $2N-1$ variables involved $(h_1, \ldots, h_N, r_1, \ldots, r_{N-1})$, and it is easily seen that there are $2N-1$ linearly independent constraints in (4.25-4.27). Moreover, $h_1 = \ldots = h_N = r_1 = \ldots = r_{N-1} = 0$ satisfies all the constraints. Thus, (4.25-4.27) define a non-empty polyhedron in $\mathbb{R}^{2N-1}$ with at least one vertex (see [65], Theorem 2.8), which means that a solution to (4.24-4.27) exists. We assume non-degeneracy—i.e., that the solution is unique. This is not necessary for the proof, but it greatly simplifies notation. Now suppose that, for this solution $\mathbf{h}$, there exist $p$ and $q$ such that $1 \le p < q \le N-1$, and

$$h_p \ne h_{p+1} = h_{p+2} = \ldots = h_q \ne h_{q+1}. \tag{B.17}$$

We now show that, unless $h_q = 0$ or $h_q = 1$, we can change $\mathbf{h}$ to make (4.24) smaller, and that therefore $\mathbf{h}$ cannot be a solution if $0 < h_q < 1$. We will be changing $h_{p+1}, \ldots, h_q$, and so let us write out the portion of (4.24) which depends on them:

$$-sh_q + \lambda(|h_q - h_p| + |h_q - h_{q+1}|), \tag{B.18}$$

where $s = \sum_{i=p+1}^{q} u_i$.
**Case 1.** Suppose that

$$h_p < h_q \text{ and } h_{q+1} < h_q. \tag{B.19}$$

Then the $h_q$-dependent portion of Equation (B.18) is:

$$h_q(2\lambda - s). \tag{B.20}$$

If $2\lambda > s$, make $h_q = \max(h_p, h_{q+1})$, which will reduce $h_q$ and therefore also reduce (B.20). It will also make either $h_p = h_q$ or $h_q = h_{q+1}$, violating our assumption (B.17). If $2\lambda < s$, make $h_q = 1$, which will also reduce (B.20). In the degenerate case $2\lambda = s$, we can go either way without changing the solution. Thus, if (B.17) and (B.19) hold, then $h_q = 1$.

**Case 2.** $h_p > h_q$ and $h_{q+1} > h_q$. This is handled similarly, with the result that $h_q = 0$.

**Case 3.** $h_p < h_q$ and $h_{q+1} > h_q$. Then the $h_q$-dependent portion of Equation (B.18) is:

$$-h_q s. \tag{B.21}$$

If $s > 0$, increase $h_q$ by setting $h_q = h_{q+1}$. If $s < 0$, reduce $h_q$ by setting $h_q = h_p$. Both will reduce (B.21) and violate (B.17). In the degenerate case $s = 0$, we can go either way.

**Case 4.** $h_p > h_q$ and $h_{q+1} < h_q$. This is handled similarly to Case 3.

The situations when $p = 0$ and/or $q = N$ are also handled similarly. The conclusion is that the optimal solution consists of zeros and ones, which means that the linear program (4.24-4.27) is equivalent to the optimization problem (4.23). ∎

# Bibliography


[1] L. Alvarez, P.L. Lions, and J.-M. Morel. Image selective smoothing and edge detection by nonlinear diffusion, II. *SIAM J. Numer. Anal.*, 29(3), 1992.

[2] M.S. Atkins and B.T. Mackiewich. Fully automatic segmentation of the brain in MRI. *IEEE Trans. on Medical Imaging*, 17(1), February 1998.

[3] M. Basseville and I.V. Nikiforov. *Detection of Abrupt Changes: Theory and Application.* Prentice Hall, 1993.

[4] D.P. Bertsekas and S.K. Mitter. A descent method for optimization problems with nondifferentiable cost functionals. *SIAM J. Control*, 11(4), November 1973.

[5] M.J. Black, G. Sapiro, D.H. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Trans. on Image Processing*, 7(3), 1998.

[6] C. Bouman and K. Sauer. An edge-preserving method for image reconstruction from integral projections. In *Proc. Conf. on Info. Sci. and Syst.*, pages 383-387, Baltimore, MD, March 1991.

[7] C. Brice and C. Fennema. Scene analysis using regions. *Artificial Intelligence*, 1, 1970.

[8] C.B. Burckhardt. Speckle in ultrasound B-mode scans. *IEEE Trans. on Sonics and Ultrasonics*, SU-25, January 1978.

[9] J. Canny. A computational approach to edge detection. *IEEE Trans. on PAMI*, PAMI-8, 1986.

[10] T.-H. Chang, Y.-C. Lin, and C.-C. J. Kuo. Techniques in texture analysis. *IEEE Trans. on Image Processing*, October 1993.

[11] T.H. Cormen, C.E. Leiserson, and R.L. Rivest. *Introduction to Algorithms.* MIT Press, 1990.

[12] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. ICCV*, pages 694–699, Cambridge, MA, 1995.

[13] R.N. Czerwinski, D.L. Jones, and W.D. O'Brien, Jr. Line and boundary detection in speckle images. *IEEE Trans. on Image Processing*, 7(12), 1998.

[14] G. Dahlquist and A. Bjorck. *Numerical Methods.* Prentice Hall, 1974.

[15] S.Z. Der and R. Chellappa. Probe-based automatic target tecognition in infrared imagery. *IEEE Trans. on Image Processing*, 6(1), 1997.

[16] P. Felzenszwalb and D. Huttenlocher. Image segmentation using local variation. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 98-104, 1998.

[17] A.F. Filippov. *Differential Equations with Discontinuous Righthand Sides.* Kluwer Academic Publishers, 1988.

[18] C.H. Fosgate, H. Krim, W.W. Irving, W.C. Karl, and A.S. Willsky. Multiscale segmentation and anomaly enhancement of SAR imagery. *IEEE Trans. on Image Processing*, 6(1), 1997.

[19] M.G. Fleming, C. Steger, J. Zhang, J. Gao, A.B. Cognetta, I. Pollak, and C.R. Dyer. Techniques for a structural analysis of dermatoscopic imagery. *Computerized Medical Imaging and Graphics*, 22, 1998.

[20] J. Froment and S. Parrino. MegaWave2 software package. CEREMADE, URA CNRS 749 de l'Université Paris-IX Dauphine, 1995.

[21] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. on PAMI*, 14(3), 1992.

[22] J.W. Goodman. Statistical properties of laser speckle patterns. In *Topics in Applied Physics, vol. 9: Laser Speckle and Related Phenomena*, 2nd edition, J.C. Dainty, Editor. Springer-Verlag, 1984.

[23] D.R. Greer, I. Fung, and J.H. Shapiro Maximum-likelihood multiresolution laser radar range imaging. *IEEE Trans. on Image Processing*, 6(1), 1997.

[24] K. Haris, S.N. Efstratiadis, N. Maglaveras, and A.K. Katsaggelos. Hybrid image segmentation using watersheds and fast region merging. *IEEE Trans. on Image Processing*, 7(12), 1998.

[25] B. Hassibi, A.H. Sayed, and T. Kailath. Linear estimation in Krein spaces–part i: Theory. *IEEE Trans. Automatic Control*, 41(1), 1996.

[26] B. Hassibi, A.H. Sayed, and T. Kailath. Linear estimation in Krein spaces–part ii: Applications. *IEEE Trans. Automatic Control*, 41(1), 1996.

[27] G.T. Herman. *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography.* Academic Press, 1980.

[28] D.V. Hinkley. Inference about the change-point in a sequence of random variables. *Biometrica*, 57(1), 1970.

[29] B.K.P. Horn. *Robot Vision.* MIT Press, 1986.

[30] B.K.P. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17, pages 185-203, 1981.

[31] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. of Comp. Vis.*, 1:321–331, 1988.

[32] I.B. Kerfoot and Y. Bresler. Theoretical analysis of multispectral image segmentation criteria. *IEEE Trans. on Image Processing*, 8(6), 1999.

[33] R.L. Kettig and D.A. Landgrebe. Classification of multispectral image data by extraction and classification of homogeneous objects. *IEEE Trans. on Geoscience Electronics*, GE-14(1), 1976.

[34] S. Kichenassamy. The Perona-Malik paradox. *SIAM J. Applied Math.*, 57, 1997.

[35] J. Koenderink. The structure of images. *Biological Cybernetics*, 50, 1984.

[36] G. Koepfler, C. Lopez, and J.-M. Morel. A multiscale algorithm for image segmentation by variational method. *SIAM J. Numer. Anal.*, 31(1), 1994.

[37] B. Kosko. *Neural Networks for Signal Processing,* pages 37-61. Prentice Hall, 1992.

[38] H. Krim and Y. Bao. A stochastic diffusion approach to signal denoising. In *Proc. ICASSP*, Phoenix, AZ, 1999.

[39] Y. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3, 1989.

[40] S.G. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, 1998.

[41] D. Marr. *Vision.* W.H. Freeman and Co., 1982.

[42] Vision and Modeling Group, MIT Media Lab. *Vision Texture Database.* http://www-white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html

[43] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation.* Birkhauser, 1995.

[44] D. Mumford and J. Shah. Boundary detection by minimizing functionals, I. In *Proc. CVPR*, pages 22–26, San Francisco, June 1985.

[45] K.M. Nagpal and P.P. Khargonekar. Filtering and smoothing in an $H^\infty$ setting. *IEEE Trans. Automatic Control*, 36(2), 1991.

[46] S. Osher and L.I. Rudin. Feature-oriented image enhancement using shock filters. *SIAM J. Numer. Anal.*, 27(4), 1990.

[47] T. Pavlidis. Segmentation of pictures and maps through functional approximation. *Computer Graphics and Image Processing*, 1, 1972.

[48] P. Perona. Orientation diffusions. *IEEE Trans. on Image Processing*, 7(3), 1998.

[49] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. on PAMI*, 12(7), 1990.

[50] P. Perona, T. Shiota, and J. Malik. Anisotropic diffusion. In [55].

[51] I. Pollak, A. S. Willsky, and H. Krim. A nonlinear diffusion equation as a fast and optimal solver of edge detection problems. In *Proc. ICASSP*, Phoenix, AZ, 1999.

[52] S. Resnick. *Adventures in Stochastic Processes.* Birkhauser, 1992.

[53] J. Rissanen. *Stochastic complexity in statistical inquiry.* World Scientific, 1989.

[54] R.T. Rockafellar. *Convex Analysis.* Princeton Univerity Press, 1970.

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

[56] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 1992.

[57] G. Sapiro. From active contours to anisotropic diffusion: connections between basic PDE's in image processing. In *Proc. ICIP*, Lausanne, 1996.

[58] K. Sauer and C. Bouman. Bayesian estimation of transmission tomograms using segmentation based optimization. *IEEE Trans. on Nuclear Science*, 39(4), August 1992.

[59] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans. on Signal Processing*, 41(2), 1993.

[60] P. Schmid. Segmentation of digitized dermatoscopic images by two-dimensional color clustering. *IEEE Trans. on Medical Imaging*, 18(2), February 1999.

[61] U. Shaked and Y. Theodor. $H_\infty$-optimal estimation: A tutorial. In *Proc. of the 31st CDC*, Tucson, Arizona, 1992.

[62] G. Strang. *Linear Algebra and Its Applications,* 3rd Edition. Harcourt Brace Jovanovich, 1988.

[63] P.C. Teo, G. Sapiro, and B. Wandell. Anisotropic smoothing of posterior probabilities. In *Proc. ICIP*, Santa Barbara, CA, 1997.

[64] P.C. Teo, G. Sapiro, and B. Wandell. Creating connected representations of cortical gray matter for functional MRI visualization. *IEEE Trans. on Medical Imaging*, 16(6), December 1997.

[65] J. Tsitsiklis. *Introduction to Mathematical Programming.* 6.251J Course Notes, MIT, 1995.

[66] H. van Trees. *Detection, Estimation, and Modulation Theory*, volume 1. Wiley, 1968.

[67] V.I. Utkin. *Sliding Modes in Control and Optimization*. Springer-Verlag, 1992.

[68] P. Viola and W.M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2), 1997.

[69] J. Weickert. Nonlinear diffusion scale-spaces: from the continuous to the discrete setting. In *Proc. ICAOS: Images, Wavelets, and PDEs*, pages 111–118, Paris, 1996.

[70] R.F. Wagner, S.W. Smith, J.M. Sandrik, and H. Lopez. Statistics of speckle in ultrasound B-scans. *IEEE Trans. on Sonics and Ultrasonics*, SU-30, May 1983

[71] M. Wertheimer. Laws of organization in perceptual forms. In *A Sourcebook of Gestalt Psychology*, W.B. Ellis, Editor. Harcourt, Brace, and Company, 1938.

[72] A. Witkin. Scale-space filtering. In *Int. Joint Conf. on AI*, pages 1019–1022, Karlsruhe, 1983.

[73] P. Wolfe. A method of conjugate gradients for minimizing nondifferentiable functions. In *Nondifferentiable Optimization*, M.L. Balanski and P. Wolfe, Editors. North-Holland Publishing, 1975.

[74] G. Wyszecki and W.S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley, 1982.

[75] S.C. Zhu and D. Mumford. Prior learning and Gibbs reaction-diffusion. *IEEE Trans. on PAMI*, 19(11), 1997.

[76] S.C. Zhu and A. Yuille. Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. on PAMI*, 18(9), 1996.

# Index