On Discontinuity-Adaptive Smoothness Priors in Computer Vision

S. Z. Li

Abstract— **A variety of analytic and probabilistic models in connection to Markov random fields (MRFs) have been proposed in the last decade for solving low level vision problems involving *discontinuities*. This paper presents a systematic study on these models and defines a general discontinuity adaptive (DA) MRF model. By analyzing the *Euler equation* associated with the energy minimization, it shows that the fundamental difference between different models lies in the behavior of *interaction* between neighboring points, which is determined by the *a priori* smoothness constraint encoded into the energy function An important necessary condition is derived for the interaction to be adaptive to discontinuities to avoid oversmoothing. This forms the basis on which a class of *adaptive interaction functions* (AIFs) is defined. The DA model is defined in terms of the Euler equation constrained by this class of AIFs. Its solution is C^1 continuous and allows arbitrarily large but bounded slopes in dealing with discontinuities. Because of the continuous nature, it is stable to changes in parameters and data, a good property for regularizing ill-posed problems. Experimental results are shown.**

Keywords— **Discontinuities, energy functions, Euler equation, computer vision, Markov random fields, minimization, regularization.**

I. Introduction

SMOOTHNESS is a generic assumption underlying a wide range of physical phenomena. It characterizes the coherence and homogeneity of matter within a scope of space (or an interval of time). It is one of the most common assumptions in computer vision models, in particular, those formulated in terms of Markov random fields (MRFs) [1], [2], [3] and regularization [4]. Its applications are seen widely in image restoration, surface reconstruction, optical flow and motion, shape from X, texture, edge detection, region segmentation, visual integration and so on.

The assumption of the uniform smoothness implies the smoothness *everywhere*. However, improper imposition of it can lead to undesirable, oversmoothed, solutions. This occurs when the uniform smoothness is violated, for example, at discontinuities where abrupt changes occur. It is necessary to take care of discontinuities when using smoothness priors. Therefore, how to apply the smoothness constraint wile preserving discontinuities has been one of the most active research areas in level vision (see, *e.g.* [5], [6], [1], [7], [3], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18]).

This paper presents a systematic study on smoothness priors involving discontinuities. The results are based on an analysis of the *Euler equation* associated with the energy minimization in MRF and regularization models. Through the analysis, it is identified that the fundamental difference among different models for dealing with discontinuities lies in their ways of controlling the *interaction* between neighboring points. Thereby, an important necessary condition is derived for any regularizers or MRF prior potential functions to be able to deal with discontinuities.

S. Z. Li is currently with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 2263. E-mail: szli@ntuix.ntu.ac.sg .

Based on these findings, a so-called *discontinuity adaptive* (DA) smoothness model is defined in terms of the Euler equation constrained by a class of *adaptive interaction functions* (AIFs). The DA solution is C^1 continuous, allowing arbitrarily large but bounded slopes. Because of the continuous nature, it is stable to changes in parameters and data. This is a good property for regularizing ill-posed problems. The results provide principles for the selection of *a priori* clique potential functions in stochastic MRF models and regularizers in deterministic regularization models. It is also shown that the DA model includes as special instances most of the existing models, such as the line process (LP) model [1], [3], weak string and membrane [10], approximations of the LP model [19], [20], minimal description length [13], biased anisotropic diffusion [16], and mean field theory approximation [17].

The study of discontinuities is most sensibly carried out in terms of analytical properties, such as derivatives. For this reason, analytical regularization, a special class of MRF models, is used as the platform for it. If we consider that regularization contains three parts [21]: the data, the class of solution functions and the regularizer, the present work addresses mainly the regularizer part. In Section II, regularization models are reviewed in connection to discontinuities. In Section III, a necessary condition for the discontinuity adaptivity is made explicit; based on this, the DA model is defined and compared with other models. In Section IV, an algorithm for finding the DA solution is presented, some related issues are discussed. Experimental results are shown in Section V. Finally, conclusions are drawn.

II. Smoothness, Regularization and Discontinuities

In MRF vision modeling, the smoothness assumption can be encoded into an energy via one of the two routes: analytic and probabilistic. In the analytic route, the encoding is done in the *regularization* framework [4], [22]. From the regularization viewpoint, a problem is said to be "ill-posed" if it fails to satisfy one or more of the following criteria: the solution exists, is unique and depends continuously on the data. Additional, *a priori*, assumptions have to be imposed on the solution to convert an ill-posed problem into a well-posed one. An important assumption of such assumptions is the *smoothness* [23]. It is incorporated into the energy function whereby the cost of the solution is defined.

From the probabilistic viewpoint, a regularized solution corresponds to the *maximum a posteriori* (MAP) estimate of an MRF [1], [24]. Here, the prior constraints are encoded into the *a priori* MRF probability distribution. The MAP solution is obtained by maximizing the posterior probability or equivalently minimizing the corresponding energy.

The MRF model is more general than the regularization model in (1) that it can encode prior constraints other than the smoothness and (2) that it allows arbitrary neighborhood systems other than the nearest ones. However, the analytic regularization model provides a convenient platform for the study of smoothness priors because of close relationships between the smoothness and the analytical continuity.

A. Regularization and Discontinuities

Consider the problem of restoring a signal f from the data $d = f + e$ where e denotes the noise. The regularization formulation defines the solution f^* to be the global minimum of an energy function $E(f)$, $f^* = \arg\min_f E(f)$. The energy is the sum of two terms

$$E(f) \stackrel{\triangle}{=} U(f \mid d) = U(d \mid f) + U(f) \tag{1}$$

The *closeness* term, $U(d \mid f)$, measures the cost caused by the discrepancy between the solution f and the data d

$$U(d \mid f) = \int_a^b \chi(x)[f(x) - d(x)]^2 dx \tag{2}$$

where $\chi(x)$ is a weighting function and a and b are the bounds of the integral. The *smoothness* term, $U(f)$, measures the cost caused by the irregularities of the solution f, the irregularities being measured by the derivative magnitudes $|f^{(n)}(x)|$. With identical independent additive Gaussian noise, $U(f \mid d)$, $U(d \mid f)$ and $U(f)$ correspond to the energies in the posterior, the likelihood the prior Gibbs distributions of an MRF, respectively [24].

The smoothness term $U(f)$, also called a *regularizer*, is the object of study in this work. It penalizes the irregularities according to the *a priori* smoothness constraint encoded in it. It is generally defined as

$$U(f) \stackrel{\triangle}{=} \sum_{n=1}^{N} U_n(f) = \sum_{n=1}^{N} \lambda_n \int_a^b g(f^{(n)}(x))dx \tag{3}$$

where $U_n(f)$ is the n^{th} order regularizer, N is the highest order to be considered and $\lambda_n \geq 0$ is a weighting factor. A *potential function* $g(f^{(n)}(x))$ is the penalty against the irregularity in $f^{(n-1)}(x)$ and corresponds to prior clique potentials in MRF models. Regularizers differ in the definition of $U_n(f)$, more specifically in the selection of g.

A.1 Standard Regularization

In the standard regularization [23], [4], the potential function takes the pure quadratic form

$$g_q(\eta) = \eta^2 \tag{4}$$

With g_q, the more irregular $f^{(n-1)}(x)$ is at x, the larger $|f^{(n)}|$, and consequently the larger potential $g(f^{(n)})$ contributed to $U_n(f)$. The standard quadratic regularizer can have a more general form

$$U_n(f, w_n) = \lambda_n \int_a^b w_n(x)[f^{(n)}(x)]^2 dx \tag{5}$$

where $w_n(x)$ are the pre-specified non-negative continuous functions [23]. It may also be generalized to multi-dimensional cases and to include cross derivative terms.

The quadratic regularizer imposes the smoothness constraint *everywhere*. It determines the constant *interaction* between neighboring points and leads to *smoothing strength* proportional to $|f^{(n)}|$, as will be shown in the next section. The homogeneous or isotropic application of the smoothness constraint inevitably leads to oversmoothing at discontinuities at which the derivative is infinite.

If the function $w_n(x)$ can be pre-specified in such a way that $w_n(x) = 0$ at x where $f^{(n)}(x)$ is infinite, then the oversmoothing can be avoided. In this way, $w_n(x)$ act as *continuity-controllers* [6]. It is further suggested that $w_n(x)$ may be discontinuous and not pre-specified [9]. For example, by regarding $w_n(x)$ as unknown functions, one could solve these unknowns using variational methods. But how well $w_n(x)$ can thus be derived remains unclear. The introduction of *line processes* [1], [3] or *weak continuity constraints* [10] provides a solution to this problem.

A.2 Line Process Model and Its Approximations

The LP model assumes *piecewise smoothness* whereby the smoothness constraint is *switched off* at points where the magnitude of the signal derivative exceeds certain threshold. It is defined on a lattice rather than on a continuous domain. Quantize the continuous interval $[a, b]$ into m uniformly spaced points $x_1, ..., x_m$ so that $f_i = f(x_i)$, $d_i = d(x_i)$ and $\chi_i = \chi(x_i)$. Introduce a set of binary line process variables $l_i \in \{0, 1\}$ into the smoothness term. If $w_i = w_1(x_i)$ in (5) also takes on a value in $\{0, 1\}$, then l_i is related it by $l_i = 1 - w_i$. The on-state, $l_i = 1$, of the line process variable indicates that a discontinuity is detected between neighboring points $i-1$ and i; the off-state, $l_i = 0$, indicates that the signal between the two points is continuous. Each turn-on of a line process variable is penalized by a quantity $\lambda\alpha$. These give the LP regularizer

$$U(f, l) = \lambda \sum_{i=2}^{m} [f_i - f_{i-1}]^2[1 - l_i] + \lambda\alpha \sum_{i=2}^{m} l_i \tag{6}$$

The energy for the LP model is

$$U(f, l \mid d) = \sum_{i=1}^{m} \chi_i [f_i - d_i]^2 + \lambda \left(\sum_{i=2}^{m} [f_i - f_{i-1}]^2[1 - l_i] + \alpha \sum_{i=2}^{m} l_i \right) \tag{7}$$

This is the *weak string* model and its extension to 2D the *weak membrane* model [10]. Eq.(7) corresponds to the energy in the posterior distribution of the surface field f and the line process field l, the distribution being of the Gibbs form

$$P(f,l \mid d) = \frac{1}{Z} \mathrm{e}^{-U(f,l \mid d)} \tag{8}$$

where Z is a normalizing constant called the partition function.

The line process variables are determined as follows: If $[f_i - f_{i-1}]^2 < \alpha$, then it is cheaper to pay the price $[f_i - f_{i-1}]^2$ and set $l_i = 0$; otherwise it is more economical to turn on the variable $l_i = 1$ to insert a discontinuity with the cost of α. This is an interpretation of the LP model based on the concept of the weak continuity constraint introduced by Blake [5] for edge labeling. An earlier idea of weak constraints can be found in Hinton's thesis work [25]. In the LP model, the *interaction* is piecewise constant (1 or 0) and the *smoothing strength* at i is either proportional to $|f_i - f_{i-1}|$ or zero – see the next section. The concept of discontinuities can be extended to model-based recognition of overlapping objects [26], [27]. There, the relational bond between any two features in the scene should be broken if the features are ascribed to two different objects.

Finding $f^* \in \mathbb{R}^m$ and $l^* \in \{0,1\}^m$ such that $U(f,l \mid d)$ is minimized is a mixture of real and combinatorial optimization. Algorithms for this can be classified as categories: stochastic [1], [3] and deterministic [19], [20], [10], [17]. Some annealing techniques are often combined into them to obtain global solutions.

In stochastic approaches, f and l are updated according to some probability distribution parameterized by a temperature parameter. For example, Geman and Geman propose to use simulated annealing with the Gibbs sampler [1] to find the global MAP solution. Marroquin [3] minimizes the energy by a stochastic update in l together with a deterministic update in f using gradient descent.

Deterministic approaches often use some classical gradient based methods. Before these can be applied, the combinatorial minimization problem has to be converted into one of real minimization. By eliminating the line process, Blake and Zisserman [10] convert the previous minimization problem into one which minimizes the following function containing only real variables

$$U(f \mid d) = \sum_{i=1}^{m} \chi_i [f_i - d_i]^2 + \lambda \sum_{i=2}^{m} g_\alpha(f_i - f_{i-1}) \tag{9}$$

where the truncated quadratic potential function

$$g_\alpha(\eta) = \min\{\eta^2, \alpha\} \tag{10}$$

shall be referred to as the *line process potential function.* Blake and Zisserman introduce a parameter p into $g_\alpha(\eta)$ to control the convexity of E, obtaining $g_\alpha^{(p)}(\eta)$. The parameter p varies from 1 to 0, which corresponds to the variation from a convex approximation of the function to its original form.

Koch *et al.* [19] and Yuille [20] perform the conversion using the Hopfield approach [28]. Continuous variables $\bar{l}_i$ in the range $[0,1]$ are introduced to replace the binary line process variables l_i in $\{0,1\}$. Each $\bar{l}_i$ is related to an internal variable $v_i \in (-\infty, +\infty)$ by a sigmoid function

$$\bar{l}_i = \theta(v_i) = 1/(1 + \mathrm{e}^{-v_i/\tau}) \tag{11}$$

with $\tau >$ as the parameter whereby $\lim_{\tau \to 0} \bar{l}_i = l_i$. The energy with this treatment is

$$\begin{aligned} U(f,\bar{l} \mid d) \quad = \quad & \sum_{i=1}^{m} \chi_i [f_i - d_i]^2 \\ & + \lambda \left(\sum_{i=2}^{m} [f_i - f_{i-1}]^2 [1 - \bar{l}_i] + \alpha \sum_{i=2}^{m} \bar{l}_i \right) \\ & + \sum_{i=2}^{m} \int_0^{\bar{l}_i} \theta^{-1}(l) dl \end{aligned} \tag{12}$$

It is shown that at stationary points where $dE/\mathrm{d}t = 0$, there are $v_i = \lambda([f_i - f_{i-1}]^2 - \alpha)$ and hence the approximated line process variables [20]

$$\bar{l}_i = \frac{1}{1 + \mathrm{e}^{-\lambda([f_i - f_{i-1}]^2 - \alpha)/\tau}} \tag{13}$$

This gives the effective potential function as

$$g_{\alpha,\tau}(\eta) = -\frac{\tau}{2} \log[1 + \mathrm{e}^{-\lambda(\eta^2 - \alpha)/\tau}] \tag{14}$$

As the temperature decreases toward zero, $\bar{l}_i$ approaches l_i: $\lim_{\tau \to 0^+} \bar{l}_i = l_i$.

Geiger and Girosi [17] approximate the line process using mean field theory. They introduce a parameter β into (8), giving an approximated posterior probability

$$P_\beta(f,l \mid d) = \frac{1}{Z_\beta} [P(f,l \mid d)]^\beta \tag{15}$$

Using the saddle point approximation method, they derive mean field equations which yield the approximated line process variables which are identical to (13). The solution is found in the limit when $\beta \to \infty$.

Li [15] proposes a continuous *adaptive regularizer* model. There, the smoothness constraint is applied without the switch-off as the LP model. Its effect is decreased as the derivative magnitude becomes larger and is completely off only at the true discontinuities where the derivative is infinite. This is an earlier form of the DA model in this work.

B. *Other Regularization Models*

Grimson and Pavlidis [7] propose an approach in which the degree of interaction between pixels across edges is adjusted in order to detect discontinuities. Lee and Pavlidis [11] investigate a class of smoothing splines which are piecewise polynomials. Errors of fit are measured after each successive regularization and used to determine whether discontinuities should be inserted. This process iterates until convergence is reached. Besl *et al.* [29] propose a smoothing window operator to prevent smoothing across discontinuities based on robust statistics. Liu and Harris

[30] develop, based on a previous work [31], a computational network in which surface reconstruction, discontinuity detection and estimation of first and second derivatives are performed cooperatively.

Mumford and Shah [8] define an energy on a continuous domain

$$U(f,\{a_i\},k \mid d) = \int_a^b [f(x)-d(x)]^2 dx + \lambda \sum_{i=0}^{k} \int_{a_i}^{a_{i+1}} [f'(x)]^2 dx + \alpha k \quad (16)$$

where λ and α are constants, k is the number of discontinuities and the sequence $a = a_0 < a_1 < ... < a_k < a_{k+1} = b$ indicates the locations of discontinuities. The minimal solution $(f^*, \{a_i\}^*, k^*)$ is found by minimizing over each value of the integer k, every sequence $\{a_k\}$, and every function $f(x)$ continuously differentiable on each interval $a_i \le x \le a_{i+1}$. The minimization over k is a hard problem.

Using the minimal description length principle, Leclerc [13] presents the following function for restoration of piecewise constant image f from noisy data d

$$U(f \mid d) = \sum_{i=1}^{m} [f_i - d_i]^2 + \lambda \sum_{i=2}^{m} [1 - \delta(f_i - f_{i-1})] \quad (17)$$

where $\delta(\cdot) \in \{0,1\}$ is the Kronecker delta function. To minimize the function, he approximates the delta function with the exponential function parameterized by τ

$$U(f \mid d) = \sum_{i=1}^{m} [f_i - d_i]^2 + \lambda \sum_{i=2}^{m} [1 - \mathrm{e}^{(f_i - f_{i-1})^2/\tau}] \quad (18)$$

and approaches the solution by continuation in τ toward 0.

III. The Discontinuity Adaptive MRF Model

By analyzing the smoothing mechanism in terms of the Euler equation, it will be clear that major difference between different models lies in their way of controlling the interaction between neighboring points and adjusting the smoothing strength. The DA model is defined based on the principle that wherever a discontinuity occurs, the interaction should diminish.

A. Defining the DA Model

We focus on the models which involve only the first order derivative $f' = f^{(1)}$ and consider the general string model

$$E(f) = U(f \mid d) = \int_a^b u(f \mid d) dx \quad (19)$$

where

$$u(f \mid d) = \chi(x)[f(x) - d(x)]^2 + \lambda g(f'(x)) \quad (20)$$

Solutions f^* minimizing $U(f \mid d)$ must satisfy the following associated Euler-Lagrange differential equation or simply the *Euler equation* [32]

$$u_f(f, f') - \frac{d}{dx} u_{f'}(f, f') = 0 \quad (21)$$

with the boundary conditions

$$f(a) = f_a, \quad f(b) = f_b \quad (22)$$

where f_a and f_b are prescribed constants. In the following discussion of solutions to the differential equation, the following assumptions are made: Both $\chi(x)$ and $d(x)$ are continuous and $f(x)$ is continuously differentiable[1]. Writing the Euler equation out yields

$$2\chi(x)[f(x) - d(x)] - \lambda \frac{d}{dx} g'(f'(x)) = 0 \quad (23)$$

A potential function g is usually chosen to be (a) even such that $g(\eta) = g(|\eta|)$ and (b) the derivative of g can be expressed as the following form

$$g'(\eta) = 2\eta h(\eta) \quad (24)$$

where h is called an *interaction function*, . Obviously, h thus defined is also even. With these assumptions, the Euler equation can be expressed as

$$\chi(x)[f(x) - d(x)] - \lambda \frac{d}{dx}[f'(x) h(f'(x))] = 0 \quad (25)$$

The magnitude $|g'(f')| = |2f'h(f')|$ relates to the *strength* with which a regularizer performs smoothing; and $h(f'(x))$ determines the *interaction* between neighboring pixels.

Let $\eta \stackrel{\triangle}{=} f'(x)$. A *necessary condition* for any regularization model to be adaptive to discontinuities is

$$\lim_{\eta\to\infty} |g'(\eta)| = \lim_{\eta\to\infty} |2\eta h(\eta)| = C \quad (26)$$

where $C \in [0, \infty)$ is a constant. The above condition with $C = 0$ entirely prohibits smoothing at discontinuities where $\eta \to \infty$ whereas with $C > 0$ allows limited (bounded) smoothing. In any case, however, the interaction $h(\eta)$ must be small for large $|\eta|$ and approaches 0 as $|\eta|$ goes to ∞. This is an important guideline for selecting g and h for the purpose of the adaptation.

Definition 1. An *adaptive interaction function* (AIF) h_γ parameterized by γ (> 0) is a function that satisfies:

$$\begin{array}{ll} \text{(i)} & h_\gamma \in C^1 \\ \text{(ii)} & h_\gamma(\eta) = h_\gamma(-\eta) \\ \text{(iii)} & h_\gamma(\eta) > 0 \\ \text{(iv)} & h'_\gamma(\eta) < 0 \quad (\forall \eta > 0) \\ \text{(v)} & \lim_{\eta\to\infty} |\eta h_\gamma(\eta)| = C < \infty \end{array} \quad (27)$$

The class of AIFs, denoted by $\mathbb{H}_\gamma$, is defined as the collection of all such h_γ. □

The continuity requirement (i) guarantees the twice differentiability of the integrand $u(f \mid d)$ in (20) with respect

[1] In this work, the continuity of $\chi(x)$ and $d(x)$ and differentiability of $f(x)$ are assumed for the variational problems defined on continuous domains [32], [33]. However, they are not necessary for discrete problems where $[a, b]$ is quantized into discrete points. For example, in the discrete case, $\chi(x_i)$ is allowed to take a value in {1,0}, indicating whether datum $d(x_i)$ is available or not.

TABLE I

Four choices of AIFs, the corresponding APFs and bands.

AIF	APF	Band
$h_{1\gamma}(\eta) = \mathrm{e}^{-\frac{\eta^2}{\gamma}}$	$g_{1\gamma}(\eta) = -\gamma \mathrm{e}^{-\frac{\eta^2}{\gamma}}$	$B_{1\gamma} = (-\sqrt{\frac{\gamma}{2}}, \sqrt{\frac{\gamma}{2}})$
$h_{2\gamma}(\eta) = \frac{1}{[1+\frac{\eta^2}{\gamma}]^2}$	$g_{2\gamma}(\eta) = -\frac{\gamma}{1+\frac{\eta^2}{\gamma}}$	$B_{2\gamma} = (-\sqrt{\frac{\gamma}{3}}, \sqrt{\frac{\gamma}{3}})$
$h_{3\gamma}(\eta) = \frac{1}{1+\frac{\eta^2}{\gamma}}$	$g_{3\gamma}(\eta) = \gamma \log(1 + \frac{\eta^2}{\gamma})$	$B_{3\gamma} = (-\sqrt{\gamma}, \sqrt{\gamma})$
$h_{4\gamma}(\eta) = \frac{1}{1+\frac{\lvert\eta\rvert}{\gamma}}$	$g_{4\gamma}(\eta) = \gamma\lvert\eta\rvert - \gamma^2 \ln(1 + \frac{\lvert\eta\rvert}{\gamma})$	$B_{4\gamma} = (-\infty, +\infty)$

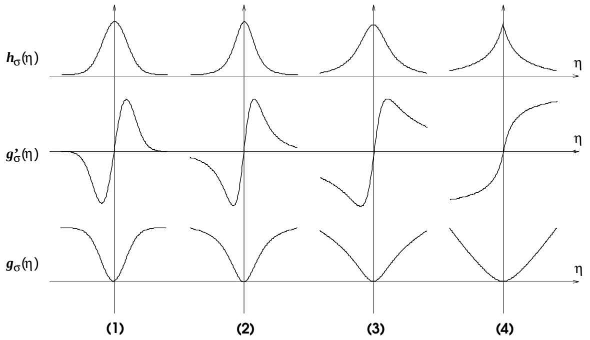

Fig. 1. The qualitative shapes of the four DA functions.

to f', a condition for the solution f to exist [32]. However, this can be relaxed to $h_\gamma \in C^0$ for discrete problems. The evenness of (ii) is usually assumed for spatially unbiased smoothing. The positive definiteness of (iii) keeps the interaction positive such that the sign of $\eta h_\gamma(\eta)$ will not be altered by $h_\gamma(\eta)$. The monotony of (iv) leads to decreasing interaction as the magnitude of the derivative increases. The bounded asymptote property of (v) provides the adaptive discontinuity control as stated earlier. Other properties follow: $h_\gamma(0) = \sup_\eta h_\gamma(\eta)$, $h'_\gamma(0) = 0$ and $\lim_{\eta\to\infty} h_\gamma(\eta) = 0$, the last meaning zero interaction at discontinuities. The above definition characterizes the properties AIFs should possess rather than instantiates some particular functions. Therefore, the following definition of the DA model is rather broad.

Definition 2. The DA solution f is defined by the Euler equation (25) constrained by $h = h_\gamma \in \mathbb{H}_\gamma$. □

The DA solution is C^1 continuous[2]. Therefore the DA solution and its derivative never have discontinuities in them. The DA overcomes oversmoothing by allowing its solution to be steep, but C^1 continuous, at point x where data $d(x)$ is steep. With every possible data configuration d, every $f \in C^1$ is possible.

There are two reasons for defining the DA in terms of the constrained Euler equation: First, it captures the essence of the DA problem; the DA model should be defined based on the h_γ therein. Second, some satisfying h_γ may not have their corresponding g_γ (and hence the energy) in closed-form, where g_γ is defined below.

Definition 3. The *adaptive potential function* (APF) corresponding to an $h_\gamma \in \mathbb{H}_\gamma$ is defined by

$$g_\gamma(\eta) = 2\int_0^\eta \eta' h_\gamma(\eta')d\eta' \tag{28}$$

□

The energy function (19) with $g = g_\gamma$ is called an *adaptive string*. The following are some properties of g_γ: Basically, g_γ is one order higher than h_γ in continuity; it is even, $g_\gamma(\eta) = g_\gamma(-\eta)$; its derivative function is odd, $g'_\gamma(\eta) = -g'_\gamma(-\eta)$; however, it is not necessary for $g_\gamma(\infty)$ to be bounded. Furthermore, g_γ is strictly monotonically increasing as $|\eta|$ increases because $g_\gamma(\eta) = g_\gamma(|\eta|)$ and $g'_\gamma(\eta) = 2\eta h_\gamma(\eta) > 0$ for $\eta > 0$. This means larger $|\eta|$ leads to larger penalty $g_\gamma(\eta)$. It conforms to the original spirit of the standard quadratic regularizers determined by g_q. The line process potential function g_α does not have such a property: Its penalty is fixed and does not increase as $|\eta|$ increases beyond $\sqrt{\alpha}$. The idea that large values of η are equally penalized is questionable [14].

In practice, it is not always necessary to know the explicit definition of g_γ. The most important factors are the Euler equation and the constraining function h_γ. Nonetheless, knowing g_γ is helpful for analyzing the convexity of $E(f)$.

For a given $g_\gamma(\eta)$, there exists a region of η within which the smoothing strength $|g'_\gamma(\eta)| = |\eta h_\gamma(\eta)|$ increases monotonically as $|\eta|$ increases and the function g_γ is convex:

$$B_\gamma = \{\eta \mid g''_\gamma(\eta) \geq 0\} \stackrel{\triangle}{=} (b_L, b_H) \tag{29}$$

The region B_γ is referred to as the *band*. The lower and upper bounds b_L, b_H correspond to the two extrema of $g'_\gamma(\eta)$, which can be obtained by solving $g''_\gamma(\eta) = 0$, and we have $b_l = -b_H$ when g is even. When $b_L < \eta < b_H$, $g''_\gamma(\eta) > 0$ and thus $g_\gamma(\eta)$ is strictly convex. For h_γ defined with $C > 0$, the bounds are $b_L \to -\infty$ and $b_H \to +\infty$.

Table I instantiates four possible choices of AIFs, the corresponding APFs and the bands. Fig.1 shows their qualitative shapes (a trivial constant may be added to $g_\gamma(\eta)$). Fig.2 gives a graphical comparison of the $g(\eta)$'s for the quadratic, the LP model and the first three DA models listed in Table I and their derivatives, $g'(\eta) = 2\eta h(\eta)$.

The fourth AIF, $h_{4\gamma} = 1/[1 + |\eta|/\gamma]$ allows bounded but non-zero smoothing at discontinuities: $\lim_{\eta\to\infty} \eta h_{4\gamma}(\eta) = \gamma$. It is interesting because $g''_{4\gamma}(\eta) = [2\eta h_{4\gamma}(\eta)]' > 0$ for all η (except at $\eta = 0$) and leads to strictly convex minimization. In fact, a positive number C in (27) leads to a convex AIF and hence energy function[3]. The convex subset of models for discontinuity adaptivity and M estimation is

[2]See [33] for a comprehensive discussion about the continuity of solutions of the class of problems to which the DA belongs.

[3]This is due to the following theorem: If $g(\cdot)$ is convex on $\mathbb{R}$, a real-valued energy function $E(f) = \int \chi(x)[f(x) - d(x)]^2 + \lambda g(f'(x))dx$ is convex w.r.t. f for all $f \in C^1$ and fixed $\chi, d \in C^1$.

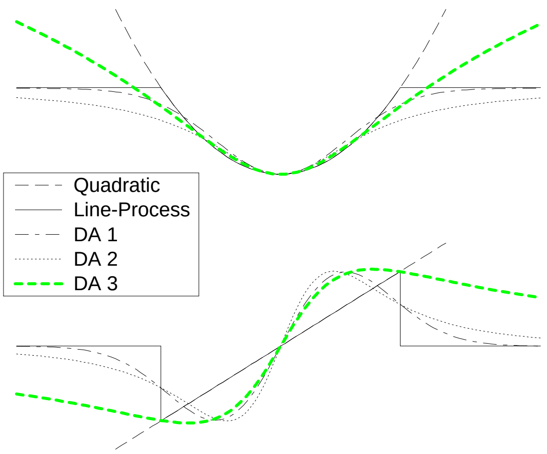

Fig. 2. Comparison of different potential functions $g(\eta)$ (top) and their first derivatives $g'(\eta) = 2\eta h(\eta)$ (bottom).

appealing because they have some inherent advantages over nonconvex models, both in stability and computational efficiency.

Fig.2 also helps us visualize how the DA performs smoothing. Like the quadratic $g_q(\eta) = \eta^2$, the DA allows the smoothing strength $|g'_\gamma(\eta)|$ to increase monotonically as η increases within the band B_γ. Outside the band, the smoothing decreases as η increases and becomes zero as $\eta \to \infty$. This differs from the quadratic regularizer which allows boundless smoothing when $\eta \to \infty$. Also unlike the LP model which shuts down smoothing abruptly just beyond its band $B_\alpha = (-\sqrt{\alpha}, \sqrt{\alpha})$, the DA decreases smoothing continuously towards zero.

B. Relations with Previous Models

The first three instantiated models behave in a similar way to the quadratic prior model when $\eta^2/\gamma \ll 1$ or $\eta^2 \ll \gamma$, as noticed by [15]. This can be understood by looking at the power series expansion of $g_\gamma(\eta)$, $g_\gamma(\eta) = \sum_{k=1}^{\infty} c_k(\eta^2/\gamma)^k$ where c_k are constants with $c_1 > 0$ (the expansion can also involve a trivial additive constant c_0). When $\eta^2/\gamma \ll 1$,

$$g_\gamma(\eta) = c_1(\eta^2/\gamma) + O(\eta^4/\gamma^2) \tag{30}$$

Thus in this situation of sufficiently small η^2/γ, the adaptive model inherits the convexity of the quadratic model.

The interaction function h also well explains the differences between various regularizers. For the quadratic regularizer, the interaction is constant everywhere

$$h_q(\eta) = 1 \tag{31}$$

and the smoothing strength is proportional to $|\eta|$. This is why the quadratic regularizer leads to oversmoothing at discontinuities where η is infinite. In the LP model, the interaction is piecewise constant

$$h_\alpha(\eta) = \begin{cases} 1 & |\eta| < \sqrt{\alpha} \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

Obviously, it inhibits oversmoothing by switching off smoothing when $|\eta|$ exceeds $\sqrt{\alpha}$ in a binary manner.

In the LP approximations using the Hopfield approach [19], [20] and mean field theory [17], the line process variables are approximated by (13). This approximation effectively results in the following interaction function

$$h_{\alpha,T}(\eta) = 1-\bar{l}(\eta) = 1-\frac{1}{1+\mathrm{e}^{-\lambda(\eta^2-\alpha)/T}} = \frac{\mathrm{e}^{-\lambda(\eta^2-\alpha)/T}}{1+\mathrm{e}^{-\lambda(\eta^2-\alpha)/T}} \tag{33}$$

As the temperature τ decreases toward zero, the above approaches $h_\alpha/2$, that is, $\lim_{\tau \to 0^+} h_{\alpha,\tau} = h_\alpha/2$. Obviously, $h_{\alpha,\tau}(\eta)$ with nonzero τ is a member of the AIF family, *i.e.* $h_{\alpha,\tau}(\eta) \in \mathbb{H}_\gamma$ and therefore the approximated LP models are instances of the DA model.

It is interesting to note an observation made by Geiger and Girosi: "sometimes a finite β solution may be more desirable or robust" ([17], pages 406-407) where $\beta = \frac{1}{T}$. They further suggest that there is "an optimal (finite) temperature (β)" for the solution. An algorithm is presented in [34] for estimating an optimal β. The LP approximation with finite $\beta < +\infty$ or nonzero $T > 0$ is more an instance of the DA than the LP model which they aimed to approximate. It will be shown in Section III-D that the DA model is indeed more stable than the LP model.

Anisotropic diffusion [35] is a scale-space method for edge-preserving smoothing. Unlike fixed coefficients in the traditional isotropic scale-space filtering [36], anisotropic diffusion coefficients are spatially varying according to the gradient information. A so-called biased anisotropic diffusion [16] model is obtained if anisotropic diffusion is combined with a closeness term. Two choices of APFs are used in those anisotropic diffusion models: $g_{1\gamma}$ and $g_{2\gamma}$.

Shulman and Herve [14] propose to use the following Huber's robust error penalty function [37] as the adaptive potential

$$g_\beta(\eta) = \min\{\eta^2, \beta^2 + 2\beta|\eta - \beta|\} \tag{34}$$

where $\beta > 0$ plays a similar role to α in g_α. The above is a convex function and has the first derivative as: $g'_\beta(\eta) = 2\eta h_\beta(\eta) = 2\eta$ for $|\eta| \le \beta$, and $g'_\beta(\eta) = 2\eta h_\beta(\eta) = 2\beta\eta/|\eta|$ for other η. Comparing $g'_\beta(\eta)$ with (24), we find that the corresponding AIF is $h_\beta(\eta) = 1$ for $|\eta| \le \beta$ and $h_\beta(\eta) = \beta/|\eta|$ for other η. This function allows bounded but nonzero smoothing at discontinuities. The same function has also been applied by [38] to curve fitting. A comparative study on the DA model and robust statistics can be found in [39], [27].

The approximation (18) of Leclerc's minimal length model [13] is in effect the same as the DA with APF 1. Eq.(17) may be one of the best cost functions for the piecewise constant restoration; for more general piecewise continuous restoration, one needs to use (18) with a *nonzero* τ, which is a DA instance. Regarding the continuity property of domains, Mumford and Shah's model [8] and Terzopoulos' continuity-controlled regularization model [9] can also be defined on continuous domains as the DA model.

C. Discrete Data and 2D Cases

When the data d is available at discrete points: $d_i = d(x_i)$, $1 \le i \le m$, the Euler equation for the adaptive string is

$$\sum_i \chi(x)[f(x) - d_i]\delta(x - x_i) = \lambda \frac{d}{dx}[f'(x)h(f'(x))] \quad (35)$$

where $\delta(\cdot)$ is the Dirac delta function. Integrating the above equation once yields

$$\sum_{i:x_i<x} \chi(x_i)[f(x_i) - d_i] = \lambda[f'(x)h(f'(x))] + c \quad (36)$$

where c is a constant. From the above, the solution f is determined by the following

$$f''(x) = 0 \qquad x_{i-1} < x < x_i,\ 1 < i \le m \quad (37)$$

$$f(x_i^+) - f(x_i^-) = 0 \qquad 1 < i < m \quad (38)$$

$$f'(x_i^+)h(f'(x_i^+)) - f'(x_i^-)h(f'(x_i^-)) = \tfrac{1}{\lambda}\chi(x_i)[f(x_i) - d_i] \qquad 1 < i < m \quad (39)$$

with appropriate boundary conditions at $x = x_1$ and $x = x_m$. Obviously, f in this case is piecewise C^1 continuous; more exactly, it is composed of consecutively joined line segments.

The adaptive string model can be extended to 2D and higher order equivalents. The 2D adaptive membrane has the following

$$\chi[f - d] - \lambda\frac{\partial}{\partial x}[f_x h(f_x)] - \lambda\frac{\partial}{\partial y}[f_y h(f_y)] = 0 \quad (40)$$

corresponding to the Euler equation (25), where data $d = d(x, y)$ and solution $f = f(x, y)$ represent images. Extending the DA to second order, one obtains the following for the adaptive rod on 1D

$$\chi[f - d] + \lambda\frac{d}{dx^2}[f''h(f'')] = 0 \quad (41)$$

and for the adaptive plate on 2D

$$\chi[f - d] + \lambda\frac{\partial}{\partial x^2}[f_{xx}h(f_{xx})] + 2\lambda\frac{\partial}{\partial xy}[f_{xy}h(f_{xy})] + \lambda\frac{\partial}{\partial y^2}[f_{yy}h(f_{yy})] = 0 \quad (42)$$

D. Solution Stability

The DA solution depends continuously on its parameters and the data whereas the LP solution does not. An informal analysis follows. Consider an LP solution f_α obtained with α. The solution is a local equilibrium satisfying

$$\chi(f_\alpha - d) - \lambda f'_\alpha h_\alpha(f'_\alpha) = 0 \quad (43)$$

In the above, $h_\alpha(f'_\alpha) = 1$ or 0, depending on α and the configuration of f_α. When $|f'_\alpha(x)|^2$ is close to α for some x, a small change $\Delta\alpha$ may flip h_α over from one state to the other. This is due to the binary non-linearity of h_α. The flip-over leads to a significantly different solution. This can be expressed as

$$\lim_{\Delta\alpha\to 0}[f_{\alpha+\Delta\alpha} - f_\alpha] \ne 0(x) \qquad \exists\alpha \quad (44)$$

where $0(x)$ is a function which is constantly zero in the domain $[a, b]$. The variation δf_α with respect to $\Delta\alpha$ may not be zero for some f_α and α, which causes instability. However, the DA solution, denoted f_γ, is stable

$$\lim_{\Delta\gamma\to 0}[f_{\gamma+\Delta\gamma} - f_\gamma] = 0(x) \qquad \forall\gamma \quad (45)$$

where f_γ denotes the DA solution. Conclusion on the stability due to changes in parameter λ can be drawn similarly.

The same is also true with respect to the data. Given α and λ fixed, the solution depends on the data, *i.e.* $f = f[d]$. Assume a small variation δd in the data d. The solution $f[d]$ must change accordingly to reach a new equilibrium $f[d + \delta d]$ to satisfy the Euler equation. However, there always exist possibilities that $h_\alpha(f'_\alpha(x))$ may flip over for some x when $|f'_\alpha(x)|^2$ is near the α, resulting in an abrupt change in the LP solution f_α. This can be represented by

$$\lim_{\delta d\to 0(x)}[f_\alpha[d + \delta d] - f_\alpha[d]] \ne 0(x) \qquad \exists f_\alpha,\ d \quad (46)$$

That is, the variation δf_α with respect to δd may not be zero for some f_α and d. However, the DA model is stable to such changes, *i.e.*

$$\lim_{\delta d\to 0(x)}[f_\gamma[d + \delta d] - f_\gamma[d]] = 0(x) \qquad \forall f_\gamma,\ d \quad (47)$$

because of its continuous nature. From the analysis, it can be concluded that the DA better regularizes ill-posed problems than the LP.

IV. Computation of DA Solutions

A. Solving the Euler Equation

The Euler equation (25) can be treated as a boundary value problem. It can also be solved by minimizing the corresponding energy (19) because a minimum of the energy is (sufficiently) a solution of the equation (With $g'(\eta) = 2\eta h(\eta)$, the form of the energy needs not to be known in order to minimize it – see below). The energy minimization approach is chosen here.

Because both $\chi(x)[f(x) - d(x)]^2$ and $g(f'(x))$ are bounded below, so is the energy. This means that a minimal solution, and hence a solution to the Euler equation, exists. If h_γ is chosen with $C = 0$, the corresponding energy $E(f)$ is non-convex with respect to f, and the minimization is subject local minima. See [10] for an analysis of convexity in string and membrane models. The following presents a discrete method for finding a local minimum.

Sample the integral interval $[a, b]$ into m uniformly spaced points: $x_1 = a$, ..., $x_m = b$. For clarity and without loss of generality, let us assume that the bounds a and b are re-scaled in such a way that the point spacing is one unit[4]

[4] More advanced numerical methods using varying spacing, such as [40] may be advantageous in obtaining more accurate solutions.

i.e. $x_i - x_{i-1} = 1$. Approximate the first derivative f'_i by the first order backward difference $f'_i = \frac{f_i - f_{i-1}}{x_i - x_{i-1}} = f_i - f_{i-1}$. Then the energy (19) is approximated by

$$E(f) = \sum_{i=1}^{m} \chi_i [f_i - d_i]^2 + \lambda \sum_{i=1}^{m} \sum_{i' \in \mathcal{N}_i} g(f_i - f_{i'}) \qquad (48)$$

where $\mathcal{N}_i$ consists of the neighbors of i. Using the gradient-descent method, we can obtain the following updating equation: for $1 < i < m$,

$$\begin{aligned} f_i^{(t+1)} \leftarrow f_i^{(t)} &- 2\mu \{ \chi_i [f_i^{(t)} - d_i] \\ &- \lambda \sum_{i' \in \mathcal{N}_i} (f_{i'}^{(t)} - f_i^{(t)})\, h(f_{i'}^{(t)} - f_i^{(t)}) \} \end{aligned} \qquad (49)$$

The solution $f^{(t)}$ takes prescribed values at the boundary points at $i = 1$ and $i = m$ to meet the boundary condition (22) and the values may be estimated from data d_i near the boundaries. With initial $f^{(0)}$, the solution is in the limit $f^* = f^{(\infty)}$.

Equation (49) helps us see more of how the DA works. The smoothing at i is due to $\lambda \sum_{i' \in N_i} (f_{i'} - f_i) h(f_{i'} - f_i)$ where $N_i = \{i-1, i+1\}$. The contributions to smoothing are from the two neighboring points. That from site i' is proportional to the product of the two factors, $f_{i'} - f_i$ and $h_\gamma(f_{i'} - f_i)$. This is why we relate $|g'(\eta)| = |2\eta h(\eta)|$ to the strength of smoothing performed by regularizers. On the other hand, $h(\eta)$ acts as an adaptive weighting function to control the smoothing due to the difference $\eta = f_{i'} - f_i$. Therefore we regard it as the interaction. Two more remarks are made: First, the contributions from the two sides, $i-1$ and $i+1$ are treated separately or non-symmetrically. Second, the sum of the contributions to smoothing is zero if the three points are aligned when $f_{i+1} - f_i = f_i - f_{i-1}$ and $h(\eta)$ is even since in this situation $\sum_{i' \in N_i} (f_{i'} - f_i) h(f_{i'} - f_i) = 0$.

The updating rule for the adaptive membrane on 2D can be easily derived. In the 2D case (40) , there is an additional term in the y dimension. This leads to the following energy

$$E(f) = \sum_{i,j} \chi_{i,j} [f_{i,j} - d_{i,j}]^2 + \lambda g(f_{i,j} - f_{i-1,j}) + \lambda g(f_{i,j} - f_{i,j-1}) \qquad (50)$$

Letting $\frac{\partial f_{i,j}}{\partial t} = -\mu \frac{\partial E}{\partial f_{i,j}}$ leads to the following updating rule

$$\begin{aligned} f_{i,j}^{(t+1)} \leftarrow f_{i,j}^{(t)} &- 2\mu \{ \chi_{i,j} [f_{i,j}^{(t)} - d_{i,j}] - \\ &\lambda \sum_{(x,y) \in N_{i,j}} (f_{x,y}^{(t)} - f_{i,j}^{(t)}) h(f_{x,y}^{(t)} - f_{i,j}^{(t)}) \} \end{aligned} \qquad (51)$$

where $N_{i,j} = \{(i-1,j), (i+1,j), (i,j-1), (i,j+1)\}$ is the set of the four neighboring points[5] of (i,j). The updating on 2D grid can be performed on the white and black sites of a checkerboard alternatively to accelerate convergence.

[5]With the 4-neighborhood system, the model considers derivatives in the horizontal and vertical directions. With the 8-neighborhood system, the regularizer also includes the diagonal derivatives weighted by $1/\sqrt{2}$.

Choose a convex $\gamma^{(0)}$; Set $f^{(0)} \leftarrow d$ and $t = 0$;
Do 1) Update $f^{(t)}$ using Eq.(49);
2) Set $t \leftarrow t + 1$;
3) If $(f^{(t)} - f^{(t-1)} < \epsilon)$
set $\gamma^{(t)} \leftarrow \max\{\gamma_{target}, \kappa\gamma^{(t-1)}\}$;
Until $(f^{(t)} - f^{(t-1)} < \epsilon)$ and $(\gamma^{(t)} = \gamma_{target})$;
Set$f^* \leftarrow f^{(t)}$

Fig. 3. A GNC algorithm for finding the DA solution.

There are three parameters in (49) which shall be determined for the DA model: λ, γ and μ. Parameter μ is related to the convergence of the relaxation algorithm. There is an upper bound for μ for the system to be stable. There also exists an optimal value for μ [10]. Optimal choices of λ for quadratic regularization may be made using cross-validation [41], [42]. In [43], a least squares method [44] is presented for estimating MRF clique potentials for a LP model. Automated selection of the λ and γ parameters for the DA model is an unsolved problem. They are currently chosen in an *ad hoc* way.

B. Non-Convex Minimization

The DA model with $C = 0$ leads to non-convex dynamic systems and direct minimization using gradient descent only guarantees to find a local minimum. A GNC-like algorithm can be constructed for approximating the global solution.

Because a convex g guarantees a convex energy function $E(f)$, it is useful to study the convexity of E by analyzing the convexity of g_γ. The expansion (30) illustrates that when $\max_i \eta_i^2/\gamma = \max_i (f_i - f_{i-1})^2/\gamma$ is small enough, the DA model behaves in a similar way to the quadratic regularizer and hence is convex. An analysis in [10] shows that it is sufficient to guarantee the convexity if γ is chosen large enough to satisfy

$$g''_\gamma(f_i - f_{i-1}) \geq c^* \qquad \forall i \qquad (52)$$

where c^* is some real number (see an analysis in [10] for determining c^*). The exact value of c^* for the convexity depends on g and χ. It can be safely assumed that $c^* = 0$ in any case.

Therefore, a convex $\gamma^{(0)}$ can be chosen for which $g''_{\gamma^{(0)}}(f_i^{(0)} - f_{i-1}^{(0)}) \leq 0$ for all i. This is equivalent to choosing a $\gamma^{(0)}$ such that $f_i^{(0)} - f_{i-1}^{(0)} \in B_{\gamma^{(0)}}$ where B_γ is the band introduced in the section. For APFs 1, 2 and 3, the $\gamma^{(0)}$ must be larger than $2v$, $3v$ and v, respectively, where $v = \max_i [f_i^{(0)} - f_{i-1}^{(0)})]^2$.

The graduation from an initially convex approximation of g_γ to its target form can be implemented by continuation in γ from a big value to the target value. A GNC-like algorithm using this heuristic is outlined in Fig.3. Given d, λ and a target value γ_{target} for γ, the algorithm aims to construct a sequence $\{\gamma^{(t)}\}$ with $\gamma^{(\infty)} \to \gamma_{target}$, and thus

$\{f_{\gamma^{(t)}}^{(t)}\}$ to approach the global minimum $f^* = \lim_{t\to\infty} f_{\gamma^{(t)}}^{(t)}$ for which $E(f^*) = \min$. In the algorithm, ϵ is a constant for judging the convergence, and κ is a factor for decreasing $\gamma^{(t)}$ towards γ_{target}, $0 < \kappa < 1$. The choice of κ controls the balance between the quality of the solution and the computational time. In principle, $\gamma^{(t)}$ should vary continuously to keep a good track of the global minimum. In discrete computation, we choose $0.9 \le \kappa \le 0.99$. More rapid decrease of $\gamma^{(t)}$ (with smaller κ) is likely to lead the system to an unfavorable local minimum. Witkin *et al.* [45] present a more sophisticated scheme for decreasing γ by relating the step to the energy change: $\gamma^{(t+1)} = \gamma^{(t)} - c_1 e^{-c_2|\nabla E|}$ where c_1 and c_2 are constants. This seems reasonable.

C. Analog Network

The computation can be performed using an analog network. Let f_i be the potential of neural cell i. Let $C_i = C = 1/2\mu$ be the membrane capacitance and $R_i = 1/\chi_i$ the membrane resistance. Let

$$T_{i,i'} = T_{i',i} = \lambda h_\gamma(f_i - f_{i'}) \tag{53}$$

be the conductance or synaptic efficacy between neurons i and i' where $|i - i'| = 1$ and $h_\gamma \in \mathbb{H}_\gamma$. If the exponential $g_{1\gamma}$ is used, then

$$T_{i,i'} = T_{i',i} = \lambda \exp\{-[f_i - f_{i'}]^2/\gamma\} \tag{54}$$

Let d_i be the external current input to i, with $d_i = 0$ when $\chi_i = 0$. Now (49) can be written as

$$C\frac{\partial f_i}{\partial t} = -\frac{1}{R_i} f_i + d_i + T_{i-1,i}[f_{i-1} - f_i] + T_{i+1,i}[f_{i+1} - f_i] \tag{55}$$

The above is the dynamic equation at neuron i of the network. The diagram of the network circuit is shown in Fig.4. The synaptic current from i to i' is

$$I_{i,i'} = \lambda g'_\gamma(f_i - f_{i'}) = \lambda[f_i - f_{i'}] h_\gamma(f_i - f_{i'}) \tag{56}$$

If the exponential $g_{1\gamma}$ is used, then

$$I_{i,i'} = \lambda[f_i - f_{i'}] \exp\{-[f_i - f_{i'}]^2/\gamma\} \tag{57}$$

A plot of current $I_{i,i'}$ versus potential difference $f_i - f_{i'}$ was shown at the bottom of Fig.2. The voltage-controlled nonlinear synaptic conductance $T_{i,i'}$, characterized by the h function defined in (27), realizes the adaptive continuity control; the corresponding nonlinear current $I_{i,i'}$ realizes the adaptive smoothing. The current $I_{i,i'}$ diminishes asymptotically to zero as the potential difference between neurons i and i' reaches far beyond the band B_γ.

Fig.7 shows the behavior of the analog network under component defects such as manufacturing inadequacy, quality changes, *etc.* The defects are simulated by adding $\pm 25\%$ evenly distributed random noise into R, C, and T in (55). The data d is shown in triangles with 50% missing rate; the locations of the missing data, for which $\chi_i = 0$, are indicated by triangles at the bottom. The noise in the data is white Gaussian with standard deviation $\sigma = 10$ (left)

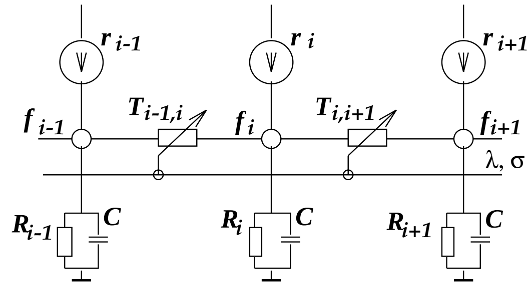

Fig. 4. Schematic diagram of the analog network circuit for the DA model.

and $\sigma = 20$ (right). The interaction function is chosen to be $h_{2\gamma}(\eta) = \frac{1}{[1+\eta^2/\gamma]^2}$. Solutions obtained with simulated component defects are shown in dashed lines in comparison with those obtained without such noise shown in thicker solid lines. The ideal signal is shown in thinner solid lines. As can be seen, there is only a little difference between the solutions obtained with and without such noise. This demonstrates not only the stability of the network circuit but also the error-tolerance property of the DA model.

V. Experiments

Two experimental results are presented in the following[6]. The first is the reconstruction of a real image of size 256×256 (Fig.5). Here, APF 1 ($g_{1\gamma}$) is used and the parameters are empirically chosen as $\lambda = 50$, $\gamma = 2$. The result shows that the reconstructed image is much cleaner with discontinuities well preserved.

The second experiment is the detection of step and roof edges from a simulated noisy pyramid image of size 128×128 (Fig.6). The detection process runs in three stages: 1) regularizing the input image and computing images of first derivatives in the two directions from the regularized image using finite difference, 2) regularizing the derivative images and 3) detecting steps and roofs by thresholding the regularized derivative images. APF 2 ($g_{2\gamma}$) is used and the parameters are empirically chosen as $\lambda = 10$, $\gamma_{target} = 10$ for the first stage of the regularization and $\lambda = 100$, $\gamma_{target} = 0.002$ for the second stage.

Edges in the horizontal and vertical directions are best detected while those in the diagonal directions are not so well done. This is because only derivatives in the two axes directions are considered in the DA discussed so far; changes in the diagonal directions are largely ignored. Regularizers using the 8-neighborhood system (see the footnote for Eq.(51) should help improve the detection of diagonal changes.

Results in Fig.7 show the behavior of the analog DA network under component defects such as manufacturing inadequacy, quality changes, *etc.* The defects are simulated by adding $\pm 25\%$ evenly distributed random noise into R, C, and T in (55). The data d is shown in triangles with 50% missing rate; the locations of the missing data, for which $\chi_i = 0$, are indicated by triangles at the bottom. The noise

[6]More results can be found in Chapter 3 of [18].

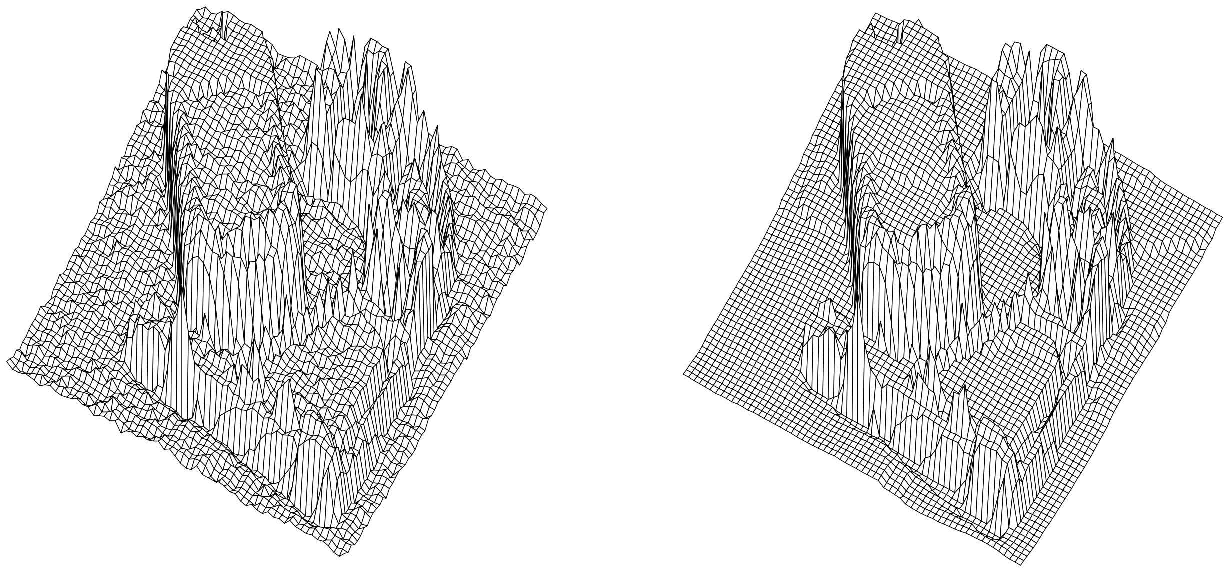

Fig. 5. 3D plots of an image (left) and its reconstruction using DA (right). The plots are drawn after sampling the images at every 4 pixels in both directions into the size of 64×64.

in the data is white Gaussian with standard deviation $\sigma = 10$ (left) and $\sigma = 20$. Solutions obtained with simulated component defects are shown in dashed lines in comparison with those obtained without such noise shown in thicker solid lines. The ideal signal is shown in thinner solid lines. As can be seen, there is only a little difference between the solutions obtained with and without such noise. This demonstrates not only the stability of the network circuit but also the error-tolerance property of the DA model.

VI. Conclusion

Through an analysis of the associated Euler equation, a necessary condition is made explicit for MRF or regularization models to be adaptive to discontinuities. On this basis, the DA model is defined by the Euler equation constrained by the class of adaptive interaction functions (AIFs). The definition provides principles for choosing deterministic regularizers and MRF clique potential functions. It also includes many existing models as special instances.

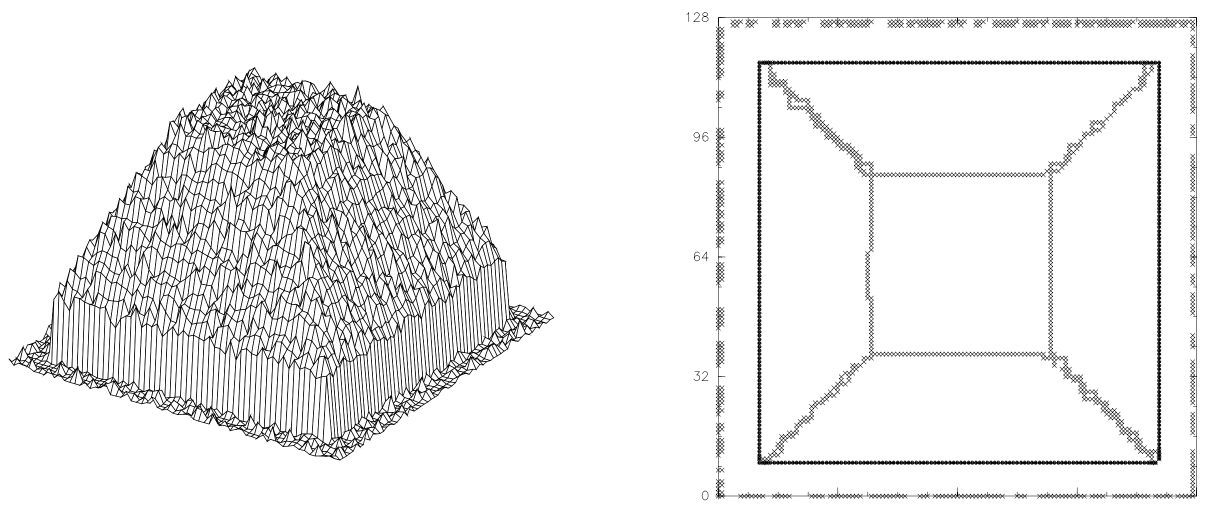

Fig. 6. Step and roof edges (right) detected from a pyramid image (left). Step edges are shown in dots and roof edges in crosses.

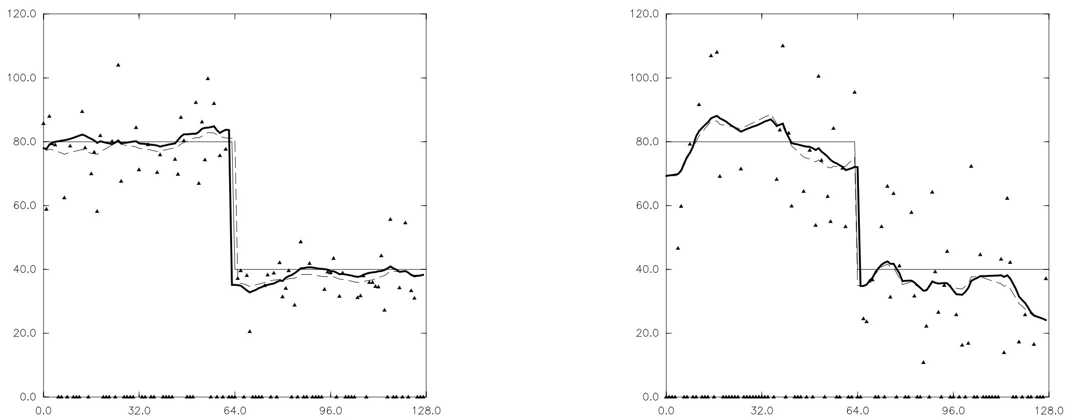

Fig. 7. Stability of the DA solution under disturbances in parameters.

The DA model has its solution in C^1 and adaptively overrides the smoothness assumption where the assumption is not valid, without the switching-on/off of discontinuities in the LP model. The DA solution never contains true discontinuities. The DA model "preserves" discontinuities by allowing the solution to have arbitrarily large but *bounded* slopes. The LP model "preserves true discontinuities" by switching between small and *unbounded* (or large) slopes.

Owing to its continuous properties, the DA model possesses some theoretical advantages over the LP model. Unlike the LP model, it is stable to changes in parameters and in the data. Therefore it is better than the LP model in solving ill-posed problems. In addition, it is able to deal with problems on a continuous domain. Furthermore, it is better suited for analog VLSI implementation.

Acknowledgments

The author is grateful to Yihong Huang, Eric Sung, Wei Yun Yau and Han Wang for their helpful comments.

References

[1] S. Geman and D. Geman, ""Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images"", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, November 1984.

[2] H. Elliott, H. Derin, R. Cristi, and D. Geman, ""Application of the Giibs distribution to image segmentation"", in *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, San Diego, March 1984, pp. 32.5.1–32.5.4.

[3] J. L. Marroquin, *Probabilistic Solution of Inverse Problems*, PhD thesis, MIT AI Lab, 1985.

[4] T. Poggio, V. Torre, and C. Koch, ""Computational vision and regularization theory"", *Nature*, vol. 317, pp. 314–319, 1985.

[5] A. Blake, ""The least disturbance principle and weak constraints"", *Pattern Recognition Letters*, vol. 1, pp. 393–399, 1983.

[6] D. T. Terzopoulos, ""The role of constraints and discontinuities in visible surface reconstruction"", in *Proc. 8th International Joint Conference on AI*, Karlsruhe, W. Germany, 1983, pp. 1073–1077.

[7] W. E. L. Grimson and T. Pavlidis, ""Discontinuity detection for visual surface reconstruction"", *Computer Vision, Graphics and Image Processing*, vol. 30, pp. 316–330, 1985.

[8] D. Mumford and J. Shah, ""Boundary detection by minimizing functionals: I"", in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, June 1985, pp. 22–26.

[9] D. T. Terzopoulos, ""Regularization of inverse visual problems involving discontinuities"", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 4, pp. 413–424, July 1986.

[10] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, MA, 1987.

[11] D. Lee and T. Pavlidis, ""One dimensional regularization with discontinuities"", in *Proc. 1st International Conference on Computer Vision*, London, England, 1987, pp. 572–577.

[12] C. Koch, ""Computing motion in the presence of discontinuities: algorithm and analog networks"", in *Neural Computers*, R. Eckmiller and Ch. c. d. Malsburg, Eds., vol. F41 of *NATO ASI Series*, pp. 101–110. Springer-Verlag, 1988.

[13] Y. G. Leclerc, ""Constructing simple stable descriptions for image partitioning"", *International Journal of Computer Vision*, vol. 3, pp. 73–102, 1989.

[14] D. Shulman and J.Y. Herve', ""Regularization of discontinuous flow fields"", in *Proc. Workshop on Visual Motion*, 1989, pp. 81–86.

[15] S. Z. Li, ""Reconstruction without discontinuities"", in *Proceedings of Third International Conference on Computer Vision*, Osaka, Japan, December 1990, pp. 709–712.

[16] N. Nordstrom, ""Biased anisotropic diffusion – a unified regularization and diffusion approach to edge detection"", in *Proceedings of the European Conference on Computer Vision*, 1990, pp. 18–27.

[17] D. Geiger and F. Girosi, ""Parallel and deterministic algorithms from MRF's: surface reconstruction"", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 5, May 1991.

[18] S. Z. Li, *Towards 3D Vision from Range Images: An Optimisation Framework and Parallel Distributed Networks*, PhD thesis, University of Surrey, Guilford, Surrey, UK, July 1991.

[19] C. Koch, J. Marroquin, and A. Yuille, ""Analog 'neuronal' networks in early vision"", *Proceedings of National Academic Science, USA*, vol. 83, pp. 4263–4267, 1986.

[20] A. L. Yuille, ""Energy function for early vision and analog networks"", *A. I. Lab. Memo* No. 987, MIT, Cambridge, MA, 1987.

[21] T. E. Boult, ""What is regular in regularization?"", in *Proceedings of First International Conference on Computer Vision*, 1987, pp. 457–462.

[22] M. Bertero, T. A. Poggio, and V. Torre, ""Ill-posed problems in early vision"", *Proceedings of the IEEE*, vol. 76, no. 8, August 1988.

[23] A. N. Tikhonov and V. A. Arsenin, *Solutions of Ill-posed Problems*, Winston & Sons, Washington, 1977.

[24] J. Marroquin, S. Mitter, and T. Poggio, ""Probabilistic solution of ill-posed problems in computational vision"", *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 76–89, March 1987.

[25] G. E. Hinton, *Relaxation and Its Role in Vision*, PhD thesis, University of Edinburgh, Edinburgh, 1978.

[26] S. Z. Li, ““A Markov random field model for object matching under contextual constraints””, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seattle, Washington, June 1994.

[27] S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, 1995.

[28] J. J. Hopfield, ““Neurons with graded response have collective computational properties like those of two state neurons””, *Proceedings of National Academic Science, USA*, vol. 81, pp. 3088–3092, 1984.

[29] P. J. Besl, J. B. Birch, and L. T. Watson, ““Robust window operators””, in *Proceedings of Second International Conference on Computer Vision*, Florida, December 1988, pp. 591–600.

[30] S. C. Liu and J. G. Harris, ““Generalized smoothing networks in early vision””, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1989, pp. 184–191.

[31] J. G. Harris, ““A new approach to surface reconstruction: The coupled depth/slope model””, in *Proceedings of First International Conference on Computer Vision*, London, England, June 1987, pp. 277–283.

[32] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, vol. 1, Interscience Publishers Inc., New York, 1953.

[33] Lamberto Cesari, *Optimization – Theory and Applications : Problems with Ordinary Differential Equations*, Springer-Verlag, New York, 1983.

[34] T. Hebert and R. Leahy, ““Statistic-based MAP image restoration from Poisson data using Gibbs priors””, *IEEE Transactions on Signal Processing*, vol. 40, no. 9, September 1992.

[35] P. Perona and J. Malik, ““Scale-space and edge detection using anisotropic diffusion””, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, July 1990.

[36] A. P. Witkin, ““Scale-space filtering””, in *Proceedings of International Joint Conference on Artificial Intelligence*, 1983, vol. 2, pp. 1019–1022.

[37] P. Huber, *Robust Statistics*, Wiley, 1981.

[38] R. Stevenson and E. Delp, ““Fitting curves with discontinuities””, in *Proceedings of International Workshop on Robust Computer Vision*, Seattle, WA, October 1-3 1990, pp. 127–136.

[39] S. Z. Li, ““Discontinuity-adaptive MRF prior and robust statistics: A comparative study””, *Image and Vision Computing*, vol. 13, no. 4, April 1995.

[40] I. Weiss, ““Shape reconstruction on a varying mesh””, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 345–362, 1990.

[41] G. Wahba, ““Spline bases, regularization, and generalized cross-validation for solving approximation problems with large quantities of noisy data””, in *Approximation Theory III*, E. Cheney, Ed. 1980, vol. 2, Academic Press.

[42] B. Shahraray and D.J. Anderson, ““Optimal estimation of contour properties by cross-validated regularization””, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 600–610, 1989.

[43] S. G. Nadabar and A. K. Jain, ““Parameter estimation in MRF line process models””, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1992, pp. 528–533.

[44] H. Derin and H. Elliott, ““Modeling and segmentation of noisy and textured images using Gibbs random fields””, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 1, pp. 39–55, January 1987.

[45] A. Witkin, D. T. Terzopoulos, and M. Kass, ““Signal matching through scale space””, *International Journal of Computer Vision*, pp. 133–144, 1987.

S. Z. Li received the B.Sc degree from Hunan University, China, in 1982, M.Sc degree from the National University of Defense Technology, China, in 1985 and Ph.D degree from the University of Surrey, UK, in 1991. All degrees are in EEE. He is currently a lecturer at Nanyang Technological University, Singapore. His research interests include computer vision, pattern recognition, image processing and optimization methods.