Correspondence

Recursive Regularization Filters: Design, Properties, and Applications

Michael Unser, Akram Aldroubi, and Murray Eden

***Abstract*—Least squares approximation problems that are regularized with specified highpass stabilizing kernels are considered. For each problem, there is a family of discrete regularization filters (R-filters) allowing an efficient determination of the solutions. These operators are stable symmetric lowpass filters with an adjustable scale factor. Two decomposition theorems for the z-transform of such systems are presented. One facilitates the determination of their impulse response, while the other allows an efficient implementation through successive causal and anticausal recursive filtering. A case of special interest is the design of R-filters for the first and second order difference operators. These results are extended for two-dimensional signals and, for illustration purposes, are applied to the problem of edge detection. This leads to a very efficient implementation (8 multiplies + 10 adds per pixel) of the optimal Canny edge detector based on the use of a separable second order R-filter.**

***Index Terms*—Approximation methods, edge detection, Gaussian filtering, recursive filters, regularization, smoothing.**

I. Introduction

Regularization theory provides a convenient way to solve ill-posed problems and to compute solutions that satisfy prescribed smoothness constraints [1]. It has been recognized for some years that this formalism provides a unified framework for studying several problems in early vision including edge detection, visual interpolation, structure from stereo, shape from shading, and the computation of optical flow [2], [3]. Most of these tasks can be formulated as global optimization problems and are solved by numerical iterative schemes. The cost function to be minimized is a combination of a quadratic error term and a stabilizing functional. This latter term usually reflects physical constraints arising within the application for which the proposed solution is a model, and acts by limiting the energy of the solutions' higher order derivatives.

For problems involving linear translation-invariant operators, it is well known that a regularized solution can be obtained by linear filtering. This is usually equivalent to the application of a smoothing operator. Such techniques have been applied for signal restoration (deconvolution) [4], signal approximation (noise reduction) [5], and signal differentiation (edge detection) [6]–[8]. In practice, the filters are either implemented in Fourier space or approximated by finite impulse response kernels, which in both cases can require substantial numbers of computations. For example, a smoothing kernel extensively used in computer vision is the isotropic Gaussian filter [7], [9], [10].

Recently, we have found that the approach by filtering can be an attractive alternative to the conventional iteration of the Gauss–Seidel method for problems requiring the estimation of multiplicative signal parameters (manuscript in preparation); for example, the estimation of optical flow in the Horn and Schunck method [11]. The main advantage of the filter approach is a substantially improved convergence rate that is independent on the magnitude of the regularization parameter; that is, the degree of smoothness imposed on the solution.

For all these applications, the issue of computational cost is extremely important and there is a strong motivation to develop fast filtering techniques. To address this issue, we present an analysis of discrete regularization in the simple case of signal approximation. We adopt a purely discrete point of view—as opposed to the more conventional continuous interpretation [2], [3]–so as to allow us to develop efficient techniques for designing, implementing, and studying the properties of such filters.

The presentation is organized as follows. In Section II, we show that simple finite impulse response (FIR) stabilizing kernels can be used to define families of parametrized regularization filters. We investigate the properties of such R-filters, emphasizing their recursive structure, which leads to fast computational algorithms. In Section III, we focus on the design of R-filters associated with the first and second order difference operators ($[-1\ 1]$ and $[-1\ 2\ -1]$, respectively). Finally, in Section IV, we extend those results to signals of higher dimensionality and identify separability conditions. We consider edge detection as an example and propose an efficient realization of two well known methods: the approximating spline [7], [8] and the Canny [12], [13] edge detection techniques. Other potential applications include the estimation of local image statistics, and fixed or adaptive image smoothing for noise reduction.

Manuscript received May 19, 1989; revised August 20, 1990. Recommended for acceptance by R. J. Woodham.

M. Unser is with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda, MD 20892, and INSERM, Unité 2, Hôpital Henri-Mondor, F-94010 Créteil, France.

A. Aldroubi and M. Eden are with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda, MD 20892.

IEEE Log Number 9040952.

II. Discrete Regularization Filters

A. Regularized Approximation of a Signal—Definition

The discrete regularized approximation of a signal $\{x(k), k = -\infty, \cdots, +\infty\}$ with respect to a stabilizing kernel $\{h(k)\}$ and a regularization parameter $\lambda \geq 0$, is defined as the signal $\{y_\lambda(k)\}$ that minimizes

$$\eta_\lambda(y) = \sum_{k=-\infty}^{+\infty} [x(k) - y(k)]^2 + \lambda \sum_{k=-\infty}^{+\infty} [y(k) * h(k)]^2. \quad (1)$$

The criterion $\eta_\lambda(y)$ combines the quadratic approximation error and a constraint functional that measures the energy of a filtered version of $y(k)$. Usually, the stabilizing kernel is chosen to be some differential operator, which is equivalent to imposing smoothness constraints on the solution.

In this study, we will consider FIR stabilizing kernels of length $(n+1)$ described in terms of their z-transform

$$h(k) \leftrightarrow H(z) = \sum_{k=-\infty}^{+\infty} h(k) z^{-k} \quad (2)$$

and their autocorrelation function

$$\varphi_h(k) = h(k) * h(-k) = \sum_{l=-\infty}^{+\infty} h(l)h(l+k) \leftrightarrow H(z)H(z^{-1}). \quad (3)$$

The only constraint that we place on $\{h(k)\}$ is that

$$\sum_{k=-\infty}^{+\infty} h(k) = H(z)|_{z=1} = 0, \quad (4)$$

the weakest condition to be satisfied by a highpass filter. A typical example of such a stabilizing kernel is the nth order difference operator (with $H_n(z) = z^{-\lceil n/2 \rceil}(1-z)^n$), which is the discrete equivalent of the nth order derivative of a continuous signal.

B. Regularization Filter

The optimal regularized approximation of the $\{x(k)\}$ is found by differentiating (1) with respect to $y(k)$ and setting the derivative to zero:

$$\frac{\partial \eta_\lambda(y)}{\partial y(k)} = -2[x(k) - y(k)] + 2\lambda \sum_{1=-n}^{+n} y(k-1)\varphi_h(l) = 0, \quad (k = -\infty, \cdots, +\infty). \tag{5}$$

This expression is a standard constant-coefficient difference equation. It follows that the regularized solution $y(k)$ may be derived from $x(k)$ by linear filtering

$$y(k) = w_\lambda(k) * x(k) \tag{6}$$

where $\{w_\lambda(k)\}$ is the so-called regularization filter. In the z-transform domain, (5) is equivalent to

$$X(z) = Y(z) + \lambda \sum_{1=-n}^{+n} Y(z) z^{-1} \varphi_h(l) \tag{7}$$

which implies that

$$W_\lambda(z) = \frac{Y(z)}{X(z)} = \frac{1}{1 + \lambda \sum\limits_{l=-n}^{+n} z^{-1}\varphi_h(l)} = \frac{1}{1 + \lambda H(z) H(z^{-1})} \leftrightarrow w_\lambda(k). \tag{8}$$

Based on this result, we may derive a certain number of elementary properties.

1) $w_\lambda(k)$ is symmetric: $w_\lambda(k) = w_\lambda(-k)$. This is a simple consequence of the fact that $\varphi_h(k)$ is symmetric which also implies that $W_\lambda(z) = W_\lambda(z^{-1})$.

2) $w_\lambda(k)$ has a sum normalized to unity. By evaluating (8) at $z = 1$ and using (4), we find that

$$\sum_{k=-\infty}^{+\infty} w_\lambda(k) = W_\lambda(z)|_{z=1} = 1. \tag{9}$$

3) $w_\lambda(k)$ is a stable filter. The frequency response of $w_\lambda(k)$ is obtained by evaluating $W_\lambda(z)$ at $z = e^{j2\pi f}$ and is also given by

$$W_\lambda(f) = \frac{1}{1 + \lambda |H(f)|^2}. \tag{10}$$

The denominator of $W_\lambda(f)$ is always greater than one which implies that $0 \le |W_\lambda(f)| \le 1$, thus implying that the filter is stable independent of the value of $\lambda \ge 0$.

4) If $h(k)$ is a highpass filter such that $|H(f)| > 0$ for $0 < f \le 1/2$ then $w_\lambda(k)$ is a lowpass filter [cf. (10)]. Moreover, the strength of the lowpass filter is modulated by the parameter λ. We clearly have the two limiting cases:

$$\lim_{\lambda \to 0} \{W_\lambda(f)\} = 1$$
$$\lim_{\lambda \to \infty} \{W_\lambda(f)\} = \begin{cases} 1 & f = \cdot\cdot, -1, 0, 1, 2, \cdots \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

which, respectively, correspond to no filtering at all and the suppression of all non-DC frequency components.

C. Decomposition Theorems

When the stabilizing kernel $\{h(k)\}$ is of length $(n+1)$, $\{w_\lambda(k)\}$ will be referred to as an *n*th order R-filter. There are two important decomposition theorems that allow a full specification of the transfer function of such systems in term of the n smallest roots $\{z_i, i = 1, \cdots, n\}$ of the characteristic polynomial

$$P_{2n}(z) = \left[\frac{1}{\lambda} + \varphi_h(0)\right] z^n + \sum_{i=1}^{n} \varphi_h(i) \left[z^{n+i} + z^{n-i}\right]. \tag{12}$$

The first theorem is especially suited for practical implementation with a minimum number of operations. The second, which uses a decomposition into partial fractions, is useful for the explicit determination of the R-filter impulse response.

Decomposition Theorem 1: The transfer function of an nth order R-filter can be decomposed as

$$W_\lambda(z) = V^+(z) V^+(z^{-1}) \tag{13}$$

where $V^+(z)$ is the z-transform of a stable causal system and is given by

$$V^+(z) = \frac{\prod\limits_{i=1}^{n} (1 - z_i)}{\prod\limits_{i=1}^{n} (1 - z_i z^{-1})}. \tag{14}$$

Furthermore, λ is related to the roots of the characteristic polynomial through the relation

$$\lambda = (-1)^n \frac{1}{\varphi_h(n)} \prod_{i=1}^{n} \frac{z_i}{(1 - z_i)^2}. \tag{15}$$

Proof: $W_\lambda(z)$ may be written as

$$W_\lambda(z) = \frac{z^n}{\lambda P_{2n}(z)} \tag{16}$$

Since $W_\lambda(z) = W_\lambda(z^{-1})$, we have that if z_i is a root of $P_{2n}(z_i)$ so is z_i^{-1} as long as $z_i \ne 0$. This means that the roots of the polynomial appear in reciprocal pairs. Furthermore, we know that the system is stable. This implies that the roots cannot be on the unit circle ($|z_i| \ne 1$). Consequently, there must be n roots with a modulus smaller than one, which we denote by $\{z_i, i = 1, n\}$. Clearly, the n reciprocal roots with a modulus greater than one are $\{z_i^{-1}, i = 1, n\}$. Hence, the characteristic polynomial can be expressed as

$$P_{2n}(z) = \varphi_h(n) \prod_{i=1}^{n} (z - z_i)(z - z_i^{-1}). \tag{17}$$

By dividing (17) by z^n and multiplying it by the product of the roots, we find that

$$W_\lambda(z) = \frac{(-1)^n \prod\limits_{i=1}^{n} z_i}{(\lambda \varphi_h(n)) \prod\limits_{i=1}^{n} (1 - z z_i)(1 - z_i z^{-1})}. \tag{18}$$

We now use the property that $W_\lambda(l) = 1$ to get (15). By substituting the corresponding value of λ in (18) and identifying the terms in z^{-1}, we finally obtain get (14). Q.E.D.

Decomposition Theorem 2: The transfer function of an nth order R-filter can be decomposed as

$$W_\lambda(z) = W^+(z) + W^+(z^{-1}) - w^+(0) \tag{19}$$

where $W^+(z)$ is z-transform of the impulse response of a stable causal system defined as

$$w^+(k) = \begin{cases} w_\lambda(k) & k \ge 0 \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

and is given by

$$W^+(z) = \sum_{j=1}^{n} \frac{a_j}{1 - z_j z^{-1}} \tag{21}$$

where

$$a_j = (-1)^n \left(\prod_{i=1}^{n} \frac{(1 - z_i)^2}{z_i} \right) \times \frac{1}{(z_j - z_j^{-1}) \prod\limits_{\substack{i=1 \\ i \ne j}}^{n} (z_j + z_j^{-1} - z_i - z_i^{-1})}. \tag{22}$$

Proof: Since $w_\lambda(k)$ is an even function, it can be decomposed as: $w_\lambda(k) = w^+(k) + w^+(-k) - w(0)$. Taking the z-transform, we get (19). Furthermore, the stability of $w(k)$ (which also implies that $|z_i| < 1$) guarantees that of $w^+(k)$. By using the results of the previous decomposition, $W_\lambda(z)$ can be rewritten as

$$W_\lambda(z) = c_0 \frac{z^n}{\prod_{i=1}^{n} (z - z_i)(z - z_i^{-1})}$$

where the constant c_0 is given by

$$c_0 = \frac{1}{\lambda \varphi_h(n)} = (-1)^n \prod_{i=1}^{n} \frac{(1 - z_i)^2}{z_i}.$$

We use a standard decomposition into simple partial fractions

$$W_\lambda(z) = \sum_{j=1}^{n} \frac{\alpha_j}{z - z_j} + \sum_{j=1}^{n} \frac{\beta_j}{z - z_j^{-1}} \tag{23}$$

for which it is relatively straightforward to establish that

$$\alpha_j = \frac{c_0 z_j}{(z_j - z_j^{-1}) \prod_{\substack{i=1 \\ i \neq j}}^{n} (z_j + z_j^{-1} - z_i - z_i^{-1})} \quad \text{and}$$

$$\beta_j = -\alpha_j z_j^{-2}. \tag{24}$$

We then note that (23) is equivalent to

$$W_\lambda(z) = \sum_{j=1}^{n} \frac{-z_j^{-1} \alpha_j}{1 - z_j z^{-1}} + \sum_{j=1}^{n} \frac{-z_j \alpha_j}{1 - z_j z} - \sum_{j=1}^{n} z_j^{-1} \alpha_j$$

which, through the use of (24), allows a direct identification of the terms in (19) and (22). Q.E.D.

D. Efficient Implementation

An efficient implementation of $\{w_\lambda(k)\}$ results directly from Theorem 1 which allows us to write $W_\lambda(z)$ as

$$W_\lambda(z) = c_1 \frac{1}{1 - b_1 z^{-1} - \cdots - b_n z^{-n}} \frac{1}{1 - b_1 z^1 - \cdots - b_n z^n} \tag{25}$$

where $c_1, b_1, \cdots, b_n$ are constant coefficients that can be obtained easily from (14). This expression tells us that $y(k)$ can be computed from a succession of two complementary causal and anti-causal recursive filters with identical coefficients:

$$y^+(k) = x(k) + b_1 y^+(k-1) + \cdots + b_n y^+(k-n) \tag{26}$$

$$y^-(k) = y^+(k) + b_1 y(k+1) + \cdots + b_n y(k+n) \tag{27}$$

$$y(k) = c_1 y^-(k). \tag{28}$$

If one neglects the last step which is just a renormalization, this filtering requires no more than $2n$ operations (1 operation = 1 add + 1 multiply) per sample value. Similarly, a separable two-dimensional nth order R-filter can be implemented with as few as $4n$ operations per pixel.

In practice, the sequences to be processed are of finite length: $\{x(k), k = 1, \cdots, K\}$. It is therefore important to provide a proper initialization of the recursive equations (26) and (27) in order to minimize border artifacts, which are especially disturbing in image processing applications. For instance, a minimum requirement is that the R-filtering of a constant signal produces the same constant. A simple approach is to use the initial conditions

$$y^+(k) = x(k)/c_0, \qquad (k = 1, \cdots, n) \tag{29}$$

$$y^-(k) = y^+(k)/c_0, \qquad (k = K - n + 1, \cdots, K) \tag{30}$$

where $c_0 = (1 - b_1 - b_2 - \cdots - b_n)$ is a proper scaling factor, which is more or less equivalent to performing a partial (one-sided) smoothing near the border regions. The remaining signal values are then computed from (26) for $k = n + 1$ to K and (27) for $k = K - n$ down to 1. Based on our experiments, we found this initialization procedure to be superior to conventional zero padding techniques in the sense that it produces no visible border artifacts.

III. Specific Examples

In this section, we use our previous results to derive the R-filters associated to the first and second order differential operators commonly used as stabilizing kernels.

A. First Order Difference (Differentiation)

The first order difference operator, which is the standard discrete approximation of the derivative, can be defined in terms of its transfer function

$$H_1(z) = (-z + 1) \quad \text{or} \quad (-1 + z^{-1}) \tag{31}$$

and correspond to the R-filter

$$W_1(z) = \frac{1}{1 + \lambda(-z^{-1} + 2 - z)} \tag{32}$$

which, according to Theorem 1, can be decomposed as $W_1(z) = V_1^+(z) V_1^+(z^{-1})$ with

$$V_1^+(z) = \frac{1 - \alpha}{1 - \alpha z^{-1}}. \tag{33}$$

The parameter α $(0 \leq \alpha \leq 1)$ is the smallest root of the characteristic polynomial (12) and is given by

$$\alpha = 1 + \frac{1}{2\lambda} - \frac{\sqrt{1 + 4\lambda}}{2\lambda}. \tag{34}$$

By using decomposition Theorem 2, it is straightforward to demonstrate that the impulse response of this filter is a symmetric exponential

$$w_1(k) = \frac{1 - \alpha}{1 + \alpha} \alpha^{|k|} \tag{35}$$

which is shown in Fig. 1(a) for various values of λ. The spread of this function, which is positive and normalized to one, is conveniently characterized by its variance. A simple way to determine this quantity is to apply the following technique which relies on the differentiation of the z-transform

$$\sigma_w^2 = \sum_{k=-\infty}^{+\infty} k^2 w_1(k) = \left. \frac{\partial^2 W_1(z)}{\partial z^2} \right|_{z=1} = 2\lambda = \frac{2\alpha}{(1 - \alpha)^2}. \tag{36}$$

This filter will have roughly the same smoothing strength as a rectangular moving average filter of size $\sqrt{12}\sigma_w$. The first order R-filter can be implemented with as few as two operations per sample value by taking $b_1 = \alpha$ in (26) and (27).

B. Second Order Difference

The discrete second order differential operator is the well known Laplacian filter described by

$$H_2(z) = H_1(z) H_1(z^{-1}) = -z + 2 - z^{-1}. \tag{37}$$

The transfer function of the corresponding second order R-filter is

$$W_2(z) = \frac{1}{1 + \lambda(z^{-2} - 4z^{-1} + 6 - 4z + z^2)} \tag{38}$$

which, according to Theorem 1, can be decomposed as $W_2(z) = V_2^+(z) V_2^+(z^{-1})$ with

$$V_2^+(z) = \frac{1 - 2\rho \cos(\omega) + \rho^2}{1 - 2\rho \cos(\omega) z^{-1} + \rho^2 z^{-2}} \tag{39}$$

where ρ and ω are the magnitude and argument of the two smallest complex conjugate roots of the characteristic polynomial ($z_1 = \rho e^{j\omega}$

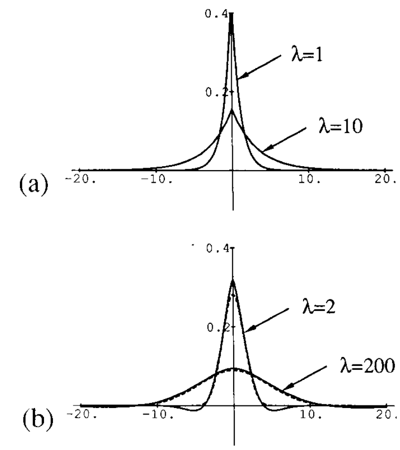

Fig. 1. (a) Impulse response of a first order R-filter for various values of λ. (b) impulse responses of a second order R-filter and its equivalent Gaussian filter (dotted line) for various values of λ.

and $z_2 = \rho e^{-j\omega}$). By determining the roots and making the necessary simplifications, we find that

$$\rho = \left(\sqrt{1+\sqrt{1+16\lambda}} - \sqrt{2}\right)\sqrt{\frac{1}{8\lambda} + \frac{1}{1+\sqrt{1+16\lambda}}} \quad (40)$$

$$\tan(\omega) = \frac{\sqrt{1+\sqrt{1+16\lambda}}}{2\sqrt{2\lambda}}. \quad (41)$$

The inverse relationship is directly obtained from (15):

$$\lambda = \frac{\rho^2}{(1-2\rho\cos(\omega)+\rho^2)^2}. \quad (42)$$

By applying the second decomposition theorem, we can show that the impulse response of this filter is

$$w_2(k) = c_0\rho^{|k|}[(\cos(\omega|k|) + \gamma\sin(\omega|k|))] \quad (43)$$

where

$$\gamma = \frac{1-\rho^2}{1+\rho^2}\frac{1}{\tan(\omega)} \quad (44)$$

and where the normalizing term c_0 is given by

$$c_0 = \frac{1+\rho^2}{1-\rho^2}\frac{1-2\rho\cos(\omega)+\rho^2}{1+2\rho\cos(\omega)+\rho^2}. \quad (45)$$

We have determined empirically that this filter resembles a normalized (zero mean and unit sum) Gaussian filter with variance

$$\sigma_w^2 \cong \sqrt{2\lambda}. \quad (46)$$

Examples of impulse responses are shown in Fig. 1(b). Their Gaussian approximations obtained using (46) have been superimposed in dotted lines and can be seen to be very similar. The main advantage of the present regularization filter is that it can be implemented with as few as four operations per sample value by taking $b_1 = \rho\cos\omega$ and $b_2 = -\rho^2$ in (26) and (27).

IV. 2-D Extensions and Applications

The results given in Section II-B carry over directly to higher dimensions. However, we have not been able to find two-dimensional equivalents of the decomposition theorems given in Section II-C. A case of special interest is when the R-filters are separable which allows a direct use of all one-dimensional results. In this section, we identify the separability conditions and consider some image processing applications.

A. Separability Conditions

The two-dimensional R-filter that minimizes the functional

$$\begin{aligned}\eta_\lambda(y) = &\sum_{k=-\infty}^{+\infty}\sum_{l=-\infty}^{+\infty}[x(k,l) - y(k,l)]^2 \\ &+ \lambda\sum_{k=-\infty}^{+\infty}\sum_{l=-\infty}^{+\infty}[y(k,l) * h_1(k)]^2 \\ &+ \mu\sum_{k=-\infty}^{+\infty}\sum_{l=-\infty}^{+\infty}[y(k,l) * h_2(l)]^2 \\ &+ \lambda\mu\sum_{k=-\infty}^{+\infty}\sum_{l=-\infty}^{+\infty}[y(k,l) * (h_1(k)h_2(l))]^2 \end{aligned} \quad (47)$$

is separable and is given by

$$w_{\lambda\mu}(k,l) = w_\lambda(k)w_\mu(l) \leftrightarrow W_{\lambda\mu}(z_1,z_2) = W_\lambda(z_1)W_\mu(z_2) \quad (48)$$

where $w_\lambda(k)$ and $w_\mu(l)$ are the one-dimensional R-filters with stabilizing kernel $h_1(k)$ and $h_2(k)$, and regularization parameters λ and μ, respectively. The functional form (47) is also necessary for separability.

Proof: By using the same procedure as in Section II-B, we can show that

$$\begin{aligned} W_{\lambda\mu}(z_1,z_2) &= \\ &\frac{1}{1+\lambda H_1(z_1)H_1\left(z_1^{-1}\right)+\mu H_2(z_2)H_2\left(z_2^{-1}\right)+\lambda\mu H_1(z_1)H_1\left(z_1^{-1}\right)H_2(z_2)H_2\left(z_2^{-1}\right)} \\ &= \frac{1}{1+\lambda H_1(z_1)H_1\left(z_1^{-1}\right)}\frac{1}{1+\mu H_2(z_2)H_2\left(z_2^{-1}\right)}\end{aligned}$$

where $H_1(z)$ and $H_2(z)$ are the z-transforms of $h_1(k)$ and $h_2(k)$, respectively. Conversely, it is easy to show that the form of the stabilizing functional in (47) is the only one that allows a separable decomposition of the filter into one-dimensional regularization operators.

The implication of this result is that commonly used 2-D stabilizing functionals such as the squared norm of the gradient [6], [8] or the energy of the Laplacian [11] cannot lead to a separable implementation. Therefore, if we are to take advantage of the separability property and the availability of fast 1-D convolution algorithms, we have to modify slightly the stabilizing functional by including not only the horizontal (k), vertical (l), but also the combined ($k \otimes l$) filtering contributions. On the other hand, this gives us somewhat more flexibility by allowing us to adjust the smoothness constraints differently in the two principal directions.

B. Application: Efficient Implementation of Edge Detectors

Edges in real images have been detected as maxima of a first-order derivative in the direction of the gradient, or as zero crossings of a second-order derivative (Laplacian). The former scheme is usually more robust and allows a better localization [7]. A natural way to compute the image gradient is to approximate the data with an analytic function and subsequently evaluate the derivatives [6], [14]. To accomplish this task, Poggio *et al.* have used regularization theory with a Laplacian stabilizing kernel which is equivalent to finding an approximating spline function [8]. The corresponding smoothing filter is usually approximated by a two-dimensional Gaussian. The vertical and horizontal components of the gradient are thereafter obtained by convolving the discrete data with the filters obtained by sampling the continuous x and y derivatives of the regularization kernel.

A similar system can be developed by applying our discrete regularization formalism. We define the components of the gradient of a discrete image to be the result of the convolution of the image with simple differential one-dimensional ($[-1\;0\;1]$) horizontal and vertical operators. If we select a stabilizing functional of the type given by (47) with a Laplacian kernel and a value of λ reflecting some *a priori*

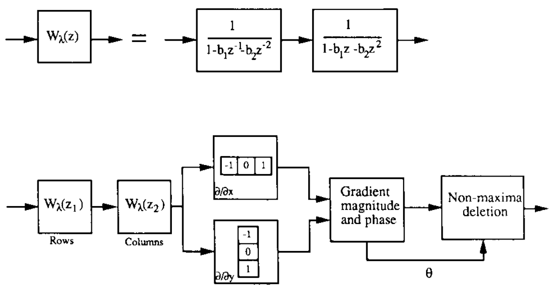

Fig. 2. Efficient implementation of the optimal Canny/Deriche edge detection algorithm using a separable two-dimensional second order R-filter.

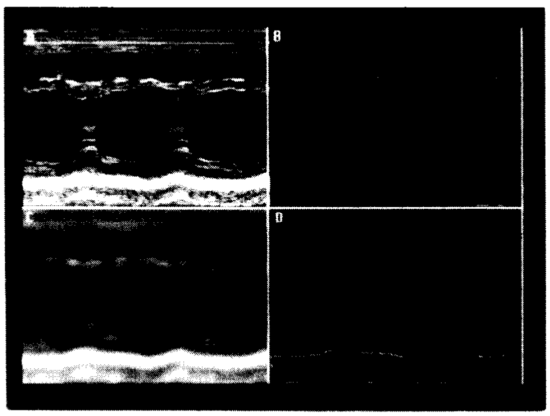

Fig. 3. Example of edge detection. (a) 231 × 157 M-mode echocardiogram ($\Delta t = 1/75$ s and $\Delta x = 1/16$ cm) [15]. (b) Output of a standard Sobel operator which is far from being satisfactory. (c) Result of smoothing with a second order R-filter with $\lambda = 40.5$ ($\sigma_w = 3$). (d) Gradient magnitude after nonmaxima deletion algorithm. These algorithms were coded in Fortran on a standard 16 MHz Apple Macintosh II personal computer. The execution time was approximately 3 s for (b), 10 s for (c), and 18 s for (d).

constraint on the smoothness of the intensity function, we obtain the edge detection algorithm shown in Fig. 2. The last two components of this block diagram correspond to a conversion of the gradient into polar coordinates and a nonmaximum deletion. This last module sets to zero the values of the gradient magnitude that are not maxima within a 3 × 1 directional window oriented in the direction of the gradient.

The smoothing operator in Fig. 2 is implemented by successive one-dimensional filtering along the rows and columns. In practice, the strength of the smoothing operator is adjusted by selecting some particular value for λ or the standard deviation of an equivalent Gaussian window [cf. (46)]. The filter coefficients in the upper block diagram in Fig. 2 are $b_1 = \rho \cos \omega$ and $b_2 = -\rho^2$ where ρ and ω are evaluated from (40) and (41). The one-dimensional filters $W_\lambda(z_1)$ and $W_\lambda(z_2)$ are implemented recursively as described in Section II-D.

The present approach has two advantages. First, signal approximation and differentiation are carried out sequentially; in other words, the output of the same smoothing R-filter can be used for the evaluation of the horizontal and vertical gradient components. Second, the computational cost is low and is independent of the strength of the smoothness constraints: only 8 multiplications (+1 optional for normalization) and 10 additions per pixel are required to compute both components of the gradient. An example of edge detection obtained with an echocardiographic image, which is typically quite noisy [15], is shown in Fig. 3.

C. Equivalence with the Canny/Deriche Operator

Interestingly enough, the system displayed in Fig. 2 is rigorously equivalent to the Canny/Deriche edge detector [12], [13], as demonstrated below. This latter operator is known to be optimal with respect to a criterion that takes into account both the efficiency of detection in the presence of noise and the reliability in localization [12]. This equivalence is consistent with the observation made by Deriche [13] who first noted the similarity between the smoothing operator associated with the Canny edge detector and the smoothing cubic spline filter derived by Poggio *et al.* using the tools of regularization theory [8].

According to Deriche [13], the transfer function of the 1-D edge detector satisfying Canny's condition of optimality with some appropriate boundary conditions is given by

$$F(z) = \frac{az^{-1}}{1 + b_1 z^{-1} + b_2 z^{-2}} + \frac{az}{1 + b_1 z + b_2 z^2} \tag{49}$$

with

$$\begin{cases} a = -ce^{-\alpha} \sin(\omega) \\ b_1 = -2e^{-\alpha} \cos(\omega) \\ b_2 = e^{-2\alpha} \end{cases} \tag{50}$$

where α and ω are the filter parameters and where c is a scaling constant. We choose to rewrite this expression as

$$F(z) = \frac{a(1 - b_2)(z^{-1} - z)}{(1 + b_1 z^{-1} + b_2 z^{-2})(1 + b_1 z + b_2 z^2)}. \tag{51}$$

If we define $c_1 = a(1 - b_2)$ and $\rho = e^{-\alpha}$, we see that

$$F(z) = W_2(z)(z^{-1} - z) \tag{52}$$

which clearly indicates that this filter corresponds to the discrete differentiation of the R-filter associated with the Laplacian (Section III-B). If we select orthogonal projections functions corresponding to the regularization kernel themselves (which incidentally is also the approach chosen by Deriche), we obtain the two-dimensional filters for the evaluation of the horizontal and vertical gradient components

$$\begin{cases} F_x(z_1, z_2) = W_2(z_1) W_2(z_2)(z_1^{-1} - z_1) \\ F_y(z_1, z_2) = W_2(z_1) W_2(z_2)(z_2^{-1} - z_2). \end{cases} \tag{53}$$

It is clear from (53) that the smoothing needs to be performed only once and that these operations are equivalent to those performed by the block diagram in Fig. 2.

However, we note that despite those similarities, our implementation (cf. Fig. 2) is computationally almost three times more parsimonious than the Deriche algorithm [13], which is also recursive but requires 12 multiplications and 10 additions per pixel and per gradient component.

D. Discussion

The formalism presented in Section II-B allows us to construct whole families of R-filters. These operators are infinite-impulse response lowpass filters. They are entirely characterized by their associated stabilizing kernel $\{h(k)\}$ and a regularization parameter λ. This latter quantity provides a convenient way to adjust the amount of smoothing.

The most important property, at least from a practical point of view, is the availability of fast recursive filtering algorithms with an amount of computation that is independent on the value of λ. This feature makes R-filters interesting alternatives to other commonly used smoothing operators. They could be useful in a large variety of standard image processing tasks such as smoothing for noise reduction [4], image interpolation, the estimation of local image statistics [16], multiresolution techniques [16], [17], or the evaluation of control parameters for adaptive processing schemes [18], [19].

One such application is the edge detection algorithm described in Section IV-C. This method has been described in greater detail because it turns out to be equivalent to the Canny edge detection algorithm, nowadays one of the most popular because of its superior performance. We believe that the coding effort for the system shown in Fig. 2 is minimal and that there is a substantial reward in algorithmic performance when compared to previous implementations.

Another application is the use of second order R-filter as a Gaussian filter. This substitution is computationally very attractive, especially for larger standard deviations. The quality of the approximation to a true Gaussian, which should be sufficient for most applications (cf. Fig. 1), can be further improved through the use of repeated convolutions (a consequence of the central limit theorem). Such an approach is suitable for the generation of multiresolution or scale-space signal representations [20], [21]. It is certainly an interesting alternative to other suggested approximation methods [22].

Our implementation of R-filters is fully recursive and some care has to be taken to avoid propagation of roundoff errors. The simplest approach, which is the one that we selected, is to use floating point arithmetic. In image processing applications where it is often necessary to save memory storage, it is sufficient to use one auxiliary 1-D real array to store intermediate filtering results. The final output of the row or column filters can be truncated and stored in standard byte or integer format. Fixed point realizations are also conceivable, provided that an error analysis be performed to determine the appropriate number of bits per sample needed to maintain the error within an acceptable range [23].

Finally, we would like to mention that the availability of fast R-filtering techniques is crucial for the design of a new class of iterative algorithms for solving linear and space variant regularization problems. It is the investigation of such filtering-based algorithms that initially motivated the present study. We are currently studying the convergence properties of such schemes. They appear to be superior to the conventional Gauss–Seidel approach, particularly for large values of λ. Further, a problem that could benefit from this approach is the area-based estimation of optical flow in the Horn and Schunk formulation [11].

V. Conclusion

In this correspondence, we have investigated the properties of R-filters, a special class of smoothing operators with an adjustable scale parameter λ. These operators provide a convenient way of solving approximation problems with certain regularization constraints. We have developed general analysis and design techniques and applied them to the study of R-filters associated with the first and second difference operators. These filters have been fully characterized in terms of their impulse response, equivalent window size, and filter coefficients, expressed as functions of the regularization parameter λ.

The R-filters that have been described here have two essential features:

1) they can be implemented recursively with a small number of operations per sample value ($2n$ operations for a one-dimensional nth order R-filter);

2) their smoothing window can be tuned to any scale through a single parameter with no effect on execution speed.

Due to these properties, R-filters stand as attractive alternatives to standard moving average and Gaussian smoothers currently used in a wide variety of image processing and computer vision applications.

References

[1] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems.* Washington, DC: Winston, 1977.

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

[3] M. Bertero, T. A. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proc. IEEE,* vol. 76, pp. 869–889, Aug. 1988.

[4] W. K. Pratt, *Digital Image Processing.* New York: Wiley, 1978.

[5] C. H. Reinsh, "Smoothing by spline functions," *Numer. Math.,* vol. 10, pp. 177–183, 1967.

[6] T. Poggio, H. Voorhees, and A. Yuille, "Regularizing edge detection," Artificial Intell. Lab., Massachusetts Inst. Technol., Cambridge, Memo 776, 1984.

[7] V. Torre and T. A. Poggio, "On edge detection," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. PAMI-8, pp. 147–163, Mar. 1986.

[8] T. Poggio, H. Voorhees, and A. Yuille, "A regularized solution to edge detection," *J. Complexity,* vol. 4, pp. 106–123, 1988.

[9] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London B,* vol. 207, pp. 187–217, 1980.

[10] D. Marr, *Vision.* San Francisco, CA: Freeman, 1982.

[11] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intell.,* vol. 17, pp. 185–203, 1981.

[12] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. PAMI-8, pp. 679–697, 1986.

[13] R. Deriche, "Using Canny's criteria to derive a recursively implemented optimal edge detector," *Int. J. Comput. Vision,* vol. 1, pp. 167–187, 1987.

[14] R. M. Haralick, "Digital step edges from zero crossing of second directional derivatives," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. PAMI-6, pp. 58–68, Jan. 1984.

[15] M. Unser, G. Pelle, P. Brun, and M. Eden, "Automated extraction of serial myocardial borders from *M*-mode echocardiograms," *IEEE Trans. Med. Imaging,* vol. MI-8, pp. 96–103, Mar. 1989.

[16] P. J. Burt, "Fast algorithms for estimating local image properties," *Comput. Graphics Image Processing,* vol. 21, pp. 368–382, 1983.

[17] M. Unser and M. Eden, "Multiresolution feature extraction and selection for texture segmentation," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 11, pp. 717–728, July 1989.

[18] J. S. Lee, "Digital image enhancement and noise filtering by the use of local statistics," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. PAMI-2, pp. 165–168, 1980.

[19] M. Unser, "Improved restoration of noisy images by adaptive least-squares post-filtering," *Signal Processing,* vol. 20, pp. 3–14, May 1990.

[20] A. P. Witkin, "Scale-space filtering," in *Proc. 4th Int. Joint Conf. Artificial Intelligence,* 1983, pp. 1019–1022.

[21] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, "Uniqueness of the Gaussian kernel for scale-space filtering," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. PAMI-8, pp. 26–33, Jan. 1986.

[22] W. M. Wells, "Efficient synthesis of Gaussian filters by cascaded uniform filters," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. PAMI-8, pp. 234–239, Mar. 1986.

[23] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

Fast B-Spline Transforms for Continuous Image Representation and Interpolation

Michael Unser, Akram Aldroubi, and Murray Eden

Abstract—**This correspondence describes efficient algorithms for the continuous representation of a discrete signal in terms of B-splines (direct B-spline transform), and for interpolative signal reconstruction (indirect B-spline transform) with an expansion factor *m*.Expressions for the *z*-transforms of the sampled B-spline functions are determined and a convolution property of these kernels is established. It is shown that both the direct and indirect spline transforms involve linear operators that are**

Manuscript received June 14, 1989; revised July 19, 1990. Recommended for acceptance by R. J. Woodham.

M. Unser is with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda, MD 20892, and INSERM, Unité 2, Hôpital Henri-Mondor, F-94010 Créteil, France.

A. Aldroubi and M. Eden are with the Biomedical and Instrumentation Program, National Institutes of Health, Bethesda, MD 20892.

IEEE Log Number 9040953.