Variational Denoising of Partly-Textured Images by Spatially Varying Constraints

Guy Gilboa*, Nir Sochen and Yehoshua Y. Zeevi

***Abstract*— Denoising algorithms based on gradient dependent regularizers, such as nonlinear-diffusion processes and total variation denoising, modify images towards piecewise constant functions. Although edge sharpness and location is well preserved, important information, encoded in image features like textures or certain details, is often compromised in the process of denoising. We propose a mechanism that better preserves fine scale features in such denoising processes. A basic pyramidal structure-texture decomposition of images is presented and analyzed. A first level of this pyramid is used to isolate the noise and the relevant texture components in order to compute spatially varying constraints based on local variance measures. A variational formulation with a spatially varying fidelity term controls the extent of denoising over image regions. Our results show visual improvement as well as an increase in the signal-to-noise ratio over scalar fidelity term processes. This type of processing can be used for a variety of tasks in PDE-based image processing and computer vision, and is stable and meaningful from a mathematical viewpoint.**

Keywords: *image denoising, texture processing, spatially varying fidelity term, nonlinear diffusion, variational image processing.*

I. Introduction

PDE-based methods have been widely used over the past decade for image denoising with edge preservation. These methods are either based on the axiomatic approach of nonlinear scale-space (nonlinear diffusions), or on the variational approach of energy functional minimization. Details regarding the interaction and close relations between these approaches can be found, for example, in [1], [35], [37].

A classical variational denoising algorithm is the total variation (TV) minimizing process of Rudin-Osher-Fatemi (ROF) [28]. This algorithm seeks an equilibrium state (minimal energy) of an energy functional comprised of the TV norm of the image I and the fidelity of this image to the noisy input image I_0:

$$E_{TV} = \int_\Omega (|\nabla I| + \frac{1}{2}\lambda(I - I_0)^2)dxdy. \tag{1}$$

This is further generalized by the Φ-formulation [10] with the functional

$$E_\Phi = \int_\Omega \left(\Phi(|\nabla I|) + \frac{1}{2}\lambda(I - I_0)^2\right) dxdy. \tag{2}$$

G. Gilboa is with the Department of Mathematics, UCLA, Los Angeles, CA 90095, USA, e-mail: gilboa@math.ucla.edu

N. Sochen is with the Department of Applied Mathematics, University of Tel-Aviv, Tel-Aviv 69978, Israel, e-mail: sochen@math.tau.ac.il

Y.Y. Zeevi is with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel. e-mail: zeevi@ee.technion.ac.il

The Euler-Lagrange (E-L) equation is

$$-F \equiv \text{div}\left(\Phi'(\cdot)\frac{\nabla I}{|\nabla I|}\right) + \lambda(I_0 - I) = 0 \tag{3}$$

where $\lambda \in \mathbb{R}$ is a *scalar* controlling the fidelity of the solution to the input image (inversely proportional to the measure of denoising). Neumann boundary conditions are assumed. The solution is usually found by a steepest descent method:

$$I_t = -F\,, \quad I|_{t=0} = I_0. \tag{4}$$

When the noise is approximated by an additive white process of standard deviation σ, the problem can be formulated as finding

$$\begin{array}{l} \min_I \int_\Omega \Phi(|\nabla I|)dxdy \\ \text{subject to } \frac{1}{|\Omega|}\int_\Omega (I - I_0)^2 dxdy = \sigma^2. \end{array} \tag{5}$$

[Note that for noise of impulsive type this method is not suitable, see e.g. [23], [6]]. In this formulation, λ can be considered as a Lagrange multiplier, computed by:

$$\lambda = \frac{1}{\sigma^2|\Omega|}\int_\Omega \text{div}\left(\Phi'(\cdot)\frac{\nabla I}{|\nabla I|}\right)(I - I_0)dxdy. \tag{6}$$

As (5) uses a scalar constraint (and a scalar λ) we refer to it as *the scalar Φ problem*. The actual function with which we work in this paper is $\Phi(z) = \sqrt{1 + \beta^2 z^2}$. The process that results from this function is an approximation of TV which is easy to implement by standard discretization of the E-L equations since it has no singularity at zero gradient. Also, when the descent is implemented by an explicit method, the time step bound is maximal (bounded by the standard CFL).

We choose it as a representative of variational denoising processes.

The performance of this, and other PDE-based methods, have shown impressive results, especially for non-textured images. The implicit assumption that underlies the formulation of these flows/equations is the approximation of images by piecewise constant functions, which are in the space of bounded variations (BV). We recall that a signal u is in BV if

$$\int_\Omega |Du| + \|u\|_{L^1} < \infty$$

where $|Du|$ is the distributional derivative (see definition in [11]). In some sense they produce an approximation of the input image as the so-called "cartoon model" and, thus, naturally dispose of the oscillatory noise while preserving edges (in some cases even enhancing them, e.g. [26]).

A good cartoon model captures much of the image important information. Yet, it has several obvious drawbacks: textures are excluded, significant small details may be left out, and even large-scale features, that are not characterized by dominant edges, may be disregarded. The purpose of this paper is to show that a relatively simple modification of the above equation yields a denoising algorithm that better preserves both the structural and textural information of the image.

Following Meyer's work [19] recent studies in the field suggested the use of weaker norms than L^2, such as G and H^{-1}, for the data fidelity term [33], [25], [3], [2]. These modern techniques can better distinguish between structural and oscillatory components and tend to reduce less contrast of the structural part. However, when used for denoising (as oppose to decomposition), these procedures still recover mainly the structural image components, where textures may be over-smoothed. The G-norm (and its various approximations) is low for oscillating patterns, and therefore does not penalize much both noise and most types of textures. The TV norm, on the other hand, penalizes strongly oscillating patterns. Therefore this type of energy minimization is still not very well adapted to capture textural parts of the image (see Fig. 6 for a comparison example).

Recent studies which perform decomposition to three categories – structure, texture and noise – by PDE's and wavelets [4], [29] seem to be an appropriate solution also for denoising of natural images. We believe that our approach can complement and improve the results presented in these papers. As the weight parameters of these methods are not spatially varying, a constant "ratio" is implicitly assumed between structures, textures and noise components throughout the image. Whereas noise can often be regarded as spatially invariant (e.g. white Gaussian noise), textures and structures are not homogenously spread in the image (in terms of variance). Another new approach is the use of Bregman iterations proposed by Osher et al. [24]. In this method as well, a scalar weight parameter λ is used. Our approach may therefore be introduced in the future also for these new types of more sophisticated (and complicated) denoising schemes.

In order to keep the presentation of the idea simple and focused, we retain in this paper the more classical variational regularization based on the gradient magnitude and L^2 fidelity term.

II. The Cartoon Pyramid Model

The cartoon model has been defined and investigated in the early 80's [12], [7], was further elaborated in [20], [21] and is widely used as the basic underlying model for many image denoising methods. In the continuous case, the cartoon has a curve Γ of discontinuities, but everywhere else it is assumed to have a small or a null gradient $|\nabla I|$. A multi-layered image representation was suggested by [18], mainly for compression purposes. However, the wavelet compression (thresholding) technique for extracting the cartoon part, by using a very high compression factor, produces blurry results which do not preserve well edges.

The TV and other gradient-dependent filters are especially good for extracting the cartoon part of the image. We use them, therefore, as a simple pyramid (scale-space) of rough image sketches at different scales. Let us define a cartoon of scale s, using the Φ process, as follows:

$$C_s \doteq I_\Phi|_{\lambda=\frac{1}{s}} \tag{7}$$

where I_Φ is the steady state of (4). See [30], [31] for a similar definition of the scale. Let us define the residue as the difference between two scales of cartoons:

$$R_{n,m} \doteq C_n - C_m \quad (n < m). \tag{8}$$

We shall refer to the Non-Cartoon part of scale s as the residue from level zero:

$$NC_s \doteq R_{0,s} = C_0 - C_s. \tag{9}$$

This cartoon and residue data structure is analogous of the pyramid of wavelet approximations. By using the definitions of (7) and (8) and integrating the E-L equation (3) we deduce the following basic properties:

Proposition 1: The cartoon pyramid model has the following scale properties:

1) The cartoon of scale 0 is the input image. $C_0 = I_0$.
2) The cartoon of scale ∞ is the mean of the input image. $C_\infty = \int_\Omega I_0(x,y)dxdy$.
3) The mean of any residue is zero. $\int_\Omega R_{n,m}dxdy = 0$.
4) A cartoon image can be built from residues of larger scales. $C_s = \sum_{n=s}^{\infty} R_{n,n+1} + C_\infty$.

Proof: To prove Property 1 we denote $M = \int_\Omega \Phi(|\nabla I_0|)dxdy$. Let us assume $I \neq I_0$ in the L^2 sense for any large λ. Specifically, there exists $h \in L^2$ such that $\|h\|_2 = \epsilon > 0$ and $E_\Phi(I) < E_\Phi(I_0)$ for $\lambda > \frac{2M}{\epsilon^2}$, where $I + h = I_0$. Then

$$\begin{aligned} E_\Phi(I_0) - E_\Phi(I) &= \int_\Omega \Phi(|\nabla I_0|) - \Phi(|\nabla I|) - \lambda\|h\|_2^2 dxdy \\ &< M - \tfrac{1}{2}\lambda\epsilon^2 \\ &< 0, \end{aligned}$$

which contradicts our assumption.

Property 2 can be proved using [1] (p. 79, Prop. 3) where we deduce that C_∞ converges to the mean image value in the L^1 sense: $\lim_{s\to\infty} \|C_s - \int_\Omega I_0 dxdy\|_{L^1} = 0$.

For Property 3 we use [1] (p. 79, Prop. 2) to show that $\int_\Omega C_s dxdy = \int_\Omega I_0 dxdy, \quad \forall s \geq 0$, and therefore

$$\int_\Omega R_{n,m}dxdy = \int_\Omega (C_n - C_m)dxdy = 0.$$

Property 4 is verified by expanding the sum expression:

$$\begin{aligned} \textstyle\sum_{n=s}^{\infty} R_{n,n+1} + C_\infty &= R_{s,s+1} + R_{s+1,s+2} + \dots \\ &\quad \dots + R_{\infty-1,\infty} + C_\infty \\ &= C_s - C_{s+1} + C_{s+1} - C_{s+2} + \dots \\ &\quad \dots + C_{\infty-1} - C_\infty + C_\infty \\ &= C_s. \end{aligned}$$

■

In Table I a cartoon pyramid example is shown. See [32] for an alternative pyramidal structure suggested recently.

In the case $\Phi(z) = z$ in Eq. (2) where we have the ROF model [28] one can relate the scale and residue directly with

the G-norm (or star norm) presented in [19] p. 30 (see also the discrete version defined in [3]). The G space is closely related to the dual of BV. The G norm $\|\cdot\|_G$ penalizes oscillatory and piece-wise constant functions in an inverse manner to TV: whereas the TV norm of oscillatory signals is high and their G norm is low, piece-wise constant signals have a low TV norm and a high G norm. For example, a signal $\sin(kx)$ in $x \in [0, 2\pi]$ and zero elsewhere has a G-norm approaching zero as $k \to \infty$, whereas its total-variation in this case approaches ∞. For more on this subject we refer the reader to [19], [33], [3], [4], [2], [25], [22].

Proposition 2: For the functional $\Phi(z) = z$, we have

1) $\|NC_s\|_G = s$.
2) $m - n \leq \|R_{n,m}\|_G \leq m + n$.
3) $\|R_{n,m}\|_G = 0, \forall n \geq \|C_0 - C_\infty\|_G$.

Proof: The first and third statements are direct consequences of our pyramid definitions and Theorem 3 and Lemma 4 in [19]. The second statement is validated by the relation $R_{n,m} = NC_m - NC_n$, using the triangle inequality and Statement 1. ■

Statement 1 states that the G-norm of the Non-Cartoon part is strictly increasing (linearly), implying that larger and less oscillatory features are incorporated in this part with the growth of s. From Statement 2, one can view $R_{n,m}$ as a texture "band" of the original image with specified upper and lower bounds of the G norm. Due to the convexity of the ROF model, the solutions for similar values of λ are sufficiently close and therefore we estimate that $\|R_{n,m}\|_G$ should be in fact closer to its lower bound $m - n$, at least when m and n are of the same order. Statement 3 shows that the telescopic buildup of the cartoon image C_s from larger texture bands, as formulated in Statement 4 of Proposition 1, is finite in practice (excluding the mean image value C_∞). For other relations connecting the scale with the G-norm see [31].

In order to construct the pyramid, the desired scales should be specified. A simple mechanism, following Gaussian and Laplacian pyramids or wavelet decompositions to detail and approximation parts, is to use predefined scales, which grow exponentially, such as $s_k = s_0\gamma^k$, where γ is some constant (e.g. $\gamma = 2$ for a dyadic scale). In Table I three levels of a pyramid are shown for $s_0 = 1$, $\gamma = 10$, $k = \{0, 1, 2\}$, as well as the zero and infinite scales.

In this pyramid larger scales retain high frequencies (edges) and one does not re-sample or decimate the image to a smaller size. This gives more freedom for choosing any set of scale values. Specifically, the multi-scale decomposition can be image driven. This topic demands more study and would not be elaborated in this paper. For some preliminary directions suggested by the authors and colleagues on how to select image-driven structure-texture splitting parameters see [15], [5].

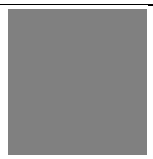
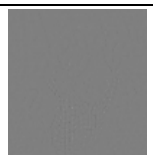
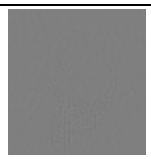
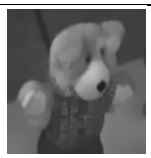
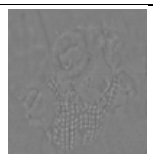
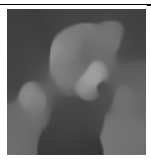
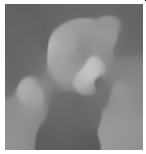
Scale	Cartoon	Non-Cartoon	Residue $R_{n,m}$
0			
1			
10			
100			
∞			

TABLE I

EXAMPLE OF A CARTOON PYRAMID. LEFT COLUMN - SCALE $s = 1/\lambda$, SECOND COLUMN - CARTOON PART C_s, THIRD COLUMN - NON-CARTOON PART NC_s, RIGHT COLUMN - RESIDUE $R_{n,m}$ (WHERE THE VALUES OF (n, m) ARE $(0, 1)$, $(1, 10)$, $(10, 100)$, $(100, \infty)$, FROM SECOND ROW DOWN, RESPECTIVELY).

A. Use for Denoising

The cartoon pyramid has a broad context and may give some more theoretical insight on issues regarding structure, texture and scale.

For our denoising purposes of the next section we use only one decomposition level which should contain the noise and the textures at a similar scale or below that of the noise. In this simpler case a good representative scale could be selected using an estimate of the noise variance.

We employ the constrained problem, similar to (5) and impose

$$\frac{1}{|\Omega|}\int_\Omega (I - I_0)^2 dxdy = \alpha\sigma^2, \tag{10}$$

where $\alpha > 0$ controls the selected scale in terms of variance. Typically $1 \leq \alpha \leq 2$ so that most noise and the relevant textures of that scale are included in the residual part. In our natural images experiment we set $\alpha = 1.5$.

Our model consists of three components: $I_0 = I_C + I_{NC} + I_n$, where $I_{orig} = I_C + I_{NC}$ is the original image, I_C is the Cartoon approximation, I_{NC} is the remainder Non-Cartoon part, and I_n is an additive noise. Note that we left the definition of "non-cartoon" part vague. It, typically, consists of textures, small-scale details, thin lines etc. The only assumption we make is that it has zero mean. Under this decomposition, the residue of the filtered image I is:

$$I_R \equiv I_0 - I = \tilde{I}_{NC} + \tilde{I}_n\ . \tag{11}$$

Note that we distinguish between the "true" non-oscillatory part and its approximation by the Φ diffusion process by the tilde upperscript.

III. The Adaptive Φ Problem

To obtain an adaptive scheme, we generalize the Φ denoising problem by imposing a spatially varying variance constraint. Let us define first a measure to which we refer as the *local variance*:

$$P_z(x,y) \equiv \frac{1}{|\Omega|}\int_\Omega (I_z(\tilde{x},\tilde{y}) - \eta[I_z])^2 w_{x,y}(\tilde{x},\tilde{y})d\tilde{x}d\tilde{y}, \quad (12)$$

where $w_{x,y}(\tilde{x},\tilde{y}) = w(|\tilde{x}-x|,|\tilde{y}-y|)$ is a normalized ($\int_\Omega w_{x,y}(\tilde{x},\tilde{y})d\tilde{x}d\tilde{y} = 1$) and radially symmetric smoothing window, $\eta[\cdot]$ is the expected value taken with respect to the probability density $w_{x,y}(\tilde{x},\tilde{y})/|\Omega|$ on the set $\Omega\times\Omega$ of all quadruples $(x,y,\tilde{x},\tilde{y})$. From the definition of the local variance it follows that $\int_\Omega P_z(x,y)dxdy = \mathcal{P}_z$, where

$$\mathcal{P}_z \equiv \text{var}(I_z). \quad (13)$$

We reformulate the scalar Φ problem, stated in Eq. (5), in the context of the adaptive Φ problem as follows:

$$\begin{array}{l} \min_I \int_\Omega \Phi(|\nabla I|)dxdy \\ \text{subject to } P_{\hat{R}}(x,y) = S(x,y), \end{array} \quad (14)$$

where $I_{\hat{R}} = (I - I_0 - C)$, C is a constant and $S(x,y) \geq 0$ is assumed to be given *a-priori*. We solve the optimization problem using Lagrange multipliers:

$$E = \int_\Omega (\Phi(|\nabla I|) + \frac{1}{2}\lambda(x,y)P_{\hat{R}}(x,y))dxdy. \quad (15)$$

The Euler-Lagrange (EL) equation for the variation with respect to I is

$$\bar{\lambda}(x,y)(I - I_0 - C) - \text{div}\left(\Phi'(\cdot)\frac{\nabla I}{|\nabla I|}\right) = 0, \quad (16)$$

where for any quantity $X(x,y)$ we define the locally averaged quantity $\bar{X}(x,y) = \int_\Omega X(\tilde{x},\tilde{y})w_{x,y}(\tilde{x},\tilde{y})d\tilde{x}d\tilde{y}$. We solve this equation for I by a gradient descent:

$$I_t = \bar{\lambda}(x,y)(I_0 - I + C) + \text{div}\left(\Phi'(\cdot)\frac{\nabla I}{|\nabla I|}\right). \quad (17)$$

In order to compute the value of λ we multiply the EL equation (16) by $(I - I_0 - C)$ and integrate over the domain Ω. After a change in the order of integrals in the λ term we get

$$\int_\Omega (\lambda(x,y)S(x,y) - Q(x,y))dxdy = 0, \quad (18)$$

where

$$Q(x,y) = (I - I_0 - C)\text{div}\left(\Phi'(\cdot)\frac{\nabla I}{|\nabla I|}\right).$$

A sufficient condition is

$$\lambda(x,y) = \frac{Q(x,y)}{S(x,y)}. \quad (19)$$

Finally, the constant C is obtained by solving $\partial_C E = 0$, yielding

$$C = \frac{\int_\Omega \lambda(x,y)(\bar{I}(x,y) - \bar{I}_0(x,y))dxdy}{\int_\Omega \lambda(x,y)dxdy}. \quad (20)$$

A. Automatic Texture Preserving Denoising

In the general case, we do not have much prior knowledge on the image that can facilitate the denoising process. In our model we assume that the noise is additive, uncorrelated with the signal (e.g. additive white Gaussian or uniform noise) and that its variance can be estimated.

Our aim is to use the Φ denoising mechanism in a more accurate and precise manner. Images which can be well represented by large scale cartoon model are the best candidates for successful denoising. Images with much fine texture and details will not benefit that much from the operation; while reducing most of the noise, this type of processing will inevitably degrade important image features. The first problem is to distinguish between good and bad candidates for Φ denoising. The task becomes even more complex if this is done adaptively. Many natural images exhibit a mosaic of piecewise smooth and texture patches. This type of image structure calls for position (spatial)-varying filtering operation.

The performance of the scalar Φ denoising process is illustrated in Fig. 1, using a typical cartoon-type and textured images. The SNR's of these three processed images are summarized in Fig. 2, and plotted as a function of the residual variance (normalized variance of the residue). Obviously, as these examples illustrate, cartoon-type images are denoised much better than textured images (both in terms of SNR and visually). Another important observation is that the maximal SNR of cartoon and non-cartoon images is reached at different levels of denoising. Whereas cartoon-type images reach their peak SNR at high denoising levels ($\mathcal{P}_R \approx \sigma^2$), non-cartoon images degrade faster and require less denoising ($\mathcal{P}_R < \sigma^2$). For deeper analysis and some bounds on the resulting SNR of Φ process denoising see [14], [13].

We present here a relatively simple method that can approximate the desired level of denoising in a region. In our above formulation (Eq. 14), the problem reduces to finding $S(x,y)$.

We use the cartoon pyramid model for this purpose. Our first aim is to differentiate between the cartoon part of the image I_C and the noise and texture parts $I_{NC} + I_n$. Our splitting parameter λ (or scale $s = \frac{1}{\lambda}$) is selected using Eq. (10). We assign

$$S(x,y) = \frac{\sigma^4}{P_R(x,y)}, \quad (21)$$

where $P_R(x,y)$ is the local variance of the residue I_R.

In the case where $I_R \approx I_n$ (basic cartoon model without textures or fine scale details) this scheme is similar to the scalar Φ process. In this case α should be close to 1. The local variance of the residue is almost constant ($P_R(x,y) \approx \sigma^2$) and hence $S(x,y) \approx \sigma^2$. We get a high quality denoising process where $I \approx I_C = I_{orig}$. In the case of most natural images, however, textures will also be filtered and included in the residue part. As the noise is uncorrelated with the signal, we can approximate the total variance of the residue as $P_{NC}(x,y) + P_n(x,y)$, the sum of local variances of the non-cartoon part and the noise, respectively. Thus, textured regions are characterized by high local variance of the residue. In order to preserve the detailed structure of such regions, the level of filtering there should be reduced over these regions.

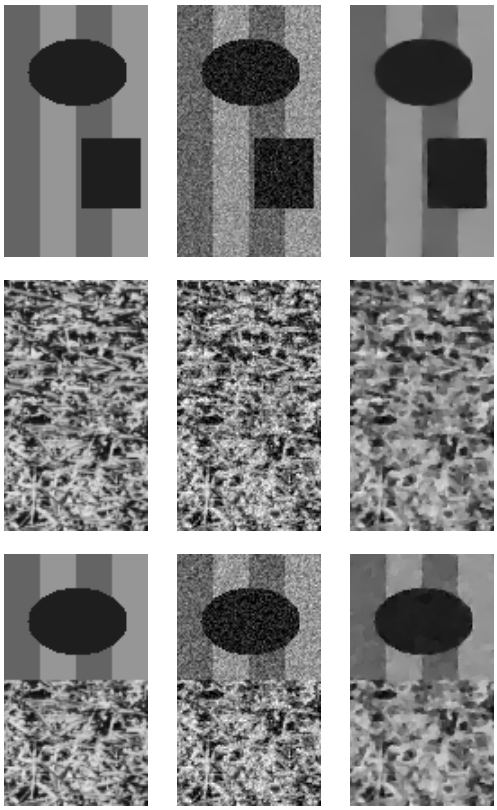


Fig. 1. Scalar Φ denoising of textured and texture-free images. Top row: Piecewise constant image, middle row: Textured image of grass, bottom row: Patches of the two types of images combined in one. Left column depicts the original images, middle column - noisy images, right column - result of scalar Φ processing (Eq. 3) at convergence ($\mathcal{P}_R = \sigma^2$).

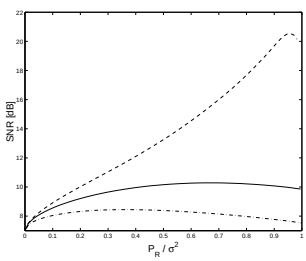


Fig. 2. SNR of scalar Φ denoising of images shown in Fig. 5. SNR is plotted as a function of the residual variance, normalized by the noise variance: $\mathcal{P}_R/\sigma^2$. Dashed line piecewise constant image, dash-dot line texture image, solid line combined image.

Let us recall the classical Wiener filter (optimal linear filter in the mean squared-error sense). Its formulation in the frequency domain is

$$G(\omega) = \frac{Ps(\omega)}{Ps(\omega) + Pn(\omega)}, \tag{22}$$

where $Ps(\omega)$ and $Pn(\omega)$ are the power spectrum of the signal and noise, respectively. The basic concept amounts to reduction in the extent of filtering ($G \to 1$) at frequencies where the signal power exceeds that of the noise.

In our case we have a similar principle, whereby reduction in the extent of filtering (i.e. $S \to 0$) is called for in regions where signal power exceeds that power of the noise. The signal is in this case that portion of the image accounting for the texture and fine details that may be filtered out by the Φ process. Formally, substituting for $P_R(x,y)$ in Eq. (21) the relation $P_R(x,y) \approx P_{NC}(x,y) + P_n = P_{NC}(x,y) + \sigma^2$, we get

$$S(x,y) \approx \sigma^2 \frac{1}{1 + P_{NC}(x,y)/\sigma^2}. \tag{23}$$

B. Algorithm

1) Separate the noise and relevant textures by minimizing $\int \Phi(|\nabla I|)$ subject to (10) and setting $I_R = I_0 - I$.
2) Compute the local variance of I_R by (12) and then compute the local constraints $S(x,y)$ by (21).
3) Solve (14) by iteratively evolving (17) and update $\lambda(x,y)$ and C according to (19) and (20).

C. Denoising with prior information

In cases where more information regarding the structure of the original signal is available, the performance of denoising process incorporating a spatially-varying fidelity constraint can be substantially ameliorated. The specifics are application-dependent and heuristic in nature. We therefore mention here only a few related ideas. To preserve specific features in the denoising process, such as long thin line or known types of textures, one can pre-process with the corresponding feature detector (Hough transform, texture detector). The value of $S(x,y)$ depends, then, locally on the feature detector response. Cases of spatially varying noise also fit the model. For example, in low-quality JPEG images, the boundaries between 8x8 pixel-blocks are often more noisy and the fidelity to the original data on these block boundaries should, therefore, be decreased (S increased). See [17] for a different solution by an adaptive window approach.

IV. Examples

The effects of adaptive- versus scalar-fidelity denoising are illustrated using a synthetic mosaic comprised of two textured patches juxtaposed with two smooth patches (Fig. 3). The scalar fidelity term requires that a global variance, equal to the noise variance, be filtered. As the Φ process is smoothing both texture and noise, more variance is filtered in the textured regions than in the originally smooth ones. This results in over-smoothing of textured regions, whereas smooth regions are not sufficiently denoised (Fig. 3, left side second row from top). The adaptive fidelity term process (second row right) applies different levels of denoising in different regions. This improves the result both visually (texture is better preserved, smooth regions are better denoised) and in terms of signal-to-noise ratio. At the third row of Fig. 3, we show how the required spatially varying variance, $S(x,y)$ (middle), depends on the value of the residue, I_R (left). The value of the adaptive fidelity term, $\lambda(x,y)$ (right), is shown for the converged process (lighter regions indicate higher value). Naturally, the values of $\lambda(x,y)$ are inversely related to the residual variance measure $S(x,y)$.

Processing a noisy version of the Barbara image (Fig. 4), it is demonstrated how the adaptive Φ method performs well on natural images. Our simple local variance criterion seems to be

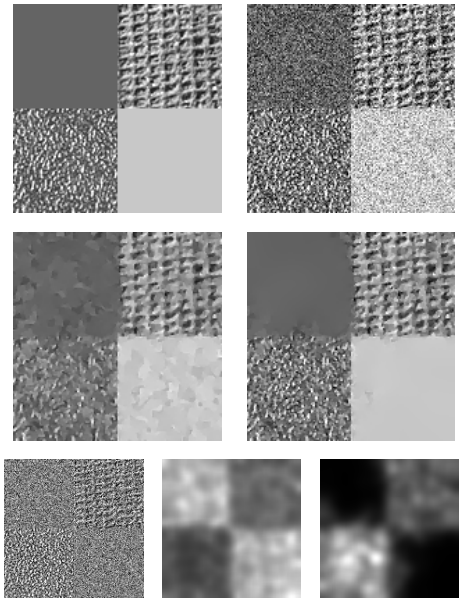


Fig. 3. Processing of a noisy mosaic of textures (fabric and metal) and smooth areas. From top: Original mosaic made of patches of fabric and metal textures, juxtaposed with two constant patches (left); Noisy version, I_0, of the original with SNR=2.4dB, $\sigma = 40$ (right); Result of processing with scalar λ - SNR=6.4dB (left), result with adaptive λ - SNR=7.6dB (right); Bottom row: Residue I_R (left); $S(x, y)$ calculated according to the residue (middle); $\lambda(x, y)$ at the convergence of the process (right).

sufficient to differentiate textured from smooth regions, even in relatively complex images. Accordingly, appropriate local requirements on the variance to be filtered are applied. In Fig. 5, Barbara's right knee is enlarged to highlight similar phenomena to those obtained in the case of the synthetic example, where textures are preserved and the denoising of smooth regions is stronger.

Fig. 6 shows the Teddy-bear from the Toys image. A comparison is performed also to the $TV - H^{-1}$ model of [25] and to [3] which implements Meyer's $TV - G$ model [19] (with a small L^2 residual). The scalar fidelity terms are chosen such that the variance of the residual is σ^2. Our algorithm diminishes the denoising in the textural parts of the bear. The $TV - G$ model is quite competitive, but still degrades the shirt textures. Similar effects can be seen in Fig. 7 where the process is compared with the regularized version of P-M [26] proposed by Catte et al. [8]. Relatively small regularization is used for the gradient computation of the diffusion coefficient (variance of Gaussian is 0.1). This causes some isolated points to remain. Stronger regularization in our experiments resulted in extensive over-smoothing of the textures. One may observe that the textural snow background is better preserved by our proposed method, while the smooth coat parts are well denoised.

In Table II we show the comparison between scalar and adaptive processes in terms of SNR. In the scalar process we show two cases. The "Standard scalar" column refers to the λ chosen according to the constrained problem (5). The "Optimal scalar" refers to choosing the parameter λ such that the maximal SNR of the recovered image is reached (out of a finite set of 30 optional values). This result naturally can be achieved only in simulations when the original clean image is at hand (see [13] for

Fig. 4. An example of processing results obtained with a natural image. From top: Original 'Barbara image' (left); Noisy version of the original image, I_0, with SNR=8.7dB, $\sigma = 20$ (right); Result of processing with scalar λ (SNR=12.6dB, left); Result of processing with adaptive λ (SNR=14.2dB, right); Residue I_R (left); $S(x, y)$ calculated according to residue (middle) $\lambda(x, y)$ at convergence of process (right).

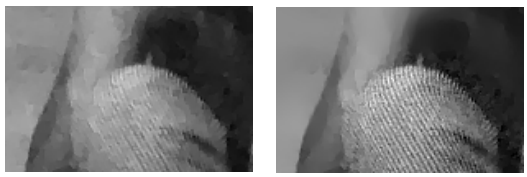


Fig. 5. Enlargement of Barbara's right knee (full images are in Fig. 3). Top - result of scalar process, bottom - result of adaptive process.

a way to approximate this parameter). Nevertheless, our algorithm consistently achieves better SNR then the optimal scalar. Note that with respect to the SNR criterion modern multiscale wavelet-based techniques achieve better performance (see e.g. [27]). This is due to their ability to denoise well also the textural parts. However, in general wavelet denoising produces less sharp results near edges and may have some oscillations. This affects the denoising quality but is less reflected by the SNR criterion. Convex gradient-based variational denoising methods admit the maximum principle and do not produce oscillatory solutions. Our algorithm retains these desired qualities.

Implementation details

We used explicit Euler schemes to implement the iterative processes. The averaging window $w(x, y)$ was selected to be a Gaussian of standard deviation $\sigma_w = 5$. The potential in all images was $\Phi(s) = \sqrt{1 + s^2}$ ($\beta = 1$). As we used gray level images with values in the range $[0, 255]$ the results are similar to TV denoising. We observed that the calculation of the constant

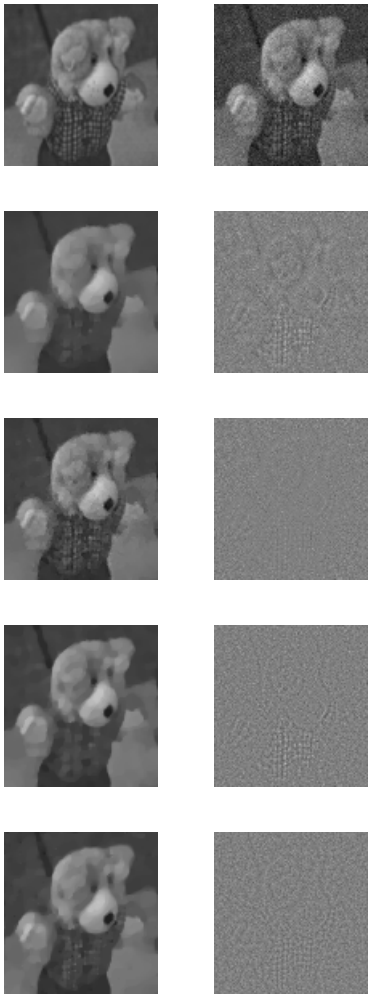


Fig. 6. Part of the Toys image. Top: original (left), noisy (right). Each row depicts the denoised image I (left) and the residual $I_0 - I$ (right) of the following models: $TV - L^2$ (scalar), our proposed adaptive $TV - L^2$, scalar $TV - H^{-1}$ [25] and scalar $TV - G$ [19,3], from second row to bottom, respectively.

C gives very little improvement. Therefore we set $C = 0$. In the experiment on natural images (results shown in Table 1) we set a constant residue variance $\mathcal{P}_R = 1.5\sigma^2$ ($\alpha = 1.5$ in (10)). Texture patches were taken from the VisTex archive [34]. All images were processed automatically with the same parameters (no tuning of parameters was performed for each image).

V. Conclusion

The widely-used variational denoising algorithms with global variance constraints perform well on simple cartoon-type images, where most of the information is represented by the simple structural approximation of the image. However, in order to preserve texture and small scale details, more subtle constraints are called for. We developed an adaptive variational scheme that controls the level of denoising by local variance constraints.

A pyramidal model of structure and texture was presented in which the structural component at any scale could be built in a telescopic manner by texture bands of subsequent higher scales.

Fig. 7. Comparison between regularized Perona-Malik and our adaptive scheme. Top: original (left), image contaminated by additive white Gaussian noise (right, $\sigma = 15$). Bottom: image denoised using regularized Perona-Malik (P-M) (left) and processing with adaptive λ. Textures and small scale features are kept better in our scheme.

Image	SNR_0	Optimal Scalar	Standard Scalar	Ours - Adaptive
Cameraman	15.86	19.56	19.32	**20.81**
Lena	13.47	18.19	17.65	**18.59**
Boats	15.61	20.23	19.83	**20.62**
Sailboat	10.36	15.51	15.16	**16.30**
Toys	10.00	17.69	17.29	**17.72**

TABLE II

Denoising results of a few classical images. From left, SNR of the noisy image (SNR_0), SNR's of scalar λ denoising ('Optimal' and 'Standard', see explanations for this table for details), SNR of our adaptive λ denoising ('Ours - Adaptive'). All experiments were done on images degraded by additive white Gaussian noise ($\sigma = 10$).

Some insight on the decomposition was given also with relation to Meyer's G norm [19].

Following this image model we use the scalar process to separate the noise and the relevant textures of the image which could be degraded in the denoising process. Regions of the residual part with higher local variance than that of the noise are treated as textured regions where denoising should be inhibited (in a soft manner). This is accomplished by introducing a new variational formulation with local variance constraints. *A-priori* knowledge on the details to be preserved can further enhance this method.

We have shown that the proposed scheme can filter noise better than the scalar constraint process over a variety of synthetic and natural images. Visually, the processed images look more natural and less 'cartoon-like'. With respect to SNR, our al-

gorithm consistently achieves higher SNR than the optimal that could be achieved with a single scalar value of λ. This study assumed a simple regularizing model based on the gradient magnitude and L^2 fidelity. The ability to effectively reduce noise from textural parts is therefore limited. Further improvement may be gained by combining PDE-based and wavelet-based methods in a spatially-varying manner, for structures and textures, respectively.

Local variance constraints can be used in almost any variational denoising schemes including ones with more sophisticated fidelity terms that are better adequate for oscillatory patterns [19], [33], [25], [3], [2].

Acknowledgments

This research has been supported in part by the Ollendorf Minerva Center, the Fund for the Promotion of Research at the Technion, Israel Academy of Science, Tel-Aviv University fund, the Adams Center and the Israeli Ministry of Science. GG is presently supported by grants from the NSF under contracts ITR ACI-0321917, DMS-0312222, and the NIH under contract P20 MH65166.

REFERENCES

[1] G. Aubert, P. Kornprobst, *Mathematical Problems in Image Processing*, Springer-Verlag, Applied Mathematical Sciences, Vol. 147, 2002.

[2] J.F. Aujol, G. Gilboa, "Implementation and parameter selection for BV-Hilbert space regularizations", UCLA CAM Report 04-66, November, 2004.

[3] J.F. Aujol, G. Aubert, L. Blanc-Féraud, A. Chambolle, "Image decomposition into a bounded variation component and an oscillating component", J. of Mathematical Imaging and Vision (JMIV), 22(1), January 2005.

[4] J.F. Aujol, A. Chambolle, "Dual norms and image decomposition models", Int. J. of Computer Vision (IJCV), 63(1), pp. 85-104, June 2005.

[5] J.F. Aujol, G. Gilboa, T. Chan, and S. Osher. "Structure-texture image decomposition – modeling, algorithms, and parameter selection", UCLA CAM Report 05-10, ftp://ftp.math.ucla.edu/pub/camreport/cam05-10.pdf, 2005.

[6] L. Bar, N. Sochen, N. Kiryati, "Image deblurring in the presence of salt-and-pepper noise", Scale-Space '05, LNCS 3459, pp. 107-118, 2005.

[7] A. Blake, A. Zisserman, *Visual Reconstruction*, MIT Press 1987.

[8] F. Catte, P. L. Lions, J. M. Morel and T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion", SIAM J. Num. Anal., vol. 29, no. 1, pp. 182-193, 1992.

[9] T.F. Chan, J. Shen, "A good image model eases restoration - on the contribution of Rudin-Osher-Fatemi's BV image model", IMA preprints 1829, Feb. 2002.

[10] R. Deriche, O. Faugeras, "Les EDP en Traitement des Images et Vision par Ordinateur", Traitement du Signal, 13(6), 1996.

[11] L.C. Evans, R.F. Gariepy, *Measure Theory and Finite Properties of Functions*, Boca Raton, FL: CRC Press, 1992.

[12] S. Geman, D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", IEEE Trans. PAMI, vol. 6, pp. 721-741, 1984.

[13] G. Gilboa, N. Sochen, Y.Y. Zeevi, "Estimation of the optimal variational parameter via SNR analysis", Scale-Space '05, LNCS 3459, pp. 230-241 ,2005.

[14] G. Gilboa, N. Sochen, Y.Y. Zeevi, "Estimation of optimal PDE-based denoising in the SNR sense", CCIT Report No. 499, Technion, August, 2004, see http://www.math.ucla.edu/∼gilboa/.

[15] G. Gilboa, N. Sochen, Y.Y. Zeevi, "Texture preserving variational denoising using an adaptive fidelity term", Proc. Variational and Level-Set Methods 2003, Nice, France, pp. 137-144, 2003.

[16] G. Gilboa, N. Sochen, Y.Y. Zeevi, "PDE-based denoising of complex scenes using a spatially-varying fidelity term", Proc. ICIP 2003, Barcelona, Spain, pp. 865-868, 2003.

[17] C. Kervrann, "An adaptive window approach for image smoothing and structures preserving", ECCV '04 (3) pp. 132-144, 2004.

[18] F.G. Meyer, A. Averbuch, R.R. Coifman, "Multi-layered image representation: application to image compression", IEEE Trans. Image Processing 11(9), pp. 1072-1080, 2002.

[19] Y. Meyer, *Oscillatory Patterns in Image Processing and Nonlinear Evolution Equations*, Vol. 22 University Lecture Series, AMS, 2001.

[20] D. Mumford, J. Shah, "Optimal approximations by piece-wise smooth functions and assosiated variational problems", Comm. Pure and Appl. Math., LII (1989), 577-685D.

[21] D. Mumford, "The Bayesian Rationale for Energy Functionals", B M ter Haar Romeny Ed., *Geometry Driven Diffusion in Computer Vision*, Kluwer Academic Publishers, 1994.

[22] A. Obereder, S. Osher, O. Scherzer, "On the use of dual norms in bounded variation type regularization", UCLA CAM Report 04-35, June 2004.

[23] M. Nikolova, "A variational approach to remove outliers and impulse noise", *JMIV*, 20(1-2):99–120, 2004.

[24] S. Osher, M. Burger, D. Goldfarb, J. Xu and W. Yin, "An Iterative Regularization Method for Total Variation Based Image Restoration", UCLA CAM Report 04-13, 2004.

[25] S. Osher, A. Sole, L. Vese, "Image decomposition and restoration using total variation minimization and the H^{-1} norm", SIAM Multiscale Modeling and Simulation, 1(3), pp. 349-370, 2003.

[26] P. Perona, J. Malik, "Scale-space and edge detection using anisotropic diffusion", PAMI 12(7), pp. 629-639, 1990.

[27] J. Portilla, V. Strela, M. Wainwright, E.P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain", IEEE Trans. Image Processing. 12(11), pp. 1338-1351, 2003.

[28] L. Rudin, S. Osher, E. Fatemi, "Nonlinear Total Variation based noise removal algorithms", Physica D 60 259-268, 1992.

[29] J.L. Starck, M. Elad, D.L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach", to appear in IEEE Trans. Image Processing, 2005.

[30] D.M. Strong, T.F. Chan, "Edge-preserving and scale-dependent properties of total variation regularization", *Inverse Problems*, 19(6), pp. S165–S187, 2003.

[31] D.M. Strong, J.F. Aujol, T.F. Chan, "Scale recognition, regularization parameter selection, and Meyer's G norm in total variation regularization", UCLA CAM Report 05-02, ftp://ftp.math.ucla.edu/pub/camreport/cam05-02.pdf, 2005.

[32] 05-02 E. Tadmor, S. Nezzar, L. Vese, "A multiscale image representation using hierarchical (BV,L2) decompositions", SIAM Multiscale Modeling and Simulation, 2(4), 554-579, 2004.

[33] L.A. Vese, S.J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing", Journal of Scientific Computing, 19(1-3), pp. 553-572, 2003.

[34] VisTex Vision Texture Archive of the MIT Media Lab, http://www-white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html

[35] J. Weickert, "A review of nonlinear diffusion filtering", B. ter Haar Romeny, L. Florack, J. Koenderink, M. Viergever (Eds.), Scale-Space Theory in Computer Vision, LNCS 1252, Springer, Berlin, pp. 3-28, 1997.

[36] J. Weickert, "Coherence-enhancing diffusion filtering", IJCV 31(2/3), 111-127, 1999.

[37] Y. You, W. Xu, A. Tannenbaum, M. Kaveh, "Behavioral analysis of anisotropic diffusion in image processing", IEEE Trans. Image Processing, 5(11), pp. 1539-1553, 1996.