PDE-BASED DENOISING OF COMPLEX SCENES USING A SPATIALLY-VARYING FIDELITY TERM

Guy Gilboa, Yehoshua Y. Zeevi

Department of Electrical Engineering,
Technion - Israel Institute of Technology
Technion city, Haifa 32000, Israel
e-mail: gilboa@tx.technion.ac.il, zeevi@ee.technion.ac.il

Nir Sochen

Department of Applied Mathematics,
University of Tel-Aviv
Ramat-Aviv, Tel-Aviv 69978, Israel
e-mail: sochen@math.tau.ac.il

ABSTRACT

The widely used denoising algorithms based on nonlinear diffusion, such as Perona-Malik and total variation denoising, modify images toward piecewise constant functions. Though edge sharpness and location is well preserved, important information, encoded in image features like textures or small details, is often lost in the process. We suggest a simple way to better preserve textures, small details, or global information. This is done by adding a spatially varying fidelity term that controls the amount of denoising in any region of the image. This form is very simple, can be used for a variety of tasks in PDE-based image processing and computer vision, and is stable and meaningful from a mathematical point of view.

1. INTRODUCTION

Nonlinear diffusion processes have been widely used over the past decade for image denoising with edge preservation. Perona and Malik [6] proposed a nonlinear diffusion equation in the form of:

$$I_t = \nabla \cdot (c(|\nabla I|)\nabla I), \quad I|_{t=0} = I_0, \quad c > 0 \qquad (1)$$

with Neumann boundary-conditions, c being a decreasing function of the gradient.

The total variation (TV) model of Rudin-Osher-Fatemi [5] was derived from the energy functional

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

where minimizing this energy by a steepest descent method results in

$$I_t = \nabla \cdot \left(\frac{\nabla I}{|\nabla I|}\right) + \lambda(I_0 - I) \qquad (3)$$

where $\lambda \in \mathbb{R}$. This is a special case of Eq. (1) with $c(s) = 1/s$ and an additional data fidelity term.

Though the performance of these, and other PDE-based methods, have shown impressive results, recently the limitations of such processes began to raise attention [1, 2]. The implicit assumption that underlies the formulation of these flows/equations is the approximation of images by piecewise constant functions (in the BV space). In some sense they produce an approximation of the input image as the so-called ”cartoon model”, thus, naturally disposing of the oscillatory noise and preserving edges (in some cases even enhancing them, see [6]).

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 thin or are not characterized by strong edges, are often being disregarded.

We show below that by a relatively simple modification of the algorithm/equation we can get a denoising algorithm that better preserves the information of the image.

2. THE CARTOON PYRAMID MODEL

The cartoon has been defined and investigated in the early 80’s and was further elaborated by Mumford and is being used since as the basic underlying model for many image denoising methods (see [4] and the references therein). In the continuous model, the cartoon has a curve Γ of discontinuities, but everywhere else it is assumed to have a small or a null gradient $|\nabla I|$ [4].

The TV and other nonlinear diffusion processes are especially good in extracting the cartoon part of the image. We use them, therefore, as a simple pyramid (scale-space) of cartoon sketches at different scales. Since the TV is a

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. YYZ is presently supported also by the Medical imaging Group of the Department of BME, Columbia University, and by the ONR MURI Program N000M-01-1-0625.

simple, single parameter (λ) process, we choose it, in this paper, as a representative of nonlinear diffusion processes. Let us define a Cartoon of scale s, using the TV process, as follows:

$$C_s \doteq I_{TV}|_{\lambda=\frac{1}{s}} \tag{4}$$

where I_{TV} is the steady state of (3). Let us define the residue between two scales as:

$$R_{n,m} \doteq C_n - C_m \quad (n < m). \tag{5}$$

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

$$NC_s \doteq R_{0,s} = C_0 - C_s. \tag{6}$$

A few basic properties of the Cartoon and Residue parts, thus defined, are listed below:

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

The TV process dissipates energy. We remark that the term $\int_\Omega (I_0 - I)^2 dxdy$ is, actually, the power of the residue.

In order to model a natural image in a simple way, yet capture its significant characteristics, we model the image as a cartoon of a single scale with its matching residue. We term the scale so chosen, to represent the cartoon part of the image, the representative cartoon scale s_r. There can be several approaches to finding a representative scale, and, in general, an image can have several such scales. We suggest to find the representative scale by examining the stability of the gradients along scales. As a cartoon consists of mainly smooth parts, divided by edges, a stable scale range $[s_1, s_2]$ is one in which the total edge length (number and size of objects) changes very slowly. As the definition of an edge is not always clear, we resort to finding the smooth regions defined as having a gradient of less then 1% of the dynamic range of the input image. The total area (length in 1D) of smooth regions is $|N_s| = \int_\Omega \chi(N_s)dxdy$ where $\chi(A)$ is the indicator function of the set A, and we define the set of smooth points as $N_s \doteq \{(x,y) : |\nabla I(x,y)| < T_s\}$. Here $T_s = (\max_\Omega(I_0) - \min_\Omega(I_0))/100$. The set of non-smooth points is $N_{ns} = \Omega - N_s$. The smoothness area $|N_s|$ is generally increasing in scale ($|N_{ns}|$ decreasing), though monotonicity of the area, and embedding of the sets, is not guaranteed. For monotone Lyapunov functionals that can indicate stability of scales see [11]. We choose the scale s_r as one of the meta-stable states of $|N_s|$ (Figs. 1,2).

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 texture, small-scale details, thin lines etc. The only assumption we make is that it has zero mean. Under this decomposition the residue of the noisy image is

$$I_R \equiv I_0 - I = \tilde{I}_{NC} + \tilde{I}_n\ . \tag{7}$$

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

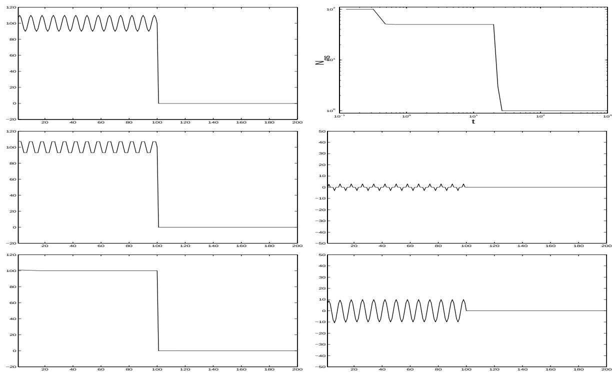

Fig. 1. Top row: Original signal I_{orig} (left), non-smoothness $|N_{ns}|$ as a function of evolution time (right), middle row: Signal (left) and residue (right) of first stable scale, bottom row: Signal (left) and residue (right) of second stable scale.

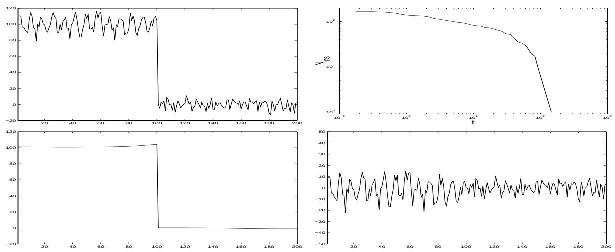

Fig. 2. Top: Noisy signal I_0 (left), non-smoothness $|N_{ns}|$ as a function of evolution time (right), bottom: signal (left) and residue (right) of stable scale.

2.1. General power-based denoising

In the general case, we do not have any significant prior knowledge on the image that can help in the denoising process. Our only assumption is that the noise is of constant power with no correlation to the signal (like additive white Gaussian or uniform noise).

Let us first define a few objects needed in the sequel: The power $P(\cdot)$ of the signal is considered here without its DC component (mean value) and is simply the signal's variance: $P(I) \equiv var(I)$. We denote $P_R \equiv P(I_R)$, $P_n \equiv P(I_n)$, $P_{orig} \equiv P(I_{orig})$, $P_0 \equiv P(I_0)$. The local power is defined as:

$$P_z(x_0, y_0) \equiv \frac{1}{|\Omega|} \int_{\Omega} (I_z(x,y) - E[I_z])^2 w_{x_0,y_0}(x,y) dxdy,$$

where $w_{x_0,y_0}(x,y) = w(|x-x_0|, |y-y_0|)$ is normalized ($\int_\Omega w_{x_0,y_0}(x,y)dxdy = 1$) and radially symmetric smoothing window. From the definition of the local spectrum it follows that $\int_\Omega P_z(x,y)dxdy = P_z$.

We begin our algorithm by finding a scale s_0 and fixing $\lambda_0 = 1/s_0$ subject to the noise constraint $P_R = P_n$. We take the respective cartoon scale C_{s_0}, and residueR_{0,s_0} of the noisy input image I_0. If the original image is a cartoon ($I_{NC} = 0$) then $C_{s_0} \approx I_C$ (as noise does not contribute, in principle, to the cartoon part). The residue R_{0,s_0} is mainly noise, except near edges, where denoising is weaker. The local power of the residue, therefore, should be similar everywhere (nearly spatially constant, like the noise power).

In the general case of natural images, however, an image has also a non-cartoon part. In such a case C_{s_0} resembles mainly the cartoon part (with some noise), where small details and texture of finer scales are severely degraded. Our aim is to preserve as much of the lost non-cartoon part as possible. This would ultimately mean a compromise between preserving some of the oscillatory part of the signal at the expanse of keeping some of the noise. We rely on a typical characteristics of natural images - textures and small details are local features. The power, therefore, of I_{NC} usually changes considerably in different areas of the image. We can model the residue of the noisy image as consisting mostly of the sum $I_{NC} + I_n$. In that case we have a new (much harder) denoising problem: Trying to separate I_{NC} from I_n. We suggest to do it according to the local power of the residue. As the noise is uncorrelated with the signal, we can approximate the total power of the residue as $P_{NC} + P_n$, the sum of powers of the non-cartoon part and the noise, respectively. Therefore a basic detector of I_{NC} is

$$g(x,y) = \frac{P_R(x,y)}{P_n}. \qquad (8)$$

The detector gets the value 1 in places of mainly noise and a larger value were the power is higher. Using a spatially varying fidelity term, we can impose different degrees of fidelity to the original data, at different locations. Since we would like to denoise less in places of texture and small details, and allow more denoising in cartoon-type regions, we suggest the following fidelity term

$$\lambda(x,y) = \lambda_0 g(x,y) = \lambda_0 \frac{P_R(x,y)}{P_n}. \qquad (9)$$

Recalling the Wiener filter formulation $G(\omega) = \frac{Ps(\omega)}{Ps(\omega)+Pn(\omega)}$, in the frequency domain, we see that Eq. (9) has similar properties. Filtering is reduced as the signal's power is stronger than the noise power (in our case it is done spatially and the preserved signal is I_{NC}, where $P_R(x,y) \approx P_{NC}(x,y) + P_n$).

Our algorithm, now, is the solution of the flow equation

$$I_t = \nabla \cdot \left(\frac{\nabla I}{|\nabla I|} \right) + \lambda(x,y)(I_0 - I) \qquad (10)$$

2.2. Denoising with prior information

When some more knowledge on the original signal is available, the performance of denoising with a spatially varying fidelity term can be substantially ameliorated. The methods here are application-dependent and more heuristic in nature. For lack of space, we will just mention a few possible ideas. To preserve specific features in the denoising process, such as long line, known type of textures etc. one can pre-process with the corresponding feature detector (Hough transform, texture detector). The value of $\lambda(x,y)$ depends locally then 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, therefore fidelity to the original data there should be decreased.

3. EXAMPLES

In Fig. 3 we show denoising in 1D of a small scale sinewave on top of one side of a step. The sinewave is better kept when the fidelity is spatially varying.

In Fig. 4 (2D case) we show denoising of a part of the Boats image (with strong noise), in comparison to two recently suggested global methods. In nonlinear diffusion evolutions, a dual problem of finding the right λ is finding the right stopping time $t = T$ when the signal is evolved without a fidelity term ($\lambda = 0$). we therefore compare our method to two advanced stopping criteria of Weickert [10] and Mrazek [3]. [1] Table 1 show the performance of the three methods on various images. Our method gives clearly better results. For all image experiments we used a Gaussian window $w_{x_0,y_0}(x,y)$ with $\sigma = 3$.

[1] One can also see our process as evolving the signal without a fidelity term and stopping the process at different times in each location.

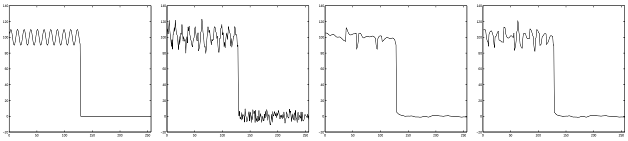

Fig. 3. From left: Original signal, noisy signal (SNR=15.1), denoised signal using scalar $\lambda = 0.02$ (SNR=20.2) and spatially varying $\lambda(x, y) = 0.02g(x, y)$ (SNR=23.4). More information on the sinewave is preserved by our method.

Image	SNR_0	Opt	Wk	Mr	Ours
Cameraman	15.8	19.2	18.9	15.7	**20.4**
Lena	13.5	18.0	17.7	17.9	**18.6**
Boats	15.6	20.0	19.6	19.8	**20.5**
Barbara	14.7	16.6	15.9	11.5	**18.1**

Table 1. Denoising results of a few classical images. From left, SNR of the noisy image (SNR_0), optimal scalar (Opt), Weickert (Wk), Mrazek (Mr) and our method. All experiments were done on images degraded by additive white Gaussian noise ($\sigma_n = 10$).

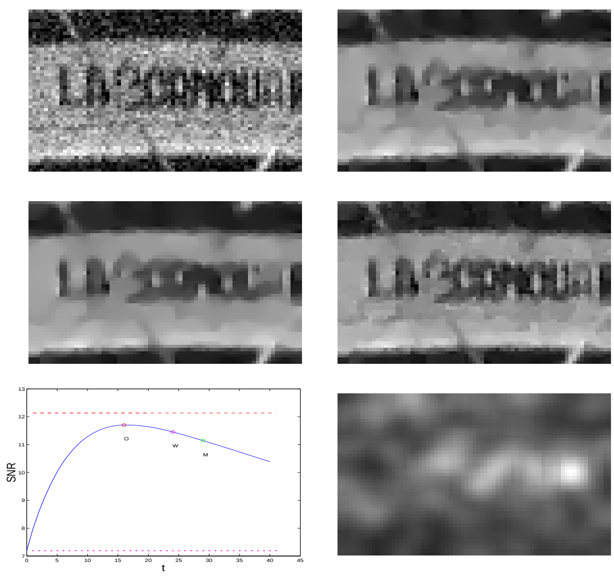

Fig. 4. From top (left to right): noisy image, processed images by Weickert, Mrazek, Our scheme. Bottom left: SNR graph as a function of the number of iterations ($\lambda = 0$). Also depicted are several important values: Dashed bottom - value of noisy input image SNR=7.2 dB. Dashed top - value of our scheme's SNR=12.1 dB. The circles on the graph mark the stopping time and consequent SNR of the various algorithms: 'O' - Optimal (SNR=11.7 dB), 'W' - Weickert (SNR=11.4 dB), 'M' - Mrazek (SNR=11.1 dB). Bottom right: $\boldsymbol{\lambda}(x, y)$.

4. CONCLUSION

A simple model for images containing cartoon and non-cartoon part was presented. Regular TV is used to extract the cartoon approximation. In order to preserve texture and small scale details, that appear locally in the image, a spatially varying fidelity term is introduced. We presented a simple mechanism based on the local power (variance) of the residue to determine the value of the fidelity term in each region. A-priory knowledge on the details to be preserved can enhance this method. Spatially varying fidelity term can be used in almost any other nonlinear diffusion method, other than TV.

We have shown that this scheme can filter noise better than modern stopping-time based mechanisms. Its performance is many times *above the optimal* possible scalar parameter (t or λ) filtering, in the SNR sense (and also visually). Further improvement may be gained in distinguishing between texture and noise by using more elaborated schemes other than the power criterion (such as transforming the residue to the Gabor/wavelet space). Work according to these ideas will be published elsewhere.

5. REFERENCES

[1] 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.

[2] L.A. Vese, S.J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing", UCLA CAM Report 02-19, May 2002.

[3] P. Mrazek, "Selection of optimal stopping time for nonlinear diffusion filtering", M. Kerckhove (ed.), Scale-Space 2001, LNCS 2106, pp.290-298, Springer-Verlag, 2001.

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

[5] L. Rudin, S. Osher, E. Fatemi, "Nonlinear Total Variation based noise removal algorithms", Physica D 60 259-268,'92.

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

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

[8] D.M. Strong, T.F. Chan, "Spatially and scale adaptive total variation based regularization and anisotropic diffusion in image processing", UCLA CAM Rep. 96-46, Nov. 1996.

[9] A. Tsai, A. Yezzi, A.S. Willsky, "Curve evolution implementation of the Mumford-Shah functional.." IEEE Trans. IP, 10(8), pp. 1169-1186, 2001.

[10] J. Weickert, "Coherence-enhancing diffusion of colour images", IVC, Vol. 17, 201-212, 1999.

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