MULTISCALE TOTAL VARIATION AND MULTISCALE ANISOTROPIC DIFFUSION ALGORITHMS FOR IMAGE DENOISING

TONY CHAN, YANG WANG, AND HAOMIN ZHOU

Abstract. As digital photography rapidly replacing the traditional film photography as the photography of choice for all but a few devoted professionals, post processing to enhance images such as denoising becomes increasingly an integral part of digital photography. In this paper we propose the multiscale total variation (MTV) and the multiscale anisotropic diffusion (MAD) algorithms for denoising. Both methods offers more flexibility than the classical TV method and the related anisotropic diffusion method. We shall discuss the algorithms as well as their implementation in details. An advantage of the MTV and the MAD methods is that an automatic stopping criterion can easilt be implemented to prevent over-processing of an image. We also raise several mathematical questions.

1. Introduction

The advent of digital imaging technologies such as digital cameras, video recorders and medical imaging devices such as MRI has changed our ways of life. Today technologies continue to evolve. The digial cameras and medical imgaing devices such as MRI continue to improve in terms of resolution and performance. On the other hand, as technologies continue to improve, consumers and patients come to expect ever more in terms of image quality. Ironically what we expect in technologies can often be contradictory. For example, as we demand more resolution in images more pixels must be packed into a sensor of a given size, which will lead to more noise in captured images. Thus to meet the expectations image denoising becomes an important challenge in digital photography and medical imaging.

Noise in images can present problems depending on the applications. For medical images noise can mask and blur important but subtle features in the images; for photography noise is a nuisance in terms of visual effect, and can ruin an otherwise memorable shot. Denoising

Key words and phrases. Denoising, total variation, wavelet TV, anisotropic diffusion, multiscale total variation, multiscale anisotropic diffusion, automatic stopping criterion.

This paper was completed while the first authors is an IPA at the National Science Foundation. The views expressed in this paper are those of the authors, and do not necessarily reflect those of the NSF. The second author is supported in part by the National Science Foundation grant DMS-0813750. The third author is supported in part by the National Science Foundation grant DMS-0410062.

of an image, however, poses challenges of its own. Most cameras have built-in denoising algorithms that often perform very poorly with soft edges and smeared out details. Post processing denoising can be more effective, but similar problems remain.

Denoising methods for monochramatic images can essentially be classified into four categories: neighborhood filters, frequency domain methods, variational PDE based methods and non-local methods. The most commonly used neighborhood filters include blurring filters, median filter or variations of it, and others such as the filters described in [32, 33]. Neighborhood filters are easy to implement and fast, and in some applications they can be effective, although in general their effectiveness is limited. In frequency domain methods a wavelet, DCT or other type of transformations is first performed. A filter is then applied to the transformation. The most common frequency domain method is the wavelet thresholding. The wavelet thresholding method assumes that noise appear in the wavelet transformation as small nonzero coefficients in the high frequency range and sets them to zero. The wavelet thresholding method is very effective in removing noise, and very fast. It is still perhaps the best "quick and dirty" denoising scheme. However, it suffers from Gibbs oscillations at discintinuities. These oscillations can be reduced, although not eliminated, by using soft wavelet thresholing [16, 17] and translational invariant wavelet thresholding [13]. Other generalizations use curvelets [3] and framelets [7]. In particular the method of iterating soft thresholded images using framelet representations in [7] is effective in removing the Gibbls oscillations. Methods that are not PDE based and do not fit the description of the other two, including some statistical methods, can be classified as non-local denoising methods, see e.g. [2, 31, 20, 25]. These methods are typically slower, and some of them assume certain statistical properties are known. Given the right images, non-local methods can yield excellent results. A particular kind of non-local method, developed in [2], works well for images with repeat patterns or large homogeneous areas, see also [20, **?**, **?**, **?**]. But in our tests on general images the success is more limited.

In this paper we propose the *multiscale total variation* (MTV) algorithm for image denoising, which has its root in the classical total variation (TV) scheme pioneered by Rudin, Osher and Fatemi [26]. Before discussing in details of the MTV algorithm we first describe denoising based on variational principles and PDE techniques. We start with a standard noisy monochromatic image model

$$z(x) = u_0(x) + n(x), \tag{1.1}$$

where $z(x)$, $u_0(x)$ and $n(x)$ are real valued functions defined on $\Omega \subset \mathbb{R}^2$, where Ω is a finite domain such as a rectangle. The function $u_0(x)$ denotes the underlying noise-free image, $z(x)$ the observed image, and $n(x)$ the noise. In our general model, we assume that $z(x)$, $u_0(x)$ and $n(x)$ are in some space of functions $\mathcal{F}$, such as $L^2(\Omega)$ or $C^1(\Omega)$. With a variational PDE based denoising method, the denoised image is the minimizer of certain energy functional $\mathbf{E}(u)$. Typically $\mathbf{E}(u)$ can be written as

$$\mathbf{E}(u) = D(u, z) + R(u), \tag{1.2}$$

where $D(u, z)$ denotes the "distance" between u and observed image z, and $R(u)$ is a regularization term that smoothes out the image. The idea of using variational method described here has been around for some time. In most cases the distance $D(u, z)$ is taken to be the L^2 distance $D(u, z) = \int_\Omega (u - z)^2$. Earlier efforts focused on least square based functionals $R(u)$'s such as $\|\Delta u\|_2^2$, $\|\nabla u\|_2^2$ and others. While noise can be effectively removed, these regularization functionals penalize discontinuity, resulting in soft and smooth reconstructed images, with subtle details lost. This is not acceptable in digital photography, as photographers often place premium emphasis on sharpness. The innovation of the total variational (TV) scheme by Rudin, Osher and Fatemi [26] is to set $R(u)$ to be the total variation $\int_\Omega |\nabla u|$ of u. With the total variation regularizer, extensive studies have shown that it does not penalize edges in u, thus it allows for sharper reconstructions, see e.g. [1, 6, 9, 15]. Among all the variational PDE based techniques, the TV minimization scheme offers one of the better combinations of noise removal and feature preservation.

It is easy to show that the TV minimization scheme leads to solving the PDE

$$\nabla \cdot \left(\frac{\nabla u}{|\nabla u|}\right) - \lambda(u - z) = 0. \tag{1.3}$$

But in practice, one introduces the time variable t and solve for $u(x, t)$ by time-marching the equation

$$u_t = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|}\right) - \lambda(u - z) = 0, \quad u(x, 0) = z(x). \tag{1.4}$$

The end result $u(x, T)$, if T is large enough, will have noise removed or reduced. In essence the time-marching by (1.4) is to use gradient flow to minimize the energy $\mathbf{E}(u)$. An important attribute of the TV minimization scheme is that it takes the geometric information of the original images into account, in that it does not penalize edges. On the contrary, significant edges are sharpened. This is similar to the anisotropic diffusion methods, see [**?**]. See also [28] and the references therein.

2. Multiscale Total Variational Method for Image Denoising

The TV minimization scheme has its own weaknesses. It is well known that when (2.4) is left running for too long a denoised image will tend to become a cartoon-like piecewise constant image, wiping out all subtle details [22, 24]. With a more gentle run, one may not remove enough noise. For optimal results it is important to have an automatic stopping criterion. This is difficult to do. Although there are some attempts in this direction [19, 23, 30], some properties of the noise (such as the variance) are assumed to be known, which is not entirely realistic for natural color photos.

A modified version of the TV denoising scheme based on wavelets was introduced in Chan and Zhou [12]. Using this scheme, Wang and Zhou [29] devised an automatic stopping criterion for the time-marching process (2.4). This criterion works surprisingly well in experiments, see [29] for a detailed discussion. In this paper we propose a couple of new methods based on the wavelet TV denoising scheme. The cartoon-like tendency in the standard TV method is a result of excessive diffusion of low frequency features in an images. The new methods, which we call *Multiscale Total Variation* (MTV) and the *Multiscale Anisotropic Diffusion* (MAD) algorithms, will concentrate the diffusion on high frequency features where the noise resides. Our tests illustrate that the MTV and the MAD methods are highly effective. Comparing to the standard TV or wavelet TV method, the two methods require fewer number of iterations to obtain comparable or better results. Figure **??** shows some comparisons.

To see how MTV and MAD algorithms work, let $\{\psi_j : \ j \in I\}$ be an orthonormal or biorthonormal wavelet basis for $L^2(\Omega)$ such as those found in [14, 27]. So we may expand any $f(x)$ as

$$f(x) = \sum_{j \in I} c_j \psi_j(x),$$

for some real (c_j). In practice, we have always used the biorthonormal 7-9 wavelet basis as our basis $\{\psi_j\}$. (The conventional notation uses two sub-indices to denote a wavelet basis. Here we use only one for brevity. There should not be any confusion.) Now expand the observed image function $z(x)$ using the basis $\{\psi_j(x)\}$,

$$z(x) = \sum_{j\in I} \alpha_j \psi_j(x).$$

Let

$$u(x,\boldsymbol{\beta}) := \sum_{j\in I} \beta_j \psi_j(x) \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_j)$. Now we set the distance functional $D(u,z)$ to be

$$D(u,z) := \sum_{j\in I} \lambda_j (\beta_j - \alpha_j)^2, \tag{2.2}$$

where $\lambda_j > 0$. The key feature is that λ_j *decreases* as the scale becomes more localized. More precisely, we use smaller values λ_j for high frequency terms and larger values for lower frequency terms. The term $R(u)$ remains to be the total variation. Thus the MTV method is to find the minimizer of the energy functional

$$\mathbf{E}_{MTV}(u,z) := \sum_{j\in I} \lambda_j (\beta_j - \alpha_j)^2 + \int_{\mathbb{R}^2} |\nabla_x u(x,\boldsymbol{\beta})| \, dx. \tag{2.3}$$

where $u = u(x,\boldsymbol{\beta})$. The idea is that since λ_j are smaller for high frequency terms the smoothing is done mostly on high frequency features. The goal of denoising is to minimize $F(u,z)$ and find the minimizer $u^* := u(x,\boldsymbol{\beta}^*)$ such that

$$\mathbf{E}_{MTV}(u^*,z) = \min_{\boldsymbol{\beta}} \mathbf{E}_{MTV}(u,z). \tag{2.4}$$

One can use simple calculus of variation to obtain the derivative of the objective functional (2.3). For $u = u(x,\boldsymbol{\beta})$ where $\boldsymbol{\beta} = (\beta_j)$,

$$\begin{aligned} \frac{\partial \mathbf{E}_{MTV}(u,z)}{\partial \beta_j} &= \int_{\mathbb{R}^2} \frac{\nabla_x u}{|\nabla_x u|} \cdot \nabla_x \psi_j dx + 2\lambda_j(\beta_j - \alpha_j) \\ &= -\int_{\mathbb{R}^2} \nabla_x \cdot \left[\frac{\nabla_x u}{|\nabla_x u|} \right] \psi_j dx + 2\lambda_j(\beta_j - \alpha_j). \end{aligned}$$

Then the Euler-Lagrange equation for the model is

$$-\int_{\mathbb{R}^2} \nabla_x \left(\frac{\nabla_x u}{|\nabla_x u|} \right) \psi_j(x) dx + 2\lambda_j(\beta_j - \alpha_j) = 0. \tag{2.5}$$

Alternatively, let $\mu_j = (2\lambda_j)^{-1}$. We may rewrite (2.5) as

$$-\mu_j \int_{\mathbb{R}^2} \nabla_x \left(\frac{\nabla_x u}{|\nabla_x u|} \right) \psi_j(x) dx + (\beta_j - \alpha_j) = 0. \tag{2.6}$$

In practice, rather than solving the Euler-Lagrange equation (2.5) or (2.6) directly for denoising, we introduce an artificial time parameter t and time-march the image using gradient flow. More precisely, set $\boldsymbol{\beta} = \boldsymbol{\beta}(t) = (\beta_j(t))$. With the Euler-Lagrange equation (2.5) we solve the following time evolution equation

$$\frac{\partial \beta_j}{\partial t} = \int_{\mathbb{R}^2} \nabla_x \cdot \left(\frac{\nabla_x u}{|\nabla_x u|} \right) \psi_j(x) dx - 2\lambda_j (\beta_j - \alpha_j), \qquad (\beta_j(0)) = (\alpha_j). \tag{2.7}$$

Alternatively, with equation (2.6) we solve the time evolution equation

$$\frac{\partial \beta_j}{\partial t} = \mu_j \int_{\mathbb{R}^2} \nabla_x \cdot \left(\frac{\nabla_x u}{|\nabla_x u|} \right) \psi_j(x) dx - (\beta_j - \alpha_j), \qquad (\beta_j(0)) = (\alpha_j). \tag{2.8}$$

Mathematically, the minimizer for the energy given by (2.3) is the steady state of either (2.7) or (2.8). However, we often stop the process before the actual minimizer is attained because, depending on the parameter λ_j the actual minimizer can be overly smoothed while noise might be effectively removed long before that. When the process stops before achieving the steady state, the two time-marching algorithms (2.7) and (2.8) are in fact two different flows. The results are subtly different as a result.

For simplicity we shall refer the algorithm given by (2.7) as MTV and the algorithm given by (2.8) as *multiscale anisotropic diffusion* (MAD). One very interesting phenomeon we have observed is that the best parameters (λ_j) for the MTV scheme for a given image do not necessarily correspond to the best parameters (μ_j) for the MAD scheme for the same image. In other words, while (λ_j) may be good for the MTV scheme, (μ_j) with $\mu_j = (2\lambda_j)^{-1}$ may not be good for the MAD scheme. In fact, as we show later, the best (μ_j) for the MAD scheme may have the opposite characteristics as $(2\lambda_j)^{-1}$. This finding is very surprising, and we do not have a complete explanation from the mathematical point of view. However, it does indicate that the denoising process using the time evolution equations stops before a steady state is reached in general. Another interesting phenomenon we have observed is that the MAD scheme seems to work better than the MTV scheme in general, often reaching a comparable level of noise removal with fewer steps. Some comparisons are given in the examples. Note that the automatic stopping criterion in [29] for the wavelet TV method is easily applicable to both MTV and MAD algorithms.

Another modification to the MTV and MAD algorithms is to allow the parameters λ_j and μ_j to depend on α_j. Rudin [**?**] has observed that the wavelet soft thresholding scheme can be improved by fixing some of the largest coefficients in the wavelet expansion. Both MTV and MAD algorithms can also be improved by allowing the parameters λ_j and μ_j to depend on α_j, as shown in our test.

3. Automatic Stopping Criterion and Examples

As we have stated earlier, one of the advantages of MTV and MAD algorithms is that there is an effective automatic stopping criterion to stop the time evolution equations (2.7) and (2.8), and thus prevent the over processing of an image. This automatic stopping criterion was first proposed in [29]. For self-containment we now describe our automatic stopping criterion for the denoising scheme with the wavelet basis $\{\psi_j : j \in I\}$. (The conventional notation for wavelet bases use two or more indices, such as $\{\psi_{jk}\}$. In this paper we only use one index for conciseness, and there should not be any confusion.) Like in the wavelet hard thresholding scheme, we first choose a threshold $\rho > 0$. Let $J_\rho = \{j \in I_D : |\beta_j(0)| = |\alpha_j| \leq \rho\}$, where $I_D \subset I$ is the index set corresponding to the diagonal portion of the highest frequency wavelet coefficients. Intuitively speaking, as in the wavelet hard thresholding scheme, the coefficients $\{\beta_j(0) : j \in J_\rho\}$ will indicate how noisy the image is. In a noise-free image these wavelet coefficients will mostly be very close to 0. But in a noisy image they will be mostly not close to zero. Define $\mu(t) = \frac{1}{|J_\rho|}\sum_{j\in J_\rho}|\beta_j(t)|$. So $\mu(t)$ measures the noise in the image at time t. The key idea is that an automatic stopping criterion of the time evolution can be designed by measuring the reduction in the value $\mu(t)$ from the orginal value $\mu(0)$.

In [29] we have described two different automatic stopping criteria, the *relative criterion* and the *absolute criterion*. Both of these can be adopted to the MTV and MAD scheme. For the relative automatic stopping criterion, we consider $\mu(t)/\mu(0)$. We will stop the time evolution whenever this value goes below a threshold b. For example, we may set $b = 0.1$. This threshold intuitively says that we stop the time evolution when we have reduced noise by 90%. For the absolute automatic stopping criterion, we stop the time evolution if $\mu(t)$ drops below a threshold c. Since in a noise-free image we expect $\mu(t)$ to be very close to zero, it is reasonable to set an absolute threshold for $\mu(t)$ to achieve a desired denoising effect.

In the actual implementation the value ρ does not seem to affect the automatic stopping time sensitively. We usually take $\rho = \frac{2}{|I_D|}\sum_{j\in I_D}|\alpha_j|$. Both the relative criterion and the absolute criterion work well, although we typically use the relative criterion. For an image with moderate noise we set the threshold b to be between 0.05 and 0.1. In more noisy cases such as the one in Figure **??**, we use smaller threshold b around 0.03. It should be pointed out that we have tested the automatic stopping time criterion on a number of noisy color images as well as monochromatic medical images. The thresholds for optimal performance stayed remarkably consistent. This is an important attribute for batch processing medical images.

References

[1] R. Acar and C. R. Vogel. Analysis of total variation penalty methods for ill-posed problems. *Inverse Prob.*, 10:1217–1229, 1994.

[2] A. Buades, B. Coll and J. M. Morel, A review of image denoising algorithms, with a new one, *Multiscale Model. Simul.*, 4 (2005), no. 2, 490–530 (electronic).

[3] E. Candès and D. Donoho, *Curvelets - A Surprisingly Effective Nonadaptive Representation for Objects with Edges*, Curves and Surfaces, L. L. Schumaker et al. Eds, Vanderbilt University Press, Nashville, TN. 1999.

[4] E. J. Candès, and F. Guo. *Edge-preserving Image Reconstruction from Noisy Radon Data*, (Invited *Special Issue of the Journal of Signal Processing on Image and Video Coding Beyond Standards.*), 2001.

[5] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier. *Nonlinear wavelet image processing: variational problems, compression and noise removal through wavelet shrinkage*. IEEE Trans. Image Processing, 7(3):319–335, 1998.

[6] A. Chambolle and P. L. Lions. Image recovery via Total Variational minimization and related problems. *Numer. Math.*, 76:167–188, 1997.

[7] R. Chan, T. Chan, L. Shen, and Z. Shen, Wavelet algorithms for high-resolution image reconstruction, *SIAM J. Sci. Comput.*, 24 (2003), no. 4, 1408–1432.

[8] T. F. Chan, S. H. Kang and J. Shen, *Euler's Elastica and Curvature Based inpainting*, SIAM J. Appl. Math., 63(2) (2002), 564-592.

[9] T. F. Chan, S. Osher, and J. Shen, *The Digital TV Filter and Nonlinear Denoising*, IEEE Trans. Image Process., 10(2), pp. 231-241, 2001.

[10] R. Chan, S. Riemenschneider, L. Shen and Z. Shen. *Tight Frame: the efficient way for high-resolution image reconstruction.* Applied and Computational Harmonic Analysis, 17(1) (2004), pp91-115.

[11] T. F. Chan and H. M. Zhou, *Total Variation Wavelet Thresholding*, to appear in J. Scientific Computing..

[12] T. F. Chan and H. M. Zhou, *Optimal Constructions of Wavelet Coefficients Using Total Variation Regularization in Image Compression*, CAM Report, No. 00-27, Dept. of Math., UCLA, July 2000.

[13] R. Coifman and D. Donoho, Translation-invariant de-noising, in *Wavelets and Statistics*, 125-150, Springer Verlag, 1995.

[14] I. Daubechies. *Ten lectures on wavelets.* SIAM, Philadelphia, 1992.

[15] D. Dobson and C.R. Vogel, *Convergence of An Iterative Method for Total Variation Denoising*, SIAM Journal on Numerical Analysis, 34 (1997), pp. 1779-1971.

[16] D. L. Donoho. *De-noising by soft-thresholding.* IEEE Trans. Information Theory, 41(3):613–627, 1995.

[17] D. L. Donoho and I. M. Johnstone. *Ideal spacial adaption by wavelet shrinkage.* Biometrika, 81:425–455, 1994.

[18] S. Durand and J. Froment, *Artifact Free Signal Denoising with Wavelets*, in Proceedings of ICASSP'01, volume 6, 2001, pp. 3685-3688.

[19] G. Gilboa, N. Sochen and Y. Zeevi, Estimation of optimal PDE-based denoising in the SNR sense, *UCLA CAM Report* 05-48, August, 2005.

[20] M. Mahmoudi and G. Sapiro, Fast image and video denoising via nonlocal means of similar neighborhoods, *IEEE Signal Processing Letter*, 12 (2005), 839-842.

[21] F. Malgouyres, *Mathematical Analysis of a Model Which Combines Total Variation and Wavelet for Image Restoration*, Journal of information processes, 2:1, 2002, pp 1-10.

[22] Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, volume 22 of *University Lecture Series*. AMS, Providence, 2001.
[23] P. Mrazek and M. Navara, Selection of optimal stopping time for nonlinear diffusion filtering, *IJCV*, v.52, no. 2/3, 2003, pp189-203.
[24] S. Osher, A. Sole, and L. Vese, *Image decomposition and restoration using total variation minimization and H_{-1} Norm*, Multiscale Modeling and Simulation 1(3), pp349-370, 2003.
[25] J Portilla, V Strela, M Wainwright, E P Simoncelli. *Image Denoising using Scale Mixtures of Gaussians in the Wavelet Domain*. IEEE Transactions on Image Processing. vol 12, no. 11, pp. 1338-1351.
[26] L. Rudin, S. Osher and E. Fatemi, *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D, Vol 60(1992), pp. 259-268.
[27] G. Strang and T. Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA, 1996.
[28] Y. Sun, P. Wu, G.W. Wei and G. Wang. *Evolution-Operator-Based Single-Step Method for Image Processing*. International Journal of Biomedical Imaging, Vol. 2006, pp 1-28.
[29] Y. Wang and H.-M. Zhou, Total Variation Wavelet Based Medical Image Denoising, *Int. J. Biomedical Imaging*, Vol 2006 (2006), 1–6.
[30] J. Weickert, Coherence-enhancing diffusion of colour images, *IVC*, Vol 17, 1999, pp 201-212.
[31] T. Weissman, M. Weinberg, E. Ordentlich, G. Seroussi and S. Verdú, A discrete universal denoiser and its application to binary images, *Proc. IEEE ICIP*, 1 (2003), 117-120.
[32] L. P. Yaroslavsky, *Digital Picture Processing – An Introduction*, Springer Verlag, 1985.
[33] L. P. Yaroslavsky and M. Eden, *Fundamentals of Digital Optics*, Birkhauser, Boston, 1996.

Department of Mathematics, University of California at Los Angeles, Los Angeles, CA 90095-1555

E-mail address: tfchan@nsf.gov

Department of Mathematics, Michigan State University, East Lanisng, MI 48824, USA.

E-mail address: ywang@math.msu.edu

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.

E-mail address: hmzhou@math.gatech.edu