J Math Imaging Vis

DOI: 10.1007/s10851-007-0652-y

A Variational Approach to Reconstructing Images Corrupted by Poisson Noise

TRIET LE
Department of Mathematics, Yale University, P.O. Box 208283, New Haven, CT 06520-8283
triet.le@yale.edu

RICK CHARTRAND*
Los Alamos National Laboratory, Theoretical Division, MS B284, Los Alamos, NM 87545
rickc@lanl.gov

THOMAS J. ASAKI*
Los Alamos National Laboratory, Computer and Computational Sciences Division, MS D413, Los Alamos, NM 87545
asaki@lanl.gov

Published online: 30 March 2007

Abstract. We propose a new variational model to denoise an image corrupted by Poisson noise. Like the ROF model described in [1] and [2], the new model uses total-variation regularization, which preserves edges. Unlike the ROF model, our model uses a data-fidelity term that is suitable for Poisson noise. The result is that the strength of the regularization is signal dependent, precisely like Poisson noise. Noise of varying scales will be removed by our model, while preserving low-contrast features in regions of low intensity.

Keywords: image reconstruction, image processing, image denoising, total variation, Poisson noise, radiography

1. Introduction

An important task of mathematical image processing is image denoising. The general idea is to regard a noisy image f as being obtained by corrupting a noiseless image u; given a model for the noise corruption, the desired image u is a solution of the corresponding inverse problem.

Many algorithms are in use for reconstructing u from f. Since the inverse problem is generally ill-posed, most denoising procedures employ some sort of regularization. A very successful algorithm is that of Rudin, Osher, and Fatemi [1], which uses total-variation regularization. The ROF model regards u as the solution to a variational problem, to minimize the functional

$$F(u) := \int_\Omega |\nabla u| + \frac{\lambda}{2}\int_\Omega |f-u|^2, \tag{1}$$

where Ω is the image domain and λ is a parameter to be chosen. The first term of (1) is a regularization term, the second a data-fidelity term. Minimizing $F(u)$ has the effect of diminishing variation in u, while keeping u close to the data f. The size of the parameter λ determines the relative importance of the two terms.

Like many denoising models, the ROF model is most appropriate for signal independent, additive Gaussian noise. See [3] for an explanation of this

*Funded by the Department of Energy under contract W-7405-ENG-36.

in the context of Bayesian statistics. However, many important data contain noise that is signal dependent, and obeys a Poisson distribution. A familiar example is that of radiography. The signal in a radiograph is determined by photon counting statistics and is often described as particle-limited, emphasizing the quantized and non-Gaussian nature of the signal. Removing noise of this type is a more difficult problem. Besbeas et al. [4] review and demonstrate wavelet shrinkage methods from the now classical method of Donoho [5] to Bayesian methods of Kolaczyk [6] and Timmermann and Novak [7]. These methods rely on the assumption that the underlying intensity function is accurately described by relatively few wavelet expansion coefficients. Kervrann and Trubuil [8] employ an adaptive windowing approach that assumes locally piecewise constant intensity of constant noise variance. The method also performs well at discontinuity preservation. Jonsson, Huang, and Chan [9] use total variation to regularize positron emission tomography in the presence of Poisson noise, and use a fidelity term similar to what we use below.

In this paper, we propose a variational, total-variation regularized denoising model along the lines of ROF, but modified for use with Poisson noise. We will see in the next section that the effect of this model is that of having a spatially varying regularization parameter. Vanzella, Pellegrino, and Torre [10] adopt a self-adapting parameter approach in the context of Mumford-Shah regularization. Wong and Guan [11] use a neural network approach in linear image filtering to learn the appropriate parameter values from a training set. Reeves [12] estimates the local parameter values from each iterate in a reconstruction process for use with Laplacian filtering, assuming locally constant noise variance. Wu, Wang, and Wang [13] estimate both the parameter values and the linear regularization operator. These methods all require separate computations be made to estimate parameter values. In our case, the spatially-varying parameter is a consequence of using the data fidelity term that matches the probabilistic noise model (see derivation below). The self-adaptation occurs automatically in the course of solving a single unconstrained minimization problem.

2. Description of the Proposed Model

In what follows, we assume that f is a given grayscale image defined on Ω, an bounded, open subset of $\mathbb{R}^2$, with Lipschitz boundary $\partial\Omega$. Usually, Ω is a rectangle in the plane. We assume f is bounded and positive. Where convenient below, we regard f as integer valued, but this will ultimately be unnecessary.

Recall the Poisson distribution with mean and standard deviation μ:

$$P_\mu(n) = \frac{e^{-\mu}\mu^n}{n!}, \quad n \geq 0. \tag{2}$$

Our discussion follows well-known lines for formulating variational problems using Bayes's Law. See [3] for an example with the ROF model.

We wish to determine the image u that is most likely given the observed image f. Bayes's Law says that

$$P(u \mid f) = \frac{P(f \mid u)P(u)}{P(f)}. \tag{3}$$

Thus, we wish to maximize $P(f|u)P(u)$. Assuming Poisson noise, for each $x \in \Omega$ we have

$$P(f(x)|u) = P_{u(x)}(f(x)) = \frac{e^{-u(x)}u(x)^{f(x)}}{f(x)!} \tag{4}$$

Now we assume that the region Ω has been pixellated, and that the values of f at the pixels $\{x_i\}$ are independent. Then

$$P(f|u) = \prod_i \frac{e^{-u(x_i)}u(x_i)^{f(x_i)}}{f(x_i)!}. \tag{5}$$

The total-variation regularization comes from our choice of prior distribution:

$$P(u) = \exp\left(-\beta \int_\Omega |\nabla u|\right), \tag{6}$$

where β is a regularization paramter.

Instead of maximizing $P(f|u)P(u)$, we minimize $-\log(P(f|u)P(u))$. The result is that we seek a minimizer of

$$\sum_i (u(x_i) - f(x_i)\log u(x_i)) + \beta \int_\Omega |\nabla u|. \tag{7}$$

We regard this as a discrete approximation of the functional

$$E(u) := \int_\Omega (u - f\log u) + \beta \int_\Omega |\nabla u|. \tag{8}$$

The functional E is defined on the set of $u \in BV(\Omega)$ such that $\log u \in L^1(\Omega)$; in particular, u must be positive almost everywhere.

The Euler-Lagrange equation for minimizing $E(u)$ is

$$0 = \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) + \frac{1}{\beta u}(f - u), \quad \text{with } \frac{\partial u}{\partial \vec{n}} = 0 \text{ on } \partial\Omega. \tag{9}$$

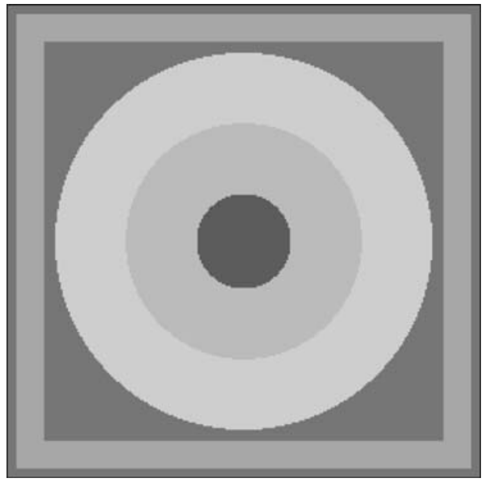

Figure 1. Circles image with frame. Image brightness has been adjusted for display, to allow the frame to be visible.

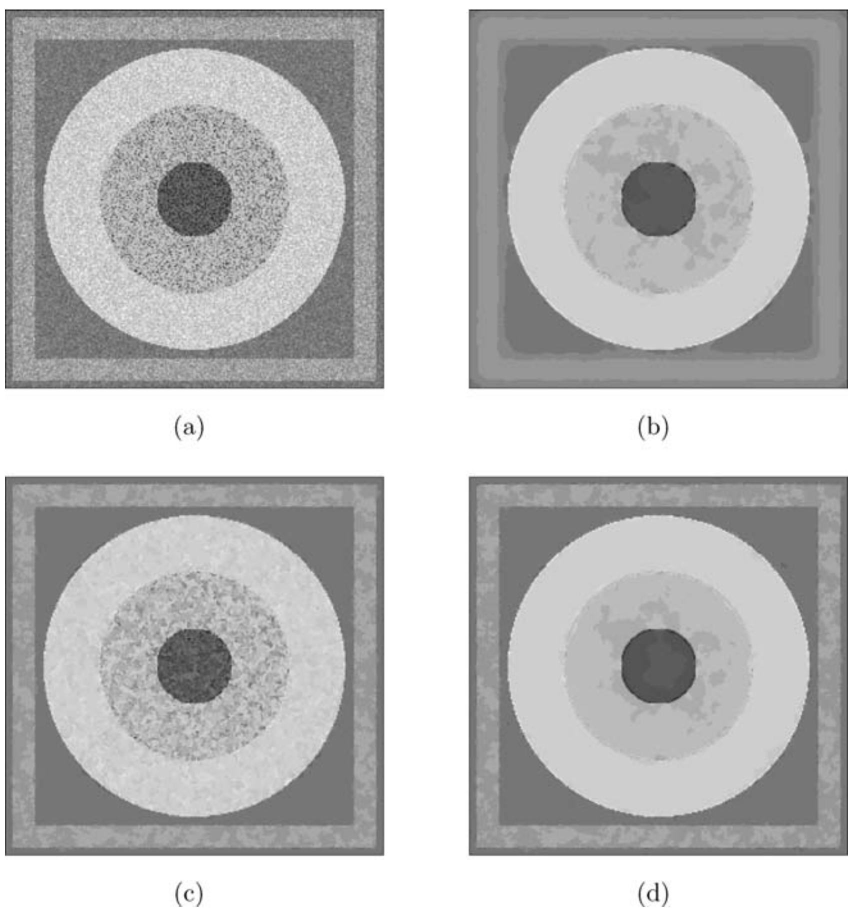

Figure 2. (a) Circles image with Poisson noise. (b) ROF denoised image. The frame is not well preserved. (c) ROF denoised image with decreased regularization strength. The frame is preserved, but the noise in the higher-intensity regions remains. (d) Image denoised with Poisson-modified total variation. Noise is removed at all scales, while preserving the frame.

Compare this with the Euler-Lagrange equation for minimizing the ROF functional (1),

$$0=\mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right)+\lambda(f-u), \quad \text{with } \frac{\partial u}{\partial \vec{n}}=0 \text{ on } \partial\Omega. \tag{10}$$

Notice that Eq. (9) is similar to Eq. (10), but with a variable $\lambda = \frac{1}{\beta u}$, which depends on the reconstructed image u. This local variation of the regularization parameter is better suited for Poisson noise because the expected noise increases with image intensity. Decreasing the value of the regularization parameter increases the denoising effect of the regularization term in the functional. We thus have a model that is similar to ROF but with a self-adjusting parameter.

3. Existence and Uniqueness

Next, we show existence and uniqueness of the minimizer for the model (8).

Theorem 1. *Let Ω be a bounded, open subset of $\mathbb{R}^2$ with Lipschitz boundary. Let f be a positive, bounded function. For $u \in BV(\Omega)$ such that $\log u \in L^1(\Omega)$, let $J(u) = \int_\Omega (u - f\log(u))$, $TV(u) = \int_\Omega |\nabla u|$, $E = \beta TV + J$. Then $E(u)$ has a unique minimizer.*

Proof: First, J is bounded below by $J(f)$, so E is bounded below. Thus we can choose a minimizing sequence $\{u_n\}$ for E. Then $TV(u_n)$ is bounded, as is $J(u_n)$. By Jensen's inequality,

$$J(u_n) \geq \|u_n\|_1 - \|f\|_\infty \log \|u_n\|_1, \tag{11}$$

so $\|u_n\|_1$ is bounded as well. This and the boundedness of $TV(u_n)$ mean that $\{u_n\}$ is a bounded sequence in the space $BV(\Omega)$. By the compactness of L^1 in BV [14, p. 176], there is $u \in BV$ such that a subsequence $\{u_{n_k}\}$ converges to u in L^1; without loss of generality, we may assume that $u_{n_k} \to u$ pointwise almost everywhere. By the lower semicontinuity of the BV norm [14, p. 172], $TV(u) \leq \liminf TV(u_{n_k})$. Since $u_{n_k} - f\log(u_{n_k})$ is bounded below (by $-\|f - f\log f\|_\infty$), we may use Fatou's Lemma to conclude that $J(u) \leq \liminf J(u_{n_k})$. Thus $E(u) \leq \liminf E(u_{n_k})$, and u minimizes E.

Clearly *TV* is a convex function. Since the logarithm is a strictly concave function and f is positive, J is strictly convex. Hence E is strictly convex. Therefore, the minimizer u is unique. □

4. Numerical Results

We use gradient descent to solve (9). We implement a straightforward, discretized version of the following PDE:

$$u_t = \mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right)+\frac{1}{\beta u}(f-u), \text{ with } \frac{\partial u}{\partial \vec{n}} = 0 \text{ on } \partial\Omega. \tag{12}$$

Derivatives are computed with standard centered-difference approximations. The quantity $|\nabla u|$ is replaced with $\sqrt{|\nabla u|^2 + \epsilon}$ for a small, positive ϵ. The time evolution is done with fixed timesteps, until the change in u is sufficiently small. A similar procedure is used to implement the ROF model (1), which we use for comparison with our proposed model.

The example in Fig. 1 consists of circles with intensities 70, 135, and 200, enclosed by a square frame of intensity 10, all on a background of intensity 5. Poisson noise is then added; see Fig. 2(a). Note that there is no parameter associated with Poisson noise, but the noise magnitude depends on the absolute image intensities. The amount of noise in a region of the image increases with the intensity of the image there.

In Figs. 2(b) and (c), the image has been denoised with the ROF (total variation) model (1). The result depends on the paramater λ. We choose λ according to the *discrepancy principle*, which says that the reconstruction should have a mean-squared difference from the noisy data that is equal to the variance of the noise. This is equivalent to the idea that of all the possible reconstructed images that are consistent with the noisy data, the image that should be chosen is the one that is most regular. In our example, we used a noise variance of the mean-squared difference between the noised image and the original image. (In cases where there is no original image available, the noise variance would have to be estimated.) The resulting λ of 0.04 gives the image in Fig. 2(b). The frame is almost completely washed out, as it differs from the background by less than the average noise standard deviation of 7.33. The frame can be preserved by increasing λ to 0.4, which has the effect of decreasing the strength of the regularization. The result, in Fig. 2(c), is that the noise in the higher-intensity regions of the image is not removed.

For our Poisson-modified total variation model, we chose the parameter β according to a suitably modified discrepancy principle: the value of the data fidelity term $\int u - f\log u$ for the reconstructed image should match that of the original image. In the example, this

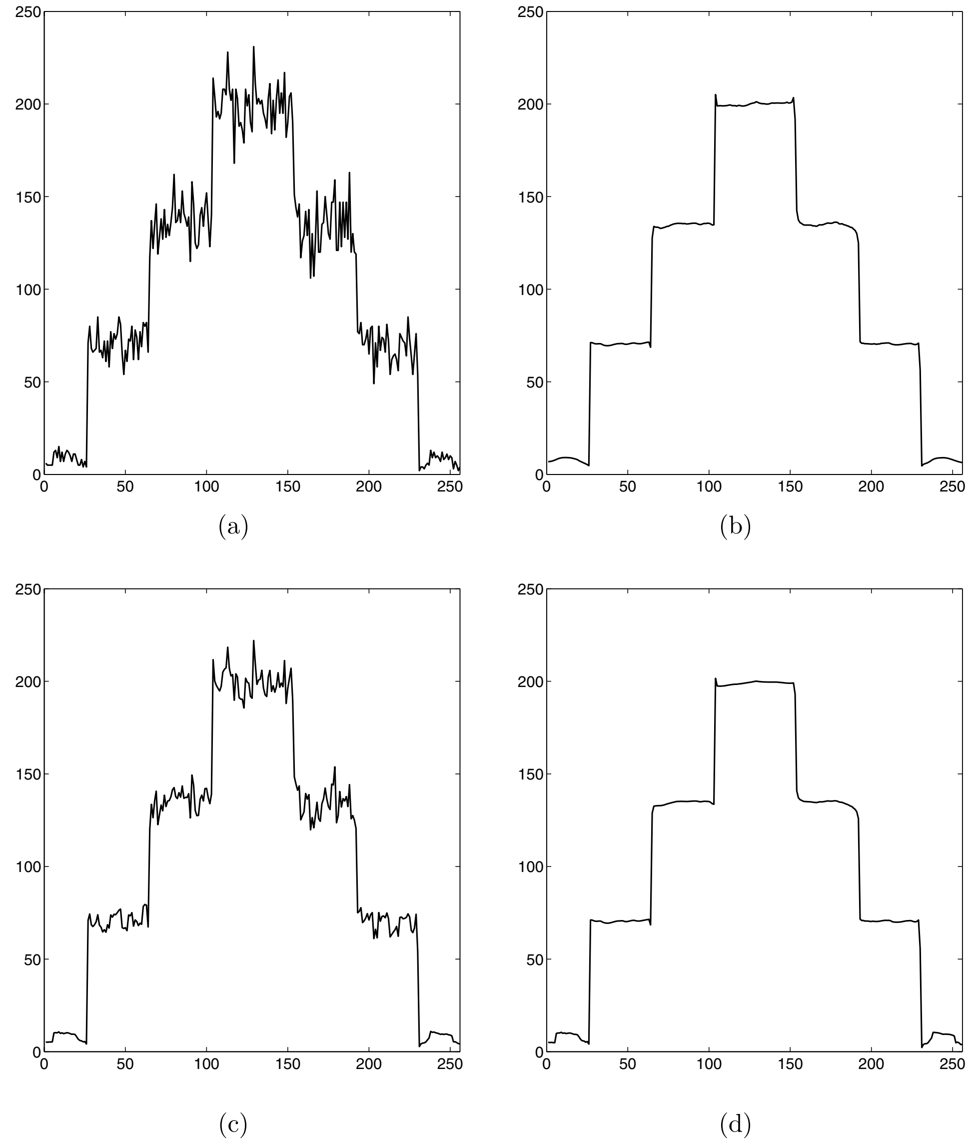

Figure 3. (a) Lineout of Poisson-noised circles image with frame. (b) Lineout of ROF denoised image. The frame is not well preserved. (c) Lineout of ROF denoised image with decreased regularization strength. The frame is preserved, but the noise in the higher-intensity regions remains. (d) Lineout of image denoised with Poisson-modified total variation. Noise is removed at all scales, while preserving the frame.

resulted in a β of 0.25. As noted above, the model behaves locally like ROF with a signal-dependent λ equal to $1/\beta u$. We thus have an effective λ of 0.8 for the background, 0.4 for the frame, and a smallest value of 0.02 in the center. Note that 0.4 was a value for λ for which the ROF model preserved the frame, while 0.02 gives a stronger regularization than that of ROF from the discrepancy principle. Therefore, it is not surprising that in Fig. 2(d), the frame is preserved as well as in Fig. 2(c), while the large-magnitude noise in the center is removed as well as in Fig. 2(b). Also see Fig. 3 for lineouts from the middle of the images, in which the qualitative properties of the results can be more clearly seen.

We can also compare our model with the ROF model by measuring the mean-squared difference between the reconstructed images and the original, noise-free image. This was 4.40 for our model, and 5.10 for the ROF model.

A second example in Fig. 4 uses an image of numbers of intensity 5, on regions of intensity 10, 85, 160, and 235. As in the previous example, the ROF model

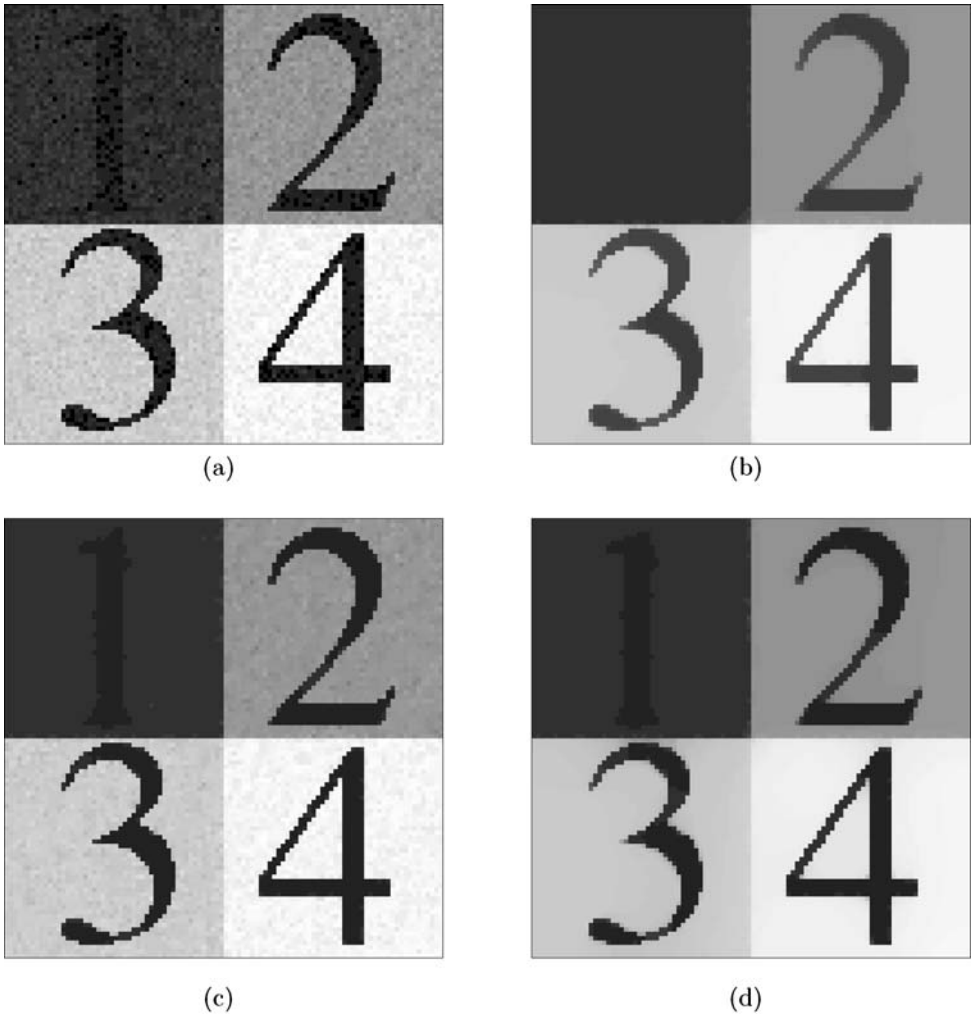

Figure 4. (a) Numbers on backgrounds of increasing intensity, corrupted by Poisson noise. (b) ROF denoised image. The '1' is obliterated. (c) Decreasing the regularization strength preserves the '1', but noise remains. (d) Our model removes noise at all scales and preserves features in low-intensity regions.

removes noise well, but eliminates low-intensity features (Fig. 4(b)). The lower noise level in the low intensity region allows the '1' to be preserved if the regularization strength is decreased (Fig. 4(c)), but then stronger noise in higher intensity regions remains. Our model removes noise at all scales; since the regularization strength self-adjusts in lower intensity regions, the '1' is preserved (Fig. 4(d)).

5. Conclusions

We have adapted the successful ROF model for total variation regularization to the case of images corrupted by Poisson noise. The gradient descent iteration for this model replaces the regularization parameter with a function. This results in a signal-dependent regularization strength, in a manner that exactly suits the signal-dependent nature of Poisson noise. From examples, we see that the resulting weaker regularization in low intensity regions of images allows for features in these regions to be preserved. If the image also contains higher intensity regions, the regularization will be stronger there and still remove the noise. This contrasts with the ROF model, whose uniform regularization strength must be chosen to either remove high intensity noise or retain low intensity features; both cannot be done.

Acknowledgment

The first author would like to thank Luminita Vese for many wonderful discussions and for introducing him to the paper [3].

References

1. L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, Vol. 60, pp. 259–268, 1992.
2. L. I. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *ICIP (1)*, pp. 31–35, 1994.
3. M. Green, "Statistics of images, the TV algorithm of Rudin-Osher-Fatemi for image denoising and an improved denoising algorithm," CAM Report 02-55, UCLA, October 2002.
4. P. Besbeas, I.D. Fies, and T. Sapatinas, "A comparative simulation study of wavelet shrinkage etimators for Poisson counts," *International Statistical Review*, Vol. 72, pp. 209–237, 2004.
5. D. Donoho, "Nonlinear wavelet methods for recovery of signals, densities and spectra from indirect and noisy data," in *Proceedings of Symposia in Applied Mathematics: Different Perspectives on Wavelets*, American Mathematical Society, 1993, pp. 173–205.
6. E. Kolaczyk, "Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds," *Statist. Sinica*, Vol. 9, pp. 119–135, 1999.
7. K. Timmermann and R. Novak, "Multiscale modeling and estimation of Poisson processes with applications to photon-limited imaging," *IEEE Trans. Inf. Theory*, Vol. 45, pp. 846–852, 1999.
8. C. Kervrann and A. Trubuil, "An adaptive window approach for poisson noise reduction and structure preserving in confocal microscopy," in *International Symposium on Biomedical Imaging (ISBI'04)*, Arlington, VA, April 2004.
9. E. Jonsson, C.-S. Huang, and T. Chan, "Total variation regularization in positron emission tomography," CAM Report 98-48, UCLA, November 1998.
10. W. Vanzella, F.A. Pellegrino, and V. Torre, "Self adaptive regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 26, pp. 804–809, 2004.
11. H. Wong and L. Guan, "Adaptive regularization in image restoration by unsupervised learning," *J. Electron. Imaging.*, Vol. 7, pp. 211–221, 1998.
12. S. Reeves, "Optimal space-varying regularization in iterative image restoration," *IEEE Trans. Image Process*, Vol. 3, pp. 319–324, 1994.
13. X. Wu, R. Wang, and C. Wang, "Regularized image restoration based on adaptively selecting parameter and operator," in *17th International Conference on Pattern Recognition (ICPR'04)*, Cambridge, UK, August 2004, pp. 602–605.
14. L.C. Evans and R.F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press: Boca Raton, 1992.

Triet M. Le received his Ph.D. in Mathematics from the University of California, Los Angeles, in 2006. He is now a Gibbs Assistant Professor in the Mathematics Department at Yale University. His research interests are in applied harmonic analysis and function spaces with application to image analysis and inverse problems.

Rick Chartrand received a Ph.D. in Mathematics from UC Berkeley in 1999, where he studied functional analysis. He now works as an applied mathematician at Los Alamos National Laboratory. His research interests are image and signal processing, inverse problems, and classification.

Tom Asaki is a staff member in the Computer and Computational Science Division at Los Alamos National Laboratory. He obtained his doctorate in physics from Washington State University. His interests are mixed-variable and direct-search optimization, applied inverse problems, and quantitative tomography.