Physica D 60 (1992) 259–268
North-Holland

PHYSICA D

Nonlinear total variation based noise removal algorithms*

Leonid I. Rudin[1], Stanley Osher and Emad Fatemi[2]
Cognitech Inc., 2800, 28th Street, Suite 101, Santa Monica, CA 90405, USA

A constrained optimization type of numerical algorithm for removing noise from images is presented. The total variation of the image is minimized subject to constraints involving the statistics of the noise. The constraints are imposed using Lagrange multipliers. The solution is obtained using the gradient-projection method. This amounts to solving a time dependent partial differential equation on a manifold determined by the constraints. As $t \to \infty$ the solution converges to a steady state which is the denoised image. The numerical algorithm is simple and relatively fast. The results appear to be state-of-the-art for very noisy images. The method is noninvasive, yielding sharp edges in the image. The technique could be interpreted as a first step of moving each level set of the image normal to itself with velocity equal to the curvature of the level set divided by the magnitude of the gradient of the image, and a second step which projects the image back onto the constraint set.

1. Introduction

The presence of noise in images is unavoidable. It may be introduced by the image formation process, image recording, image transmission, etc. These random distortions make it difficult to perform any required picture processing. For example, the feature oriented enhancement introduced in refs. [6,7] is very effective in restoring blurry images, but it can be "frozen" by an oscillatory noise component. Even a small amount of noise is harmful when high accuracy is required, e.g. as in subcell (subpixel) image analysis.

In practice, to estimate a true signal in noise, the most frequently used methods are based on the least squares criteria. The rationale comes from the statistical argument that the least squares estimation is the best over an entire ensemble of all possible pictures. This procedure is L_2 norm dependent. However it has been conjectured in ref. [6] that the proper norm for images is the total variation (TV) norm and not the L_2 norm. TV norms are essentially L_1 norms of derivatives, hence L_1 estimation procedures are more appropriate for the subject of image estimation (restoration). The space of functions of bounded total variation plays an important role when accurate estimation of discontinuities in solutions is required [6,7].

Historically, the L_1 estimation methods go back to Galileo (1632) and Laplace (1793). In comparison to the least square methods where closed form linear solutions are well understood and easily computed, the L_1 estimation is nonlinear and computationally complex. Recently the subject of L_1 estimation of statistical data has received renewed attention by the statistical community, see e.g. ref. [13].

Drawing on our previous experience with shock related image enhancement [6,7], we propose to denoise images by minimizing the total variation norm of the estimated solution. We derive a constrained minimization algorithm as a

* Research supported by DARPA SBIR Contract #DAAH01-89-C0768 and by AFOSR Contract #F49620-90-C-0011.

[1] E-mail: cogni!leonid@aerospace.aero.org.
[2] Current address: Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, MN 55455, USA.

time dependent nonlinear PDE, where the constraints are determined by the noise statistics.

Traditional methods attempt to reduce/remove the noise component prior to further image processing operations. This is the approach taken in this paper. However, the same TV/L_1 philosophy can be used to design hybrid algorithms combining denoising with other noise sensitive image processing tasks.

2. Nonlinear partial differential equations based denoising algorithms.

Let the observed intensity function $u_0(x, y)$ denote the pixel values of a noisy image for x, $y \in \Omega$. Let $u(x, y)$ denote the desired clean image, so

$$u_0(x, y) = u(x, y) + n(x, y)\,, \tag{2.1}$$

when n is the additive noise.

We, of course, wish to reconstruct u from u_0. Most conventional variational methods involve a least squares L^2 fit because this leads to linear equations. The first attempt along these lines was made by Phillips [1] and later refined by Twomey [2,3] in the one-dimensional case. In our two-dimensional continuous framework their constrained minimization problem is

$$\text{minimize} \int_\Omega (u_{xx} + u_{yy})^2 \tag{2.2a}$$

subject to constraints involving the mean

$$\int_\Omega u = \int_\Omega u_0 \tag{2.2b}$$

and standard deviation

$$\int_\Omega (u - u_0)^2 = \sigma^2\,. \tag{2.2c}$$

The resulting linear system is now easy to solve using modern numerical linear algebra. However, the results are again disappointing (but better than the MEM) with the same constraints) – see e.g. ref. [5].

The L_1 norm is usually avoided since the variation of expressions like $\int_\Omega |u|\,dx$ produces singular distributions as coefficients (e.g. δ functions) which cannot be handled in a purely algebraic framework. However, if L^2 and L^1 approximations are put side by side on a computer screen, it is clear that the L^1 approximation looks better than the "same" L^2 approximation. The "same" means subject to the same constraints. This may be at least partly psychological; however, it is well known in shock calculations that the L^1 norm of the gradient is the appropriate space. This is basically the space of functions of bounded total variation: BV. For free, we get the removal of spurious oscillations, while sharp signals are preserved in this space.

In ref. [6] the first author has introduced a novel image enhancement technique, called Shock Filter. It had analogy with shock wave calculations in computational fluid mechanics. The formation of discontinuities without oscillations and relevance of the TV norm was explored here.

In a paper written by the first two authors [7], the concept of total variation preserving enhancement was further developed. Finite difference schemes were developed there which were used to enhance mildly blurred images significantly while preserving the variation of the original image.

Additionally, in [8], Alvarez, Lions and Morel devised an interesting stable image restoration algorithm based on mean curvature motion, see also ref. [9]. The mean curvature is just the Euler–Lagrange derivative of the variation.

We therefore state that the space of BV functions is the proper class for many basic image processing tasks.

Thus, our constrained minimization problem is:

$$\text{minimize} \int_\Omega \sqrt{u_x^2 + u_y^2}\,dx\,dy \tag{2.3a}$$

subject to constraints involving u_0.

In our work so far we have taken the same two constraints as above:

$$\int_\Omega u \, dx \, dy = \int_\Omega u_0 \, dx \, dy . \tag{2.3b}$$

This constraint signifies the fact that the white noise $n(x, y)$ in (2.1) is of zero mean and

$$\int_\Omega \tfrac{1}{2}(u - u_0)^2 \, dx \, dy = \sigma^2 , \quad \text{where } \sigma > 0 \text{ is given} . \tag{2.3c}$$

The second constraint uses a priori information that the standard deviation of the noise $n(x, y)$ is σ.

Thus we have one linear and one nonlinear constraint. The method is totally general as regards number and shape of constraints.

We arrive at the Euler–Lagrange equations

$$0 = \frac{\partial}{\partial x}\left(\frac{u_x}{\sqrt{u_x^2 + u_y^2}}\right) + \frac{\partial}{\partial y}\left(\frac{u_y}{\sqrt{u_x^2 + u_y^2}}\right) - \lambda_1 - \lambda_2(u - u_0) \quad \text{in } \Omega, \text{ with} \tag{2.4a}$$

$$\frac{\partial u}{\partial n} = 0 \quad \text{on the boundary of } \Omega = \partial\Omega . \tag{2.4b}$$

The solution procedure uses a parabolic equation with time as an evolution parameter, or equivalently, the gradient descent method. This means that we solve

$$u_t = \frac{\partial}{\partial x}\left(\frac{u_x}{\sqrt{u_x^2 + u_y^2}}\right) + \frac{\partial}{\partial y}\left(\frac{u_y}{\sqrt{u_x^2 + u_y^2}}\right) - \lambda(u - u_0), \quad \text{for } t > 0, \; x, y \in \Omega , \tag{2.5a}$$

$$u(x, y, 0) \quad \text{given} , \tag{2.5b}$$

$$\frac{\partial u}{\partial n} = 0 \quad \text{on } \partial\Omega . \tag{2.5c}$$

Note, that we have dropped the first constraint (2.3b) because it is automatically enforced by our evolution procedure (2.5a–c) if the mean of $u(x, y, 0)$ is the same as that of $u_0(x, y)$.

As t increases, we approach a denoised version of our image.

We must compute $\lambda(t)$. We merely multiply (2.5a) by $(u - u_0)$ and integrate by parts over Ω. If steady state has been reached, the left side of (2.5a) vanishes. We then have

$$\lambda = -\frac{1}{2\sigma^2}\int \left[\sqrt{u_x^2 + u_y^2} - \left(\frac{(u_0)_x u_x}{\sqrt{u_x^2 + u_y^2}} + \frac{(u_0)_y u_y}{\sqrt{u_x^2 + u_y^2}}\right)\right] dx \, dy . \tag{2.6}$$

This gives us a dynamic value $\lambda(t)$, which appears to converge as $t \to \infty$. The theoretical justification for this approach comes from the fact that it is merely the gradient-projection method of Rosen [14].

We again remark that (2.5a) with $\lambda = 0$ and right part multiplied by $|\nabla u|$ was used in ref. [8] as a model for smoothing and edge detection. Following ref. [9] we note that this equation moves each level curve of u normal to itself with normal velocity equal to the curvature of the level surface divided by the magnitude of the gradient of u. Our additional constraints are needed to prevent distortion and to obtain a nontrivial steady state.

We remark that Geman and Reynolds, in a very interesting paper [10], proposed minimizing various nonlinear functionals of the form

$$\int_\Omega \varphi(\sqrt{u_x^2 + u_y^2}) \, dx \, dy$$

with constraints. Their optimization is based on simulated annealing, which is a computationally slow procedure used to find the global minimum. We, by contrast, seek a fast PDE solver that computes a "good" local minimum of the TV functional. There is reason to believe that the local extrema approach is more relevant to this image processing task.

Finally, we note that we originally introduced this method in two confidential contract reports [11,12].

The numerical method in two spatial dimensions is as follows. We let:

$$x_i = ih, \quad y_j = jh, \quad i, j = 0, 1, \ldots, N, \quad \text{with}$$
$$Nh = 1, \tag{2.7a}$$

$$t_n = n\,\Delta t, \quad n = 0, 1, \ldots, \tag{2.7b}$$

$$u_{ij}^n = u(x_i, y_j, t_n), \tag{2.7c}$$

$$u_{ij}^0 = u_0(ih, jh) + \sigma\varphi(ih, jh). \tag{2.7d}$$

The modified initial data are chosen so that the constraints are both satisfied initially, i.e. φ has mean zero and L^2 norm one.

The numerical approximation to (2.5), (2.6) is

$$u_{ij}^{n+1} = u_{ij}^n + \frac{\Delta t}{h}\left[\Delta_-^x\left(\frac{\Delta_+^x u_{ij}^n}{\left((\Delta_+^x u_{ij}^n)^2 + (m(\Delta_+^y u_{ij}^n, \Delta_-^y u_{ij}^n))^2\right)^{1/2}}\right) + \Delta_-^y\left(\frac{\Delta_+^y u_{ij}^n}{\left((\Delta_+^y u_{ij}^n + (m(\Delta_+^x u_{ij}^n, \Delta_-^x u_{ij}^n))^2\right)^{1/2}}\right)\right] - \Delta t\,\lambda^n(u_{ij}^n - u_0(ih, jh)), \tag{2.8a}$$

for $i, j = 1, \ldots, N$,

with boundary conditions

$$u_{0j}^n = u_{1j}^n, \quad u_{Nj}^n = u_{N-1,j}^n, \quad u_{i0}^n = u_{iN}^n = u_{i,N-1}^n. \tag{2.8b}$$

Here

$$\Delta_\mp^x u_{ij} = \mp(u_{i\mp1,j} - u_{ij}) \tag{2.9a}$$

and similarly for $\Delta_\mp^y u_{ij}$.

$$m(a, b) = \text{minmod}(a, b) = \left(\frac{\text{sgn } a + \text{sgn } b}{2}\right)\min(|a|, |b|) \tag{2.9b}$$

and λ^n is defined discreetly via

$$\lambda^n = -\frac{h}{2\sigma^2}\left[\sum_{i,j}\left(\sqrt{(\Delta_+^x u_{ij}^n)^2 + (\Delta_+^y u_{ij}^n)^2} - \frac{(\Delta_+^x u_{ij}^0)(\Delta_+^x u_{ij}^n)}{\sqrt{(\Delta_+^x u_{ij}^n)^2 + (\Delta_+^y u_{ij}^n)^2}} - \frac{(\Delta_+^y u_{ij}^0)(\Delta_+^y k u_{ij}^n)}{\sqrt{(\Delta_+^x u_{ij}^n)^2 + (\Delta_+^y u_{ij}^n)^2}}\right)\right]. \tag{2.9c}$$

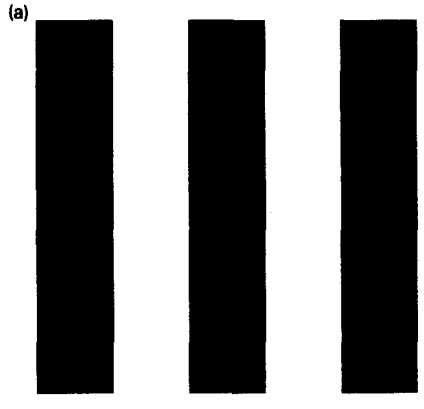
A step size restriction is imposed for stability:

$$\frac{\Delta t}{h^2} \leqslant c. \tag{2.9d}$$

3. Results

We have run our two-dimensional denoising algorithm on graphs and real images.

The graphs and images displayed take on in-



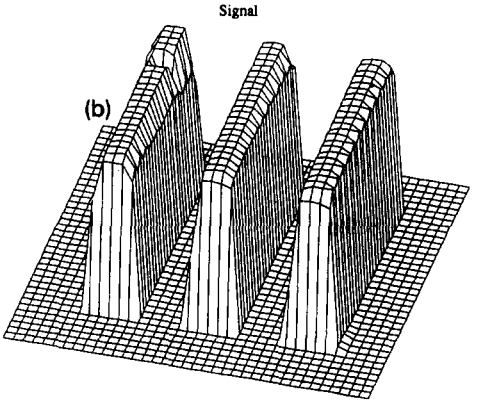


Fig. 1. (a) "Bars". (b) Plot of (a). (c) Plot of noisy "bars", SNR = 1.0. (d) Noisy "bars", SNR = 1.0. (e) Plot of the reconstruction from (d). (e) TV reconstruction from (d). (g) Plot of the reconstruction error.

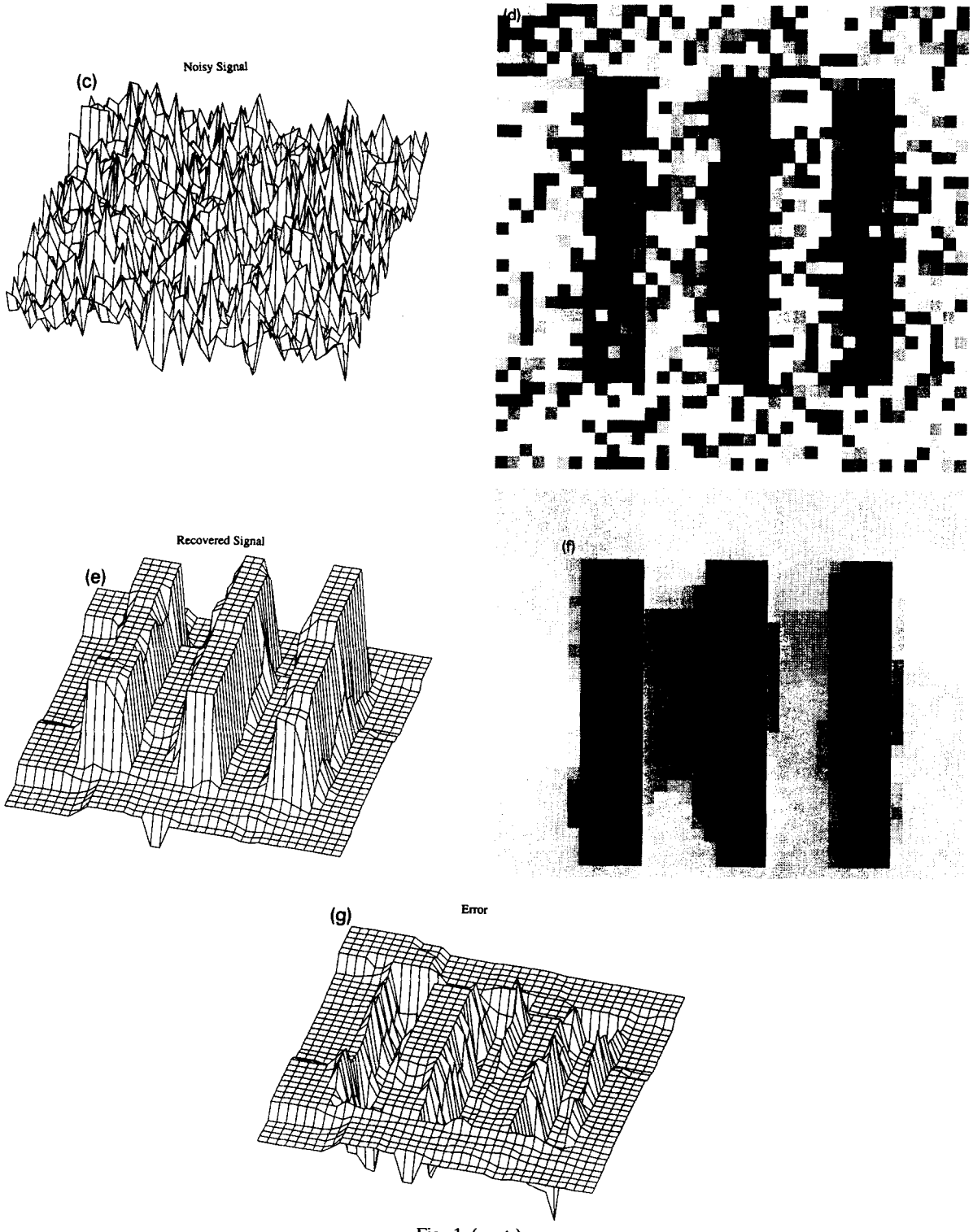


Fig. 1 (cont.).

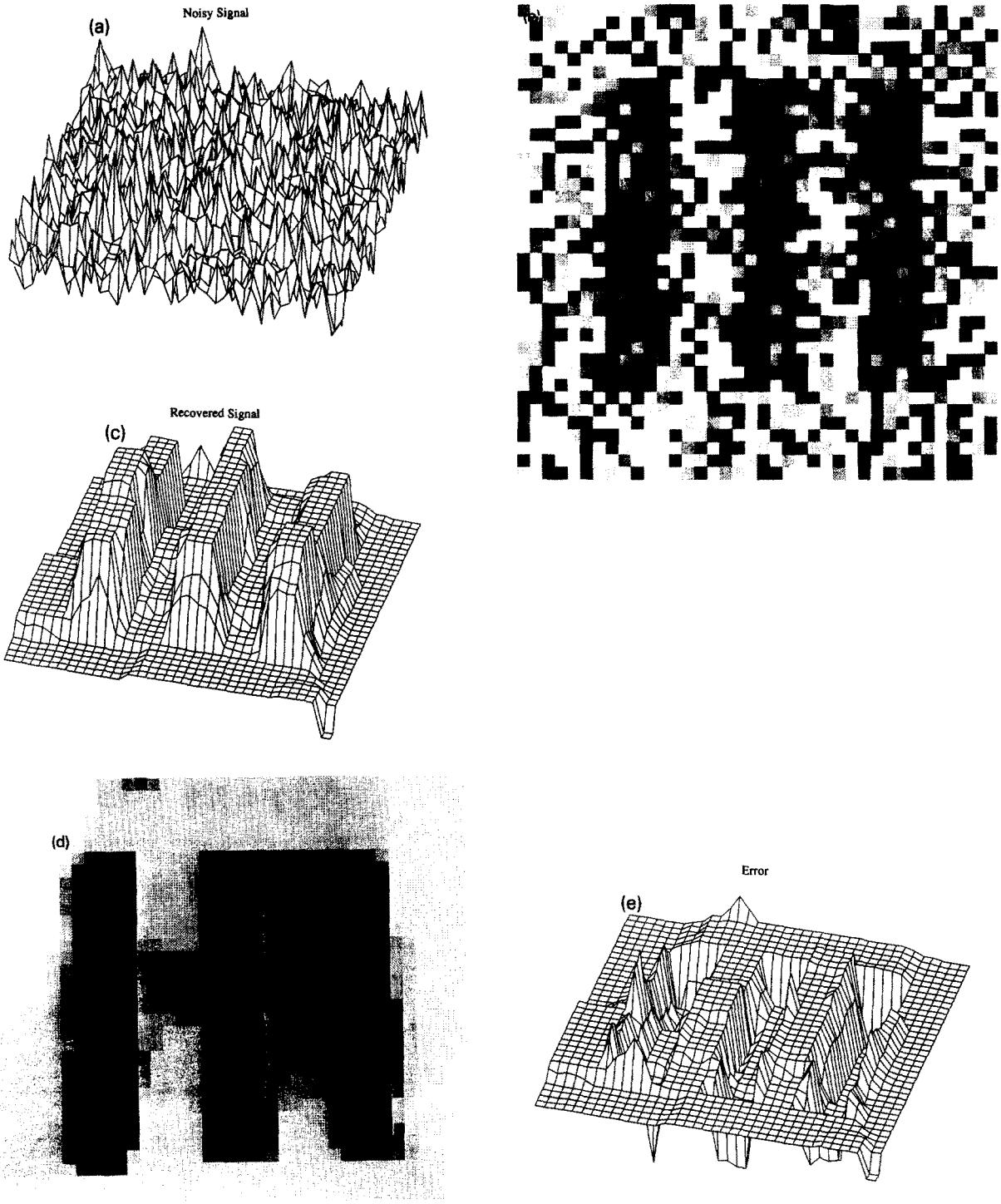


Fig. 2. (a) Plot of fig. 1a plus noise, SNR = 0.5. (b) Noisy fig. 1a, SNR = 0.5. (c) Plot of the reconstruction from (b). TV reconstruction from (b). (e) Plot of the reconstruction error.

teger values from 0 to 255. When Gaussian white noise is added the resulting values generally lie outside this range. *For display purposes only* we threshold; however, the processing takes place on a function whose values generally lie arbitrarily far outside the original range.

Signal to noise ratio (SNR) is defined by:

$$\mathrm{SNR} = \frac{\Sigma_\Omega (u_{ij} - \bar{u})^2}{\Sigma_\Omega (n_{ij})^2}, \tag{3.1}$$

where $\bar{u}$ is the mean of the signal u_{ij} and n_{ij} is the noise.

Figs. 1 and 2 concern three parallel steps taken

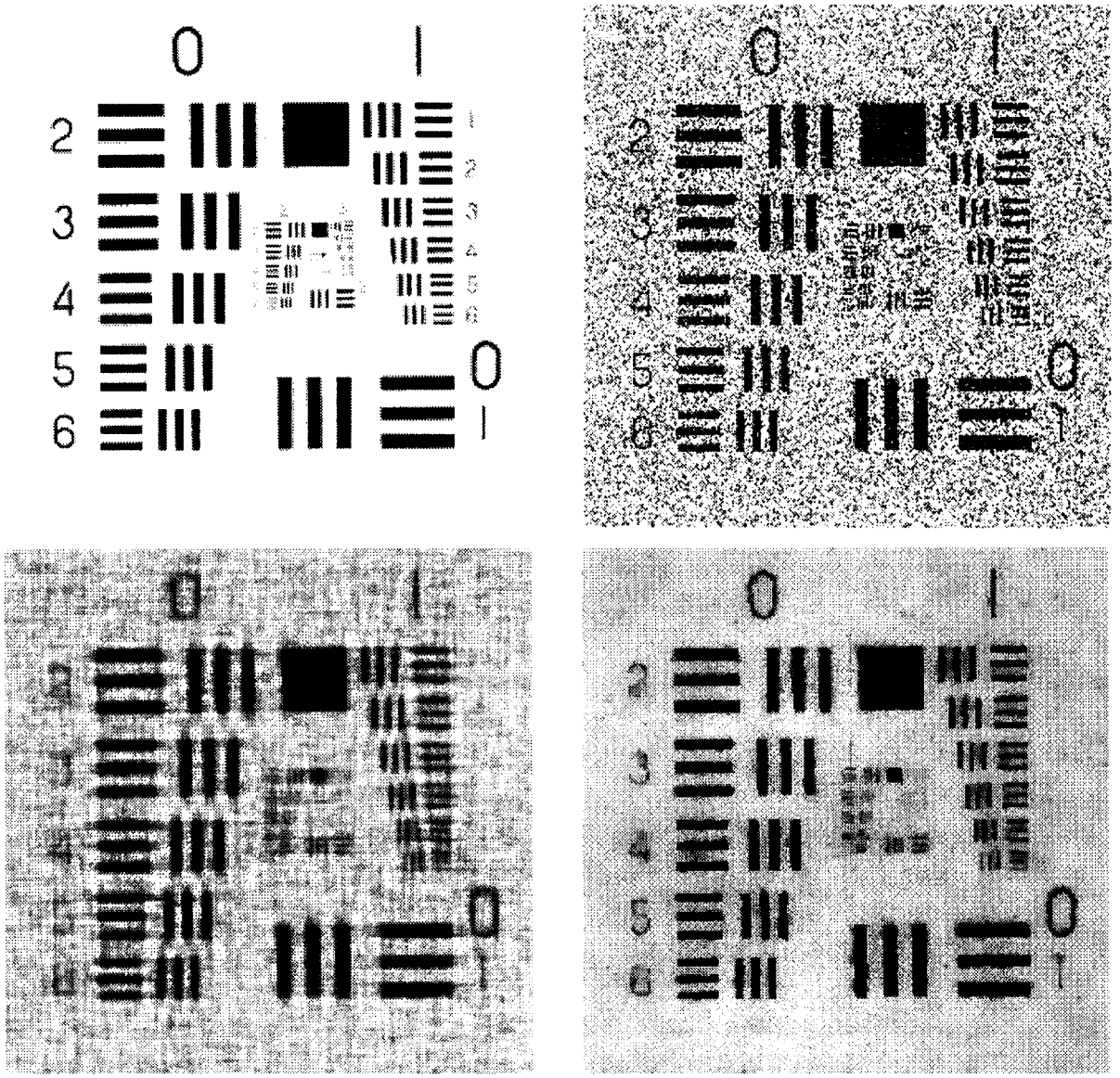


Fig. 3. (a) "Resolution Chart". (b) Noisy "Resolution Chart", SNR = 1.0. (c) Wiener filter reconstruction from (b). (d) TV reconstruction from (b).

from fig. 3a. This is 38 by 38 pixels wide 256 gray levels black and white original image. Fig. 1a shows the original signal. Fig. 1b shows its intensity plot. Fig. 1c shows the intensity of the noisy signal with additive Gaussian white noise, signal to noise ratio SNR 1. Fig. 1d shows the noisy signal. Fig. 1e shows a graph of the recovered sharp signal and fig. 1f shows the recovered signal. Finally, fig. 1g shows the error which is fairly "hollow". It is zero both within the original steps and also beyond a few pixels outside of them. Fig. 2a shows the intensity plot of a noisy signal when SNR = $\frac{1}{2}$, twice as much Gaussian white noise as signal. Fig. 2b shows the noisy

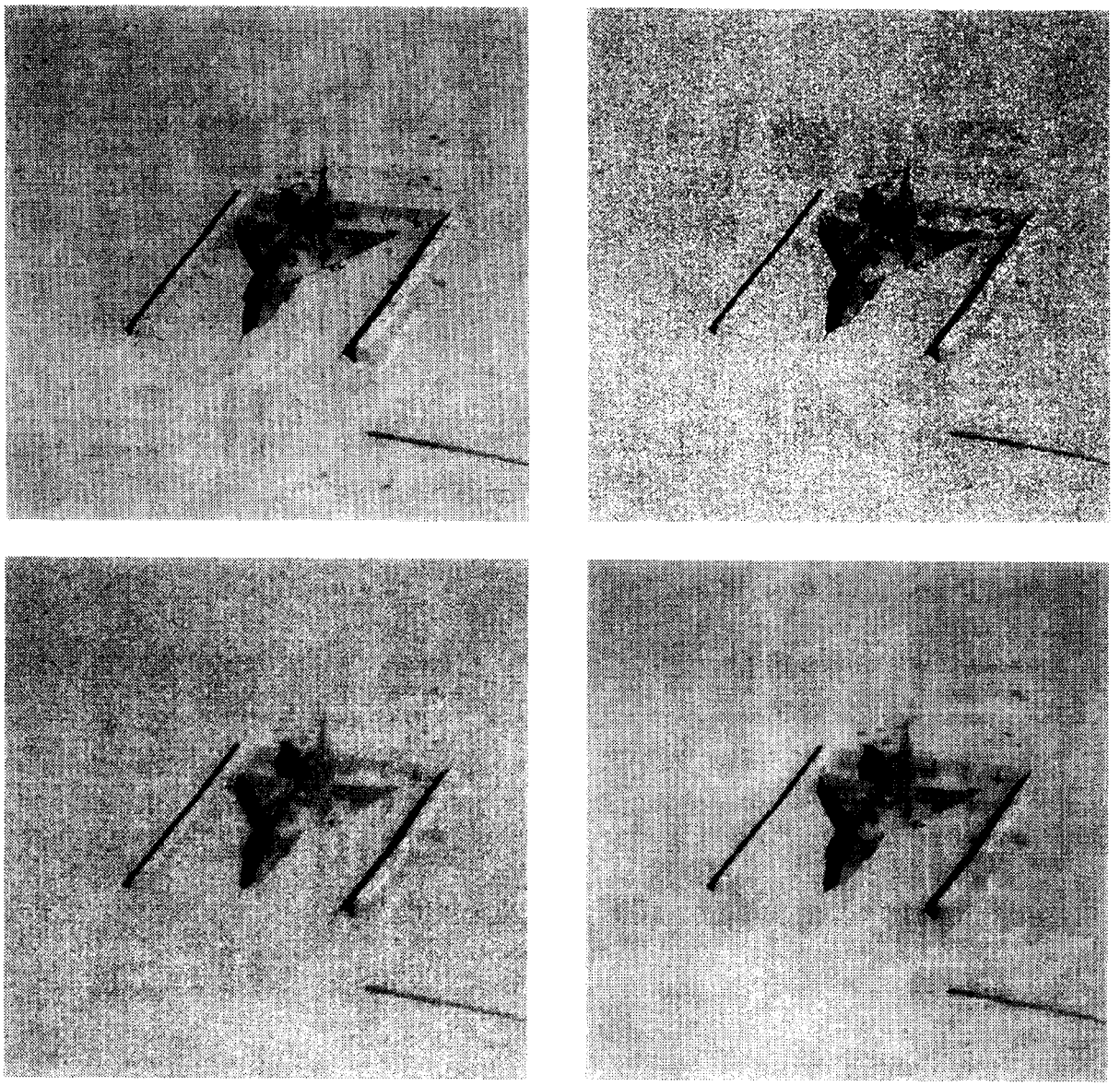


Fig. 4. (a) "Airplane". (b) Noisy "Airplane", SNR = 1.0. (c) Wiener filter reconstruction from (b). (d) TV reconstruction from (b).

image. Fig. 2c shows the intensity plot of the recovered signal and fig. 2d shows the recovered image. Finally, fig. 2e shows the almost "hollow" error.

It appears that our denoising procedure beats the capability of the human eye – see figs. 1b, 2b and 2c.

The remaining figures are 256 gray level standard black and white images taken from the USC IPI image data base. Fig. 3a shows the original 256 × 256 pixels resolution chart. Fig. 3b shows the result of adding Gaussian white noise, SNR 1. Fig. 3c shows the result of our denoising algorithm. Finally fig. 3d shows the result of using a Weiner filter where the power spectrum was estimated from fig. 3b. Notice fig. 3d has a

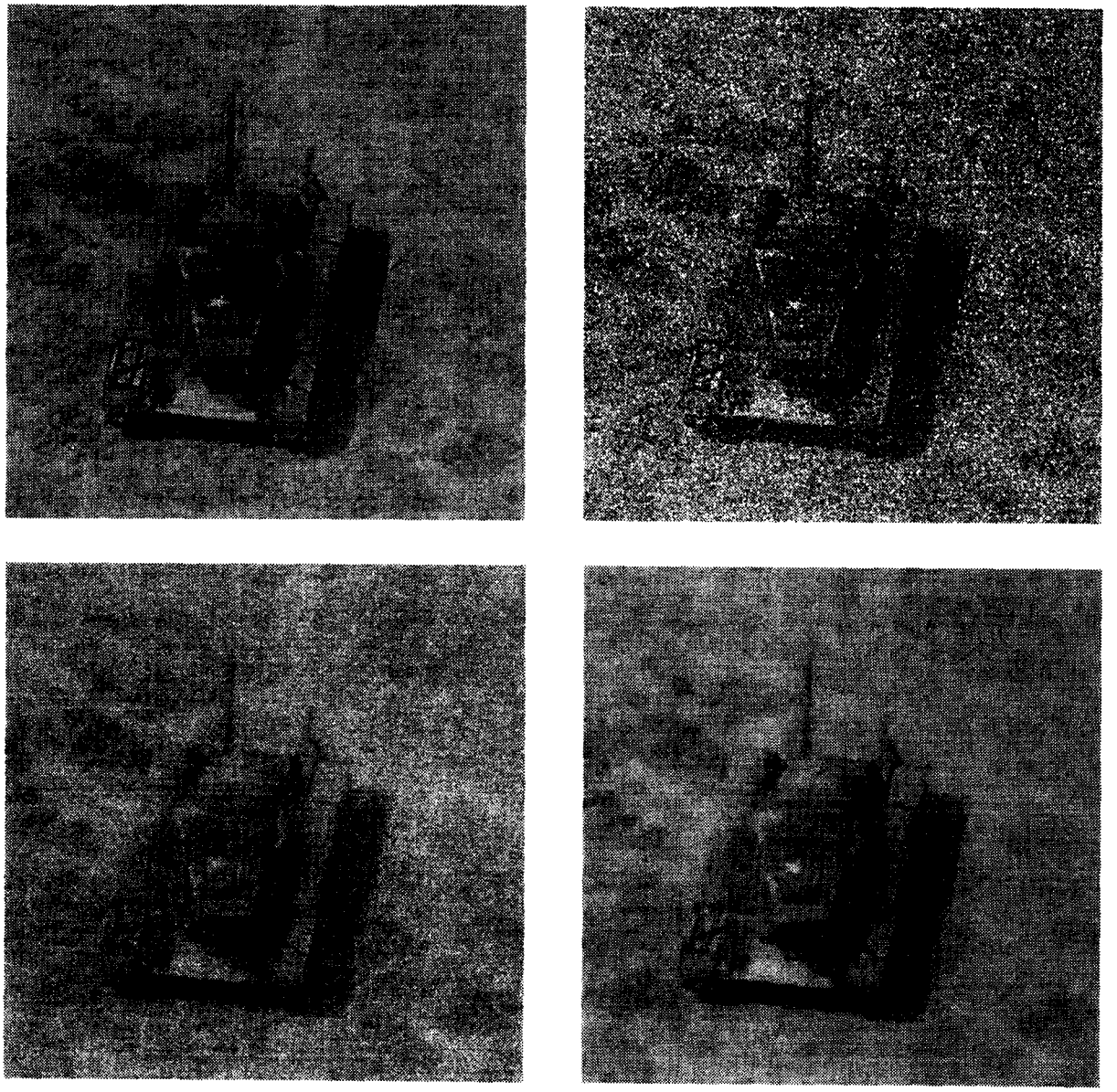


Fig. 5. (a) "Tank". (b) Noisy "Tank", SNR = 1.0. (c) Wiener filter reconstruction from (b). (d) TV reconstruction from (b).

lot of background noise which makes it problematic for automatic processing. Fig. 4a shows a 256×256 airplane in the desert (clean image). Fig. 4b shows the result of adding Gaussian white noise – SNR 1. Fig. 4c shows the result of a denoising via our algorithm. Fig. 4d shows the result of a Weiner filter denoising with the true spectrum estimated from the noisy image, via a moving average. Fig. 5a shows the original 512×512 picture of a tank. Fig. 5b shows the result of adding Gaussian white noise SNR 4. Fig. 5c shows a Weiner filter denoising with spectrum estimates from fig. 5b. Fig. 5d shows our algorithm applied to the same window. Notice that the discontinuities are much clearer in the last case. Also Wiener restoration has oscillatory artifacts.

Our recent experiments indicate that the use of more constraints (information about the noise and the image) in this method will yield more details of the solution in our denoising procedure.

References

[1] B.R. Frieden, Restoring with maximum likelihood and maximum entropy, J. Opt. Soc. Am. 62 (1972) 511.

[2] D.L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, J. ACM 9 (1962) 84.

[3] S. Twomey, On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature, J. ACM 10 (1963) 97.

[4] S. Twomey, The application of numerical filtering to the solution of integral equations encountered in indirect sensing measurements, J. Franklin Inst. 297 (1965) 95.

[5] B.R. Hunt, The application of constrained least squares estimation to image restoration by digital computer, IEEE Trans. Comput. 22 (1973) 805.

[6] L. Rudin, Images, numerical analysis of singularities and shock filters, Caltech, C.S. Dept. Report #TR: 5250:87 (1987).

[7] S. Osher and L.I. Rudin, Feature oriented image enhancement using shock filters, SIAM J. Num. Anal. 27 (1990) 919.

[8] L. Alvarez, P.L. Lions and J.M. Morel, Image selective smoothing and edge detection by nonlinear diffusion, SIAM J. Num. Anal. 29 (1992) 845.

[9] S. Osher and J. Sethian, Fronts propagating with curvature dependent speed: Algorithms based on a Hamilton–Jacobi formulation, J. Comput. Phys. 79 (1985) 12.

[10] D. Geman and G. Reynolds, Constrained restoration and the recovery of discontinuities, preprint (1990).

[11] E. Fatemi, S. Osher and L.I. Rudin, Removing noise without excessive blurring, Cognitech Report #5, (12/89), delivered to DARPA US Army Missile Command under contract #DAAH01-89-C-0768.

[12] L.I. Rudin and S. Osher, Reconstruction and enhancement of signals using non-linear non-oscillatory variational methods, Cognitech Report #7 (3/90), delivered to DARPA US Army Missile Command under contract #DAAH01-89-C-0768.

[13] Y. Dodge, Statistical data analysis based on the L_1 norm and related methods (North-Holland, Amsterdam, 1987).

[14] J.G. Rosen, The gradient projection method for nonlinear programming, Part II, nonlinear constraints, J. Soc. Indust. Appl. Math. 9 (1961) 514.