Signal Processing 20 (1990) 3-14
Elsevier

IMPROVED RESTORATION OF NOISY IMAGES BY ADAPTIVE LEAST-SQUARES POST-FILTERING

Michael UNSER*

Biomedical Engineering and Instrumentation Branch, National Institutes of Health, Bethesda, MD 20892, U.S.A.

Received 10 July 1989
Revised 30 October 1989

Abstract. This paper describes a class of post-filtering algorithms that adaptively compute a linear combination between a noisy image and a restored version of it obtained by linear filtering. A set of optimal weighting coefficients is derived by minimizing the quadratic error between the output of this system and a noise-free signal. The only a priori information that is required is the noise variance. The local application of this optimization principle leads to the definition of the constrained and non-contrained adaptive least-squares filters (ACLSF and ALSF, respectively). These algorithms, and particularly the ACLSF, can be implemented in an extremely efficient way by using a fast recursive updating strategy. We then consider the particular case of a moving average as the initial filter and compare this application of the ALSF and ACLSF with Lee's adaptive noise filtering algorithm [Lee, 1980]. We also present some simulations and experimental examples illustrating the capability of these algorithms to reduce noise efficiently while preserving image details.

Zusammenfassung. Eine Klasse von Nachfilterungs-Algorithmen wird beschrieben, die adaptiv eine Linearkombination eines verrauschten Bildes und einer durch linear Filterung erhaltenen restaurierten Version berechnet. Ein Satz optimaler Gewichtskoeffizienten wird gewonnen, indem der quadratische Fehler zwischen dem Ausgangssignal dieses Systems und dem rauschfreien Bild minimiert wird. Als einzige Vorinformation wird die Varianz des Rauschens benötigt. Die lokale Anwendung dieses Optimierungsprinzips führt auf die Definition adaptiver LS-Filter mit und ohne Nebenbedingungen (ACLSF, bzw. ALSF). Diese Algorithmen, insbesondere der Typ ACLSF, können auf äußerst effiziente Weise mit Hilfe schneller rekursiver Nachführstrategien realisiert werden. Dann betrachten wir den Sonderfall der gleitenden Mittelung als urprüngliches Filter und vergleichen ALSF und ACLSF in dieser Anwendung mit Lee's adaptivem Rauschfilterungs-Algorithmus [Lee, 1980]. Wir stellen auch einige Simulationen und Versuchsbeispiele dar, um zu illustieren, wie diese Algorithmen Rauschen wirkungsvoll unterdrücken und gleichzeitig Bilddetails erhalten können.

Résumé. Cette correspondance décrit une classe d'algorithmes de post-traitement évaluant de façon adaptive une combinaison linéaire entre une image bruitée et une version restorée de celle-ci. Les coefficients de pondération sont obtenus en minimisant l'erreur quadratique entre la sortie du système et le signal non-bruité. La variance du bruit est la seule information a priori requise. L'application locale de ce principe d'optimisation donne lieu à la définition de filtres adaptifs aux moindres carrés avec ou sans contraintes (ACLSF et ALSF). Ces filtres, en particulier ACLSF, peuvent être réalisés de façon très efficace par remise à jour itérative. Le cas particulier d'un filtrage initial par moyennage local est considéré et les versions correspondantes des ACLSF et ALSF sont comparées avec l'algorithme de filtrage adaptatif de Lee [Lee, 1980]. Des exemples simulés et expérimentaux illustrent la capacité de ces algorithmes de réduire le bruit tout en préservant les détails dans une image.

Keywords. Adaptive noise smoothing, noise filtering, image restoration, least squares estimation.

1. Introduction

Diminishing noise is a fundamental task in image processing. This problem has been studied by many researchers and numerous noise filtering techniques are now available [12, 13]. These approaches can be broadly classified into optimal (or model driven) and heuristic methods. Two-dimensional extensions of the well-known Wiener and Kalman filtering techniques fall into the first

* Permanent address: INSERM, Unité U-138, Hôpital Henri-Mondor, 94010 Créteil, France.

0165-1684/90/$3.50 1990, Elsevier Science Publishers B.V.

category. Early methods concentrated on non-recursive techniques implemented in the Fourier domain [13], while more recent work has centered on two-dimensional recursive implementations of the Kalman filter [17, 18]. Their effectiveness depends heavily on the validity of a particular image model.

The second category includes methods performing nonlinear (mostly adaptive) smoothing in the spatial domain. Typical examples are the various generalizations of median and rank order filtering [2, 4, 11], the K-nearest neighbor averaging [3], gradient inverse weighted smoothing [16], sigma filtering [10] and Lee's adaptive algorithm for additive noise [8, 9]. They have become increasingly popular, mainly because of their ease of implementation, computational efficiency, conceptual simplicity, and their ability to diminish the effect of noise while preserving sharp edges. Among these approaches, Lee's adaptive filter appears to be particularly efficient [12]. It computes a linear weighted sum of the noisy image itself and the output of a moving average filter, which provides an estimate of the local mean. Recently, Kuan et al. [7] have shown that this procedure approximates a minimum mean square error filter when the signal satisfies a particular form of non-stationary mean Gaussian image model initially proposed by Hunt and Cannon [5] and is corrupted by additive white Gaussian noise. Although these theoretical results are noteworthy, we remark that such an image model goes against our intuitive perception of noise in that it considers an uncorrelated Gaussian random component as an intrinsic part of the signal itself. These authors also make the assumption that the signal is locally ergodic and stationary, which allows them to use local image statistics to estimate unknown ensemble statistics.

In this paper, we extend the idea underlying Lee's algorithm to that of computing a signal estimate from a weighted sum of the observed (noisy) signal and the output of an arbitrary initial restoration filter; that is, after the image has been transformed in a way judged to be appropriate to the class of images being processed. This procedure, which is illustrated in Fig. 1, is called postfiltering and is aimed at improving the image quality obtained in the preliminary step. It is clear that the merit of this technique is heavily dependent on the performance of the initial restoration filter, the choice of which is important but is not considered further in this paper. What we want to emphasize is that this system is capable of correcting some of the initial filter's shortcomings, for example, excessive edge smoothing. An important feature of the system shown in Fig. 1 is that the weighting coefficients are optimized locally and that the algorithm has the potential of achieving better performance than conventional space invariant filtering techniques (e.g. the Wiener filter).

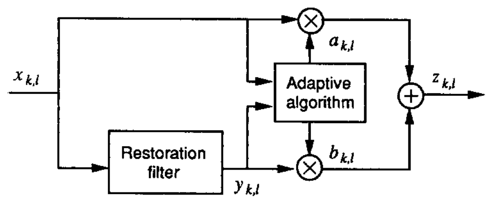

Fig. 1. General structure of an adaptive post-filtering algorithm.

In contrast with previous work, the present approach makes no assumptions about the signal but looks at a particular filter structure. This point of view naturally leads to the derivation of weighting coefficients that are optimal in the least squares sense. We show that this solution can be implemented efficiently based on a knowledge of the noise statistics alone. An interesting result is that Lee's adaptive filter is very similar to a particular version of our algorithm, but is sub-optimal because of differences in the evaluation of local statistics.

2. The least squares regression equations

Our method is applicable to the restoration of a signal, $\{\mu_{i,j}\}$, corrupted by zero mean additive stationary noise $\{n_{i,j}\}$:

$$x_{i,j} = \mu_{i,j} + n_{i,j} \tag{1}$$

with known variance $\sigma^2 = \mathrm{E}\{n_{i,j}^2\}$. As with most restoration techniques, the signal and noise components are assumed to be mutually decorrelated. The observed noisy signal, $\{x_{i,j}\}$, is initially restored by means of a suitably chosen operator, which for convenience, is assumed to be a linear space invariant filter:

$$y_{i,j} = x_{i,j} * h_{i,j} = \mu'_{i,j} + n'_{i,j}, \tag{2}$$

whose impulse response $\{h_{i,j}\}$ has a spatial sum that is normalized to unity to provide an unbiased signal estimate whenever $\mu_{i,j} = \text{const}$. The knowledge of $\{h_{i,j}\}$ and the noise second order moments (autocorrelation function) allows the determination of the residual noise correlation coefficient which is defined as

$$\rho = \mathrm{E}\{n_{i,j} n'_{i,j}\} / \sigma^2, \tag{3}$$

where $n'_{k,l}$ denotes the filtered noise component. For additive white noise, it is straightforward to show that $\rho = h_{0,0}$.

Our goal is to attempt, in a post-processing phase, to provide a better signal estimate, $z_{i,j}$, by constructing a weighted sum of the original noisy signal and its filtered version:

$$z_{i,j} = a_{i,j} x_{i,j} + b_{i,j} y_{i,j} \tag{4}$$

as shown in Fig. 1. The motivation of this operation is the following. The filter is usually chosen to improve the signal-to-noise ratio in most parts of the image. However, in areas of heavy signal activity such as edges or texture, filtering may tend to degrade the signal more than it actually reduces noise and one would be better off keeping the original noisy measurement. In this respect, (4) has the potential of making the best use of both worlds. In the particular case where the initial restoration filter is chosen to be the Wiener filter, we obtain an algorithm that is similar to the signal equivalent approach of Abramatic and Silverman [1]. The essential difference is that these authors determine the weighting coefficients using an empirical masking function whereas the present approach is based on a well-defined least squares principle.

Although the coefficients a and b are allowed to vary as a function of the spatial index (i, j), we first consider the case where these quantities are constant over a given region R in an image, and determine the values that provide the best approximation of $\mu_{k,l}$ in the least square sense. This is achieved by minimizing the quadratic error:

$$\varepsilon^2 = \frac{1}{N_R} \sum_{(k,l) \in R} (z_{k,l} - \mu_{k,l})^2, \tag{5}$$

where N_R is the total number of pixels in R. These basic results are then used in Section 3 to derive the structure of the adaptive filter. We note that these equations can be readily extended to less restrictive conditions (initial non-linear restoration and non-stationary noise) provided that the noise variance and residual correlation coefficient are known over the entire image.

2.1. Basic solution

By substituting (4) in (5) and setting the partial derivatives with respect to a and b to zero, we find that the optimal coefficients are given by the following matrix equation:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix}^{-1} \begin{bmatrix} S_{x\mu} \\ S_{y\mu} \end{bmatrix}. \tag{6}$$

The notation S_{uv} is used to designate a normalized sum of squares defined as

$$S_{uv} = \frac{1}{N_R} \sum_{(k,l) \in R} u_{k,l} v_{k,l}, \tag{7}$$

where u and v stand for either x, y or μ. Of course, $\mu_{k,l}$ is not known so that $S_{x\mu}$ and $S_{y\mu}$ cannot be computed from (7). Instead, we will make use of our a priori knowledge of the noise statistics to compute estimates of these quantities. This point is discussed in Section 3.

2.2. Constrained solution

A useful constraint is to require that

$$a + b = 1, \tag{8}$$

which essentially guarantees that $\mathrm{E}\{z_{k,l}\} = \mu_{k,l}$

whenever $E\{y_{k,l}\} = \mu_{k,l}$. The minimization of the quadratic approximation error subject to this no-bias condition leads to the slightly modified system of equations:

$$\begin{bmatrix} a \\ b \\ \lambda \end{bmatrix} = \begin{bmatrix} S_{xx} & S_{xy} & 1 \\ S_{xy} & S_{yy} & 1 \\ 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} S_{x\mu} \\ S_{y\mu} \\ 1 \end{bmatrix}, \tag{9}$$

where λ is a dummy Lagrange multiplier.

A good stability criterion for our system would be to require that $z_{k,l}$ is always between $x_{k,l}$ and $y_{k,l}$, independently of these values. It can be shown that this requirement is equivalent to (8) plus the additional constraint: $0 \leq a \leq 1$.

3. Fast adaptive implementation

In our adaptive filter implementation, R is a $(2r+1)\times(2s+1)$ estimation window centered on the current pixel index (i,j). The optimal coefficients $a_{i,j}$ and $b_{i,j}$ are computed from (6) (or (9)) for every position of the estimation window and the output of the filter is evaluated from (4). Accordingly, (7) can be rewritten as

$$S_{uv}(i,j) = \frac{1}{(2r+1)(2s+1)} \times \sum_{k=-r}^{r} \sum_{l=-s}^{s} u_{i+k,j+l} v_{i+k,j+l}. \tag{10}$$

The $S_{uv}(i,j)$ are computed by applying a moving average filter to the image product: $\{u_{i,j}v_{i,j}\}$. The running average may be implemented efficiently with 4 operations per pixel by using an updating algorithm that is successively applied to the rows and columns of the image [15].

Using (1) and (2), $S_{x\mu}(i,j)$ and $S_{y\mu}(i,j)$ can be decomposed as

$$\begin{aligned} S_{x\mu}(i,j) &= S_{xx}(i,j) - S_{xn}(i,j) \\ &= S_{xx}(i,j) - S_{nn}(i,j) - S_{\mu n}(i,j), \end{aligned} \tag{11}$$

$$\begin{aligned} S_{y\mu}(i,j) &= S_{xy}(i,j) - S_{yn}(i,j) \\ &= S_{xy}(i,j) - S_{\mu' n}(i,j) - S_{n' n}(i,j). \end{aligned} \tag{12}$$

These quantities are estimated by replacing the unknown terms of the right-hand side of (11) and (12) by their expected values:

$$\begin{aligned} \hat{S}_{x\mu}(i,j) &= S_{xx}(i,j) - E\{n^2\} \\ &= S_{xx}(i,j) - \sigma^2, \end{aligned} \tag{13}$$

$$\begin{aligned} \hat{S}_{y\mu}(i,j) &= S_{xy}(i,j) - E\{\mu' n\} - E\{nn'\} \\ &= S_{xy}(i,j) - \rho\sigma^2. \end{aligned} \tag{14}$$

Note that the condition that the noise and signal are decorrelated is necessary so that $E\{\mu n\} = 0$ and $E\{\mu' n\} = 0$. Finally, by replacing these expressions in (6) and evaluating the matrix inverse, we find that the regression coefficients in the non-constrained case are given by

$$a_{i,j} = \frac{S_{xx}(i,j)S_{yy}(i,j) - S_{xy}^2(i,j) - \sigma^2 S_{yy}(i,j) + \rho\sigma^2 S_{xy}}{S_{xx}(i,j)S_{yy}(i,j) - S_{xy}^2(i,j)}, \tag{15}$$

$$b_{i,j} = \frac{\sigma^2 S_{xy}(i,j) - \rho\sigma^2 S_{xx}}{S_{xx}(i,j)S_{yy}(i,j) - S_{xy}^2(i,j)}. \tag{16}$$

These equations defined the so-called adaptive least squares filter (ALSF). In the constrained case, these quantities take the much simpler form:

$$a_{i,j} = 1 - \frac{(1-\rho)\sigma^2}{P(i,j)}, \tag{17}$$

$$b_{i,j} = 1 - a_{i,j}, \tag{18}$$

where $P(i,j)$ is given by

$$\begin{aligned} P(i,j) &= S_{xx}(i,j) + S_{yy}(i,j) - 2S_{xy}(i,j) \\ &= S_{(x-y)(x-y)}(i,j) > 0, \end{aligned} \tag{19}$$

which is a local estimate of the variance of the difference between the noisy and filtered signal which we will refer to as the residue. This leads to the particularly simple implementation of the adaptive constrained least square filter (ACLSF) shown in Fig. 2. Some insight into the way this filter operates is gained by noticing that $(1-\rho)\sigma^2$ is the variance of the residue when filtering does not modify the signal component, that is, when $\mu_{k,l} = \mu'_{k,l}$ and the residue variation is due to noise

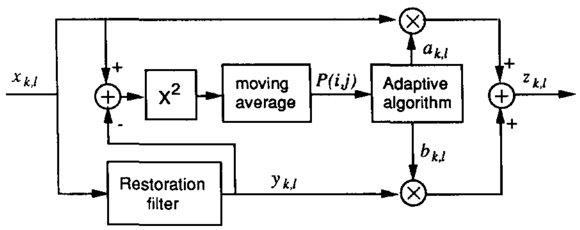

Fig. 2. Efficient implementation of the adaptive constrained least squares filter (ACLSF).

alone. Therefore, whenever the residue energy is small and close to this reference value, our adaptive scheme will give a predominant weight to the filtered signal. Conversely, when the residue energy is greater than this threshold the weight is progressively shifted to the unfiltered signal with the extreme being no filtering at all. This is consistent with the fact that an abnormally large value of $P(i,j)$ is an indication that filtering tends to degrade the signal. Since there is no justification for a value of $a_{i,j}$ lower than zero other than errors due to statistical fluctuations, it is legitimate to clip this quantity to zero which slightly improves the robustness of the algorithm.

4. Case study: improving upon the moving average filter

In this section, we consider the particular case where the initial restoration filter is a simple $(2m+1)\times(2n+1)$ moving average filter:

$$y_{i,j}=\frac{1}{(2m+1)(2n+1)}\sum_{k=-m}^{m}\sum_{l=-n}^{n}x_{i+k,j+l}, \tag{20}$$

which is characterized by the residual noise correlation coefficient $\rho=[(2m+1)(2n+1)]^{-1}$.

4.1. Constrained versus unconstrained algorithm

From the previous section, it is quite apparent that the ACLSF is much simpler to implement than the ALSF. For the former, the coefficients of the adaptive filter are entirely determined from the value of $P(i,j)$ which is associated to a $(2r+1)\times(2s+1)$ window centered on the current spatial index:

$$P(i,j;\, r,s)=\frac{1}{(2r+1)(2s+1)}\times\sum_{k=-r}^{r}\sum_{l=-s}^{s}(x_{i+k,j+l}-y_{i+k,j+l})^2. \tag{21}$$

$P(i,j;\, r,s)$ is computed by averaging the squared residual signal that can be stored in an auxiliary bidimensional array. The unconstrained algorithm requires the use of at least three such auxiliary arrays for the evaluation of $S_{xx}(i,j)$, $S_{yy}(i,j)$ and $S_{xy}(i,j)$. The question is whether or not the use of this latter scheme, which results in a threefold increase in computational complexity, could have some practical benefits.

To address this issue, we consider the simplified case of the filtering of a constant signal $\{\mu_{i,j}=\mu\}$, corrupted by additive white noise with variance σ^2. We also assume, for simplicity, that the extent of the estimation window approaches infinity so that the sums of squares can be replaced by their expected values:

$$S_{xx}=\mu^2+\sigma^2, \tag{22}$$

$$S_{yy}=\mu^2+\frac{\sigma^2}{(2m+1)(2n+1)}=\mu^2+\rho\sigma^2, \tag{23}$$

$$S_{xy}=\mu^2+\rho\sigma^2, \tag{24}$$

$$P=S_{xx}+S_{yy}-2S_{xy}=(1-\rho)\sigma^2. \tag{25}$$

For the unconstrained case, substitution of these quantities in (15) and (16) yields

$$a_{\mathrm{u}}=0,\qquad b_{\mathrm{u}}=\frac{\mu^2}{\mu^2+\rho\sigma^2}, \tag{26}$$

and the corresponding minimum mean square error is given by

$$\min\{\varepsilon^2\}=(\rho\sigma^2)\left(\frac{\mu^2}{\mu^2+\rho\sigma^2}\right)=\rho\sigma^2 b_{\mathrm{u}}. \tag{27}$$

Similarly, for the constrained case, we find that

$$a_c = 0, \qquad b_c = 1 \quad \text{and} \quad \min\{\varepsilon^2\} = \rho\sigma^2. \tag{28}$$

In both cases, the estimation of the signal is entirely based on the output of the averaging filter, which is a reasonable result. The difference between both approaches is only significant when the signal energy (μ^2) is small and of the same order of magnitude as the variance of the residual noise component ($\rho\sigma^2$). As expected, the estimation error is somewhat smaller for the ALSF. What is less satisfactory is that this improvement of performance is achieved at the cost of a biased signal estimate. On the other hand, the ACLSF is non-biased. Furthermore, its performance does not depend on the magnitude of the signal component, which is a desirable property in image processing.

4.2. Comparison with Lee's adaptive filter

When $\{y_{i,j}\}$ is defined by (20), the ACLSF is very similar to the filter described by Lee [8], with the difference that for the latter, the coefficient that is applied to the observed signal is determined from the following equation:

$$a'_{i,j} = 1 - \frac{\sigma^2}{Q(i,j)}, \tag{29}$$

where

$$Q(i,j) = \frac{1}{(2m+1)(2n+1)} \times \sum_{k=-m}^{m} \sum_{l=-n}^{n} (x_{i+k,j+l} - y_{i,j})^2$$
$$= \left(\frac{1}{(2m+1)(2n+1)} \times \sum_{k=-m}^{m} \sum_{l=-n}^{n} (x_{i+k,j+l})^2\right) - y_{i,j}^2. \tag{30}$$

The complementary coefficient $\{b'_{i,j}\}$ is also computed from (18). The similarity between (17) and (29), as well as between (21) and (30), calls for the following comments.

(i) In Lee's initial formulation, $y_{i,j}$ and $Q(i,j)$ are local statistics used to estimate the a priori mean ($\bar{x}_{i,j} = \mu_{i,j}$) and variance ($\mathrm{E}\{[x_{i,j} - \bar{x}_{i,j}]^2\}$) of $x_{i,j}$; the signal is assumed to be locally stationary so that these ensemble statistics can be approximated by spatial averages. As a consequence, there is no distinction between the averaging kernels in (20) and (30) and the author only considers the case where $m = r$ and $n = s$. In our formulation, $\{y_{i,j}\}$ can be an arbitrary filter and there is no conceptual connection between the extent of its impulse response and the $(2r+1) \times (2s+1)$ estimation window over which the least squares estimates of $\{a_{i,j}\}$ and $\{b_{i,j}\}$ are computed. In fact, it is r and s (and not m and n) that determine the adaptability of our algorithm.

(ii) When $m = r$ and $n = s$, $Q(i,j)$ and $P(i,j; m, n)$ are equivalent only when $y_{i,j}$ is constant over the estimation window. This is generally not the case and $Q(i,j)$ will probably be greater than $P(i,j; m, n)$. Kuan et al. [7] suggest modifying Lee's algorithm by replacing $Q(i,j)$ by a modified variance estimate equivalent to $P(i,j; m, n)$, which is in much closer agreement with (17). They argue that the use of this particular estimate gets rid of the constraint of a locally stationary mean and they present some evidence of improved performance. We remark, however, that this argument is contradicted by their use of (20) to estimate $\bar{x}_{i,j}$.

(iii) The additional difference between (29) (or the modified algorithm described by Kuan et al.) and (17), is the presence of $(1-\rho)$ in the numerator. As remarked earlier, $(1-\rho)\sigma^2$ is the variance of $(x_{i,j} - y_{i,j})$ as well as the expected value of $P(i,j; r, s)$, when filtering does not distort the signal component. Therefore, comparing $P(i,j; r, s)$ against σ^2 instead of $(1-\rho)\sigma^2$ tends to bias the filter towards using a value of $a_{i,j}$ that is slightly below the one prescribed by the constrained least squares solution. For instance, when considering the simplified case studied in Section 4.1, neglecting ρ in (17) leads to the weighting coefficients $a' = -\rho/(1-\rho)$ and $b' = 1/(1-\rho)$, and results in a somewhat increased estimation error: $\varepsilon^2 = \rho\sigma^2/(1-\rho)^2$. We note, however, that the corresponding loss of performance is significant only when ρ is large, that is, when the spatial extent of the filter is small (e.g. 3×3).

4.3. Experimental examples

To support the observations made in the two preceding sections, we first consider simulation experiments allowing an objective quantitative performance assessment. The original mandrill image is shown in Fig. 3a. This image was degraded with various levels of additive independent Gaussian noise and initially processed using both 3×3 and 5×5 moving average filters. The degraded image with $\sigma^2=225$ and its filtered version obtained using a 5×5 averaging kernel are shown in Fig. 3b,c. We then applied the ALSF and ACLSF using various estimation window sizes, as well as Lee's adaptive filter. The results of this processing are summarized in Table 1 in terms of the quadratic signal-to-noise ratio (QSNR) computed using the original image (Fig. 3a) as reference. This performance criterion is defined as

$$\mathrm{QSNR}=\frac{\sum_{k=1}^{K}\sum_{l=1}^{L}(\mu_{k,l}-\mu)^2}{\sum_{k=1}^{K}\sum_{l=1}^{L}(y_{k,l}-\mu_{k,l})^2}, \tag{31}$$

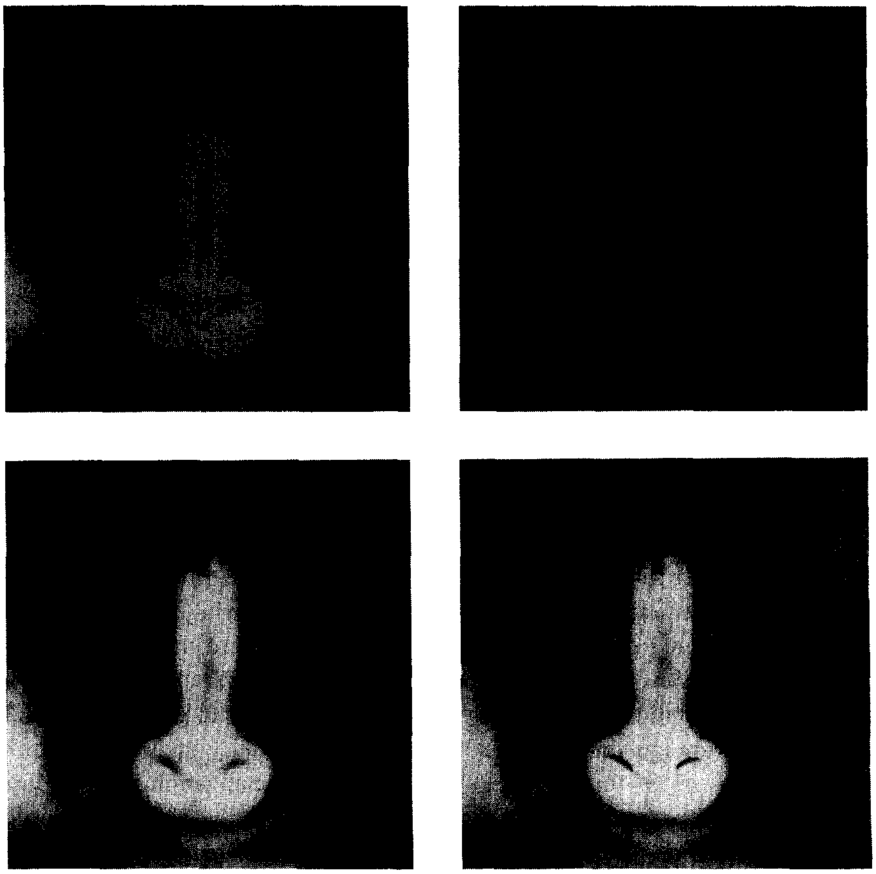

Fig. 3. Adaptive noise filtering example. (a) Original 256×256 image; (b) degraded image using white Gaussian noise with $\sigma^2=225$, QSNR = 12.36 (10.9 dB); (c) initial filtering using a 5×5 moving average, QSNR = 9.17 (9.62 dB); (d) ACLSF with input images (b) and (c) using a 7×7 estimation window, QSNR = 25.00 (13.9 dB).

Table 1

Quadratic signal-to-noise ratio for various post-filtering algorithms using different parameter settings

	$m=n=3$ $\sigma^2=100$	$m=n=5$ $\sigma^2=100$	$m=n=3$ $\sigma^2=225$	$m=n=5$ $\sigma^2=225$
initial	27.79	27.79	12.36	12.36
mov. ave	12.93	9.32	12.13	9.177
Lee	40.44	39.49	23.04	23.08
$r=s=3$				
ACLSF	43.61 (43.34)[a]	41.26	24.09 (24.15)	23.11
ALSF	42.43 (41.95)	41.49	22.90 (22.88)	22.97
$r=s=7$				
ACLSF	45.22 (45.01)	43.44	25.77 (25.55)	24.92
ALSF	45.70 (45.33)	43.71	25.87 (25.56)	25.01
$r=s=11$				
ACLSF	45.39 (44.74)	43.08	25.98 (25.55)	25.00
ALSF	45.54 (45.20)	43.34	26.00 (25.72)	25.07
$r=s=17$				
ACLSF	44.84 (44.15)	42.35	25.77 (25.32)	24.61
ALSF	44.95 (44.68)	42.50	25.79 (25.53)	24.68
$r=s=65$				
ACLSF	42.20 (41.90)	39.36	24.48 (24.14)	22.89
ALSF	42.28 (42.15)	39.39	24.52 (24.31)	22.93

[a] The values in parentheses correspond to the approximation $\rho=0$.

where μ is the average gray level value of the noise free image: $\{\mu_{k,l}(k=1,\ldots,K;\,l=1,\ldots,L)\}$. All adaptive algorithms provide a substantial increase of the QSNR. This is quite remarkable considering the fact that the initial moving average filter performs very poorly at the global level and tends to degrade the subjective image quality more than it actually reduces the noise. We also notice that the various versions of the LS algorithm always provide an improvement over Lee's initial scheme. The ALSF is slightly superior to the ACLSF, which is in accordance with our predictions, except for small window sizes where the use of a constrained solution makes the ACLSF somewhat less sensitive to errors due to random fluctuations. For $m=n=3$, there is a slight advantage in using $\rho=1/(2m+1)(2n+1)$ instead of $\rho=0$; the QSNR values corresponding to this latter case are given in parentheses. However, there is no noticeable difference for larger averaging kernels. It is also interesting to see that for a given noise variance there is an optimal estimation window size that does not seem to be related to the size of the averaging kernel ($m\times n$). For $\sigma^2=225$, the optimal window size is approximately 9×9 and the corresponding image obtained using the ACLSF with a 5×5 moving average filter is shown in Fig. 3d. The improvement over the moving average filter (Fig. 3b) is striking in the highly textured areas (hair) and for small image details such as the eyes. Similarly, a close inspection of uniform regions such as the nose reveals significant noise reduction when compared to the initial noisy input. Figure 4 displays a series of rectified residual noise images coresponding to the case $m=n=3$ with $\sigma^2=100$. It is quite apparent that the moving average filter (Fig. 4b) tends to globally degrade the signal more than it actually reduces the noise. The areas in which the restoration filter performs well (e.g. central region of the face) are clearly identified by the ACLSF which leads to appreciable noise reduction (Fig. 4d). The residual noise component for Lee's adaptive filter (Fig. 4c) has also been included. The qualitative comparison of Figs. 4c and 4d is

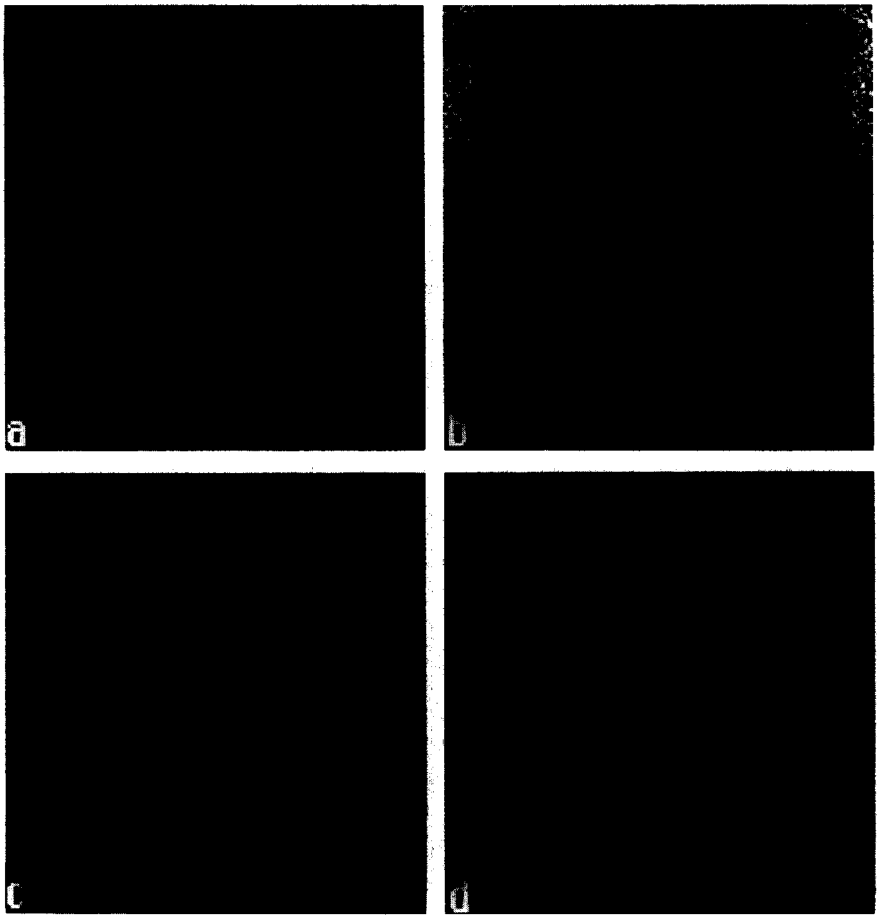

Fig. 4. Rectified residual noise components. (a) Initial noise component with $\sigma^2 = 100$; (b) residue of the noisy filtered image using a 3×3 moving average; (c) residue of the noisy filtered image using Lee's algorithm; (d) residue of the noisy filtered image using the ACLSF with a 7×7 estimation window.

clearly in favor of the ACLSF, although the performance improvement as measured by the QSNR is not dramatic.

Figure 5 shows an application of this post-filtering technique to the improvement of high resolution electron micrographs. The original image represents coated vesicles isolated from bovine brain [14]. It was obtained on the Brookhaven scanning transmission electron microscope (STEM). This micrograph is noisy due to the use of low-dose techniques designed to preserve insofar as possible the integrity of the specimen. In a preliminary stage, the noise variance was estimated from a small 32×32 area of uniform background. The electron micrograph was then filtered with a 5×5 moving average filter and processed by means of the ACLSF using a 7×7 estimation window. The quality of the restored image (Fig. 5c) is subjectively superior both to the original noisy observation and its smoothed version. The comparison between these images is made easier by looking at the enlarged 64×64 sub-regions in Fig. 5d. The noise has been substantially reduced in the background area where the output of the filter is close to the smoothed image. On the other hand, the sharpness of the specimens is preserved. This is due to the fact that for regions of high signal activity the output of the filter is closer to the noisy observation.

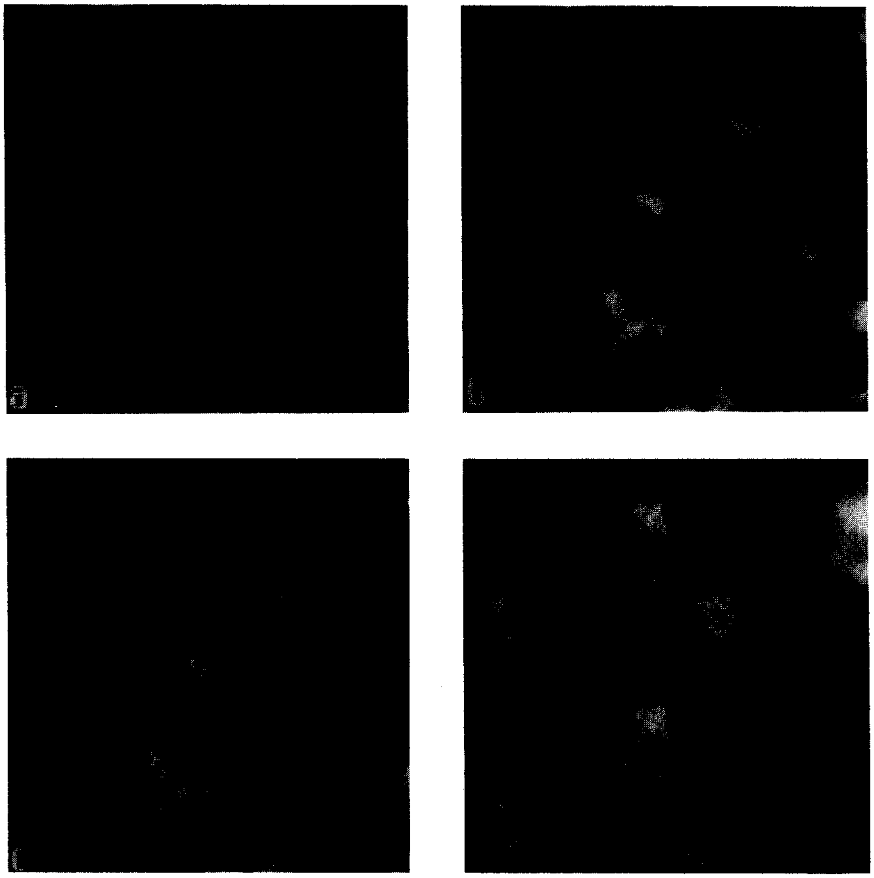

Fig. 5. Processing of a noisy electron micrograph. (a) Original 256 × 256 micrograph (initial variance estimate $\sigma^2 = 36$;) (b) initial filtering using a 5 × 5 moving average; (c) ACLSF with input images (a) and (b) using a 7 × 7 estimation window; (d) 200% enlargement of a 64 × 64 region from images (a), (b) and (c).

5. Discussion

The various adaptive post-filtering algorithms that have been considered have a simple structure and yet perform very efficiently, as illustrated by these examples. The most interesting property of this approach is that the final restored image has a signal-to-noise ratio that is greater than any of the inputs of the algorithm.

The computational requirement is minimal, particularly for the ACLSF, which in the present implementation requires no more than 8 operations per pixel. For a 256 × 256 image such as the ones shown in Figs. 3 and 5, the full restoration using a 3 × 3 moving average and the ACLS takes no more than 26 s on an Apple Macintosh II computer (7 s for the moving average and 17 s for the ACLSF). This time is independent of the size of both smoothing ($n \times m$) and estimation ($r \times s$) windows. The ALSF is not as fast and takes about 45 s of CPU.

Although the basic structure of the ALSF or ACLSF is very close to Lee's adaptive filter for additive white noise, the approach taken here is conceptually quite different. Instead of assuming a particular signal model as is the case in [7, 8], our starting point is the particular filter structure shown in Fig. 1 and the a priori knowledge of the noise statistics. We then derive the optimal filter coefficients that locally minimize the mean square error. This approach is less restrictive since it is applicable to all kinds of signals. Furthermore, there are no approximations involved such as the replacement of the a priori mean and signal variances by estimates provided by local statistics. The major benefits of this new formulation are the following.

(i) The initial filtering (typically, moving average) and the adaptive least squares estimation process are fully decoupled. This means that any restoration filter or procedure can be applied in a preliminary step and that there is no reason to limit ourselves to the use of a simple moving averaging filter. The only prerequisite is that the restoration filter performs well—meaning that it reduces noise more than it degrades the signal—in at least some regions of the image. Furthermore, the size of the estimation window is not necessarily equal to the size of the initial smoothing kernel and can be adjusted for optimal performance.

(ii) The ALSF and ACLSF compute a signal estimate that is locally optimal in the least squares sense. This explains the increase of performance over Lee's initial adaptive filter. A closer comparison between (17) and (29) shows that the ACLSF is essentially a refinement of this latter scheme, that is, it improves the already excellent performance of Lee's algorithm [12]. The derivations given in this paper also suggest using this type of approach with any type of noisy signal, regardless of the underlying model.

The size of the estimation window determines the adaptability of our post-filtering algorithms. From our experiments, it appears that there usually is an optimal window size whose performance is related to intrinsic image properties as well as to the level of noise. Using a small estimation window will improve the adaptability of the algorithm which is very good in non-stationary image regions (edges). However, it will also make the filter coefficients more sensitive to statistical errors and degrade performance in stationary image regions (uniform or texture). Consequently, there is a compromise to be found which depends on the noise statistics. We remark, however, that even the choice of the largest possible estimation window (the size of the picture itself) should not produce a result that is worse than either the initial noisy picture or its filtered version.

We have considered two different post-filters, namely the ALSF and ACLSF. The ALSF, which is unconstrained, usually performs somewhat better than the ACLSF but not enough to justify the added computational complexity. As shown in Section 4.1, it also has the disadvantage of providing a signal estimate slightly biased towards zero. Therefore, the ACLSF seems to be a better choice for most practical applications.

The use of this post-filtering technique is not restricted to the restoration of images corrupted by additive white Gaussian noise. It is applicable to any type of stationary noise (gaussian or non-gaussian, correlated or non-correlated) provided that the variance of this noise is known. This approach may also be used indirectly with certain types of signal dependent noise that can be transformed into signal independent noise through the use of an appropriate variance stabilizing transformation [6]. For example, Poisson signals can be processed by using a square-root transformation to produce a sequence of random variables that are approximately Gaussian with constant variance: $\sigma^2=\frac{1}{4}$. The final signal estimate is then obtained by applying the inverse transformation to the filtered data.

Finally, we note that the use of $\rho=0$ in (17) instead of the value prescribed by (3) usually does not degrade performance in a significant way provided that the averaging kernel is sufficiently large.

6. Conclusion

In this paper, we have considered the design of post-filters that recursively compute a weighted sum between the output of an initial restoration filter and the noisy image itself. These filters have the potential of improving the signal-to-noise ratio and the sharpness of the restored image. We have derived two adaptive algorithms that perform an optimal computation of the filter coefficients over a local sliding window. This process requires the specification of the noise statisics but, unlike some previous approaches, makes no assumptions about the signal.

To illustrate these results, we have concentrated on the particular case where the initial restoration filter is a simple moving average filter. We have compared our approach with Lee's adaptive noise filtering algorithm and emphasized the major differences. Despite the similarities between the structure of these filters, we found a consistent performance improvement over this latter scheme which we attribute to our more rigorous and more general derivation. An important filter parameter is the size of the estimation window which governs the adaptability of the algorithm to local image structures and which should be chosen in an optimal fashion.

The constrained algorithm (ACLSF) has a particularly simple structure and can be implemented in a very efficient way. It is almost as efficient as the unconstrained filter (ALSF). Due to its simplicity and its computational economy, this approach should be useful in many image processing applications such as the restoration of high resolution electron micrographs.

Acknowledgment

I would like to thank Dr. Murray Eden, BEIB, Chief, for his constant support and his valuable suggestions.

References

[1] J.-F. Abramatic and L.M. Silverman, "Nonlinear restoration of noisy images", *IEEE Trans. Pattern Analysis Machine Intelligence*, Vol. PAMI-4, No. 2, 1982, pp. 141–149.

[2] A.C. Bovick, T.S. Huang and D.C. Munson, "A generalization of median filtering using linear combinations of order statistics", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-31, December 1983, pp. 1342–1350.

[3] L.S. Davis and A. Rosenfeld, "Noise cleaning by iterated local averaging", *IEEE Trans. Systems Man Cybernet.*, Vol. SMC-8, 1978, pp. 705–710.

[4] B.R. Frieden, "A new restoring algorithm for the preferential enhancement of edge gradients", *J. Opt. Soc. Amer.*, Vol. 66, 1976, pp. 280–283.

[5] B.R. Hunt and T.M. Cannon, "Nonstationary assumptions for Gaussian models of images", *IEEE Trans. Systems Man Cybernet.*, Vol. SMC-6, 1976, pp. 876–881.

[6] M.G. Kendall and A. Stuart, "*The Advanced Theory of Statistics, Vol III, Design and Analysis, and Time Series,* Griffin, London, England, 1966.

[7] D.T. Kuan, A.A. Sawchuk, T.C. Strand and P. Chavel, "Adaptive noise smoothing for images with signal-dependent noise", *IEEE Trans. Pattern Analysis Machine Intelligence*, Vol. 7, No. 2, 1985, pp. 165–177.

[8] J.S. Lee, "Digital image enhancement and noise filtering by the use of local statistics", *IEEE Trans. Pattern Analysis Machine Intelligence*, Vol. 2, No. 2, 1980, pp. 165–168.

[9] J.S. Lee, "Refined filtering of images using local statistics", *Comput. Vision Graphics Image Process.*, Vol. 15, 1981, pp. 380–389.

[10] J.S. Lee, "Digital image smoothing and the sigma filter", *Comput. Vision Graphics Image Process.*, Vol. 24, 1985, pp. 255–269.

[11] Y.H. Lee and A.S. Kassam, "Generalized median filtering and related nonlinear filtering techniques", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-33, June 1985, pp. 672–683.

[12] G.A. Mastin, "Adaptive filters for digital image noise smoothing: An evaluation", *Comput. Vision Graphics Image Process.*, Vol. 31, 1985, pp. 103–121.

[13] W.K. Pratt, *Digital Image Processing*, Wiley, New York, 1978.

[14] A.C. Steven, J.F. Hainfeld, J.S. Wall and C.J. Steer, "Mass distributions of coated vesicles isolated from liver and brain: Analysis by scanning transmission electron microscopy", *J. Cell Biology*, Vol. 97, December 1983, pp. 1714–1723.

[15] M. Unser, "Recursion in short-time signal analysis", *Signal Process.*, Vol. 5, No. 3, May 1983, pp. 229–240.

[16] D.C.C. Wang, A.H. Vagnucci and C.C. Li, "Gradient inverse weighted smoothing scheme and the evaluation of its performance", *Comput. Graphics Image Process.*, Vol. 15, 1981, pp. 167–181.

[17] J.W. Woods and C.H. Radewan, "Kalman filtering in two dimensions", *IEEE Trans. Inform. Theory*, Vol. IT-23, No. 4, 1977, pp. 473–482.

[18] J.W. Woods and V.K. Ingle, "Kalman filtering in two-dimensions: Further results", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-29, 1981, pp. 188–197.