Sequential anisotropic Wiener filtering applied to 3D MRI data

Marcos Martin-Fernandez, Carlos Alberola-Lopez, Juan Ruiz-Alzola, Carl-Fredrik Westin

*Laboratory of Mathematics in Imaging (LMI), Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115, USA
Laboratorio de Procesado de Imagen (LPI), Universidad de Valladolid, 47011 Valladolid, Spain
Departamento de Investigación y Tecnología, Instituto de Tecnología de Canarias, 38009 Santa Cruz de Tenerife, Spain

Received 21 April 2006; accepted 14 May 2006

Abstract

We present three different sequential Wiener filters, namely, isotropic, orientation and anisotropic. The first one is similar to the classical Wiener filter in the sense that it uses an isotropic neighborhood to estimate its parameters. Here we present a sequential version of it. The orientation Wiener filter uses oriented neighborhoods to estimate the structure orientation present at each voxel, giving rise to a modified estimator of the parameters. Finally, the anisotropic Wiener filter combines both approaches adaptively so that the appropriate approach is locally selected. Several synthetic experiments are presented showing the performance of the filters with respect to their parameters. A mean square error analysis is performed using a publicly available magnetic resonance imaging (MRI) brain phantom and a comparison with other filtering approaches is carried out. In addition, results from filtering real MRI data are presented.

© 2007 Elsevier Inc. All rights reserved.

Keywords: MMSE criterion; Wiener filter; Parameter estimation; Sequential filter; Orientation filter; Anisotropic filter; MSE performance

1. Introduction

Noise image removal or, equivalently, the estimation of the underlying noiseless signal, has been reported in the literature for years. One may think several reasons for performing filtering; reducing the level of noise can, for example, help a segmentation algorithm extract finer details more robustly. Filtering is also useful for improving the visual appearance. The obvious solution for noise reduction is time averaging; its associated advantage is noise reduction without affecting the spatial resolution. A drawback with this technique is that several samples of the same data need to be acquired, which drastically increases the acquisition time (specially in the case of magnetic resonance images), a fact that, generally speaking, is hardly acceptable in a clinical setting. Additionally, data must be perfectly aligned, which, in real practice, poses additional practical problems.

In this article, we investigate how spatial averaging using Wiener theory [1,2] can be used to decrease the noise level in the image data. This is a well-known technique (see, for instance, [3] and references therein) that spatially filters data giving relative importance to neighbors of a given pixel according to the correlation structure; it has the main advantage of conceptual simplicity, but it has the drawbacks derived from the assumptions on which it is based and, particularly, from the need of (at least local) stationarity. If this is held, the Wiener solution is the minimum mean squared error estimator (MMSEE), provided that data are Gaussian and, if not, it is the linear MMSEE [4]. However, the assumption of stationarity will only be valid in fairly homogeneous areas of the image and will fail at boundaries between such regions resulting in interregion blurring and thus removing important details at edges. Adaptive versions of this filter [5–7] provide a good solution in the homogeneous regions while maintaining the discontinuities fairly unchanged, giving rise to a global reduction of the noise without a noticeable decrease in the spatial resolution. A further attempt in this direction is proposed in [8], in which a visibility function is defined to account for the presence of gradients in the surrounding of the pixel under test; if a high gradient is declared present, the filter becomes the identity function. Although these adaptive approaches implement a spatially varying smoothing, the resulting filtering is isotropic. Our contribution is to investigate on the merit of
combining an isotropic approach with an anisotropic one where the latter is only applied in the vicinity of boundaries.

Needless to say, many techniques have been proposed for image denoising. Apart from the Wiener filter itself and other geostatistical related approaches [9], one may find many others. Wavelet-based filters [10–12] have been popular, perhaps more enthusiastically in the last decade. In these approaches, a wavelet transform is applied and then coefficients are thresholded. Clearly, the design of the thresholding function and the choice of the mother wavelet are key issues in these methods. In any case, these methods are considered Wiener-like methods [11,13] in the wavelet domain so what we describe in this article does not compete with these ideas, but it is quite complementary since a similar implementation can be carried out in the wavelet domain.

Other linear filtering approaches have been reported as well; tensor-controlled quadrature filters have been proposed to that end [14,15]. These filters may have trouble, however, when the local simplicity assumption of the signal is not satisfied [14]. Not far from this idea is the filter proposed by Yang [16], in which a Gaussian kernel is tuned according to the local structure to be filtered. This idea is used as the starting point in [17]; the authors claim that as for magnetic resonance imaging (MRI) data, the Gaussianity assumption gives biased results due to the well-known fact that data have as first-order distribution a Rice probability density function (PDF). Then the authors fit a local surface model to the “signal parameter” of the Rice PDF, and the parameters of this surface model are identified via maximum likelihood estimation (MLE) using this PDF. Apparently this surface model is then smoothed with a filter whose shape is tuned very much as in [16], but with gradients derived from the parameters of the fitted surface model instead (i.e., not directly from the noisy data). Nice results seem to be obtained; however, results are not thoroughly explained since the experimental conditions are not fully reported (sample sizes, neighborhood shape, etc.).

More closely related to our work, we should mention [18] and [19]. In the former, the authors design a procedure to cause an offset to a filtering kernel to push it away from the boundaries. The authors intend to be competitive with anisotropic diffusion schemes in terms of speed, while maintaining performance sufficiently close to this alternative paradigm. In this attempt, however, the kernel itself seems to remain unchanged while its position is shifted according to the vector field calculated from the data. As for the latter, the filtering is preceded by a hypothesis test stage in which each image region is declared either homogeneous or nonhomogeneous. The way data are filtered depends on the result of this test: for homogeneous regions, data are simply averaged, while for those declared nonhomogeneous, some final value is obtained as a weight between the original value and either the minimum or the maximum value within the neighborhood of the pixel under test.

Among the methods described for image removal, probably those currently receiving more attention are the techniques classified under the “diffusion” procedure, that is, those spatial filtering methods based on solving the diffusion equation [20,21]. An extension of the original idea introduces directionality in the filtering process resulting in smoothing mostly along the direction of the edge or discontinuity wherever it is present [22]. Accurate estimation of the parameters involved in the filter is of importance [23]. The capacity of noise removal associated to this sort of filters is widely acceptable. On the other side, time consumption could be an issue due to their iterative nature [18]. Additionally, the approach is deterministic, that is, the underlying signal model is not stochastic; this is not any drawback per se, but it makes theoretical comparisons with stochastic methods hard, that is, it is difficult to assess whether such procedures are optimum in any statistical sense.

A key question at this point is, why, after all the material reported, Wiener approaches should be further researched? In our opinion, the inclusion of anisotropy in this type of filters is worth taking. To that end we will describe a fast solution that is competitive with other more sophisticated approaches, both in terms of mean square error (MSE) and in terms of computational load. We will provide three iterative Wiener filters specially designed to accomplish these goals for 3D volume data sets at a suitable computational effort [16]. To our knowledge, no other anisotropic Wiener-like filters similar to ours have been reported in the literature.

The article is structured as follows: Section 2 introduces the terminology and presents the unbiased Wiener solution in a general context. The next three sections present the three Wiener filters. At the end of these sections an iterative algorithm is provided. Section 6 presents some synthetic experiments with the goal of analyzing the performance of the proposed techniques both qualitatively and quantitatively. The analysis of the sensitivity of the filters results with respect to the parameters will allow us to draw some conclusions not only about the performance of the three filters, but also about how to set up the parameters properly. Finally, Section 7 shows some results on real data and Section 8 concludes the paper summarizing the achieved results.

2. Wiener image filtering

We assume the following model for our image data\(^1\)

\[
Y = X + N
\]

(1)

where \(Y\) is the observed data corrupted by additive noise, \(X\) the ground truth, that is, the signal without noise, and \(N\) the noise. Under the Bayesian paradigm, the ground truth is regarded as a sample from a stochastic process [4], the

\(^1\) The material in this section is well known. It is included here for the article to be self-contained.
parameters of which are assumed known or may be estimated from data. This is the way to include prior knowledge in such a framework. Our goal is to estimate \( \mathbf{X} \) using the observation \( \mathbf{Y} \). We hereafter assume the data have been rearranged as a column vector with \( \mathbf{N} \) elements. We also assume that the noise has zero mean and covariance matrix \( \mathbf{C}_N \) that is \( N \times N \), symmetric and positive semidefinite. We can then write\(^2\)

\[
E[\mathbf{N}] = 0 \quad E(\mathbf{NN}^T) = \mathbf{C}_N
\]

With respect to the signal \( \mathbf{X} \), we will assume a vector of means \( \eta_X \) with size \( N \times 1 \) and covariance matrix \( \mathbf{C}_X \) which is \( N \times N \), symmetric and positive semidefinite. We can then write

\[
E[\mathbf{X}] = \eta_X \quad E((\mathbf{X} - \eta_X)(\mathbf{X} - \eta_X)^T) = \mathbf{C}_X
\]

Finally, we will assume that the signal \( \mathbf{X} \) and the noise \( \mathbf{N} \) components are uncorrelated. As the noise has zero mean, the signal and the noise are orthogonal as well; thus, we can write

\[
E(\mathbf{XN}^T) = 0
\]

The Wiener filter is a space-variant linear filter, which is optimum in the MMSE sense \([1]\). Let \( \mathbf{Z} \) denote the filter output. As the filter is linear the output is given by

\[
\mathbf{Z} = \mathbf{W}^T \mathbf{Y}
\]

where \( \mathbf{W} \) is an \( N \times N \) matrix of coefficients to be determined. We are interested in filters to be unbiased, that is, the expected value at the output is desired to be equal to the expected value of the input

\[
E[\mathbf{Z}] = E(\mathbf{W}^T \mathbf{Y}) = \mathbf{W}^T E(\mathbf{Y}) = \mathbf{W}^T \eta_X = E(\mathbf{Y}) = \eta_X
\]

This is simply achieved by removing the mean and adding it back after the filtering operation, which has been termed as common practice in signal processing \([9]\), that is,

\[
\mathbf{Z} = \mathbf{W}^T (\mathbf{Y} - \eta_X) + \eta_X
\]

Then, well-known procedures \([4]\) lead to the solution

\[
\mathbf{Z} = \mathbf{C}_X (\mathbf{C}_X + \mathbf{C}_N)^{-1} (\mathbf{Y} - \eta_X) + \eta_X
\]

3. Sequential isotropic Wiener

Eq. (8) is the optimum unbiased linear filter under the MMSE criterion. However, generally speaking the parameters \( \eta_X, \mathbf{C}_X \) and \( \mathbf{C}_N \) are unknown and have to be estimated from the observation data \( \mathbf{Y} \). In practice, as only one sample for \( \mathbf{Y} \) is available, direct estimation resorts to iterative refinements \([24,25]\). Alternative solutions make use of simplifying assumptions \([4]\). One common assumption is that the data elements are mutually independent, both for the signal and for the noise. This means that the covariance matrices \( \mathbf{C}_X \) and \( \mathbf{C}_N \) are diagonal, the elements of which are the vector component variances. This simplification leads us to a local filtering approach instead of the global filtering given by Eq. (8). The local Wiener filter can be written as \([7]\)

\[
Z(n) = \frac{\sigma_X^2(n)}{\sigma_X^2(n) + \sigma_N^2} [Y(n) - \eta_X(n)] + \eta_X(n)
\]

where \( Y(n) \) is the \( n \)-th element of the vector \( \mathbf{Y} \), \( \sigma_X^2(n) = \mathbf{C}_X(n,n) \) is the signal variance and \( \sigma_N^2(n) = \mathbf{C}_N(n,n) \) the noise variance. This is the same as considering the matrix of coefficients \( \mathbf{W} \) diagonal, that is, instead of \( N^2 \) unknown coefficients, here we have \( N \) unknown coefficients one at each data position \( n \). We can say that the local filter is linear and space-variant memoryless.

It is worth noting that the assumed model both for the signal and the noise, although simplified, is still nonstationary as both means and variances are functions of the element index \( n \). To assume that the signal is stationary is in general far too restrictive, but this assumption is most times satisfied by the noise. Consequently, we will assume that the noise is stationary but that the signal is not. In this case, we can drop the index \( n \) for the noise variance, that is, we have only one noise variance \( \sigma_N^2 \). Thus, Eq. (9) can be rewritten as

\[
Z(n) = \frac{\sigma_X^2(n)}{\sigma_X^2(n) + \sigma_N^2} [Y(n) - \eta_X(n)] + \eta_X(n)
\]

Here, the number of unknowns to estimate have been considerably reduced. This number is equal to \( 2N + 1 \): \( N \) signal means \( \eta_X(n) \), \( N \) signal variances \( \sigma_X^2(n) \) and the noise variance \( \sigma_N^2 \). If these parameters were known, Eq. (10) would give the local Wiener solution; however, and generally speaking, this is not usually the case and we have to find out how to estimate those parameters in advance from the observation \( \mathbf{Y} \).

In order to step forward we have to rely on some assumption about the signal \( \mathbf{X} \). Though we said that this signal is clearly nonstationary, it is reasonable to assume that the local variation of both the mean and the variance within a given neighborhood is less than the global variation, that is, we will assume that the signal has only local interactions \([26]\). Our assumption is that the signal can be considered locally ergodic so as to use spatial averages to estimate the unknown parameters.

Let \( \mathbf{d}(n) \) denote a set of neighbors corresponding to the site (voxel) \( n \) of size \( L \). In Fig. 1, we can see four different isotropic neighborhoods corresponding to sizes \( L = 7 \), \( L = 19 \), \( L = 27 \) and \( L = 33 \), respectively. We called these neighborhoods isotropic as there is no preferred orientation, that is, they are symmetric with respect to the volume coordinate axis. We will further assume that \( \mathbf{d}(n) \) is equal for any site \( n \), with \( 1 \leq n \leq N \).

The idea here is to use the neighborhoods \( \mathbf{d}(n) \) to estimate the \( 2N + 1 \) unknowns. As the neighborhoods are isotropic we will refer to this filter as isotropic Wiener, that

\(^2\) Superscript “\( T \)” stands for matrix transposition.
Fig. 1. Isotropic neighborhood system denoted by $\mathcal{A}(n)$ with (A) $L=7$ neighbors, (B) $L=19$ neighbors, (C) $L=27$ neighbors and (D) $L=33$ neighbors.

is, we are assuming that the data are isotropic. That is the first approach we will present. In subsequent sections we will introduce anisotropic models that will be able to deal with anisotropic data.

Assuming local ergodicity in the set $\mathcal{A}(n)$ for site $n$, we can estimate the local mean $\eta_X(n)$ as the local sample means of the data $Y$ for that neighborhood given by\footnote{For a given set $A$ the operator $|A|$ stands for its cardinal, i.e., number of elements in set $A$.}

$$\eta_X(n) = \eta_Y(n) \approx \frac{1}{|\mathcal{A}(n)|} \sum_{m \in \mathcal{A}(n)} Y(m)$$

We have made use of the fact that the mean of the noise is zero. The estimation of the variance is given by the following unbiased estimator

$$\sigma_X^2(n) \approx \sigma_Y^2(n) \approx \frac{1}{|\mathcal{A}(n)| - 1} \times \sum_{m \in \mathcal{A}(n)} \left[ Y(m) - \frac{1}{|\mathcal{A}(m)|} \sum_{p \in \mathcal{A}(m)} Y(p) \right]^2$$

These estimations are approximations to the true mean and variances. If some prior knowledge about the real distribution of the data is provided, the estimation of the mean and the variance can be improved using the true distribution. It is well known that for MRI the true distribution is Rician \cite{27}. In this case the mean will overestimate the deterministic parameter of the distribution whenever the signal-to-noise ratio takes on small values. However, for MRI of the brain, the percentage of voxels for which the signal-to-noise ratio is small is not significant whenever the interest is focused on brain tissues. For white matter and gray matter, the signal-to-noise ratio is considerably higher than 1.91, which is known to be the signal-to-noise value for the Rayleigh case with no signal component. For other brain tissues as for the cerebrospinal fluid and for nonbrain tissues within the patient’s head the signal-to-noise ratio can be closer to that bound and, thus, a correction for the bias would be needed. This can be accomplished by, for instance, using the MLE of the deterministic component parameter of the Rice distribution \cite{28,29} instead of Eq. (11) and calculating the sample variance of these estimates instead of Eq. (12).

The estimation of the noise variance $\sigma_N^2$ is more involved and will have a great influence in the filter output as we will see later. The problem is due to the fact that the variances add up as

$$\sigma_Y^2(n) = \sigma_X^2(n) + \sigma_N^2$$

One possible solution is to look for a region in the image for which the signal component is a priori known to be zero \cite{30}. In this case, $\sigma_Y^2(n)$ will be zero and we will have that $\sigma_Y^2(n) = \sigma_X^2$ in that region. Thus, the sample variance in that region would give us a reasonable estimation for the noise variance. In practice, this requires either user interaction to select this region or an automated selection, being the latter always a matter of objection.

Let look at the minimum of the local variances given by

$$\sigma_{\min}^2 = \min_{1 \leq n \leq N} \sigma_Y^2(n)$$

If the estimation of the local variances $\sigma_Y^2(n)$ were exact, that minimum variance $\sigma_{\min}^2$ will give us a good estimator for the noise variance which corresponds to a site $n$ for which the signal variance $\sigma_Y^2(n)$ is minimum and probably very close to zero. However, we do not know the exact variance $\sigma_Y^2(n)$ but its estimation, which is given by Eq. (12). Consequently, the minimum will underestimate due to the presence of outliers. We can consider this minimum as a lower bound for the noise variance. An upper bound for the variance will be given by the average of the local variances

$$\sigma_{ave}^2 = \frac{1}{N} \sum_{n=1}^{N} \sigma_Y^2(n)$$

If there were no signal component in the data $Y$, that estimator would lead to an accurate estimation for the noise variance. However, this is never the case, so the average $\sigma_{ave}^2$ can only represent an upper bound.

The solution here is to introduce a free parameter $\lambda$ ranging in the interval $(0,1)$ that will be hereafter referred to as regularization parameter \cite{31}, which would allow us to fine tune the amount of regularization required by means of a selection of the noise variance between the given bounds as

$$\sigma_N^2 = \lambda \sigma_{\min}^2 + (1 - \lambda) \sigma_{ave}^2$$

We can say the higher the $\lambda$ value, the lower the noise variance $\sigma_N^2$ and, consequently, the lower the regularization.
This parameter may be either manually set or learnt by training for a specific target application. We have tuned the filter, as indicated in Section 6, for MRI data.

To conclude, we can say that as the Eqs. (11), (12) and (16) are coarse approximations, the formal strength of the method is not accompanied by powerful results. Therefore, we propose here to sequentially repeat the filtering approach described above up to five times (this choice is justified in Section 6). The final approach will be:

1. Select a regularization value \( \lambda \equiv (0,1) \).
2. Set counter to 1.
3. Estimate the \( N \) unknown means \( \eta_X(n) \) using Eq. (11).
4. Estimate the \( N \) unknown variances \( \sigma_X^2(n) \approx \sigma_Y^2(n) \) using Eq. (12).
5. Estimate the noise variance \( \sigma_N^2 \) using Eqs. (14), (15) and (16).
6. Filter the image \( Y \) using the estimations above and Eq. (10) to get a new image \( Z \).
7. Set \( Y = Z \).
8. Increment counter.
9. Go to Step 3 or exit if counter has reached its maximum value.

4. Sequential orientation Wiener

The Wiener filter described in the previous section works as expected only in homogeneous regions. In a homogeneous region, the local variance \( \sigma_X^2(n) \) will be relatively small compared to the noise variance \( \sigma_N^2 \), then Eq. (10) can be read as \( Z(n) \approx \eta_X(n) \); that is, the noise would be eliminated in that region. In regions for which a clear discontinuity is present, \( \sigma_X^2(n) \) will be relatively large compared to the noise variance \( \sigma_N^2 \), then Eq. (10) would yield \( Z(n) \approx Y(n) \), that is, the original data would remain untouched. That was the reason we called that filter isotropic.

A more evolved approach would be that one which allows filtering in regions where a discontinuity is present. In these cases the filter should do its work only along the discontinuity without mixing up both regions. In order to do so, we need to find out for each site \( n \), first, whether or not a discontinuity is present, and second, its orientation in the case that a discontinuity is present. In this section, we will deal with the latter issue, while the former will be discussed in the next section. In what follows we are going to assume that a discontinuity is present at each site \( n \) and we will estimate its orientation. In next section, we will decide whether or not a discontinuity is present, comparing isotropic estimations with their anisotropic (i.e., oriented) counterparts. We will say that the method that will be described in the next section is a combination of the method described here and the one described in the previous section. The method that will be now presented will be hereafter referred to as orientation Wiener, while the method to be presented in the next section will be referred to as anisotropic Wiener.

In order to take into account data orientation we propose to split the neighborhood system \( \mathcal{B}(n) \) defined at each site \( n \) with \( L = 27 \) neighbors in six neighborhood subsystems with six different spatial orientations as shown in Fig. 2. We will index orientations with counter \( k \); thus, \( \mathcal{B}_k(n) \) will denote the \( k \)-th neighborhood subsystem (out of 6) for site \( n \).

![Fig. 2. Anisotropic neighborhood subsystems with \( L = 27 \) encoding six orientations (red sites).](image-url)
We estimate the unbiased local variance for each \( k \) orientation as
\[
\sigma_k^2(n) \approx \frac{1}{|B_k(n)| - 1} \left( \sum_{m \in B_k(n)} Y(m) - \frac{1}{|B_k(n)|} \sum_{p \in B_k(n)} Y(p) \right)^2
\]
(17)

Whenever the orientation of the data aligns with one of the neighborhood subsystem, the corresponding variance \( \sigma_k^2(n) \) will attain a minimum value. This fact will allow us to estimate the orientation, given by the integer \( q(n) \) at each site \( n \), as
\[
q(n) = \arg \min_{1 \leq k \leq 6} \sigma_k^2(n)
\]
(18)

Under the assumption that the voxel size is small, we have assessed that six orientations are usually enough. Once the orientation has been estimated, we can easily rewrite Eqs. (11) and (12) for the new situation using the sites belonging to the subsystem \( B_{q(n)}(n) \) for each site \( n \). In particular, the mean is given by
\[
\eta_X(n) = \eta_Y(n) \approx \frac{1}{|B_{q(n)}(n)|} \sum_{m \in B_{q(n)}(n)} Y(m)
\]
(19)
and the variance by
\[
\sigma_X^2(n) \approx \sigma_Y^2(n) = \sigma_{q(n)}^2(n)
\]
(20)

For the same reason as stated in the previous section, Eqs. (16), (17), (18), (19) and (20) are not so accurate; we have therefore resorted to the following sequential approach:

1. Select a regularization value \( \lambda \in (0,1) \).
2. Set counter to 1.
3. Estimate the \( N \) unknown local orientations \( q(n) \) using Eqs. (17) and (18).
4. Estimate the \( N \) unknown means \( \eta_X(n) \) using Eq. (19).
5. Estimate the \( N \) unknown variances \( \sigma_X^2(n) = \sigma_{q(n)}^2(n) \) using Eq. (20).
6. Estimate the noise variance \( \sigma_N^2 \) using Eqs. (14), (15) and (16).
7. Filter the image \( Y \) using the estimations above and Eq. (10) to get a new image \( Z \).
8. Set \( Y = Z \).
9. Increment counter.
10. Go to Step 3 or exit if counter has reached its maximum value.

5. Sequential anisotropic Wiener

In the previous section we assumed that a discontinuity (boundary) was always present at each site. This is obviously too restrictive. Here, we will deal with the problem of deciding whether or not a boundary is present at a given site position \( n \). We will use the orientation-bearing variances \( \sigma_k^2(n) \), given by Eq. (17), to, first, estimate the presence of a boundary, and second, select either the isotropic model described in Section 3 if no boundary is present or the orientation model described in Section 4 if it is. This model which combines the isotropic and the orientation models, as previously stated, will be hereafter referred to as anisotropic Wiener.

In order to estimate the presence of a boundary at a given site \( n \), we will study the variability of the orientation-bearing variances \( \sigma_k^2(n) \) among the neighborhood subsystems as \( k \) changes. We are going to define what we will call orientation variability. To do so, let us define the minimum value for \( \sigma_k^2(n) \) as
\[
\sigma_{\min}^2(n) = \min_{1 \leq k \leq 6} \sigma_k^2(n)
\]
and the average as
\[
\sigma_{\text{ave}}^2(n) = \frac{1}{6} \sum_{k=1}^{6} \sigma_k^2(n)
\]
(22)

Both parameters will allow us to define the orientation variability \( \gamma(n) \) for each site \( n \) as
\[
\gamma(n) = \frac{\sigma_{\text{ave}}^2(n) - \sigma_{\min}^2(n)}{\sigma_{\text{ave}}^2(n)}
\]
(23)
which will range within the interval \((0,1)\). For a given site \( n \) with no boundary the variability among the variances \( \sigma_k^2(n) \) will be small, yielding a value for \( \gamma(n) \) close to zero. Whenever a boundary, following any orientation, is present there will be a lack of homogeneity in the neighborhood, which will lead to higher variability and consequently will give rise to a value of \( \gamma(n) \) approaching one. In order to select the isotropic or the orientation models, a hard decision will be necessary, which means a threshold operation. If no other prior information is available, there would be no reason to choose a different threshold for different sites, so we will use a global (homogeneous) threshold to select which model to apply. This global threshold will be denoted by \( \gamma_0 \) and should range in the interval \((0,1)\). Depending on the sharpness of the boundaries present in the data we have to set higher or lower threshold; the sharper the boundaries the higher the threshold. If the threshold is not properly selected, the misclassification rate can increase. In particular, an underestimated value will lead to some isotropic sites being considered as boundaries and an overestimation will mean to miss some smooth boundaries. Thus, the anisotropic model has two parameters; namely, first the amount of expected regularization, set by \( \lambda \) and, secondly, the threshold \( \gamma_0 \), which allows to characterize the boundaries.

\[^4\text{The maximum value can be used instead of the average to achieve similar performance, but the maximum is in general more dependent on the presence of noise.}\]
Fig. 3. First synthetic experiment: (A) original image, (B) with added noise, (C) sequential isotropic Wiener solution, (D) sequential orientation Wiener solution and (E) sequential anisotropic Wiener solution.

Fig. 4. Second synthetic experiment: (A) original image, (B) with added noise, (C) sequential anisotropic Wiener solution and (D) boundary mask.
The final value for the mean \( \eta_X(n) \) is given by

\[
\eta_X(n) = \eta_Y(n) \approx \begin{cases} 
\frac{1}{|\mathcal{A}(n)|} \sum_{m \in \mathcal{A}(n)} Y(m) & \gamma(n) > \gamma_0 \\
\frac{1}{|\mathcal{A}(n)|} \sum_{m \in \mathcal{A}(n)} Y(m) & \gamma(n) \leq \gamma_0 
\end{cases}
\]  
(24)

where \( q(n) \) is the orientation given by Eq. (18) at site \( n \). The final value for the variance \( \sigma_X^2(n) \) is similarly given by

\[
\sigma_X^2(n) \approx \sigma_Y^2(n) \approx \begin{cases} 
\sigma_{q(n)}^2(n) & \gamma(n) > \gamma_0 \\
\left[ \frac{1}{|\mathcal{A}(n)| - 1} \sum_{m \in \mathcal{A}(n)} \left( Y(m) - \frac{1}{|\mathcal{A}(m)|} \sum_{p \in \mathcal{A}(m)} Y(p) \right) \right]^2 & \gamma(n) \leq \gamma_0 
\end{cases}
\]  
(25)

Finally, we will implement a sequential version of the filter in order to deal with the lack of accuracy in the estimation of the parameters as presented above. Specifically, the sequential anisotropic Wiener filter is being given by:

1. Select a regularization value \( \lambda \approx (0,1) \).
2. Select a threshold \( \gamma_0 \approx (0,1) \).
3. Set counter to 1.
4. Estimate the \( 6N \) unknown orientation-bearing variances \( \sigma^2_X(n) \) using Eq. (17).
5. Estimate the \( N \) unknown orientation variability \( \gamma(n) \) using Eqs. (21), (22) and (23).
6. Estimate the \( N \) unknown local orientations \( q(n) \) using Eqs. (17) and (18).
7. Estimate the \( N \) unknown means \( \eta_X(n) \) using Eq. (24).
8. Estimate the \( N \) unknown variances \( \sigma^2_X(n) \approx \sigma^2_Y(n) \) using Eq. (25).
9. Estimate the noise variance \( \sigma_N^2 \) using Eqs. (14), (15) and (16).

---

Fig. 5. Second synthetic experiment: convergence of the filters and sensitivity with respect to the neighborhood size \( L \). (A) Global MSE value. (B) MSE value for the boundaries. (C) MSE value in homogeneous regions.
10. Filter the image \(Y\) using the estimations above and Eq. (10) to get a new image \(Z\).
11. Set \(Y=Z\).
12. Increment counter.
13. Go to Step 3 or exit if counter has reached its maximum value.

6. Synthetic evaluation

In this section, we will present the results for two volume data sets synthetically generated. For the first one, the goal is to visually determine the performance of the three proposed Wiener filtering schemes: isotropic, orientation and anisotropic. To that end, we have created a volume data set with size of \(164 \times 164 \times 164\). A spherical object has been generated, the radial profile of which follows a sawtooth function with one discontinuity. Fig. 3A shows a quadrant (90°) of a slice as a gray-level image. We can see in that image a clear circle-shaped discontinuity. We have modified this original 3D image by adding colored noise that has zero mean, Gaussian distribution and Laplacian frequency spectrum. The gray-level image for the same region of interest is shown in Fig. 3B. Fig. 3C shows the filter output for the sequential isotropic Wiener filter after 10 iterations. In this case, as the object structure is large and the amount of noise present in the image is high we have set \(k=0.9\). As we can see in Fig. 3C, the filter has performed an adequate job in the two ramp-shaped regions; however, the edge has not been filtered. This is due to the fact that this filter was designed to deal with continuous-intensity profiles and not to deal with discontinuities. Fig. 3D shows the result obtained after 10 iterations using the orientation Wiener with \(\lambda=0.9\). In this case, an opposite effect can be seen: the boundary is well filtered, but false edges have arisen in the ramp-shaped regions. This effect should have been expected as this filter can deal with discontinuities, but not with continuous intensity profiles (a ramp is a continuous function that can be considered as locally constant). Finally,
Fig. 3E shows the result achieved using the anisotropic Wiener filter after 10 iterations with $\lambda=0.9$. We have set $\gamma_0=0.5$ as the filter has to deal both with discontinuities and with continuous regions. In this case, the continuous regions are similar to the ones in Fig. 3C and the boundary is similar to that in Fig. 3D, although more variation can be seen in the continuous regions with some false boundaries and some blurring has appeared in the boundary. This is due to the great amount of noise present in the noisy data as shown in Fig. 3B. In the three cases the neighborhood size was set to $L=27$ neighbors.

In the second experiment the goal is to study the sensitivity of the method with respect to the neighborhood size $L$, the regularization parameter $\lambda$ and the threshold parameter $\gamma_0$. In this case the synthetic data sets were generated using a MRI brain simulator, which resembles to some extent real MRI brain data [32]. In [33], the authors have reported that their simulator can be used to evaluate image processing algorithms. The simulator is online at http://www.bic.mni.mcgill.ca/brainweb/. Our first data set is for normal brain, T1, $1\times1\times1$ mm, with 0% noise and 0% nonuniformity. The size of this data set is $212\times176\times110$. Fig. 4A shows a 3D view for this phantom data. We will consider this data set as our ground truth. A second data set has also been generated, a noisy one, for which the noise parameter in the web site simulator was 9%. Fig. 4B shows the same 3D view as before, but for the noisy data. Fig. 4C shows one of the results.

As we pursue to quantify the performance of the filters with respect to the above-mentioned parameters, a measure of error has to be provided. As the ground truth is known and given by the data set associated to Fig. 4A, the MSE is defined, for a given image, as the average value of the square of the difference between this image and the ground truth image. We are not only interested in quantifying the results achieved in continuous regions, but also at the boundaries present in the image. We will consequently define three different MSE values: a global MSE using all the voxels in the image, MSE at boundaries using only these voxels at which a boundary is present and MSE in homogeneous regions using only these voxels at which no boundary is present. To that end, we have created a binary mask by thresholding a gradient image, with the latter calculated out of a smoothed version of the ground truth. This mask is shown in Fig. 4D. Using this mask, or its complement to one, we can easily determine which voxels have to be used for the MSE at boundaries and which for the MSE in homogeneous regions. In any case the voxels belonging to the background have never been considered.

Fig. 5 shows the achieved results as far as convergence and neighborhood size $L$ are concerned. Fig. 5A shows the results for the global MSE, Fig. 5B for the MSE at boundaries and Fig. 5C for the MSE in homogeneous regions. The solid horizontal line represents the MSE of the noisy data shown in Fig. 4B, the line with circles the filtering result for $L=7$, the one with asterisks for $L=19$ and the one with squares for $L=27$. The filter used for this experiment was the isotropic with $\lambda=0.5$. We can see that the larger the neighborhood, the faster the convergence. For instance, for $L=7$, the convergence is achieved after about eight iterations, while for $L=27$ the convergence is usually attained for 3 or 4. For each iteration, the larger the neighborhood the greater the execution time. As with larger neighborhoods the needed iterations are reduced to one half, the overall effect is that the larger the neighborhood the less the total execution time, so larger neighborhoods, up to $L=27$ neighbors, are preferable.

We can also see using the results shown in Fig. 5 that for the isotropic filter the convergence is faster over the continuous regions than for the boundaries and that the attained MSE is considerably lower: around 150 at the boundaries and slightly less than 100 over the continuous regions. A third conclusion we can draw is that more iterations do not necessarily mean better MSE performance. A problem that arises here is how to choose an optimum stopping point. As the graphs for smaller neighborhoods are more flat, the final MSE is less dependent on the stopping point. The downside in this case is that more iterations are needed and, consequently, more execution time. A trade-off between MSE accuracy and execution time can be made. Generally speaking, we can say that if one is more interested in achieving better results for continuous regions the stopping point should be after three or four iterations; however, if one is more interested in the boundaries the stopping point can be set after five to eight. Finally, when setting the neighborhood size to $L=27$ we need less iterations, but an error at estimating a proper stopping point would cost an increase on the final MSE. Decreasing the neighborhood size to $L=19$ or to $L=7$ would need more iterations to attain the same MSE, but the stopping point is less sensitive to an error.
Fig. 6 presents some results using the same synthetic data shown in Fig. 4 about the sensitivity of the filters with respect to the \( k \) parameter. MSE results are shown after five iterations with \( L = 27 \) as a function of the parameter \( \lambda \). Fig. 6A shows the results for the global MSE, Fig. 6B for the MSE at boundaries and Fig. 6C for the MSE in homogeneous regions. In each case, the solid line represents the MSE of the noisy data shown in Fig. 4B, the one with circles the filtering result for the isotropic Wiener, the one with asterisks for the orientation Wiener and the one with squares for the anisotropic Wiener. For the anisotropic Wiener the threshold was set to \( \gamma_0 = 0.25 \). If one is to focus on the boundaries, high values for \( \lambda \) are clearly encouraged, ranging from 0.45 to 0.8. If removing more noise over continuous regions is more important, lower values for \( \lambda \) should be chosen, ranging from 0.3 to 0.5. In any case, the

![Fig. 6.](image)

Fig. 8. Experiment with real MRI data: (A) original noisy data, (B) sequential isotropic Wiener solution and (C) sequential anisotropic Wiener solution.
isotropic Wiener filter is always worse than any of the other two. Taking into consideration only the MSE value, both the orientation and the anisotropic filter performances are quite similar, even it seems that the orientation is slightly better than the anisotropic. In practice, due to the false boundary artifact shown in Fig. 3D, we can say that, speaking in terms of MSE, both schemes are similar, but the anisotropic filter improves the continuous regions avoiding that the false boundary artifact appears.

Fig. 7 deals with the sensitivity of the anisotropic Wiener filter with respect to the threshold parameter $\gamma_0$. For the MSE results shown in this figure we have set $\lambda = 0.5$ and $L = 27$ and stopped the filter after five iterations. The line with circles represents the global MSE, the one with asterisks the MSE at boundaries and the one with squares the MSE in homogeneous regions. $\gamma_0 = 1$ corresponds to the isotropic Wiener and $\gamma_0 = 0$ to the orientation Wiener. Any point in between is the anisotropic Wiener. Taking in consideration only the results in this Fig. 7, it seems that the lower values of $\gamma_0$ are better. That is true only in terms of MSE, but if we recall Fig. 3D, the boundary artifact will start to appear in continuous regions for low values of $\gamma_0$ (orientation Wiener case). We have to establish a trade-off between the allowed boundary artifact and the increase in MSE. The lower the $\gamma_0$ parameter the lower the MSE, but the more likely the boundary artifact. In addition, the MSE
graphs are very flat in the range (0.1,0.3) and the probability that a continuous region is considered as a boundary decreases considerably in that range, so a good trade-off in the $\gamma_0$ parameter could be 0.3. This can also depend on the amount of noise present in the image, so a training stage for which some images are filtered under supervision is definitely recommended. Fig. 4C shows the result using the best parameter setting yielding the optimum solution using the anisotropic Wiener filter after five iterations and with $L=27$, $\lambda=0.5$ and $\gamma_0=0.25$.

7. Filtering real MRI data

7.1. Visual results

Here we will present some results using real MRI data. The MR volume has been acquired on a GE-Signa 1.5-T scanner (General Electric, Milwaukee, WI). Coronal T1-weighted images were acquired with a 3D volumetric radiofrequency spoiled gradient echo (SPGR) series with the following scan parameters: $TR=35–45$ ms, $TE=5–7$ms, flip angle=45$^\circ$, NEX=1. The resolution is $0.9375 \times 0.9375 \times 0.9375$ mm. The final volume size is $176 \times 201 \times 111$.

Fig. 8 shows the results obtained. In particular, Fig. 8A shows the original noisy data in three views: axial, sagittal and coronal. A specific region of interest has been magnified for each view to highlight the filter performance. Fig. 8B shows the same view for the output of the isotropic Wiener filter after five iterations for which $L=27$ and $\lambda=0.05$ were set. As we are interested in keeping all the fine details and due to the fact that the level of noise is not very high, we have reduced the $\lambda$ value correspondingly. Fig. 8C shows the same view for the anisotropic Wiener filter output using the same parameter set as before but with threshold $\gamma_0=0.25$. It is clear that both filters worked very well in the continuous (homogeneous) regions; however, the isotropic...
filter did not work properly at the boundaries (discontinuities) as some amount of blur is always introduced. Concerning the anisotropic filter, we can see in zoomed subimage shown in Fig. 8C that in this case the boundaries are not blurred and are much clearer.

7.2. Comparison with other methods

Fig. 9 shows a performance comparison, in terms of the global MSE, of the Wiener filtering procedure described in this article and other methods reported in the literature. Specifically, both Wiener filters (bars labeled W1 and W2) use \( \lambda = 0.5 \) and five iterations. For the former \( \gamma_0 = 0.2 \) and for the latter \( \gamma_0 = 0.3 \). Bars labeled M1 and M2 show the results of 3D median filter [7] with window length three and five voxels in each dimension; bars labeled G1–G3 correspond to an anisotropic diffusion filter [20] and bars labeled F1–F3 are those from a flux-diffusion filter [22].

Clearly, in terms of MSE, the best choice are filters F1–F3; it is remarkable, however, that the simple median filter M1 attains a satisfactory MSE figure. However, using the MSE as the only quality measure is, as it is well known, objectionable. Specifically, Fig. 10 shows both an original image shown in Fig. 10A and its noisy counterpart shown in Fig. 10B used for the experiment carried out to get Fig. 9. As for the filtered images, we show the results of M1, F1 and W1, respectively, in Fig. 11A–C.

Which one is the best is clearly a matter of discussion, but it is our opinion that the result of M1 has a noisier resemblance than the result of W1. Additionally, the result of W1 does not show a higher feeling of blur than both M1 and F2. This being the case, and taking into consideration that the filter F2 takes 27 s per iteration (in a Sun UltraSparc 10) as opposed to W1, which takes 14 in the same machine (not to mention the higher amount of memory needed in the former), makes our approach a candidate to be taken in consideration.

8. Conclusions

We have presented three iterative filtering algorithms that depart from two design conditions, namely, unbiasedness and optimality in the mean square sense, that is, Wiener filters. The difference among them relies on how to estimate the filter parameters, namely, mean and variances, at each voxel site. The filters were termed as isotropic, orientation and anisotropic. Several synthetic experiments have been presented focusing on the sensitivity of the MSE with respect to the different parameters in play: number of iterations, neighborhood size \( L \), amount of regularization \( \lambda \) and threshold \( \gamma_0 \). We have defined three MSE values: global MSE, MSE at boundaries and MSE in homogeneous regions, which have allowed us to study several aspects of the filter performances. Finally, some results using real MRI data have been presented. The anisotropic Wiener filter seems to be the better solution within the proposed filters. This filter can be used both for denoising purposes and as a preprocessing step before segmentation. In the latter case, it is really important not to blur the boundaries, as for any segmentation algorithm, the shape of the boundaries are as least as important as the amount of noise reduction achieved by the filter. This filter not only preserves the boundaries, but also filters them out along the boundary direction. Comparison to other methods have shown that, despite our filters are not the best in terms of final MSE, it does achieve a nice compromise between computational cost, amount of memory used and performance.

Acknowledgments

The authors acknowledge NIH for research grant P41-RR13218, the Comision Interministerial de Ciencia y Tecnologia, Spain, for research grants TIC2001-3808-C02 and TEC2004-06647-C03 and the European Commission for the funds associated to the Network of Excellence SIMILAR (FP6-507609). The first author acknowledges the Fulbright Commission for the postdoctoral grant FU2003-0968.

References


