Weighted median filters: A tutorial

Article *in* IEEE Transactions on Circuits and Systems II Analog and Digital Signal Processing · April 1996

DOI: 10.1109/82.486465 · Source: IEEE Xplore

Circuits and Systems Exposition

Weighted Median Filters: A Tutorial

Lin Yin, Ruikang Yang, *Member, IEEE,* Moncef Gabbouj, *Senior Member, IEEE,* and Yrjö Neuvo, *Fellow, IEEE*

***Abstract*— Weighted Median (WM) filters have attracted a growing number of interest in the past few years. They inherent the robustness and edge preserving capability of the classical median filter and resemble linear FIR filters in certain properties. Furthermore, WM filters belong to the broad class of nonlinear filters called stack filters. This enables the use of the tools developed for the latter class in characterizing and analyzing the behavior and properties of WM filters, e.g. noise attenuation capability. The fact that WM filters are threshold functions allows the use of neural network training methods to obtain adaptive WM filters. In this tutorial paper we trace the development of the theory of WM filtering from its beginnings in the median filter to the recently developed theory of optimal weighted median filtering. The following one and multidimensional applications are presented in this paper: idempotent weighted median filters for speech processing, adaptive weighted median and optimal weighted median filters for image and image sequence restoration, weighted medians as robust predictors in DPCM coding and Quincunx coding, and weighted median filters in scan rate conversion in normal TV and HDTV systems.**

I. Introduction

LINEAR filters have been the dominating filter class through the history of signal processing, due mainly to the sound theoretical basis provided by the theory of linear systems and the computational efficiency of linear filtering algorithms. Despite the elegant linear system theory, not all signal processing problems can be satisfactorily addressed through the use of linear filters. Linear filters tend to blur sharp edges, fail to remove heavy tailed distribution noise effectively, and perform poorly in the presence of signal-dependent noise [78], [79]. Many researchers seem to have the view that it is difficult to obtain major breakthroughs in signal processing without resorting to nonlinear methods. Neural networks serve as an example of the potential of the nonlinear approach.

On the other hand, due to the lack of sound underlying theory, nonlinear methods tend to be *ad hoc* and application specific making it very difficult to assess their performance and optimize the designs. A good underlying theory covering a subclass of nonlinear filters is needed for the research to have a common foundation and for the methods to be transferable among applications.

A nonlinear filter class that has been proven very useful is the class of median based filters. During the last few years their theory has also been developing quite fast. It is the opinion of the authors that median based filters today give one sound approach to nonlinear filtering. The success of median filters is based on two intrinsic properties: edge preservation and efficient noise attenuation with robustness against impulsive-type noise. Neither property can be achieved by traditional linear filtering techniques without resorting to time-consuming and often *ad hoc* data manipulations. Edge preservation is essential in image processing due to the nature of visual perception. Edges also occur in biomedical signals when the "system" moves from one state to another. In addition to being able to effectively eliminate impulsive noise, encountered, e.g., in communications systems and TV images [52], [76], it is worth noting that the median filter is the optimal filter for biexponential noise in a similar manner as the sliding average is for Gaussian noise.

The median filter is not, however, a perfect filtering operation nor it is very flexible. It may cause edge jitter [11], streaking [12] and may remove important image details [2], [66]. The main reason is that the median filter uses only rank-order information of the input data within the filter window, and discards its original temporal-order information. For constant signals embedded in i.i.d. noise, temporal-order is irrelevant. This, however, is not true for nonconstant signals corrupted by noise [72]. In order to utilize both rank- and temporal-order information of input data, several classes of rank order based filters have been developed in recent years, such as FIR-median hybrid (FMH) filters [40], [66], [45], Ll filters [53], [72], weighted median (WM) and weighted order statistics (WOS) filters [13], [42], [51], [98], [104], [115], [120], [123], stack filters [16], [24], [27], [54], [58], [81], [91], [98], [113], [125], and Boolean filters [10], [41], [55], [119]. FMH filters and Ll filters make use of the desirable properties of both linear filters and median filters. Several surveys have been written to trace the development from median to stack filtering, see for instance, [18], [19], [30], [32].

Other advances in this area brought about the development of Vector Median filters and Matched Median filters [5], [6].

Manuscript received March 10, 1994; revised November 1, 1995. This paper was recommended by Associate Editor S. Kiaei.

L. Yin and R. Yang are with the Nokia Research Center, Tampere, Finland.
M. Gabbouj is with the Department of Information Technology, Pori Tampere University of Technology, Finland.
Y. Neuvo is with Nokia Mobile Phones R&D Center, Espoo, Finland.

Publisher Item Identifier S 1057-7130(96)02269-0.

1057-7130/96$05.00 © 1996 IEEE

The vector median was introduced for multispectral signal processing to utilize correlation between different components [5]. Multispectral satellite images and color images in TV systems are typical examples. Vector median filters have been applied for impulsive noise attenuation in color images and cross luminance and cross color elimination [70], [71]. Matched median filters were introduced for detection in communication systems and defined both for baseband and passband systems [6], [101]. It was shown that the matched median filter outperforms the linear matched filter in the case of binary antipodal signaling under impulsive noise.

In this paper, we will concentrate on WM filters which are the simplest rank order based filters among those mentioned above. An N-length WM filter can be described by N parameters and implemented using a sorting operation with the same order of computations as the same size median filter. On the other hand, WM filters offer much greater flexibility in design specifications than the median filter. The weights control the filtering behavior. The relation of WM filters to the median filter is comparable to that of linear FIR filters to the average filter.

The aim of this paper is to summarize the recent developments in the theory and applications of WM filtering. In Section II, median and WM filters are introduced. Some very useful deterministic and statistical properties of median filters are summarized in the same section, and the analogy between WM filters and linear FIR filters is highlighted. In Section III, the connection between WM filters and Positive Boolean Functions (PBF) is reviewed under the threshold decomposition framework. Section IV explores the main deterministic and statistical properties of WM filters. The problem of optimal design of WM filters is addressed in Section V. An important extension of WM filters, FIR-WOS hybrid filters are reviewed in Section VI. Several 1-, 2-, and 3-D applications of WM filters are outlined in Section VII, ranging from speech processing to image sequence processing. Section VIII contains some conclusions. Interested readers are referred to the extensive bibliography for more detailed discussions about the theory, extensions, and applications of WM and related filters.

II. Median and Weighted Median Filters

Following a brief discussion of the basic deterministic and statistical properties of median filters, WM filters are formally defined and an interesting analogy is drawn between the latter and linear FIR filters.

A. Median Filtering

In an abstract published in EASCON'74, J. Tukey mentioned the "running" median as one of those scale invariant nonlinear smoothers which might be useful in smoothing time series [92]. Since then, numerous papers, monographs, book chapters (see for instance [22], [32], [78], [94]) have explored the median operation, analyzing its behavior and proposing new and viable extensions. To compute the output of a median filter, an odd number of sample values are sorted, and the middle or median value is used as the filter output. If the filter length is $N = 2K + 1$, the filtering procedure is denoted as

$$Y(n) = \mathrm{MED}[X(n-K), \cdots, X(n), \cdots, X(n+K)] \quad (2.1)$$

where $X(n)$ and $Y(n)$ are the nth sample of the input and output sequences, respectively.

In most cases, it is reasonable to assume that the signal is of finite length, say L, consisting of samples from $X(0)$ to $X(L-1)$. To be able to filter also the outmost input samples, where parts of the filter window fall outside the input signal, the latter is appended to the required size. The appending of the input signal is commonly performed by replicating the outmost input samples as many times as needed. This appending strategy is referred to as the first and last values carry-on appending strategy, in the literature [29], and will be used throughout the paper.

An immediate extension of the median filter is the class of rank order filters. These filters operate exactly like the median filter except that they put out the ith largest sample in the window, the median being a special case

$$\begin{aligned} Y(n) = &\, i\text{th largest value of the set} \\ &[X(n-K), \cdots, X(n), \cdots, X(n+K)]. \end{aligned} \quad (2.2)$$

Two special rank order filters are the minimum and the maximum filters which output the minimum value and the maximum value in the window, respectively. These two operators are the fundamental operations in discrete morphological filtering, with flat structuring elements, namely erosion and dilation [34], [62], [63], [84].

Another simple extension of the median filter is the recursive median filter [69]. The recursive median of window with $2K+1$ is defined by replacing some of the input samples in (2.1) by previously derived output samples as follows:

$$\begin{aligned} Y(n) = &\,\mathrm{MED}[Y(n-K), Y(n-K+1), \cdots, \\ &Y(n-1), X(n), \cdots, X(n+K)]. \end{aligned} \quad (2.3)$$

In recursive median filtering, the filtering operation is performed "in-place" so that the output of the filter replaces the old input value before the filter window is moved to the next position. With the same amount of operations, the recursive median filter usually provides better smoothing capability than the nonrecursive filter, at the expense of increased distortion.

Frequency analysis and impulse response have no meaning in median and rank order filtering; the impulse response of a median filter is zero! As a result, new tools had to be developed to analyze and characterize the behavior of these nonlinear filters, deterministically and statistically. The basic descriptor of the deterministic properties of median filters is their root signal set, that is a set of signals which are invariant to further filtering. The basic statistical descriptor of median filters is the set of output distributions which are used to study the noise attenuation properties of median filters. These properties of median filters are briefly reviewed next.

1) Deterministic Properties of Median Filters: Root Signals: Since the output of the median filter is always one of the input samples, it is conceivable that certain signals could pass through the median filter unaltered. This has been shown to hold for median and many median-based filters. Since these signals define the signature of a filter, they are referred to as root signals. For a median filter of length $N = 2K + 1$, this means that

$$X(n) = \text{MED}[X(n-K), \cdots, X(n), \cdots, X(n+K)]. \quad (2.4)$$

If the above condition is satisfied for all n, $X(n)$ is called a root signal of that particular median filter.

The fact that some signals are invariant to median filtering offers interesting possibilities. In noise filtering, a problem is often how to preserve some desired signal features while attenuating noise. An optimal situation would arise if the filter could be designed so that the desired features were invariant to the filtering operation and only noise would be affected. Since the median filter is a nonlinear filter, and the superposition principle does not apply, this of course can never be fully obtained. However, when a signal consists of constant areas and stepwise changes between these areas, a similar effect is achieved. Noise will be attenuated, but stepwise changes will remain. In image filtering, a common approach is to design a median filter such that certain image patterns, e.g., lines, are root signals and thus not disturbed by the filtering operation.

The characterization of root signals is based on local signal structures, as defined in [33] and [93]. Those used in [33] are summarized here for a median filter of window width $N = 2K + 1$.

A *constant neighborhood* is a region of at least $K + 1$ consecutive identically valued samples.

An *edge* is a monotonically rising or falling set of samples surrounded on both sides by constant neighborhoods of different values.

An *impulse* is a set of at most K samples whose values are different from the surrounding regions and whose surrounding regions are identically valued constant neighborhoods.

An *oscillation* is any signal structure which is not part of a constant neighborhood, an edge or an impulse.

Note that the definition of these concepts is tied to the filter length, e.g., a constant neighborhood of a length N filter is also a constant neighborhood for all filters of lengths less than N.

With these concepts, Gallagher and Wise proved that an arbitrary finite length signal is a median root if it consists of constant neighborhoods and edges only [33] (see also Tyan [93]). They also proved that repeated median filtering of any finite length signal will result in a root signal after a finite number of passes. This property of median filters is very significant and is called the "convergence property". The tightest known bound on the number of passes of the median filter necessary to reach a root can be found in [100]: if the filter window width is $2K + 1$ and the signal has length L, then at most

$$3\left\lceil \frac{L-2}{2(K+2)} \right\rceil$$

passes of the filter are required to produce a root signal. This bound is rather conservative in practice. Typically, after 5-10 filterings only slightly, if any, changes take place.

All rank order filters possess the convergence property of the median filter. However, all root signals of rank order filters, other than the median filter, are constant signals, i.e., trivial. Recursive median filters do also possess the convergence property. Furthermore, they are idempotent, i.e., they produce root signals after a single pass. A recursive median filter has exactly the same set of root signals as the (nonrecursive) median filter of the same window width, but a given input signal may be filtered to different roots by the two filters. Root signals of median related filters have been used for speech and image coding in [3] and [4], [96], [97].

2) Statistical Properties of Median Filters: Median filters are used as robust smoothers and the filtering results are often evaluated in a qualitative way. The performance of the filters depends on how well it can suppress undesired parts of the signal and, equally important, how well desired information is retained. There is no general numerical measure to combine these two properties and thus evaluate the performance of the filter. In the case of linear filters, the two objectives can be achieved if and only if the noise and the signal occupy different frequency bands.

The recently developed tools make the analysis of the statistical behavior of median filters possible. Assuming that the samples inside the filter window are a combination of a constant (desired) signal corrupted by some additive white noise, as done often in the analysis of linear filters, one can evaluate the output distribution of the filter and hence estimate the output variance. Filter optimization is possible based on this approach, but yields even better results when combined with other types of constraints (e.g., structural constraints on the filter's behavior [27]) and applied to extensions of median filters which allow much more design flexibility than just the filter length. We shall first review these statistical properties of median filters and later those of weighted median filters.

Let the input signal of a median filter be white noise, modeled here by independent identically distributed random variables with mean μ and variance σ^2. The corresponding distribution and density functions are denoted by $\Phi(t)$ and $\phi(t)$, respectively. Results from probability theory [60] give the distribution, $\Psi_{med}(t)$, and the density, $\psi_{med}(t)$, of the median filter output as

$$\Psi_{med}(t) = \sum_{i=K+1}^{N} \binom{N}{i} \Phi(t)^i (1 - \Phi(t))^{N-i} \quad (2.5)$$

and

$$\psi_{med}(t) = \frac{N!}{K!K!} \phi(t) \Phi(t)^K (1 - \Phi(t))^K, \quad (2.6)$$

respectively. These two expressions provide the basis for quantitative analysis of the noise attenuation of median filters. Usually, numerical methods must be used to apply these equations.

Under some general assumptions, see [20], the asymptotic distribution (for large N) of the median filter output, when the input is white noise, is normal with mean μ_{med} and variance σ_{med}^2,

$$\mu_{med} = t_{0.5} \quad (2.7)$$

$$\sigma_{med}^2 = \frac{1}{4N[\phi(t_{0.5})]^2} \quad (2.8)$$

TABLE I
ASYMPTOTIC (FOR LARGE N) OUTPUT VARIANCES OF THE AVERAGE AND THE MEDIAN FILTERS OF LENGTH N FOR DIFFERENT INPUT NOISE DISTRIBUTIONS. INPUT IS WHITE NOISE WITH ZERO MEAN AND σ^2 VARIANCE

Noise probability density	Average	Median
Uniform $\phi(x) = \begin{cases} \frac{1}{\sqrt{12\sigma^2}}, & -\sqrt{3\sigma^2} \leq x \leq \sqrt{3\sigma^2} \\ 0, & \text{otherwise} \end{cases}$	$\frac{\sigma^2}{N}$	$\frac{3\sigma^2}{N+2}$
Gaussian $\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}$	$\frac{\sigma^2}{N}$	$\frac{\pi\sigma^2}{2N}$
Laplacian $\phi(x) = \frac{1}{\sqrt{2\sigma^2}} e^{\frac{-\sqrt{2}\lvert x\rvert}{\sigma}}$	$\frac{\sigma^2}{N}$	$\frac{\sigma^2}{2N}$

where $t_{0.5}$ is defined by $\Phi(t_{0.5}) = 0.5$. This result explains the robust noise smoothing properties of the median filter. Regardless of the input distribution, the median operation produces an unbiased estimate of the distribution median $t_{0.5}$. Moreover, the estimate is always consistent, i.e., $\lim_{N \to \infty} \sigma_{med}^2 = 0$. The output variance does not depend on the input variance directly, but on $\phi(t_{0.5})$ instead. With heavy-tailed or impulsive noise distributions, the variance of the distribution grows with the amplitude of the impulses, whereas $\phi(t_{0.5})$ does not necessarily change. Note that the sample mean does not possess this property.

Table I lists asymptotic output variances for some common probability distributions. The results for the average filter and the median filter, for uniform noise, are valid for all values of N. Note, that in the case of Laplacian noise distribution, the asymptotic variance of the median operation is half of that of the sample average, i.e., the sample average is 3 dB worse than the median. Also note that for Gaussian noise, the median is only about 2 dB worse than the average.

The good performance of the median filter with Laplacian distributed noise is due to an important optimality property of the median operation. The sample median is the maximum likelihood estimate for the location parameter of the Laplacian density. This property shows that the median filter is optimal in the mean absolute error sense for attenuating double exponentially distributed noise. This analogy results from alternate definitions for the average and the median of N values $X_1, \cdots, X_N$. Consider the expression

$$L(\beta) = \sum_{i=1}^{N} |X_i - \beta|^\gamma. \tag{2.9}$$

The median of $X_1, \cdots, X_N$ can be defined as the value β minimizing $L(\beta)$ in (2.9) when $\gamma = 1$. The definition always produces one of the samples X_i as a result. The sample average minimizes the same expression when $\gamma = 2$.

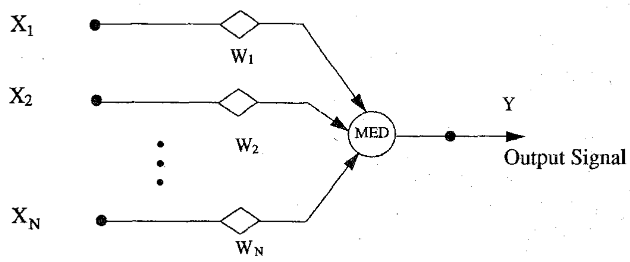


Fig. 1. Weighted median filter.

B. Weighted Median Filters

The weighted median (WM) filter was first introduced as a generalization of the standard median filter, where a nonnegative integer weight is assigned to each position in the filter window [13], [51]. In this subsection, we give two alternative definitions of WM filters and discuss the analogy between WM filters and linear FIR filters.

As shown in Fig. 1, the structure of a WM filter is quite similar to that of a linear FIR filter. For real-valued signals, WM filters can be defined in two different but equivalent ways. The first definition can be used in the common case of positive integer weights.

Definition 2.1: For the discrete-time continuous-valued input vector $\underline{X} = [X_1, X_2, \cdots, X_N]$, the output Y of the WM filter of span N associated with the integer weights

$$\underline{W} = [W_1, W_2, \cdots, W_N] \tag{2.10}$$

is given by

$$Y = \text{MED}[W_1 \diamond X_1, W_2 \diamond X_2, \cdots, W_N \diamond X_N] \tag{2.11}$$

where MED[$\cdot$] denotes the median operation and $\diamond$ denotes duplication, i.e.,

$$K \diamond X = \overbrace{X, \cdots, X}^{K \text{ times}}. \tag{2.12}$$

This filtering procedure can be stated as follows: sort the samples inside the filter window, duplicate each sample X_i to the number of the corresponding weight W_i and choose the median value from the new sequence.

Example 2.1: Consider a length 5 WM filter with integer weights [1, 2, 3, 2, 1]. Now apply the filter to the following sequence so that the window is centered at the sample value 8,

$$\underline{X} = [-1, 5, 8, 11, -2].$$

After sorting and duplication, the samples inside the filter window are 11, 11, 8, 8, 8, 5, 5, -1, -2. The filter output is $Y = 8$, whereas, the 5-point median would have produced the result $Y = 5$.

The second definition of the WM operation also allows positive noninteger weights to be used.

Definition 2.2: The weighted median of $\underline{X}$ is the value β minimizing the following expression

$$L(\beta) = \sum_{i=1}^{N} W_i |X_i - \beta|. \tag{2.13}$$

TABLE II
WEIGHTED MEDIAN FILTERS FOR WINDOW WIDTH 1 THROUGH 5

Window width	Weighted median
1	[1]
2	
3	[1,1,1]
4	[2,1,1,1]
5	[1,1,1,1,1], [3,2,2,1,1] [2,2,1,1,1], [3,1,1,1,1]

Here, β is guaranteed to be one of the samples X_i because $L(\beta)$ is piecewise linear and convex, if $W_i \geq 0$ for all i.

The output of the WM filter for real positive weights can be calculated as follows: sort the samples inside the filter window; add up the corresponding weights from the upper end of the sorted set until the sum just exceeds half of the total sum of weights (i.e., $\geq \frac{1}{2}\sum_{i=1}^{N} W_i$); the output of the WM filter is the sample corresponding to the last weight added.

Example 2.2: Consider a length 5 WM filter with weights $\underline{W}$ and an input $\underline{X}$:

$$\underline{W} = [0.1, 0.2, 0.3, 0.2, 0.1], \quad \underline{X} = [1, 5, 8, 11, 2].$$

After sorting, we get the sorted input set with the corresponding weights

$$\begin{array}{ccccc} 0.2 & 0.3 & 0.2 & 0.1 & 0.1 \\ 11 & 8 & 5 & 2 & 1 \end{array}$$

Starting from the left, add the weights until the sum reaches or exceeds 0.45. The first weight is smaller than 0.45, so add the next weight. The sum is now 0.5 exceeding 0.45. The weighted median is therefore 8.

One fact must be emphasized that there is only a finite number of WM filters for a given finite length window, even though there exists an infinite number of real weight vectors [14], [67]. As already indicated by the above two examples, multiplying the weights by a positive constant does not change the filter. For a detailed treatment of WM filters, particularly with real weights, interested readers are referred to [67]. In particular, it was shown that any real-weight WM corresponds (is identical) to an integer weight WM. Table II lists all WM filters for window width 1 through 5, excluding weight permutations. The number of WM filters grows very rapidly with the window size. For instance, there are 114 WM filters for window width 7, and 172 958 WM filters for window width 9.

As the median filter is extended to a rank order filter, the WM filter can be extended to a weighted order statistic (WOS) filter [98].

Definition 2.3: For the discrete-time continuous-valued input $\underline{X} = [X_1, X_2, \cdots, X_N]$, the output Y of the WOS filter of span N associated with the weights $\underline{W}$ and the threshold T_h is given by

$$Y = T_h \text{ :th largest element of the set } [W_1 \diamond X_1, W_2 \diamond X_2, \cdots, W_N \diamond X_N]. \quad (2.14)$$

Note that the WM filter is a WOS filter with the same weights and a threshold $T_h = \frac{1}{2}\sum_{i=1}^{N} W_i$ ($T_h = \frac{1}{2}(1 + \sum_{i=1}^{N} W_i)$ for integer weights). The WOS filter has some special properties compared to the WM filter. For example, some WOS filters called asymmetric median filters can eliminate negative impulses (or positive impulses), but preserve positive impulses (or negative impulses) [98].

WM filters can also operate recursively. The output of a recursive WM filter is given by

$$Y(n) = \text{MED}[W_{-K} \diamond Y(n-K+1), \cdots, W_{-1} \diamond Y(n-1), \mathbf{W}_0 \diamond X(n), \cdots, W_K \diamond X(n+K)]. \quad (2.15)$$

Recursive WM filters have a certain similarity to linear IIR filters. Although useful in practice, they are in general difficult to handle theoretically.

1) Analogy Between WM Filters and Linear FIR Filters: As the median filter is analogous to the average filter, there exists an interesting analogy between WM filters and linear FIR filters. Let $H_i, i = 1, 2, \cdots, N$, be the impulse response coefficients of a 1-D FIR filter of size N. If $H_i \geq 0$ for all i, $1 \leq i \leq N$ and $\sum_{i=1}^{N} H_i = 1$ then the output Y of the FIR filter can be calculated also as

$$Y = \sum_{i=1}^{N} H_i X_i = \text{average}\{W_1 \diamond X_1, W_2 \diamond X_2, \cdots, W_N \diamond X_N\} \quad (2.16)$$

where W_i, $1 \leq i \leq N$, are positive integers such that $H_i = W_i / \sum_{i=1}^{N} W_i$. If we replace the average in (2.16) with the median then the definition of the WM filter is obtained.

As mentioned earlier, the median is the maximum likelihood estimate of the signal level in the presence of uncorrelated additive biexponentially distributed noise; while, the arithmetic mean is that for Gaussian distributed noise. This means that the sample median of $X_i, i = 1, \cdots, N$, is the value β which minimizes (2.17) when $\gamma = 1$ and the sample mean is the value β which minimizes (2.17) when $\gamma = 2$:

$$L(\beta) = \sum_{i=1}^{N} |X_i - \beta|^{\gamma}. \quad (2.17)$$

If we introduce weights W_i in (2.17) as follows:

$$L(\beta) = \sum_{i=1}^{N} W_i |X_i - \beta|^{\gamma} \quad (2.18)$$

then, in the case $\gamma = 2$, the value β minimizing (2.18) can be expressed as

$$\beta = \sum_{i=1}^{N} W_i X_i \Big/ \sum_{i=1}^{N} W_i \quad (2.19)$$

which can be taken as a normalized FIR filter, i.e., a weighted average. Likewise, if $\gamma = 1$, the value β minimizing (2.18) is

$$\beta = \text{MED}[W_1 \diamond X_1, \cdots, W_N \diamond X_N]$$

i.e., a weighted median.

In the sequel, other interesting analogies will be drawn between WM filters and FIR filters. The similarity between them may suggest that WM filters would play as a fundamental role among median type filters as FIR filters among linear filters.

III. Threshold Decomposition and Weighted Median Filters

In 1984, a powerful theoretical tool called threshold decomposition for analyzing rank order based filters was developed by Fitch *et al* [26]. Using this tool, the analysis of rank order based filters reduces to studying their effects on binary signals.

A. Threshold Decomposition and Median Filtering in the Binary Domain

Threshold decomposition of an M-valued signal vector $\underline{X}$, where the samples are integer-valued, $0 \leq X_i < M$, means decomposing it into $M-1$ binary signal vectors, $\underline{x}^1, \underline{x}^2, \cdots, \underline{x}^{M-1}$, according to the following rule

$$x_i^m = T^m(X_i) = \begin{cases} 1, & \text{if } X_i \geq m; \\ 0 & \text{else.} \end{cases} \tag{3.1}$$

This thresholding scheme can be applied also to noninteger signals or, in general, to any signal which is quantized to a finite number of arbitrary levels.

Example 3.1: Consider a five-valued integer signal vector

$$\underline{X} = [0,0,2,3,3,3,2,1,0,4,0,1,0,0].$$

The threshold decomposition of this signal vector into four binary signal vectors results in

$$\underline{x}^4 = [0,0,0,0,0,0,0,0,0,1,0,0,0,0],$$

$$\underline{x}^3 = [0,0,0,1,1,1,0,0,0,1,0,0,0,0],$$

$$\underline{x}^2 = [0,0,1,1,1,1,1,0,0,1,0,0,0,0],$$

$$\underline{x}^1 = [0,0,1,1,1,1,1,1,0,1,0,1,0,0].$$

The original multivalued signal samples X_i's can be reconstructed from the threshold levels by adding these together

$$X_i = \sum_{m=1}^{M-1} x_i^m, \tag{3.2}$$

where X_i is the ith sample of $\underline{X}$.

A very important property of median filters was observed by Fitch *et al.* [26]: Applying a median filter to an M-valued signal is equivalent to decomposing the signal to $M-1$ binary threshold signals, filtering each binary signal separately with the corresponding binary median filter, and then adding the binary output signals together. That is, if the multivalued input samples $\underline{X} = [X_1, \cdots, X_N]$ are inside the median filter window, then the multilevel median output

$$Y = \mathrm{MED}[\underline{X}] = \sum_{m=1}^{M-1} \mathrm{MED}[\underline{x}^m], \tag{3.3}$$

where

$$\underline{x}^m = [x_1^m, \cdots, x_N^m]. \tag{3.4}$$

This is a *weak superposition* property which holds for the *nonlinear* median operation. The importance of the property arises from the fact that binary signals are much easier to analyze than multivalued signals. The median operation on binary samples reduces to a simple Boolean operation. For example, if the samples inside a length-three binary median filter are denoted by x_1, x_2, and x_3; then the median filter is equivalent to the Boolean expression $x_1x_2 + x_2x_3 + x_1x_3$ where $+$ denotes the logical OR operation and x_ix_j is the logical AND operation of variables x_i and x_j.

The approach of threshold decomposition and binary filtering has led to a general class of nonlinear filters called stack filters [98], whose output is given by

$$S(\underline{X}) = \sum_{m=1}^{M-1} f(\underline{x}^m) \tag{3.5}$$

where $f(.)$ is a Boolean function which satisfies the stacking property. A Boolean function f possesses the stacking property if whenever two input vectors $\underline{u}$ and $\underline{v}$ stack, i.e., $u_i \geq v_i$ for each $i \in \{1, \cdots, N\}$, then also their respective outputs stack, $f(\underline{u}) \geq f(\underline{v})$. A binary function possesses the stacking property if and only if it can be expressed as a Boolean expression which contains no complements of input variables [35]. Such functions are called positive Boolean functions (PBF's).

B. Weighted Median Filters and Linearly Separable Selfdual PBF

WM filters belong to the class of stack filters. In other words, any WM filter can be represented by a PBF in the binary domain. For example, the WM filter $\underline{W} = [1,2,3,2,1]$ corresponds to the following PBF in the binary domain

$$f(x_1,x_2,x_3,x_4,x_5) =$$
$$x_2x_3 + x_3x_4 + x_1x_3x_5 + x_1x_2x_4 + x_2x_4x_5.$$

However, a stack filter defined by a PBF $f(\underline{x})$ is a WM filter if and only if $f(\underline{x})$ is selfdual and linearly separable. In the following we discuss the concepts of selfduality and linear separability.

Definition 3.1: The dual of a Boolean function $f(\underline{x})$ is given by $f^D(x_1, \cdots, x_N) = \overline{f}(\overline{x}_1, \cdots, \overline{x}_N)$, where $\overline{x}_i$ denotes the complement of x_i. A Boolean function $f(\underline{x})$ is selfdual if $f(\underline{x}) = f^D(\underline{x})$.

Denote the onset of $f(.)$ by $f^{-1}(1) = \{\underline{x} \in \{0,1\}^N | f(\underline{x}) = 1\}$, and the offset of $f(.)$ by $f^{-1}(0) = \{\underline{x} \in \{0,1\}^N | f(\underline{x}) = 0\}$ [85]. Then $f(\underline{x})$ is a selfdual Boolean function if and only if the following holds [64]:

$$\underline{x} \in f^{-1}(1) \Longleftrightarrow \overline{\underline{x}} \in f^{-1}(0). \tag{3.6}$$

This equation states that selfdual Boolean functions satisfy a symmetry property. Indeed, as is shown in the next section, stack filters defined by selfdual PBF's are statistically unbiased in the sense of median, i.e., the median of the input is also the median of the output in the case of i.i.d. input variables.

In addition to being selfdual, PBF's which define WM filters must also be linearly separable. These Boolean functions can be realized by threshold logic gates efficiently [43], [57], [64], [85].

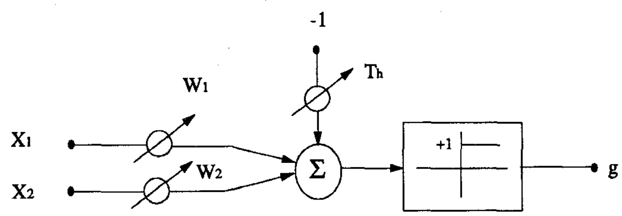


Fig. 2. A two-input threshold logic gate (or neuron).

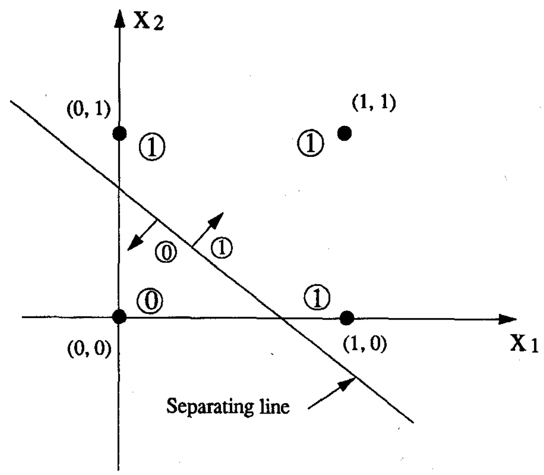


Fig. 3. Linear separability of a Boolean function.

Definition 3.2: A Boolean function $g(\underline{x})$ is said to be linearly separable if it can be expressed in the form

$$g(\underline{x}) = U(\underline{W}\underline{x}^T - T_h), \tag{3.7}$$

where ^{T} denotes transpose, $U(X)$ is the unit step function

$$U(X) = \begin{cases} 1, & \text{if } X \geq 0; \\ 0 & \text{else,} \end{cases} \tag{3.8}$$

$\underline{x}$ is a binary vector whose entries are understood here to be real 1's and 0's, T_h is a threshold value, and $\underline{W}$ is a set of positive weights.

Fig. 2 shows a two input threshold logic gate, and Fig. 3 shows all possible binary inputs for a two-input threshold logic gate in an input vector space. In this space, the coordinate axes are the components of the input vector. It is worth mentioning that threshold logic gates are also called neurons in neural networks [103].

When linearly separable Boolean functions are positive, their weights W_i and the threshold T_h can be restricted to be positive real numbers without loss of generality, based on the following theorem [85].

Theorem 3.1: For a linearly separable PBF, there exists at least one threshold logic gate whose weights and threshold are nonnegative. Conversely, the threshold logic gate with nonnegative weights and threshold represents a linearly separable PBF.

The WM filter is a special case of a WOS filter with $T_h = \sum_{i=1}^{N} W_i/2$. The PBF's corresponding to WM filters can be expressed as

$$f(\underline{x}) = U(\underline{W}\underline{\xi}^T), \tag{3.9}$$

where

$$\underline{\xi} = [\xi_1, \xi_2, \cdots, \xi_N], \tag{3.10}$$

$$\xi_i = 2x_i - 1. \tag{3.11}$$

Using the threshold decomposition and linearly separable selfdual PBF's, we have another definition of WM filters.

Definition 3.3: For the discrete-time M-valued input $\underline{X} = [X_1, X_2, \cdots, X_N]$, the output Y of the WM filter of span N associated with the nonnegative weights $\underline{W} = [W_1, W_2, \cdots, W_N]$ is given by

$$Y = \sum_{m=1}^{M-1} U(\underline{W}\underline{\xi}^{mT}). \tag{3.12}$$

The definition of WM filters in the binary domain is extremely useful for the theoretical analysis of WM filters, for many properties of WM filters are easily derived based on this definition. In particular, optimal WM filtering algorithms under the MSE and the MAE criteria were developed using this definition [111], [112], [115].

C. Conversions Between WM Filters and PBF's

Conversion between WM filters and PBF's means finding the PBF corresponding to a WM filter and vice versa [64], [67], [85], [120], [123]. Using PBF to represent WM filters proved to simplify the analysis of their properties. Some statistical properties of WM filters (i.e., output probability density functions) were derived in [120] using PBF representation. Certain deterministic properties of stack filters (root structures, convergence rates) were found using this form of representation [29], [98]. However, the threshold decomposition architecture may not be advantageous for implementation purposes due to the large number of threshold signals (255 binary signals for 8-bit samples). In this case, we may need to convert the PBF to the corresponding WM filter (if the PBF is selfdual and linearly separable). Other advantages of converting a PBF to a WM (whenever this is possible) include reducing the number of filter parameters; e.g. a window width 17 PBF may require a huge number of minterms; whereas a same size WM needs only 17 parameters.

In order to find a PBF corresponding to a WM filter, we consider a binary signal that is filtered with a WM filter. By Definition 3.2 or 3.3, the output is 1 if there is a combination of input variables with value 1 such that the sum of their corresponding weights is greater than the threshold value ($T_h = \sum_{i=1}^{N} W_i/2$ for WM filters). The PBF corresponding to the WM filter can be found by listing all such combinations treating them as logical products and summing the products logically. The resulting disjunctive form can then be minimized using rules of Boolean algebra. Sheng [85] described an optimal algorithm to find the Minimum Sum-of-Products (MSP) form of a PBF corresponding to a threshold logic gate (WOS filter in the integer domain).

The problem of finding the WM filter which is equivalent to a stack filter defined by a given PBF $g(\underline{x})$ is more complicated than its converse. We have to find weights W_i such that the function $g(\underline{x})$ is realized by (3.9) (assuming it is possible).

Since $g(.)$ is a PBF,

$$\underline{a} \in g^{-1}(1) \Rightarrow \underline{b} \in g^{-1}(1), \forall\, \underline{b} \geq \underline{a} \tag{3.13}$$

and

$$\underline{a} \in g^{-1}(0) \Rightarrow b \in g^{-1}(0), \forall\, \underline{b} \leq \underline{a}. \tag{3.14}$$

That is, the onset of $g(.)$, $g^{-1}(1)$, is uniquely specified by its minimal elements, say, $\underline{a}_1, \underline{a}_2, \cdots, \underline{a}_r$, which correspond to product terms of the MSP form of $g(.)$. Likewise the offset of $g(.)$, $g^{-1}(0)$ is uniquely specified by its maximal elements, say, $\underline{b}_1, \cdots, \underline{b}_s$. The maximal elements of $g^{-1}(0)$ can be found using the following lemma.

Lemma 3.1 [64]: Let g be as above and g^D the dual of g. Then $\underline{a}$ is a minimal element of $(g^D)^{-1}(1)$ iff $\overline{\underline{a}}$ (the complement of $\underline{a}$) is a maximal element of $g^{-1}(0)$.

Let $\underline{W} = [W_1, \cdots, W_N]$ where $W_i \geq 0$, ($i = 1, \cdots, N$). Let also $\underline{x} = [x_1, \cdots, x_N]$, where x_i, ($i = 1, \cdots, N$) are binary variables which are handled here as real 1's and 0's. If $\underline{W}\ \underline{a}_j^T \geq T_h$, ($j = 1, \cdots, r$), ($\underline{a}_1, \cdots, \underline{a}_r$) being the minimal elements of $g^{-1}(1)$, then $\underline{W}\underline{x}^T \geq T_h$ for any $\underline{x} \in g^{-1}(1)$. Likewise, if $\underline{W}\underline{b}_j^T < T_h$, ($j = 1, \cdots, s$), ($\underline{b}_1, \cdots, \underline{b}_s$) being the maximal elements of $g^{-1}(0)$, then $\underline{W}\underline{x}^T < T_h$ for any $\underline{x} \in g^{-1}(0)$.

Based on Lemma 3.1, the problem of finding a WOS filter corresponding to a PBF $g(x_1, \cdots, x_N)$ can be solved by finding the values $W_1, \cdots, W_N$ and T_h satisfying the inequalities derived using the MSP forms of g and g^D. Each logical product (say) $x_a \cdots x_b$, in the MSP form of g generates an inequality

$$W_a + \cdots + W_b \geq T_h. \tag{3.15}$$

Likewise, each logical product (say) $x_c \cdots x_d$, in the MSP form of g^D generates an inequality

$$W_e + \cdots + W_f < T_h, \tag{3.16}$$

where $W_e + \cdots + W_f$ is the sum of weights corresponding to the Boolean variables $x_e, \cdots, x_f$ not existing in the logical product $x_c \cdots x_d$. For integer weights, inequality (3.16) can be further transformed to

$$W_e + \cdots + W_f \leq T_h - 1. \tag{3.17}$$

The minimal WOS filter equivalent to the stack filter defined by the PBF g is found by minimizing the sum of all weights subject to all constraints (3.15) and (3.17). This is a standard linear programming problem. The formulation of the linear programming problem to find a threshold gate realization of a PBF can be found in [57], [64], [67], [85].

The problem of finding the WM filter corresponding to a PBF can be solved in a similar way as that of finding a WOS filter; we only need to set the threshold $T_h = (1 + \sum W_i)/2$. However, special properties of WM filters can be used to make the linear programming problem simpler. All WM filters correspond to selfdual PBF's. In case of a selfdual PBF ($g = g^D$) the dualization phase described above can be omitted.

D. Cascaded Weighted Median Filters

Like in linear filtering, it is common to use cascades of median based filters for, e.g. analysis and filtering purposes [7]–[9], [37], [120]. Theoretical analysis of WM filter cascades is difficult. Fortunately, cascades of WM filters can often be represented by a single WM filter.

By cascade connection of filters F and G we mean that the original input signal $\mathbf{X}$ is filtered by filter F which produces an intermediate signal $\mathbf{Y}$. This in turn is filtered by filter G which produces the output signal $\mathbf{Z}$. Cascaded filters F and G can be represented as a single filter H which produces the output $\mathbf{Z}$ directly from the input $\mathbf{X}$. The maximum window length of filter H is $N_H = N_F + N_G - 1$, where N_F and N_G are window lengths of filters F and G, respectively. For the cascade connection of filters F and G, we use the notation

$$[W_{-L_F}, \cdots, \mathbf{W_{0_F}}, \cdots, W_{R_F}]$$
$$[W_{-L_G}, \cdots, \mathbf{W_{0_G}}, \cdots, W_{R_G}],$$

(F is applied first). For equivalence of filters F and G, we use the notation

$$[W_{-L_F}, \cdots, \mathbf{W_{0_F}}, \cdots, W_{R_F}] = [W_{-L_G}, \cdots, \mathbf{W_{0_G}}, \cdots, W_{R_G}],$$

which means that the filters correspond to the same PBF (weights need not be equal). For P passes of filter F, we use the notation

$$[W_{-L_F}, \cdots, \mathbf{W_{0_F}}, \cdots, W_{R_F}]^P.$$

It is defined that

$$[W_{-L_F}, \cdots, \mathbf{W_{0_F}}, \cdots, W_{R_F}]^0 = [\mathbf{1}].$$

The exact procedure to realize a cascade of two WM filters as a single WM filter (if it is possible) is described in [120]. The two original filters are first expressed as PBF's f and g. The next step is to find the composition of f and g and reduce it to its minimum sum of product (MSP) form. The WM filter coefficients corresponding to the new PBF must then be searched using linear programming. The resulting PBF is not always linearly separable in which case no solutions can be found. The result is always a stack filter defined by a selfdual PBF. Several observations concerning cascaded WM filters are presented in [120], most stemming from results of compositions of PBF.

Fact 1: The cascade operation is not commutative.

Fact 2: The cascade operation is associative.

Fact 3: Different cascade combinations can give the same result; however, the intermediate results need not be the same, for example

$$[1, 2, \mathbf{1}, 2, 1][1, 2, \mathbf{3}, 2, 1] = [1, 1, \mathbf{1}, 1, 1][1, \mathbf{1}, 1] = [1, 2, 3, \mathbf{3}, 3, 2, 1]$$

Fact 4: The P-times iterated 3-point median filter has been found to be

$$\begin{aligned} [1, \mathbf{1}, 1]^P &= [\alpha_0, \alpha_1, \cdots, \alpha_P, \boldsymbol{\beta}_{\mathbf{P}}, \alpha_P, \cdots, \alpha_1, \alpha_0] \\ \alpha_i &= 2\alpha_{i-1} + \alpha_{i-2}, \alpha_0 = 0, \alpha_1 = 1; \\ \beta_i &= 2\beta_{i-1} + \beta_{i-2}, \beta_0 = 1, \beta_1 = 1. \end{aligned} \tag{3.18}$$

Some other cascade WM filters are shown in Table III [120].

TABLE III
SOME CASCADE WM FILTERS

$[1,\mathbf{1},1]^2$	=	$[1,2,\mathbf{3},2,1]$
$[1,\mathbf{1},1]^3$	=	$[1,2,5,\mathbf{7},5,2,1]$
$[1,\mathbf{1},1]^4$	=	$[1,2,5,12,\mathbf{17},12,5,2,1]$
$[1,\mathbf{1},1]^5$	=	$[1,2,5,12,29,\mathbf{41},29,12,5,2,1]$
$[1,\mathbf{1},1]^6$	=	$[1,2,5,12,29,70,\mathbf{99},70,29,12,5,2,1]$
$[1,1,\mathbf{1},1,1][1,\mathbf{1},1]^1$	=	$[1,2,3,\mathbf{3},3,2,1]$
$[1,1,\mathbf{1},1,1][1,\mathbf{1},1]^2$	=	$[1,3,7,10,\mathbf{11},10,7,3,1]$
$[1,1,\mathbf{1},1,1][1,\mathbf{1},1]^3$	=	$[1,2,7,16,23,\mathbf{25},23,16,7,2,1]$
$[1,1,\mathbf{1},1,1][1,\mathbf{1},1]^4$	=	$[1,2,5,17,39,56,\mathbf{61},56,39,17,5,2,1]$
$[1,1,\mathbf{1},1,1][1,\mathbf{1},1]^5$	=	$[1,2,5,12,41,94,135,\mathbf{147},135,94,41,12,5,2,1]$
$[1,\mathbf{1},1]^1[1,1,\mathbf{3},1,1]$	=	$[1,1,4,\mathbf{5},4,1,1]$
$[1,\mathbf{1},1]^2[1,1,\mathbf{3},1,1]$	=	$[1,1,4,9,\mathbf{13},9,4,1,1]$
$[1,\mathbf{1},1]^3[1,1,\mathbf{3},1,1]$	=	$[1,1,4,9,22,\mathbf{31},22,9,4,1,1]$
$[1,\mathbf{1},1]^4[1,1,\mathbf{3},1,1]$	=	$[1,1,4,9,22,53,\mathbf{75},53,22,9,4,1,1]$
$[1,\mathbf{1},1]^5[1,1,\mathbf{3},1,1]$	=	$[1,1,4,9,22,53,128,\mathbf{181},128,53,22,9,4,1,1]$
$[1,1,2,\mathbf{3},2]^1$	=	$[1,1,2,\mathbf{3},2]$
$[1,1,2,\mathbf{3},2]^2$	=	$[1,1,2,5,7,12,\mathbf{19},12,2]$
$[1,1,2,\mathbf{3},2]^3$	=	$[1,1,2,5,7,12,31,43,74,\mathbf{117},74,12,2]$

IV. PROPERTIES OF WEIGHTED MEDIAN FILTERS

Like median filters, WM filters are usually characterized by their deterministic and statistical properties. Based on the threshold function representation, Yli-Harja *et al* [120] analyzed the statistical properties of WM filters. The root signal properties of several subclasses of WM filters have been studied by several researchers [31], [37], [87], [99], [109], [122]. Without reference to the threshold decomposition, Prasad *et al* derived a number of deterministic properties of WM filters, which show equivalence between two WM filters [80], and Yang *et al.* derived a simple and intuitive expression of the output distribution of WM filters [104], [105].

A. Root Properties of Weighted Median Filters

Recall that a root of a filter is a signal that is invariant to this filter. The question of whether a WM filter will make a signal of finite length converge to a root in a finite number of passes is crucial to the usefulness of that WM filter because a filter which does not possess this convergence property could have quite a chaotic filtering behavior.

Although the root properties of median filters are well known, root signals of WM filters are much more complicated than those of median filters. The convergence property of WM filters is only known for some special cases [15], [37], [87], [99], [109], [122]. The main reason is perhaps the fact that a WM filter with window width $2K + 1$ can preserve details lasting less than $K + 1$ samples. For example, consider the filter [1, 2, 2, 5, 2, 2, 1]. It is easy to check that the signal [1, 1, 1, 2, 2, 0, 2, 0, 0] is invariant to this filter.

Since symmetric WM filters are widely used in practice to avoid biased effects, we shall confine our discussion on root properties of symmetric WM filters.

Definition 4.1: A WM filter defined by $[W_{-K}, \cdots, W_{-1}, \mathbf{W}_0, W_1, \cdots, W_K]$ is symmetric if

$$W_{-i} = W_i \quad i = 1, 2, \cdots, K. \tag{4.1}$$

Based on the convergence properties of symmetric threshold automata [36], Wendt proved the following theorem.

Theorem 4.1 [99]: Symmetric WM filters with $2K + 1$ variables, which preverve all the roots of median filter of window width $2K+1$, will make any appended input sequence converge to a root, or a cycle of period 2, in a finite number of passes.

For example, the following class of filters obeys the conditions of Theorem 4.1 and makes all appended inputs converge to roots in a finite number of passes.

$$W_i = W_{-i} = K + 1 - i, \quad i = 0, 1, 2, \cdots, K. \tag{4.2}$$

It is well known that the recursive median filter will make any appended input converge to a root in just one pass [69]. One would expect that a recursive WM filter would have a better convergence behavior than its nonrecursive counterpart. Wendt proved the following theorem.

Theorem 4.2 [99]: Recursive symmetric WM filters, which preserve roots of the median filter, will make every appended input signal converge to a root in a finite number of passes.

Note that in Theorem 4.1, the conditions under which symmetric WM filters will converge are not specified. Furthermore, Theorem 4.1 is not valid for the whole class of symmetric WM filters but for a subclass of symmetric WM filters which can preserve the roots of the median filter. The condition of preserving the roots of the median filter was proven redundant in [110]. In other words, any symmetric WM filter will make any appended input sequence converge to a root, or a cycle of period 2, in a finite number of passes. In order to find conditions under which WM filters will converge, researchers have studied the root properties of several subclasses of WM filters [15], [37], [87], [109], [122].

Theorem 4.3 [122]: Symmetric WM filters will make any input signal converge to a root in a finite number of passes if

$$\mathbf{W}_0 \geq \sum_{i=1}^{K} W_i$$

or

$$\mathbf{W}_0 \leq 2 \min\{W_1, W_2, \cdots, W_K\}.$$

This theorem suggests that if a symmetric WM filter does not possess the converge property, then it must satisfy

$$2 \min\{W_1, W_2, \cdots, W_K\} < \mathbf{W}_0 < \sum_{i=1}^{K} W_i.$$

Theorem 4.4 [109]: Symmetric WM filters satisfying the following condition will converge to a root in a finite number of passes

$$\mathbf{W}_0 < 2W_p - 2 \sum_{i=p+2}^{K} W_i$$

where p, $1 \leq p \leq K$, is called the *feature value* of a symmetric WM filter.

In [109], it is shown that this subclass of symmetric WM filters is an important subclass of WM filters. Many optimal WM filters preserving a given length of pulses fall into this subclass. The cardinality of the root signal set of this subclass was also derived in [109].

Center WM filters [87], [88], which give more weight only to the center value of the window, are proven to possess the convergence property.

Theorem 4.5 [87]: Filtering by any center WM filter, any appended finite length signal will converge to a root after a finite number of passes.

Note that no filter subclass in Theorem 4.3 through Theorem 4.5 is a proper subset of any of the others.

Currently, the convergence rate for symmetric WM filters is only known for some special cases [15], [37]. Among these figure the idempotent weighted median filters which need only a single pass over the input signal to filter it to a root, [31], [37]. Two classes of idempotent WM filters have been found by Gabbouj *et al.* [31].

Assume that the sum of the filter weights, denoted by W_s, is odd,

$$W_s = \sum_{i=-K}^{K} W_i. \tag{4.3}$$

Gabbouj *et al.* proved the following theorem.

Theorem 4.6 [31]: Any recursive symmetric WM filter whose $W_s = 2\mathbf{W_0} + 1$, where $\mathbf{W_0} \geq 1$, is idempotent.

WM filters satisfying the hypothesis of Theorem 4.6 are called Class 1 WM filters in [31]. For example, the recursive WM filter with weight vector $\underline{W} = [1, 1, \mathbf{3}, 1, 1]$ is an idempotent filter. When the filter window is positioned at input sample X_i, there are always a number of samples corresponding to nonzero weights inside the filter window. We will denote the set of the indices of these samples as Ω_i. For example, when the filter window is of length 5, say $\underline{W} = [1, 1, \mathbf{3}, 1, 1]$, the index set Ω_i is

$$\Omega_i = \{i-2, i-1, i, i+1, i+2\}. \tag{4.4}$$

Gabbouj *et al.* [31] presented another class of idempotent WM filters, called Class 2 WM filters, which is a proper subset of the first class defined in Theorem 4.6.

Theorem 4.7 [31]: Any WM filter (recursive and nonrecursive) whose window is symmetric, $W_s = 2\mathbf{W_0} + 1$, where $\mathbf{W_0} \geq 1$, and where for each input sample X_i and for all $j \in \Omega_i$, there exists k such that

$$k \in \Omega_i, k \in \Omega_j \tag{4.5}$$

where $j \neq k, j \neq i, k \neq i$, is idempotent.

For example, the WM filter with weight vector $\underline{W} = [1, 1, \mathbf{3}, 1, 1]$ is idempotent. These two classes of idempotent WM filters have the very important property that apart from removing impulses, they have only a minimal effect on the signal. All filters in the first class filter out impulses from input signals, when used in recursive mode. All filters in the second class filter out impulses from the input signals, when used in the recursive and the nonrecursive modes. Among all idempotent filters, the above two classes of filters belong to the class of "gentle" filters. Their noise attenuation consists exclusively of suppressing those " isolated" impulses in the signal. On the other hand, recursive median filters, belong to the class of "strong" filters. They tend to attenuate fine details by reacting relatively slowly to changes in the signal. This may lead, among other things, to edge jitter.

Other root signal properties of recursive WM filters were presented by Han *et al.* [39]. They found a set of conditions on the weights for a recursive WM filter to preserve monotone or locally monotone sequences. Convergence results for the larger class of stack filters have been considered in [28], [29], in which the authors have successfully characterized the convergence properties of some subclasses of stack filters.

B. Statistical Properties of Weighted Median Filters

As a class of robust smoothers, the performance of WM filters is usually described by their noise attenuation properties. These properties depend on the output distribution functions of WM filters expressed in terms of input distributions. The first expression of the output distribution of WM filters was derived by Yli-Harja *et al.* [120]. However, it is neither intuitive nor efficient since it is based on the PBF representation of WM filters in the binary domain. Recently, by introducing a set of parameters called M_i, Yang *et al.* derived useful expressions of the output distribution and output moments of WM filters, which are stated in the following theorems. An optimal WM filtering theory with structural constraints was developed in [104], [107] using these expressions.

Theorem 4.8 [104]: Let the inputs of a WM filter, with window width $N = 2K + 1$, be independent and identically distributed with a common distribution function $\Phi(t)$. The output distribution of the WM filter $\Psi_{wm}(t)$ has the following form

$$\Psi_{wm}(t) = \Psi_{med}(t) + \sum_{i=1}^{K} M_i \left\{ \Phi(t)^i \left(1 - \Phi(t)\right)^{N-i} - \Phi(t)^{N-i} \left(1 - \Phi(t)\right)^i \right\} \tag{4.6}$$

where $\Psi_{med}(t)$ is the output distribution of the median filter with the same window width and M_i's represent the number of different weight vector combinations, in which there are i weights and the sum of these i weights is not less than the threshold T_h, for $i = 1, 2, \cdots, K$.

Example 4.2: Given a WM filter with $\underline{W} = [1, 3, 8, 2, 3]$, then the threshold $T_h = 9$. It is obvious that

$$M_1 = 0,$$

$$M_2 = 4 : \{[8,3]; [8,3]; [8,2]; [8,1]\},$$

$$M_3 = 6 : \{[8,3,3]; [8,3,2]; [8,3,1]; [8,3,2]; [8,3,1]; [8,2,1]\}$$

$$M_4 = 5 : \{[8,3,3,2]; [8,3,3,1]; [8,3,2,1]; [8,3,2,1]; [3,3,2,1]\}$$

$$M_5 = 1 : \{[8,3,3,2,1]\}.$$

The cardinality of the multisets of weights above yields the corresponding M_i. Based on Theorem 4.8, the unbiasness property of WM filters has been proved.

TABLE IV
THE L-TABLE FOR $N = 3$ TO $N = 25$ WITH UNIFORM DISTRIBUTION

N	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}	L_{11}	L_{12}
3	4.014-E1											
5	3.433-E1	5.714-E2										
7	2.395-E1	4.762-E2	9.524-E3									
9	1.711-E1	3.030-E2	7.792-E3	1.732-E3								
11	1.273-E1	1.958-E2	4.662-E3	1.399-E3	3.330-E4							
13	9.812-E2	1.319-E2	2.797-E3	7.992-E4	2.664-E4	6.660-E5						
15	7.789-E2	9.243-E3	1.745-E3	4.525-E4	1.469-E4	5.289-E5	1.371-E5					
17	6.333-E2	6.708-E3	1.135-E3	2.654-E4	7.938-E5	2.835-E5	1.083-E5	2.887-E6				
19	5.254-E2	5.012-E3	7.666-E4	1.622-E4	4.423-E5	1.474-E5	5.670-E6	2.268-E6	6.186-E7			
21	4.433-E2	3.839-E3	5.349-E4	1.030-E4	2.564-E5	7.867-E6	2.855-E6	1.165-E6	4.841-E7	1.345-E7		
23	3.794-E2	3.004-E3	3.834-E4	6.776-E5	1.545-E5	4.359-E6	1.468-E6	5.711-E7	2.447-E7	1.049-E7	2.958-E8	
25	3.287-E2	2.393-E3	2.824-E4	4.594-E5	9.652-E6	2.510-E6	7.823-E7	2.845-E7	1.171-E7	5.230-E8	2.300-E8	6.574-E9

TABLE V
THE L-TABLE FOR $N = 3$ TO $N = 25$ WITH GAUSSIAN DISTRIBUTION

N	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}	L_{11}	L_{12}
3	5.449-E1											
5	4.909-E1	5.395-E2										
7	3.907-E1	4.631-E2	7.639-E3									
9	3.197-E1	3.230-E2	6.372-E3	1.267-E3								
11	2.694-E1	2.312-E2	4.067-E3	1.037-E3	2.297-E4							
13	2.323-E1	1.723-E2	2.638-E3	6.221-E4	1.856-E4	4.408-E5						
15	2.039-E1	1.329-E2	1.783-E3	3.743-E4	1.062-E4	3.528-E5	8.803-E6					
17	1.816-E1	1.055-E2	1.254-E3	2.342-E4	6.029-E5	1.948-E5	6.992-E6	1.810-E6				
19	1.635-E1	8.567-E3	9.129-E4	1.528-E4	3.545-E5	1.055-E5	3.753-E6	1.429-E6	3.808-E7			
21	1.487-E1	7.095-E3	6.843-E4	1.035-E4	2.171-E5	5.888-E6	1.954-E6	7.495-E7	2.992-E7	8.153-E8		
23	1.362-E1	5.971-E3	5.256-E4	7.240-E5	1.383-E5	3.421-E6	1.045-E6	3.779-E7	1.539-E7	6.382-E8	1.771-E8	
25	1.256-E1	5.094-E3	4.123-E4	5.210-E5	9.117-E6	2.066-E6	5.799-E7	1.947-E7	7.549-E8	3.229-E8	1.382-E8	3.895-E9

Corollary 4.1 [104]: If the i.i.d. input distribution $\Phi(t)$ is symmetric with respect to its mean m, then $\Psi_{wm}(t)$ is also symmetric w.r.t. m, i.e., the output mean m_{wm} equals the input mean m. That is, WM filters are unbiased estimators of the mean.

In fact, it has been shown that all stack filters which are defined by selfdual PBF's are statistically unbiased [120]. In practice, the unbiasedness of stack filters is a typical constraint.

The output moments of WM filters can be obtained using their output distribution function.

Theorem 4.9 [104]: Given a WM filter with window width $N = 2K + 1$, for i.i.d. inputs with common distribution function $\Phi(t)$ and density function $\phi(t)$. The output γ-norm central moment σ_{wm}^{γ} of the WM filter can be expressed as:

$$\sigma_{wm}^{\gamma} = \sigma_{med}^{\gamma} + \sum_{i=1}^{K} M_i L_i(N, \Phi, \gamma) \tag{4.7}$$

where σ_{med}^{γ} is the γ-norm central moment of the standard median filter with the same window size

$$L_i(N, \Phi, \gamma) = \int_{-\infty}^{+\infty} \Gamma(\Phi(t))|t - m_{wm}|^{\gamma}\phi(t)dt \geq 0 \tag{4.8}$$

for symmetric $\Phi(t)$, $\gamma \geq 0$, $i = 1, \cdots, K$, and

$$\begin{aligned}\Gamma(\Phi(t)) = &\Big(i - N\Phi(t)\Big)\Phi(t)^{i-1}\Big(1 - \Phi(t)\Big)^{N-i-1} \\ &+ \Big(i - N\Big(1 - \Phi(t)\Big)\Big)\Phi(t)^{N-i-1}\Big(1 - \Phi(t)\Big)^{i-1}\end{aligned} \tag{4.9}$$

This theorem demonstrates that the output central moment of the WM filter consists of two parts: the first corresponds to that of the standard median filter and the second quantifies the contribution of the weights. If all weights are equal, the second term vanishes. It is also obvious that

$$\sigma_{wm}^{\gamma} \geq \sigma_{med}^{\gamma} \tag{4.10}$$

since M_i and $L_i(N, \Phi, \gamma)$ are nonnegative. This fact states that for i.i.d. inputs, the standard median filter achieves the best noise attenuation. Recall that the running average filter has the best noise attenuation among linear FIR filters, another interesting analogy.

Theorem 4.9 is easily applied to evaluate the noise attenuation capability of WM filters [104]. Let the output variance be a measure of noise attenuation. We recast (4.7) when $\gamma = 2$,

$$\sigma_{wm}{}^2 = \sigma_{med}{}^2 + \sum_{i=1}^{K} L_i(N, \Phi, 2)M_i \tag{4.11}$$

where $\sigma_{med}{}^2$ denotes the variance of the standard median output. In a vector form, (4.11) becomes

$$\sigma_{wm}{}^2 = \sigma_{med}{}^2 + \underline{M}\ \underline{L}^T \tag{4.12}$$

where

$$\underline{L} = [L_1\ L_2\ \cdots\ L_K] \tag{4.13}$$

and

$$\underline{M} = [M_1\ M_2\ \cdots\ M_K] \tag{4.14}$$

denote the L-vector and M-vector, respectively. For input signals having the same noise distribution function Φ, one can calculate the M-vectors and L-vectors and compare the variances of different filters. Since the L-vectors are independent of the weights, they can be tabulated for different noise distributions and different filter window sizes. Tables IV–VI, called L-tables, show the values of the L-vectors for window

TABLE VI
THE L-TABLE FOR $N = 3$ TO $N = 25$ WITH BIEXPONENTIAL DISTRIBUTION

N	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}	L_{11}	L_{12}
3	4.746-E1											
5	4.407-E1	3.391-E2										
7	3.770-E1	2.982-E2	4.085-E3									
9	3.290-E1	2.226-E2	3.473-E3	6.119-E4								
11	2.925-E1	1.705-E2	2.352-E3	5.087-E4	1.032-E4							
13	2.639-E1	1.348-E2	1.626-E3	3.210-E4	8.447-E5	1.876-E5						
15	2.409-E1	1.094-E2	1.166-E3	2.048-E4	5.053-E5	1.517-E5	3.588-E6					
17	2.219-E1	9.080-E3	8.640-E4	1.356-E4	3.024-E5	8.708-E6	2.876-E6	7.124-E7				
19	2.059-E1	7.668-E3	6.590-E4	9.315-E5	1.875-E5	4.947-E6	1.597-E6	5.669-E6	1.455-E7			
21	1.922-E1	6.573-E3	5.147-E4	6.609-E5	1.208-E5	2.903-E6	8.690-E7	3.063-E7	1.151-E7	3.037-E8		
23	1.804-E1	5.704-E3	4.102-E4	4.822-E5	8.051-E6	1.769-E6	4.866-E7	1.608-E7	6.081-E8	2.392-E8	6.456-E9	
25	1.700-E1	5.003-E3	3.326-E4	3.604-E5	5.534-E6	1.117-E6	2.827-E7	8.649-E8	3.095-E8	1.240-E8	5.064-E9	1.393-E9

sizes $N = 3$ to $N = 25$ for three commonly used noise distributions, uniform, Gaussian, and Laplacian. With the aid of these tables, one can easily compare the noise attenuation capability of different WM filters.

Example 2.3: Compare the noise attenuation of the following two WM filters

$$\underline{W}_1 = [2, 2, 2, 7, 12, \mathbf{13}, 12, 7, 2, 2, 2]$$

$$\underline{W}_2 = [4, 4, 4, 9, 14, \mathbf{25}, 14, 9, 4, 4, 4]$$

for uniformly distributed noise.

It is easy to obtain the M-vectors of these two filters as follows:

$$\underline{M}(\underline{W}_1) = [0, 0, 5, 47, 136],$$

$$\underline{M}(\underline{W}_2) = [0, 0, 5, 34, 161].$$

From L-Table IV we find

$$\underline{L} = [10.4905, 1.6316, 0.3885, 0.1166, 0.02787] \times 10^{-3}.$$

Substituting $\underline{L}$, $\underline{M}(\underline{W}_1)$ and $\underline{M}(\underline{W}_2)$ into (4.12), it is easy to see that $\underline{W}_2$ has better noise attenuation than $\underline{W}_1$ for uniformly distributed noise. In fact, one can easily check that $\underline{W}_2$ has better noise attenuation performance than $\underline{W}_1$ for Gaussian and Laplacian distributed noise, as well. Note that if the two M-vectors were ordered, say $\underline{M}(\underline{W}_1) \leq \underline{M}(\underline{W}_2)$, then we can immediately conclude that $\underline{W}_1$ would have better noise attenuation capability than $\underline{W}_2$, without resorting to the L-vector.

V. OPTIMAL WEIGHTED MEDIAN FILTERING

In practice, the filter may be of little use if the weights of a WM filter cannot be determined in some "optimal" fashion. In order to find optimal WM filters under a certain error criterion, two different approaches have been developed recently, the estimation approach [16], [17], [61], [82], [89], [90], [111], [112], [114] ,[126] and the structural approach [17], [27], [50], [68], [107], [118], [124]. The two approaches are used to find a WM filter which minimizes noise and preserves the desired details of the anticipated signals.

In this section, we will review these two optimal WM filtering approaches.

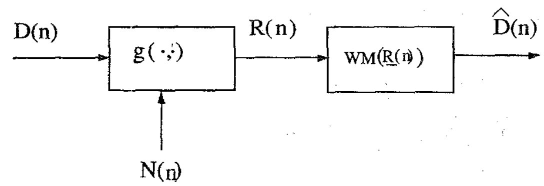


Fig. 4. The optimal weighted median filtering problem.

A. Problem Formulation

The optimal filtering problem over the class of WM filters can be stated as in Fig. 4. The process $R(n)$ at the input of a WM filter is assumed to be a corrupted version of some desired process $D(n)$. The corruption may be caused by a noise process $N(n)$ through a memoryless, possibly nonlinear, operation $g(.,.)$.

A window of width N slides across the input process $R(n)$. Let $\underline{R}(n)$ be the vector containing the N samples in the window at time n,

$$\underline{R}(n) = [R(n-K) \ldots, R(n), \ldots, R(n+K)], \qquad (5.1)$$

where $2K + 1 = N$. The WM filter output is an estimate, called $\hat{D}(n)$, of the desired process $D(n)$; thus

$$\hat{D}(n) = \text{MED}[W_1 \diamond R(n-K), W_2 \diamond R(n-K+1), \cdots, W_N \diamond R(n+K)]. \qquad (5.2)$$

For convenience, we first assume that $R(n)$ and $D(n)$ are M-valued signals. This assumption, in fact, is not necessary and is later relaxed. Under this assumption, the WM filter can also be implemented in the binary domain. That is, $\hat{D}(n)$ can be expressed as

$$\hat{D}(n) = \sum_{m=1}^{M-1} U(\underline{W}\underline{\xi}^{mT}(n)), \qquad (5.3)$$

where $U(\cdot)$ is the unit step function. $\underline{\xi}^m(n)$ is obtained from the binary input $\underline{r}^m(n) = T^m(\underline{R}(n))$, as follows:

$$\underline{\xi}^m(n) = [\xi^m(n-K), \ldots, \xi^m(n), \ldots, \xi^m(n+K)], \quad (5.4)$$

where

$$\xi^m(n) = 2r^m(n) - 1, \qquad (5.5)$$

for $m = 1, 2, \cdots, M-1$. The goal of optimal WM filtering is to find a WM filter from the class of window width N WM filters such that the MAE or MSE between the filter's output and some desired signal is minimized. Although both MAE and MSE criteria have been used in optimal WM filtering, this paper will concentrate on optimal WM filtering algorithms under the MAE criterion. It is straightforward to extend the MAE algorithms to the MSE criterion.

If $D(n)$ and $R(n)$ are jointly stationary, then the MAE to be minimized is

$$J(\underline{W}) = E[|D(n) - \hat{D}(n)|] = E\left[\left|\sum_{m=1}^{M-1}(d^m(n) - U(\underline{W}\underline{\xi}^{mT}(n))\right|\right] \quad (5.6)$$

where $d^m(n) = T^m(D(n))$.

It was shown that, due to the stacking property of WM filters, the optimization problem for finding optimal WM filters under the MAE criterion can be simplified as [16], [111]

$$\min_{\underline{W}\in\Im} J(\underline{W}) = \sum_{m=1}^{M-1} E[(\zeta^m(n) - \mathrm{sgn}(\underline{W}\underline{\xi}^{mT}(n)))^2] \quad (5.7)$$

where

$$\zeta^m(n) = 2d^m(n) - 1, \quad (5.8)$$

$$\mathrm{sgn}(X) = 2U(X) - 1, \quad (5.9)$$

and

$$\Im = \{\underline{W} : W_i \geq 0, \quad i = 1, 2, \cdots, N\}. \quad (5.10)$$

Obviously, traditional recursive algorithms cannot be used to solve this optimization problem since sgn(.) is not a differentiable function of the weights $\underline{W}$. One way to overcome this difficulty is to approximate the sign function with a differentiable function, [21]. Two different differentiable functions, a linear function and a sigmoidal function, have been used to approximate the sign function [111], [112], [114]. Usually, algorithms based on the sigmoidal approximation give better performance than those based on the linear approximation, [89], [90], since the former is a closer approximation of the sign function. However, with the linear approximation, nonadaptive and adaptive algorithms can be derived.

B. MAE Optimal WM Filtering

1) Nonadaptive LMA WM Filtering: If the sign function is replaced by a linear function, we have

$$\min_{\underline{W}\in\Im} \{\underline{W}\mathbf{R}\underline{W}^T - 2\underline{W}\underline{R}^{sT} + \sum_{m=1}^{M-1} E[(\zeta^m(n))^2]\}, \quad (5.11)$$

where

$$\mathbf{R} = \sum_{m=1}^{M-1} E[\underline{\xi}^m(n)^T \underline{\xi}^m(n)] \quad (5.12)$$

$$\underline{R}^s = \sum_{m=1}^{M-1} E[\zeta^m(n)\underline{\xi}^m(n))]. \quad (5.13)$$

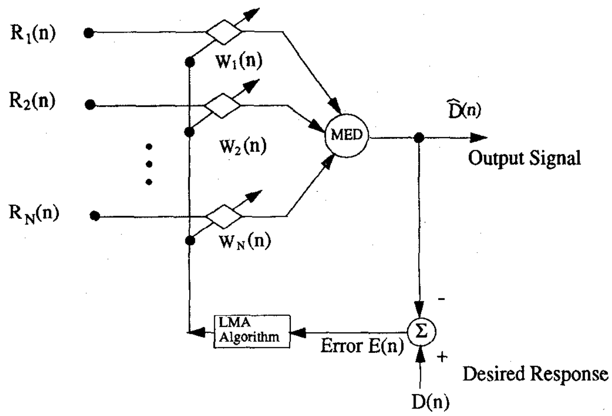


Fig. 5. Adaptive weighted median filter.

One may notice that after replacing the sign function with a linear function, WM filters become in fact linear FIR filters. That is, the optimization problem in (5.7)–(5.10) reduces to finding an optimal linear FIR filter with nonnegative weights. In contrast to traditional LMS linear filtering [102], this optimal FIR filter minimizes the sum of squared errors incurred at the levels of the threshold decomposition architecture.

If no constraints are imposed on $\underline{W}$, the solution is obtained by solving the linear system of equations

$$\underline{W} = \mathbf{R}^{-1}\underline{R}^{sT}. \quad (5.14)$$

With linear inequality constraints, however, this optimization problem can be solved by the gradient projection method as was done in [113]. The method consists of three steps:

- Step 1) Initialization: set all weights in $\underline{W}(0)$ to arbitrary positive values and give a small positive value to α_t. Here α_t is an iteration tolerance.
- Step 2) Iterations:

$$\underline{W}^T(k+1) = P[(I - \mu\mathbf{R})\underline{W}^T(k) + \mu\underline{R}^{sT}]$$

$$0 < \mu < 2/(\lambda_1 + \lambda_N) \quad (5.15)$$

  where $P(.)$ is a projection operation which does not affect the positive entries of the weight vector and maps its negative entries to zeros. λ_1 and λ_N are the smallest and the largest eigenvalues of $\mathbf{R}$, respectively, and μ is a fixed stepsize.
- Step 3) Convergence: if $\sum_{i=1}^{N}(W_i(k+1) - W_i(k))^2 \leq \alpha_t$, $\underline{W}(k+1)$ is accepted as the solution. Otherwise go to Step 2.

It has been proved that if the autocorrelation matrix $\mathbf{R}$ is nonsingular, $\underline{W}(k+1)$ will converge to the optimal solution $\underline{W}^*$ as $k \to \infty$.

In practice, $\mathbf{R}$ and $\underline{R}^s$ are estimated as follows, given the observation of $\underline{R}(j)$ and $D(j), j = 0, 1, \cdots, L-1$ [113]:

$$\hat{R}_{i,j} = \frac{1}{L}\sum_{n=0}^{L-1}(R_{max}(n) - R_{min}(n) - 2|R(n-K+i) - R(n-K+j)|)$$
$$i, j = 1, 2, \ldots N \quad (5.16)$$

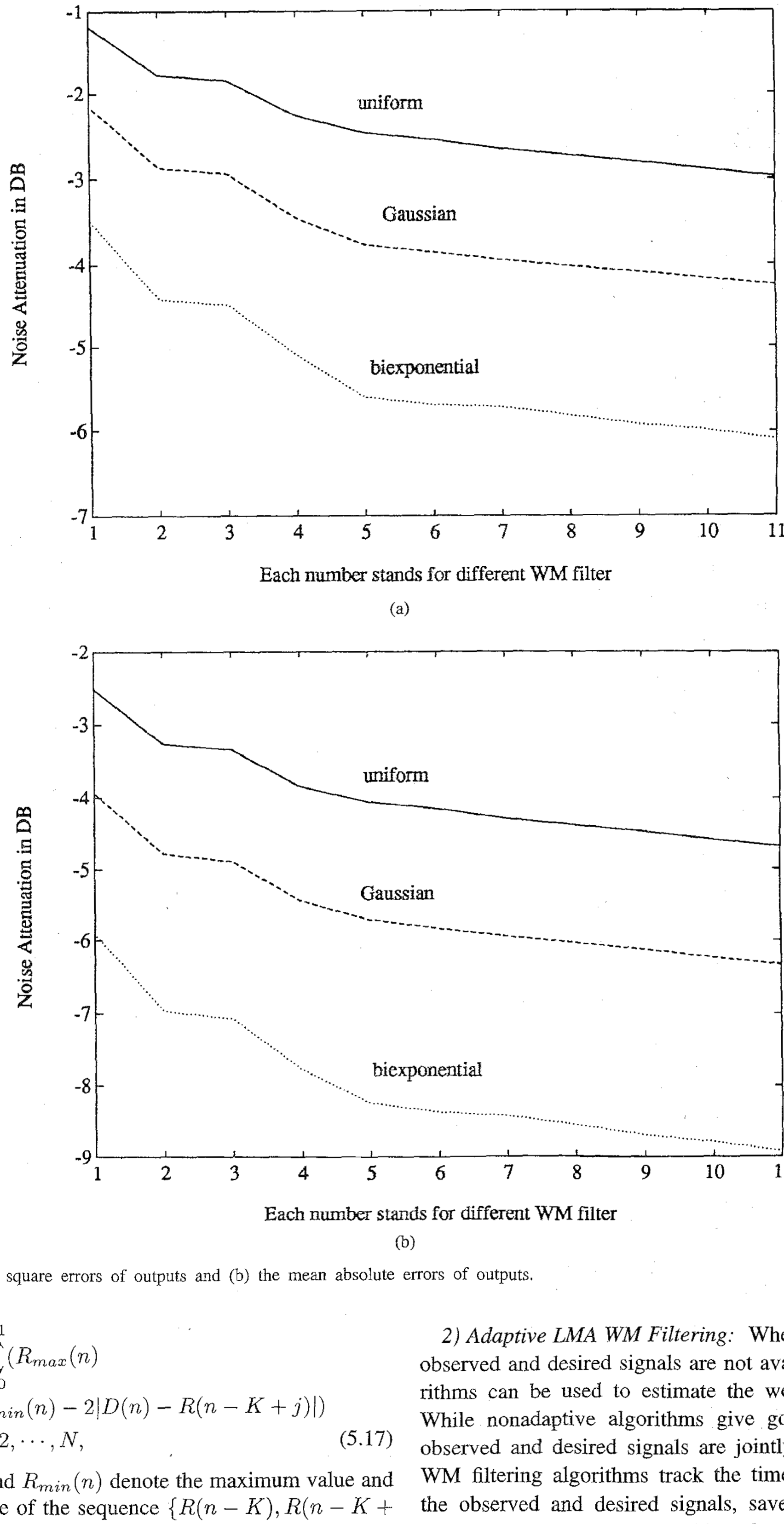


Fig. 6. (a) The mean square errors of outputs and (b) the mean absolute errors of outputs.

$$\begin{aligned}\widehat{R}_j^s = &\frac{1}{L}\sum_{n=0}^{L-1}(R_{max}(n) \\ &- R_{min}(n) - 2|D(n) - R(n-K+j)|) \\ j = &1, 2, \cdots, N, \end{aligned} \tag{5.17}$$

where $R_{max}(n)$ and $R_{min}(n)$ denote the maximum value and the minimum value of the sequence $\{R(n-K), R(n-K+1), \cdots, R(n+K)\}$, respectively.

2) Adaptive LMA WM Filtering: When the *statistics* of the observed and desired signals are not available, adaptive algorithms can be used to estimate the weights of WM filters. While nonadaptive algorithms give good results when the observed and desired signals are jointly stationary, adaptive WM filtering algorithms track the time-varying statistics of the observed and desired signals, save memory space, and are easy to implement. In this subsection we review two types of adaptive LMA WM filtering algorithms, based on approximating the sign function by a linear function and a sigmoidal function [111], [114], [115].

a) Linear approximation: Using a linear approximation of the sign function, an adaptive LMA WM filtering algorithm has been derived [115].

$$\underline{W}(n+1) = P[\underline{W}(n) + 2\mu(n) \sum_{m=R_{min}(n)}^{R_{max}(n)} (\zeta^m(n) - \underline{W}(n)\underline{\xi}^{mT}(n))\underline{\xi}^m(n)], \tag{5.18}$$

where $\mu(n)$ is an adaptation stepsize which can be a positive constant or time-varying sequence. Its bounds were derived in [113].

This algorithm can be implemented in the real domain [115]

$$W_i(n+1) = W_i(n) + 2\mu(n)[(R_{max}(n) - R_{min}(n) - 2|D(n) - R(n-K+i)| - \sum_{j=1}^{N} W_j(n)(R_{max}(n) - R_{min}(n) - 2|R(n-K+i) - R(n-K+j)|) \tag{5.19}$$

$$W_i(n+1) = 0 \quad \text{if} \quad W_i(n) < 0. \tag{5.20}$$

The convergence of this algorithm is guaranteed under certain conditions [113].

b) Sigmoidal approximation: The sigmoidal function is another choice to approximate the sign function. This function is widely used in neural networks because of its good approximation ability to the sign function. When the sigmoidal function is used, the following adaptive algorithm is obtained.

$$\underline{W}(n+1) = P[\underline{W}(n) + 2\mu(n) \sum_{m=1}^{M-1} (\zeta^m(n) - \hat{\zeta}^m(n))\delta^m(n)\underline{\xi}^m(n)] \tag{5.21}$$

where

$$\delta^m(n) = (1 + \hat{\zeta}^m(n))(1 - \hat{\zeta}^m(n)) \tag{5.22}$$

$$\hat{\zeta}^m(n) = \operatorname{sgn}_s(\underline{W}(n)\underline{\xi}^{mT}(n)) \tag{5.23}$$

and $\operatorname{sgn}_s(.)$ is a sigmoidal function defined as

$$\operatorname{sgn}_s(X) = \frac{2}{1+e^{-X}} - 1. \tag{5.24}$$

This adaptive algorithm yields better performance than the one based on linear approximation. However, implementing this algorithm in the binary domain leads to very large computational complexity. Fortunately, it can be simplified by exploiting the properties of $\delta^m(n)$ and the fact that there are at most $N+1$ different binary vectors $\underline{\xi}^m(n)$ for each N -length observation vector $\underline{R}(n)$, leading to a very simple adaptive WM filtering algorithm [117]

$$W_i(n+1) = P[W_i(n) + 2\mu(D(n) - \hat{D}(n))\operatorname{sgn}(R_i(n) - \hat{D}(n))], \quad i = 1, \cdots, N. \tag{5.25}$$

There is an intuitive explanation of how this algorithm operates. When the actual output of the WM filter is smaller than the desired output, the weights corresponding to samples which are larger than the actual output are incremented. Fig. 5 shows the adaptive WM filtering structure.

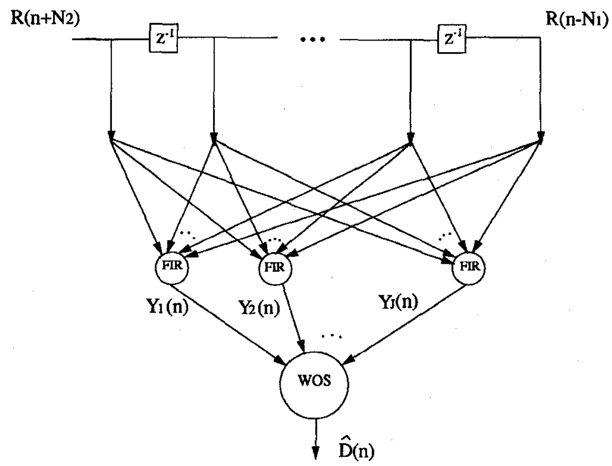


Fig. 7. The general structure of the FWH filter.

One may notice that it is very similar to the adaptive linear filtering structure. In fact, updating the weights of adaptive WM filters is easier than those of adaptive LMS filters. Naturally, the actual filtering operations are quite different. Other adaptive WOS filtering algorithms for signal and image processing have been developed in [89], [90].

Besides its simplicity, this algorithm does not have any specific requirements on input signals and desired signals. However, these adaptive WM filters cannot be guaranteed to converge to the globally optimal solution because the optimization problem based on the sigmoidal approximation does not have unique minima.

C. Optimal WM Filtering Under Structural Constraints

It has often been observed that noise attenuation alone does not necessarily yield acceptable signal quality in images. This happens specifically when the training signals are unavailable or incomplete. It would be appropriate and very useful if this estimation approach would be somehow combined with the requirement to preserve certain image details which are known by the designers to be part of the desired signal. That is, the optimal filter sought would attenuate noise optimally, and in the meantime, it would preserve certain signal features. This approach was first developed by Gabbouj *et al.* [27] for stack filters and called optimal filtering under structural constraints. This approach was later tuned and applied to WM filters in [68]. In case of WM filters, the signal structures can easily be expressed in terms of linear inequalities. Based on the optimal WM filtering algorithms discussed above, a structural approach has been developed to design a WM filter which minimizes the MAE subject to a predetermined set of structural constraints on the filter's behavior [68]. In the special case of white noise corrupting a constant signal, optimal WM filters which preserve certain length signal details have been derived [107]. This approach is extended to 2-D filtering, where the image details to be preserved were lines in certain directions, areas and combinations, see [105], [106].

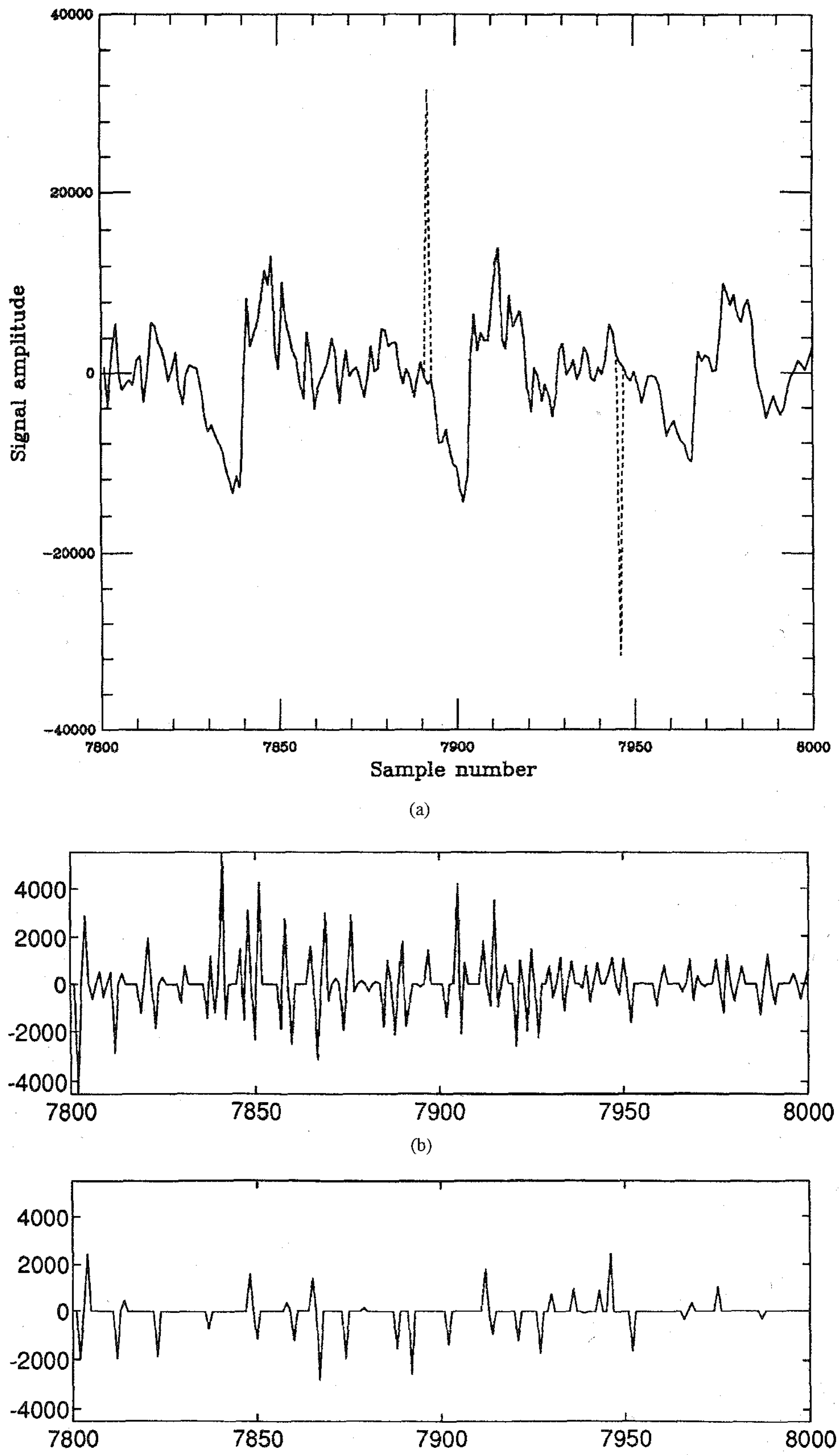
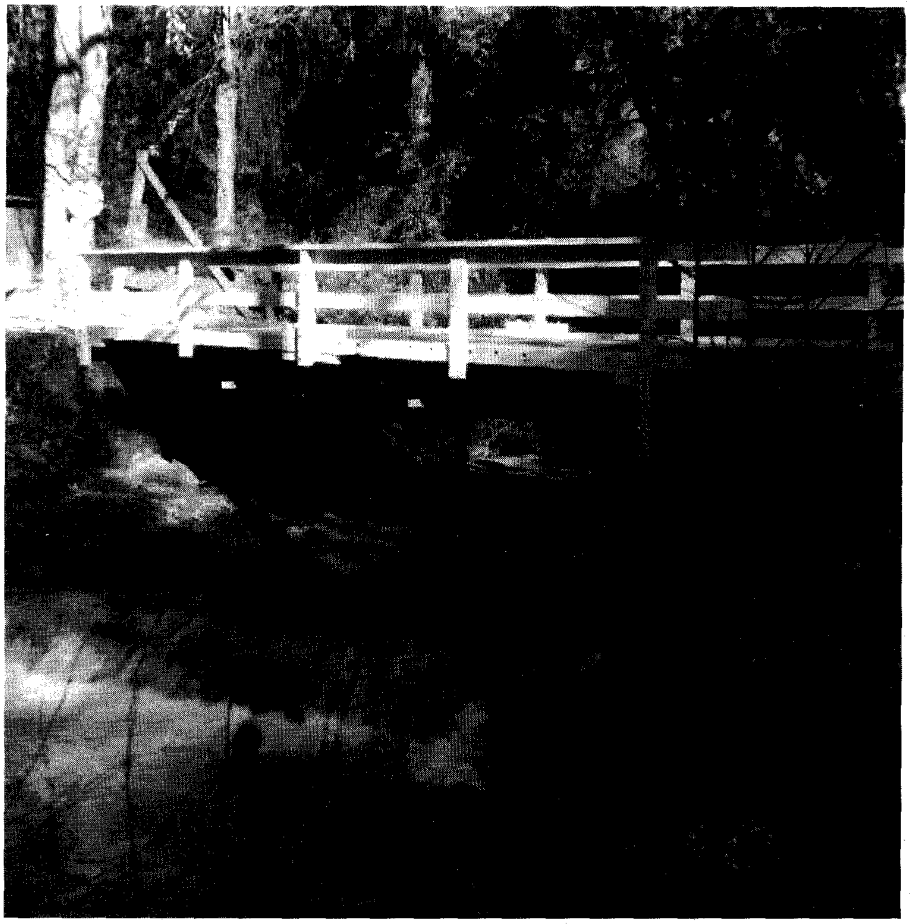


Fig. 8. (a) 200 samples of a high quality speech sample. The speech sample (continuous line) is PCM coded and has a sampling frequency of 8 kHz. The length of the speech sample is 80 000 samples, i.e., 10 s. Each sample is represented by 16 b. Random bit errors were introduced to the speech sample with the probability 0.05% resulting in the signal shown with the dotted line. (b) The difference signal between the original and the filtered signal by a 3-point median filter. The smearing of the speech signal that is clearly visible and the speech quality has been severely decreased. (c) The difference signal between the original and the filtered signal by an idempotent filter given by (7.1) The smearing of the speech signal is minimal and the speech quality is very close to the original sample.



(a)

Fig. 9. The restoration of image bridge, (a) origianl image, (b) noisy image, (c) the Wiener filter, (d) the Stack filter, (e) the nonadaptive WM filter, (f) the structural WM filter, (g) the Median filter, and (h) the FWH1 filter.

Using mathematical notation, one can express the problem of finding optimal WM filters under structural constraints as follows

$$J_\gamma(\underline{W}) = E[|D(n) - \hat{D}(n)|^\gamma], \quad \gamma > 0 \tag{5.26}$$

subject to

$$W_i \geq 0, \quad \text{for} \quad i = 1, 2, \cdots, N \tag{5.27}$$

$$\mathbf{A}\underline{W}^T \geq \underline{B}^T, \tag{5.28}$$

where γ is a positive constant, $\gamma = 1, 2$ corresponding to the MAE and the MSE criteria, respectively. $\mathbf{A}$ is an $I \times N$ matrix, and $\underline{B}$ is an I dimensional vector of constants. Equation (5.28) denotes I inequalities which represent the structural constraints.

For example, given a WM filter with $\underline{W} = [W_1, W_2, \mathbf{W_3}, W_4, W_5]$. Suppose that constant pulses of length two or more shall be preserved, then the constraints on the weights should be

$$W_2 + W_3 \geq T_h,$$

$$W_3 + W_4 \geq T_h.$$

Replacing T_h by $1/2\sum_{i=1}^{5} W_i$, we can rewrite the above inequalities as follows:

$$W_2 + W_3 \geq W_1 + W_4 + W_5,$$

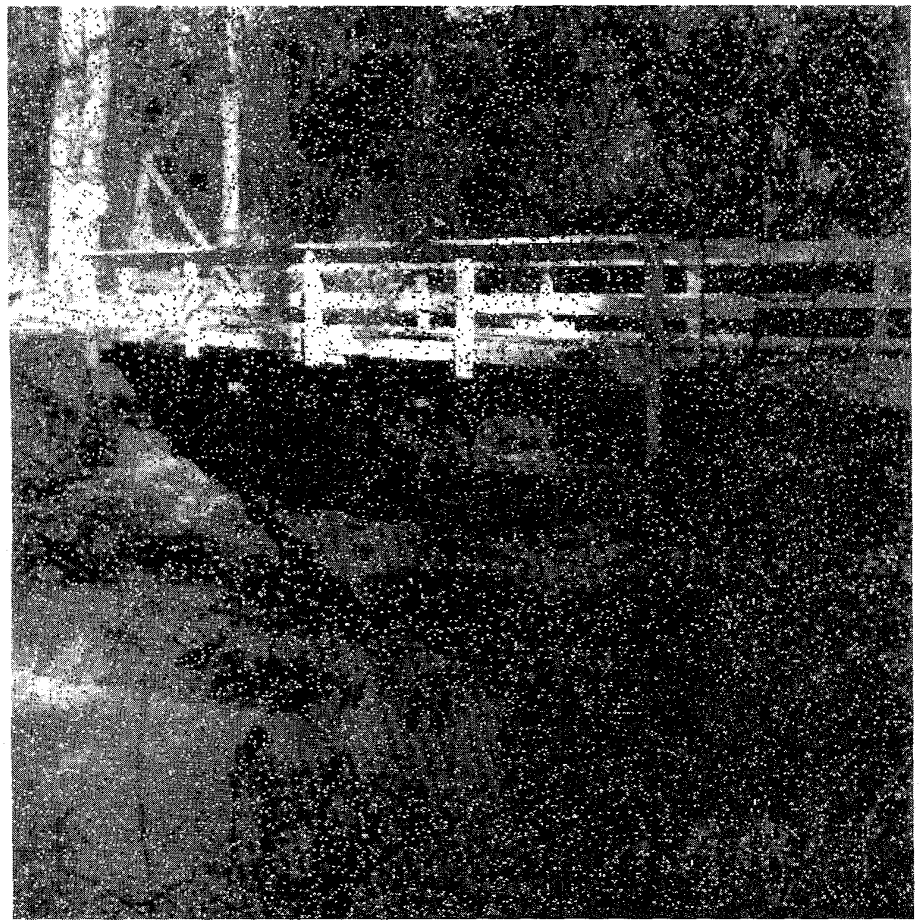
$$W_3 + W_4 \geq W_1 + W_2 + W_5.$$

In general, it is difficult to solve the optimization problem in (5.26)–(5.28). However, in certain cases optimal WM filters with structural constraints can be obtained even in closed form [105], [107].

When the desired signal is a constant signal, the MSE and the MAE of the WM filter are the variance and the absolute deviation of the output signal. Assume the constant signal is contaminated by i.i.d. noise with common distribution function $\Phi(t)$. Then the cost function of the optimization problem becomes

$$J_\gamma(\underline{W}) = \sigma_{med}^\gamma + \sum_{i=1}^{K} M_i L_i(N, \Phi, \gamma), \tag{5.29}$$

where $N = 2K + 1$, M_i and $L_i(N, \Phi, \gamma)$ are defined in (4.6) and (4.8).



(b)

Fig. 9. (*Continued.*)

It is straightforward to see that if there are no structural constraints on the weights, the optimal WM filter is the standard median filter.

Several algorithms are provided in [107], [108] for solving the optimization problem in (5.29). Using these algorithms, the optimal symmetric WM filters which preserve length 2 details and length 3 details have been found.

Given a WM filter with window size $N = 2K + 1$,

$$\underline{W} = [W_{-K}, \cdots, \mathbf{W}_0, \cdots, W_K]$$

the optimal WM filters preserving length 2 details are expressed by the following explicit form

$$\underline{W} = [1, 1, \cdots, 1, K, \mathbf{2K-1}, K, 1, \cdots, 1, 1]. \tag{5.30}$$

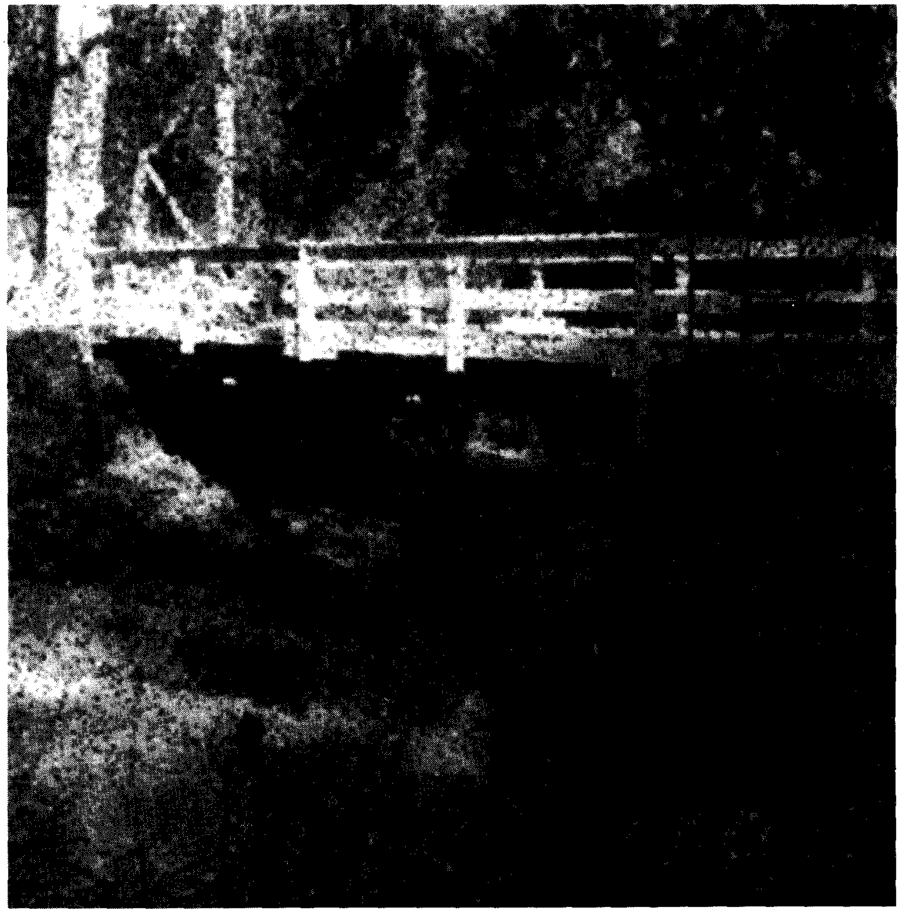
More importantly, this class of optimal filters is proven to have the following properties:

- These filters are optimal for different noise distributions.
- These filters are optimal for any positive γ. This means that they are optimal both in the sense of MSE criterion and MAE criterion.

TABLE VII
ALL WM FILTERS PRESERVING PULSES OF LENGTH 2 FOR $N = 7$

No.	M_1 M_2 M_3	Filters
1	0, 6, 15	[1, 1, 1, **5**, 1, 1, 1]
2	0, 4, 15	[1, 3, 3, **11**, 3, 3, 1]
3	0, 4, 14	[1, 3, 3, **9**, 3, 3, 1]
4	0, 3, 13	[1, 1, 5, **5**, 5, 1, 1]
5	0, 2, 15	[1, 1, 2, **5**, 2, 1, 1]
6	0, 2, 14	[1, 3, 4, **9**, 4, 3, 1]
7	0, 2, 13	[1, 1, 4, **5**, 4, 1, 1]
8	0, 2, 12	[1, 3, 6, **9**, 6, 3, 1]
9	0, 2, 11	[1, 3, 7, **9**, 7, 3, 1]
10	0, 2, 10	[1, 3, 5, **9**, 5, 3, 1]
11	0, 2, 9	[1, 1, 3, **5**, 3, 1, 1]

To compare the performance of optimal WM filters with other WM filters, Fig. 6 shows the MSE's and the MAE's of all possible symmetric WM filters with window size 7 which preserve length 2 details. The input is a constant plus white noise. Three noise distributions are used, uniform, Gaussian, and Laplacian, with zero mean and unit variance. Table VII lists all these WM filters used in Fig. 6. One can observe that the optimal WM filter (no. 11 in Table VII) achieves the best noise attenuation for the different noise distributions both in the MSE sense and MAE sense.



(c)

Fig. 9. (*Continued.*)

Although optimal symmetric WM filters which preserve length 2 details have a simple expression, there seems not to be one for optimal WM filters preserving length 3 details. We have to find the filter for each window size. Table VIII lists some optimal WM filters preserving length 3 details.

TABLE VIII
OPTIMAL WEIGHTED MEDIAN FILTERS PRESERVING PULSES OF LENGTH 3

N=7	$M_1 = 0$, $M_2 = 0$, $M_3 = 5$	[3, 4, 5, **7**, 5, 4, 3]
N=9	$M_1 = 0$, $M_2 = 0$, $M_3 = 5$ $M_4 = 29$	[2, 2, 5, 6, **9**, 6, 5, 2, 2]
N=11	$M_1 = 0$, $M_2 = 0$, $M_3 = 5$, $M_4 = 34$, $M_5 = 161$	[3, 3, 3, 7, 11, **19**, 11, 7, 3, 3, 3]
N=13	$M_1 = 0$, $M_2 = 0$, $M_3 = 5$, $M_4 = 53$, $M_5 = 209$, $M_6 = 484$	[2, 2, 2, 2, 9, 10, **17**, 10, 9, 2, 2, 2, 2]
N=15	$M_1 = 0$, $M_2 = 0$, $M_3 = 5$, $M_4 = 65$, $M_5 = 321$, $M_6 = 955$, $M_7 = 1905$	[2, 2, 2, 2, 2, 11, 12, **21**, 12, 11, 2, 2, 2, 2, 2]
N=17	$M_1 = 0$, $M_2 = 0$, $M_3 = 5$, $M_4 = 77$, $M_5 = 457$, $M_6 = 1662$, $M_7 = 4136$, $M_8 = 7447$	[2, 2, 2, 2, 2, 2, 13, 14, **25**, 14, 13, 2, 2, 2, 2, 2, 2]

VI. AN IMPORTANT EXTENSION OF WM FILTERS: FIR-WOS HYBRID FILTERS

From the discussion above, we have seen that WM filters can be easily designed to suppress non-Gaussian noise and preserve signal structures such as edges and lines. However, they cannot be designed in general to retain or restore some desired signal frequencies and reject others. On the other hand, linear filters do not consider rank ordering of the observations in determining the output, leading to poorer performance at signal edges and in the presence of non-Gaussian noise. To exploit the desirable properties of both filter classes, following the success of the initial work on FIR-Median hybrid (FMH) filters [66], FIR-WOS hybrid (FWH) filters were introduced in [117]. Other applications of FIR-WOS filters can be found in [45].

The general structure of the FWH filter is shown in Fig. 7. In the first stage of FWH filters, the input signal $R(n)$ is filtered with J linear FIR filters. Let N be the length of the FIR filters. The outputs of the FIR filters are

$$Y_i(n) = \underline{H}_i \underline{R}^T(n), \quad i = 1, 2, \cdots, J \tag{6.1}$$

where ^{T} denotes the transpose, $\underline{H}_i$ is the coefficient vector of the i'th FIR filter

$$\underline{H}_i = [H_{i,1}, H_{i,2}, \cdots, H_{i,N}], \tag{6.2}$$

(d)

Fig. 9. (*Continued.*)

$\underline{R}(n)$ is the input vector of the FWH filter at time n,

$$\underline{R}(n) = [R(n - N_1), R(n - N_1 + 1), \cdots, R(n + N_2)] \quad (6.3)$$

N_1, N_2 are nonnegative integers, and $N_1 + N_2 + 1 = N$. In the second stage, the outputs of the FIR filters are filtered by a WOS filter with the nonnegative weight vector $\underline{W} = [W_1, W_2, \cdots, W_J]$ and the threshold T_h. If the weights of the WOS filter are integers, the output of the FWH filter is given by

$$F(\underline{R}(n)) = T_h : th \text{ largest value of the set } [W_1 \diamond Y_1(n), \cdots, W_J \diamond Y_J(n)]. \quad (6.4)$$

It is obvious that the FWH filters include FIR filters, WOS filters, and FMH filters as special cases. They can be easily extended to multistage FIR-WOS hybrid filters which include Ll filters and hybrid order statistic filters [73], [74]. In fact, FWH filters rather than WM filters are widely used in practice, as shown in the next section.

The significance of FWH filters is that an adaptive algorithm is available for finding optimal FWH filters under the MAE criterion [117]. The algorithm is similar to the backpropagation algorithm which is widely used in neural networks

$$W_0(n + 1) = P\left[W_0(n) - 2\mu_2(D(n) - \hat{D}(n))\right] \quad (6.5a)$$

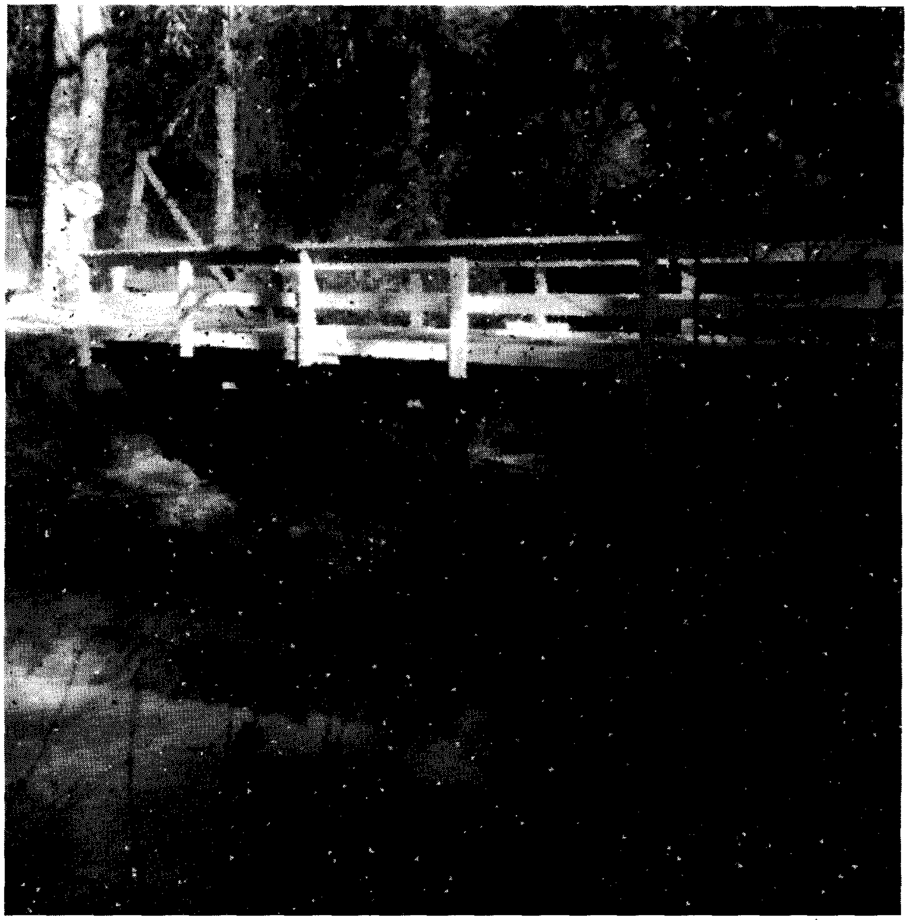
$$W_i(n+1) = P\left[W_i(n) + 2\mu_2(D(n) - \hat{D}(n))\operatorname{sgn}(Y_i(n) - \hat{D}(n))\right]$$

$$i = 1, 2, \cdots, J \quad (6.5b)$$

$$\underline{H}_i(n + 1) = \underline{H}_i(n) + 2\mu_1(D(n) - \hat{D}(n))W_i(n)(1 + \operatorname{sgn}_s(Y_i(n) - \hat{D}(n)) (1 - \operatorname{sgn}_s(Y_i(n) - \hat{D}(n)))\underline{H}(n), \quad (6.6)$$

where μ_1 and μ_2 are adaptation stepsizes.

As with most backpropagation based algorithms, this algorithm cannot be guaranteed to find global optimal FWH filters. However, experimental results have demonstrated that the adaptive FWH filters converge to some nontrivial solutions [117].



(e)

Fig. 9. (*Continued.*)

VII. Applications of Weighted Median Filters

With the development of WM filtering theory, the applications of WM filters have been explored in 1-, 2-, and 3-D signal processing. WM filters are basically smoothers with unit gain, and perform similar to FIR filters with positive coefficients summing up to unity. This is why they are mainly used as robust noise smoothers and predictors.

A. Speech Signal Processing

An early application of median filtering was introduced in [48], where Jayant studied the use of a short median filter for improving DPCM coded speech quality in the presence of transmission errors. It was found that if the probability of a bit error is of the order of 1% or more, the result of smoothing with a 3-point (or a 5-point) median filter turns out to be perceptually desirable in spite of the speech smearing caused by the filter.

With small bit error rates, the smearing of the speech waveform caused by the 3-point median filter is unacceptable if compared to the advantage gained. However, even a bit error rate as low as 0.05% is very disturbing if the speech quality is otherwise good. Idempotent WM filters [31] have the property that apart from removing single impulses they have only a minimal effect on the signal. Therefore, they are good candidates for the current problem.

Haavisto *et al.* [37] applied 1-D idempotent WM filters to PCM coded speech samples contaminated by random bit errors and compared their performance to that of the median and average filters. The 1-D idempotent WM filter they used can be represented in the following form:

$$\underline{W} = [\overbrace{1, 1, \cdots, 1}^{K\text{times}}, 2K - 1, \overbrace{1, \cdots, 1, 1}^{K\text{times}}] \tag{7.1}$$

where $2K + 1$ is the filter length.

It was found that with $K = 5$, it was very difficult to distinguish between the original signal and the filtered signal. Therefore, only a minor additional advantage could be achieved by making the filter longer. With a bit error rate of 0.05%, this filter performed very well. The quality of the filtered speech signal was very close to the original; whereas, the smearing caused by the 3-point median filter was quite severe. This is illustrated in Fig. 8. When the bit error rate is

(f)

Fig. 9. (*Continued.*)

increased to 0.5%, the idempotent filter significantly improves the speech quality and is much better than the 3-point median filter. When the error rate is further increased to 2.5% the median filter performs better and a shorter idempotent filter should be used. However, in this example the 2.5% bit error rate means that there is on average at least one erroneous bit in every third data sample, and acceptable speech quality can no longer be achieved.

B. Adaptive Filtering of Still Images

Adaptive WM and FWH filters have been used to restore noisy images by Yin *et al* [111]–[115]. It was found that with much less computations and memory, adaptive WM filters work as well as adaptive stack filters and adaptive FWH filters perform better than both, with additional complexity.

Fig. 9(a) and (b) shows the original image "bridge" and the image corrupted with impulsive noise, respectively. The probability of impulses is 0.125, and if it occurs it can be positive or negative with equal probability. The impulses were set to a height of ± 200. In the experiment, the upper left quarter of the original and noisy images, Fig. 9(a) and (b), were used to train the LMS Wiener filter, Lin's LMA adaptive stack filter [59], WM filters, and FWH filters. Two adaptive FWH filters are used: FWH1 and FWH2. FWH1 is shown in Fig. 10. FWH2 has the same structure as FWH1 but its FIR filter is the Wiener filter under the MSE criterion. Thus, in FWH2, only the WOS section is updated during the training process.

When training signals are not available, the approach presented in Section V-C can be used to design optimal WM filters under structural constraints. In this experiment, we also include one WM filter which achieves maximum noise attenuation and preserves vertical, horizontal, and diagonal lines [104].

The filters resulting from the training algorithm and the structural approach were used to filter the noisy image. The MAE values and the MSE values of the restored images are listed in Table IX. Results produced by the adaptive WM filters with the linear approximation are not given here because they produced almost the same result as the nonadaptive WM filters.

Fig. 9 shows the images obtained by (c) the Wiener filter, (d) the stack filter, (e) the nonadaptive WM filter, (f) the structural WM filter, (g) the median filter, and (h) the FWH1 filter. As can be seen by comparing the errors introduced by these filters, all rank order based filters give a much better

(g)

Fig. 9. (Continued.)

TABLE IX
IMPULSIVE NOISE REDUCTION FOR IMAGE

Filter	Mean Square Error	Mean Absolute Error
Wiener	473.3	16.45
Adaptive Stack	194.6	4.09
Nonadaptive WM	224.4	4.20
Adaptive WM (S)	179.0	4.28
Structural WM	230.0	4.10
Median	165.3	7.45
Adaptive FWH1	132.1	3.32
Adaptive FWH2	169.6	4.32

visual performance than the Wiener filter. Furthermore, the WM filter works as well as the adaptive stack filter; while, FWH1 gives the best performance. One may also notice that the performance of the structural WM filter is comparable to that of the adaptive WM filters although it is derived on the basis of constant signals corrupted by additive noise. In order to compare these rank order based filters, we display their parameters in Table X. Note that the first parameter in the FWH filter is the weight corresponding to the linear FIR filter substructure and the last is the threshold of the WOS filter.

It is interesting to notice that the FIR filter in FWH1 is a high-pass filter while the Wiener filter in FWH2 is a lowpass one. In fact, the Wiener filter is useless in FWH2 as its corresponding weight is zero. However, this does not happen in FWH1. The FIR filter plays a very interesting role in FWH1. As one can anticipate, in conventional adaptive WOS filters, the weight at the center of the window tends to be large in order to preserve image details. However, when the center sample is corrupted by impulsive noise, the adaptive WOS filter sometimes cannot remove this impulse, e.g., due to the presence of other impulses in the same window. In FWH1, this problem is alleviated by the FIR filter. As a high-pass filter, the FIR filter in FWH1 changes the polarity of the impulse corrupting the center sample in the window. The two impulses of opposite signs tend to cancel each other's effect at the output of the WOS filter. Due to this, FWH1 gives better performance than FWH2 and WOS filters.

Other experiments have been performed to test optimal WM filters for different noisy images using different window sizes [111]–[115].

(h)

Fig. 9. (*Continued.*)

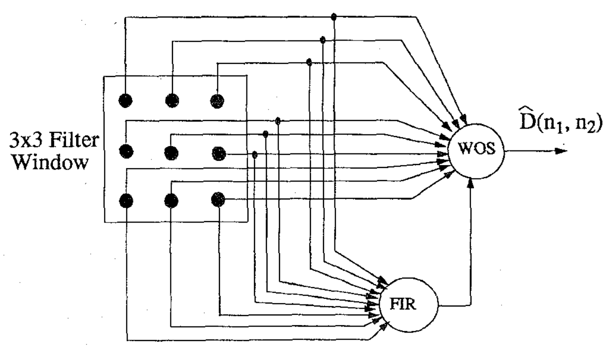


Fig. 10. The structure of the adaptive FWH1.

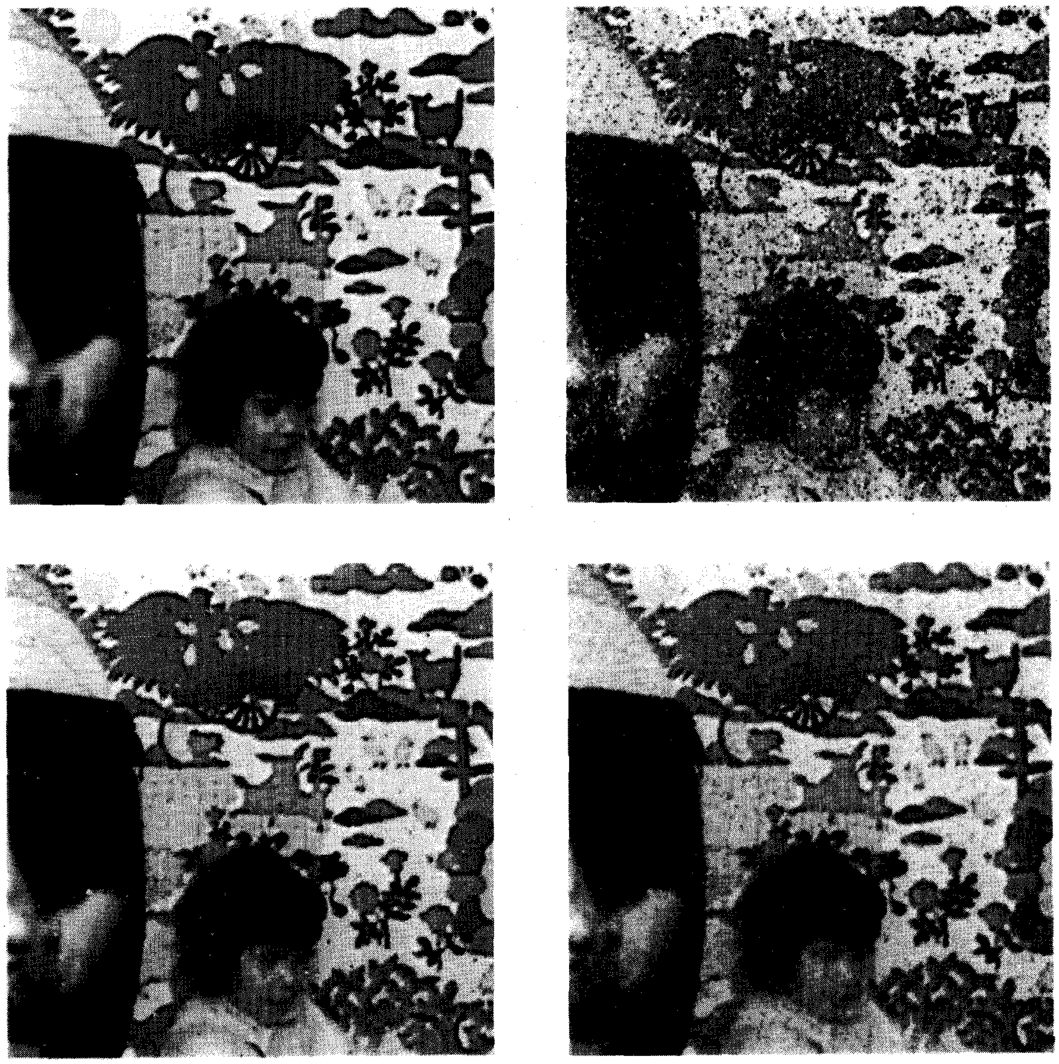
C. Image Sequence Filtering

In image sequences, there are nonmoving and moving image regions. In nonmoving image regions, noise can be reduced without impairing the spatial resolution by performing filtering along the time axis, i.e., using strictly temporal filters. However, in moving image regions, temporal filters introduce severe blurring. Small moving objects may totally disappear from a filtered sequence. Motion estimation and compensation methods have been used to reduce the degradation of moving objects [44]. However, accurate motion detection in a noisy image sequence is difficult. In addition, motion detection and specially motion estimation are time consuming. Viero *et al* [94], [95] found that in nonmoving regions, the following WM filter

$$\begin{aligned} Y(n_1, n_2, n_3) = &\mathrm{MED}[X(n_1, n_2, n_3 - 1), \\ &X(n_1, n_2 - 1, n_3), X(n_1 - 1, n_2, n_3) \\ &3 \diamond X(n_1, n_2, n_3), X(n_1 + 1, n_2, n_3), \\ &X(n_1, n_2 + 1, n_3), X(n_1, n_2, n_3 + 1)] \end{aligned} \tag{7.2}$$

retains all spatial details independent of their size and orientation (if the sequence is assumed to contain no noise). On the other hand, in moving regions the WM filter reduces to a spatial filter under certain assumptions [95],

$$\begin{aligned} &Y(n_1, n_2, n_3) = \mathrm{MED}[X(n_1, n_2 + 1, n_3), \\ &X(n_1 + 1, n_2, n_3) 3 \diamond X(n_1, n_2, n_3), \\ &X(n_1 - 1, n_2, n_3), X(n_1, n_2 - 1, n_3)]. \end{aligned} \tag{7.3}$$



(c) (d)

Fig. 11. Image sequence restoration: (a) The seventh frame of the original sequence; (b) The image degraded by impulsive noise (the probability of impulses is 0.1 and the height is 200); (c) The image restored by the WM filter; (d) The image restored by the linear filter.

To assess the performance of the above filter, it has been applied to the eight frames long "Costgirls" sequence. Fig. 11(a) and (b) shows the seventh frame of the original and noisy sequence, respectively. The probability of impulses is 0.1 and the height is ± 200. The noisy image sequence was filtered by the WM filter in (7.2) and a linear average filter which uses motion detection. In nonmoving image regions, it performs a 3-point temporal average and in moving regions, a spatial average inside a plus-shaped window of size five. The MAE and MSE values of the restored image sequences are listed in Table XI.

Fig. 11(c) and (d) shows the images restored by the WM filter and the linear filter. The WM filter retains stationary and moving details, but it did not remove all the impulses present in the noisy image sequence. The linear filter retained large structures of the original image sequence quite well, but the average smeared impulses and the resulting image sequence were of unacceptable quality, as can be seen from Fig. 11(d).

D. Prediction of a Square Waveform

Like linear FIR filters, WM filters can also be used as predictors and interpolators. Here we apply adaptive FWH filters to predict a square waveform both in Gaussian and non-Gaussian noise [117]. The waveform to be restored is a symmetric square wave with a period of 16 samples and magnitude of ± 50. The corrupting noise is a contaminated Gaussian noise whose distribution is $\Phi(\epsilon, \sigma_1, \sigma_2)$, that is, with probability $1-\epsilon$ the noise is normally distributed with standard deviation σ_1, and with probability ϵ it is normally distributed with standard deviation σ_2 ($\sigma_2 > \sigma_1$).

The order of the linear filter and WOS filter is 33, implying that samples $X(n-33), \cdots, X(n-1)$ are used to estimate

TABLE X
OPTIMAL FILTERS FOR REMOVING IMPULSIVE NOISE

Filter	FIR Part	WOS Part
Wiener	$\begin{pmatrix} 0.10 & 0.12 & 0.09 \\ 0.12 & 0.19 & 0.12 \\ 0.09 & 0.12 & 0.10 \end{pmatrix}$	
Non-adaptive WM		$\begin{pmatrix} 0.06 & 0.10 & 0.05 \\ 0.11 & 0.41 & 0.11 \\ 0.08 & 0.12 & 0.08 \end{pmatrix}$
Adaptive WM (S)		$\begin{pmatrix} 0.45 & 0.71 & 0.44 \\ 0.84 & 2.33 & 0.85 \\ 0.42 & 0.68 & 0.45 \end{pmatrix}$
Structural WM		$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 1 \end{pmatrix}$
FWH1	$\begin{pmatrix} 0.119 & 0.118 & 0.128 \\ 0.122 & -1.50 & 0.136 \\ 0.115 & 0.127 & 0.141 \end{pmatrix}$	$2.89, \begin{pmatrix} 0.58 & 1.02 & 0.53 \\ 1.08 & 4.71 & 1.10 \\ 0.55 & 1.03 & 0.56 \end{pmatrix}, 6.33$
FWH2	$\begin{pmatrix} 0.098 & 0.124 & 0.093 \\ 0.119 & 0.188 & 0.121 \\ 0.092 & 0.122 & 0.095 \end{pmatrix}$	$0.00, \begin{pmatrix} 0.68 & 1.14 & 0.61 \\ 1.25 & 3.63 & 1.26 \\ 0.63 & 1.12 & 0.66 \end{pmatrix}, 5.53$

TABLE XI
IMPULSIVE NOISE REDUCTION FOR IMAGE SEQUENCES

Filter	Mean Square Error	Mean Absolute error
WM Filter	83.305	1.740
Linear Filter	243.849	7.492

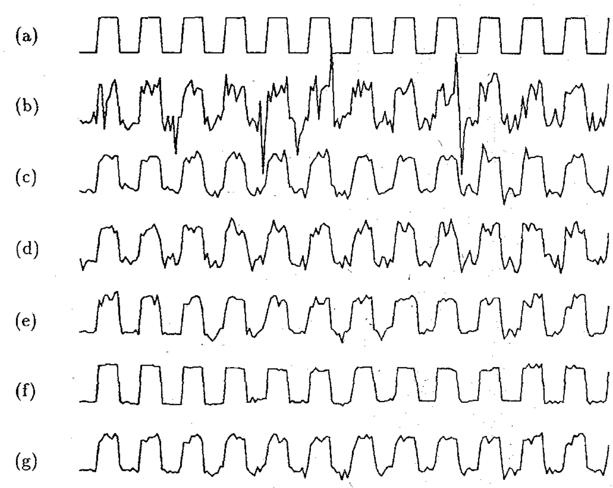


Fig. 12. Segment of desired, noisy, and filtered signals for a square wave and $\Phi(0.05, 10, 150)$ distributed additive contaminated Gaussian noise. (a) 200 samples of a square waveform, (b) noisy signal. (c) adaptive FIR filter, (d) adaptive Ll filter, (e) adaptive WOS filter, (f) adaptive FWH1, and (g) adaptive FWH2.

$X(n_1-1, n_2-1, t)$	$X(n_1, n_2-1, t)$	$X(n_1+1, n_2-1, t)$	line n_2-1, odd field t
$X(n_1-1, n_2, t-1)$	$X(n_1, n_2, t-1)$	$X(n_1+1, n_2, t-1)$	line n_2, even field t-1
$X(n_1-1, n_2+1, t)$	$X(n_1, n_2+1, t)$	$X(n_1, n_2+1, t)$	line n_2+1, odd field t

Fig. 13. Window for intrafield predictive coding. The pixel $X(n_1, n_2+1, t)$ is predicted from the previously reconstructed pixels $X(n_1-1, n_2-1, t), X(n_1, n_2-1, t), X(n_1+1, n_2-1, t)$, and $X(n_1-1, n_2+1, t)$.

$X(n)$. The FWH filter used has the following simple structure,

$$\underline{Y}(n) = WOS[X(n-33), \cdots, X(n-1), Y_{FIR}(n)]^T$$

where $Y_{FIR}(n)$ is the output of an FIR filter of the same window size and at same location. That is, in this FWH filter, only one FIR filter and one WOS filter are to be optimized. The FWH filter is called FWH1 if the FIR and WOS filters are adapted simultaneously. If the FIR filter is a fixed Wiener filter and only the WOS filter is updated, the FWH filter is called FWH2. In this simulation, 10 000 noisy samples are used. The MAE and the MSE values of the filters estimated for four noisy cases are listed in Table XII.

As the above table shows, the FIR filter performs better than the WOS filter in Gaussian noise because the WOS filter has poorer ability to suppress Gaussian noise. However, as the impulsiveness of the noise increases, the performance of the WOS filter becomes better than the linear filter. Even in Gaussian noise, on the other hand, FWH1 and FWH2 give better performance than the FIR filter. This demonstrates again that FWH filters can perform better than the best FIR filters for signals, where edges are common.

The most interesting fact may be that in all the cases, FWH1 performs better than FWH2. That is, optimizing the FIR and WOS filters of the FWH filter independently leads to a suboptimal solution. In fact, checking the parameters of FWH1 and FWH2, it is found that the FIR filter plays different roles in both structures. Instead of a Wiener filter as in FWH2, the FIR filter in FWH1 is designed as an overshooting filter which compensates for the effect of the out of phase samples in the window of the WOS filter [117].

Fig. 12(a) shows a segment of the desired square waveform and Fig. 12(b) shows one observation sequence with contaminated Gaussian noise $\Phi(0.05, 20, 150)$. Estimates of the desired signal produced by the linear FIR filter, Ll filter, WOS filter, and the FWH filters operating on the observation are shown in Fig. 12(c)–(g), respectively.

E. Weighted Median Predictive Coding

Differential Pulse Code Modulation (DPCM) is an effective method to reduce bit rates in digital signal transmission. In the basic DPCM coder, a predictor is used to predict the present sample based on the previous coded information that has already been transmitted. The difference between the

(b)

Fig. 14. Visual evaluation of the reconstructed images, BER=1%. (a). The original picture of the 6th frame of the sequence. The reconstructed frames by (b) the linear predictor. (c). the adaptive linear predictor and (d) the WM predictor.

predicted and the actual values of the sample is then quantized and transmitted. The predictors used in DPCM coders are often linear, as linear prediction theory is well established. However, in linear DPCM coders, image transmission errors tend to propagate and severely degrade the reconstructed image quality [49]. In this case, better results can be achieved

(d)

Fig. 14. *(Continued.)*

by using median type predictors which isolate and do not propagate transmission errors [56], [83], [86].

With the 3 × 3 pixel window shown in Fig. 13, Salo *et al.* found that the following FWH predictor gives the best results [83]

$$\hat{X}(n_1, n_2+1, t) = \mathrm{MED}[X(n_1, n_2-1, t), X_l, X_r, 2 \diamond X(n_1-1, n_2+1, t)] \quad (7.4)$$

TABLE XII
MAE (MSE) VALUES FOR ENHANCING A SQUARE WAVE

Filter Type	Contaminated Gaussian Noise Parameters				
	$\Phi(0,10,0)$	$\Phi(0.02,10,75)$	$\Phi(0.05,10,75)$	$\Phi(0.02,20,150)$	$\Phi(0.05,20,150)$
Identity	7.95 (99.67)	8.98 (198.5)	10.6 (359.8)	18.0 (794.0)	21.1 (1439)
FIR	3.87 (23.22)	4.76 (45.10)	6.08 (79.00)	8.96 (153.2)	11.2 (244.0)
WOS	4.33 (33.91)	4.53 (43.29)	4.86 (59.36)	8.92 (148.1)	9.36 (174.1)
FWH1	2.50 (9.98)	2.60 (11.02)	2.75 (14.30)	5.46 (57.32)	6.16 (90.46)
FWH2	3.46 (19.37)	3.94 (25.98)	4.33 (39.98)	7.83 (105.9)	8.86 (143.0)

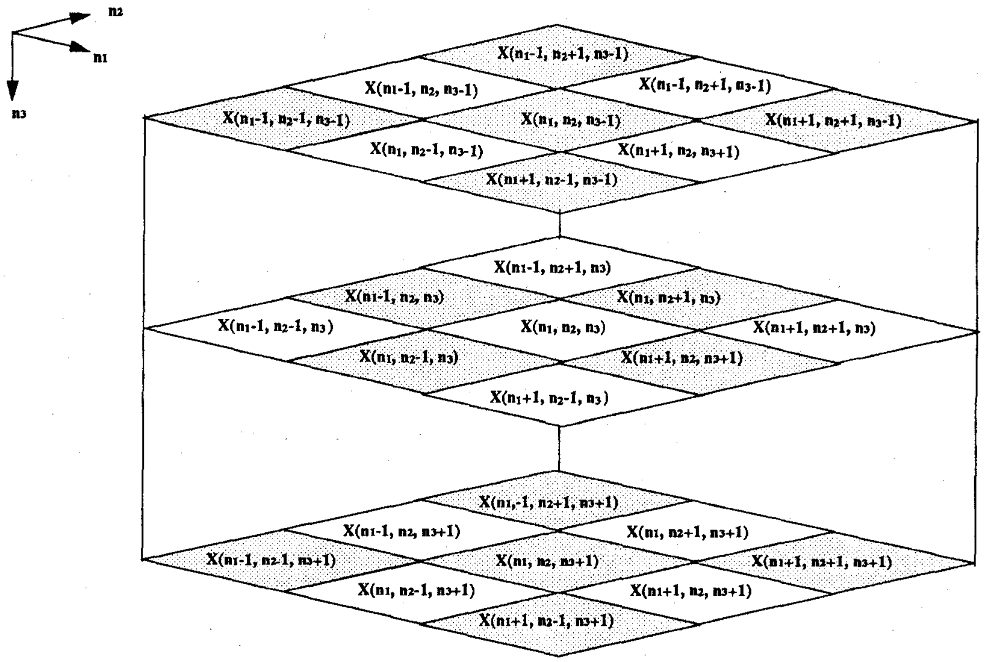


Fig. 15. Three-dimensional quincunx sampling structure corresponding to three picture frames.

where X_l is an overall signal level linear predictor

$$X_l = X(n_1-1, n_2+1, t)/2 + (X(n_1, n_2-1, t) + X(n_1+1, n_2-1, t))/4 \tag{7.5}$$

X_d is an upper nearest neighbor level predictor,

$$X_d = (X(n_1-1, n_2+1, t) + X(n_1, n_2-1, t))/2 \tag{7.6}$$

and X_r is a ramp predictor at -45^0

$$X_r = 2X_d - X(n_1-1, n_2-1, t). \tag{7.7}$$

This nonlinear predictive coder was tested using a digital video sequencer. It is compared to the linear predictive coder defined by (7.5) which is considered as a classical predictor structure in DPCM coding [77].

The test sequences are the moving sequence "Costgirls" and the still image sequence "Home". Table XIII shows the RMS errors for the two DPCM coders. It can be seen that the RMS errors are considerably smaller for the FWH predictor than for the linear predictor.

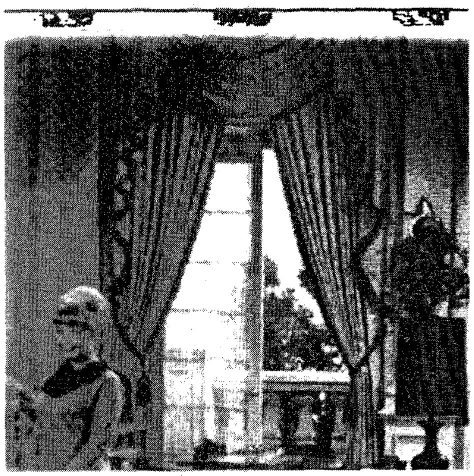
TABLE XIII
MS ERRORS FOR THE DPCM CODERS CALCULATED FOR TWO NATURAL IMAGES TAKEN FROM THE MOVING SEQUENCE "COSTGIRLS" AND THE STILL IMAGE SEQUENCE "HOME"

Coding Method	RMS for Costgirls	RMS for Home
Linear DPCM (3 bits)	11.97	16.37
FWH DPCM (3 bits)	8.68	8.89
Linear DPCM (2 bits)	15.20	20.71
FWH DPCM (2 bits)	11.76	11.31

In the presence of image transmission errors, WM based predictors demonstrate much better performance than the linear predictors which tend to propagate transmission errors. Fig. 14(a)–(d) shows the original sixth frame of the sequence "Costgirls" and the reconstructed images using the linear, adaptive linear and adaptive WM predictors [86], respectively. The bit error rate is 1%.

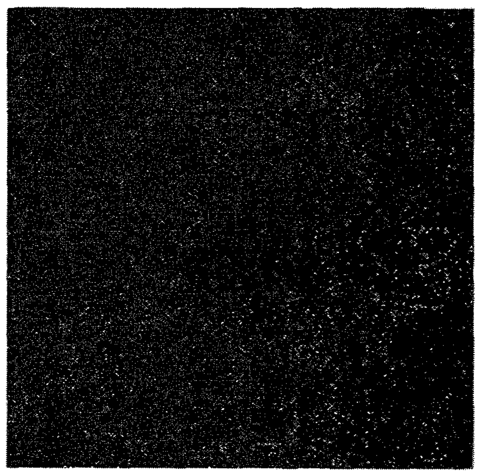
F. Quincunx Coding for Picture Memories

In addition to the above DPCM method, there are some other less used coding approaches. One of them is interpolative



(a)

(b)



(c)

Fig. 16. Quincunx coding of images (a) the original, (b) Pirsch DPCM (16-level quantizer) errors magnified by 50, (c) Field Quincunx coding using the weighted median interpolator, errors magnified by 50.

coding, where a subset of pixels (every other pixel, one pixel out of four, etc.) is transmitted or stored. At the receiver, the missing pixels are interpolated. When every other pixel is taken from offset positions, interpolative coding is called quincunx coding (QC) [1]. This coding method is simpler to implement than DPCM, since no predictors, quantizers, nor feedback loops are needed.

Jarske *et al.* [46], [47] developed several types of median based interpolators and compared the results with linear ones for still images and image sequences. Fig. 15 shows the 3-D sampling structure, where the clear pixels correspond to the image points that need to be reconstructed and the shaded pixels to the available image points. In the case of still images, it was found that the field-quincunx coder with the following WM interpolator gives best results

$$X(n_1, n_2, n_3) = \mathrm{MED}[X(n_1 - 1, n_2, n_3), X(n_1 + 1, n_2, n_3), X(n_1, n_2 - 1, n_3)$$

$$X(n_1, n_2 + 1, n_3), 3 \diamond X(n_1, n_2, n_3 - 1)]. \quad (7.8)$$

Fig. 16 (a) shows part of the original image "home". Fig. 16(b) and (c) shows the difference images between the original and the coded images using the linear interpolator and the WM interpolator, respectively. Certainly, the performance of this interpolator can be further improved by using the samples in frame $n_3 + 1$ at the expense of increased memory.

G. Motion Adaptive Scan Rate Up-Conversion

In current television systems, low field rate and interlaced display standards result in visible artifacts such as line flicker and line crawl [65]. One way to partially remove these anomalies is to up-convert the picture scan rate in the receiver. Two methods are typically used to perform this task: field rate doubling and interlaced to progressive conversion. These tasks have been implemented with fixed and motion compensated interpolators. Methods based on motion vectors are computationally demanding and are currently considered too complex to be used in domestic receivers [25], [75].

(a)

Fig. 17. The "car" sequence: (a) One frame from the sequence (the rectangle marks the part used in the next comparisons); (b) The boxed part is up-converted by the three-point median filter; and (c) The boxed part is up-converted by the weighted median interpolator.

Recently, median based approaches have been introduced [23], [83]. These algorithms do not include motion compensation, thus the complexity is low. However, due to the edge preserving property of median operation, the overall spatial resolution is better than that obtained by commonly used averaging or line repetition. The simplest median filter for scan rate up-conversion is the three-point vertical median filter [23]. The problem with this interpolator is that it causes disturbing serration effects on moving diagonal edges.

To alleviate the above problem, Haavisto *et al.* [38] presented up-conversion algorithms based on WM filter structures and a simple motion detector. The algorithms preserve edges and narrow lines in the images and do not cause visible serration effects. Pixels from two consecutive fields, like the three-point vertical median filter, are used. The estimate of pixel n_1 at the missing line n_2 in field t (from the existing lines n_2-1 and n_2+1 in field t and line n_2 in the previous field $(t-1)$) is given by

$$\begin{aligned} X(n_1,n_2,t) =& \mathrm{MED}[X(n_1-1,n_2-1,t), \\ & X(n_1,n_2-1,t), X(n_1+1,n_2-1,t) \\ & X(n_1-1,n_2+1,t), X(n_1,n_2+1,t), \\ & X(n_1,n_2+1,t), \alpha \diamond X(n_1,n_2,t-1), \beta \diamond \overline{X}] \end{aligned} \quad (7.9)$$

where

$$\overline{X} = \frac{1}{2}[X(n_1,n_2-1,t) + X(n_1,n_2+1,t)]. \quad (7.10)$$

The filter weights α and β were adapted according to the output of the motion detector. The adaptive nature of the filter assures good performance in both stationary and moving picture areas. In fact, the WM interpolator gives quite good results even when no motion detection is used.

The algorithm was tested using an image sequence "car". Fig. 17(a) shows one picture from the sequence. The rectangle marks the part used in the following figures. Fig. 17(b) shows the magnified part extracted from the "car" sequence which is up-converted by the three-point median filter. The problem with the three-point median filter is that it is not able to preserve diagonal narrow lines and, more importantly, it causes disturbing serration effects on moving diagonal edges. These serration artifacts are clearly visible in real sequences. From Fig. 17(b), one can observe that the diagonal edge formed by the windscreen of the car appears staircased as if it were made of blocks larger than one pixel. Fig. 17(c) is the same picture when the motion adaptive up-conversion algorithm using a WM interpolator was applied. The pixel size appears smaller than in Fig. 17(b).

(b)

Fig. 17. *(Continued.)*

VIII. Conclusion

The theory and major applications of weighted median filters are reviewed in this paper. The behavior and performance of WM filters are assessed based on their deterministic and statistical properties. The former is related to the concept of root signals and deterministic convergence; while, the latter deals with, e.g., output distributions and noise attenuation capability. Several interesting similarities between WM filters and linear FIR filters were pointed out.

As WM filters belong to the broader class of stack filters, tools, e.g., the threshold decomposition, developed for the latter are used in the analysis of WM filters. In the binary domain, WM filters are self-dual, linear separable positive Boolean functions. Conversion algorithms between WM filters and positive Boolean functions are reviewed.

Based on this threshold representation, optimal adaptive and nonadaptive weighted median filtering algorithms have been developed in the literature and summarized here. The optimization goal is to minimize the mean absolute error or the mean square error. Furthermore, under a given set of structural constraints, the resulting optimal WM filter will attenuate noise maximally while preserving certain desired signal structures.

To demonstrate the effectiveness of weighted median operations, several important applications are reviewed in the paper, including speech processing, image and image sequence restoration, DPCM coding and quincunx coding, scan rate up-conversion in normal TV and HDTV systems.

Even though a remarkable development has been achieved, the theory of WM filtering is far from mature. There are many open questions and research problems of considerable interest. The following are some examples. First, the root structures of WM filters are not clear yet, except for several special cases. Even for symmetric WM filters, the limits on the number of passes to a root are not known. Full understanding of the root structures of WM filters will certainly make it possible to design optimal WM filters under root structural constraints.

In optimal WM filtering, the reader may notice that the structural approach is convenient because it does not require training signals. However, in cases where nonlinear programming must be used to find the optimal solution, the algorithm can only be efficiently used for relatively small size WM filters with a small number of constraints. Therefore, it is an important research direction to develop computationally more

(c)

Fig. 17. *(Continued.)*

efficient algorithms for the design of optimal WM filters under structural constraints.

In the applications reviewed in the paper, one may notice that the filter window tends to be rather small making it relatively easy to find a good filter. Applications in 1-D signal processing seem to be not as well developed calling for both better design methods for long WM filters and work around novel applications.

Finally, we would like to emphasize that although WM filters have some similarities with linear filters, WM filters cannot replace linear filters and vice versa. This is because WM filters cannot be designed in general to retain or restore some desired signal frequencies and reject others. This is connected with the fact that the weights in a WM filter are nonnegative. On the other hand, linear filters lead to poorer performance at signal edges and in the presence of non-Gaussian noise. Therefore, linear-weighted order statistic hybrid filters may be attractive in many cases. Although these adaptive hybrid filters have produced some interesting and promising results, more work needs to be done along this direction.

ACKNOWLEDGMENT

The authors would like to thank Dr. P. Haavisto, Dr. X. Song, Dr. T. Viero, and T. Jarske for providing their simulation results. Special thanks are also due to G. Colacchio, IEEE Senior Editor, for her patience during the editing of this long paper.

REFERENCES

[1] B. Alp, J. Juhola, Y. Neuvo, and T. Jarske, "Multidimensional reconstruction of quincunx coded image sequences," in *Proc. Picture Coding Symp.*, Cambridge, MA, Mar. 1990, pp. 4.5-1-4.5-3.

[2] G. R. Arce and R. E. Foster, " Detailed preserving ranked-order based filters for image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 83-98, Jan. 1989.

[3] G. R. Arce and N. C. Gallagher, Jr, "State description for root-signal set of median filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, pp. 894-902, Dec. 1982.

[4] ——, "BTC image coding using median filter root," *IEEE Trans. Commun.*, vol. COM-31, pp. 784-793, June 1983.

[5] J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters," *Proc. IEEE*, vol. 78, pp. 690-710, Apr. 1990.

[6] J. Astola and Y. Neuvo, "Matched median filtering," *IEEE Trans. Commun.*, vol. 40, Apr. 1992.

[7] A. Bangham, "Properties of a series of nested median filters, namely the data sieve," *IEEE Trans. Signal Process.*, vol. 41, pp. 31-42, Jan. 1993.

[8] A. Bangham, T. G. Campbell, and M. Gabbouj, "The quality of edge preservation by nonlinear filters," in *Proc. IEEE Workshop Visual Signal Processing Commun., VSPC*, 1992, pp. 37-39.

[9] A. Bangham, T.G. Campbell and R.V. Aldridge, "Multiscale median and morphological filters used for 2D pattern recognition," *Signal Processing*, vol. 38, pp. 387-415, 1994.

[10] K. E. Barner and G. R. Arce, "Permutation filters: a class of nonlinear filters based on set permutations," *IEEE Trans. Signal Process.*, vol. 42, pp. 782-798, Apr. 1994.

[11] A. C. Bovik, T. S. Huang, and D. C. Munson, Jr., "The effect of median filtering on edge estimation and detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 181-194, Mar. 1987.

[12] A. C. Bovik, "Streaking in median filtered images," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, pp. 493-503, Apr. 1987.

[13] D. R. K. Brownrigg, "The weighted median filter," *Commun. ACM*, vol. 27, no. 8, pp. 807-818, Aug. 1984.

[14] ——, "Generation of representative members of an RrSst weighted median filter class," *IEE Proc. Part. F, Commun., Radar, Signal Process.*, vol. 133, pp. 445-448, Aug. 1986.

[15] H. Chen, R. Yang, and M. Gabbouj, "On root structures and convergence properties of weighted median filters," *Circuits, Syst., Signal Process.*, vol. 14, no. 6, 1995.

[16] E. J. Coyle and J. H. Lin, "Stack filters and the mean absolute error criterion," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, pp. 1244-1254, Aug. 1988.

[17] E. J. Coyle, J.-H. Lin, and M. Gabbouj, "Optimal stack filtering and the estimation and structural approaches to image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 2037-2066, Dec. 1989.

[18] E. J. Coyle, M. Gabbouj, and J.-H. Lin, "From median filters to optimal stack filtering," in *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 9-12, June 1991.

[19] E. J. Coyle, M. Gabbouj, and J. Yoo, "Order statistics- based nonlinear filters: stack filters and weighted median filters," (Invited Tutorial) in *Proc. IEEE Winter Workshop Nonlinear Digital Signal Process.*, Tampere, Finland, Jan. 1993, pp. T-3.1–9.

[20] H. Cramer, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946.

[21] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*. New York: Prentice Hall Int., 1982.

[22] E. R. Dougherty and J. Astola, *An Introduction to Nonlinear Image Processing*. New York: SPIE Press, vol. TT16, 1994.

[23] T. Doyle and M. Looymans, "Progressive scan rate conversion using edge information," in *Proc. of 3rd Int. Workshop on HDTV*, vol. III, Torino, Italy, 1989.

[24] K. Egiazarian, "Efficient algorithms for digital filtering and pattern recognition," Ph.D. dissertation, Tampere Univ. Technol., Publication 132, 1994.

[25] G. M. X. Fernando and D. W. Parker, "Motion compensated display conversion," in *Proc. 2nd Int. Workshop Signal Process. HDTV*, L'Aquila, Italy, Feb. 1988, pp. 3–9.

[26] J. P. Fitch, E. J. Coyle, and N. C. Gallagher, Jr., "Median filtering by threshold decomposition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-32, pp. 1183-1188, Dec. 1984

[27] M. Gabbouj and E. J. Coyle, "Minimum mean absolute error stack filtering with structural constraints and goals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, pp. 955-968, June, 1990.

[28] M. Gabbouj, P.-T. Yu, and E. J. Coyle, "Convergence behavior and convergence rates of stack filters," in *Proce. IEEE Int. Symp. Circuits Syst.*, pp. 143-154, June, 1991.

[29] ——, "Convergence behavior and root signal sets of stack filters," *Circuits, Syst. Signal Process.*, Special Issue Median and Morphological Filtering, vol. 11, no. 1, pp. 171-194, 1992.

[30] M. Gabbouj, E. J. Coyle, and N. C. Gallagher, Jr., "An overview of median and stack filtering," *Circuits, Syst. Signal Process.*, Special Issue Median and Morphological Filtering, vol. 11, no. 1, pp. 7-46, 1992.

[31] M. Gabbouj, P. Haavisto, and Y. Neuvo, "Recent advances in median filtering," Invited, in *Proc. 1990 Bilkent Conf. New Trends in Commun., Contr., Signal Process.*, Bilkent Univ., Ankara, Turkey, July 2-5, 1990, Vol. II, E. Arikan Ed. New York: Elsevier Science, B. V., 1990, pp. 1080-1094.

[32] M. Gabbouj, "Weighted median filtering-Striking analogies to FIR filters," *IEEE Circuits Syst. Tutorials*, pp. 5-21, 1994, ch. 1.11.

[33] N. C. Gallagher, Jr. and G. W. Wise, "A theoretical analysis of the properties of median filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, pp. 1136-1141, Dec. 1981.

[34] C. R. Giardina and E. R. Dougherty, *Morphological Methods in Image and Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1988.

[35] E. N. Gilbert, "Lattice theoretic properties of frontal switching functions," *J. Math. Physics*, Vol. 33, pp. 57-67, 1954.

[36] E. Goles, F. Fogelman-Soulie, and D. Pellegrin, "Decreasing energy functions as a tool for studying threshold networks," *Discrete Applied Math.*, vol. 12, 1985.

[37] P. Haavisto, M. Gabbouj, and Y. Neuvo, "Median based idempotent filters," *J. Circuits, Syst., Comput.*, vol. 1, no. 2, pp. 125-148, June 1991.

[38] P. Haavisto, J. Juhola, and Y. Neuvo, "Motion adaptive scan rate upconversion," Invited paper, *Multidimensional Syst. Signal Process. J., Special Issue Multidimensional Process. Video Signals*, vol. 3, no. 2/3, pp. 113-130, May, 1992.

[39] Y. Han, I. Song, and Y. Park, "Some root properties of recursive weighted median filters," *Signal Process.*, vol. 25, no. 3, pp. 337-344, Dec. 1991.

[40] P. Heinonen and Y. Neuvo, "FIR-median hybrid filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, pp. 832-838, June 1987.

[41] R. C. Hardie and K. E. Barner, "Rank conditioned rank selection filters for signal restoration," *IEEE Trans. Image Process.*, vol. 3, pp. 192-206, Mar. 1994.

[42] N. Himayat and S. A. Kassam, "A structure for adaptive order statistics filtering," *IEEE Trans. Image Process.*, vol. 3, pp. 265-280, May 1994.

[43] S. T. Hu, *Threshold Logic*. Berkeley, CA: Univ. California Press, 1965.

[44] T. S. Huang and Y. P. Hsu, "Image sequence enhancement," in T.S. Huang, Ed., *Image Sequence Analysis*. Springer-Verlag, Berlin, 1981, ch. 1.

[45] S. Inkinen, *Nonlinear readout signal processing for particle detectors*, Ph.D. dissertation, Tampere Univ. Technol., Publication 161, 1995.

[46] T. Jarske, K. Saarinen, J. Juhola, and Y. Neuvo, " Quincunx coding for picture memories," in *Proc. 131st SMPTE Tech. Conf.*, Los Angeles, CA, Oct. 21-25, 1989, pp. 1–18.

[47] T. Jarske and Y. Neuvo, "Adaptive DPCM with median type predictors," *IEEE Trans. Consumer Electron.*, vol. 37, pp. 348-352, Aug. 1991.

[48] N. S. Jayant, "Average and median-based smoothing technique for improving quality in the presence of transmission errors," *IEEE Trans. Commun.*, vol. COM-24, pp. 1043-1045, Sept. 1976.

[49] N. S. Jayant and P. Null, *Digital Coding of Waveforms: Principles and Application to Speech and Video*. Englewood Cliffs, NJ: Prentice Hall, 1984.

[50] B. Jeong and Y. H. Lee, "Design of weighted order statistic filters using the perceptron algorithm," *IEEE Trans. Signal Process.* vol. 42, pp. 3264-3269, Nov. 1994.

[51] B. I. Justusson, "Median filtering: statistical properties," in *Two-Dimensional Digital Signal Processing II*, T. S. Huang, Ed. New York: Springer-Verlag, 1981.

[52] K. J. Kerpez, "Minimum mean squared error impulse noise estimation and cancellation," *IEEE Trans. Signal Process.*, vol. 43, pp. 1651-1662, July 1995.

[53] C. Kotropoulos and I. Pitas, "Multichannel L filters based on marginal data ordering," *IEEE Trans. Signal Process.* vol. 42, pp. 2581-2595, Oct. 1994.

[54] P. Kuosmanen, "Statistical analysis and optimization of stack filters," Ph.D. dissertation, Tampere Univ. Technol., Publication EI 77, 1994.

[55] K. D. Lee and Y. H. Lee, "Threshold Boolean filters," *IEEE Trans. Signal Process.* vol. 42, pp. 2022-2036, Aug. 1994.

[56] Y. H. Lee, D. H. Kang, J. H. Choi, and K. D. Lee, "DPCM with median predictors," *Nonlinear Image Processing III*, E. R. Dougherty, J. Astola, and C. Boncelet, Jr., Eds., *Proc. SPIE* vol. 1658, pp. 199-209. Feb. 1992.

[57] P. M. Lewis II and C. L. Coates, *Threshold Logic*. New York: Wiley, 1967.

[58] J. H. Lin and E. J. Coyle, "Minimum mean absolute error estimation over the class of generalized stack filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, pp. 663-678, Apr. 1990.

[59] J. H. Lin, T. M. Sellke, and E. J. Coyle, "Adaptive stack filtering under the mean absolute error criterion," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, pp. 938-954, June 1990.

[60] B. W. Lindgren, *Statistical Theory*, 3rd ed. New York: Macmillan, 1976.

[61] T. Loupas, W. N. McDicken, and P. L. Allan, "An adaptive weighted median filter for speckle suppression in medical ultrasonic images," *IEEE Trans. Circuits Syst.*, vol. 36, pp. 129-135, Jan 1989.

[62] P. Maragos and R. W. Schafer, "Morphological filters - Part I: Their set theoretic analysis and relations to linear shift-invariant filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, pp. 1153-1169, Aug. 1987.

[63] ——, "Morphological filters - Part II: their relations to linear median, order-statistic, and stack filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, pp. 1170-1184, Aug. 1987.

[64] S. Muroga, *Threshold Logic and Its Applications*. New York: Wiley Intersci., 1971.

[65] Y. Neuvo, "Interpolators for TV scanning rate conversions," in *1st World Electronic Symp. ITU COM 89*, Palexpo, Geneva, Switzerland, Oct. 4-7, 1989, pp. 203-207.

[66] A. Nieminen, P. Heinonen, and Y. Neuvo, " A new class of detail-preserving filters for image processing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 74-90, Jan. 1987.

[67] J. Nieweglowski, M. Gabbouj, and Y. Neuvo, "Weighted medians-positive Boolean functions conversion," *Signal Processing*, vol. 34, no. 2, pp. 149-162, Nov. 1993.

[68] J. Nieweglowski, L. Yin, M. Gabbouj, and Y. Neuvo "Optimal WOS filters under structural constraints," in *Proc. ISCAS 92, IEEE Int. Symp.*

Circuits Syst., San Diego, CA, May 10-13, 1992, pp. 2621–2624.

[69] T. A. Nodes and N. C. Gallagher, Jr., "Median filters: some modifications and their properties,"*IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, pp. 739-746, Oct. 1983.

[70] K. Öistämö, and Y. Neuvo, "Vector median operation for color image processing," in *Proc. 1990 SPIE/SPSE Symp. Electronic Imaging*, Santa Clara, CA, Feb. 11-16, 1990, pp. 2-12.

[71] ——, "Video signal processing using vector median," in *SPIE Conf. Visual Commun. Image Process.*, Lausanne, Switzerland, Oct. 2-4, 1990, pp. 1171-1183.

[72] F. Palmieri and C. G. Boncelet, Jr., "Ll-filter - A new class od order statistic filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol 37, pp. 691-701, May 1989.

[73] F. Palmieri, "A backpropagation algorithm for multilayer hybrid order statistic filters," in *Proc. 1989 IEEE Int. Conf. Acoust., Speech Signal Process.*, Glasgow, Scotland, 1989, pp. 1179-1182.

[74] F. Palmieri and R. E. Croteau, "Hybrid order statistic filters for adaptive image restoration," in *1990 SPIE/SPSE's Symp. Electronic Imaging: Science Technol.*, San Jose, CA, Feb. 1990.

[75] D. Pele, P. Siohan, and B. Choquet, "Field rate conversion by motion estimation/compensation," in *Proc. 2nd Int. Wkshp. Signal Process. HDTV*, L'Aquila, Italy, Feb. 1988, pp. 3–12.

[76] S. S. Perlman, S. Eisenhandler, P. W. Lyons, and M. J. Shumila, "Adaptive median filtering for impulse noise elimination in real-time TV signals," *IEEE Trans. Commun.*, vol. 35, pp. 646-652, June 1987.

[77] P. Pirsch, "A new predictor design for DPCM coding of TV signals," in *Conf. Rec. ICC'80*, June 1980, pp. 31.2.1-31.2.5, also "Design of a DPCM codec for VLSI realization in CMOS technology," *Proc. IEEE*, vol. 73, pp. 592-598, Apr. 1985.

[78] I. Pitas and A. N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*. Boston, MA: Kluwer Academic, 1990.

[79] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1991.

[80] M. K. Prasad and Y. H. Lee, "Weighted median filters: generation and properties," in *Proc. IEEE Int. Symp. Circuits Syst., ISCAS-89*, pp. 425-428.

[81] ——, "Stack filters and selection probabilities," *IEEE Trans. Signal Process.,* vol. 42, pp. 2628-2643, Oct. 1994.

[82] K. Saarinen and Y. Neuvo, "An adaptive weighted median filter based on LMS algorithm," in *Proc. 1990 Bilkent Int. Conf. New Trends Commun., Contr. Signal Process.*, Ankara, Turkey, June 1990, pp. 1241-1248.

[83] J. Salo, Y. Neuvo, and V. Hämeenaho, "Improving TV picture quality with linear median type operations," *IEEE Trans. Consumer Electron.*, vol. 34, pp. 373-379, Aug. 1988.

[84] J. Serra, *Image Analysis and Mathematical Morphology*. London: Academic, 1982.

[85] C. L. Sheng, *Threshold Logic*. New York: Academic, 1969.

[86] X. Song, "Video coding using weighted median prediction and nonlinear pyramid vector quantization," Ph.D. dissertation, Tampere Univ. Technol., Finland, Sept. 1993.

[87] T. Sun, M. Gabbouj, and Y. Neuvo, "Center weighted median filters: some properties and application in image processing," *Signal Process.*, vol. 35, no. 3, pp. 213-229, Feb. 1994.

[88] ——, "Deterministic properties of center weighted median filters," in *Proc. 1992 Int. Conf. Commun. Technol.*, Beijing, 1992, pp. 06.06.1-4.

[89] I. Tabus, M. Gabbouj, and L. Yin, "Real domain adaptive WOS filtering using neural network approximation," in *Proc. IEEE Winter Workshop Nonlinear Digital Signal Process.*, Tampere, Finland, Jan. 1993, pp. 3.2-2.1–4.

[90] I. Tabus and M. Gabbouj, "Applications of neural networks training methods to optimal WOS image filtering," in *Proc. IEEE Visual Signal Process. Commun.*, Melbourne, Australia, Sept. 20-22, 1993, pp. 155-158.

[91] I. Tabus, "Training and model based approaches for optimal stack and boolean filtering with applications in image processing," Ph.D. dissertation, Tampere Univ. Technol., Publication 155, 1995.

[92] J. W. Tukey, "Nonlinear (Nonsuperposable) methods for smoothing data," in *Conf. Rec. EASCON'74*, 1974, p. 673.

[93] S. G. Tyan, "Median filtering: deterministic properties," in *Two Dimensional Digital Signal Processing, II: Transforms and Median Filters*, vol. 42, ch.5, pp. 197-217; *Topics in Applied Physics*, T. S. Huang, Ed. New York: Springer-Verlag, 1981.

[94] T. Viero and Y. Neuvo, "3-D median structures for image sequence filtering and coding," in *Motion Analysis and Image Sequence Processing*, M. I. Sezan and R. L. Lagendijk, Eds. Boston, MA: Kluwer Academic, 1993.

[95] ——, "Non-moving regions preserving median filters for image sequence filtering," in *Proc. IEEE Int. Conf. Syst. Eng.*, Dayton, OH, Aug. 1991, pp. 245–248.

[96] Q. Wang, M. Gabbouj, and Y. Neuvo, "Root properties of morphological filters,"*Signal Process.*, vol. 34, no. 2, pp. 131-148, Nov. 1993.

[97] ——, "Root-signal sets of morphological filters and their use in variable-length BTC image coding," *J. Mathemat. Imaging Vision*, vol. 2, pp. 155-171, 1992.

[98] P. D. Wendt, E. J. Coyle, and N. C. Gallagher, Jr., "Stack filter," *IEEE Trans. Acoust., Speech, Signal Process.* vol. ASSP-34, pp. 898-911, Aug. 1986.

[99] P. D. Wendt, "Nonrecursive and recursive stack filters and their filtering behavior," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, pp. 2099-2107, Dec. 1990.

[100] P. D. Wendt, E. J. Coyle, and N. C. Gallagher, Jr., "Some convergence properties of median filters," *IEEE Trans. Circuits Systems*, vol. CAS-33, pp. 276-286, Mar. 1986.

[101] R. Wichman, J. Lilleberg, and Y. Neuvo, "Matched median detectors for multiple-access channels," *Signal Process.,* Special Issue Nonlinear Signal Process., vol. 38, pp. 345-357, Aug. 1994.

[102] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1985.

[103] B. Widrow, R. G. Winter, and R.A. Baxter, "Layered neural nets for pattern recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, pp. 1109-1118, July 1988.

[104] R. Yang, L. Yin, M. Gabbouj, J. Astola, and Y. Neuvo, "Optimal weighted median filtering under structural constraints," *IEEE Trans. Signal Process.*, vol. 43, pp. 591-604, Mar. 1995.

[105] R. Yang, L. Yin, M. Gabbouj, and Y. Neuvo, "2-D optimal weighted median filters," in *Proc. 1992 IEEE Workshop Visual Signal Process. Commun.*, North Carolina State Univ., Raleigh, NC, Sept. 1992, pp. 36-39.

[106] R. Yang, L. Yin, M. Gabbouj, J. Astola, and Y. Neuvo, "Performance of detail-preserving weighted median filters for image processing," in *Proc. IEEE Winter Workshop Nonlinear Digital Signal Process.*, Tampere, Finland, Jan. 1993, pp. 5.1-3.1-5.1-3.6.

[107] ——, "Optimal weighted median filters under structural constraints," in *Proc. IEEE Int. Symp. Circuits Syst.*, Chicago, IL, May 2-6, 1993, pp. 942-945.

[108] R. Yang, M. Gabbouj, and Y. Neuvo, "Fast algorithms for analyzing and designing weighted median filters," *Signal Process.*, vol. 41, pp. 135-152, 1995.

[109] ——, "On convergence behavior of symmetric weighted median filters," in *Proc. Euro. Conf. Circuit Theory Design*, Davos, Switzerland, Aug. 30–Sept. 3, 1993, pp. 1053-1058.

[110] ——, "Parametric analysis of weighted order statistics filters," *IEEE Signal Process. Lett.*, vol. 1, pp. 95–98, June 1994.

[111] L. Yin, J. Astola, and Y. Neuvo, "Adaptive weighted order statistic filters," in *Int. Wkshp. Algorithms Parallel VLSI Architectures*, Part B, Pont-a-Mousson, France, June 10-16, 1990, pp. 327-332; included in *Algorithms and Parallel VLSI Architectures*, E. F. Deprettere, Ed.

[112] ——, "Optimal weighted order statistic filters under the mean absolute error criterion," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, Canada, May 1991, pp. 2529-2532.

[113] ——, "Adaptive stack filtering with application to image processing," *IEEE Trans. Signal Process.*, vol. 41, pp. 162-184, Jan. 1993.

[114] ——, "Adaptive multistage weighted order statistic filters based on the back-propagation algorithm," *IEEE Trans. Signal Process.*, vol. 42, pp. 419-422, Feb. 1994.

[115] ——, "Adaptive weighted median filtering under the mean absolute error criterion," in *Proc. IEEE Wkshp. Visual Signal Process. Commun.*, June 6-7, 1991, Hsinchu, Taiwan, pp. 184-187.

[116] ——, "Optimal generalized weighted order statistic filters," in *Int. Conf. Visual Commun. Image Process. '91: Sixth in a Series*, Nov. 10-13 1991, Boston, MA.

[117] L. Yin and Y. Neuvo, "Fast adaptation and performance characteristics of FIR-WOS hybrid filters," *IEEE Trans. Signal Process.*, vol. 42, pp. 1610-1628, July 1994.

[118] L. Yin, "Stack filter design: A structural approach," *IEEE Trans. Signal Process.*, vol. 43, pp. 831-840, Apr. 1995.

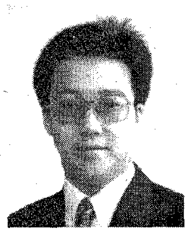
[119] L. Yin, J. Astola, and Y. Neuvo, "Adaptive Boolean filters for noise reduction," in *Proc. Int. Conf. Artificial Neural Networks,* Espoo, Finland, June 24-28, 1991, pp. 759-764.

[120] O. Yli-Harja, J. Astola, and Y. Neuvo, "Analysis of the properties of median and weighted median filters using threshold logic and stack filter representation," *IEEE Trans. Signal Process.*, vol. 39, pp. 395-410, Feb. 1991

[121] P.-T. Yu and E. J. Coyle, "The classification and associative memory capability of stack filters," *IEEE Trans. Signal Process.*, vol. 40, pp. 2483-2497, Oct. 1992.

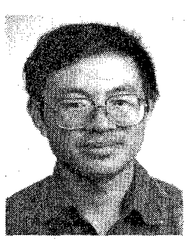
[122] P.-T. Yu and W.-H. Liao, "The convergence behavior of symmetric

weighted median filters," in *Proc. IEEE Winter Workshop Nonlinear Digital Signal Process.*, Tampere, Finland, Jan. 1993, pp. 5.1-4.1-5.1-4.6.
[123] ——, "Weighted order statistics filters -Their classification, some properties, and conversion algorithm," *IEEE Trans. Signal Process.*, vol. 42, pp. 2483-2497, Oct. 1994.
[124] B. Zeng, "Optimal median-type filtering under structural constraints," *IEEE Trans. Image Process.*, vol. 4, pp. 921-931, July 1995.
[125] B. Zeng, M. Gabbouj, and Y. Neuvo, "A unified design method for rank order, stack and generalized stack filters based on classical Bayes decision," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 1003-1020, Sept. 1991.
[126] B. Zeng and Y. Neuvo, "Optimal parallel stack filtering under the mean absolute error criterion," *IEEE Trans. Image Process.*, vol. 3, pp. 324-327, May 1994.



Lin Yin received the B.S.E.E. degree from the Shandong University of Technology, China, in 1983, the M.S.E.E. degree from the Xi'an Jiaotong University, China, in 1986, and the Doctor of Technology degree in electrical engineering from the Tampere University of Technology, Finland, in 1991.

From 1986 to 1989 he was a Teacher with the Department of Information and Control Engineering, Xi'an Jiaotong University. From 1989 to 1992 he was with the Signal Processing Laboratory, Tampere University of Technology, as a Researcher. In 1993 he worked in the R&D Center, Nokia Mobile Phones, Oulu, Finland, as a DSP engineer. Since 1995 he has been with Audio Systems Laboratory, Nokia Research Center, Tampere, Finland, as a senior research engineer, where he heads a team working on audio compression algorithms. His research interests include nonlinear signal processing and audio compression.



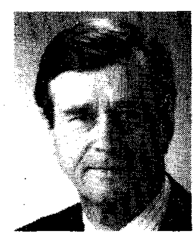
Ruikang Yang (S'93–M'95) received the B.S.E.E. and M.S.E.E. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 1983 and 1986, respectively, and the Doctor of Technology degree in electrical engineering from the Tampere University of Technology, Finland, in 1994.

From 1988 to 1991 he was a lecturer with the Department of Electronic Engineering, University of Electronic Science and Technology of China. He was with Coforce Computer Corporation during 1987–1991. In 1991 he joined the Signal Processing Laboratory, Tampere University of Technology, Finland, as a Researcher. Since 1994, he has been a senior research engineer at Nokia Research Center, Tampere, Finland, where he conducts research on speech recognition in noise environments. His current research interests include speech recognition, speech coding and nonlinear signal processing.

Moncef Gabbouj (S'85–M'90–SM'95) received the B.S. degree in electrical engineering in 1985 from Oklahoma State University, Stillwater. He received the M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN in 1986 and 1989, respectively.

From 1994 to 1995, he was an Associate Professor with the Signal Processing Laboratory of Tampere University of Technology, Tampere, Finland. From 1990 to 1993, he held the position of Senior Research Scientist with the Research Institute for Information Technology, Tampere, Finland. Currently, he is a Professor with the Department of Information Technology of Tampere University of Technology, Pori, Finland. His research interests include nonlinear signal and image processing and analysis, mathematical morphology, and neural networks.

Dr. Gabbouj is an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, and Guest Editor of the European journal, *Signal Processing*, Special Issue on Nonlinear Digital Signal Processing (August 1994). He is the past Chairman of the IEEE Circuits and Systems Society, Technical Committee on Digital Signal Processing, and the IEEE SP/CAS Chapter of the IEEE Finland Section. He is also the DSP Track Chair of the 1996 IEEE ISCAS and the Program Chair of NORSIG'96. He is the Director of the International University Program in Digital Signal Processing in the Signal Processing Laboratory at Tampere University of Technology. He is a member of Eta Kappa Nu, Phi Kappa Phi, and the IEEE SP and CAS societies. He was a co-recipient of the Myril B. Reed Best Paper Award from the 32nd Midwest Symposium on Circuits and Systems. He was also co-recipient of the NORSIG'94 Best Paper Award from the 1994 Nordic Signal Processing Symposium.



Yrjö Neuvo (S'70–M'74–SM'82–F'89) received the Diploma Engineer and Licentiate of Technology degrees from Helsinki University of Technology, in 1968 and 1971, respectively, and the Ph.D degree in electrical engineering from Cornell University, Ithaca, NY, in 1974.

He held various research and teaching positions at Helsinki University of Technology, the Academy of Finland, and Cornell University from 1968 to 1976. Since 1976 he has been a Professor of electrical engineering at the Tampere University of Technology, Tampere, Finland. The academic year 1981–1982 he was with the University of California, Santa Barbara as a Visiting Professor. Since 1984 he has held the position of National Research Professor at the Academy of Finland. Since 1993 he has been with Nokia Corporation, currently as Senior Vice President of R&D at Nokia Mobile Phones. His main research interests are in digital signal and image processing algorithms and their implementations.

Dr. Neuvo is a member of EURASIP, Phi Kappa Phi, the Finnish Academy of Technical Sciences, Foreign Member of the Royal Swedish Academy of Technical Sciences, and a member of the Academiae Europae. He is Honorary Doctor of Medicine from Tampere University. He is a member of the Editorial Board of *Signal Processing*, Associate Editor of *Circuits, Systems and Signal Processing*, and Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He serves as a memeber of the Signal Processing Committee of the IEEE Circuits and Systems Society. He is a member of EU's Science and Technology Assembly (1994-present). He was a Distinguished Lecturer of the IEEE Circuits and Systems Society during 1991–1995. He was a member of the Committee for Development of Science and Technology in Europe during 1990–1994. He was the President of the Society of Electronics Engineers in Finland in 1978-1980 and he was the General Chairman of the 1988 IEEE International Symposium on Circuits and Systems held in Helsinki, Finland.