AFFINE ORDER STATISTIC FILTERS *: A DATA-ADAPTIVE FILTERING FRAMEWORK FOR NONSTATIONARY SIGNALS †

Alexander Flaig[1] *Gonzalo R. Arce*[1] *Kenneth E. Barner*[1 2]

[1] Department of Electrical Engineering
University of Delaware, Newark, DE 19716
flaig@ee.udel.edu

[2] A.I. duPont Institute/University of Delaware
Applied Science & Engineering Laboratories
Wilmington, DE 19899

ABSTRACT

We introduce a novel, data-adaptive, and robust filtering framework: *affine order-statistic filters*. Affine order-statistics relate classical order-statistics to observations in their natural order and thus inherently yield a meaningful data representation. Affine order-statistic filters exploit this notion to *adaptively* process *non*stationary signals. Affine order-statistic filters overcome many of the limitations associated with traditional order-statistic filters, in particular: filters in this class are *parsimonious in the number of filter coefficients*, they are *statistically efficient* in a wide range of signal statistics, and they admit real-valued filter weights leading to a *wide-range of filtering characteristics*. The class of affine order statistic filters contains two families: the *WOS affine filter* class whose structure can adapt, according to the observed data, from an FIR linear filter to a WOS filter, and the *FIR affine filter* class whose structure can adapt from an L-filter to an FIR-filter. In this paper we introduce the *median affine filter* and the *center affine filter* as representatives of each class, and show their performance in two applications where the signals are non-stationary in nature.

1. INTRODUCTION

Order-statistics have found considerable attention in robust signal processing. The running median and L-filters are frequently cited examples of filters based on order-statistics. These filters overcome most of the limitations associated with linear filters in non-Gaussian environments as they can neglect outliers, closely follow signal discontinuities, and effectively preserve details in multi–dimensional imaging signals. Weighted order statistic (WOS) filters [1], constitute a more flexible class of order-statistic filters as they resemble linear FIR filters in certain respects. However, WOS filters are tuned to time-series obeying double-exponential distributions and are therefore very inefficient in processing Gaussian or near Gaussian signals.

Several hybrid filtering algorithms have been proposed to overcome these limitations. All of them in some fashion combine the sorting and linear-combination structures of order-statistic and linear filters, respectively. Among these, FIR-WOS hybrid filters [2] and permutation $L\ell$ filters have generated interest [3]. Unfortunately, FIR-WOS hybrid filters and permutation Lℓ have limitations of their own: the multi-stage structure of FIR-WOS hybrid filters typically requires a large number of filter coefficients and the WOS aspect restricts them to be smoothers. While permutation filters can be designed to have a wide-range of filtering characteristics, they require many additional filter coefficients compared to linear filters and, moreover, their computational complexity increases very rapidly with the filter window size.

In this paper we introduce the class of *affine order-statistic filters*, which overcome many of the limitations of the previously cited schemes while preserving all their desirable characteristics. Affine order-statistic filters are *robust*, they *exploit* both *temporal and rank-order characteristics* of the data, and they admit real-valued filter coefficients leading to a *wide-range of filtering characteristics*. Furthermore, their *complexity is parsimonious* using as few as $N+1$ *filter weight coefficients*. More importantly, affine order-statistic filters have an additional property not found in either FIR-WOS hybrid filters or permutation filters: they allow *data-adaptive* filtering. It is well known that data-adaptiveness is imperative for effective non-stationary signal processing. Affine order-statistic filters accommodate to possibly time-varying signal characteristics by adjusting their weights automatically. In particular, it can be shown that the behavior of affine order-statistic filters can resemble that of linear filters in near-Gaussian environments and that of order-statistic filters in impulsive noise [4]. In ambiguous cases affine order-statistic filters operate like hybrid structures, considering both temporal-and rank-order information simultaneously.

2. AFFINITY MEASURES

The processing of signals typically involves the sliding of an observation window over an input sequence, and at each window location forming an estimate of some underlying process. As observation samples are generally corrupted from some combination of the acquisition, transmission, and storage processes, the estimate of an underlying process must consider the *reliability* of individual observation samples. This is especially important for processes corrupted by heavy tailed noise. For such signals, sample rank is the most widely used and researched measure of reliability. Rank, however, is a crude integer domain measure that ignores the dispersion of samples. In the following we define a real–valued measure of reliability reflecting sample dispersion based on the affinity to a reliable reference sample.

Consider a set of N real–valued observation samples $x_1, x_2, \ldots, x_N$ in their natural order and ordered according to rank $x_{(1)}, x_{(2)}, \ldots, x_{(N)}$, where $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(N)}$, then x_i corresponds to $x_{(r_i)}$ if r_i is the rank of x_i.

Numerous filtering algorithms, including median, order statistic, α–trimmed mean, $L\ell$, and WOS filters, rely on *ranking* to identify unreliable samples. The reliance on rank explicitly assumes that centrally ranked samples are reliable and samples in the extremes of the ordered set are unreliable. For instance, an α–trimmed mean filter considers the

* FOR MORE INFORMATION ON AFFINE ORDER–STATISTIC FILTERS AND AN INTERACTIVE JAVA DEMONSTRATION SEE HTTP://WWW.EECIS.UDEL.EDU/EE/SIGNALS/ROBUST/INDEX.HTML

† THIS WORK WAS SUPPORTED IN PART BY THE ARMY RESEARCH LABORATORY UNDER COOPERATIVE AGREEMENT NO. DAAL01-96-2-0002, AND BY A FELLOWSHIP FROM THE GERMAN ACADEMIC EXCHANGE SERVICE.

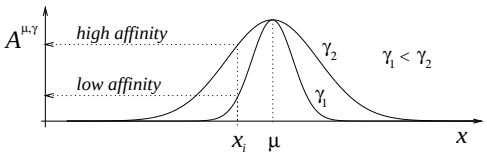

Figure 1. The affinity function $\mathcal{A}^{\mu,\gamma}$ assigns a low or high affinity to the sample x_i for $\gamma_1 < \gamma_2$ respectively.

α smallest and largest samples unreliable and the $N+1-2\alpha$ centrally ranked samples reliable. The unreliable samples are discarded, and an estimate is formed from the reliable samples, treating them all equally. While other order statistic based methods, such as $L\ell$ filters, vary the weight given to a sample according to its rank, the strict reliance on rank as a characterization of reliability can lead to poor performance [4].

An effective measure of reliability should reflect not only the rank but also the *dispersion* of the observation samples. Such a measure can then be used to gage the reliability of an observation and accordingly modify its contribution to the estimate in the filtering process. The approach developed here is based on the *affinity* of samples to a *reliable reference point* μ, for instance the sample median. The proximity of each sample to μ is checked by an *affinity function* that returns a real–value affinity measure. Those samples close to μ are said to have a high affinity, and those samples distant a low affinity.

The affinity function $\mathcal{A}^{\mu,\gamma}$ is defined as a mapping of the set of observation samples to the interval $[0,1]$:

$$\mathcal{A}^{\mu,\gamma} : x_i \mapsto \mathcal{A}_i^{\mu,\gamma} \in [0,1] \tag{1}$$

for $i = 1, 2, \ldots, N$, where the parameters $\mu \in (-\infty, \infty)$ and $\gamma \in [0, \infty)$ control the location and spread of the affinity function, respectively. In general, $\mathcal{A}^{\mu,\gamma}$ is a nonlinear function. While many forms of affinity measures can be adopted, we impose the following restrictions: (a) The affinity function is *unimodal* with mode μ and assumes unity value at the mode, (b) the affinity function is a *nondecreasing function of* γ, i.e., $\mathcal{A}_i^{\mu,\gamma_1} \geq \mathcal{A}_i^{\mu,\gamma_2}$ for $\gamma_1 \geq \gamma_2$, and (c) the affinity function reduces to a *delta function* at the mode for $\gamma \to 0$, and is *uniform* for all inputs, for $\gamma \to \infty$.

These restrictions have the following intuitive interpretations. The parameter μ defines a reference location. Samples proximal to this location, or having similar value, are assigned high affinity (≈ 1) while distant samples are assigned low affinity (≈ 0). The scale on which the transition from proximal to distant occurs is controlled by γ. For small γ ($\to 0$), only samples equivalent to μ are proximal, while for γ large ($\to \infty$), all samples are proximal. In terms of observation samples, the affinity function defines a neighborhood around μ, within which samples are considered reliable.

In this paper we restrict ourselves to the *Gaussian affinity function*, which defines a smooth transition between high and low affinity, as shown in Fig. 1:

Definition: For $x \in R$, the *Gaussian affinity function* performs the mapping

$$\mathcal{A}^{\mu,\gamma} : x \mapsto e^{-\frac{(x-\mu)^2}{\gamma}}, \tag{2}$$

where $\mu \in (-\infty, \infty)$ and $\gamma \in (0, \infty)$.

The parameter γ, allows for tuning of the transition from high to low affinity and therefore modification of the neighborhood around μ with which we associate reliability.

3. THE MEDIAN AFFINE FILTER

The median affine filter forms its estimate based on N observations taking into consideration the *reliability* of each of them by an affinity function which relates each sample to the median, the most reliable order statistic. Since reliability is a fuzzy measure, represented by a real-valued number, the median affine filter weights each observation to control its impact on the estimate, rather than discarding single samples by performing hard decisions.

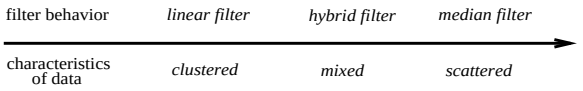

Figure 2. Behavior of the median affine filter depending on the local characteristics of the data.

Definition: The output $\hat{d}$ of the *median affine filter* is defined as

$$\hat{d}^{\gamma} = \frac{\sum_{i=1}^{N} \mathcal{A}_i^{\gamma} x_i w_i}{\sum_{i=1}^{N} \mathcal{A}_i^{\gamma} w_i}, \tag{3}$$

where x_i is the i^{th} temporally-ordered observation, $\mathcal{A}_i^{\gamma}$ denotes its associated affinity measure, and the w_i's are filter coefficients.

Note, that the product $\mathcal{A}_i^{\gamma} x_i$ approximates x_i whenever x_i is located within the fuzzy neighborhood defined by the affinity function. When x_i is distant from $x_{(med)}$, which is the case when x_i is an outlier, for instance, the product $\mathcal{A}_i^{\gamma} x_i$ tends to zero. This has the following intuitive interpretation: Whenever x_i is classified *reliable* its value is translated into the product $\mathcal{A}_i^{\gamma} x_i \approx x_i$. If x_i is considered *un*reliable, its affinity measure forces the product $\mathcal{A}_i^{\gamma} x_i$ to zero, and therefore limits the impact of x_i on the estimate.

The filter coefficients w_i have a similar meaning as in linear FIR filters. However, in the proposed filter, they operate on the affinity weighted samples rather than the unmodified observations.

We shall see soon that the values of the affinity measures depend on the dispersion of the data. These values can be such that an offset would be introduced by a simple summation of the weighted samples, therefore, a normalization as in (3) is necessary to guarantee *unbiasedness* of the filter as a location estimator.

In the following we find it useful to collect the samples with a high affinity to the median in the *set of median affine observations*: $\mathcal{MAO} = \{x_i : \mathcal{A}_i^{\gamma} \approx 1\}$ [1].

Tunable Filter Function: The median affine filter emerges from a synthesis of the median filter and the linear FIR filter. The parameter γ controls the impact of these structures on the behavior of the median affine filter. While a small γ puts little weight on the natural order of the observations and strong emphasis on their rank-order, a large value of γ stresses the linear part and uses rank-order only to reject outliers. In particular, we observe the following limiting cases: For $\gamma \to \infty$ the median affine filter reduces to a *normalized linear FIR filter*, and for $\gamma \to 0$ it reduces to the classical *median filter*. For any other positive value of γ the median affine filter behaves like an *hybrid filter*, utilizing both temporal and rank-order information simultaneously.

Data-adaptiveness: Suppose γ is fixed, then the *average* behavior of the median affine filter is governed by the statistics of the affinity measures [4]. The *local* behavior, however, is a function of the *dispersion* of the current observations relative to the spread of the affinity function. Provided that $0 < \gamma < \infty$, we can distinguish three cases:

[1] Readers familiar with fuzzy set theory may notice that the set of median affine observations is actually a *fuzzy set*, where the membership value is given by the affinity measure of each observation.

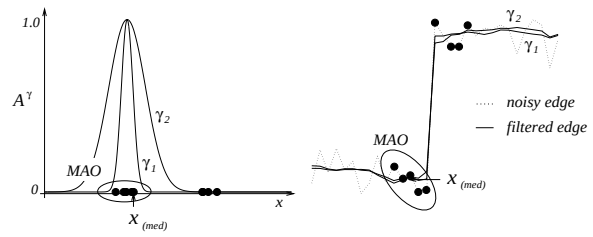

Figure 3. Edge preservation property: (left) Assigning affinity measures to observations from a noisy step edge. (right) Noisy (dotted) and median affine filtered edge (solid) for γ_1 and γ_2, where $\gamma_1 < \gamma_2$. The larger value of γ results in increased smoothing left and right of the edge.

(a) when the observations are *clustered* around the sample median, all of them are contained in the set $\mathcal{MAO}$, thus the median affine estimator reduces to a *linear FIR filter*; (b) conversely, when the data is widely *scattered*, the set $\mathcal{MAO}$ contains no sample *but* the sample median. Thus, the filter reduces to a classical *median filter* ; (c) for the *mixed* case, we observe that the set $\mathcal{MAO}$ contains the sample median and the order-statistics in its proximity. Its estimate is a *linear combination* of the *central order statistics* weighted according to their *natural order* and thus clearly exhibits *hybrid* character. Figure 2 gives an overview of the filter behavior as a function of data dispersion.

Note, that as the observation window is shifted, the affinity function is in general relocated. Thus, the characteristics of the current output are determined by the dispersion of the data relative to the *re*located affinity function, and can, therefore, be different from the previous one. Because of this dynamic change of filter characteristics, the median affine filter is well suited to process signals with rapidly changing statistics.

Frequency-selectiveness: We know that for $\gamma \to \infty$, the median affine filter reduces to a normalized linear FIR filter whose frequency behavior is determined by the filter coefficients w_i. Similarly, for an appropriately chosen value of γ and well-behaved data the median affine filter will exhibit the same frequency behavior as the normalized FIR filter with identical coefficients. In impulsive noise, however, the outlying samples are rejected and the filter coefficients apply to the reliable samples only. This results in a *robust frequency-selective* behavior as illustrated in Section 5.

The median affine filter can be shown to be **translation invariant** and **linear-trend-preserving**. The last property, however, requires the filter to be *symmetric*, meaning that the filter coefficients satisfy $w_i = w_{N-(i-1)},\ i = 1, 2, \ldots, N$ and the affinity function satisfies $\mathcal{A}^{\mu,\gamma}(\mu - x) = \mathcal{A}^{\mu,\gamma}(\mu + x),\ \gamma \in [0, \infty), \mu \in (-\infty, \infty)$. The proofs of these properties are straightforward and are therefore omitted.

Edge preservation: The median affine filter preserves an *ideal* step edge only if $\gamma = 0$. The blurring effect, introduced for $\gamma > 0$, however, is minimal and can in general be neglected. For *noisy* edges the smoothing effect of the median affine filter *prior* and *after* the edge is superior to that of the median filter, as we show in the following.

A noisy, raising edge can be modeled as the sequence $\{x_1, x_2, \ldots, x_M, x_{M+1} + a, \ldots, x_N + a\}$, where a is any positive constant, and the x_i are distributed such that $min\{x_{M+1} + a, \ldots, x_N + a\} >> max\{x_1, x_2, \ldots, x_M\}$. For $M \geq (N+1)/2$ and an appropriately chosen value of γ, the set of $\mathcal{MAO}$'s consists of the observations which were taken prior to the edge (Fig. 3): $\mathcal{MAO} = \{x_1, x_2, \ldots, x_M\}$. Correspondingly, for $M < (N+1)/2$ we have $\mathcal{MAO} = \{x_{M+1} + a, x_{M+2} + a, \ldots, x_N + a\}$. Thus, the estimate prior to the edge is formed as $\hat{d} \approx \sum_{i=1}^{M} \mathcal{A}_i^\gamma w_i x_i / \sum_{i=1}^{M} \mathcal{A}_i^\gamma w_i$. A similar expression can be found for the samples following the edge. In either case, the output of the median affine filter is a *linear combination of all samples which are located on the same side of the edge as the sample median.* Thus, the order statistics are effectively exploited to attenuate the corrupting noise, while preserving the (desired) discontinuity. Figure 3 illustrates this property.

Impulse suppression: Like the median filter, the median affine filter suppresses pulses of width less than $(N+1)/2$ samples, where N is the sample size, and outputs a *linear combination of median affine observations.* This estimate is a linear combination of either $N-(l-k)$ low or $N-(l-k)$ high order statistics, all of which are weighted according to their temporal order. Intuitively, impulse suppression and smoothing should be superior over a filter structure which uses all of the order statistics, no matter if they are part of an undesired pulse or not (L-filter).

4. THE CENTER AFFINE FILTER

In filter applications where *feature enhancing* is needed, one is not necessarily interested in finding the most reliable sample, but those observations which clearly stand out from a noisy background. A suitable filter operation for problems of this nature can be achieved by utilizing the affinity function in the L-filter framework. By positioning the mode of the affinity function on the *center observation sample*, rather than the median, all observations are related to the central sample. This concept is realized in the *center affine filter* – a representative of the FIR affine filter class –, which is defined as a *linear combination of affinity weighted order statistics*:

Definition:

$$\hat{d} = \frac{\sum_{i=1}^{N} \mathcal{A}_{(i)}^\gamma x_{(i)} w_{(i)}}{\sum_{i=1}^{N} \mathcal{A}_{(i)}^\gamma w_{(i)}}, \tag{4}$$

where $x_{(i)}$ is the i^{th} largest sample, $\mathcal{A}_{(i)}^\gamma$ is its associated affinity measure, and the $w_{(i)}$'s are filter coefficients.

The heart of the center affine filter is the product of the i^{th} order statistic $x_{(i)}$ with its associated affinity measure. Whenever $x_{(i)}$ is close to the central sample, this product approximates $x_{(i)}$. If $x_{(i)}$ is distant from x_c, however, $\mathcal{A}_{(i)}^\gamma x_{(i)}$ tends to zero. Thus, the center affine filter forms its estimate based on those samples which are close to the central observation sample, provided that $0 < \gamma < \infty$. Like in the L-filter, the filter coefficients $w_{(i)}$ put additional weight on each sample according to its rank.

The center affine filter reduces to its basic structures for extreme values of γ: in particular we observe that for $\gamma \to 0$ the center affine filter reduces to an *identity operation*, whereas for $\gamma \to \infty$ it emulates an *L-filter.*

For a thorough analytical treatment of the center affine filter we refer the reader to [4] – its performance, however, is illustrated in the following section.

5. APPLICATIONS

Figure 4 illustrates the performance of a median affine bandpass filter operating on a chirp signal. Figure 4a and 4b show the clean chirp and the output of a linear FIR bandpass filter of order $N = 44$, respectively. To illustrate the power of the median affine structure, the chirp has been contaminated by alpha-stable noise ($\alpha = 1.5$), as shown in Fig. 4c (some impulses are truncated). The output of the linear FIR and the median affine bandpass filters with the same filter coefficients operating on the noisy chirp are displayed in Fig. 4d and 4e, respectively. Clearly, the impulses

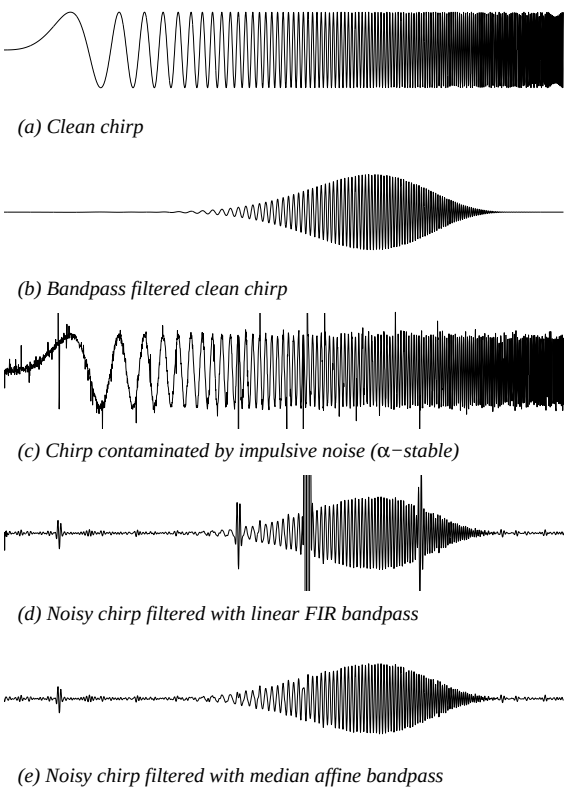

Figure 4. Performance of the median affine bandpass on a noisy chirp.

in the bandpass are fatal for the linear filter. The median affine filter exhibits a robust performance within the entire bandpass.

The feature enhancing abilities of the center affine filter are illustrated in Fig. 5. Figure 5a shows an inverse synthetic aperture radar (ISAR) image of a B727. A linear filter destroys the weak features of the plane, making it unusable for target recognition (Fig. 5b). The L-filter does a better job in preserving the tail and the body of the plane, but is not sensitive enough to keep the wings and tip (Fig. 5c). The center affine filter preserves the plane features excellently while smoothing the heavy background noise, as can be seen in Fig. 5d.

6. OPTIMIZATION

There are several ways to *design* the affine order-statistic filters. In many applications they have to obey certain frequency or rank-order characteristics putting constraints on the filter coefficients w_i, or $w_{(i)}$, respectively. Here, we will assume that these coefficients are known. Thus the design problem reduces to choosing a good value of γ. This choice, however, is crucial to an effective behavior of the filter.

The parameter γ can be designed such that the affine order statistic estimate minimizes a cost function $J(\gamma)$. Frequently $J(\gamma)$ is chosen to be the expected value of the squared error signal $e = d - \hat{d}$, i.e. $J(\gamma) = E\{(d-\hat{d})^2\}$. In this case the filter performs *optimal* in the *mean square error sense*. A closed form expression for the expected value does, in general, not exist. $J(\gamma)$, however, exhibits certain properties [4] which allow for an adaptive, gradient based approach as shown next:

Let γ_{opt} be the value of γ which minimizes $J(\gamma)$. To find γ_{opt} iteratively,

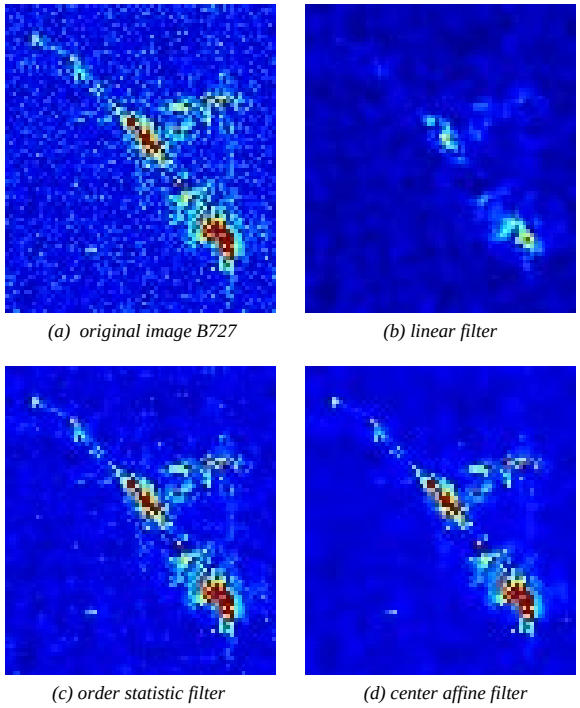

Figure 5. Feature enhancing using the center affine filter.

(1) start with an initial guess of γ, $\gamma(0)$. Using $\gamma(0)$, compute the gradient of the cost function

$$\frac{\partial J(0)}{\partial \gamma} = -2e(0)\frac{\partial \hat{d}(0)}{\partial \gamma(0)}. \tag{5}$$

(2) Update the previous estimate:

$$\gamma(n+1) = \gamma(n) - \frac{\mu}{2}\frac{\partial J(n)}{\partial \gamma(n)}, \tag{6}$$

where the stepsize $\mu > 0$ and $\partial J(n)/\partial\gamma(n)$ can be shown to be

$$\frac{\partial J(n)}{\partial \gamma(n)} = -2e(n)\frac{\sum_{i=1}^{N} w_i \frac{\partial \mathcal{A}_i^{\gamma}(n)}{\partial \gamma(n)}(x_i(n) - \hat{d}(n))}{\sum_{i=1}^{N} w_i \mathcal{A}_i^{\gamma}(n)}, \tag{7}$$

where $e(n) = d(n) - \hat{d}(n)$ is the present error.

This algorithm was applied to estimate γ for the ISAR image in Section 5: After 1000 iteration steps a γ yielding a result similar to the presented one was reached.

REFERENCES

[1] P. Ghandi and S. A. Kassam, "Design and performance of combination filters," *IEEE Transactions on Acoust. Speech, Signal Proc.*, vol. ASSP-37, July 1991.

[2] P. Heinonen and Y. Neuvo, "fir-median hybrid filters," *IEEE Trans. Acoust., Speech,Signal Process.*, vol. ASSP-35, June 1987.

[3] Y.-T. Kim and G. R. Arce, "Permutation filter lattices: A general nonlinear filtering framework," *IEEE Transactions on Signal Processing*, vol. 42, Sept. 1994.

[4] A. Flaig, G. R. Arce, and K. E. Barner, "Affine order statistic filters: a data-adaptive filtering framework for nonstationary signals," *submitted to IEEE Trans. Acoust., Speech,Signal Process.*, 1997.