A Generalization of Median Filtering Using Linear Combinations of Order Statistics

ALAN C. BOVIK, STUDENT MEMBER, IEEE, THOMAS S. HUANG, FELLOW, IEEE, AND DAVID C. MUNSON, JR., MEMBER, IEEE

***Abstract*—We consider a class of nonlinear filters whose output is given by a linear combination of the order statistics of the input sequence. Assuming a constant signal in white noise, the coefficients in the linear combination are chosen to minimize the output MSE for several noise distributions. It is shown that the optimal order statistic filter (OSF) tends toward the median filter as the noise becomes more impulsive. The optimal OSF is applied to an actual noisy image and is shown to perform well, combining properties of both the averaging and median filters. A more general design scheme for applications involving nonconstant signals is also given.**

I. Introduction

MEDIAN filtering has recently been recognized as an effective alternative to the linear smoother for some applications [1]-[4]. In particular, the moving median of a time or spatial series has been shown to preserve edges or monotonic changes in trend, while eliminating impulses of short duration. In these respects, the median smoother is superior to the linear filter.

The median is a particular case of the ith order statistic (or rank statistic) of a finite set of real numbers. The ith order statistic of N real numbers $x_1, \cdots, x_N$ where N is usually odd for digital filtering applications, is defined as the ith largest number in algebraic value. Here we shall denote the ith order statistic by $x_{(i)}$ in keeping with the mathematical literature. The minimum is then $x_{(1)}$, the maximum $x_{(N)}$, and the median $x_{((N+1)/2)}$.

Little work has been done in digital filtering applications using order statistics other than the median. It is the object of this paper to present an order statistic filter design scheme, where the output of the filter is a linear combination of the order statistics of several input samples considered simultaneously. The order statistic filter (OSF) is nonlinear due to the ordering process, which considerably complicates the analysis.

Before introducing the OSF, a short review of previous work in median filtering follows.

Manuscript received November 22, 1982; revised April 27, 1983. This work was supported by the Joint Services Electronics Program under Contract N00014-79-C-0424. Portions of this manuscript were presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Paris, France, May 3-5, 1982.

The authors are with the Coordinated Science Laboratory and the Department of Electrical Engineering, University of Illinois, Urbana, IL 61801.

A. Median Filtering

The median filter was introduced in 1974 by Tukey [1], who used the moving median as a smoothing technique in time series analysis. Rabiner *et al.* [2] used the median filter and series combinations of linear filters and median filters to smooth speech waveforms in a qualitative study, and reported favorable results. They found that the median filter was generally superior to a Hamming window of similar length for smoothing several waveforms such as log input energy of a speech signal, zero-crossing rate, and pitch period. The median smoother was noted to preserve discontinuities of sufficient duration while eliminating local roughness in the signal, whereas the linear smoother was seen to be inadequate in that much information was lost due to smearing. They deemed the median-linear series combination to be yet more effective for their application, but there is no evidence to support this for the general application.

Jayant [3] performed a similar study involving the supression of impulse noise due to bit errors in the transmission of digital speech signals. He compared moving average and moving median-based filters in computer simulations and informal listening tests for various filter lengths. He concluded that for independently occurring errors the two techniques performed similarly. However, he noted that for dependent error occurrences, which were interpreted as clusterings of the errors, the averager was generally superior.

Median filtering has also come into recent use for enhancing images. Pratt [4] made a rather qualitative study of two-dimensional median filters of various sizes and shapes, and examined application of one-dimensional medians to a picture corrupted by impulsive noise. He concluded that although the median filter is extremely useful for suppressing impulsive and "salt-and-pepper" noise, it should be considered an ad hoc method dependent upon the particular application.

There has been some more recent work in actual implementation of median filtering. Huang *et al.* [5] have devised a fast algorithm to implement two-dimensional median filters. Their algorithm is based upon the fact that as the filter window is moved from column to column across the image, most of the pixels are retained within the filter window. The algorithm updates the histogram corresponding to the new pixel values entering the window. A considerable reduction in computer time is attainable when compared with conventional sorting algorithms. In particular, for an $n \times n$ filter window

0096-3518/83/1200-1342$01.00

the number of comparisons required for computation of each pixel is about $(2n + 10)$; using ordinary sorting methods the number of comparisons is significantly higher.

An algorithm for real-time median filtering has been developed by Ataman, Aatre, and Wong [6].[1] This algorithm allows on-line computation of the running median.

It is possible that some of these methods, or variations, could be used for general order statistic filtering; however, that is beyond the scope of this paper, and will not be discussed here.

B. The Order Statistic Filter

The output of an order statistic filter of length N operating on a sequence $\{x_j\}$ for N odd is given by

$$y_k = \mathrm{OSF}(\{x_j\}_{j=k-M}^{k+M}) = \sum_{i=1}^{N} \alpha_i x_{(i)}^k \tag{1}$$

where $M = (N - 1)/2$ and $x_{(i)}^k$ are the order statistics of $x_{k-M}, \cdots, x_k, \cdots, x_{k+M}$. The α_i are constants that may be chosen for a particular application. We can generalize this to two dimensions by considering the points within the window as the values to be ranked and linearly combined, regardless of the shape and size of the window. Generally the window is symmetric about its center and we replace the center pixel value with the output value.

The median filter is a particular case of (1), with coefficients

$$\alpha_i = \begin{cases} 1; & i = (N+1)/2 \\ 0; & \text{otherwise.} \end{cases}$$

We can also define a maximum filter, for example, by taking

$$\alpha_i = \begin{cases} 0; & i = 1, \cdots, N-1 \\ 1; & i = N. \end{cases}$$

As a generalization, we may constrain all of the coefficients of the order statistics to be zero except for the ith, which we set to unity. The result is called an ith ranked-order operation. Nodes and Gallagher [7] have found that an $(N - 1)$th ranked-order operation is effective for digital AM detection both with and without corruption by impulse noise. Similarly, an Nth ranked-order operation (the maximum filter) can be used for peak detection.

In this paper we consider the design of a general OSF. In Section II we assume a constant signal in additive white noise and derive an explicit expression for the coefficients in the OSF minimizing the output MSE. The optimal coefficients are given for different filter lengths for several common noise distributions.

In Section III we compute the optimal OSF coefficients for a number of generalized exponential noise distributions. It is shown that as the noise becomes more impulsive (i.e., the tails of the noise density become heavier), the optimal OSF tends toward the median filter.

In Section IV, the optimal OSF for a constant signal is applied to an actual noisy image. The OSF performs well and is found to share properties of both the averaging and median filters.

Finally, Section V provides a more general design scheme for applications involving nonconstant signals.

II. Optimal OSF Coefficients for a Constant Signal in White Noise

We consider a constant signal s corrupted by zero-mean additive white noise; thus, the OSF input samples are of the form

$$x_j = s + n_j$$

where the n_j are independent, identically distributed random variables satisfying $E\{n_j\} = 0$. In addition, it will be assumed that the noise distribution is symmetric.

In the following, we will use the MSE as an optimality criterion, but we will also insist that the order statistic estimator be unbiased, i.e.,[2]

$$s = E\{y_k\} = E\left\{\sum_{i=1}^{N} \alpha_i x_{(i)}\right\} = s \sum_{i=1}^{N} \alpha_i + \sum_{i=1}^{N} \alpha_i E\{n_{(i)}\}. \tag{2}$$

The noise distribution is symmetric so that the optimal α_i are symmetric and

$$E\{n_{(i)}\} = -E\{n_{(N-i+1)}\}.$$

Therefore, the unbiasedness condition (2) reduces to

$$\sum_{i=1}^{N} \alpha_i = 1. \tag{3}$$

A. Minimization of the Mean Squared Error

The MSE is given by

$$\mathrm{MSE} = E\{(y_k - s)^2\} = E\left\{\left(\sum_{i=1}^{N} \alpha_i x_{(i)} - s\right)^2\right\}$$

which can be simplified to

$$\mathrm{MSE} = E\left\{\left(\sum_{i=1}^{N} \alpha_i n_{(i)}\right)^2\right\} \tag{4}$$

by substituting $x_{(i)} = s + n_{(i)}$ and using (3). Expanding (4) gives the form

$$\mathrm{MSE} = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j H_{ij} \tag{5}$$

where

$$H_{ij} = E\{n_{(i)} n_{(j)}\}. \tag{6}$$

Equation (5) is a quadratic form that can be expressed as

$$\mathrm{MSE} = \boldsymbol{\alpha}^T H \boldsymbol{\alpha} \tag{7}$$

where H is the $N \times N$ correlation matrix of the random vector $(n_{(1)}, n_{(2)}, \cdots, n_{(N)})^T$ and $\boldsymbol{\alpha}$ is the constant vector $(\alpha_1, \alpha_2, \cdots, \alpha_N)^T$.

[1] Also, see [6] for a more complete list of references on median filtering.

[2] To simplify notation we will write $x_{(i)}$ instead of $x_{(i)}^k$, and $n_{(i)}$ instead of $n_{(i)}^k$.

The minimization of (7) subject to (3) is a straightforward quadratic optimization problem that can be solved using Lagrange multipliers. The Lagrangian function is given by (where $\boldsymbol{e}$ denotes a column of ones)

$$F(\boldsymbol{\alpha}, \lambda) = \boldsymbol{\alpha}^T H \boldsymbol{\alpha} + \lambda(1 - \boldsymbol{e}^T \boldsymbol{\alpha}).$$

Setting the derivative with respect to $\boldsymbol{\alpha}$ equal to zero yields

$$2H\boldsymbol{\alpha} - \lambda \boldsymbol{e} = 0. \tag{8}$$

The correlation matrix H is generally positive definite (it must be at least nonnegative definite). In this case (7) is strictly convex and a unique solution can be obtained by multiplying (8) by $\boldsymbol{e}^T H^{-1}$ and using (3) in the form $\boldsymbol{e}^T\boldsymbol{\alpha} = 1$ to give

$$\lambda = 2/[\boldsymbol{e}^T H^{-1} \boldsymbol{e}].$$

Substituting into (8) yields the optimal coefficients

$$\boldsymbol{\alpha} = H^{-1}\boldsymbol{e}/[\boldsymbol{e}^T H^{-1} \boldsymbol{e}]. \tag{9}$$

A more general expression for $\boldsymbol{\alpha}$ has been derived by Lloyd [8], using the Gauss-Markov least-squares theorem. This expression does not require that the parent distribution be symmetric as we assumed in deriving (9), but does require computation of the expected values of the ordered noise variates.

In order to compute the optimal coefficients using (9), it is first necessary to compute the elements of H, given by (6). We now consider this problem.

B. Computation of the Correlation Matrix

Evaluation of the H_{ij} in (6) requires expressions for the marginal and bivariate densities of the $n_{(i)}$. For this, we refer to the monograph on order statistics by David [9]. Denoting the parent distribution and density of the noise as $F_n(\cdot)$ and $f_n(\cdot)$, respectively, the density of $n_{(i)}$ for $i = 1, \cdots, N$ is given by

$$g_{n_{(i)}}(x) = K_i F_n^{i-1}(x)\,[1 - F_n(x)]^{N-i} f_n(x) \tag{10}$$

where $K_i = N!/[(i-1)!(N-i)!]$.

The joint density of $n_{(i)}$ and $n_{(j)}$ for $i, j = 1, \cdots, N\ (i<j)$ is

$$g_{n_{(i)}n_{(j)}}(x, y) = K_{i,j} F_n^{i-1}(x)\,[F_n(y) - F_n(x)]^{j-i-1} \cdot [1 - F_n(y)]^{N-j} f_n(x) f_n(y) \tag{11}$$

where $K_{i,j} = [N!/[(i-1)!(j-i-1)!(N-j)!]$.

Using the above notation, the H_{ij} are given by

$$H_{ij} = \iint_{-\infty}^{\infty} xy\, g_{n_{(i)}n_{(j)}}(x, y)\, dx\, dy \qquad (i<j) \tag{12}$$

$$H_{ii} = \int_{-\infty}^{\infty} x^2 g_{n_{(i)}}(x)\, dx. \tag{13}$$

Due to the complexity of (10) and (11), numerical integration is generally required for evaluation of the H_{ij}, even for simple parent distributions of the original noise. The number of required integrations can be approximately quartered, however, by making use of the following symmetry relations:

a) $H_{ij} = H_{ji}$

b) $H_{ij} = H_{N-i+1, N-j+1}$

where b) assumes a symmetric parent density f_n.

Equations (12) and (13) were used to compute H for six different noise distributions and all odd values of N ranging from 3 to 25. The results are quite voluminous and hence are tabulated in [10], rather than here. The noise distributions chosen were the U-shaped, uniform, parabolic, triangular, normal, and Laplacian. These have been considered by Sarhan [11]-[13] for some small values of N (both odd and even) in estimating population means and variances in the mid-1950's. The U-shaped and parabolic densities are given by

$$f_U(x) = \sqrt{27/125}\,(3x^2/2\sigma^2); \qquad |x| < \sqrt{5/3}\,\sigma$$

and

$$f_P(x) = 3(\sqrt{5}\,\sigma + x)(\sqrt{5}\,\sigma - x)/20\sqrt{5}\,\sigma^3; \qquad |x| < \sqrt{5}\,\sigma$$

where σ is the standard deviation. We have taken $\sigma = 1$ for all distributions. This is without loss of generality, since Sarhan has shown that the optimal OSF coefficients are independent of σ.

C. Optimal Coefficients and Comparison of MSE

The resulting optimal values of the coefficients $\{\alpha_i\}_{i=1}^N$ were computed for each of the six noise distributions and for all odd values of N ranging from 3 to 25, and are available in [10]. Tables I-III list the optimal coefficients for $N = 3$, 9, and 25.

The results obtained for the uniform and normal distributions are expected, since the midpoint ($\alpha_1 = \alpha_N = 1/2$) and the average ($\alpha_i = 1/N$; $i = 1, \cdots, N$) are the respective maximum likelihood estimators for these distributions. The resulting α_i for the Laplacian case are not surprising either, as we see that most of the weight is located in the center α_i. In fact, we note that the weight in the central coefficient becomes more pronounced as the noise distribution grows heavier tailed.

Table IV compares the MSE of the optimal OSF with that of both the median filter and the averaging filter for $N = 3$. It can be shown that the averaging filter is the best linear unbiased estimator (BLUE) for the problem under consideration. Note that the optimal OSF performs at least as well as either the median or BLUE for any noise distribution. This occurs because both the median and BLUE are OSF filters.

We next consider a family of generalized exponential noise distributions to further explore the relationship between the tail behavior of the noise and the properties of the OSF.

III. A Family of Generalized Exponential Noise Distributions

Consider the noise density

$$f_n(x) = ke^{-\gamma |x|^\beta}; \qquad |x| < \infty \tag{14}$$

where γ and β are positive, and where k is chosen such that

$$\int_{-\infty}^{+\infty} f_n(x)\, dx = 1.$$

Integration yields

$$k = (\beta\gamma^{1/\beta})/2\Gamma(1/\beta)$$

TABLE I
OPTIMAL OSF COEFFICIENTS FOR $N = 3$

DISTRIBUTION	α_1	α_2
U-shaped	0.54000	-0.08000
Uniform	0.50000	0.00000
Parabolic	0.44048	0.11905
Triangular	0.39456	0.21088
Normal	0.33333	0.33333
Laplacian	0.15168	0.69663
	α_3	α_2

TABLE II
OPTIMAL OSF COEFFICIENTS FOR $N = 9$

DISTRIBUTION	α_1	α_2	α_3	α_4	α_5
U-shaped	0.55478	-0.02627	-0.01460	-0.00988	-0.00806
Uniform	0.50000	0.00000	0.00000	0.00000	0.00000
Parabolic	0.34031	0.06134	0.04428	0.03623	0.03568
Triangular	0.25913	0.04914	0.05734	0.08392	0.10094
Normal	0.11111	0.11111	0.11111	0.11111	0.11111
Laplacian	-0.01899	0.02904	0.06965	0.23795	0.36469
	α_9	α_8	α_7	α_6	α_5

with

$$\gamma = [\Gamma(3/\beta)/\Gamma(1/\beta)]^{\beta/2}\sigma^{-\beta}$$

where Γ is the ordinary gamma function and σ, the standard deviation, will be taken to be 1. Obviously, if β is small, the above density will have heavy tails and the resulting noise will be impulsive in nature. Conversely, if β is large, the tails will be shallow and the noise will be comparatively smooth.

The techniques used in Section II can be applied to obtain the optimal coefficients for various β. As β decreases, the time required for computation of the correlation matrix, H, increases dramatically since very small intervals are required in the numerical integration routine. This limitation is not particularly disturbing, however, since the optimal OSF will be seen to approach the median for such extremely impulsive noise.

The Laplacian case with $\beta = 1$ and the normal case with $\beta = 2$ have already been considered. In this section we also consider $\beta = 1/2$, $3/4$, $3/2$, and 3. The correlation matrices were computed for $N = 3$ and are available in [10]. Filter lengths larger than $N = 3$ would have entailed prohibitive computational time for the cases with $\beta < 1$. The resulting optimal OSF coefficients were computed from (9) and are listed in Table V.

Note that as β grows smaller, there is a dramatic tendency towards the median; the value of α_2 approaches one very rapidly, and the value of $\alpha_1 = \alpha_3$ approaches zero. This strengthens the notion that median and almost-median type filters are very effective for suppressing additive noise which is more impulsive than the Laplacian case, at least for a nearly constant background signal.

TABLE III
OPTIMAL OSF COEFFICIENTS FOR $N = 25$

DISTRIBUTION	α_1	α_2	α_3	α_4	α_5
U-shaped	0.49732	0.00177	0.00320	0.00387	0.00394
Uniform	0.50000	0.00000	0.00000	0.00000	0.00000
Parabolic	0.28292	0.04758	0.03169	0.02393	0.01936
Triangular	0.21685	0.03614	0.02411	0.01811	0.01457
Normal	0.04000	0.04000	0.04000	0.04000	0.04000
Laplacian	0.00550	0.00335	-0.00427	-0.00101	-0.00008
	α_{25}	α_{24}	α_{23}	α_{22}	α_{21}

DISTRIBUTION	α_6	α_7	α_8	α_9	α_{10}
U-shaped	0.00081	-0.00122	-0.00197	-0.00206	-0.00180
Uniform	0.00000	0.00000	0.00000	0.00000	0.00000
Parabolic	0.01641	0.01447	0.01316	0.01204	0.01147
Triangular	0.01248	0.01190	0.01341	0.01786	0.02610
Normal	0.04000	0.04000	0.04000	0.04000	0.04000
Laplacian	0.00065	0.00314	0.01064	0.02907	0.06499
	α_{20}	α_{19}	α_{18}	α_{17}	α_{16}

DISTRIBUTION	α_{11}	α_{12}	α_{13}
U-shaped	-0.00162	-0.00150	-0.00146
Uniform	0.00000	0.00000	0.00000
Parabolic	0.01096	0.01068	0.01066
Triangular	0.03684	0.04643	0.05039
Normal	0.04000	0.04000	0.04000
Laplacian	0.11835	0.17195	0.19541
	α_{15}	α_{14}	α_{13}

TABLE IV
COMPARISON BETWEEN MSE OF THE OPTIMAL OSF FILTER, THE MEDIAN FILTER, AND THE BLUE

DISTRIBUTION	MSE(OPTIMAL) / MSE(MEDIAN)	MSE(OPTIMAL) / MSE(BLUE)
U-shaped	0.306	0.749
Uniform	0.500	0.900
Parabolic	0.603	0.963
Triangular	0.667	0.988
Normal	0.742	1.000
Laplacian	0.900	0.862

Fig. 1 shows a plot of MSE(optimal)/MSE(median) and MSE(optimal)/MSE(BLUE) versus β for the above values of β. It is clear that the optimal OSF performs significantly better than the BLUE for $\beta < 1$, corresponding to impulse noise, and that the OSF performs better than the median filter for $\beta > 1$, corresponding to shallow-tailed noise.

In a related study, Gastwirth and Cohen [14] considered the behavior of several robust estimators, most of which are OSF's. In particular, they found the optimal coefficients $\boldsymbol{\alpha}$ given by (9) for the contaminated normal densities $CN(\epsilon, D)$ given by

TABLE V
OPTIMAL OSF COEFFICIENTS FOR GENERALIZED EXPONENTIAL NOISE AND $N = 3$

β	α_1	α_2
3.0	0.4054	0.1891
2.0	0.3333	0.3333
1.5	0.2626	0.4748
1.0	0.1517	0.6966
0.75	0.0773	0.8453
0.5	0.0052	0.9897
	α_3	α_2

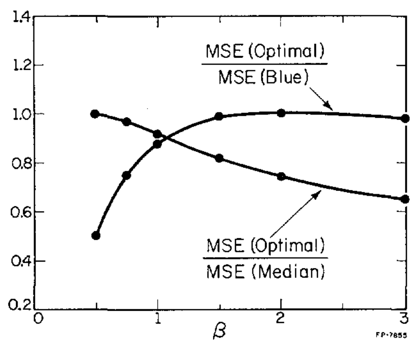


Fig. 1. Comparison between the MSE of the optimal OSF filter, the median filter, and the BLUE for generalized exponential noise.

$$f_{\epsilon,D}(x) = \frac{1}{\sqrt{2\pi}} \left\{ (1-\epsilon) e^{-x^2/2} + \frac{\epsilon}{D} e^{-x^2/2D^2} \right\}$$

for $\epsilon = 0.01$, 0.05, 0.10 and $D = 3$. This corresponds to an assumed standard normal noise contaminated by a fraction ϵ of $\eta(0, 3)$ noise. They found that for reasonably small samples the optimal $\boldsymbol{\alpha}$ were very similar to the coefficients associated with the δ-trimmed mean. The δ-trimmed mean corresponds to an OSF with coefficients

$$\alpha_i = \begin{cases} 1/(N - 2[\delta N]); & [\delta N] + 1 \leqslant i \leqslant N - [\delta N] \\ 0; & \text{otherwise} \end{cases}$$

where N is the sample size, δ is the fraction trimmed from each end of the ordered sample, and $[\mu]$ denotes the integer part of μ.

The densities $CN(\epsilon, D)$ do not represent a class of increasingly heavy-tailed densities like those in (14), but rather represent a degree of contamination suitable for robustness studies. Gastwirth and Cohen found that in small samples ($N < 20$) both the OSF's for the contaminated normal densities and suitably trimmed means were very robust over densities ranging from the normal to the Laplacian, where robustness was defined in terms of the efficiency relative to the optimal OSF for each density. They concluded that for simplicity and reliability a moderately trimmed mean ($\delta \sim 0.20$) is very effective for small to moderate sample sizes.

IV. Application of a Designed Filter to a Noisy Image

In this section an optimal OSF for a constant signal is applied to an image corrupted by pseudorandom computer-generated noise. The image under consideration consists of 240 × 240 pixel values with eight bits of resolution per pixel. The original, uncorrupted image is shown in Fig. 2(a). Zero-mean Laplacian noise ($\sigma^2 = 100$) was added to the image, as shown in Fig. 2(b). The optimal 3 × 3 (9 point) Laplacian OSF, with coefficients given in Table II, was then applied with the result shown in Fig. 2(c).

A 3 × 3 averaging filter and a 3 × 3 median filter were also applied for comparison. The averaging filter was noted to produce a slightly blurrier image than either the OSF or the median filter, whereas the median filter produced sharp edges in the image but also introduced some blotches in areas of fairly constant value. The Laplacian OSF combined qualities of both the other filters, yielding sharp edges but with a slight smoothing effect reminiscent of the averager. The differences in the filtered versions were rather slight, however, and were deemed to be too small to show in reproduction; hence the filtered versions corresponding to the averager and the median are not shown.

In order to better preserve discernible differences after reproduction, Fig. 3(a)-(e) compares a single line of the original image with the filtered versions.[3] The averaging filter smoothed out some information-carrying features that are present in the original, particularly edge-type structures. The median filter preserved edges well, but flattened narrow peaks that were present in the original image. The Laplacian OSF yielded a compromise between the other two filters.

A useful quantitative comparison of the performances of the three filters is the empirical mean squared error given by

$$e = \frac{1}{K^2} \sum_{i=1}^{K} \sum_{j=1}^{K} (f_{ij} - o_{ij})^2$$

where K is the number of pixels per line (240), o_{ij} are the pixel values in the original image, and f_{ij} are the pixel values in the filtered image. Table VI lists the empirical MSE for the Laplacian OSF, the median filter, and the averaging filter. The Laplacian OSF performed better than the median filter and significantly better than the averaging filter.

V. Generalization for Nonconstant Signals

Most practical applications deal with unknown signals, but unfortunately, given a statistical description of such a signal, the design of an optimal OSF is generally intractable. There are cases though, such as in detection, where the signal may be nonconstant, but known. In this section we extend the results of Section II to the case of a nonconstant known signal.

Consider a known signal s_j corrupted by white noise n_j, resulting in the sequence

[3]Note that for a 3 × 3 window, the filtered versions also depend on the two lines adjacent to that shown in Fig. 3(b).

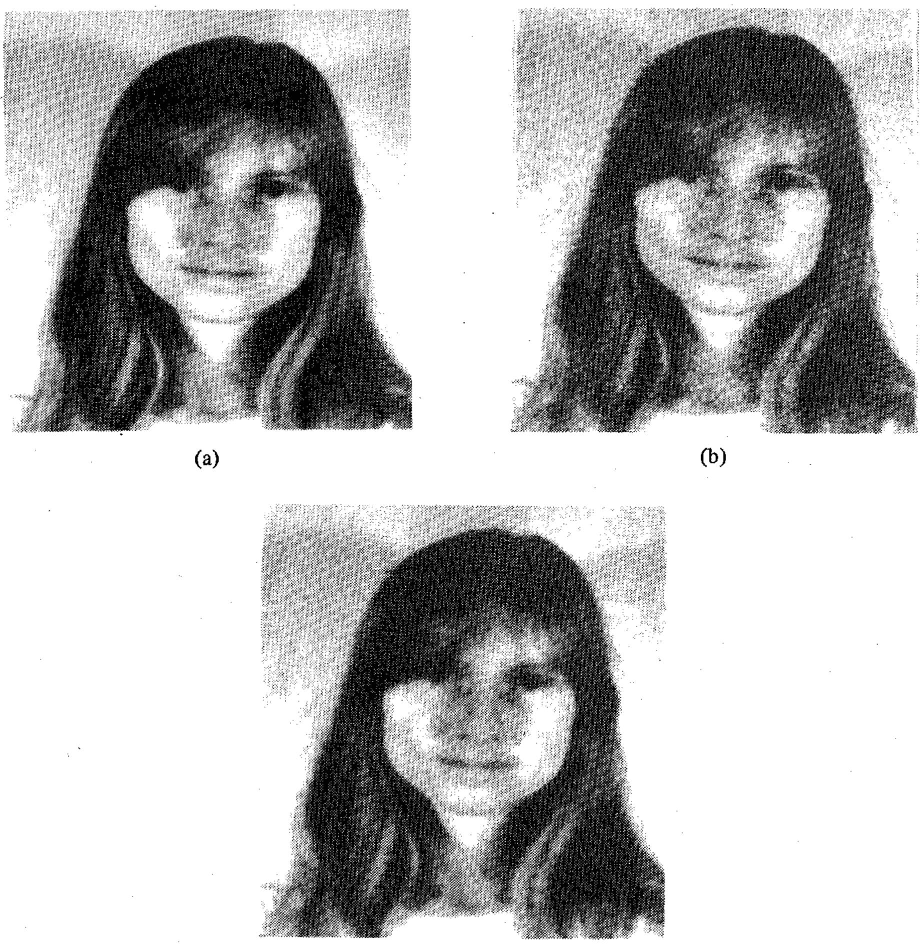


Fig. 2. (a) Original image. (b) Image corrupted by zero-mean Laplacian noise ($\sigma^2 = 100$). (c) Noisy image filtered by Laplacian OSF.

$$x_j = s_j + n_j.$$

In this case the optimal OSF is time or spatially varying with coefficients $\boldsymbol{\alpha}_k = (\alpha_1^k, \alpha_2^k, \cdots, \alpha_N^k)^T$. The filter output is

$$y_k = \sum_{i=1}^{N} \alpha_i^k x_{(i)}^k$$

so that the MSE for the kth output is

$$\mathrm{MSE}_k = E\{(y_k - s_k)^2\} = E\left\{\left(\sum_{i=1}^{N} \alpha_i^k x_{(i)}^k - s_k\right)^2\right\}.$$

As before, this can be rewritten using a quadratic form. Let H_k be the correlation matrix with elements $h_{ij}^k = E\{x_{(i)}^k x_{(j)}^k\}$ and let $\boldsymbol{\mu}_k$ be the mean vector with elements $\mu_i^k = E\{x_{(i)}^k\}$ for $i, j = 1, \cdots, N$. The MSE can then be expressed as

$$\mathrm{MSE}_k = \boldsymbol{\alpha}_k^T H_k \boldsymbol{\alpha}_k - 2 s_k \boldsymbol{\alpha}_k^T \boldsymbol{\mu}_k + s_k^2. \tag{15}$$

Again, we will insist that the estimate be unbiased, i.e.,

$$s_k = E\{y_k\} = \sum_{i=1}^{N} \alpha_i^k E\{x_{(i)}^k\} = \boldsymbol{\alpha}_k^T \boldsymbol{\mu}_k. \tag{16}$$

Substitution of the unbiasedness condition into (15) yields

$$\mathrm{MSE}_k = \boldsymbol{\alpha}_k^T H_k \boldsymbol{\alpha}_k - s_k^2. \tag{17}$$

The minimization of (17) subject to (16) can be solved as before using the Lagrange multiplier method. The Lagrangian function is

$$F(\boldsymbol{\alpha}_k, \lambda) = \boldsymbol{\alpha}_k^T H \boldsymbol{\alpha}_k - s_k^2 + \lambda(s_k - \boldsymbol{\alpha}_k^T \boldsymbol{\mu}_k).$$

Setting the derivative with respect to $\boldsymbol{\alpha}_k$ equal to zero yields

$$2 H_k \boldsymbol{\alpha}_k - \lambda \boldsymbol{\mu}_k = 0$$

which, combined with the unbiasedness constraint (16), gives the optimal coefficients

$$\boldsymbol{\alpha}_k = s_k H_k^{-1} \boldsymbol{\mu}_k / \boldsymbol{\mu}_k^T H_k^{-1} \boldsymbol{\mu}_k.$$

Notice that unlike the constant signal case, the optimal coefficients now depend on the signal.

Computation of $\boldsymbol{\mu}_k$ and H_k requires both the marginal and bivariate densities of the $x_{(i)}^k$. The densities given by (10) and (11) cannot be used, since the N independent observations $x_{k-M}, \cdots, x_k, \cdots, x_{k+M}$ are no longer identically distributed

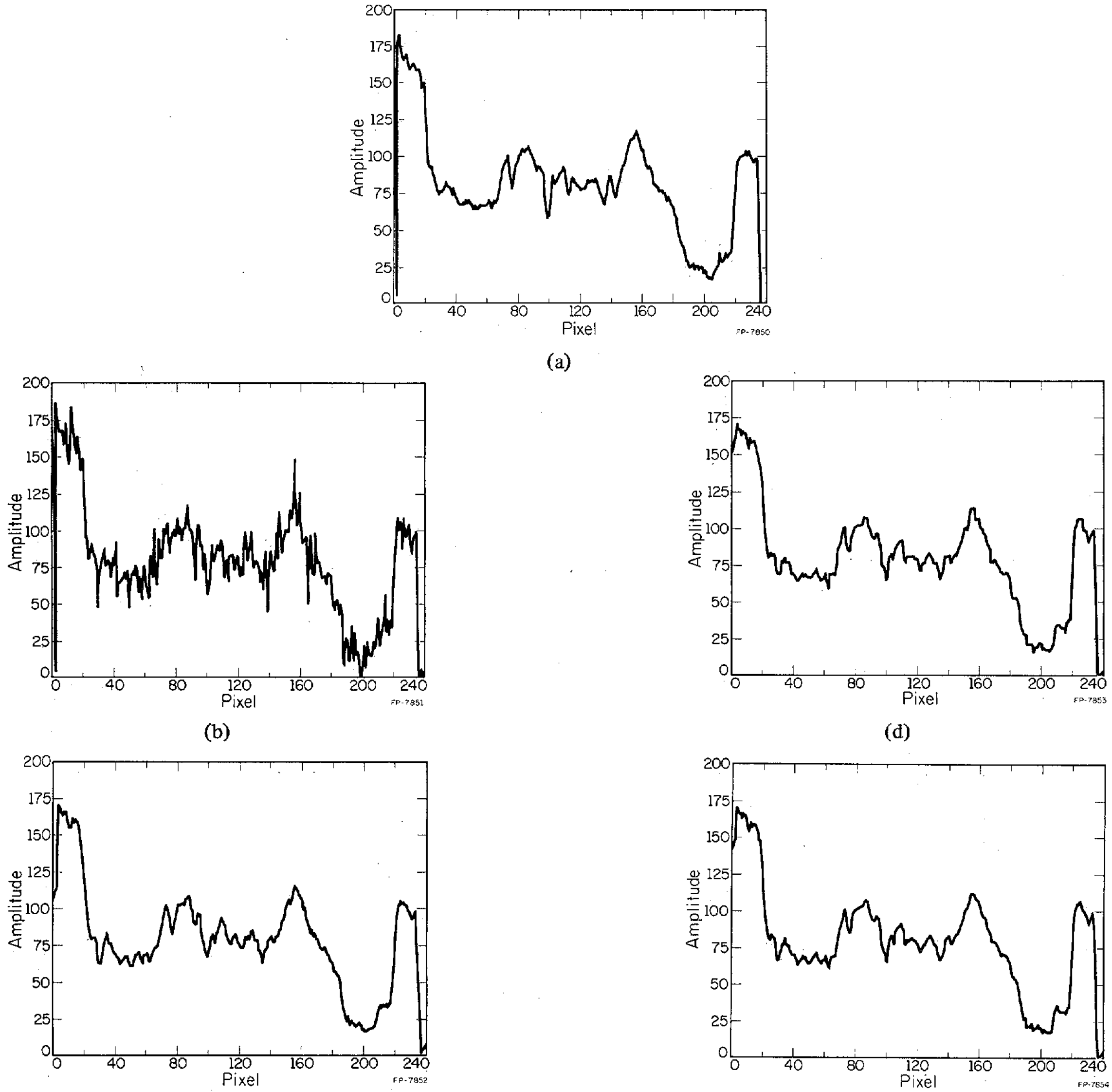


Fig. 3. (a) Horizontal line through mouth region in Fig. 2(a) (uncorrupted original). (b) Line from Fig. 3(a) with Laplacian noise ($\sigma^2 = 100$) added. (c) Noisy image filtered by averaging filter (horizontal line through mouth region). (d) Noisy image filtered by median filter (horizontal line through mouth region). (e) Noisy image filtered by Laplacian OSF (horizontal line through mouth region).

TABLE VI
COMPUTED EMPIRICAL MSE FOR FILTERED VERSIONS OF IMAGE CORRUPTED BY LAPLACIAN NOISE
($\sigma^2 = 100$)

Filter	MSE
Laplacian OSF	60.4
Median Filter	66.0
Averaging Filter	74.1

(the means are different). Denoting the densities of the unordered observations by $f_j^k(x)$ and the distributions by $F_j^k(x)$ for $j = 1, \cdots, N$, the marginal density of $x_{(i)}^k$ for $i = 1, \cdots, N$ is given by

$$g_i^k(x) = L_i\, {}^+|A_i^k(x)|^+$$

where

$$L_i = \frac{1}{(N-i)!(i-1)!}$$

and

$$A_i^k(x) = \begin{bmatrix} F_1^k(x) & \cdots & F_N^k(x) \\ \vdots & & \vdots \\ F_1^k(x) & \cdots & F_N^k(x) \\ f_1^k(x) & \cdots & f_N^k(x) \\ 1 - F_1^k(x) & \cdots & 1 - F_N^k(x) \\ \vdots & & \vdots \\ 1 - F_1^k(x) & \cdots & 1 - F_N^k(x) \end{bmatrix} \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} i - 1 \text{ rows} \\ \left.\right\} 1 \text{ row} \\ \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} N - i \text{ rows} \end{array}$$

where ${}^+|\cdot|^+$ denotes the *permanent* of the matrix, which is simply the determinant with all of the signs in the expansion positive rather than alternating [15].

Similarly, the joint density of $x_{(i)}^k$ and $x_{(j)}^k$ where $i<j$ is given by

$$g_{ij}^k(x,y) = L_{ij}{}^{+}\left|A_{ij}^k(x,y)\right|^{+}$$

where

$$L_{ij} = \frac{1}{(i-1)!(j-i-1)!(N-j)!}$$

and

$$A_{ij}^k(x,y) =$$

$$\left[\begin{array}{ccc}
F_1^k(x) & \cdots & F_N^k(x) \\
\vdots & & \vdots \\
F_1^k(x) & \cdots & F_N^k(x) \\
f_1^k(x) & \cdots & f_N^k(x) \\
F_1^k(y)-F_1^k(x) & \cdots & F_N^k(y)-F_N^k(x) \\
\vdots & & \vdots \\
F_1^k(y)-F_1^k(x) & \cdots & F_N^k(y)-F_N^k(x) \\
f_1^k(y) & \cdots & f_N^k(y) \\
1-F_1^k(y) & \cdots & 1-F_N^k(y) \\
\vdots & & \vdots \\
1-F_1^k(y) & \cdots & 1-F_N^k(y)
\end{array}\right]
\begin{array}{l}
\left.\begin{array}{l} \\ \\ \\ \end{array}\right\} i-1 \text{ rows} \\
\left.\begin{array}{l} \\ \end{array}\right\} 1 \text{ row} \\
\left.\begin{array}{l} \\ \\ \\ \end{array}\right\} j-i-1 \text{ rows} \\
\left.\begin{array}{l} \\ \end{array}\right\} 1 \text{ row} \\
\left.\begin{array}{l} \\ \\ \\ \end{array}\right\} N-j \text{ rows.}
\end{array}$$

It should be noted that these expressions are valid for the order statistics of any set of N independent nonidentically distributed random variables.

VI. Conclusion

We have studied a new class of nonlinear filters whose output is given by a linear combination of the order statistics of the input sequence. Both the median and averaging filters are members of this class. Assuming a constant signal in additive white noise, an explicit expression was derived for the optimal OSF coefficients. Both qualitative and quantitative comparisons suggested that OSF's (designed for a constant signal) can probably perform better than the median and averaging filters in some applications.

Computation of the optimal OSF coefficients requires the correlation matrices of the order statistics. A number of these matrices are available in a separate report [10] and may also be of use in other endeavors.

The dependence of the OSF coefficients on the impulsivity of the additive noise was investigated using a family of generalized exponential distributions. For noise more impulsive than the Laplacian case, the optimal OSF quickly tends toward the median.

A generalization of the OSF design scheme was given for known nonconstant signals. This could prove useful for applications such as signal detection.

Finally, more work needs to be done in this area, particularly on the development of performance measures for comparison of different nonlinear filtering schemes.

References

[1] J. W. Tukey, "Nonlinear (nonsuperposable) methods for smoothing data," in *Conf. Rec., 1974 EASCON*, p. 673.

[2] L. R. Rabiner, M. R. Sambur, and C. E. Schmidt, "Applications of a nonlinear smoothing algorithm to speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 552–557, Dec. 1975.

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

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

[5] T. S. Huang, G. J. Yang, and G. Y. Tang, "A fast two-dimensional median filtering algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 13–18, Jan. 1979.

[6] E. Ataman, V K. Aatre, and K. M. Wong, "A fast method for real-time median filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 415–420, Aug. 1980.

[7] T. M. Nodes and N. C. Gallagher, Jr., "Median filtering," in *Proc. 18th Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 8–10, 1980, pp. 926–934.

[8] E. H. Lloyd, "Least-squares estimation of location and scale parameters using order statistics," *Biometrika*, vol. 39, pp. 88–95, 1952.

[9] H. A. David, *Order Statistics*. New York: Wiley, 1981.

[10] A. C. Bovik, "Nonlinear filtering using linear combinations of order statistics," Coordinated Sci. Lab., Univ. Illinois, Urbana, Rep. R-935, 1982.

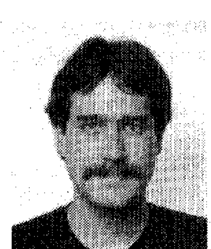
[11] A. E. Sarhan, "Estimation of the mean and standard deviation by order statistics," *Ann. Math. Statist.*, vol. 25, pp. 317–328, 1954.

[12] —, "Estimation of the mean and standard deviation by order statistics, part II," *Ann. Math. Statist.*, vol. 26, pp. 505–511, 1955.

[13] —, "Estimation of the mean and standard deviation by order statistics, part III," *Ann. Math. Statist.*, vol. 26, pp. 576–592, 1955.

[14] J. L. Gastwirth and M. L. Cohen, "Small sample behavior of some robust linear estimators of location," *J. Amer. Statist. Ass.*, vol. 65, pp. 946–973, 1970.

[15] R. J. Vaughan and W. N. Venables, "Permanent expressions for order statistic densities," *J. Roy. Statist. Soc.*, vol. 34, pp. 308–310, 1972.

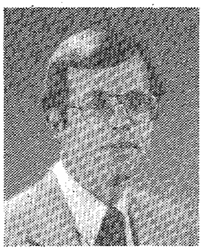


Alan C. Bovik (S'80) was born in Kirkwood, MO, on June 25, 1958. He received the B.S. degree in computer engineering and the M.S. degree in electrical engineering from the University of Illinois, Urbana-Champaign, in 1980 and 1982, respectively. He is currently completing the Ph.D. degree in electrical engineering at the same university.

He has served as both Research Assistant in the Coordinated Science Laboratory and as Teaching Assistant in the Department of Electrical Engineering at the University of Illinois. His current research interests are in the areas of statistical signal and image processing, computer vision, and computational aspects of human vision.

Mr. Bovik is a member of Phi Kappa Phi.

Thomas S. Huang (S'61–M'63–SM'76–F'79), for a photograph and biography, see p. 649 of the June 1983 issue of this Transactions.



David C. Munson, Jr. (S'75-M'79) was born in Red Oak, IA, on October 19, 1952. He received the B.S. degree in electrical engineering from the University of Delaware, Newark, in 1975, and the M.S., M.A., and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 1977, 1977, and 1979, respectively.

Since August 1979, he has been with the University of Illinois, Urbana-Champaign, IL, where he is currently an Associate Professor in the Department of Electrical Engineering and an Associate Research Professor in the Coordinated Science Laboratory.

His research interests include the effects of finite register length in digital signal processing, signal processing architectures, multidimensional and time-varying signal processing, and image reconstruction in sensor array processing systems such as synthetic aperture radar.

Dr. Munson is a member of Eta Kappa Nu and Tau Beta Pi.

Two-Dimensional Root Structures and Convergence Properties of the Separable Median Filter

THOMAS A. NODES, MEMBER, IEEE, AND NEAL C. GALLAGHER, JR., MEMBER, IEEE

Abstract—**The root (signals invariant to filtering) structures of the two-dimensional separable median filter are derived and presented. In addition, it is proved that with rare exception, after repetitive passes of the separable median filter any two-dimensional signal will be reduced to a signal containing only root structures.**

I. Introduction

THE median filter is an easily implemented operation in which a windowing function is passed over a signal, and at each position, the output of the filter is taken to be the value of the mid or median valued point out of all the points inside the window. Thus, the output value y_i of a one-dimensional median filter of window width $2 \cdot N + 1$ is

$$y_i \equiv \text{the median value of } \{X_{i-N}, X_{i-N+1}, \cdots, X_{i+N}\}$$

where $\{X_j\}$ is the input signal. There are a number of ways to extend the median operation to two dimensions. One way would be to pass a two-dimensional window, such as a square or cross-shaped window, over the two-dimensional signal and again take as the output at each position the median valued point of all the points inside the window. An alternative to using a two-dimensional window is the separable median filter. As presented by Narendra [1], the separable median filter consists of a two-pass operation in which each row of a two-dimensional signal is first filtered by a horizontally oriented one-dimensional median filter; then, each column of the resulting signal is filtered by a vertically oriented median filter. Specifically, the output value $y_{i,j}$ of a separable median filter is

$$y_{i,j} = \text{the median value of } \{Z_{i-N,j}, Z_{i-N+1,j}, \cdots, Z_{i+N,j}\}$$

where

$$Z_{pq} = \text{the median value of } \{X_{p,q-N}, X_{p,q-N+1}, \cdots, X_{p,q+N}\}$$

where $\{X_{i,j}\}$ is the two-dimensional input signal. This operation requires the ranking of fewer points than does an equivalently sized two-dimensional windowed filter, and thus tends to be faster. Both of these types of two-dimensional median filters have been used with success in digital image processing where they are known for their speed and their ability to track signal edges and reduce impulsive type noise [1]-[5]. Unfortunately, even though these operations are easy to implement, they are nonlinear with memory and thus very difficult to analyze. For this reason, most of the past designs utilizing median operations have been ad hoc in nature. However, some progress has recently been made in developing both deterministic [5]-[8] and nondeterministic [9]-[12] properties of

Manuscript received September 27, 1982. This work was supported inpart by the Air Force Office of Scientific Research under Grant AFOSR 78-3605.

T. A. Nodes is with TRW, Military Electronics Division, San Diego, CA 92101.

N. C. Gallagher, Jr. is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907.

0096-3518/83/1200-1350$01.00 © 1983 IEEE