Design of Weighted Order Statistic Filters Using the Perceptron Algorithm

by
Byeongjang Jeong and Yong Hoon Lee

Dept. of Electrical Engineering
Korea Advanced Institute of Science and Technology
373-1 Kusong-dong Yusong-gu Taejon 305-701 Korea
phone: +82-42-869-3437, fax: +82-42-869-3410
e-mail: yohlee@ee.kaist.ac.kr

Abstract

In this paper, we observe that the design of optimal weighted order statistic(WOS) filters under the mean absolute error criterion can be thought of as a two-class linear classification problem. Based on this observation, the perceptron algorithm is applied to design WOS filters. It is shown, through experiments, that the perceptron algorithm can find optimal or near optimal WOS filters in practical situations.

1. Introduction

The *weighted order statistic*(WOS) filter is a nonlinear digital filter[1]. It has a window moving over an input sequence, and its output is calculated by duplicating each input sample $X_i, 1 \leq i \leq b$, within a window to the number of the corresponding weight W_i; sorting the resulting array of $\sum_{i=1}^{b} W_i$ points and then choosing the T-th largest value from the sorted data, where T is a threshold value. If $\sum_{i=1}^{b} W_i$ is odd and $T = (\sum_{i=1}^{b} W_i + 1)/2$, then the WOS filter becomes a *weighted median*(WM) filter[1]-[3]. While the class of WOS filters encompasses WM, median and rank order filters[4], it is a subclass of stack filters[5].

In [6] and [7], it is shown that an optimal stack filter can be designed under the mean absolute error(MAE) criterion by using linear programming(LP). Although the WOS filter is a special case of stack filters, it cannot be optimized through LP. In [8], WOS filters are designed by using the method of steepest descent after approximating the MAE criterion. The design method is simple to implement, but cannot produce an optimal WOS filter minimizing the MAE.

In this paper, we observe that the problem of designing optimal WOS filters under the MAE criterion can be thought of as a two-class linear classification problem[9]. Based on this observation, the *perceptron* algorithm is applied to design WOS filters. The proposed method can find optimal or nearly optimal WOS filters in practical situations, and it is simple to implement.

The organization of this paper is as follows. In Section 2, we shall briefly review the definitions and properties of stack and WOS filters and introduce our notation. The proposed design method for WOS filters is described in Section 3. Finally, in Section 4 WOS filters are designed using the proposed method and applied to enhance noisy images.

2. Review on Stack and WOS Filters

Let $X(n)$ be an input signal, $Y(n)$ be an output signal and $\mathbf{X}(n)$ be a vector consisting of b samples which lie within the window at time index n: $\mathbf{X}(n) = [X(n-b_1), ..., X(n), ..., X(n+b_2)]^t \triangleq [X_1(n), ..., X_j(n), ..., X_b(n)]^t$ where $b = b_1 + b_2 + 1, X_j(n) = X(n - b_1 - 1 + j)$ and the superscript t denotes transposition. From now on, the time index n will be dropped from $\mathbf{X}(n)$, $X_j(n)$ and $Y(n)$ to simplify notation.

A. Stack Filters

The class of stack filters encompasses all filters that can be expressed as a composition of local MIN/MAX operations. For example, the filter represented by $Y = MAX\{MIN(X_1, X_2), MIN(X_2, X_3)\}$ is a stack filter. Let $X_i, 1 \leq i \leq b$, take an integer value from $\{0, 1, ..., M-1\}$ and $F(\mathbf{X})$ denote the output of a stack filter. Then

$$Y = \sum_{m=1}^{M-1} T_m[F(\mathbf{X})] = \sum_{m=1}^{M-1} F[T_m(X_1), ..., T_m(X_b)] \quad (1)$$

where $T_m(X_i)$ is a function which takes the value 1 if $X_i \geq m$ and 0, othewise. The second equality in (1) comes from the fact that a composite function of MIN/MAX operations commute with nondecreasing functions[10]. When the output of a filter for a multilevel input can be decomposed into a sum of binary filter outputs as in (1), such a filter is said to possess the *threshold decomposition* property. Since $T_m(X_i)$ is binary, $F[T_m(X_1), T_m(X_2), ..., T_m(X_b)]$ is also binary and this function is a Boolean function. Moreover, this Boolean function is a positive Boolean function(PBF), because the MIN/MAX operations for binary signals are equivalent to logical AND/OR operations. Therefore, stack filters are expressed as PBF's in the binary domain. For example, the stack filter in the example presented at the beginning of this section is represented by $y = x_1x_2 + x_2x_3$ where $x_1, \ldots, x_3$ are binary-valued inputs, and multiplications(additions) denote logical $AND(OR)$ operations. The class of stack filters encompasses all filters expressed as PBF's on the binary domain.

B. WOS Filters

The output Y of a WOS filter is written as

$$Y = T\text{-}th\ largest\{\overbrace{X_1, \cdots, X_1}^{W_1\ times}, \overbrace{X_2, \cdots, X_2}^{W_2\ times}, \cdots, \overbrace{X_b, \cdots, X_b}^{W_b\ times}\} \quad (2)$$

where $W_i, 1 \leq i \leq b$, is a positive integer. Let $X_{(i)}$ be the i-th $largest\{X_1, X_2, \cdots, X_b\}$ and $W_{(i)}$ be the corresponding weight. Then a necessary and sufficient condition for $X_{(k)}, 1 \leq k \leq b$, being the output of a WOS filter is

$$k = min\{j | \sum_{i=1}^{j} W_{(i)} \geq T\}. \quad (3)$$

WOS filters can be defined by using (3): in such a definition $W_1, W_2, \ldots, W_b$ and T are not necessarily limited to positive integers but can take arbitrary nonnegative real numbers. A WOS filter defined by (3) can always be converted into the expression in (2). Using (3), the output $f(\mathbf{x})$ of a WOS filter for binary inputs x_i is written as

$$f(\mathbf{x}) = \begin{cases} 1, & \text{if } \sum_{i=1}^{b} W_i x_i \geq T \\ 0, & \text{otherwise} \end{cases} \quad (4)$$
$$\triangleq U(\mathbf{a}^t\mathbf{z})$$

where $\mathbf{a} = [W_1, W_2, \ldots, W_b, T]^t, \mathbf{z} = [x_1, x_2, \ldots, x_b, -1]^t$ and $U(\cdot)$ is the unit step function. The function $f(\mathbf{x})$ in (4) is a special case of Boolean functions, and is called the *threshold* function[11]. A threshold function becomes a PBF if $W_i \geq 0$ and $T \geq 0$. Since WOS filters have nonnegative W_i's and T, they are stack filters.

3. Design of WOS filters

The optimization of WOS fiters under the MAE criterion is a slight modification of the optimization of stack filters. In this section, the design of WOS filters will be considered after briefly reviewing the procedure for designing optimal stack filters.

Suppose that a stack filter $F(\mathbf{X})$ is used to estimate a desired signal S, where the time index n of the desired signal is dropped. In [6] and [7] it is shown that the MAE between S and $F(\mathbf{X})$ is given by $MAE \triangleq E\{|S - F(\mathbf{X})|\} = \sum_{m=1}^{M-1} E\{|T_m(S) - F[T_m(X_1), T_m(X_2), \cdots, T_m(X_b)|\}$ and that this equation becomes $MAE = \sum_{j=0}^{2^b-1} c_j f(\mathbf{x}^j) + C$ where $f(\mathbf{x}^j)$ is a filter output for the j-th binary input vector $\mathbf{x}^j$, and c_j's and C are constants determined by

$$c_j = \sum_{m=1}^{M-1} \{Prob[s_m = 0, \mathbf{x}_m = \mathbf{x}^j \ at\ level\ m] - Prob[s_m = 1, \mathbf{x}_m = \mathbf{x}^j \ at\ level\ m]\}, \quad (5)$$

$$C = \sum_{j=0}^{2^b-1} \sum_{m=1}^{M-1} Prob\{s_m = 1, \mathbf{x}_m = \mathbf{x}^j \ at\ level\ m\}.$$

Here $s_m = 1$ if $S \geq m$ and 0, otherwise, $\mathbf{x}_m$ is the binary input vector at level m and $\{\mathbf{x}^j | 0 \leq j \leq 2^b - 1\}$ is the set of all possible binary vectors of dimension b. We denote $\mathbf{x}^j = [x^j(1), \cdots, x^j(b_1 + 1), \cdots, x^j(b)]^t$. Now, the optimal stack filter can be obtained through the following optimization:

Find out $f(\cdot)$ *minimizing*

$$J(f) = \sum_{j=0}^{2^b-1} c_j f(\mathbf{x}^j) \quad (6)$$

subject to the constraints

$$f(\mathbf{x}^j) \geq f(\mathbf{x}^i) \quad if \quad \mathbf{x}^j \geq \mathbf{x}^i. \quad (7)$$

In this problem, the inequlity in (7) is the stacking constraints which restricts $f(\cdot)$ to be a PBF. This optimization problem can be solved by LP, but the complexity of LP increases exponentially as b increases.

From (4) and (6) the optimization of WOS filters is described as follows:

Find out a vector **a** *minimizing*

$$J(\mathbf{a}) = \sum_{j=0}^{2^b-1} c_j U(\mathbf{a}^t \mathbf{z}^j) \quad (8)$$

subject to the constraints

$$W_i \geq 0 \quad for\ all\ i, \quad 1 \leq i \leq b \quad (9)$$
$$T \geq 0.$$

Due to the nonlinear function $U(\cdot)$, the optimization in (8) and (9) cannot be solved through LP. Next we shall see that the problem in (8) can be thought of as a two-class linear classification problem. The objective function $J(\mathbf{a})$ in (8) is minimized if $U(\cdot)$ takes the value 1(0) whenever c_j is negative(positive). That is, $J(\mathbf{a})$ is minimized if

$$U(\mathbf{a}^t\mathbf{z}^j) = \begin{cases} 1, & \text{if } c_j < 0 \\ 0, & \text{if } c_j > 0. \end{cases} \quad (10)$$

Therefore, if a vector **a** satisfying (10) exists, the optimization in (8) and (9) is equivalent to the following:

Find out a vector **a** *such that*

$$\begin{cases} \mathbf{a}^t\mathbf{z}^j \geq 0, & \text{if } c_j < 0 \\ \mathbf{a}^t\mathbf{z}^j < 0, & \text{if } c_j > 0 \end{cases} \quad (11)$$

under the constraints in (9).

The problem in (11) is a two-class linear classification problem. To be more precise, define $Z^+ \triangleq \{\mathbf{z}^j | c_j < 0\}$ and $Z^- \triangleq \{\mathbf{z}^j | c_j > 0\}$. Here, the vectors with $c_j = 0$ are neglected. Finding out a vector **a** satisfying (10) is equivalent to obtaining a linear discriminant function classifying Z^+ and Z^-. If Z^+ and Z^- are linearly separable, then a solution vector of (11) exist and can be obtained by the perceptron algorithm[9]. In this case, the designed WOS filter is equivalent to the optimal stack filter.

On the other hand, when Z^+ and Z^- are not separable, the solution to (11) does not exist and the optimization in (8) cannot be solved through (11). Use of the perceptron algorithm for the nonseparable case will result in a suboptimal solution to (8).

When an input signal X is equal to the desired signal S, the optimal filter is naturally the identity filter, which is a WOS filter. If this happens Z^+ and Z^- are linearly separable, as shown below.

Observation 1 : If $X = S$, a vector **a** separating Z^+ and Z^- is always exist.

Proof: If $X(n) = S(n)$, for any binary vector $\mathbf{x}_m(n) = [x_m(n-b_1), \cdots, x_m(n), \cdots, x_m(n+b_2)]^t$ for $0 \leq j \leq 2^b - 1$, $x_m(n) = s_m(n)$. Therefore if $x^j(b_1+1) = 1$, $Prob\{s_m = 1, \mathbf{x}_m = \mathbf{x}^j\ at\ level\ m\} = Prob\{\mathbf{x}_m = \mathbf{x}^j\ at\ level\ m\} \geq 0$ and $Prob\{s_m = 0, \mathbf{x}_m = \mathbf{x}^j\ at\ level\ m\} = 0$. Hence we have $c_j = -\sum_{m=1}^{M-1} Prob\{\mathbf{x}_m(n) = \mathbf{x}^j\ at\ level\ m\} \leq 0$ for all j for which $x^j(b_1+1) = 1$. Similarly we can show that $c_j \geq 0$ for all j for which $x^j(b_1+1) = 0$. Therefore, $Z^+ = \{\mathbf{z}^j | c_j < 0\} \subset \{\mathbf{z}^j | x^j(b_1+1) = 1\}$ and $Z^- = \{\mathbf{z}^j | c_j > 0\} \subset \{\mathbf{z}^j | x^j(b_1+1) = 0\}$. And a vector separating these two sets always exists. For instance Z^+ and Z^- are linearly separated by the identity filter for which $\mathbf{a} = [0, \cdots, 0, 1, 0, \cdots, 0, 1]^t$. ∎

From this obervation, it is deduced that Z^+ and Z^- tends to become linearly separable as the observation X closes with the desired signal S. This deduction is verified through computer simulation in Section 4.

Now we describe the proposed algorithm for designing WOS filters. The algorithm is the same as the standard perceptron algorithm except for the constraints in (9). As mentioned before, this algorithm finds out an optimal WOS filter when Z^+ and Z^- are linearly separable, and a suboptimal WOS filter, otherwise. The perceptron criterion function that will be used is

$$J_p(\mathbf{a}) = \sum_{\mathbf{z}^j \in Z_m} c_j \mathbf{a}^t \mathbf{z}^t \quad (12)$$

where $Z_m = \{\mathbf{z}^j | c_j \mathbf{a}^t \mathbf{z}^j > 0\}$ is the set of misclassified vectors. It is obvious that the criterion function $J_p(\mathbf{a})$ always has a nonnegative value.

Proposed Algorithm

Let $\mathbf{a}(k)$ be the weight vector at k-th iteration.

Step 1: Set $\mathbf{a}(1)$ to an arbitrary nonnegative $(b+1) \times 1$ vector and $k = 1$.

Step 2: Replace $\mathbf{a}(k)$ with

$$\mathbf{a}(k+1) = \mathbf{a}(k) - \rho_k \sum_{\mathbf{z}^j \in Z_m(\mathbf{a}(k))} c_j \mathbf{z}^j$$

where $Z_m(\mathbf{a}(k)) = \{\mathbf{z}^j | c_j \mathbf{a}^t(k)\mathbf{z}^j > 0\}$ and ρ_k is a sequence satisfying the limit conditions in [9, p.146]. Then replace all negative elements of $\mathbf{a}(k+1)$ with zero(0).

Step 3: Stop if $\sum_{i=1}^{b+1} |a_i(k+1) - a_i(k)| < \epsilon$. Otherwise, go to Step 2 after increasing k by 1. Here ϵ is a positive number which is sufficiently small, and $a_i(k)$ is the i-th element of $\mathbf{a}(k)$.

4. Experimental Results

Now we present a design example illustrating the performance of the proposed algorithm. This example concerns designing of a 2-dimensional 3×3 WOS filter for enhancing a noisy image corrupted by impulses. The original image shown in Fig. 1 is the 256×256 boat image with 8-bits of resolution. The image is corrupted by positive and negative impulses with values 200 and 50, respectively. Following the approach in [12] and [13] we assume that a portion of the original image and noisy image are given; the c_j's in (5) are estimated from the given data. Suppose that $N^j_{m,0}(N^j_{m,1})$ is the number of occurrences that binary vector $\mathbf{x}^j$ is observed in the noisy image and at the same time, $s = 0(1)$ in the original image on level m. Then, $\hat{c}_j = \sum_{m=1}^{M-1}(N^j_{m,0} - N^j_{m,1})/(I \times J)$ where $I \times J$ is the size of a left quarter of image(that is, $I \times J = 128 \times 128$ in this experiment) and $M = 256$. Using the perceptron algorithm with $\epsilon = 10^{-5}$ and $\rho_k = 1/(1+0.1k)$, we obtained the vectors $\mathbf{a}$ and the number of binary vectors which cannot be separated linearly by the obtained vector $\mathbf{a}$ while changing the probability of occurence of impulses, P_e. These results are summarized in Table I. When $P_e = 0$ the resulting WOS filter is the identity filter, and the WOS filters approch the median filter as P_e increases. It is interesting to note that the number of vectors which are not linearly separable increases as P_e increases. In this experiment, for $P_e \leq 0.025$ the designed WOS filter is the optimal one minimizing the MAE, and thus it is equal to the optimal stack filter.

In order to examine the noise suppression characteristics of the designed WOS filters, they are applied to the noisy images and MAE's between the original and recoverd images are evaluated. The results are summarized in Table II. For comparison, the MAE's obtained for 3×3 median and optimal stack filters are also shown in Table II. It is noted that the designed WOS filters outperform the median filter and that their performances are close to those of the optimal stack filters. The images for $P_e = 0.05$ are shown in Fig 2. Visually, the results for the WOS and the optimal stack filters look similar. The median filter suppressed impulses somewhat better than the others but caused severe blurring.

5. Conclusions

The perceptron algorithm has been applied to design WOS filters based on the observation that the optimization of WOS filters under the MAE criterion can be thought of as a two-class linear classification problem. Through computer experiments, it has been shown that the percetron algorithm can find optimal or near optimal WOS filters in practical situations.

References

[1] 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 Processing*, vol. SP-39, pp395-410, Feb. 1991.

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

[3] M. K. Prasad and Y. H. Lee, "Analysis of weighted median filters based on inequalities relating the weights," *Circuits Systems Signal Process*, vol. 11, no. 1, pp. 115-136, Jan. 1992.

[4] T. K. Nodes and N. C. Gallagher Jr., "Median filters: Some modifications and their properties," *IEEE Trans. Acoust., Speech, Signal Pro-*

cessing, vol. ASSP-29, pp. 739-746, Oct. 1982.

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

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

[7] E. J. Coyle, J. H. Lin, and M. Gabbouj, "Optimal stack filtering and structural approches to image processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, pp. 2037-2066, Dec. 1989.

[8] L. Yin, J. Astola, and Y. Neuvo, "Optimal weighted order statistic filters under the mean absolute error criterion," *Proc. of IEEE ICASSP*, pp. 2529-2532, Toronto, Canada, May 1991.

[9] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 1973.

[10] A. T. Fam and Y. H. Lee, "Selection filters and commutativity with memoryless nonlinearities," *Proc. of IEEE ISCAS*, pp. 1743-1746, New Orleans, Louisiana, May 1990.

[11] S. Mujroga, *Threshold Logic and Its Applications*, John Wiley and Sons, New York, 1971.

[12] B. Zeng, M. Gabbouj, and Y. Neuvo, "A unified design methods for rank order, stack and generalized stack filters based on classical Bayes decision," *IEEE Trans. Circuits Syst.*, vol. CAS-38, pp. 1003-1020, Sep, 1991.

[13] B. Zeng, H. Zhou, and Y. Neuvo, "Synthesis of optimal detail-restoring stack filters for image processing," *Proc. of IEEE ICASSP*, pp. 2533-2536, Toronto, Canada, May 1991.

TABLE I

The weight vectors of WOS filters and the number of vectors which cannot be separated linearly by the vector **a** (for all **a**, T is set at 1)

Pe	0.0	0.0125	0.025	0.05	0.1	0.2
Weight vectors	$\begin{bmatrix} 0.060 & 0.228 & 0.044 \\ 0.048 & 1.033 & 0.040 \\ 0.095 & 0.308 & 0.100 \end{bmatrix}$	$\begin{bmatrix} 0.459 & 0.408 & 0.057 \\ 0.003 & 0.997 & 0.001 \\ 0.006 & 0.065 & 0.005 \end{bmatrix}$	$\begin{bmatrix} 0.081 & 0.240 & 0.112 \\ 0.058 & 0.895 & 0.047 \\ 0.113 & 0.321 & 0.104 \end{bmatrix}$	$\begin{bmatrix} 0.124 & 0.259 & 0.125 \\ 0.124 & 0.754 & 0.123 \\ 0.123 & 0.246 & 0.123 \end{bmatrix}$	$\begin{bmatrix} 0.173 & 0.276 & 0.174 \\ 0.102 & 0.624 & 0.101 \\ 0.101 & 0.275 & 0.174 \end{bmatrix}$	$\begin{bmatrix} 0.180 & 0.274 & 0.182 \\ 0.181 & 0.456 & 0.182 \\ 0.182 & 0.274 & 0.182 \end{bmatrix}$
The number of vectors violating linear separability	0	0	0	5	24	38

TABLE II

MAE's between the original and filtered images

Pe	MAE		
	Median filter	WOS filters designed by proposed algorithm	Optimal stack filters
0.0	3.1962	0.0000	0.0000
0.0125	3.2549	0.4090	0.4090
0.025	3.3314	0.6458	0.6458
0.05	3.4742	1.0443	1.0435
0.1	3.7833	1.9120	1.8878
0.2	4.5890	3.3674	3.3613

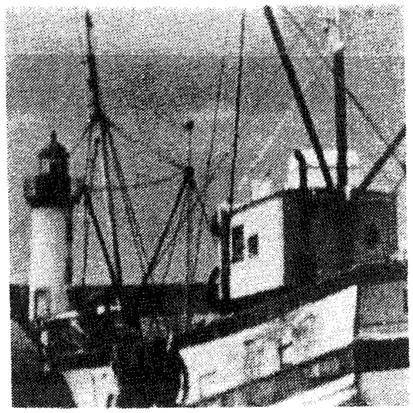

Fig. 1. Original image: Boat.

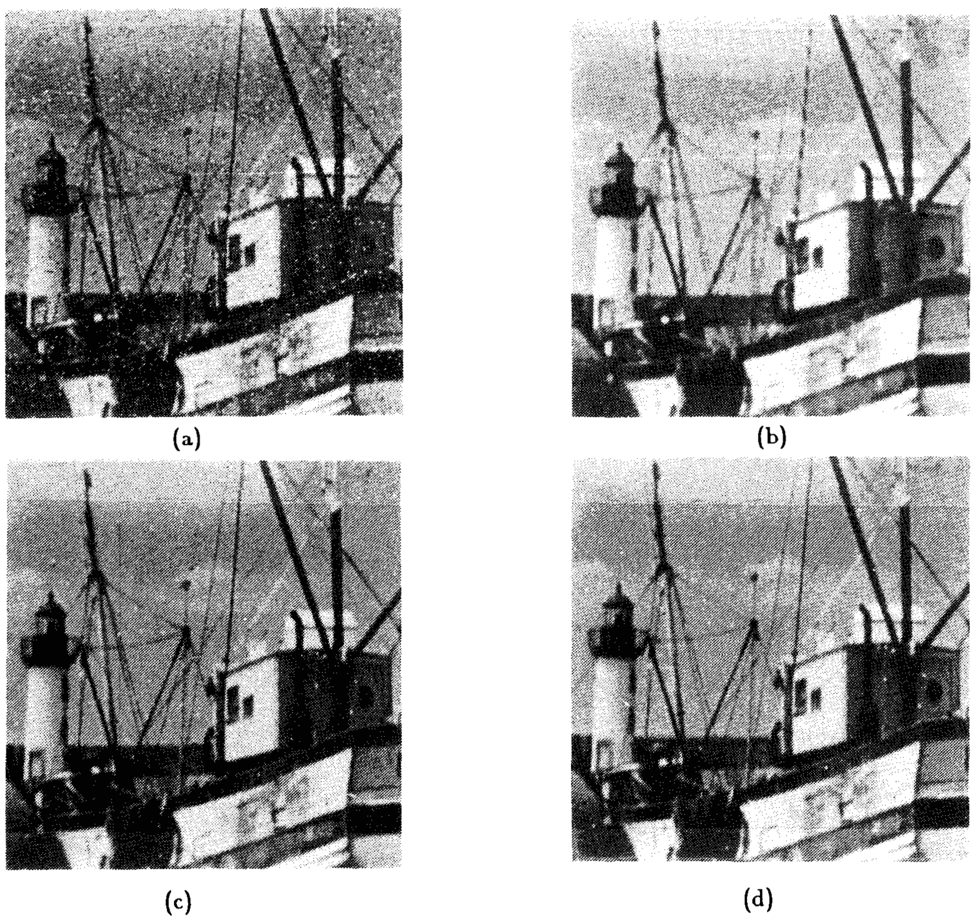

Fig. 2. Images for $P_e = 0.05$. (a) Noisy image. (b) Median filtered image. (c) Designed WOS filtered image. (d) Optimal stack filtered image.