Neural Network Training Methods for Optimal Weighted Order Statistics Filters Design

Ioan Tabus

Department of Automatic Control and Computers
Chair of Process Computers and Automation
"POLITEHNICA" University of Bucharest
313, Splaiul Independentei,
77206 Bucharest
ROMANIA

Moncef Gabbouj

Signal Processing Laboratory
Tampere University of Technology
SF-33 101 Tampere
FINLAND

Abstract: The problem of nonlinear filter design using Neural Network (NN) training methods has recently attracted the attention of a growing number of researchers. This paper aims to shortly review and unify the problem setting in which various training NN methods were used to desig optimal Weighted Order Statistics filters and to compare the performance obtained with each method.

Ioan Tabus was born in Romania, on January 6, 1957 . He received the M.S. degree in Electrical Engineering in 1982 and the Ph.D degree in Automatic Control in 1993, from the Polytehnical University of Bucharest.
In the academic year 1992-1993 he was a visiting researcher at the Technical University of Tampere, Finland.
From 1984 he has been with the Polytechnical University of Bucharest holding teaching positions: from 1984 to 1990 he was a Teaching Assistant and since 1990 a Teaching Lecturer in the Department of Automatic Control and Computers.
His research interests include neural networks for signal processing applications, system identification and adaptive nonlinear filtering.
He is co-author of three books and of numerous reports and papers in the fields of automatic control, system identification and signal processing.
In 1991 Dr. Tabus was co- recipient of the Award of the Romanian Academy.

Moncef Gabbouj received the B.S. degree in Electrical Engineering in 1985 from Oklahoma State University, Stillwater, and the M.S. and Ph.D degrees in Electrical Engineering from Purdue University, West Lafayette, IN in 1986 and 1989, respectively.
Since 1990, he has been with the Research Institute of Information Technology, Tampere, Finland, where he is currently a Senior Research Scientist. He also holds a teaching position in the Signal Processing Laboratory at Tampere University of Technology, Finland. His research interests include nonlinear signal and image processing, mathematical morphology, neural networks and artificial intelligence.
Dr. Gabbouj is the Director of the International University Program in Digital Signal Processing in the Signal Processing Laboratory at Tampere University of Technology.
He is member of Eta Kappa Nu and Phi Kappa Phi. Dr. Gabbouj was co-recipient of the Myril B.Reed Best Paper Award from the 32nd Midwest Symposium on Circuits and Systems.

1. Introduction

Weighted Order Statistics (WOS) filters are nonlinear filters which proved to be very effective in filtering images perturbed with nonGaussian noise. The operating principle of these filters may be seen as a generalization of the method known as "Order Statistics filtering" which produces the smoothing of a signal taking as output certain order statistics of a fixed length window sliding along the signal; WOS filters allow, in addition, emphasizing some samples in the window (e.g. admitting multiple copies of the samples).

The properties of this filter class are more easily analysed if the input signal (restricted to be M-integer valued) is first decomposed into M binary signals (this process being referred to as "threshold decomposition"). The overall filtering effect is obtained as follows: first, every binary signal is processed by a binary filter (a Boolean function which is positive and linearly separable), the resulting outputs being summed up to reconstruct a M- valued integer which is the output of the WOS filter.

The association of WOS filters with Neural Networks was used in the past aiming at two different purposes: as a tool for the fast implementation of WOS filters using VLSI structures [4] and as a way to design optimal WOS filters under the Mean Absolute Error (MAE) criterion [1, 2, 3].

This paper aims to shortly review and unify the problem setting in which various NN training methods were used to design optimal Weighted Order Statistics filters and to compare the performance obtained with these methods.

2. WOS Filters Structures

A WOS filter processes at any time t the input values situated in a window X(t) of length $N = N_1 + N_2 + 1$, including the current input x(t),

$$X(t) = \left[x\left(t-N_1\right) \ldots x(t) \ldots x\left(t+N_2\right)\right] =$$

$$= \left[X_1 X_2 \ldots X_N\right](t) \qquad (1)$$

The WOS filter parameters are the real valued weights $W(X_i) = W_i$ associated with each entry X_i in the window and the threshold weight, W_{N+1}, which form the parameter vector

$$W = \left[W_1 \ldots W_N W_{N+1}\right] \qquad (2)$$

Definition 1: *Real input and weights WOS filters*

A WOS filter processes the input as follows:

- The samples in the window are ordered decreasingly, resulting in an ordered vector

$$OX(t) = \left[X_{(1)} \ldots X_{(N)}\right](t) = \left[X_{i_1} \ldots X_{i_N}\right](t) \qquad (3)$$

where $X_{(k)}$ denotes the k'th element in the ordered string;

- For every $r = 1,\ldots,N$; the inequality $\sum_{i=1}^{r} W\left(X_{(i)}(t)\right) \leq W_{N+1}$ is tested and the first r for which the inequality holds is selected (denote it r_1);
- The filter output will be the r_1'th value in the ordered vector, $y(t) = WOS(W,X(t)) = X_{\left(r_1\right)}(t)$.

Definition 2: *Integer weights WOS filters with real inputs*

If the weights are integer numbers, then the output of a WOS filter can be obtained as follows:

- Build an extended window duplicating W_i, times each sample X_i, in the input window;
- Sort in a decreasing order the samples in the extended window;
- Select the W_{N+1}'th element in the ordered window as the output of the filter.

If the input of the filter is an integer value less than M, then definitions 1 and 2 can be rephrased using the "threshold at level m" operator, $T_m(.)$ which acts on the integer value I to give

$$T_m(I) = \begin{cases} 1, & \text{if } I \geq m; \\ 0, & \text{if } I \leq m. \end{cases} \qquad (4)$$

It is obvious that the original input can be easily recovered from its thresholded versions, $\{I^1,\ldots,I^{M-1}\}$ summing up all the binary versions

$$I = \sum_{m=1}^{M-1} T_m(I) = \sum_{m=1}^{M-1} I^m \qquad (5)$$

Definition 3: *Real weights WOS filters with integer inputs*

A WOS filter processes the input in four stages:

- The samples from the input window are thresholded at all levels m between 1 and M-1, resulting in M-1 windows with binary elements

$$X^m(t) = T_m(X(t)) = \left\{T_m\left(X_1\right),\ldots,T_m\left(X_N\right)\right\} \qquad (6)$$

- For every window $X^m(t)$, $m = 1, \ldots, M-1$ the binary output $y^m(t) = WOS(W,X^m(t))$, is computed as follows

$$y^m(t) = \begin{cases} 1, & \text{if } \sum_{i=1}^{N} W_i X_i^m(t) \geq W_{N+1}; \\ 0, & \text{else} \end{cases} \qquad (7)$$

- The binary outputs are summed-up to obtain the output of the filter

$$y(t) = \sum_{m=1}^{M-1} y^m(t) \qquad (8)$$

All the three Definitions given above express various structures which can be used for implementing WOS filters. Since these definitions are equivalent when applied to integer weights and inputs, it follows that we can use different structures in the design stage and in the implementation stage. We shall focus here on the advantages which can be taken of a particular structure in solving the optimal WOS filter problem in the design stage.

3. Optimal WOS Filter Design Problem

3.A Optimality under MAE criterion

One way to approaching the optimal design of order statistics type filters is to make some assumptions concerning data and noise models and then, under stationarity hypothesis, to find the parameters which minimize the Mean Absolute Error between the original signal and the filter output obtained when the corrupted signal is at the filter input.

Problem 1 *Given*

- the noise and signal model classes, $s(\bullet)\in S, n(\bullet)\in N$,

find the filter parameters W* which minimize the criterion

$$J(W) = E[\,|\,s(t) - WOS(W,s(t)+n(t))\,|\,] \qquad (9)$$

This approach was intensively used in the past, but the solution is suboptimal in practice because in most of the applications the stationarity hypothesis is not fulfilled. The same situation is met with in linear prediction theory, where the optimality criterion is sometimes formulated in terms of the expectation of the squared error, leading to the "autocorrelation" solution, while formulating the optimality as the minimum sum of square errors leads to the "autocovariance" solution. Here we shall not pursue any further the expectation based criterion, but we shall seek for an alternative best suited to practical applications.

3.B Optimality under Least Absolute Error criterion

An approach to the optimal WOS filter design closer to "learning" principles is to consider as known some representative sets of inputs and desired outputs and to find the parameters which bring the filter closest to the specified desired behavior. More formally, we define

Problem 2 *Given*

- the input set $\left\{X(t)\right\}_{t=1}^{T}$ (the real signal x(t)

is already arranged in windows as in (1));

- the desired output set $\left\{d(t)\right\}_{t=1}^{T}$

find the filter parameters W* which minimize the criterion

$$J(W)=\frac{1}{T}\sum_{t=1}^{T}\left|d(t)-WOS(W,X(t))\right| \qquad (10)$$

We shall use, in the following, the superscript "m" to denote the thresholded (at level m) version of a variable and the notation wos (W,q) for the (binary) output of the filter whose input is the binary window q.

Lemma 1 The criterion (10) can be written in the following form:

$$J(W)=\frac{1}{T}\left(C+\sum_{q\in\{0,1\}^N} c(q)wos(W,q)\right) \qquad (11)$$

where the coefficients C,c(q) can be computed from the data set using the threshold decomposition of the data as an intermediate step.

Proof See Appendix.

We can give now another, but equivalent formulation of Problem 2:

Problem 3 *Given*

- the coefficients $\left\{c(q)\right\}_{q\in\{0,1\}^N}$ (carryin

information about the original data set)

find the filter parameters W* which minimize the criterion

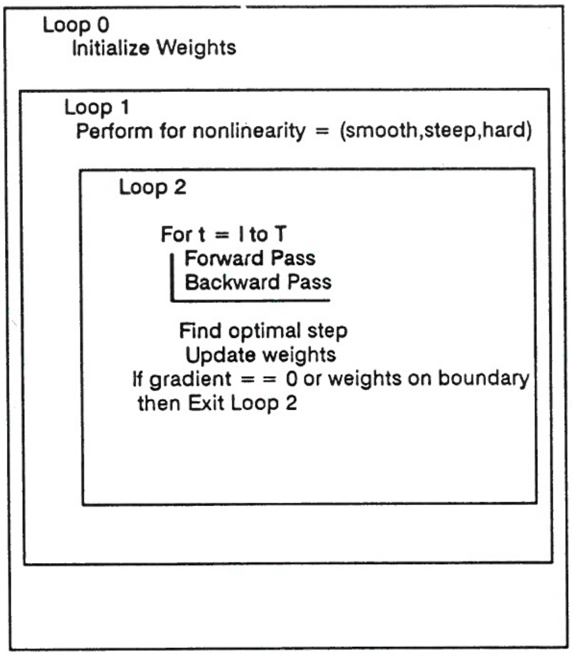
$$J_1(W)=\sum_{q\in\{0,1\}^N} c(q)wos(W,q) \qquad (12)$$

Remark

Both problems deal with the same unknown parameters and therefore the dimensionality of the problem is established only by the length of the "data set", i.e. for Problem 2, the dimensionality is T and, for Problem 3, the dimensionality is 2^N. For small values of N, it is simpler to solve Problem 3, but for large values of N, Problem 3 is not any more tractable.

4. Optimal Design Using Neural Network Training Methods

The WOS filter structure presented in Definition 1 can be easily implemented in the form of a four layer Neural Network [3] as in Figure 1. Since the nonlinearities in the third and fourth layers are step functions, the classical Backpropagation (BP) training method cannot be directly applied to this



structure. The next procedure [2] tries to overcome this difficulty.

Procedure # 1 (First *attempt* at solving Problem 2)

1. Replace the hard nonlinearity in layers 3 and 4 by sigmoidal nonlinearities.

2. Apply the Backpropagation procedure to find the parameters W.

□

This procedure presents two strong drawbacks.

- As it is well- known the BP procedure stops in a local minimum of the criterion function (which sometimes happens to be the global optimum). But for Problem 2, it was shown in [3] that there was a large number of local minima, most of them being worst than the simple classical median filter;
- The structure obtained at the end of the training is not exactly a WOS filter since the nonlinearity is sigmoidal and no control has been imposed on its steepness.

A way to confining the drawbacks of the preceding procedure is to use a more structured training procedure, including the principles of Monte Carlo search (for looking at as many local minima as possible) and the principles of deterministic "simulated annealing").

Procedure # 2 (solving Problem 2 by deterministic simulated annealing over stages of BP search).

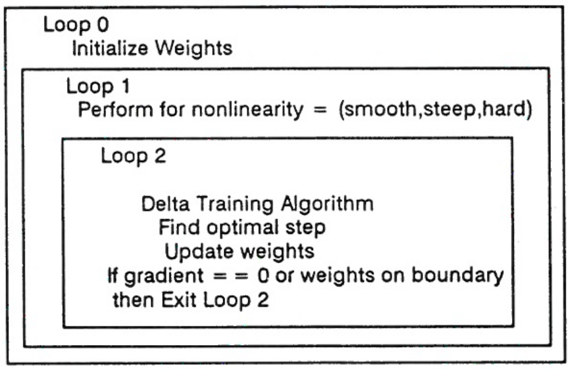
Select the best local minimum

□

This procedure was shown in [3] to be effective in finding parameters close to the global optimum (see [3] for details).

Next we shall describe a procedure (with the same structuring hierarchy as the preceding one) for solving Problem 3 in order to take advantage of the preprocessing level, which compresses the information in the data set (see also remarks under Problem 3 which indirectly discuss the compression ratio obtained).

Procedure # 3 (Solving Problem # 3 by deterministic simulated annealing over stages of delta rule search)



Select the best local minimum

□

The structure of the iterative Delta Training algorithm is depicted in Figure 2.

Remarks

1. The global minimum of criterion (12) could be obtained if

$$\mathrm{wos}(W,q) = \begin{cases} 1, & \text{if } c(q) < 0; \\ 0, & \text{if } c(q) \geq 0 \end{cases} \tag{13}$$

or equivalently, if

$$\begin{aligned} W_q^T &> W_{N+1}, \quad \text{if } c(q) < 0; \\ W_q^T &\leq W_{N+1}, \quad \text{if } c(q) \geq 0 \end{aligned} \tag{14}$$

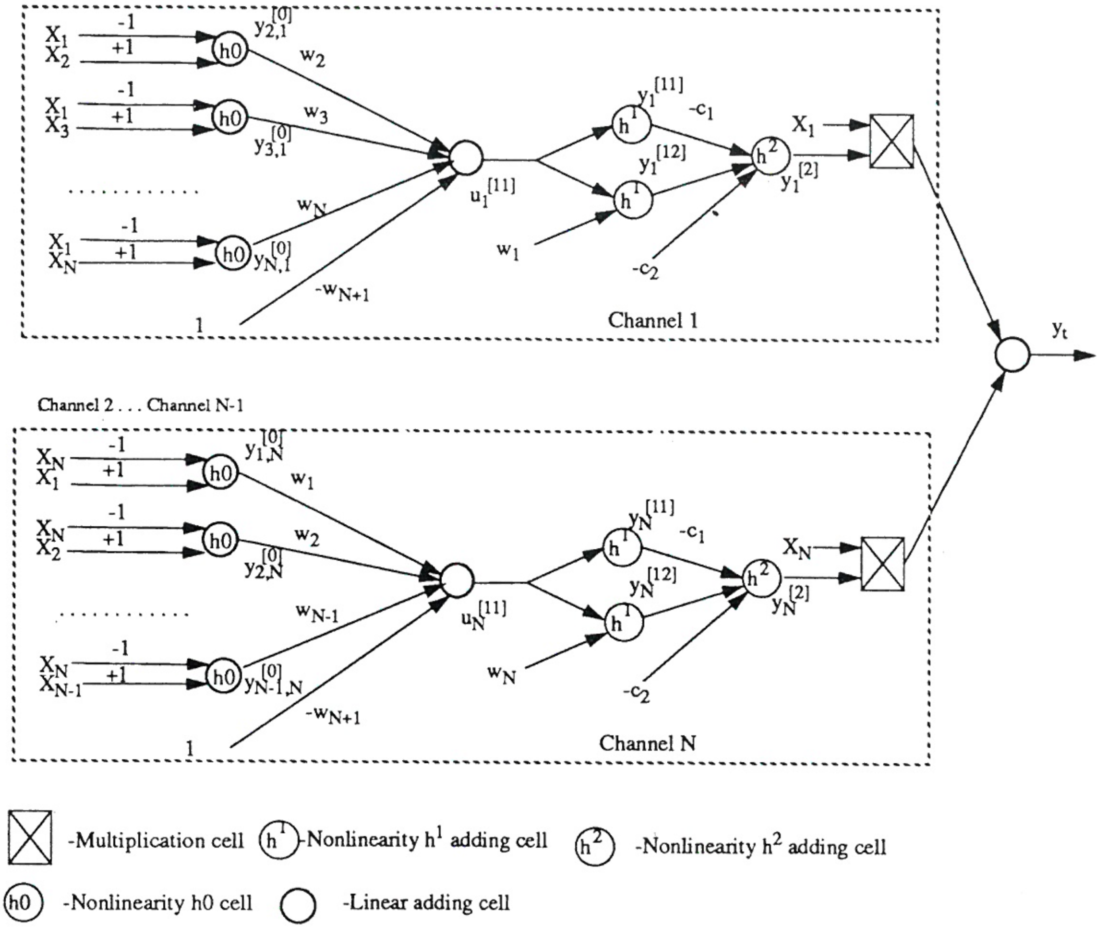


Figure 1. Neural Network Structure Implementing Real Input WOS Filters

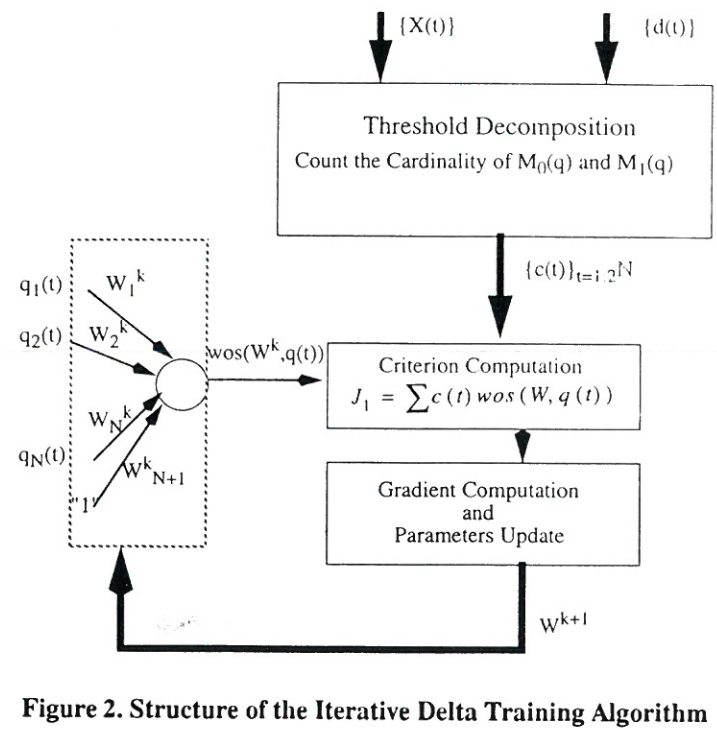


Figure 2. Structure of the Iterative Delta Training Algorithm

for all vectors $q \in \{0,1\}^N$. There are 2^N such inequalities and only $N+1$ free parameters; there is no reason why to believe that the set of inequalities is consistent (i.e. there is a vector W which satisfies all of them). If the set of inequalities is consistent, Problem 3 can be restated as follows:

Problem 4 *Given*

- The input set $\{q \mid q \in \{0,1\}^N\}$;
- The desired set $\{d(q) \mid d(q)=1 \text{ for } c(q)<0$ and $d(q)=0 \text{ for } c(q) \geq 0\}$

find the perceptron per(W*, q), which minimizes

$$J_2(W)=\frac{1}{T}\sum_{t=1}^{T}\left|d(t)-per(W,q)\right| \tag{15}$$

This is the standard formulation of the problem which is solved by the Perceptron algorithm (see[6] for details about the algorithm).

2. If the set of the inequalities (14) is not consistent, the Perceptron will still offer a solution, but it is a suboptimal one. The drawback is that this solution does not take into account the value of c(q), as criterion (12) does (in criterion (15) only the sign of the coefficient c(q) is significant). It follows that Problem 4 can hardly be considered a good approach to optimal WOS filter design under the MAE criterion (as claimed in [1]).

Conclusions

The problem of WOS filter design is well suited to be solved using NN training methods.

This problem differs from most other applications of NN in two major aspects: first, the structure of the NN which must be used is known in advance (there is no need for structure selection criteria) and second, the nonlinearities in the final network *must* be step functions (in order to ensure the Order Statistics structure to the network, desirable for other reasons than that of only the minimization of MAE criterion, i.e. good behaviour in edge preservation).

This paper presented some NN structures and training procedures effective in solving the problem of WOS filter design. The choice among various structures and methods is mostly restricted by the length of the window size.

For small length size the procedures using threshold decomposition are simpler, including a preprocessing stage which reduces a lot of the computational load in the iterative stage.

For large window sizes, the only effective procedure is # 2, based on real input domain processing of the training set.

The usefulness of these design techniques stems not only from their straightforward implementation as filters but also from the means they offer for continuing the study of optimal nonlinear filters whose behaviour is still hidden in many aspects, in part due to the lack of effective optimal design methods.

Appendix

Proof of Lemma 1

$$T \cdot J(W) = \sum_{t=1}^{T}\left|d(t)-WOS(W,X(t))\right| =$$

$$=\sum_{t=1}^{T}\sum_{m=1}^{M-1}\left|d^m(t)-wos(W,X^m(t))\right| \tag{16}$$

The last equality holds due to the following property of threshold decomposition (4)

$$d \geq (\leq) y \Rightarrow \forall m,\ d^m \geq (\leq) y^m \tag{17}$$

from which follows

$$\left|d-y\right| = \sum_{m=1}^{M-1}\left|d^m-y^m\right|.$$

We shall regroup the elements which appear in the summation (16), emphasizing the terms for which the binary windows $X^m(t)$ are identical vectors, q (the set of all such vectors is $\{0,1\}^N$). Then, we shall split the sum according to the partition into two complementary sets:

- $M_0(q)$ – the set of all pairs (t, m) for which $d^m(t)=0$ and $X^m(t)=q$; denote $N_0(q)$ the cardinality of this set;
- $M_1(q)$ – the set of all pairs (t, m) for which $d^m(t)=1$ and $X^m(t)=q$; denote $N_1(q)$ the cardinality of this set;

$$T \cdot J(W) = \sum_{t=1}^{T}\sum_{m=1}^{M-1}\left|d(t)-wos(W,X^m(t))\right| =$$

$$= \sum_{q\in\{0,1\}^N}\left[\sum_{(t,m)\in M_0(q)}\left|d(t)-wos(W,X^m(t))\right| + \right.$$

$$\left. + \sum_{(t,m)\in M_1(q)}\left|d(t)-wos(W,X^m(t))\right|\right] =$$

$$= \sum_{q\in\{0,1\}^N}\left[\sum_{(t,m)\in M_0(q)}\left|-wos(W,q)\right| + \right.$$

$$\left. + \sum_{(t,m)\in M_1(q)}\left|1-wos(W,q)\right|\right] =$$

$$= \sum_{q\in\{0,1\}^N}\left[N_1 + \left(N_0(q)-N_1(q)\right) wos(W,q)\right] =$$

$$= C + \sum_{q\in\{0,1\}^N} c(q)\, wos(W,q) \qquad (18)$$

with

$$c(q) = N_0(q) - N_1(q) \qquad (19)$$

$$C = \sum_{q\in\{0,1\}^N} N_1(q) \qquad (20)$$

q. e. d.

REFERENCES

1. JEONG, B. and LEE,Y. H., **Design of Weighted Order Statistic Filters Using the Perceptron Algorithm**, Proc. IEEE Winter Workshop on Nonlinear Digital Signal Processing,. Tampere, Finland, January 1993,pp. 3.1-1.1-6.
2. YIN, L., ASTOLA, J. and NEUVO, Y. , **Adaptive Weighted Order Statistic Filter Using Backpropagation Algorithm**, Proc. IEEE Int. Symp. Circ. Syst., New Orleans, Louisiana, 1990, pp.499-502.
3. TABUS,I.,GABBOUJ, M. and YIN, L., **Real Domain WOS Filtering Using Neural Network Approximations**, Proc. IEEE Winter Workshop on Nonlinear Digital Signal Processing, Tampere,Finland,January 1993, pp. 7.2-1.1-6.
4. VAINIO, O., YIN, L. and NEUVO, Y., **On the Properties and VLSI Implementation of Generalized Median Operations**, Proc. IEEE Workshop on VLSI and Sign. Processing,San Diego, CA., November 1990, pp. 167-176.
5. GABBOUJ,M., COYLE, E.J.and CALLAGHER,C., **An Overview of Median and Stack Filtering**, CIRCUITS, SYSTEMS, SIGNAL PROCESSING, Vol.11, 1, 1992, pp. 7-45.
6. MINSKY, M. and PAPERT, S., **Perceptrons. An Introduction to Computational Geometry**, MIT PRESS, Cambridge, Massachusetts, 1988.