From Median Filters to Optimal Stack Filtering

Edward J. Coyle, Moncef Gabbouj†, and Jean-Hsang Lin‡

School of Electrical Engineering
Purdue University
West Lafayette, IN 47907, USA

†Signal Processing Laboratory
Tampere Univ. of Technology
Tampere SF-33101, FINLAND

‡Dept. of Electrical Engineering
University of Delaware
Newark, DE 19716, USA

Abstract

Within the last two decades a useful, nontrivial theory of nonlinear signal processing has been built around the median filter. We outline the development of this theory from its beginnings in the study of the noise removal properties and structural behavior of the median filter to the recently developed theory of optimal stack filtering.

A recent application of stack filters is provided to demonstrate the effectiveness of this new theory.

1. Introduction

Linear filters have long been the primary tool for signal and image processing. They are easy to implement and analyze and, perhaps most importantly, the linear filter which minimizes the mean squared error criterion can usually be found in closed form. Furthermore, they are optimal among the class of all filtering operations when the noise is additive and Gaussian.

Unfortunately, a small deviation from this Gaussian assumption sometimes leads to a severe deterioration in the performance of linear filters. In the many applications in which non-Gaussian noise arises, linear methods have thus proven to be inadequate for signal smoothing and noise reduction. One such case occurs in the presence of speckle noise. Other types of non-Gaussian and/or signal-dependent noise also cause problems. We believe that these cases occur more frequently than not; therefore, linear methods are not completely satisfactory when dealing with real signals and noise rather than simply computer simulations.

So, what should be done? The obvious answer is to use a filter that is not linear. There are, however, many classes of non-linear filters, and the task of choosing the right class is itself a challenge. The user could consult some type of look-up table to determine which filter or class of filters best fits the problem at hand [PiV].

One filter that would certainly appear in any such catalogue would be the median filter. The median filter, or "running median" as it was called in the first publications in which it appeared [Tuk], consists of a window, usually of odd width, which is stepped one sample at a time along a signal. At each position of the window, the sample values inside are ranked according to their magnitude and the middle element in this ranking is defined to be the output.

Typically, the window is assumed to have width $2N+1$ where N is any positive integer. Suppose that the window is centered on the k'th sample in the input sequence and that the $2N+1$ points in the window, in time-order, are specified by the vector

$$(x_{k-N}, x_{k-N+1}, \cdots, x_k, \cdots, x_{k+N}).$$

We want to find y_k, where

$$y_k = median(x_{k-N}, x_{k-N+1}, \cdots, x_{k+N}) \tag{1.1}$$

which is the output of the median filter when the window is centered on the k'th sample of the signal.

First, the samples in the window are reordered according to their rank, with $x_{(i)}$ denoting the sample of i^{th} rank. The samples in the window, in rank-order, would then be

$$(x_{(1)}, x_{(2)}, \cdots, x_{(2N+1)}).$$

Suppose, for instance, that N=2 and that the samples, in time order, in the window are

$$(x_{k-2}, x_{k-1}, x_k, x_{k+1}, x_{k+2}) = (8,1,6,4,1);$$

in rank-order they would be

$$(x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}, x_{(5)}) = (1,1,4,6,8).$$

The median value of $2N+1$ samples is given by $x_{(N+1)}$, which for the example just given would be: $x_{(3)} = 4$. We thus have $y_k = 4$. The window is then moved so it is centered over the k+1'st sample in the input sequence and the output value y_{k+1} is computed by following the above procedure.

Figure 1.1 shows the input and output for a window width three median filter, and also shows how the window moves along the signal. Note from the figure that the median filter can preserve edge structures while deleting impulsive structures.

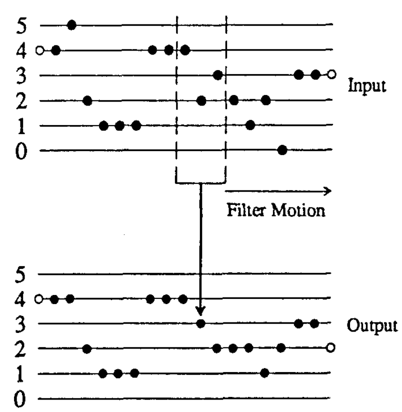

Fig. 1.1: The window width 3 (WW3) median filter.

Several questions can now be asked. Since the median filter is nonlinear, what can we say about it when we can't use the traditional tools of linear analysis to understand it? Are there whole classes of filters with the median's ability to preserve details such as edges while removing noise?

CH 3006-4/91/0000 – 0009 $1.00 © IEEE

These questions have attracted the attention of a growing number of researchers over the last few decades. The result of this attention is a large number of papers, dissertations and books written on median and median-related filters.

In the following subsections, we present a very brief literature review on this subject. The glaring omissions are order statistic filters, weighted median filters, vector median filters, selection filters, rank order processors, and Ξ-filters. Other related areas not covered here are morphology, cellular automata and neural networks. For a full review, see [PiV] and [GCG].

2. Theoretical Analysis of Median Filters

In the 1970's, J.W. Tukey [Tuk, pg 210] introduced the "running median" as a tool for smoothing discrete data. Since then, its name has changed to "median filter," and it has been used in several areas of digital signal processing, including speech processing, image enhancement, and seismic data analysis.

Some of the earliest theoretical results on the behavior of the median filter concerned the existence and nature of its invariant signals, or *root signals*. An example of a root signal is provided in Figure 2.1, in which it is easy to see that the input and output signals are the same.

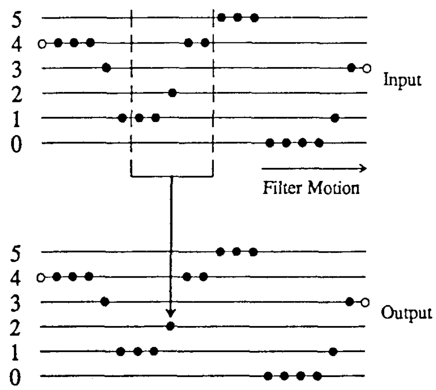

Fig. 2.1: Root signal of WW5 median filter.

The locally monotone nature of these finite-length fixed points was proven for all median filters of odd window width in [Tya] and [GaW]. In [GaW] it was also shown that any finite length signal is filtered to a root signal after a finite number of passes of a median filter of a fixed window width. This is an important result -- it ties median-type filters to neural nets since such convergence behavior is the primary characteristic of associative memories.

The many studies of the root signals and convergence behavior of the median filter form a significant part of what is now called the study of the structural behavior of median filters and median-type filters (see [CLG][GCG] for a review). The goal is to determine the type and number of root signals, whether every signal is filtered to a root signal, and which structures are preserved, created, modified or deleted by median-type filters.

The early studies of the structural behavior of the median filter complemented results known for quite some time in the statistics literature: the median of iid samples of a double exponential random variable is the maximum likelihood estimator of the mean of that random variable; and the conditional median at each time instant t is the minimum mean absolute error estimator of the signal value at time t, where the conditioning is on the past history up to time t of the noise corrupted observations of the signal.

With the above theoretical results, the success of median filters was becoming well understood and it was becoming clear which applications were well-suited for median filtering. For instance, its ability to preserve edges and delete impulsive structures combined with its robustness with respect to different noise types is why it has been used so extensively in image processing.

All of the above praise might lead one to believe that the median is the perfect filtering operation, which is not true. The median can introduce edge jitter and streaking into an image (see, for example, [BHM]).

These shortcomings, and the desire to find other filters similar to the median but which allow more design choices than just the window width, has motivated many generalizations of the median filter (see [GCG][PiV] for a history).

3. From Median Filters to Stack Filters

In 1984, an important theoretical tool for analyzing median and median-type filters was developed by Fitch et al, [FCG1]. They showed [FCG1][FCG2] that all rank-order-based filters possess a weak superposition property called the *threshold decomposition,* which says that median filtering any sequence whose elements take on values in the set $Q = \{0,1, \cdots ,M-1\}$ is equivalent to decomposing the signal into binary sequences by thresholding at each level from 1 through $M-1$, filtering each resulting binary sequence by a (binary) median filter, and then adding up the results.

To express this property more precisely, define $T_i(\cdot)$ to be the operator which thresholds its argument at the level i:

$$T_i(x) = \begin{cases} 0 & \textit{if } x < i \\ 1 & \textit{if } x \geq i \end{cases} \tag{3.1}$$

For instance, $T_3(5) = 1$ and $T_3(2) = 0$. We will also apply thresholding to vectors, and will define it as follows:

$$T_i\left[(x_1, x_2, \cdots, x_n)\right] = \left[T_i(x_1), T_i(x_2), \cdots, T_i(x_n)\right] \tag{3.2}$$

For example, we would have $T_3\left[(6,2,1,3,7)\right] = (1,0,0,1,1)$.

Then, suppose a median filter of window width 3 whose input sequence takes on values in Q. The threshold decomposition states that for every k

$$\begin{aligned} Med(x_{k-1},x_k,x_{k+1}) &= Med\left[\sum_{i=1}^{M-1} T_i\left[(x_{k-1},x_k,x_{k+1})\right]\right] \qquad (3.3) \\ &= \sum_{i=1}^{M-1} Med\left[T_i\left[(x_{k-1},x_k,x_{k+1})\right]\right]. \end{aligned}$$

This weak superposition property, which is illustrated in the example provided in Figure 3.1, allows the analysis of the effects of the median on multi-valued sequences to be reduced to the study of its effects on binary sequences.

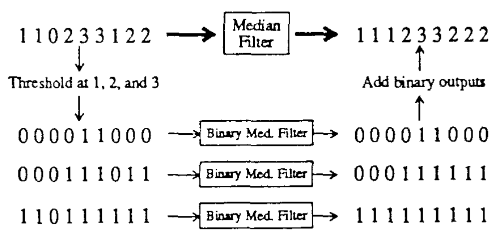

Fig. 3.1: Threshold decomposition of WW3 MF.

Stack filters were defined as the class of digital filters having the threshold decomposition property and the ordering property [NaR], called the stacking property in [WCG], exhibited by the median filter. The stacking property led to a connection to positive Boolean functions [Gil].

In Figure 3.2, we show the threshold decomposition architecture of a specific stack filter known as the asymmetric median filter. Mathematically, this figure states the following superposition property for the stack filter $S_f(\cdot)$ based on the positive Boolean function $f(\cdot)$:

$$S_f(x_{k-1}, x_k, x_{k+1}) = S_f\left[\sum_{i=1}^{M-1} T_i((x_{k-1}, x_k, x_{k+1}))\right] \quad (3.4)$$
$$= \sum_{i=1}^{M-1} S_f\left[T_i((x_{k-1}, x_k, x_{k+1}))\right]$$
$$= \sum_{i=1}^{M-1} f\left[T_i((x_{k-1}, x_k, x_{k+1}))\right].$$

The difference between this filter and the median filter shown in Figure 3.1 is entirely in the Boolean function used on each

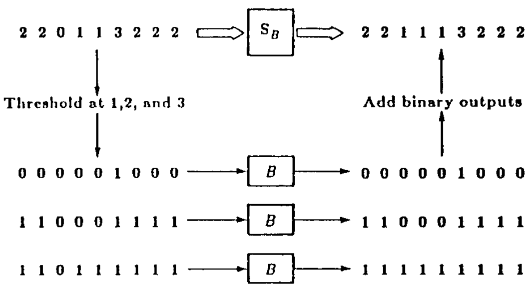

Fig. 3.2: A window width 3 stack filter S_B based on the positive Boolean function B.

threshold level. As long as this Boolean function is positive, the filter possesses the stacking property and the threshold decomposition property. The fact that these filters were maxima of minima was then proven via morphology in [MaS]; it is a consequence of [NaR], which shows that maxima and minima are the multi-level generalization of the logical *and* and logical *or* operations used in positive Boolean functions.

One of the primary advantages of stack filters is that a composition of stack filters is again a stack filter, which follows since the composition of two positive Boolean functions is another positive Boolean function.

4. Optimal Stack Filtering

Perhaps the most important advantage of stack filters is the fact that the optimal stack filter can be found via a linear program when the goal is to minimize the mean absolute error criterion [CoL][CLG].

The process $R(j)$ at the input of a stack filter is assumed to be a corrupted version of some desired process $S(j)$. The corruption may be caused either by a noise process $N(j)$ or by some intentional operation, such as a modulation scheme.

At each time instant j, the stack filter output is an estimate, called $\hat{S}(j)$, of the desired process $S(j)$. This estimate is based on the observed sequence $\vec{R}_n(j)$ in the window of the stack filter; thus,

$$\hat{S}(j) = S_f(\vec{R}_n(j)).$$

The goal is to pick a stack filter from the class of window width n stack filters such that the average mean absolute error per time unit between the filter's output and the desired signal is minimized. If $S(j)$ and $R(j)$ are jointly stationary, then the cost to be minimized is

$$E\left[\, |S(j) - S_f(\vec{R}_n(j))|\, \right]. \quad (4.1)$$

One rationale for using the absolute error criterion is that it nicely reduces the estimation error of the stack filter to the sum of the decision errors incurred by the Boolean filters on each level of the threshold decomposition architecture. To illustrate this fact, let $s_k(j) = T_k(S(j))$; then

$$E\left[\,|\, S(j) - S_f(\vec{R}_n(j))\,|\,\right] = E\left[\,|\sum_{k=1}^{M-1} s_k(j) - f(T_k(\vec{R}_n(j)))\,|\,\right]$$

by threshold decomposition,

$$= E\left[\sum_{k=1}^{M-1} |\, s_k(j) - f(T_k(\vec{R}_n(j)))\,|\,\right]$$

by the stacking property,

$$= \sum_{k=1}^{M-1} E\left[\,|\, s_k(j) - f(T_k(\vec{R}_n(j)))\,|\,\right]. \quad (4.2)$$

Because of the above reduction, and since a stack filter is completely defined by the Boolean function on each level of its architecture, the cost function to be minimized can be expressed in terms of a linear function of the output variables of the Boolean function $f(\cdot)$, [CoL][LiC].

Let the output of $f(\cdot)$, when the length n binary sequence $\vec{s}_j$ is at its input, be called the decision variable $P_f(1|\vec{s}_j) \in \{0, 1\}$. This decision variable specifies the probability that the filter output is a 1 when the vector $\vec{s}_j$ appears in the filter's window. The Boolean function $f(\cdot)$ can therefore be represented as a length 2^n vector $\vec{P}_f$, whose k'th entry is $P_f(1|\vec{s}_k)$. With these definitions, the mean absolute error in the last equation in (4.2) can be reformulated as the following cost function:

$$Cost = \sum_{j=1}^{2^n} C_j \cdot P_f(1|\vec{s}_j). \quad (4.3)$$

C_j can be interpreted as the cost incurred by $f(\cdot)$ for deciding a 1 when seeing $\vec{s}_j$. Note that this cost function is a linear function of the filter decision variables.

The stacking constraints can be expressed as a set of inequalities in terms of these decision variables; specifically,

$$P_f(1|\vec{s}_i) \le P_f(1|\vec{s}_j) \qquad \text{if} \quad \vec{s}_i \le \vec{s}_j, \quad (4.4)$$

where for any two length n real sequences $\vec{s}$ and $\vec{t}$, $\vec{s} \le \vec{t}$ if and only if each entry of $\vec{s}$ is less than or equal to the corresponding entry in $\vec{t}$.

The optimal filtering problem over the class of stack filters under the mean absolute error criterion can therefore be formulated as an zero-one integer linear program -- the goal being to determine whether each $P_f(1|\vec{s}_j)$ should be 0 or 1.

By exploiting the structure of the constraint matrix, which is totally unimodular, this zero-one integer linear program can be formulated as the following linear program [CoL]:

$$\textit{minimize} \quad \sum_{j=1}^{2^n} C_j \cdot P_f(1|\vec{s}_j) \quad (4.5)$$

subject to the constraints:

$$P_f(1|\vec{s}_i) \le P_f(1|\vec{s}_j) \qquad \text{if} \quad \vec{s}_i \le \vec{s}_j \quad (4.6)$$

$$0 \le P_f(1|\vec{s}_i) \le 1 \quad \forall\, i. \quad (4.7)$$

An adaptive algorithm is now available for minimum mean absolute error stack filtering [LSC]; it eliminates the need for complex computation of the statistics of the signal and noise processes, provided an appropriate training sequence is available. J.-H. Lin has shown that it now takes approximately 8 minutes of CPU time on a Sparc station to train a 3x3 stack filter on a 256x256 eight bit image.

The theory of root signals for stack filters has been combined with the theory of minimum mean absolute error stack filtering developed above [GaC]. This unified approach allows the designer to pick a stack filter which minimizes noise subject to constraints on its structural behavior.

4. Application

One recent application of stack filters is to the problem of edge detection in noisy images [YBDC]. Very good algorithms exist if the noise is Gaussian, but they perform very poorly if there is even a small contamination of heavy-tailed noise. A stack filter based edge detector works very well in such situations, and does almost as well in purely Gaussian noise as the edge operators which are optimal in that situation.

Figures 4.1 and 4.2 show an example in which a stack filter based edge detector was used to detect the edges in an image corrupted by a combination of Gaussian noise of standard deviation 25 and salt and pepper noise. Figure 4.2 is the output edge map, in which it can be seen that there are few false edges and good detection and localization of true edges.

5. References

[BHM] A.C. Bovik, T.S. Huang, and D.C. Munson, Jr., "The effect of median filtering on edge estimation and detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 2, pp. 181-194, March 1987.

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

[CoL] E.J. Coyle and J.-H. Lin, "Stack filters and the mean absolute error criterion," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-36, no. 8, pp. 1244-1254, August 1988.

[FCG1] J.P. Fitch, E.J. Coyle and N.C. Gallagher, Jr., "Median filtering by threshold decomposition," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, no. 6, pp. 1183-1188, December 1984.

[FCG2] ____, "Threshold decomposition of multidimensional ranked order operations," *IEEE Trans. on Circuits and Systems*, vol. CAS-32, no. 5, pp. 445-450, May 1985.

[GaC] M. Gabbouj and E.J. Coyle, "Minimum mean absolute error stack filtering with structural constraints and goals," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-38, no. 6, pp. 955-968, June 1990.

[GaW] N.C. Gallagher, Jr., and G.L. Wise, "A theoretical analysis of the properties of median filters," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-29, no. 6, pp. 1136-1141, December 1981.

[GCG] M. Gabbouj, E.J. Coyle, and N.C. Gallagher, "An Overview of Median and Stack Filtering," *Circuits, Systems, and Signal Processing*, Special Issue on Median and Morphological Filters, to appear.

[Gil] E.N. Gilbert, "Lattice theoretic properties of frontal switching functions," *J. Math. Physics*, vol. 33, April 1954.

[Jus] B.I. Justusson, "Median filtering: statistical properties," *Two-Dimensional Digital Signal Processing II: Transforms and Median filters*, Chap. 4, vol 42, *Topics in Applied Physics*, T.S. Huang, Ed., New York: Springer Verlag, 1981.

[LSC] J.-H. Lin, T.M. Sellke, and E.J. Coyle, "Adaptive stack filtering under the mean absolute error criterion," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-38, no. 6, pp. 938-954, June 1990.

[MaS] P.A. Maragos and R.W. Schafer, "Morphological filters - Part I: their set theoretic analysis and relations to linear shift invariant filters,", and *"Morphological filters - Part II: their relations to median, order statistic and stack filters," IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 8, pp. 1153-1184, August 1987.

[NaR] Y. Nakagawa and A. Rosenfeld, "A note on the use of local min and max operations in digital picture processing," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. SMC-8, no. 8, pp. 632-635, August 1978.

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

[Tuk] J.W. Tukey, *Exploratory Data Analysis*, Addison-Wesly: Reading, MA, 1977.

[Tya] S.G. Tyan, "Median filtering: deterministic properties," *Two-Dimensional Digital Signal Processing II: Transforms and Median filters*, Chap. 5, vol 42, *Topics in Applied Physics*, T.S. Huang, ed. New York: Springer-Verlag, 1981.

[WCG] P.D. Wendt, E.J. Coyle, and N.C. Gallagher, Jr., "Stack filters," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 4, pp. 898-911, August 1986.

[YBD] J. Yoo, C.A. Bouman, E.J. Delp, and E.J. Coyle, "Intensity Edge Detection with Stack Filters," *Proc. of SPIE 1991 Symposium on Electronic Imaging*, San Jose, CA, Feb. 1991.

Fig. 4.1: Aerial image with Gaussian and impulsive noise.

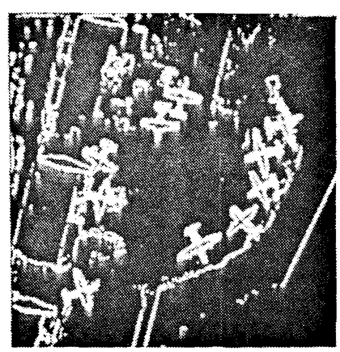

Fig. 4.2: Output of stack filter based edge detector.