RANK-ORDER FILTERS AND BAYES POSTERIOR DECISION

Bing Zeng, Moncef Gabbouj, and Yrjö Neuvo

Signal Processing Laboratory
Tampere University of Technology
P.O. Box 527, SF-33101 Tampere
Finland

Abstract. This paper gives the optimal stack filtering theory under the mean absolute error (MAE) criterion a completely new meaning in terms of the *a posteriori* Bayes minimum-cost decision. It is shown that under certain conditions this always leads to a rank-order filter (ROF) as the best filter in the minimum MAE sense. It is further shown that for a mostly practical case, the solution becomes the median filter.

The ROFs produced by this approach are subjected to a sensitivity analysis to quantify their dependency upon the cost coefficients. Several design examples will be provided.

1. INTRODUCTION

Rank-order filters (ROFs) include the well-known median filter and form a subclass of stack filters [1]-[3]. ROFs can be defined both in the multi-level domain and in the binary domain. In the multi-level domain, the r-th ROF of window width b, denoted as $R_{b,r}(\cdot)$, is defined as

$$Y(n) = R_{b,r}(\vec{X}(n)) = \text{the } r\text{th largest sample in } \vec{X}(n) \quad (1)$$

for $r = 0, 1, \cdots, b, b+1$, where $\vec{X}(n)$ (n is a time index) is a window vector which contains b samples successive in the time domain: $\vec{X}(n) = \{X(n-N_1), \cdots, X(n), \cdots, X(n+N_2)\}$ where $N_1 + N_2 + 1 = b$. The following are several special cases: $r = 0$ corresponds to the 1 filter, $r = 1$ is the max filter, $r = b$ is the min filter, $r = b+1$ corresponds to the 0 filter, and $r = (b+1)/2$ (for odd b) gives the median filter.

Since ROFs are stack filters, they can also be defined through threshold decomposition and positive Boolean functions [1], [2]. Actually, for the r-th ROF of window width b, the number of min-terms included in the corresponding Boolean function (in the MSP form) is exactly $\binom{b}{r}$ (a binomial number) for $r = 1, 2, \cdots, b$. Fig. 1 illustrates how a ROF $R_{3,2}$ (a window width three median filter) can be realized through (i) threshold decomposing the multi-level input signal to binary signals, (ii) filtering the binary signals using the Boolean function $x_1x_2 + x_2x_3 + x_3x_1$ which corresponds to the ROF, and (iii) adding the binary outputs on all threshold levels to obtain the multi-level output. Such an implementation has proven to be attractive to VLSI implementation [8].

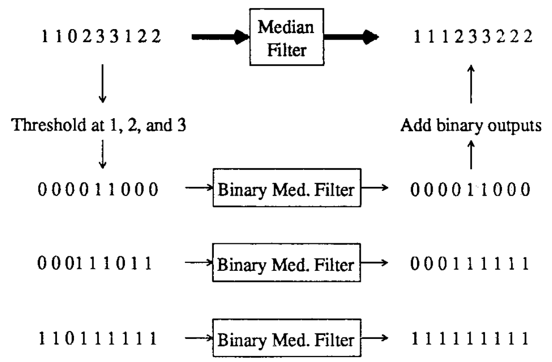

Fig. 1. An example showing how a ROF (3-point median filter) can be implemented through threshold decomposition and Boolean filtering.

In this paper, we consider how to design a ROF which is optimal under the MAE criterion. We show that such a problem is equivalent to the *a posteriori* Bayes decision with consistency constraints and the optimal solution can be obtained without resorting to the use of a linear program (LP). We further show that for a special case (in fact, the most practical case), the optimal solution becomes the median filter.

The paper is organized as follows. Section 2 reviews the theory of optimal stack filtering under the MAE criterion (as ROFs are also stack filters). In Section 3, we interpret the equivalence between this theory and the classical *a priori* and *a posteriori* Bayes decisions. The *a priori* problem has been solved in [6]. Here, we consider the *a posteriori* problem whose solution is shown in Section 4 to be a ROF with an appropriately chosen order. In Section 5, the ROFs produced are subjected to a sensitivity analysis in order to quantify their dependency upon the cost coefficients. Several design examples are provided in Section 6, and the conclusion of this paper is drawn in Section 7.

2. MINIMUM MAE STACK FILTERING

The optimal stack filtering theory under the MAE criterion [4] deals with an optimization problem stated as follows:

$$\begin{aligned}\text{minimize}: & \ C(S_f) = E\{|D(n) - S_f(\vec{X}(n))|\} \\ \text{subject to}: & \ \text{the stacking constraints}\end{aligned}$$

where $\vec{X}(n)$ denotes the received window process (supposed to be stationary), $D(n)$ is the desired signal, and S_f represents a window width b stack filter.

Let us assume that $D(n)$ and all elements of $\vec{X}(n)$ take values in $\{0, 1, \cdots, M\}$. Using the threshold decomposition technique, the above optimization can be reduced to (see [4]-[7])

Optimization 1:

minimize :

$$C(S_f) = \sum_{j=1}^{2^b} \Big[P_f(0|\vec{w}_j) E_j(0) + P_f(1|\vec{w}_j) E_j(1) \Big]$$

subject to :

$P_f(0|\vec{w}_j), P_f(1|\vec{w}_j) = 0$ or 1

$P_f(0|\vec{w}_j) + P_f(1|\vec{w}_j) = 1$

$P_f(1|\vec{w}_i) \geq P_f(1|\vec{w}_j)$ whenever $\vec{w}_i \geq \vec{w}_j$,

for all $\vec{w}_i, \vec{w}_j \in \{0,1\}^b$

where

$$E_j(0) = \sum_{l=1}^{M} C_l(1, 0, \vec{w}_j) \pi_l(1|\vec{w}_j) \pi(\vec{w}_j) \tag{2}$$

and

$$E_j(1) = \sum_{l=1}^{M} C_l(0, 1, \vec{w}_j) \pi_l(0|\vec{w}_j) \pi(\vec{w}_j). \tag{3}$$

Various notations used in the above optimization and equations are explained as follows. $\vec{w}_j$ denotes a binary state of length b. $P_f(0|\vec{w}_j)$ or $P_f(1|\vec{w}_j)$ is the output of the Boolean function $f(\cdot)$ for $\vec{w}_j$. $C_l(1, 0, \vec{w}_j)$ and $C_l(0, 1, \vec{w}_j)$ are the cost coefficients associated with wrong decisions for $\vec{w}_j$ on level l whose values can be set by designers. (In this paper, we assume that the cost coefficients for correct decisions are zero.) $\pi_l(\vec{w}_j)$ is the limiting probability of the event that $\vec{w}_j$ is observed on level l, and $\pi_l(0|\vec{w}_j)$ or $\pi_l(1|\vec{w}_j) = 1 - \pi_l(0|\vec{w}_j)$ is some conditional probability which has two different definitions, leading to the *a priori* and the *a posteriori* interpretations, respectively. This will be discussed in the following section.

3. BAYES DECISION INTERPRETATION OF OPTIMAL STACK FILTERING

Optimization 1 can be interpreted as a Bayes decision. This way, we see that $C(S_f)$ is the cost function of the decision where the cost coefficients for correct decision are zero and those for wrong decisions are $C_l(1, 0, \vec{w}_j)$ and $C_l(0, 1, \vec{w}_j)$ when $\vec{w}_j$ is observed on level l. Since a stack filter upholds the stacking property in its outputs, the equivalent Bayes decision must be consistent from one threshold level to another. For instance, the decision cannot be such that the signal is 0 on level 1, while deciding the signal is 1 on level 2.

Now, let us consider the definition of $\pi_l(0|\vec{w}_j)$. The *a priori* approach defines that $\pi_l(0|\vec{w}_j)$ is the probability of the event that the *desired signal* (i.e., $D(n)$) is less than l when $\vec{w}_j$ is observed in the *received window process* (i.e., $\vec{X}(n)$) threshold decomposed on level l. This leads to the design of optimal stack filters and generalized stack filters under the MAE criterion first solved by Coyle *et al.* using linear program techniques [4], [5], and recently simplified by Zeng *et al.* [6], [7]. In this paper, we introduce another definition of $\pi_l(0|\vec{w}_j)$ which provides a posterior interpretation. Specifically, we define

$$\begin{aligned}\pi_l(0|\vec{w}_j) =& \text{Prob}\{\text{filtered output value on level } l \text{ is} \\ & 0 \mid \vec{w}_j \text{ is observed}\} \\ =& \frac{N_b^{(0)}(\vec{w}_j)}{N_b}\end{aligned} \tag{4}$$

where N_b is the total number of positive Boolean functions with b variables and $N_b^{(0)}(\vec{w}_j)$ is the number of positive Boolean functions of b variables which produce a 0 output for the binary state $\vec{w}_j$ on level l. Similarly, $\pi_l(1|\vec{w}_j)$ is defined as

$$\pi_l(1|\vec{w}_j) = \frac{N_b^{(1)}(\vec{w}_j)}{N_b}. \tag{5}$$

It is easy to show that $\pi_l(0|\vec{w}_j) + \pi_l(1|\vec{w}_j) = 1$ for all $\vec{w}_j$'s.

It should be pointed out that the posterior meaning defined above is different from the standard one where the decision is based on the probability model of the outputs. However, the definition of (4) does give the *a posteriori* interpretation because it is based on the observations after the filtering pass. With this definition, the previously formed Bayes decision thus becomes the *a posteriori* decision.

4. SOLUTION OF THE A POSTERIORI BAYES DECISION

By Eq. (4), it is easy to see that $\pi_j(0|\vec{w}_j)$ is independent of level l, i.e.,

$$\pi_l(0|\vec{w}_j) = \pi_{l+1}(0|\vec{w}_j) \qquad \forall \vec{w}_j. \tag{6}$$

With Eqs. (4) and (5), the optimization procedure is simplified greatly, as will be shown in the following.

Lemma 1: Probabilities $\pi_l(0|\vec{w}_j)$'s possess the stacking property, i.e.,

$$\pi_l(0|\vec{w}_i) \leq \pi_l(0|\vec{w}_j) \quad \text{whenever} \quad \vec{w}_j \leq \vec{w}_i. \tag{7}$$

Proof: Let us show that $\pi_l(1|\vec{w}_i) \geq \pi_l(1|\vec{w}_j)$ which is equivalent to Inequality (7). Suppose that we find a positive Boolean function $f(\cdot)$ such that $f(\vec{w}_j) = 1$. Since $\vec{w}_i \geq \vec{w}_j$, $f(\vec{w}_i) = 1$. This implies that $N_b^{(1)}(\vec{w}_j) \leq$

$N_b^{(1)}(\vec{w}_i)$. Using Eq. (5), we thus proved that $\pi_l(1|\vec{w}_i) \geq \pi_l(1|\vec{w}_j)$.

Lemma 2: Let $\vec{w}_i$ and $\vec{w}_j$ denote two binary states on threshold level l. If $\vec{w}_i \cdot \vec{1}^t = \vec{w}_j \cdot \vec{1}^t$, then

$$\pi_l(1|\vec{w}_i) = \pi_l(1|\vec{w}_j). \tag{8}$$

Proof: Since $\vec{w}_i \cdot \vec{1}^t = \vec{w}_j \cdot \vec{1}^t$, both states contain the same number of 1's. Assume, without loss of generality, that $\vec{w}_i$ contains K 1's located at bit positions $n_1, n_2, \cdots, n_K$ and $\vec{w}_j$ contains K 1's located at bit positions $m_1, m_2, \cdots, m_K$. We must show that the number of positive Boolean functions which produce a 1 for $\vec{w}_i$ is equal to the number of positive Boolean functions which produce a 1 for $\vec{w}_j$. Define the following one-to-one mapping $\Psi : \Psi(n_k) = m_k \; \forall k = 1, 2, \cdots, K$. Now, any positive Boolean function (in its MSP form) that produces a 1 for $\vec{w}_i$ must contain at least one term whose literals are some or all of the bits n_k's. Using the one-to-one mapping Ψ, we obtain another positive Boolean function which produces a 1 for $\vec{w}_j$. Therefore, for each positive Boolean function that produces a 1 for $\vec{w}_i$, there exists a positive Boolean function that produces a 1 for $\vec{w}_j$.

Lemma 3: Given two binary states $\vec{w}_i$ and $\vec{w}_j$. If they meet the condition $\vec{w}_i \cdot \vec{1}^t + \vec{w}_j \cdot \vec{1}^t = b$, then

$$\pi_l(0|\vec{w}_i) = \pi_l(1|\vec{w}_j). \tag{9}$$

Proof: First, let us show that $\pi_l(0|\vec{w}_i) = \pi_l(1|\vec{w}_i^c)$ where $\vec{w}_i^c = \vec{1} - \vec{w}_i$ (i.e., the 0 bits of $\vec{w}_j$ become 1's and vice versa). This is equivalent to showing that $\pi_l(1|\vec{w}_i) + \pi_l(1|\vec{w}_i^c) = 1$. Using the definition given by (5), this equation becomes

$$N_b^{(1)}(\vec{w}_i) + N_b^{(1)}(\vec{w}_i^c) = N_b. \tag{10}$$

However, from the definition of $N_b^{(1)}(\vec{w}_i)$, it is easy to show that (10) is true for any $\vec{w}_i$. Therefore, we proved that $\pi_l(0|\vec{w}_i) = \pi_l(1|\vec{w}_i^c)$. Next, it is easy to see that $\vec{w}_j$ and $\vec{w}_i^c$ have the same number of ones, so we can apply the result of Lemma 2. This completes the proof.

Combining Lemma 1 and Lemma 3, we can easily show the following lemma.

Lemma 4: Given a binary states $\vec{w}_j$. (1) For odd b, if $\vec{w}_j$ contains no more (no less) than $(b-1)/2$ ($(b+1)/2$) ones, then $\pi(1|\vec{w}_j) < (>) 0.5$. (2) For even b, if $\vec{w}_j$ contains less (more) than $b/2$ ones, then $\pi(1|\vec{w}_j) < (>) 0.5$. Moreover, $\pi(1|\vec{w}_j) = \pi(0|\vec{w}_j) = 0.5$ if $\vec{w}_j$ contains exactly $b/2$ ones.

With these lemmas and assuming that the same cost coefficients $C(1,0)$ and $C(0,1)$ are used for all binary states on all threshold levels, we present an important theorem concerning the solution of the *a posteriori* Bayes minimum-cost decision as follows.

Theorem 1: The solution of the *a posteriori* Bayes minimum-cost decision is a ROF which can be obtained without solving an LP.

Proof: As the conditional probabilities $\pi_l(0|\vec{w}_j)$'s and $\pi_l(1|\vec{w}_j)$'s are independent of l, the expected costs $E_j(0)$ and $E_j(1)$ for binary state $\vec{w}_j$ reduce to

$$E_j(0) = C(1,0)\pi(1|\vec{w}_j) \sum_{l=1}^{M} \pi_l(\vec{w}_j) \tag{11}$$

and

$$E_j(1) = C(0,1)\pi(0|\vec{w}_j) \sum_{l=1}^{M} \pi_l(\vec{w}_j). \tag{12}$$

It is easy to check that these costs satisfy the stacking property defined in [6] (i.e., if $E_j(0) < E_j(1)$ for $\vec{w}_j$, then the same inequality must hold for any binary state stacking on top of $\vec{w}_j$). Therefore, the optimal solution yielding the minimum MAE can be obtained by using the suboptimal routine proposed in [6] which does not require an LP and whose computational requirements are merely comparisons between the costs. If $E_j(0) < E_j(1)$, we decide that $P_f(0|\vec{w}_j) = 1$; otherwise $P_f(1|\vec{w}_j) = 1$. On the other hand, according to Lemma 2, we know that all $\pi(0|\vec{w}_j)$'s or $\pi(1|\vec{w}_j)$'s are equal for the binary states having an equal number of ones. Therefore, the decision outputs for these binary states must also be the same. Obviously, such a solution is a ROF.

It should be pointed out that from the above proof, one can see that the solution actually does not depend on the priori probability models. This is because $C(0,1)$, $C(1,0)$, and $\pi_l(0|\vec{w}_j)$ are all independent of the probability models.

Finally, an important consequence of Theorem 1 is presented for the case when $C(0,1)$ and $C(1,0)$ are equal. Obviously, such a case is very practical.

Consequence 1: Suppose that b is odd. If the cost coefficients $C(0,1)$ and $C(1,0)$ are equal, then the solution of the *a posteriori* Bayes minimum-cost decision is the median filter.

Proof: When $C(1,0)$ and $C(0,1)$ are equal, we need only check if $\pi(1|\vec{w}_j) < \pi(0|\vec{w}_j)$ in order to decide the output value for $\vec{w}_j$. This, however, is equivalent to checking if $\pi(1|\vec{w}_j) < 0.5$. If this is true, then the decision output for $\vec{w}_j$ is 0; otherwise the output is 1. But, based on Lemma 4, it is easy to see that the decision output is zero if $\vec{w}_j \cdot \vec{1}^t < (b+1)/2$ and will become one as long as $\vec{w}_j \cdot \vec{1}^t \geq (b+1)/2$. Therefore, the solution is the median filter.

5. SENSITIVITY ANALYSIS

In the optimization procedure discussed earlier, it is possible that a small, and sometimes, even a large variation in the cost coefficients does not change the solution. In this section we attempt to quantify these variations. It is obvious that the larger these acceptable variations are, the more robust the solution of the optimization procedure is. The following theorem gives the range in which the cost coefficients $C(0,1)$ and $C(1,0)$ are allowed to vary while the optimal ROF remains unchanged.

Theorem 2: The rth ROF is optimal under the MAE criterion iff

$$\pi(1|\vec{w}_i) < \frac{C(0,1)}{C(0,1)+C(1,0)} < \pi(1|\vec{w}_j) \qquad (13)$$

for any $\vec{w}_i \in \Gamma_{r-1}$ and $\vec{w}_j \in \Gamma_r$ where $\Gamma_q = \{\vec{w}_m : \vec{w}_m \cdot \vec{1}^t = q\}$.

Proof: To prove necessity, suppose that, for some cost coefficients $C(0,1)$ and $C(1,0)$, the optimal ROF is the rth ROF. Then, for any $\vec{w}_i \in \Gamma_{r-1}$, according to the optimization procedure, we have

$$C(1,0)\pi(1|\vec{w}_i) < C(0,1)\pi(0|\vec{w}_i)$$

which is equivalent to

$$\pi(1|\vec{w}_i) < \frac{C(0,1)}{C(0,1)+C(1,0)}.$$

Similarly, we can show that for any $\vec{w}_j \in \Gamma_r$, we have

$$\pi(1|\vec{w}_j) > \frac{C(0,1)}{C(0,1)+C(1,0)}.$$

These two results imply that (13) holds.

To prove sufficiency, assume that $C(0,1)$ and $C(1,0)$ vary, but (13) still holds. According to the optimization procedure, (13) implies that the Boolean function representing the ROF will produce a 0 output for any $\vec{w}_i \in \Gamma_{r-1}$ and a 1 output for any $\vec{w}_j \in \Gamma_r$. Extending this result to the multi-level case, one can see that the solution is still the rth ROF.

6. DESIGN EXAMPLES

From Definitions (4) and (5), we can easily see that in order to design optimal ROFs, one must first find the number of positive Boolean functions for a given window width and also the numbers of subsets of these functions which produce a 1 output for some specific binary states. This task is, however, not such an easy one for relatively large window widths [9]. Here, we only consider cases $b = 3$ and $b = 4$. Tables 1 and 2 present the designed ROFs which yield minimum MAE for $b = 3$ and $b = 4$, respectively. Note that all filters given in these tables are independent of the prior probability model of the signal and noise processes and therefore might be used directly in certain practical cases.

Table 1: Window Width Three Optimal ROFs ($C(0,1) = 1.0$)

$C(1,0)$	Filters
0-1/19	$r=4$
1/19-3/7	$r=3$
3/7-7/3	$r=2$
7/3-19	$r=1$
>19	$r=0$

Table 2: Window Width Four Optimal ROFs ($C(0,1) = 1.0$)

$C(1,0)$	Filters
0-1/167	$r=5$
1/167-5/37	$r=4$
5/37-1	$r=3$
1-37/5	$r=2$
37/5-167	$r=1$
>167	$r=0$

7. CONCLUSION

In the present paper, optimal rank-order filtering under the MAE criterion was shown to be equivalent to the *a posteriori* Bayes decision. It was also shown that finding the minimum MAE ROF does not require an LP, thus dramatically reducing the complexity of the algorithm presented in [3]. Furthermore, the median filter was shown to be the optimal solution (in the minimum MAE and the *a posteriori* sense) for a very practical case. The robustness of the designed ROFs w.r.t. the cost coefficients was analyzed, which, supported by the independence of optimal solution on the prior statistics of the signal and noise processes, suggests the potential of ROFs in practical applications.

References

[1] J. P. Fitch, E. J. Coyle, and N. C. Gallagher, Jr., "Median filtering by threshold decomposition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, pp. 1183-1188, Dec. 1984.

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

[3] E. J. Coyle "Rank order operators and the mean absolute error criterion," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-36, pp. 63-76, Jan. 1988.

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

[5] J.-H. Lin and E. J. Coyle, "Minimum mean absolute error estimation over the class of generalized stack filters," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. ASSP-38, pp. 663-678, Apr. 1990.

[6] B. Zeng, M. Gabbouj, and Y. Neuvo, "Optimal stack filtering and classical Bayes decision," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Toronto, May 1991.

[7] B. Zeng, M. Gabbouj, and Y. Neuvo, "Design of optimal generalized stack filters under the mean absolute error criterion," *Proc. IEEE Workshop Visual Signal Process. & Communications.*, Taiwan, June 1991.

[8] K. Chen, "Realizations of a class of nonlinear filters using a bit-serial approach," *Proc. IEEE Int. Symp. Circuits Systems*, pp. 1749-1752, Helsinki, June 1988.

[9] M. Gabbouj, "Optimal stack filter examples and positive Boolean functions," M.S. Thesis, School of Elec. Eng., Purdue Univ., West Lafayette, IN., Dec. 1986.