OPTIMIZATION OF STACK FILTERS USING SAMPLE SELECTION PROBABILITIES

Ilya Shmulevich, Vladimir Melnik, Karen Egiazarian

Signal Processing Laboratory, Tampere University of Technology, Tampere, Finland
E-mail: ilya@cs.tut.fi, melnik@cs.tut.fi, karen@cs.tut.fi

ABSTRACT

We propose an optimization procedure of stack filters, which takes into consideration the filter's sample selection probabilitites. A statistical optimization of stack filters can result in a class of stack filters all of which are statistically equivalent. Such a situation arises in cases of non-symmetric noise distributions or in the presence of constraints. Among the set of equivalent stack filters, our method constructs a statistically optimal stack filter, whose sample selection probabilities are concentrated in the center of its window. This leads to improvement of detail preservation.

1. INTRODUCTION

Stack filters constitute an important class of nonlinear filters based on monotone Boolean functions [1]. A design method for stack filters based on minimization of the mean absolute error was demonstrated in [2]. Statistical properties of stack filters have been studied in terms of output distributions and moments for i.i.d. input signals [3],[4],[5]. Consequently, it becomes possible to optimize stack filters in the mean square sense [6]. In other words, the knowledge of the input distribution allows one to find a stack filter or a set of stack filters all of which minimize the output variance. Such an optimization paradigm is advantageous for several reasons.

Firstly, it is not necessary to perform a search over the set of stack filters, the cardinality of which grows very quickly [7]. Secondly, the optimization procedure does not suffer from high computational complexity, since usually a linear programming or quadratic programming problem needs to be solved, in which the dimensionality of the search space is linearly related to the window width of the stack filter. Finally, various constraints can be easily incorporated into the optimization framework. For example, structural constraints as well as statistical constraints, such as unbiasedness (in the mean sense), can be easily included. Moreover, robustness constraints can also be easily expressed using rank selection probabilities [8].

Rank selection probabilities (RSP) and sample selection probabilities (SSP) are probabilities that the output equals a sample with a certain rank and certain time-index in the filter window, respectively. The output distribution of a stack filter can be expressed in terms of its RSPs. On the other hand, SSPs give us information about the temporal behavior of stack filters [8]. This information is important for examining the detail preservation properties of stack filters. Efficient spectral algorithms exist for the computation of the selection probabilities of stack filters [9].

It is well known that the median filter is optimal among the set of all stack filters for minimizing the output variance when the input distribution is symmetric and no constraints are imposed [6]. However, in many situations, a symmetric noise distribution cannot be assumed [10]. Moreover, we may wish to require that the mean of the output be equal to some particular value, constituting a constraint. In such cases, the optimization results not in one stack filter, but rather in a set of stack filters all of which are statistically equivalent. That is, they are all optimal in the sense of minimizing the output variance and have identical RSPs. Therefore, the set of stack filters can be decomposed into equivalence classes, where two stack filters are equivalent if they have the same RSPs and consequently, output distributions.

More generally, optimization of stack filters can result in an entire class of stack filters, all of which are statistically optimal under the chosen criterion. Nevertheless, from a given class of filters, we may wish to select one filter which is best in some other sense. In this paper, we propose a method of finding the stack filter which preserves signal details better than all other filters in its class. This notion is captured by the filter's SSPs.

We begin by noting that detail preservation depends on how much sample selection probability is concentrated in the center of the filter window. As an extreme example, the identity filter, which has the best possible detail preservation, has a probability 1 of selecting the center sample. On the other hand, the median filter has equal SSPs for each of its samples in the window. Therefore, among the class of statistically equivalent optimal stack filters, we wish to select one filter which has the highest concentration of sample selection probability in the center of the window. Despite efficient algorithms for computing SSPs for stack filters, determining the SSPs for each stack filter and then selecting the best one has a major drawback: there may be an immense number of stack filters in a given equivalence class. This notwithstanding, such an approach still requires one to generate each stack filter and then compute its SSPs.

We propose a direct approach which *constructs* a filter with the highest concentration of sample selection probability in the center of the window from the statistically optimal class of filters. In Section 2, we give a brief review of statistical stack filter optimization as well as present some notation and definitions. Section 3 discusses the proposed SSP based optimization algorithm.

2. BACKGROUND

Let E^n represent the n-cube. The k^{th} level $(0 \leq k \leq n)$ of the cube contains only those vectors with exactly k components equal to 1 or, equivalently, with Hamming weight $w(x) = k$. The set of all vectors on the k^{th} level will be denoted by $E^{n,k}$. For any two vectors $\alpha, \beta \in E^n$, $\alpha \preceq \beta$ means $\alpha_i \leq \beta_i$ for $1 \leq i \leq n$. A Boolean function $f : E^n \to E^1$ is called *monotone* (positive) if for any two primes α and β such that $\alpha \preceq \beta$, we have $f(\alpha) \leq f(\beta)$. Each stack filters corresponds to a monotone Boolean function [1]. A continuous stack filter is obtained by replacing conjunction and disjunction operations by min and max, respectively. Thus, it operates on the real-valued domain. For example, the Boolean function $f(x_1, x_2, x_3) = x_1 x_2 + x_2 x_3$ (where $\cdot$ means conjunction and $+$ means disjunction) corresponds to

$$S_f(X_1, X_2, X_3) = \max\{\min\{X_1, X_2\}, \min\{X_2, X_3\}\}$$

where X_1, X_2, X_3 are real-valued variables. Suppose that the input variables of some stack filter $S_f(\cdot)$ are i.i.d. random variables with distribution

$$F(t) = \Pr\{X_i \leq t\}$$

Then, it is well known [5] that the output

$$Y = S_f(X_1, \ldots, X_n)$$

of the stack filter has output distribution

$$\Psi(t) = \sum_{i=0}^{n-1} A_i (1 - F(t))^i \cdot F(t)^{n-i} \tag{1}$$

where

$$A_i = \left|\left\{x \in E^{n,i} : f(x) = 0\right\}\right|$$

This sets the stage for the mean square optimization of stack filters. The variance $\mu_2 = E\left\{(Y - E\{Y\})^2\right\}$ of the output Y of the stack filter can be written as [6]

$$\mu_2 = \sum_{i=0}^{n-1} A_i M(F, 2, n, i) - \left(\sum_{i=0}^{n-1} A_i M(F, 1, n, i)\right)^2 \tag{2}$$

where

$$M(F, k, n, i) = \int_{-\infty}^{\infty} x^k \frac{d}{dx}\left((1 - F(t))^i \cdot F(t)^{n-i}\right)$$

So, for a given noise distribution, the goal becomes to find parameters A_i such that the objective function in (2) is minimized. For non-symmetric (around zero) distributions, such an optimization results in a whole class of stack filters all of which are statistically equivalent, since they all possess the same parameters A_i. This paper is concerned with selecting one among that set that would result in best detail preservation.

Let $s_j = \Pr\{Y = X_j\}$ be the jth SSP and $s = [s_1, \ldots, s_n]$ be the vector of SSPs. In [9], it was shown that $s_j = d_j(1) - d_j(0)$, where

$$d_j(k) = \sum_{x \in f^{-1}(1) | x_j = k} \left[n \binom{n-1}{w(x) - k}\right]^{-1} \tag{3}$$

and $f^{-1}(k) = \{x \in E^n : f(x) = k\}$. The SSPs capture the temporal behavior of stack filters and are thus related to detail preservation capability. A higher concentration of probability in the center of the window leads to better detail preservation.

3. PROPOSED OPTIMIZATION METHODOLOGY USING SSP

In this section, let us focus on non-symmetric noise distributions, since it is well known [6] that in the class of stack filters, the optimal filter for zero-mean symmetric distributions is the median filter. Suppose that the statistical optimization algorithm, for a given distribution $F(t)$, results in parameters $A_0, A_1, \cdots, A_{n-1}$.

This implies that we must select $\binom{n}{m} - A_m$ vectors from sets $E^{n,m}$, $m = 0, \dots, n-1$. Suppose for the moment that we wish to maximize the SSP s_j. Since $s_j = d_j(1) - d_j(0)$, this corresponds to simultaneously maximizing $d_j(1)$ and minimizing $d_j(0)$. Let us rewrite equation (3) as

$$\begin{aligned} d_j(k) &= \sum_{m=0}^{n} \sum_{x \in f^{-1}(1) \cap E^{n,m} | x_j = k} \left[n \binom{n-1}{m-k} \right]^{-1} \\ &= \sum_{m=0}^{n} \left[n \binom{n-1}{m-k} \right]^{-1} \cdot \\ &\quad \cdot \sum_{x \in f^{-1}(1) \cap E^{n,m} | x_j = k} 1 \\ &= \sum_{m=0}^{n} \left[n \binom{n-1}{m-k} \right]^{-1} \cdot \\ &\quad \cdot \left| \left\{ x \in f^{-1}(1) \cap E^{n,m} : x_j = k \right\} \right| \end{aligned}$$

So, $d_j(1)$ can be maximized by making

$$\left| \left\{ x \in f^{-1}(1) \cap E^{n,m} : x_j = 1 \right\} \right|$$

as large as possible and $d_j(0)$ can be minimized by making

$$\left| \left\{ x \in f^{-1}(1) \cap E^{n,m} : x_j = 0 \right\} \right|$$

as small as possible. Note that,

$$\left| \left\{ x \in f^{-1}(1) \cap E^{n,m} : x_j = k \right\} \right| + A_m \leq \binom{n}{m} \quad (4)$$

necessarily holds. Furthermore, observe that

$$\begin{aligned} & \left\{ x \in f^{-1}(1) \cap E^{n,m} \right\} \\ = & \bigcup_{k \in \{0,1\}} \left\{ x \in f^{-1}(1) \cap E^{n,m} : x_j = k \right\} \end{aligned}$$

and

$$\bigcap_{k \in \{0,1\}} \left\{ x \in f^{-1}(1) \cap E^{n,m} : x_j = k \right\} = \emptyset$$

for any j, which imply that

$$\begin{aligned} & \left| \left\{ x \in f^{-1}(1) \cap E^{n,m} \right\} \right| & (5) \\ = & \sum_{k \in \{0,1\}} \left| \left\{ x \in f^{-1}(1) \cap E^{n,m} : x_j = k \right\} \right| \\ = & \binom{n}{m} - A_m & (6) \end{aligned}$$

Since A_m is a constant resulting from statistical optimization, it follows from equation (5) that maximizing $d_j(1)$ and minimizing $d_j(0)$ is, in fact, the same thing.

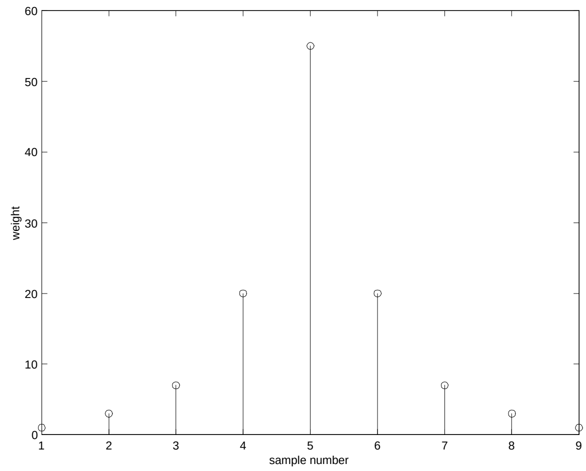

Figure 1: Weight vector for SSP optimization, $n = 9$.

For a given j and $k = 1$, the bound in (4) may not be attained. In that case, it becomes necessary to add other vectors $x \in E^{n,m}$ for which $x_j = 0$. Nevertheless, our goal is to maximize $d_j(1) - d_j(0)$ for j as close to $\frac{n+1}{2}$ as possible, or equivalently, to concentrate as much SSP in the center of the window as possible. This can be achieved by the following procedure.

We assign weights $v = [v_1, \dots, v_n]$ to each of the n coordinates of binary input vectors $x = [x_1, \dots, x_n]$, with maximal weight in the center and monotonically decreasing weights toward the ends of the window. Figure 1 shows a typical weight vector constructed from a bi-exponential function. However, any similar function can be used instead. The weight $v(x)$ of a binary vector x can then be defined as $x \cdot v^T$. Furthermore, we define the weight $v(\Omega)$ of a set Ω of vectors as $\sum_{x \in \Omega} v(x)$. We then select the required number of binary vectors from $E^{n,m}$, as dictated by the parameters A_m, with the maximum weight. In other words, we select a set $\Omega_m \subseteq E^{n,m}$, $|\Omega_m| = \binom{n}{m} - A_m$ so that $v(\Omega_m)$ is maximum. Even a straightforward sorting procedure of $v(x)$, $x \in E^{n,m}$ is an extremely efficient way to select the set Ω_m. Clearly, this approach produces those vectors $x \in E^{n,m}$ which contain more coordinates equal to 1 towards the center of the window. We now give an example of the proposed optimization algorithm, showing its effect on SSP and consequently, detail preservation.

3.1. Optimization Example

Suppose that the input noise density $g(t)$ is a mixture of two Gaussian densities $f_1(t)$ and $f_2(t)$. That is,

$$g(t) = p \cdot f_1(t) + (1-p) \cdot f_2(t)$$

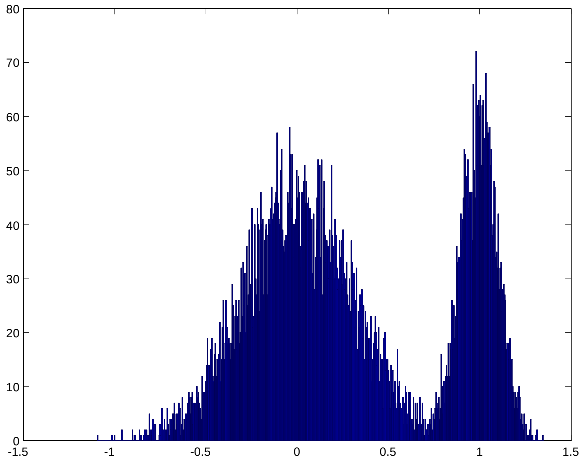

Figure 2: Histogram of the input noise

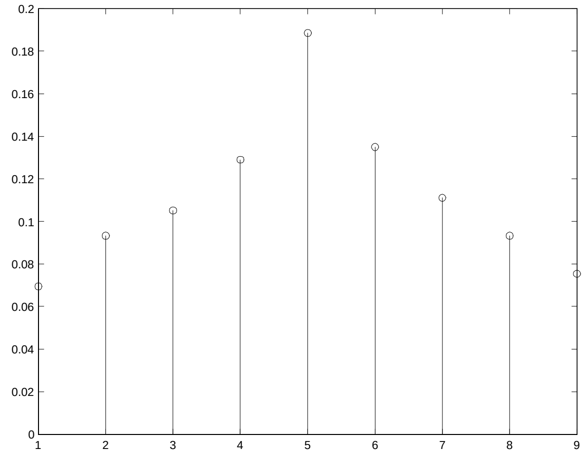

Figure 3: SSPs of the optimal filter constructed using the proposed algorithm

In this example, $p = 0.7$ and the mean and standard deviation of the two Gaussian densities are

$$\begin{aligned} \mu_1 &= 0 \text{ and } \sigma_1 = 0.3 \\ \mu_2 &= 1 \text{ and } \sigma_2 = 0.1 \end{aligned} \quad (7)$$

Such a model is often encountered when the noise originates from two separate sources. The second density, which has a non-zero mean and small variance, represents additive impulsive-type noise in addition to the usual additive zero-mean Gaussian noise modeled by the first density. Figure 2 illustrates a histogram produced from this mixture density.

Suppose that we wish to design an optimal stack filter which would minimize the output variance with the added constraint of the output being zero-mean. For our noise distribution, the set of optimal stack filters with window width $n = 9$ all of which minimize the output variance and whose output is zero-mean, is given by the vector of parameters

$$A = [1, 9, 36, 84, 126, 126, 45, 0, 0, 0] \quad (8)$$

According to equation (2), the output variance of an optimal stack filter should be $\mu_2 = 0.0445$. The rank selection probabilities of this filter are: $r_3 = 0.5357$ and $r_4 = 0.4643$. From this, we see that the filter prefers the third and fourth ranks, since besides minimizing variance, it must also maintain zero-mean output. From (8), we can see that 39 vectors must be selected from $E^{9,6}$, since $84 - 45 = 39$. By monotonicity, 27 of the 36 vectors on $E^{9,7}$ get selected. Therefore, we must select 9 more vectors on $E^{9,7}$, for a total of 48 vectors. We select these vectors according to the procedure described above so as to maximize detail preservation, using the weight vector shown in Figure 1. The resulting vector of SSPs is shown in Figure 3.

In order to demonstrate the advantages of the proposed algorithm, we considered two statistically equivalent optimal stack filters with the parameters given in (8). Thus, the two filters have identical RSPs and output variance. However, one stack filter, referred to as Filter A, was optimized using the proposed method and has SSPs as shown in Figure 3. The other stack filter, Filter B, was constructed without the use of the algorithm described herein. That is, the minimal vectors were selected randomly on each level of E^9. The SSPs of Filter B are shown in Figure 4. As can be seen

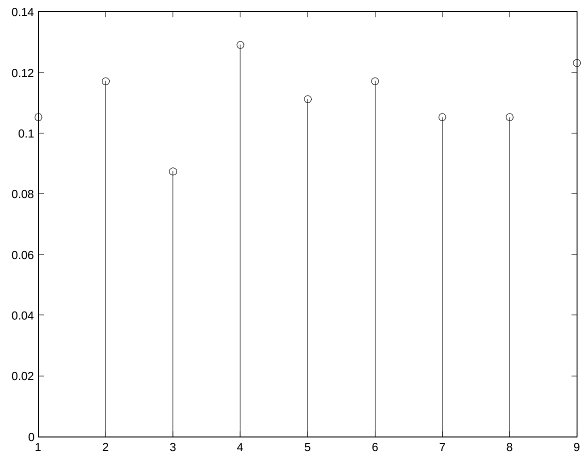

Figure 4: SSPs of a stack filter constructed without the use of the proposed algorithm

from this figure, SSP is not concentrated around the center of the window. Consequently, we should expect the detail preservation ability of Filter B to be inferior to the Filter A. To check this, we applied both filters to a noisy signal shown in Figure 5. The noise used was presented above with parameters given in (7). We

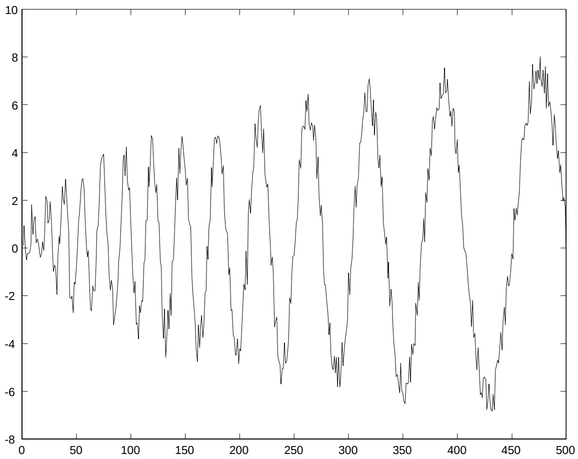

Figure 5: Test signal

compared the MSE of the two stack filters. The filter constructed using our method, which concentrates SSP in the center of the window produced an MSE of 0.1960. Filter B, however, produced a larger MSE equal to 0.2339. Since both filters were statistically equivalent, the smaller MSE of Filter A was due to its superior detail preservation ability.

4. CONCLUSION

We proposed an optimization procedure of stack filters, which takes into consideration the filter's sample selection probabilities. Among the set of equivalent stack filters, our method constructs a statistically optimal stack filter, whose sample selection probabilities are concentrated in the center of its window. As an example, we used a mixture of two noise sources with Gaussian distributions. We demonstrated that the constructed stack filter leads to detail preservation improvement. Our approach is direct in that it is constructive. That is, it does not require the computation of SSPs of each filter in a given equivalence class, and thus is extremely efficient computationally.

5. REFERENCES

[1] P. Wendt, E. Coyle, N. Gallagher, "Stack Filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 898-911, 1986

[2] E. Coyle, J. Lin, "Stack filters and the mean absolute error criterion," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 36, no. 4, pp. 1244-1254, 1988

[3] O. Yli-Harja, "Formula for the joint distribution of stack filters," *Signal Processing Letters*, vol. 1, pp. 129-130, 1994

[4] P. Kuosmanen, *Statistical analysis and optimization of stack filters*, Acta Polytechnica Scandinavica, Electrical Engineering Series, 77, Helsinki, 1994.

[5] S. Agaian, J. Astola, K. Egiazarian, *Binary Polynomial Transforms and Digital Filters*, Marcel Dekker, New York, 1995.

[6] J. Astola, P. Kuosmanen, *Fundamentals of Nonlinear Digital Filtering*, CRC Press, 1997.

[7] I. Shmulevich, T.M. Sellke, M. Gabbouj, E.J. Coyle, "Stack Filters and Free Distributive Lattices," *Proceedings of 1995 IEEE Workshop on Nonlinear Signal Processing*, Halkidiki, Greece, 1995.

[8] M. Prasad, Y. Lee, "Stack Filters and Selection Probabilities," *IEEE Trans. on Signal Processing*, vol. 42, no. 10, 1994.

[9] K. Egiazarian, P. Kuosmanen, J. Astola, "Boolean derivatives, weighted Chow parameters, and selection probabilitites of stack filters," *IEEE Trans. on Signal Processing,* vol. 44, no. 7, 1996.

[10] Lukin V.V., Melnik V.P., Chemerovsky V.I., Astola J.T., Kuosmanen P., Saarinen K.P. "Adaptive α-trimmed filters for speckle image processing," *Proceedings of IS&T/SPIE Symposium on Electronic Imaging: Science and Technology*, Vol. 3304, San Jose, California, USA, Jan, p. 209-219, 1998.