EFFICIENCY OF REAL-TIME GAUSSIAN TRANSIENT DETECTORS: COMPARING THE KARHUNEN-LOÈVE AND THE WAVELET DECOMPOSITIONS

Francisco M. Garcia and Isabel M. G. Lourtie

ISR -Instituto de Sistemas e Robótica, IST - Instituto Superior Técnico
Torre Norte, Av. Rovisco Pais, 1049-001 Lisboa, Portugal
E-mail : fmg@isr.ist.utl.pt

ABSTRACT

In general, finite-dimensional discrete-time representations of continuous-time Gaussian transients is not complete. Such representations typically lead to suboptimal detectors, where the compromise between computational complexity and processor performance requires optimization, specially when real-time processing is mandatory. This paper proposes a procedure for the optimization of the processor parameters, using the Bhattacharyya distance to evaluate the resemblance between the original continuous-time signal and its finite dimensional discrete representation. Two different decompositions are analyzed and compared, namely the Karhunen-Loève decomposition (KLD) and the discrete wavelet transform (DWT). It is shown that the DWT presents serious advantages when the signals to detect have a large number of important eigenvalues, which is often the case in some applications such as passive sonar.

1. INTRODUCTION

In recent years, passive detection has known an increasing interest in underwater acoustic applications. In particular, the need to detect small wideband transients, such as man-made metallic noises, mammal sounds or bubbles bursts arising from the sea floor in seismic regions [6] require the use of nonstationary stochastic models with a large number of relevant eigenvectors that increase the robustness of the processor. The classical solution to this problem is based on the Karhunen-Loève decomposition (KLD), where the observation process is decomposed under a small number of uncorrelated coefficients. This procedure has proven to be adequate when the signals to detect are stationary, and the corresponding KLD is the Fourier transform, approximated in the discrete-time domain by the FFT after a correct filtering and sampling procedure. The resulting processor, consisting mainly on a FFT decomposition stage and a log-likelihood test (LLT) is thus efficient since the FFT is computationally low-cost and the covariance matrix in the quadratic form of the LLT is diagonal. When the signals to detect are transients with a large number of eigenvectors, the signal KLD is no more the FFT and the decomposition step is performed by discrete-time internal products between the samples of the observation process and the signal covariance matrix eigenvectors. As the number of eigenvectors increase, the computational complexity (CC) of the decomposition step also increases. In real-time applications, the KLD leads, in many cases, to prohibitive costly algorithms.

Using Mallat's recursive algorithm for image decomposition in its one-dimensional form, the discrete wavelet transform (DWT) strongly reduces the decomposition stage CC of the processor. The LLT stage CC, however, increases since the resulting decomposition coefficients are in general correlated, even though a convenient choice of the mother wavelet and sampling interval lead to sparse coefficients covariance matrices.

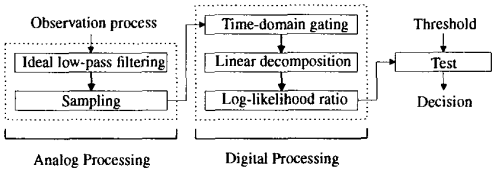

Figure 1: Processing scheme.

This paper proposes methods to find the parameters that optimize the performance of the processor given a fixed CC, for both cases where either the KLD or the DWT are used. It also presents an example showing that in some realistic situations, the DWT can present a huge efficiency gain comparing with the KLD.

2. PROBLEM FORMULATION

The detection problem is formulated as a simple binary test. The observation process $r(t)$ is defined as

$$r(t) = \begin{cases} s(t) + n(t), & \text{under hypothesis } H_1 \\ n(t), & \text{under hypothesis } H_0, \end{cases} \quad (1)$$

where $s(t)$ and $n(t)$ correspond to continuous-time zero mean Gaussian distributed processes with autocorrelation functions $k_s(t_1, t_2)$ and $\sigma^2\delta(t_1 - t_2)$, respectively, and $\delta(t)$ and σ^2 represent the Dirac delta and the proportionality coefficient of the noise variance. It is assumed that most of the energy of the signal $s(t)$ ($s(t) \in L^2(\mathbb{R})$) lies in compact support bands either in the time and frequency domains.

The processing scheme is presented in fig. 1. The first two blocks consist in filtering and sampling stages. The lowpass filter is assumed ideal with cutoff frequency ω_c equal to half the sampling frequency ω_s ($\omega_c = \omega_s/2$), this restriction being necessary to ensure that the filtering and sampling stages are equivalent to a linear decomposition. The samples are then truncated in the time-domain, and decomposed in a small number of coefficients with low cross-correlation. Afterwards, the log-likelihood ratio is computed and compared with a threshold. Clearly, the filtering, sampling, gating and linear decomposition operations reduce the

0-7803-6293-4/00/$10.00 ©2000 IEEE.

signal information represented by the coefficients covariance matrix. It is possible to reduce this negative effect by choosing an almost-complete signal representation (high sampling rates, large observation intervals and many decomposition coefficients) but the corresponding processor CC becomes untracktable, specially for real-time applications. It is thus necessary to determine the best tradeoff by choosing the parameters of the scheme in fig. 1 that optimize the processor performance, while maintaining the CC at a reasonable level.

For real-time applications, the rate at which the likelihood ratios are sequentially computed also influences the tradeoff CC/processor performance. In this scenario, it is assumed that a given process is observed (filtered and sampled) for a long time interval. However, only the coefficients corresponding to a comparative small sliding window (the size of the gating in fig. 1, i.e. the approximate length of the transient signal), are computed and the resulting likelihood test is evaluated. As new samples of the observation process arrive, the window is shifted and a new likelihood test is performed. Defining by N_t the shift interval of the sliding window, two consecutive likelihood tests are separated by N_t sampling intervals. Thus, at times $t - N_t$ and t, the sliding windows correspond respectively to the intervals $[t - N_a - N_t + 1; t - N_t]$ and $[t - N_a + 1; t]$, where N_a stands for the sliding window length. Clearly, the performance of the real-time processor is the best when $N_t = 1$, i.e., when the likelihood tests are computed at every sampling interval and thus reducing as much as possible the probability of missing a signal between two consecutive tests. However, the processor's CC is strongly reduced if N_t corresponds to several sampling intervals (the CC is proportional to $1/N_t$). Although this effect degrades the processor performance, quadratic processors are relatively robust to small arriving times shifts [3], even when the signals to detect are small transients, nonstationary in nature.

3. OPTIMIZATION PROBLEM

The optimization problem can be stated as follows: let $\{\gamma\}$ be the set of parameters that uniquely define a given processor, and let $\alpha(\{\gamma\})$ and $\beta(\{\gamma\})$ be measures of, respectively, the performance and the CC of the processor. We wish to determine the optimum set of parameters $\{\gamma_{\text{opt}}\}$ such that

$$\{\gamma_{\text{opt}}\} = \arg\max_{\{\gamma\}} \{\alpha(\{\gamma\}) | \beta(\{\gamma\}) = \beta_{\text{ref}}\}, \qquad (2)$$

where β_{ref} stands for the desired CC.

The CC $\beta(\{\gamma\})$ is represented by the number of multiplications per time unity executed by the processor. The expression of $\beta(\{\gamma\})$ depends on the linear decomposition used and on the likelihood ratio expression. Since these two terms differ substantially on the KLD and the DWT cases, the corresponding CCs are derived individually for each case in the next two sections.

The best measure to evaluate the processor's performance in detection problems is the probability of error (PE). However, the PE between two hypotheses does not have a closed form and is difficult to compute, specially in multidimensional optimization problems. An alternative way is to use the Bhattacharyya distance (BD) [2] to obtain an upper bound, $\epsilon_u(H_0, H_1)$, of the PE between two hypothesis H_0 and H_1. However, in the present case, one needs to measure the negative impact of the model mismatch in the detector performance. One important component that contributes to this mismatch is the shift error that results from the non-coincidence between the signal arriving times and LLT time instants computation. To deal with this situation, the BD is used instead to evaluate the resemblance between the covariance matrices of the true and approximated signals in hypothesis H_1. For simplicity, it is assumed that the random "shift error", corresponding to a diagonal shift on all the elements in the covariance matrix of the approximated signal, is maximum when the signal arrives at half the length of the sliding window, and the detector performance is evaluated through the BD in this worst case condition. The BD requires furthermore that i) a finite-dimensional discrete description of the complete and approximated signals is available, ii) the corresponding coefficients are expressed in the same subspace and iii) the covariance matrices are positive definite. Although a finite-dimensional description of a continuous-time process cannot in general be obtained, a very good approximation (obtained by observing the transient signal on a large observation window, using very high sampling rates and a large number of eigenvalues) is used as a "discrete reference signal". The approximated signal is computed by sequentially reducing the number of eigenvectors and the observation window length, and increasing both the sampling interval and the likelihood test interval (i.e., the diagonal shift between the approximated covariance matrix and the reference one). Since it is assumed that the lowpass filtering is ideal, the filtering and sampling processes correspond to a linear decomposition, as are the KLD and the DWT. Reducing the observation interval signifies only that some of the coefficients of the filtering and sampling decomposition are kept and the others are discarded. Thus, the approximated signal coefficients subspace is only a shifted linear subspace of the reference signal and it is possible to obtain a projection of the approximated signal covariance matrix in the subspace of the reference signal to satisfy the above referred point ii). Regarding point iii), the white noise present in the binary hypothesis problem ensures that the covariance matrices of the reference and approximated signals plus noise are positive definite. Finally, we evaluate the processor performance through

$$\alpha(\{\gamma\}) = \epsilon_u(H_r, H_a), \qquad (3)$$

where H_r and H_a stand for the hypotheses where the reference and approximated signals plus noise are present, the corresponding covariance matrices being $\boldsymbol{C}_{H_r}$ and $\boldsymbol{C}_{H_a r}$, this latter corresponding to the projection of the low order approximated covariance matrix $\boldsymbol{C}_{H_a}$ on the higher dimension reference signal subspace. When $\alpha(\{\gamma\})$ is close to 0.5, the resemblance between the approximated and reference signals is high and a negligible detector performance degradation can be expected.

4. KARHUNEN-LOÈVE OPTIMIZATION

When the decomposition used is the KLD, the set of parameters $\{\gamma\}$ that need to be optimized are i) N_s^{ar} - the relation between the approximated and reference sampling intervals ($N_s^{ar} \in \mathbb{R}$, $N_s^{ar} > 1$); ii) N_w^r - The size of the sliding window, or the gating in figure 1, expressed in number of reference signal samples. For simplicity, it is assumed that N_w^r is an integer multiple of N_s^{ar}; iii) M_a - The number of eigenvalues and eigenvectors considered in the KLD. It is important to note that these do not correspond to the largest original eigenvalues of the reference signal, but with the filtered and truncated ones ($M_a \in \mathbb{N}$); iv) N_t - The LLT is performed every N_t approximated signal sampling intervals ($N_t \in \mathbb{N}$).

As it was referred to before, the CC of a processor depends on i) the decomposition stage and ii) the likelihood ratio decomposi-

tion. When the KLD is used, the decomposition stage corresponds to M_a internal products between each of the eigenvectors and the observation process. Since the KLD coefficients are uncorrelated, the likelihood ratio is defined as

$$\ell = \sum_{i=1}^{M_a} r_i^2 k_i \qquad \text{with} \qquad k_i = \frac{\lambda_i}{\sigma_a^2(\lambda_i + \sigma_a^2)} \tag{4}$$

where r_i is the coefficient resulting from the decomposition of the sampled observation process under the i-th eigenvector, λ_i is the i-th eigenvalue and σ_a^2 is the noise spectral height after lowpass filtering ($\sigma_a^2 = \sigma^2/T_s^a$), T_s^a being the approximate signal sampling interval. Therefore, the number of multiplications per reference sampling interval is

$$\beta_{\text{KLD}}(N_s^{ar}, N_w^r, M_a, N_t) = \frac{M_a\left(\frac{N_w^r}{N_s^{ar}} + 2\right)}{N_t N_s^{ar}}. \tag{5}$$

Since a closed-form expression for the CC is available, it is possible to transform the optimization problem of (2) to an unrestricted optimization problem. For this purpose, using the restriction in (2) and solving (5) in order to N_s^{ar} (the only parameter that takes values in $\mathbb{R}$) we get

$$N_{s_{\text{ref}}}^{ar} = \frac{M_a + \sqrt{M_a^2 + N_t\beta_{\text{ref}} M_a N_w^r}}{N_t \beta_{\text{ref}}}, \tag{6}$$

and the new unrestricted optimization problem is rewritten as

$$[N_w^r, M_a, N_t]_{\text{opt}} = \arg \max_{N_w^r, M_a, N_t} \{\alpha(N_w^r, N_{s_{\text{ref}}}^{ar}, M_a, N_t)\}, \tag{7}$$

which is easily maximized using a three steps iterative algorithm: optimization is performed sequentially on each single parameter until convergence is achieved.

5. DISCRETE WAVELET TRANSFORM OPTIMIZATION

The DWT decomposes iteratively an initial discrete sequence c_n^0 under the subsequences d_k^j, $j = 1, \cdots, J$ and a lowpass residual sequence c_k^J. By Mallat's algorithm [1, 5]

$$c_k^j = \sum_n h(n-2k)c_n^{j-1}, \qquad d_k^j = \sum_n g(n-2k)c_n^{j-1}, \tag{8}$$

where $h(n)$ and $g(n)$ are, respectively, lowpass and highpass elements of a quadrature mirror filter (QMF) pair of compact support. The decompositions in (8) are equivalent to internal products of the original sequence c_n^0 with orthogonal filters $h_k^j(n)$ and $g_k^j(n)$. One important feature of the DWT consists in the translation property [3]. Letting $n_0 = l.2^J$, then a n_0 shift in the original sequence c^0 returns shifted coefficients, such that

$$\begin{aligned} \tilde{d}_k^j &= < c_{n-n_0}^0, g_k^j(n) > = < c_n^0, g_{k+l.2^{J-j}}^j(n) > = d_{k+l.2^{J-j}}^j \\ \tilde{c}_k^J &= < c_{n-n_0}^0, h_k^J(n) > = < c_n^0, h_{k+l}^J > = c_{k+l}^J. \end{aligned} \tag{9}$$

In real-time detection problems, one of the most attractive assets of the DWT consists in the fact that if the likelihood tests are carried out at an integer multiple of 2^J, it is only necessary to calculate the new coefficients corresponding to the non-overlapping zone of two consecutive sliding windows. Furthermore, the filters $h(n)$ and $g(n)$ are, in general, of small size and the decomposition CC of (8) is very low. The drawback consists on the fact that, usually, the coefficients covariance matrices are not diagonal, although, with a correct choice of the filters $h(n)$ and $g(n)$, sparsity can be achieved. The resulting likelihood ratio is thus

$$\ell = \boldsymbol{r}' \boldsymbol{M} \boldsymbol{r}, \qquad \boldsymbol{M} = \boldsymbol{C}_{H_0}^{-1} - \boldsymbol{C}_{H_1}^{-1}, \tag{10}$$

where $\boldsymbol{r}$ represents the DWT coefficients vector. In opposition to the KLD case, where the decomposition is completely defined by the signal covariance matrix, for the DWT case there is an infinity of possibilities for the filters $h(n)$ and $g(n)$, among which a single pair must be chosen. The optimization problem is thus more complex than in the KLD case, and is carried out in two steps: i) Choose a small number of combinations of sampling intervals and QMF pairs that, in some way, fit best to the signal present in hypothesis H_1; ii) Choose among the results of i) and the other parameters that need optimization, the situation that returns the best detector performance for a given CC. Step i) of the optimization process is performed according to a frequency functional discussed in [4].

For a given scale j and a filter length N_f, the filter coefficients and the sampling interval T_s^a are chosen in order to maximize the functional

$$\mathcal{F} = \int_{-\frac{\pi}{T_s^a}}^{+\frac{\pi}{T_s^a}} P(\omega)|G^j(\omega T_s^a)|^2 d\omega \quad P(\omega) = \int_{-\infty}^{\infty} S(t,\omega)dt, \tag{11}$$

where $S(t,\omega) = \text{FT}_\tau[k_s(t, t-\tau)](\omega)$, $\text{FT}[\cdot]$ denotes the Fourier transform and $|G^j(\omega)| = |\text{FT}[g_k^j(n)]|$ is independent of the translation parameter k [3]. The maximization of $\mathcal{F}$ is performed for a limited number of scales (typically $j = 1, 2, 3$). For larger scales, the optimum sampling interval T_s^a becomes too small and the solutions are not efficient. Regarding the filter coefficients, [7] presents parameterizations of orthonormal QMFs of length $2N_f$ from $N_f - 1$ free parameters $\theta_k \in [0; 2\pi]$, $k = 1, \cdots, N_f - 1$. The maximization of $\mathcal{F}$ is performed in the following steps. i) The Daubechies family of filters [1] are used to initialize the values of θ_k; ii) $\mathcal{F}$ is maximized according to T_s^a; iii) $\mathcal{F}$ is maximized according to the coefficients θ_k; iv) return to step ii) until convergence is achieved. In steps ii) and iii), MATLAB optimization algorithms were used.

In (11), the maximization of $\mathcal{F}$ is performed in a single scale j to achieve the maximum sparsity in the DWT coefficients covariance matrix. However, a convenient description of the signal may require more scales than only the scale j. Furthermore, although the resulting matrix $\boldsymbol{M}$ is sparse, the evaluation of (10) considers only the largest N_m non-zero terms of $\boldsymbol{M}$. This procedure requires some care since it is necessary to ensure that the resulting matrix is symmetric and all the cross-terms correspond to elements that are also present in the diagonal (i.e, the elements (i,j) and (j,i) of $\boldsymbol{M}$ can be nonzero, if the elements (i,i) and (j,j) are nonzero too). This procedure ensures that the corresponding approximated signal covariance matrix is semipositive definite. Under these conditions, the parameters to tune in the second step of the optimization algorithm are: a) The filter length N_f; b) The optimization scale j; c) The initial and final DWT decomposition scales J_1 and J_2; d) The binary parameter f_{J_2} that determines whether the lowpass DWT residue at scale J_2 is computed ($f_{J_2} = 1$) or not ($f_{J_2} = 0$); e) The number of nonzero terms in the matrix $\boldsymbol{M}$, N_m; f) The shift parameter k_t (meaning that the likelihood ratios are performed at every $k_t.2^{J_2}$ approximated sampling intervals T_s^a). For a given optimization scale j and filter length N_f, step i) of the optimization procedure has already determined the corresponding best filter

Table 1: KLD optimization results.

β_{ref}	N_s^{ar}	N_w^r	M_a	N_t
50	4.4452	293	36	11
150	3.8628	328	60	9

coefficients and approximated and reference sampling intervals ratio N_s^a. The resulting CC is given by

$$\beta_{DWT} = \frac{\left(\sum_{l=1}^{J_2-1} 2^l + \sum_{l=0}^{J_2-J_1} 2^l + f_{J_2}\right) N_f k_t + N_{mc} + 2N_{md}}{k_t 2^{J_2} N_s^a}, \tag{12}$$

where N_{mc} and N_{md} correspond, respectively, to the nonzero cross and diagonal elements of $\boldsymbol{M}$. Given a desired CC β_{ref}, we have

$$N_{mt} = \left\lfloor \left[2^{J_2} N_s^a k_t \beta_{ref} - k_t N_f \left(\sum_{l=1}^{J_2-1} 2^l + \sum_{l=0}^{J_2-J_1} 2^l + f_{J_2} \right) \right] \right\rfloor, \tag{13}$$

where $N_{mt} = N_{mc} + 2N_{md}$ and $\lfloor(x)$ stands for the largest even integer smaller than x. A given set of parameters $\{\gamma\} = \{j, J_1, J_2, N_s^a, k_t, f_{J_2}, N_f\}$ uniquely define the value of N_{mt}, but there are still a large number of combinations of pairs $\{N_{mc}, N_{md}\}$ to choose from. To find the best combination, we evaluate the value of $\alpha(\{\gamma\})$ with the restriction given in (13), from the limit case where $N_{md} = N_{mt}/2$ and $N_{mc} = 0$ (only diagonal elements are present in the covariance matrix), up to the other limit case where $N_m = N_{md} + N_{mc} = N_{md}^2 \Leftrightarrow N_{md}^2 + N_{md} - N_{mt} = 0$ (we have a matrix with smaller dimension, but with no zero elements), and choose the maximum. From one limit case to the other, we decrease N_{md} by one (taking out the smaller element of the diagonal) and increase N_{mc} by the two largest symmetric cross elements (N_{mt} remains the same).

6. SIMULATION RESULTS

This example illustrates the case where the transient signal $s(t)$ in hypothesis H_1 is represented by a large number of eigenvalues. $s(t)$ corresponds to a small chirplike stochastic transient of small duration (approximatively 0.5s.), with 98 nonzero eigenvalues presented in fig 2 a). This signal shows that a relatively large number of orthonormal eigenvectors sound alike to the human ear, and thus may be generated by a single source. The discrete reference signal corresponding to $s(t)$ is sampled at a frequency of 1600 Hz., with a gating window of 800 samples. The noise variance at the reference signal sampling frequency is $\sigma^2 = 10$. The processors using the KLD and DWT were optimized respectively for the cases where $\beta_{ref} = \{10; 50\}$ and $\beta_{ref} = \{50; 150\}$. The parameter optimization results are presented respectively in tables 1 and 2. Using the optimum parameters for each case, the receiver operating characteristics (ROCs) were drawn from Monte-Carlo simulation with 100000 runs, and the results are presented in fig. 2 b).

The ROCs in fig. 2 b) show that the performances of the DWT with $\beta_{ref} = 10$ and $\beta_{ref} = 50$ are better respectively to the KLD ones with $\beta_{ref} = 50$ and $\beta_{ref} = 150$. Thus, in the present situation, the DWT requires at least 3 to 5 times less multiplications to achieve better results than the KLD. This result is due mainly to

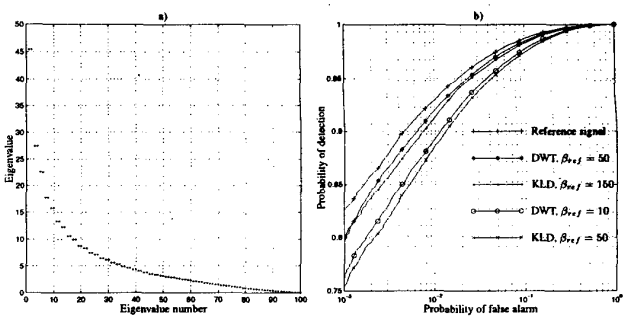

Figure 2: a) Eigenvalues of $s(t)$. b) Receiver operating characteristics.

Table 2: DWT optimization results.

β_{ref}	N_f	j	J_1	J_2	j_{J_2}	N_{md}	N_{mc}	k_t
10	3	1	1	2	0	55	233	2
50	8	2	1	3	1	110	544	1

the low-cost CC of the DWT decomposition stage, together with the sparsity of the resulting coefficients covariance matrices.

7. CONCLUSION

This paper reports on the real-time detection of Gaussian transient signals in noise. It presents methods for the optimization of the processor parameters either when the KLD or the DWT are used. It is known that for stationary and/or strongly correlated signals the KLD is efficient. The example presented shows that when the original signal has a large number of eigenvalues, the DWT may present a huge efficiency increase comparing to the KLD.

8. REFERENCES

[1] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, XLI:909–996, 1988.

[2] K. Fukunaga. *Introduction to Statistical Pattern Recognition, Second Edition*. Academic Press, 1990.

[3] F. M. Garcia and I. M. G. Lourtie. An arrival time estimator for transient stochastic signals using the wavelet transform. In *38^{t}h Midwest Symposium on Circuits and Systems*, Rio de Janeiro, Brazil, August 1995.

[4] F. M. Garcia and I. M. G. Lourtie. A wavelet transform frequency classifier for stochastic transient signals. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3058–3061, Atlanta, USA, May 1996.

[5] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, July 1989.

[6] R. J. Vaccaro. The past, present and future of underwater acoustic signal processing. *Signal Processing Magazine*, 15(4):21–51, July 1998.

[7] H. Zou and A. H. Tewfik. Parametrization of compactly supported orthonormal wavelets. *IEEE Transactions on Signal Processing*, 41(3):1428–1431, March 1993.