ENHANCED SIGNATURES FOR EVENT CLASSIFICATION : THE PROJECTOR APPROACH

Vidya Venkatachalam and Jorge L. Aravena

Department of Electrical and Computer Engineering
Louisiana State University, Baton Rouge, LA 70803-5901
Tel : (504) 388-5537, Fax : (504) 388-5200
Email : aravena@ee.lsu.edu

ABSTRACT

The classification of nonstationary signals of unknown duration is of great importance in areas like oil exploration, moving target detection, and pattern recognition. In an earlier work, we provided a solution to this problem, based on the wavelet transform, by defining representations called *pseudo power signatures* for signal classes which were independent of signal length, and proposed a simple approach using the Singular Value Decomposition to generate these signatures. This paper offers a new approach resulting in more discriminating signatures. The enhanced signatures are obtained by solving a nonlinear minimization problem involving an inverse projection. The problem formulation, solution procedure, and computational algorithm are presented in this work. The efficacy of the projection signatures in separating highly correlated signal classes is demonstrated through a simulation example.

1. INTRODUCTION AND PREVIOUS WORK

Consider the following classification problem :
Signals are obtained by propagating electromagnetic waves through several layers of different classes of materials. The goal is the determination of the various classes present and the thickness of each layer. We denote the presence of a particular class as the occurrence of an event.
The problem was described as a generic classification problem in a previous work ([1]), and was solved by introducing the concept of *pseudo power signatures.* For signals in each class, these signatures capture information at different scales, independent of the signal duration. Essentially, the signatures characterize the scale power distribution in a manner independent of time. In [1], we proposed an approach to determine the signatures by performing a Singular Value Decomposition (SVD) of the Continuous Wavelet Transform (CWT), and extracting the principal component. Specifically, for $x \in L^2(\Re)$ with CWT, $c_\psi^x \in H = L^2(\Re^2, C_\psi^{-1}\frac{da\,db}{a^2})$, where ψ is an admissible wavelet, we approximated $c_\psi^x(a,b)$ by a separable element of the form [1]

$$c_\psi^x(a,b) \approx s_\psi^x(a) r_\psi^x(b)$$

[1] We have shown ([2]) that there do not exist any admissible wavelets that admit a separable CWT function. Thus, we can only approximate a CWT function by a separable form.

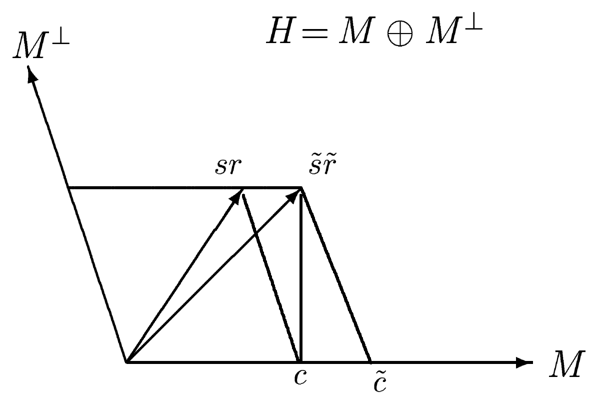

Figure 1: Graphical representation of the SVD and projection vectors

where $s_\psi^x \in S = L^2(\Re, C_\psi^{-1}\frac{da}{a^2})$, and $r_\psi^x \in R = L^2(\Re, db)$. The normalized function s_ψ^x corresponds to the pseudo power signature of x.
Now, it is known that the principal component of the SVD provides the optimal separable approximation to an element in the standard Hilbert space $L^2(\Re^2, dbda)$. However, H is a Hilbert space with a weighted inner product, and the principal component obtained using the traditional SVD analysis does not provide the closest separable approximation to an element in H. The situation is described graphically in Figure 1. In the figure, M denotes the closed subspace of CWT functions, and $M^\perp$ its orthogonal subspace in H. The element, $c \in M$, is to be approximated by a separable element of H. In theory, the best separable approximation is provided by the separable element in H that orthogonally projects onto c. The traditional SVD analysis yields the element $\tilde{s}\tilde{r} \in H$ as the best separable approximation. Note however, that in the sense of the weighted inner product defined in H, the element $\tilde{s}\tilde{r}$ does not orthogonally project onto c, but rather onto $\tilde{c} \in M$. If $\| c - \tilde{c} \|_H$ is large, then we can intuitively see that the element $\tilde{s}\tilde{r}$ is a poor approximation to c. What we need to determine is the element $sr \in H$ that orthogonally projects onto c. The normalized function s can then be used to denote the pseudo power signature of the function $x \in L^2(\Re)$ whose CWT is given by c. Since this signature is obtained as a result of a projection,

it is referred to as a projection signature. The following sections describe in detail a method to create the projection signatures, the computational algorithms used, and some experimental results on the classification of signals using the approach.

2. PROBLEM FORMULATION

The first step in the determination of the projection signature is the definition of a suitable orthogonal projection operator $\mathcal{K} : H \to M$.

Theorem 2.1 *There exists an orthogonal projection operator $\mathcal{K} : H \to M$ defined as follows. Given any $c \in H$,*

$$\mathcal{K}[c](a,b) = C_\psi^{-1} \int_\alpha \int_\beta \overline{c_\psi^{\psi_{ab}}(\alpha,\beta)} c(\alpha,\beta) \frac{d\beta d\alpha}{\alpha^2}$$

The proof is fairly simple, and can be obtained from [2]. Moreover, denoting $\Gamma : L^2(\Re) \to H$ as the CWT operator, and $\Gamma^* : H \to L^2(\Re)$ as the adjoint operator (Γ^* is effectively the inverse CWT operator extended to the whole of H by defining $M^\perp$ to be its null space), we can easily show that $\mathcal{K} = \Gamma\Gamma^*$ ([2]). A result which readily follows from Theorem 2.1 is given by :

Corollary 2.1 *To every $c \in H$, there corresponds one and only one $\hat{x} \in L^2(\Re)$ such that the CWT of $\hat{x}$, $c_\psi^{\hat{x}}$, is given by*

$$c_\psi^{\hat{x}}(a,b) = \mathcal{K}[c](a,b)$$

For any given $x \in L^2(\Re)$, let $c_\psi^x \in M$ denote its CWT with respect to an admissible $\psi \in L^2(\Re)$. Consider the element $sr \in H$. Let $\hat{c} = \mathcal{K}[sr] \in M$, and $\hat{x} \in L^2(\Re)$ the element associated with sr by Corollary 2.1. It intuitively follows that if we determine $sr \in H$ such that it minimizes $\| c_\psi^x - \hat{c} \|_M$, then the sr effectively minimizes $\| x - \hat{x} \|_2$. Hence, we can expect that $\hat{c}$, and consequently, sr, will better characterize the intrinsic properties of x. However, we do not know if the orthogonal projection operator $\mathcal{K}$, when restricted to the set of separable elements in H, is one-one. Consequently, there may be more than one separable element $sr \in H$ with the same projection $\hat{c} \in M$. Thus, in order to ensure the determination of a *unique* projection signature, we add a regularizing term $\alpha \| sr \|$ to the minimization problem. For analysis purposes, we choose $\alpha = 1$. The minimization problem is then represented as follows :
For a given $c_\psi^x \in M$, find the decomposition $s_\psi^x r_\psi^x \in H$ that minimizes the index

$$J(s_\psi^x, r_\psi^x) = \left\{ \| c_\psi^x - \mathcal{K}[s_\psi^x r_\psi^x] \|_M^2 + \| s_\psi^x r_\psi^x \|_H^2 \right\}$$

This is an infinite dimensional nonlinear minimization, and requires the solution of the inverse projection problem. The problem formulation and solution procedure for the infinite dimensional case can be found in [2]. However, for a practical application, the problem needs to be reduced to a finite dimensional one. For this, we need to find a finite dimensional representation for $\mathcal{K}$, and determine a suitable discretization for the elements $c_\psi^x, \mathcal{K}[s_\psi^x r_\psi^x] \in M$, $s_\psi^x \in S$, and $r_\psi^x \in R$.

2.1. Discrete approximation to the minimization problem

Using the concept of frames and frame operators, and a wavelet $\psi \in L^2(\Re)$ that arises from a multiresolution ([3]), we can determine the set of discretized coefficients $\{c_\psi^x(2^l, n)\}_{ln}$ using the Shensa algorithm ([4]). For most practical applications, c_ψ^x has near compact support in the time-scale plane. For a signal of finite time support, and a suitably chosen ψ, (where ψ has compact time support), it can be well approximated using finitely many discretized CWT coefficient values. This implies that there exists L, N such that $c_\psi^x(2^l, n) \approx 0$, for all $l > L$ and for all $n > N$. One can represent this using a finite dimensional matrix $C_\psi^x(l,n) = [c_\psi^x(2^l,n)]$ of dimension $L \times N$.
The problem of finding a discrete approximation to the orthogonal projection operator $\mathcal{K} : H \to M$, is more involved. The approach followed here approximates $\mathcal{K}$ by using a successive application of the inverse and forward Shensa algorithms. Let $S : l^2(\mathcal{Z}) \to l^2(\mathcal{Z}^2)$ denote the forward Shensa operator. Then the adjoint operator $S^* : l^2(\mathcal{Z}^2) \to l^2(\mathcal{Z})$ is given by the extended inverse Shensa operator. Clearly, $S^*S = I$. Under certain assumptions, we can show that SS^* is the discrete approximation to $\mathcal{K}$.
Define an operator, $F_2 : H \to l^2(\mathcal{Z}^2)$ as

$$F_2[c](l,n) = \langle c, \delta_{l,n} \rangle_H, \quad c \in H$$

where the sampling function $\delta_{l,n} = \delta(a - 2^l, b - n)$. The adjoint operator $F_2^* : l^2(\mathcal{Z}^2) \to H$ is then defined as

$$F_2^*[h](a,b) = \sum_l \sum_n h(l,n) q_{l,n}(a,b), \quad h \in l^2(\mathcal{Z}^2)$$

where

$$q_{l,n}(a,b) \quad = \quad \begin{cases} 1, & 2^l \le a < 2^{l+1}, \quad n \le b < n+1 \\ 0, & elsewhere \end{cases}$$

and $F_2^* F_2 c$ is a piecewise constant approximation to $c \in H$ of the form

$$F_2^* F_2[c](a,b) \approx \sum_{l=1}^{L} \sum_{n=1}^{N} c(2^l, n) q_{l,n}(a,b)$$

It is possible to show that if we make the assumption that the CWT is piecewise constant, the orthogonal projection operator $\mathcal{K} = \Gamma\Gamma^*$ can be approximated as

$$\mathcal{K} \quad \approx \quad F_2^* S S^* F_2$$

With $s_d(l) = s_\psi^x(2^l)$, $s_d \in \mathcal{C}^L$, and $r_d(n) = r_\psi^x(n)$, $r_d \in \mathcal{C}^N$, the element $\mathcal{K}[s_\psi^x r_\psi^x]$ can be approximated by the finite dimensional matrix $\tilde{C} \in \mathcal{C}^{L \times N}$ resulting from the operation $SS^*[s_d r_d^T]$. The infinite dimensional minimization then reduces to the following problem :
Given a matrix $C_\psi^x \in \mathcal{C}^{L \times N}$ of samples on the Shensa grid of the CWT of $x \in L^2(\Re)$, determine the rank one matrix $s_d r_d^T \in \mathcal{C}^{L \times N}$ such that the following functional is minimized

$$J(s_d, r_d) = \| C_\psi^x - SS^* \left[s_d r_d^T \right] \|_2^2 + \| s_d r_d^T \|_2^2$$

3. SOLUTION TO THE MINIMIZATION

We solve the finite dimensional minimization problem using the following steps :

1. Proving existence of a global minimum.
2. Establishing necessary conditions for optimality.
3. Developing an iterative algorithm involving successive quadratic minimizations with respect to the vectors $s_d \in \mathcal{C}^L$, and $r_d \in \mathcal{C}^N$.

The discrete orthogonal projection approximation operator SS^* on the separable finite dimensional Hilbert space $l^2(\mathcal{C}^{L\times N})$, is isometrically equivalent to a matrix $\tilde{K} : \mathcal{C}^{LN} \to \mathcal{C}^{LN}$. The isometry, T, essentially rearranges matrices as vectors by stacking its rows. With this notation, we can define the following two subproblems. Let I_L denote the identity matrix of size L, and $c = TC_\psi^x$ a column vector representation for C_ψ^x. Then, with $\otimes$ denoting the standard Kronecker product, for a fixed r_d, we obtain a minimization with respect to s_d as shown.

$$\min_{s_d} J(s_d) = \| c - \tilde{K}(r_d \otimes I_L)s_d \|_2^2 + \| (r_d \otimes I_L)s_d \|_2^2 \quad (1)$$

Similarly, denoting $c_T = TC_\psi^{x^T}$, and I_N as the identity matrix of size N, for a fixed s_d, we can define the minimization problem with respect to r_d as

$$\min_{r_d} J(r_d) = \| c_T - \tilde{K}(s_d \otimes I_N)r_d \|_2^2 + \| (s_d \otimes I_N)r_d \|_2^2 \quad (2)$$

It can be readily shown that each of the above subproblems has a unique solution, since J is separately convex in s_d, and r_d. Let $B_L = \{s_d \in \mathcal{C}^L; \ \| s_d \|_2 \leq 1\}$ denote the closed unit ball in $l^2(\mathcal{Z})$ of dimension L, which is compact. For a fixed s_d of unit norm, $s_d \in B_L$. Let $r_d(s_d)$ be the solution to Equation 2. Then, the functional $J(s_d, r_d(s_d))$ is effectively a function of $s_d \in B_L$.

Theorem 3.1 *There exists $\bar{s}_d \in B_L$, and $\bar{r}_d = r_d(\bar{s}_d) \in \mathcal{C}^N$ such that*

$$J(\bar{s}_d, \bar{r}_d) = \inf_{s_d \in B_L, r_d \in \mathcal{C}^N} J(s_d, r_d)$$

The proof of the theorem can be found in [2], and is based on the result that J is continuous on the compact set B_L. Once, the existence of the minimal solution has been established, we can develop necessary conditions for minimization using Calculus of Variations. Taking variations with respect to s_d in Equation 1, and r_d in Equation 2, the necessary conditions for minimization with respect to s_d and r_d respectively are given by

$$(r_d \otimes I_L)^*(\tilde{K} + I)(r_d \otimes I_L)s_d - (r_d \otimes I_L)^* c = 0 \quad (3)$$

$$(s_d \otimes I_N)^*(\tilde{K} + I)(s_d \otimes I_N)r_d - (s_d \otimes I_N)^* c_T = 0 \quad (4)$$

In order to determine the projection signature, we developed an iterative procedure which involves successively solving Equations 3 and 4. Owing to the convexity of J in each variable s_d, r_d separately, the first order necessary conditions also become sufficient for each minimization. Thus, we obtain a monotonically decreasing cost sequence $\{J^i\}_i$.

Theorem 3.2 *There exists $\bar{s}_d \in B_L$, and $\bar{J} \geq 0$ such that the sequence $\{J^i\}_i$ converges to $\bar{J} = J(\bar{s}_d)$.*

The proof of the theorem is based on the continuity of $J(s_d)$ on the compact set B_L, and the sequential compactness of B_L. Complete details are given in [2]. Thus, we have a solution technique to compute the projection signatures, with a guaranteed convergence to the limiting solution, though there is no assurance that the convergence is to the global minimum.

3.1. Computational algorithm

The computational algorithm to determine the projection signatures, based on the iterative procedure developed, is as follows.

1. Select a wavelet $\psi \in L^2(\Re)$ which arises from a multiresolution, and the number of levels L to be used in the filter bank corresponding to the multiresolution.
2. For the given finite discrete input signal $x \in \mathcal{C}^{N_x}$, determine the discretized CWT coefficient matrix $C_\psi^x \in {}^{L\times N}$ using the forward Shensa algorithm.
3. Based on the scalogram $SC_\psi^x(l,n) = | C_\psi^x(l,n) |^2$ values, modify L such that $C_\psi^x(l,n) \approx 0$, for all $l \geq L$, $n \geq N$, and recompute C_ψ^x using the modified value of L.
4. Pick random vectors $s_d^0 \in \mathcal{C}^L$, $r_d^0 \in \mathcal{C}^N$, and set a value for *tol*.
   - At the i th stage, set $r_d = r_d^{i-1}$. Using the conjugate gradient technique, with gradient

     $$\Delta_{s_d} = \Re\{(s_d r_d^{i-1^T} + SS^*\left[s_d r_d^{i-1^T}\right] - C_\psi^x) r_d^{i-1^c}\}$$

     solve the minimization problem for s_d. Let $\tilde{s_d}^i$ denote the solution.
   - Set $s_d^i = \frac{\tilde{s_d}^i}{\|\tilde{s_d}^i\|}$.
   - Next, with $s_d = s_d^i$, using the conjugate gradient technique with gradient

     $$\Delta_{r_d} = \Re\{(s_d^i r_d^T + SS^*\left[s_d^i r_d^T\right] - C_\psi^x)^T s_d^{i^c}\}$$

     solve the minimization problem for r_d . Let r_d^i denote the solution.
   - Compute the cost function $J(s_d^i, r_d^i)$. If $(\ J(s_d^{i-1}, r_d^{i-1}) - J(s_d^i, r_d^i)\ ) < tol$, terminate.
5. end.

The projection signature of the input signal $x \in \mathcal{C}^{N_x}$ is given by the function s_d^i obtained at the termination of the algorithm.

4. SIMULATION RESULTS

In this section, we will consider the same example presented in [1], and show the improvement obtained using the projection signatures. The signals x_1, x_2, x_3 from three different signal classes and their projection signatures are displayed in Figure 2. For purposes of comparison, the SVD signatures obtained from the principal component of the SVD of the discretized CWT matrix of the signals x_1, x_2, x_3 are also displayed in the same figure. All the signatures were generated using the wavelet $Db4$. Observe how well the projection signatures S_1 and S_2 separate the highly correlated signal classes x_1 and x_2. The signal to be classified x is shown in Figure 3. The signal x was obtained by concatenating segments of each of the three signal classes. The classification process we used here was a very simple one. We determined the correlation of each signature with the discretized CWT $[c_\psi^x(2^l, n)]$ of the signal x for each n. The correlation graphs are also shown in Figure 3. Observe that the transition points at -50 and 50 are clearly marked, and the signal x can be classified with a high degree of confidence based on the correlation graphs. This example is representative of several classification tests we ran using projection signatures. In most cases, we obtained signatures of very good quality, with unambiguous classification results.

From a computational standpoint, the use of the conjugate gradient technique guarantees convergence in at most P steps where P is the size of the vector over which we are minimizing, which makes this technique very efficient. In fact, our experiments show that convergence usually occurs in far fewer iterations, no more than 15 for signals with 30000 sample points. One would expect from the above discussion, that this computational technique would fail when the assumption on the piecewise constant nature of the CWT is violated. However, our results on randomly selected sample signals indicate that this approach does in fact provide highly discriminating signatures for diverse classes of signals.

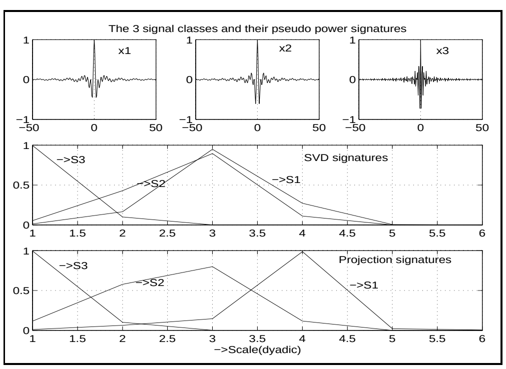

Figure 2: The 3 known signals and their SVD and projection signatures

5. CONCLUSIONS

In this paper, we proposed a technique using the CWT to compute pseudo power signatures for signal classes based on projections. The technique involved a nonlinear minimization, and we provided the complete solution to the minimization problem. We also developed an efficient algorithm for the computation of the signatures using an iterative procedure with fast convergence, and illustrated the quality of the projection signatures through a representative example. It is important to note that the actual classification can be done very quickly, since the signatures are vectors of very small dimension. This approach has wide applicability, in areas as diverse as oil exploration, hidden mine detection, moving target detection, system identification, and pattern recognition.

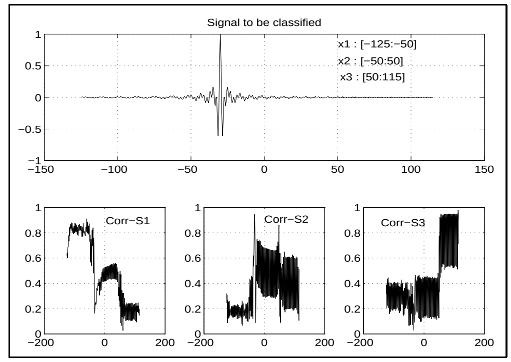

Figure 3: The signal to be classified and its correlation graphs

6. REFERENCES

[1] Vidya Venkatachalam and Jorge L. Aravena, " Nonstationary Signal Classification Using Pseudo Power Signatures," *Proc. ISCAS 98*, Monterey, CA.

[2] Vidya Venkatachalam, "Nonstationary Signal Analysis And Classification Using Pseudo Power Signatures," *Ph.D Dissertation*, Louisiana State University, Dec. 1998.

[3] I. Daubechies, " Ten Lectures on Wavelets," *Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, 1992.

[4] Mark J. Shensa, " The Discrete Wavelet Transform: Wedding the À Trous and Mallat Algorithms," *IEEE Trans. on Signal Processing*, vol. 40, pp. 2464-2482, Oct. 1992.