STRUCTURAL RISK MINIMIZATION FOR REDUCED-BIAS TIME-FREQUENCY-BASED DETECTORS DESIGN

C. Richard, R. Lengellé

Laboratoire LM2S, Université de Technologie de Troyes
12 rue Marie Curie, BP 2060, 10010 Troyes Cedex, FRANCE
cedric.richard@univ-troyes.fr

ABSTRACT

Detectors design requires substantial knowledge of the observation statistical properties, conditionally to the competing hypotheses H_0 and H_1. However, many applications involve complex phenomena, in which few a priori information is available. Several methods of designing time-frequency-based (TF) receivers from labeled training data have been proposed. Unfortunately, the resulting detectors have large biases, particularly when the number of training samples is small against the data dimension. The method presented here is based on the Structural Risk Minimization principle developed by Vapnik, and consists in locally adjusting the resolution of TF-based detectors to the information carried by each TF location. This operation, controlled by a measure of H_0 and H_1 separability, allows to advantageously reduce receivers complexity and solutions bias. The resulting reduced-bias TF-based detectors can yield a substantial improvement in detection performances.

1. INTRODUCTION

Cohen's class time-frequency (TF) representations have been extensively used for detection in applications ranging from radar to machine fault diagnostics, due to the need for dealing with non-stationary signals. These bilinear representations are parameterized in time and frequency terms, and describe non-stationary signals characteristics via their time-varying spectral content. Most of the TF-based detectors which have been proposed are linear structures operating in the TF domain, and are merely equivalent to quadratic receivers usually defined in the time domain [1]. Recently, a promising TF-based quadratic detection theory has been introduced: in [9], Sayeed and Jones identify several scenarios in which detectors are both optimum and fully exploit the many degrees of freedom available in TF representations.

It is noteworthy that these approaches require substantial knowledge of signals whereas phenomena are complex and poorly understood in many applications [8]. In that context, Richard and Lengellé have presented a method of obtaining TF-based detectors directly from labeled training data [5][6]. The resulting receivers offer better performances than those obtained via the Fisher criterion maximization [4]. Unfortunately, detectors designed from training data often have a large bias, particularly when the number of training samples is small against the data dimension [3]. This experimental evidence has been theoretically studied by Vapnik and Chervonenkis, who have exhibited links between the generalization performances of receivers, their complexity (VC-dimension), and the size of the training sets [12]. Then methodologies such as Structural Risk Minimization (SRM) [11] and Minimum Description Length (MDL) [7] principles have been proposed to match the complexity of classification structures to the amount of available examples.

In this paper, the method that we present is based on the SRM principle, and consists in locally adjusting the resolution of TF-based detectors to the information carried by each TF location. This operation, controlled by the optimization of a measure of the hypotheses H_0 and H_1 separability, allows to advantageously reduce the VC-dimension of receivers and the bias of solutions. The resulting reduced-bias detectors, by virtue of their adapted resolution in the TF domain, can yield a substantial improvement in detection performances. The paper is organized as follows. First, we briefly describe an efficient method of designing TF-based detectors from training data. Then, we discuss the issue of obtaining reduced-bias TF-based receivers. An example illustrates the efficiency of our approach. Finally, we present some conclusions regarding the results presented.

2. TIME-FREQUENCY APPROACH TO DECISION PROBLEMS

2.1. Linear detection in the TF domain

The detection problem we consider is as follows. Given a discrete-time signal $\underline{x}$ received over the interval (D), where $\underline{x} = [x(0), ..., x(d-1)]^T$, one must decide between the two competing hypotheses H_0 and H_1:

$$\begin{cases} H_0: & C_x(k,f;\Pi) = C_n(k,f;\Pi), \\ H_1: & C_x(k,f;\Pi) = C_{s+n}(k,f;\Pi), \end{cases} \qquad (1)$$

where $\underline{s}$ is the underlying (complex) signal to be detected, and $\underline{n}$ some additive (complex) noise. C_x denotes a discrete Cohen's class TF representation of the signal $\underline{x}$, and Π the autocorrelation domain kernel.

By analogy with the classical matched filter theory, one can consider the following class of TF-based detectors:

$$\lambda(\underline{x}\,;\,h) = \sum_k \sum_f h(k,f)\, C_x(k,f;\Pi)\,. \qquad (2)$$

The two-dimensional function h is a TF reference to be determined, using the *a priori* known characteristics of the signal $\underline{s}$ and the noise $\underline{n}$. Such a determination can be achieved from training data by using the optimal approach proposed in the next section.

2.2. Detector design from training data

We now discuss a method of obtaining optimum TF-based detectors from training data (i.e. which minimize an estimation of the error probability), regardless of the signals distributions. This approach has been introduced by Fukunaga to design linear discriminants in the context of Pattern Recognition [3], and adapted by Richard and Lengellé to design optimum TF-based detectors [5][6].

Using the expression of the statistic (2), the detection problem (1) can be rewritten as follows:

$$\begin{cases} \text{if } \lambda(\underline{Y}\,;\,h) = \underline{V}^T \underline{Y} \geq \gamma \ \text{ then } H_1 \\ \text{else } H_0, \end{cases} \qquad (3)$$

where

$$\underline{Y} = \left[C_x(1,1;\Pi) \ldots C_x(d,d;\Pi)\right]^T$$
$$\underline{V} = \left[h(1,1) \ldots h(d,d)\right]^T. \qquad (4)$$

The design work consists in finding the optimum vector $\underline{V}$ and threshold value γ in the sense of a preselected criterion, and for a given data set. Using a minimal *a priori* knowledge, the statistic λ can be characterized by the following expected values and variances:

$$\eta_i = E\left\{\lambda(\underline{Y}\,;\,h) \middle| H_i\right\} = \underline{V}^T \underline{M}_i + \gamma,$$
$$\sigma_i^2 = \mathrm{Var}\left\{\lambda(\underline{Y}\,;\,h) \middle| H_i\right\} = \underline{V}^T \Sigma_i \underline{V}, \qquad (5)$$

where

$$\begin{cases} M_i = E\left\{\underline{Y} \middle| H_i\right\}, \\ \Sigma_i = E\left\{(\underline{Y} - M_i)(\underline{Y} - M_i)^T \middle| H_i\right\}. \end{cases} \qquad (6)$$

M_i and Σ_i must be estimated from training samples.

Let $f(\eta_0, \eta_1, \sigma_0^2, \sigma_1^2)$ be any measure (depending only on these parameters) of H_0 and H_1 separability in the λ-space. This criterion must be optimized so that the derivatives of f, with respect to $\underline{V}$ and γ, must be equated to zero. The resolution of these two equations provides a interesting analytical solution for the detector (2) design:

$$\underline{V}_\alpha = \left[\alpha \Sigma_0 + (1-\alpha)\Sigma_1\right]^{-1} (M_1 - M_0), \qquad (7)$$

$$\alpha = \frac{\partial f / \partial \sigma_0^2}{\partial f / \partial \sigma_0^2 + \partial f / \partial \sigma_1^2}\,. \qquad (8)$$

Thus, the optimum TF-based receiver has the form (7) regardless of the selection of f: the criterion effect only appears in the parameter α ($0 \leq \alpha \leq 1$). In our case, we choose the α value which minimizes an estimation of the error probability, using an iterative procedure. Note that the obtained detector offers better performances than those determined via the maximization of the Fisher criterion ($\alpha = P\{H_0\}$) and the signal to noise ratio ($\alpha = 1$) [4][8].

As a conclusion, this method allows to determine the optimal TF-based receiver in the sense of the best criterion $f(\eta_0, \eta_1, \sigma_0^2, \sigma_1^2)$, without setting it up.

3. OPTIMIZATION OF TF-BASED DETECTION STRUCTURES

3.1. Complexity regularization

Achieving good performances with detectors designed from training samples requires matching their complexity to the amount of available data: receivers with a too large number of adjustable parameters may exhibit poor generalization performances, whereas those with an insufficient complexity may not be able to learn the training examples. In between, there is an optimal complexity which yields the best generalization error E_{gene} for a given size of the training set. This problem, which has particularly been studied by Vapnik *et al.* [12], is now briefly discussed.

Let C be a set of detectors, and let V_C be its VC-dimension. This parameter characterizes the complexity of the receivers contained in C: it is defined as the maximum number of training samples they can learn without error, and for all possible binary labelings. In some cases such as generalized linear classifiers, V_C corresponds to the number of free parameters available in the structure (e.g. if $C = \{(\lambda\,;\,\gamma)\}$, then $V_C = d^2 + 1$). But generally, its analytic determination is quite difficult. As shown in [11], the VC-dimension of a receiver allows to derive an upper bound of the generalization error from the size N of the design set, and from the training error. With a probability close to one, and simultaneously for all statistics which belong to C, it has been demonstrated that:

$$E_{gene} \leq E_{train} + O\left(\sqrt{V_C \ln(N)/N}\right). \qquad (9)$$

E_{train} denotes the frequency of errors on the training set.

The method of Structural Risk Minimization consists in matching V_C to the amount of training data in order to get the best compromise between the competing terms E_{train} and $O(.)$: reducing V_C causes $O(.)$ to decrease, but E_{train} to

Fig.1: WD of the signal s (assumed unknown)

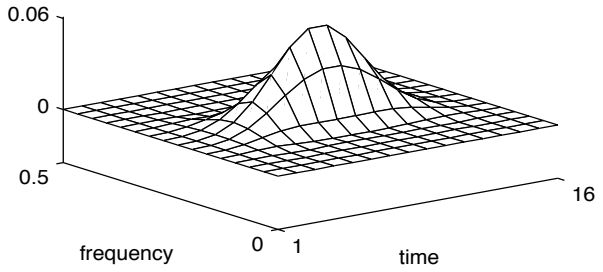
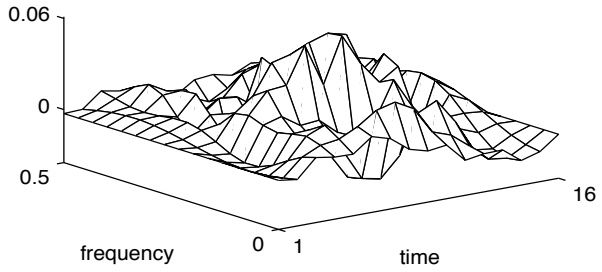


Fig.2: Detector config. (Section 2.2)



increase. In order to give a precise statement of the VC-dimension selection problem, we assume the following nested sequence of subsets C_i in the class C:

$$C_1 \subset \ldots \subset C_r \subset \ldots \Rightarrow V_{C_1} \leq \ldots \leq V_{C_r} \leq \ldots \quad (10)$$

The problem is then: given a training set, and the sequence of receivers $(\lambda_1 ; \gamma_1)$, ... determined by a learning algorithm, select C_{opt} such that $(\lambda_{opt} ; \gamma_{opt})$ minimizes the resulting generalization error. Because E_{gene} cannot be computed in general, one can consider its estimation E_{test} on a separate test set (Cross-Validation). Some approximations of the generalization error, solely based on E_{train} and on additive complexity penalty terms, have also been introduced [7][11].

3.2. Reduced-bias TF-based detectors design

As it as been shown before, one must balance the VC-dimension of a TF-based detector with its goodness of fit to the training data, in order to optimize E_{gene}. Let us define the resolution of the receiver (2) over a TF area (A) as the number of free parameters h(k, f) available over it. The approach we now investigate consists in locally adapting the detector resolution to the information carried by each TF location: the more discriminant the information over (A) is, the higher the local resolution is. Let C_r be the following class of receivers whose resolution, over the whole TF domain, is equal to r:

$$C_r = \left\{ (\lambda_{WV}(.;h);\gamma) \,\middle|\, h(k,f) = \alpha_j \text{ on } A_j, j = 1,\ldots,r \right\}, \quad (11)$$

where $\{A_1, ..., A_r\}$ denotes a partition of the TF domain. This condition imposes on the reference h to roughly exploit the structure of TF representations (correlation between adjoining TF locations [2], non-existence of impulses in the TF domain, ...).

It is noteworthy that $C_1, C_2, ..., C_{d^2}$ define a nested sequence of subsets in the class C of linear receivers operating in the TF domain, from the smallest resolution (the energy detector, when r = 1) to the highest one ($r = d^2$). The VC-dimension of each subset C_r is equal to (r + 1). As a consequence, the strategy discussed in §(3.1) can be advantageously used to match the complexity of TF-based detection structures to the amount of available training data. Thus, the target detector $(\lambda_{opt} ; \gamma_{opt})$ satisfies:

$$(\lambda_{opt} ; \gamma_{opt}) = \text{argmin}_r \left\{ E_{test}\left((\lambda_r ; \gamma_r)\right) \right\} \quad (12)$$

where

$$(\lambda_r ; \gamma_r) = \text{argmin}_{\{A_1,\ldots,A_r\}} \left\{ E_{train}\left((\lambda;\gamma) \in C_r\right) \right\}. \quad (13)$$

A possible way of representing the partition $\{A_1, ..., A_r\}$ comes from the Voronoï diagram theory. Consider a finite number of points a_1, ..., a_r (the Voronoï sites) of the TF domain. To each site a_i is associated the set A_i of all TF locations for which the closest Voronoï site is a_i, termed Voronoï cell. Thus, the resolution of eq.(13) can be completed by optimizing the location of the Voronoï sites in the TF domain. Unfortunately, E_{train} as a function of $(a_1, ..., a_r)$ is piecewise constant due to the TF domain discretization, which leads the standard gradient-based methods to perform badly. In the context of mechanical inclusions identification [10], Schoenauer *et al.* have shown the efficiency of Evolutionary algorithms for Voronoï diagrams optimization. In Part 4, a simple example illustrates the excellent performances of this approach to design reduced-bias TF-based detectors.

4. EXPERIMENTAL RESULTS

In the case of detecting the presence or absence of $s(k) = \sin(0.44\pi k + \theta) \times (1 - \cos(2\pi k/15))$, $k \in \{0, ..., 15\}$, in zero mean white Gaussian noise, with phase θ a uniform random variable, the optimal receiver is well known to correspond to the inner product of the Wigner distribution (WD) of the signal s (Fig.1) with that of the observation x. In order to illustrate our approach, several experiments of blind detector design from training data were conducted with 270 realizations of the hypotheses H_0 and H_1, in such way that the training set size was relatively small compared to the problem dimension ($16^2 + 1$). A test set containing 150 realizations of the signal to be detected plus noise, and noise only, was also generated. These samples were dedicated to the selection of α_{opt} and r_{opt} in equations (7) and (12). The TF reference h, resulting from the direct

Fig.3: Test error vs resolution

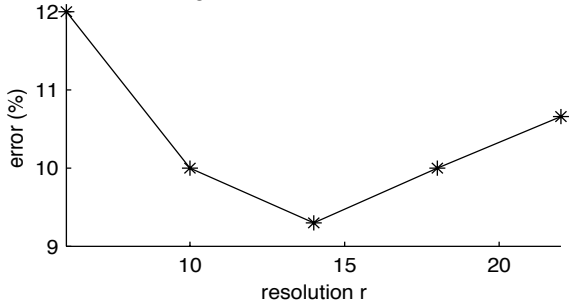
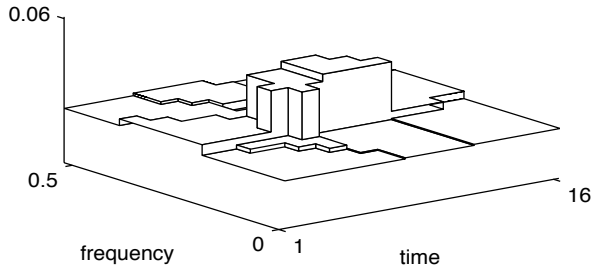


Fig.4: Reduced–bias detector config. (Section 3.2)



application of the method introduced in Part 2, is shown in Fig.2 ($r = d^2$): the presence of the signal component is not very apparent because few training examples were available. The design of the optimal reduced-bias TF-based detector was also performed. The competition between VC-dimension and training error results in a minimum of the error test, obtained for r equal to 14 (Fig.3). For this optimal value of r, the TF reference h is presented in Fig.4: it has roughly the same structure as the WD of the signal to be detected (Fig.1), and strongly synthesizes the information carried by the TF representation.

The generalization performance of these detectors were estimated by applying them to 2000 realizations each of signal present or absent. Using the quadrature matched filter, which constitute the optimal detector, the error probability was 8.50%. The performance of the receiver resulting from the method introduced in Part 2 was 11.05%. This result must be compared to 8.90% obtained with the reduced-bias TF-based detector. Consequently, these experiments clearly demonstrate the ability of the proposed method to closely approach the performance of the optimal quadratic detector, even if the size of the training set is relatively small compared to the problem dimension.

5. DISCUSSION

Bilinear TF representations have been widely used for detection in non-stationary environments. However, most of the TF-based detectors which have been proposed are linear structures operating in the TF domain, merely equivalent to quadratic receivers usually defined in the time domain. Moreover, their design requires substantial knowledge of signals whereas phenomena are poorly understood in many applications. In order to improve the design stage, we have developed a method of determining the optimal TF-based detector, i.e. which minimizes an estimation of the error probability, directly from labeled training data. Unfortunately, it is well-known in Pattern Recognition that sample-based classifiers can have large biases, particularly when the size of the training set is small against the data dimension. The method developed here, which is based on the Structural Risk Minimization principle, fully exploit the structure of TF representations to advantageously reduce the VC-dimension of the receiver. The obtained reduced-bias detectors, by virtue of their adapted resolution in the TF domain, can yield a substantial improvement in generalization performances. However, this efficient process is computationally expensive, and further works on discovering practical algorithms are needed.

6. BIBLIOGRAPHY

[1] P. FLANDRIN, "A Time-Frequency Formulation of Optimum Detection," IEEE Transactions on ASSP, vol. 36, pp. 1377-1384, 1988.

[2] P. FLANDRIN, Temps-Fréquence. Paris, France: HERMES, 1993.

[3] K. FUKUNAGA, Introduction to Statistical Pattern Recognition. London, GB: Academic Press, 1990.

[4] D. L. JONES and A. M. SAYEED, "Blind Quadratic and Time-Frequency based Detectors from Training Data," IEEE ICASSP, 1995.

[5] C. RICHARD and R. LENGELLE, "Joint Time and Time-Frequency Optimal Detection," IEEE UK-TFTS, 1997.

[6] C. RICHARD and R. LENGELLE, "Une nouvelle approche pour la détection linéaire optimale dans le plan temps-fréquence," Colloque GRETSI, Grenoble (F), 1997.

[7] J. RISSANEN, Complexity in Statistical Inquiry. Teaneck, USA: World Scientific, 1989.

[8] B. SAMIMY, G. RIZZONI, A. M. SAYEED, and D. L. JONES, "Design of Training Data-Based Quadratic Detectors with Application to Mechanical Systems," IEEE ICASSP, 1996.

[9] A. M. SAYEED and D. L. JONES, "Optimal Detection using Bilinear Time-Frequency and Time-Scale Representations," IEEE Transactions on SP, vol. 43, pp. 2872-2883, 1995.

[10] M. SCHOENAUER, L. KALLEL, and F. JOUVE, "Mechanical Inclusions Identification by Evolutionary Computation," Revue Européenne des Eléments Finis, vol. 5, pp. 619-648, 1996.

[11] V. VAPNIK, Estimation of Dependences Based on Empirical Data. New York, USA: Springer Verlag, 1982.

[12] V. N. VAPNIK and A. Y. CHERVONENKIS, "On the Uniform Convergence of Relative Frequencies of Events to their Probabilities," Theory of Probabilities and its Applications, vol. 16, pp. 264-280, 1971.