ADAPTIVE DIFFUSION AND DISCRIMINANT ANALYSIS FOR COMPLEXITY CONTROL OF TIME-FREQUENCY DETECTORS

J. Gosme(1), *P. Gonçalvès*(2), *C. Richard*(1), *R. Lengellé*(1)

(1) Laboratoire LM2S, Université de Technologie de Troyes
B.P. 2060, 10010 Troyes Cedex - FRANCE

(2) Projet IS2, INRIA Rhône-Alpes
655 avenue de l'Europe, 38334 St Imier Cedex - FRANCE

ABSTRACT

Achieving good performance with data-driven detectors requires matching their complexity to the amount of available training data. Receivers with a too large number of adjustable parameters often exhibit poor generalization performance whereas those characterized by an insufficient complexity cannot learn all the discriminant information carried by training samples. This paper deals with the complexity control of data-driven time-frequency detectors. Our approach is based on a locally adaptive filtering technique for time-frequency representations, called adaptive diffusion. It consists in using discriminant analysis to design the conductance function that controls the diffusion process. The resulting filtering scheme preserves discriminant information while acting on the complexity of the time-frequency detector. Simulation examples illustrate the efficiency of our approach.

1. INTRODUCTION

Because time-frequency representations (TFRs) give evolutions of signals with respect to time and frequency, they have been largely used to deal with nonstationary situations. Among the host of solutions that have been proposed, Cohen class encloses bilinear TFRs that are covariant with respect to time shifts and frequency shifts. Any cohen class TFR can be interpreted as resulting of a filtering scheme applied to a central distribution, the Wigner distribution (WD). Since it satisfies intersting localisation properties, the WD has been widely used despite the presence of awkward interference terms (ITs) limiting its clarity. Various smoothing schemes have been proposed to remove these ITs. One can note homogeneous smoothing approaches [1, 2] and, more recently, locally adaptive techniques [3, 4]. The last reference presents a diffusion based technique with a conductance function that locally adapts the smoothing strength to the analyzed signal.

By virtue of their rich structure, TFRs have been extensively used for detection [5, 6]. In particular they provide an interesting point of view to control the complexity of detectors when *a priori* information on competing hypotheses is available through a small-sized learning set, as it is explained now. In such a situation, achieving good performance with a data-driven detector requires matching its complexity to the amount of training samples. This complexity is directly related to the dimension of the space spanned by training samples, or by their TFR within the context of TF based detection. Smoothing TFRs then enables to adjust the complexity of the associated TF detectors and improves its performance [7]. Recently, several signal-dependent smoothing schemes have been proposed for a better control of the space dimension reduction. In [8] for example, a smoothing kernel is optimized in order to maximize a distance between the competing hypotheses. The approach proposed in [9] takes advantage of the local adaptivity of a diffusion process. It uses diffusion of the WD of each observation x, adaptively controlled by a conductance function that only depends on a TFR of x.

Here we propose a new method for designing the conductance function. It relies on the extraction of discriminant information from the learning set, which is then used via the conductance function to preserve most of discriminant information during the diffusion process. This paper is organized as follows. First, emphasis is placed on theoretical background related to the complexity control of TF detectors. Next, our procedure for designing the conductance function is introduced. Finally the efficiency of the resulting diffusion process for obtaining TF detectors with improved performance is illustrated.

2. COMPLEXITY CONTROL OF TF DETECTORS

2.1. Principle

In order to exhibit the need for controlling the complexity of TF detectors, we consider the following problem that deals with the detection of a nonstationary signal $S(t)$ embedded in an additive noise $B(t)$:

$$\begin{cases} H_0 \;:\; X(t) \;=\; B(t) \\ H_1 \;:\; X(t) \;=\; S(t) \;+\; B(t). \end{cases} \tag{1}$$

Classical statistical detection theories lead to the fundamental result that the optimum solution to this problem consists in comparing any strictly monotonic function of the likelihood ratio to a threshold value. In many practical applications, implementing such a test may be impossible because of incomplete specification of the conditional probability densities. Therefore, we are often led to consider a simpler procedure for designing detectors. If one

This work was supported by the GdR-PRC ISIS, CNRS, France, through the program "Jeunes Chercheurs".

can collect labeled signals resulting from experimental observation, a possible alternative consists in choosing a detector structure and adjusting its free parameters according to a contrast criterion. However, adopting such an approach requires matching the detector complexity to the amount of training data for achieving good generalization performance. This has been theoretically studied in [10], where the authors define a measure of complexity for detectors, the dimension of Vapnik-Chervonenkis (VC-dim). Note that for linear classifiers, the VC-dim is equal to $N+1$, where N is the dimension of the space spanned by the learning samples, or by their TFRs within the context of TF based detection. This measure of complexity can be used to compute a confidence interval for the error probability of the designed detector d. It can be shown that the following inequality is satisfied with a probability equals to $(1-\epsilon)$:

$$|P_e(d) - P_{emp}(d, \mathcal{A}_n)| \leq E(n, V_C, \epsilon) \tag{2}$$

$$E(n, V_C, \epsilon) \equiv \sqrt{\frac{V_C}{n}\left(1 + \ln\frac{2n}{V_C}\right) - \frac{1}{n}\ln\frac{\epsilon}{4}}, \tag{3}$$

where $\mathcal{A}_n$ denotes any given n-sample learning set, $P_e(d)$ the error probability of d, $P_{emp}(d, \mathcal{A}_n)$ its estimation using $\mathcal{A}_n$, and V_C the VC-dim of d. One can see that minimizing the error probability $P_e(d)$ requires matching V_C to n. In the particular case of linear classifiers, we then have to control the dimension of the space spanned by training data.

2.2. Detectors design without complexity control

Let us consider Problem (1) with $S(t) = s(t)e^{j\phi_0}$, where $s(t)$ denotes a deterministic known signal to be detected and ϕ_0 an initial random phase uniformly distributed between $-\pi$ and π. Let B be an independent, identically distributed gaussian noise. Under these statistical hypotheses, it can be shown that an optimal detector consists in comparing the following detection statistic $\Lambda(x)$ to a threshold ν_0 [11]:

$$\Lambda(x) \;=\; \left|\sum_t x[t]\, s^*[t]\right|^2 \;\underset{d(X)=0}{\overset{d(X)=1}{\gtrless}}\; \nu_0. \tag{4}$$

Since the discrete WD satisfies the Moyal relation if its definition is properly chosen [12], this detection structure can be rewritten as a linear detector operating in the TF domain, called TF matched filter [6]:

$$\Lambda(x) = \sum_t \sum_f W_x[t,f]\, W_s[t,f] \underset{d(X)=0}{\overset{d(X)=1}{\gtrless}} \nu_0. \tag{5}$$

Let us suppose that our knowledge relative to System (1) reduces to a learning set. Since conditional probability density functions are unknown, an optimum solution cannot be determined. However, to address this problem, we have chosen to optimize a linear detector operating on the WD of the observation x

$$\Lambda(x) = \sum_t \sum_f W_x[t,f]\, a[t,f] \underset{d(X)=0}{\overset{d(X)=1}{\gtrless}} \nu_0. \tag{6}$$

Such a receiver has been widely used in the literature since it provides a flexible and meaningful quadratic decision function. The free parameters of $\Lambda(x)$, represented by the reference a, have been adjusted to maximize the Fisher criterion that is defined as:

$$\rho_{Fisher}(\eta_0, \eta_1, \sigma_0^2, \sigma_1^2) = \frac{(\eta_1 - \eta_0)^2}{\sigma_0^2 + \sigma_1^2}, \tag{7}$$

where η_i and σ_i^2 are conditional means and variances of Λ. It can be shown [13] that (5) maximizes the Fisher criterion, which means that optimizing ρ_{Fisher} with respect to (6) should lead us to $a \equiv W_s$. However, Figure 2 shows that the linear TF detector (6), optimized as explained above from 50 independent realizations of H_0 and H_1, performs poorly compared to the optimum detection structure (5). This illustrates the effect of a small-sized training set on the performance of a detector, and emphasizes the need for controlling the VC-dim of receivers during learning stages.

3. ADAPTIVE DIFFUSION FOR COMPLEXITY CONTROL OF TF DETECTORS

In order to control the complexity of the linear classifier (6), we have to adjust the dimension of the space spanned by the TFRs of the learning samples. As shown in [7], smoothing TFRs reduces the dimension of the space they spawn. Rather than smoothing blindly TFRs, we propose here an adaptive smoothing technique preserving discriminant information. This method relies on diffusion smoothing, which can act on the complexity of the designed detector, associated with a carefully chosen conductance function that preservers valuable information.

3.1. Diffused time-frequency representations

Inspired by the multi-scale analysis introduced within image processing context, adaptive diffusion has been used recently in the context of TF analysis as a locally adaptive smoothing technique [4]. Written as

$$\begin{cases} D_x(t,f;\tau=0) \;=\; TFR_x(t,f) \\ \frac{\partial D_x(t,f;\tau)}{\partial \tau} \;=\; \mathrm{div}_{t,f}(c_x(t,f)\nabla_{t,f} D_x(t,f;\tau)), \end{cases} \tag{8}$$

this iterative technique enables to adjust, at every location (t,f), the smoothing strength by mean of the conductance function denoted $c_x(t,f)$. In [4], the aim of adaptive diffusion is to remove ITs while preserving auto-components (ACs). The conductance function then has to discriminate between the presence of ITs or ACs at every location (t,f). In order to do so, the authors have proposed two approaches.

The first scheme is based on the interference-free property of the spectrogram to characterize ACs. The following conductance function enables to restrain diffusion on ACs:

$$c_x(t,f) = \left(1 + \left(\frac{S_x(t,f)}{\delta}\right)^{\alpha}\right)^{-1}, \tag{9}$$

where $\alpha \geq 0$ and $\delta > 0$ are adjustable parameters and S_x is the spectrogram of the signal x. As stressed on in [4, 9], the spectrogram also introduces its robustness to noise in the conductance function and therefore in the adaptive diffusion.

The second scheme uses the WD to be smoothed itself. Since ITs are oscillating terms, and therefore take negative values, the sign of the WD can be used in the conductance function. Areas such that $W_x[t,f] < 0$ are smoothed whereas others are preserved if ones uses the following conductance function:

$$c_x(t,f) = \chi_{(-\infty,0)}(W_x(t,f)). \qquad (10)$$

Here $\chi_{(-\infty,0)}(x)$ is a function equal to 1 when x is negative, and 0 otherwise. An possible extension of this technique consists of making the conductance function dependent on the diffusion time. This dependency can be introduced by using $D_x(t,f;\tau-1)$ instead of $W_x(t,f)$ in (10), where $D_x(t,f;\tau-1)$ denotes the result of the diffusion process at the preceding time $(\tau-1)$.

Because adaptive diffusion is an iterative technique, a stopping criterion is needed. In order to quantify the trade-off between the removal of ITs and the degradation of ACs, an entropy measure has been proposed in [4]. The process is stopped when this criterion reaches a minimum value.

3.2. Extraction of discriminant information

There are many possible techniques for extracting discriminant information from a learning set [14]. Linear discriminant analysis is frequently used in practice. This approach consists in finding the linear transformation operating on data that maximizes discriminant information measured by $J = \text{trace}(S_w^{-1}S_b)$. Here, S_w denotes the within-class scatter matrix, which shows the scatter of samples around their respective class expected vectors. The between-class scatter matrix S_b is the scatter of expected vectors around the mixture mean m_0. For L-hypothesis problems, S_w and S_b are expressed by

$$\begin{aligned} S_w &= \sum_{i=1}^{L} P_i\, E\{(X-m_i)\,(X-m_i)^T|H_i\} \\ S_b &= \sum_{i=1}^{L} P_i\,(m_i-m_0)\,(m_i-m_0)^T. \end{aligned}$$

with $m_0 = \sum_{i=1}^{L} P_i\; m_i$. In these expressions, the X's represent the observations, the P_i's denote the *a priori* probability of each hypothesis H_i, and the m_i's are the conditional expected vectors $E\{X|H_i\}$. In practice, the above-mentioned expected values are estimated from training data. For L-hypothesis problems, explanations on the maximisation of the criterion J may be found in [14]. For two-hypothesis problems such as (1), it can be shown that the linear application maximizing J consists in projecting observations onto the vector $z = S_w^{-1}(m_1 - m_0)$, which is also the eigenvector associated with the only non-zero eigenvalue of $S_w^{-1}S_b$.

3.3. Complexity control with preservation of discriminant information

As introduced previously, diffusion permits to adjust the complexity of TF detectors, and then to improve their performance. Obviously, preserving most of discriminant information provided by the learning set during the smoothing process is a condition of success. Then its seems legitimate to expect improved detection performance if one could combines discriminant analysis with adaptive diffusion.

Instead of adapting diffusion to preserve ACs while smoothing ITs, we propose to design the conductance function such that discriminant information is preserved. Let z be the vector resulting from the linear discriminant analysis of the training set. We propose to use a TFR of z for controlling diffusion at every location (t,f), e.g., the WD of z since this distribution satisfies interesting localization properties. Note that the higher $W_z(t,f)$ is, the more discriminant the information carried by $W_x(t,f)$ is and the less it should be diffused. Therefore, we suggest to use the following conductance function

$$c(W_z(t,f)) = \left(1 + \left(\frac{W_z(t,f)}{\delta}\right)^{\alpha}\right)^{-1} \qquad (11)$$

and to proceed as follows for designing a TF detector with improved performance:

1. perform a linear discriminant analysis of training data and compute the conductance function (11);
2. iterate the diffusion process for every training sample using the conductance function (11);
3. use the diffused TFRs of training samples for designing the optimum linear detector (6) with respect to the Fisher criterion;
4. estimate the probability of error of the detector. If it is decreasing, go back to Step 2. Else stop the process.

In order to illustrate this approach, we have considered the same detection problem as in Subsection 2.2. As shown in Figure 1, the lowest error rate 24% has been obtained for a VC-dim equals to 29, which corresponds to $\tau = 350$ iterations for the diffusion process. This performance must be compared to 44%, which is the error rate of the detector that has been obtained without any complexity control. Note that the latter corresponds to iteration $\tau = 0$ in Figure 1. Figure 2 presents the receiver operating characteristics (ROCs) of these two detectors and of the TF matched filter (5). One can observe that the performance of the detector with optimized complexity are closed to the performance of the optimal detector.

This demonstrate that our technique uses the smoothing action of adaptive diffusion to successfully control the complexity of the designed detector, and its adaptivity to preserve most of the discriminant information.

4. CONCLUSION

In this paper we have proposed a method of designing TF detectors with optimized complexity. The complexity control technique uses a recently introduced smoothing procedure, the adaptive diffusion of TFRs, to reduce the dimension of the space spanned by training samples. To preserve most of discriminant information available in the training set during smoothing process, we have shown that linear discriminant analysis provides valuable information for designing the conductance function that rules diffusion. We have successfully experimented our approach on simulated data. It may offer an helpful support for designing efficient detectors without prior knowledge of phenomena.

5. REFERENCES

[1] Y. Zhao, L. E. Atlas, and R. J. Marks, "The use of cone-shaped kernels for generalized time-frequency representations of nonstationnary signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 1084–1091, 1990.

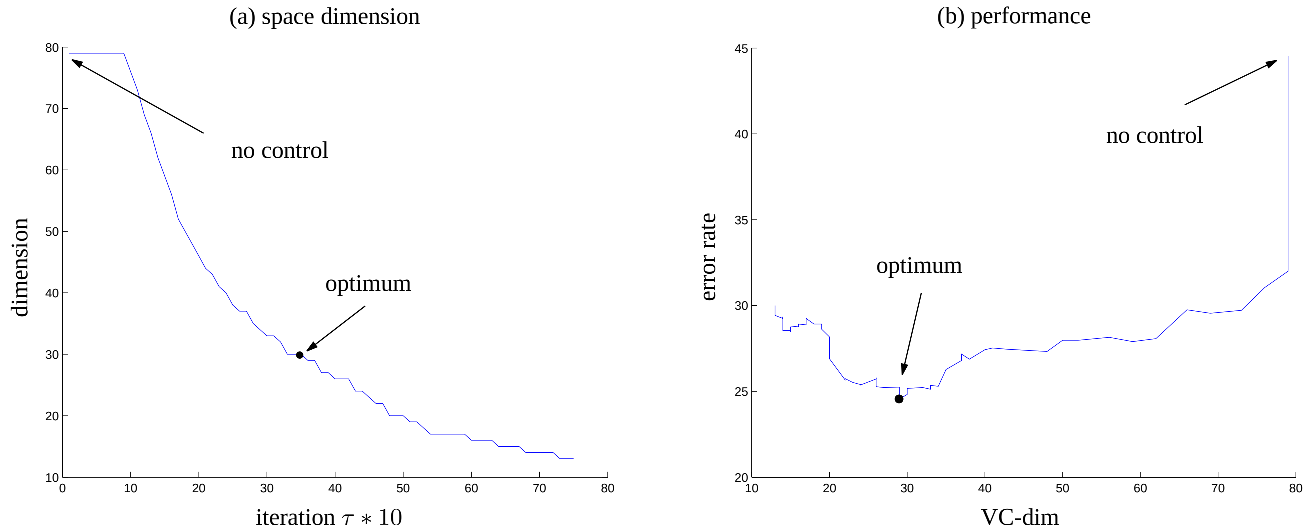

Fig. 1. Optimization of a linear TF detector from 50 realizations of each competing hypothesis (SNR = -1 dB). (a) shows the variation of the VC-dim of the detector, and therefore the dimension of the space spanned by the TFRs of learning samples, as a function of the diffusion time τ. (b) shows the variation of the error rate of the designed detector with respect to its VC-dim. The minimum error 24% is obtained for a VC-dim equals to 29. This optimum corresponds to $\tau = 350$ iterations for the diffusion process.

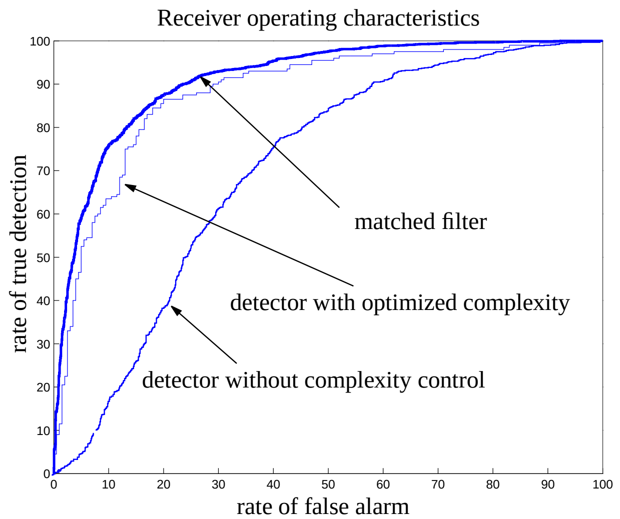

Fig. 2. Comparison of the performance of various detectors. The VC-dim of the detector with optimized complexity equals 29. It corresponds to $\tau = 350$ iterations for the diffusion process. One can observe how positively our technique for controlling complexity affects the performance of the detector.

[2] R. G. Baraniuk and D. L. Jones, “Signal dependent time-frequency representations: optimal kernel design,” *IEEE Transactions on Signal Processing*, vol. 41, pp. 1589–1602, 1993.

[3] D. L. Jones and R. G. Baraniuk, “An adaptive optimal kernel time-frequency representation,” *IEEE Transactions on Signal Processing*, vol. 43, pp. 2361–2371, 1995.

[4] P. Gonçalvès and E. Payot, “Adaptive diffusion equation for time-frequency representations,” in *Proc. IEEE Digital Signal Processing Workshop*, 1998.

[5] A. M. Sayeed and D. L. Jones, “Optimal detection using bilinear time-frequency and time-scale representations,” *IEEE Transactions on Signal Processing*, vol. 43, pp. 2872–2883, 1995.

[6] P. Flandrin, “A time-frequency formulation of optimum detection,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1377–1384, 1988.

[7] C. Richard and R. Lengellé, “Sur le contrôle de la complexité des détecteurs opérant dans le domaine temps-fréquence par le biais de la fonction de paramétrisation,” in *Proc. dix-septième colloque GRETSI*, 1999, pp. 905–908.

[8] M. Davy, C. Doncarli, and G.F. Boudreaux-Bartel, “Improved optimization of time-frequency based signal classifiers,” *IEEE Signal Processing Letters*, vol. 8, pp. 52–57, 2001.

[9] J. Gosme, C. Richard, P. Gonçalvès, and R. Lengellé, “Complexity control of time-frequency detectors using adaptive diffusion,” *Submitted to the IEEE International Conference on Acoustics, Speech, Signal Processing*, 2002.

[10] V. N. Vapnik, *Estimation of dependences based on empirical data*, Springer-Verlag, New York, 1982.

[11] H. V. Poor, *An introduction to signal detection in noise*, Springer-Verlag, New York, 1994.

[12] C. Richard, “Discrétisation des détecteurs temps-fréquence : Problèmes en découlant et eléments de solution,” in *Proc. dix-huitième colloque GRETSI*, 2001.

[13] C. Richard, R. Lengellé, and F. Abdallah, “Bayes-optimal detectors design using relevant second-order criteria.,” *IEEE Signal Processing Letters (to appear)*, 2002.

[14] K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, San Diego, 1990.