SVM CLASSIFICATION APPLYING WAVELETS TO PATTERNS HIDDEN BY NOISE

JAIME GOMEZ
Sener Ingenieria y Sistemas SA
C/ Severo Ochoa, 4. 28760 Tres Cantos. Madrid
SPAIN
jaime.gomez@sener.es

CARMEN MORATO
Depatamento de Matemática Aplicada. ETSIA.
Universidad Politécnica de Madrid
Ciudad universitaria s/n. 28040 Madrid
SPAIN
mariadelcarmen.morato@upm.es

TERESA CASTELLANOS
Depatamento de Matemática Aplicada. ETSIA.
Universidad Politécnica de Madrid
Ciudad universitaria s/n. 28040 Madrid
SPAIN
maite.castellanos@upm.es

JUAN SEIJAS
Departamento de Señales Sistemas y Radiocomunicaciones. ETSIT.
Universidad Politécnica de Madrid
Ciudad universitaria s/n. 28040 Madrid
SPAIN
juan.seijas@ieee.org

Abstract: Current approaches in signal processing for pulse detection (radar, communications, etc.) or, more general, for time series characterization, use domain transformations for frequency related information concentration. Those transformations allow extracting information that can be distinguishable (in the frequency domain) from noise. Wavelet transforms are a step forward because they work in the frequency domain keeping the time domain information. Wavelets are usually described in two main sets: continuous wavelet transforms and discrete wavelet transforms. Although they share mathematical motivation, both transformations have different algorithms and properties. The property we will focus on is time invariance. Continuous wavelets are time invariant, but they are also very expensive to calculate. Real time systems will find it hard to process all that information with enough accuracy. On the other hand, uniformly sampling the translation parameter as input to the discrete wavelet process destroys this time invariance. In this paper we will introduce experimental results that show an alternative way to generate a time invariant representation of a signal using discrete wavelet transforms. This algorithm upgrades the computational cost-efficiency of pulse detection capability with respect to the basic approach.

Key-Words: - Support Vector Machine, wavelet, computational complexity, signal detection, signal-to-noise ratio, probability of detection.

1 Introduction

Signal processing (and filtering) is an attempt to find a better expression of some information included in a bigger set of data, either by reshaping it or filtering out selected parts (those parts are named as noise). Wavelet transform is one of the most successful methods in signal de-noising processes. With the wavelet transform a time series can be viewed in multiple resolutions. Each resolution reflects a different frequency. The wavelet technique takes averages and differences of a signal, breaking the signal down into spectrum.

Wavelets applications [17] include detection of long term evolution, suppression of mixed signals, compression and de-noising, etc. In [14] a signal detection algorithm is proposed. It is based on comparing the component with maximum absolute value of its Discrete Wavelet Transform (DWT) coefficients, for a given scale, with a certain threshold. In [4] it is studied a linear Support Vector Machine as an alternative algorithm to overcome input SNR dependence.

On previous work [5,11] the authors have developed an algorithm for pulsed signal detection. This algorithm could be applied to several real-life problems like radar, signal anlysis or time series characterization, obtaining a process gain better than using previous approaches with wavelets [14]. Wavelets are generated by the scale and translation of a single prototype function called wavelet mother:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right) \quad (1)$$

where ψ is the mother wavelet, s is the scaling factor and τ is the translation factor. They are building blocks of wavelet transform for different scales and translations, just as trigonometric functions of different frequencies are building blocks of Fourier transform [12].

In the case of discrete wavelet transform (DWT), τ and s also take discrete values, given by

$$s = a_0 \quad (2)$$

$$\tau = n\tau_0 a_0^m \quad (3)$$

$$\psi_{m,n}(t) = a_0^{-m/2}\psi\left(a_0^{-m}t - n\tau_0\right) \quad (4)$$

A particular class of wavelets are orthonormal wavelets which are linearly independent, complete and orthogonal. This means that there is no "redundant" data from the original signal in more than one wavelet. In [3] Daubechies developed conditions under which wavelets form orthonormal bases. Thus the Discrete Wavelet Coefficients are the inner products of the signal and wavelet function. That is:

$$f(t) = \sum_{m,n}\psi_{m,n}(t) < \psi_{m,n}(t), f(t) > \quad (5)$$

In [7] Mallat developed a fast wavelet algorithm based on the pyramid algorithm developed by Burt and Adelson [1]. The basic components at each stage of the pyramid are two analysis filters: a low-pass filter **h** and high-pass filter **g**, and a decimation by two operation.

2 Continuous vs Discrete Wavelet

Wavelet transforms are usually described in two main sets: Continuous wavelet transforms and discrete wavelet transforms. In the first set, the input f(n), the translation factor τ and the scale factor s are continuous functions, and its transformation formulas (1) are described in the previous section. This set includes dyadic wavelets

$$\psi_{2^j,\tau}(t) = \frac{1}{\sqrt{|2^j|}}\psi\left(\frac{t-\tau}{2^j}\right), \quad (6)$$

for j being a natural number, where only the scale factor is discretized and the input f(n) is still a continuous function. It has most of the properties of basic continuous wavelets but its scale constraints allow an easier implementation of the algorithm.

On the other hand, discrete wavelet transforms have discrete values both for the input signal f(n) and the scale and translation factors. The fast orthogonal wavelet transform algorithm is used to calculate the coefficients that represent the wavelet basis functions decomposition. But in the input signal discretization and decimation process the translation-invariant property in continuous wavelet transforms is destroyed [10]. There is a loss of information throughout the whole process that avoids perfect reconstruction of the input continuous signal. There are several algorithms that try to introduce the translation invariance property on DWT. They usually involve either high computational resources or representation redundancy [2,6,8,9,13].

To obtain a quasi translation invariant representation using discrete wavelet transforms we should increase the discretization rate. At the limit, when this rate is close to infinite, then the discrete wavelet transform can be similar to a numerical approximation of a dyadic wavelet, which is translation-invariant. Given the sampling interval s • u_0 of a given scale s (if u_0 is the discretization rate) and the rate of variation of the coefficients of the transform for that scale (f • $\psi_s(t)$), if the sampling interval is large with respect to the rate of variation of the output coefficients, then it is easily observed that there is no relationship between translated functions transformations, and we will surely lose useful reconstruction information for a later pattern recognition problem [10].

On real-life engineering, increasing the sampling rate is not always an easy task. It usually depends on expensive hardware resources that may be unstable for state-of-the-art sampling rates.

In this paper we shall try an alternative way for obtaining all the signal information in the wavelet coefficients using software processing only, described in the next sections.

3 Discrete Wavelet Cycle

Suppose we have a sampled input signal, H samples size, using some appropriate u_0 sampling rate. There is a N-samples pulse, N≤H, hidden inside white Gaussian noise all over the H-samples window. Our objective is to detect and locate the pulse within the window (see [5] for a detailed description). This is a very common problem for radar as well as communications signal processing.

We are able to generate a machine-learning representation that uses the wavelet coefficients set for a N-sample window to classify it either as having a complete N-samples pulse or not. [4,11]. We are also able to generate a similar machine-learning representation that classifies a N-samples window either as having a M-samples piece of the pulse and noise on the m remaining samples (m=N-M), or not. Figure 1 shows an example.

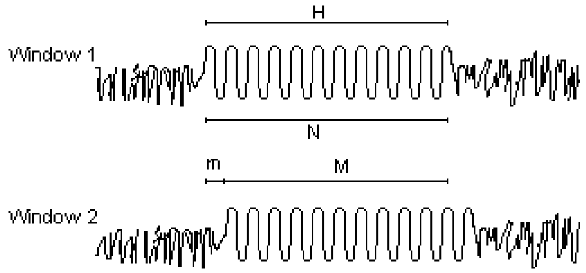

Fig. 1: Input signal windowing (somehow exaggerated for visibility purposes). The x-axis is time and y-axis is power. Window 1 is *0-shift* and window 2 is *m-shift*.

If the wavelet transformation function were translation-invariant, then both machine-learning classifiers should also be somehow translated (for instance, if we use a linear classifier we might expect its coefficients to be translated), and most important, they will have very similar performances. On the experiments section we shall see performance greatly differ between incomplete pulse windows.

The output of each m-shift classifier is a Gaussian random variable. Its mean and deviation for both positive and negative examples will provide the performance of the classifier in terms of probability of detection (P_d) and probability of false alarm (P_{fa}). In figure 2 the correlation coefficients for a subset of these random variables with respect to the complete pulse window (0-shift) is shown for the setup described on the experiments section. A high degree of uncorrelance can be observed between very close variables, i.e. between wavelet representations whose input data is less than 0.5% samples away one with respect to the other.

In this figure we can also observe there is a cycle in the correlation coefficients 16 samples long. This value is directly related to the decimation step in the discrete wavelet transform algorithm. The scale used in the experiments was $s = a^j = 2^4$.

The correlation coefficients tell us that the different random variables (we call them sources) carry different amount of information, and, as we shall see on next section, they carry complimentary information.

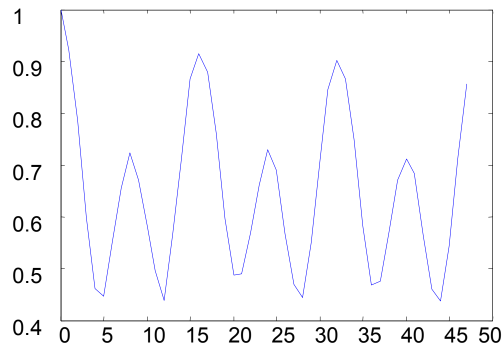

Fig. 2: Cross correlation coefficients between 0-shift random variable and m-shift random variables with m $\in$ [0,47] in a 1024 sample pulse.

4 Multiple sources classifier

Each m-shift classifier Gaussian variable has some discriminating information. If all the single sources have some amount of discriminating information that is not correlated (being translation-variant), then we might expect a joint classifier to be a better discriminator than any of the single sources alone. We define a k-multiple-source classifier as a linear classifier having as input k single-source Gaussian variables for a given input signal (positive examples) and noise (negative examples). Obviously, each single source corresponds to a different m_i value trying to detect the same event: the 0-shift window contains or not the complete pulse.

One of our objectives is to analyse the maximum improvement that can be achieved using multiple-source classifiers. The correlation cycle is directly related to the scale, so a maximum number for k is easily determined. Nevertheless, the improvement obtained every time we add a new source to the classifier is not linear.

Our second objective is to find the value of k having the best relationship between available computing resources and classification performance. The most expensive step in this algorithm is the wavelet calculation for each source, which is O(H) at a sampling rate u_0. This complexity increases linearly with the inverse of the sampling rate. A new sampling rate $u_1 = u_0 \bullet a^j$ would generate the same computing needs as a^j wavelet calculations having sampling rate u_0.

But as $k < a^j$ without a significant loss on performance, we are able to decrease up to one order of magnitude the required computational resources having a similar classification performance than a single wavelet processing with sampling rate u_1.
Therefore, we could assert that the advantage obtained by increasing the sampling rate can be achieved by software processing using around one order of magnitude less computational resources.

5 Experimental results

In the experiments shown in this paper we have used the same setup as in previous work [11]. The main features of this setup are: Daubechies 5 mother wavelet calculated on a 1024 samples chirp pulse, using scale d4. We used Support Vector Machines [15] as the machine-learning tool, with a linear kernel (i.e. a linear classifier). The use of SVM on this paper has not been deemed relevant, because the special properties we used form this state-of-the-art algorithm were already discussed in [5]. However, the implementation of SVM as the machine learning tool provides better performances than any other Neural Network learning algorithm like those proposed by [16]

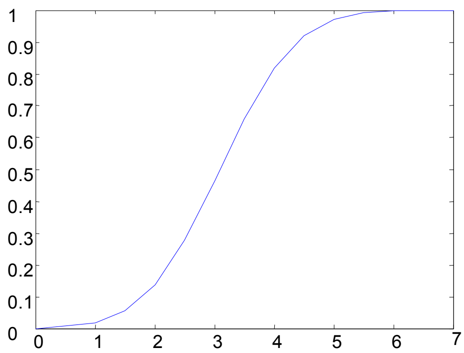

Fig. 3: The *mean difference* axis transformation. The y-axis is P_d while x-axis is the mean difference (see text).

In the figures shown in this section, the performance measure is not described in terms of probability of detection, but in a closely related function. We define a "mean difference" as the distance between the means of the signal Gaussian distribution and noise Gaussian distribution. We translate the P_d obtained through the machine learning process to two different distributions with deviation 1 for a $P_{fa} = 10^{-3}$. It is just a simple y-axis scale change. The reason for this change is to have a better description of the effect of using multiple sources in limit values, as when the probability of detection approaches 100%. For instance, if the mean difference is 0, then $P_d = P_{fa}$. Figure 3 shows the relationship between both measures.

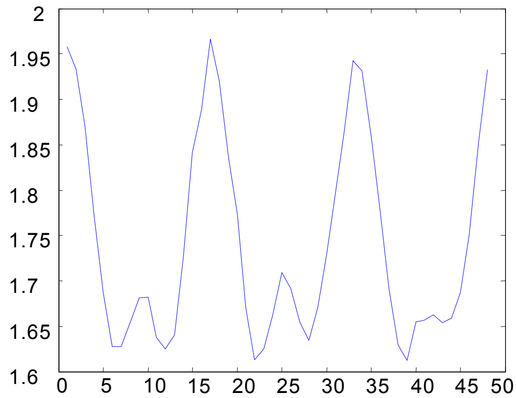

Fig. 4: Performances for each source alone form 0-shift to 47-shift (x-axis) measured as the mean difference (y-axis).

In figure 4 we can observe the performance of each source alone. One of the first unexpected results is that the 0-shift source, i.e. the wavelet transformation of the complete pulse window, is not performing best. One of the peaks on the other cycles is performing better. There is an obvious random effect in the modelling process and so these results are not necessarily extended to other experiment setups. But still, this case is most probably not an exception.

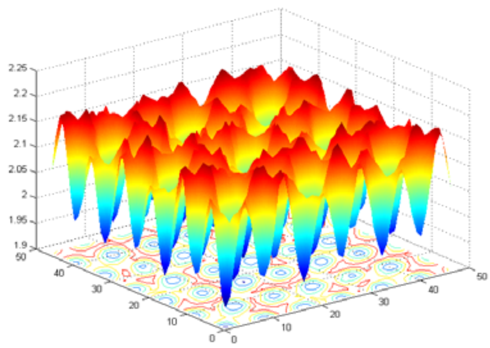

Fig. 5: Performances for the integration of three sources, one of them is 0-shift and the other two vary from 0 to 47-shift, measured as the mean difference. Best value is x=5, y=27, with z=2.1867.

In figure 5 an example is shown around the multiple source integration performance. It calculates the performance, measured as the mean difference for three sources, one of them fixed to the 0-shift source, and the other two varying from 0 to 47-shift. Note that the function is almost symmetric along the main diagonal, only small random effects because of the stochastic machine learning algorithm prevent the function from being completely symmetric. The best pulse detection set is {0,5,27}, confirming that first

cycle sources (those having better pulse information input) are not necessarily the ones having better performance, neither alone nor together with other sources.
The sources being best at one experiment will still be so when integrating more sources, as expected.
On figure 6 the best performance values for each multiple source is shown. Note that the behaviour of the function finds a limit after a small number of sources are added to the integration step. Moreover, in our experiments multiple source integration with more than 8 sources did not beat this last value because of stochastic machine learning behaviour. There is no doubt that non-linear classifiers will increase performance in those more complex cases. For this test setup, and available resources, the best k values were around $2 \leq k \leq 3$.

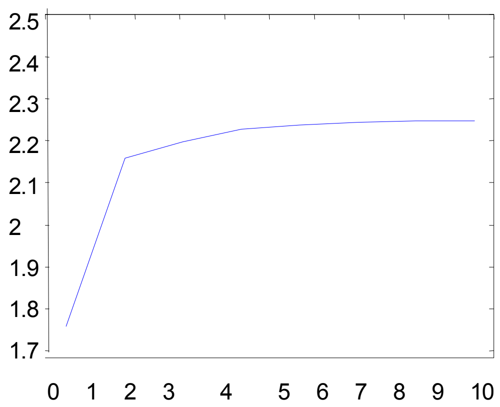

Fig. 6: Best multiple source linear integration performance from single source to 8 sources (x-axis), measured as mean difference (y-axis). The highest value is 2.2431.

6 Conclusions and Future Work

Radar emitter classification is a subset of the data clustering problem which tries to discover some hidden structure in the input data. Most of the algorithms applied for solving this problem fit into two main categories: Statistical modeling and Neural Networks. We have proposed a novel approach where the neural networks paradigm is improved by the SVM and in this paper we have shown that additional uncorrelated information can be found in discrete wavelet algorithms when processing translated input.
The number of additional sources of information that are useful in an integration step has a limit. This limit is related with the number of decimations executed to obtain the final representation, i.e. with the scale. Our results show that the use of this approach increases the system detection capability as much as the difference between the best and the worst single source wavelet description.
The next step is to compare this algorithm to a single source discrete wavelet transform with an increase on the sampling rate with different approaches. We expect the results will be fairly similar, using linear and non-linear classifiers.
Then we will be able to analyse mathematical similarities between the discrete and dyadic wavelets, and the way to achieve computational cost-efficient discrete translation invariant transforms.

References:

[1] P.J. Burt, The Laplacian pyramid transforms for image coding, *IEEE Transactions on Communications*, No 31, 1983, pp532-540.
[2] I. Cohen, S. Raz, D. Malah, Orthonormal Shift-Invariant Wavelet Packet Decomposition and Representation, *Signal Processing*, Vol 57, 1997, pp 251-270.
[3] I. Daubechies, Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math*., Vol 41, 1988, pp 909–996.
[4] J. Gomez, I. Melgar, J. Seijas, D. Andina, Sub-Optimum Signal Linear Detector Using Wavelets and Support Vector Machines, *WSEAS Trans. on Communications*, issue 4, vol 2, 2003, pp426-431.
[5] J. Gomez, I. Melgar, J. Seijas, Wavelet Time Shift Properties Integration with Support Vector Machines, *LNAI-3131 Modeling Decisions for Artificial Intelligence*, 2004, pp49-59.
[6] R. Kronland-Martinet, J.Morlet, A.Grossman, Analysis of sound patterns through wavelet transforms, *Int. J. Pattern Recognition and Artificial Intelligence*, Vol 1, 1987, pp 273-301.
[7] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol 11, 1989, pp 674–693.
[8] S. Mallat, Zero crossing of a wavelet transform, *IEEE Trans. Information Theory*, Vol 37, 1991.
[9] S. Mallat and Z.Zhang, Matching Pursuit with time-frequency dictionaries, *IEEE Trans. Signal Processing*, Vol 41, 1993.
[10] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.
[11] I. Melgar, J. Gomez, J. Seijas, Optimum Signal Linear Detector in the Discrete Wavelet Transform-Domain, *WSEAS Trans. on Communications*, issue 3, Vol 2, 2003, pp253-258.
[12] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill 1984.

[13] J.C Pesquet, H. Hrim, H. Carfantan, Time-Invariant Orthonormal Wavelet Representations, *IEEE Trans. Signal Processing*, Vol 44, 1996, pp 1964-1970.

[14] J. Torres, A. Vega, S. Torres, D. Andina, Chirp Detection Through Discrete Wavelet Transform, *Proceedings of ISPRA'02*, 2002, pp1971-1975.

[15] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

[16] Jun Liu, P. Jim Y. Lee, Lingjie Li, Zhi-Quan Luo and K. Max Wong, Online Clustering Algorithms for Radar Emitter Classification, *IEEE Trans. on Pattern Analysis and Machine Learning*, Vol 27, No 8, 2005, pp 1185-1196.

[17] C, Chui, *Wavelet Analysis and its Applications*, *Vol. 1*, Academic Press, Boston, MA, 1992.