OPTIMIZATION OF TIME AND FREQUENCY RESOLUTION FOR RADAR TRANSMITTER IDENTIFICATION

Bradford W. Gillespie and Les E. Atlas

Interactive System Design Laboratory, Department of Electrical Engineering, University of Washington
Seattle, Washington 98195-2500, USA

ABSTRACT

An entirely new set of criteria for the design of kernels for time-frequency representations (TFRs) has been recently proposed [1], [2], [3]. The goal of these criteria is to produce kernels (and thus, TFRs) which will enable accurate classification without explicitly defining, *a priori*, the underlying structure that differentiates individual classes. These kernels, which are optimized to discriminate among multiple classes of signals, are referred to as signal class-dependent kernels, or simply class-dependent kernels. Until now, our technique has utilized the Rihaczek TFR as the base representation, deriving the optimal smoothing in time and frequency from this representation. Here the performance of the class-dependent approach is investigated in relation to the choice of the base representation. Classifier performance using several base TFRs is analyzed within the context of radar transmitter identification. It is shown that both the Rihaczek and the Wigner-Ville distributions yield equivalent results, far superior to the short-time Fourier transform. In addition, a correlation reduction step is presented here. This improves performance and extensibility of the class-dependent approach.

1. INTRODUCTION

Recently, we have devised a method that allows the classification task to, given adequate and representative training data, ascertain the relative role of time and frequency resolution in classification. This allows the optimally smoothed time-frequency representation (TFR) to be constructed.

Our approach is based on the premise that automatic detection and classification systems should be provided with only enough input resolution to achieve needed performance. Namely, resolution that is too great will potentially require a large detector or classifier training set and will be sensitive to irrelevant features and/or noise. Large dimensionality detectors and classifiers are also computationally expensive and slow. It should be noted that we are not referring to or bound by implicit Heisenburg or window-related resolution limitations — we are instead explicitly limiting the resolution to optimize accurate classification.

To generate this optimally smoothed TFR a base representation is used to determine the optimal smoothing kernel. Therefore, this approach can be thought of as determining an optimal smoothing kernel *given* the base representation. Until now, our technique has utilized the Rihaczek TFR as the base representation. Here the effect of this choice is explored within the context of radar transmitter identification.

The goal of radar transmitter identification is to determine the particular transmitter from which a signal originated, using *only* the received waveform. No localization information is exploited to accomplish this task. Each transmitter must be identified in the presence of other transmitters of the same type (*i.e.* same model number but different serial number). Individual transmitter identification can be accomplished by exploiting the unintentional modulation present in these radar signals. This modulation is a result of subtle variations between particular transmitter components, and acts as a signature for an individual radar station.

2. BACKGROUND

Modern TFR research often begins by selecting a kernel (*i.e.* generating function) $\Phi[n,\tau]$ that operates upon an instantaneous autocorrelation function:

$$R[n,\tau]=\sum_{n'=n-N}^{n+N} x[n']x[n'+\tau]\,. \tag{1}$$

The resultant TFR, $P[n,k]$, arises from the discrete Fourier transform (in τ) of the result of multiplying the kernel (in τ) and convolving the kernel (in n) with the instantaneous autocorrelation function, $R[n,\tau]$. As an alternative, a discrete Fourier transform (in n) can be applied to the instantaneous autocorrelation function, to yield an ambiguity function:

$$A[\eta,\tau]=\boldsymbol{F}_n\{R[n,\tau]\}=\sum_{n=0}^{M-1} R[n,\tau]\mathrm{e}^{-j\frac{2\pi}{M}n\eta}\,. \tag{2}$$

There is an equivalent kernel, $\phi[\eta,\tau]$, which operates multiplicatively in both dimensions upon the ambiguity function, $A[\eta,\tau]$. These two kernels are also related by a discrete Fourier transform (in n):

$$\phi[\eta,\tau]=\boldsymbol{F}_n\{\Phi[n,\tau]\}=\sum_{n=0}^{M-1} \Phi[n,\tau]\mathrm{e}^{-j\frac{2\pi}{M}n\eta}\,. \tag{3}$$

Any non-zero extent of $\phi[\eta,\tau]$, in η and/or τ can effect a smoothing on $P[n,k]$ in time and/or frequency respectively. For example, if $\phi[\eta,\tau]=0$ for all values except those on the $\eta=0$ axis, then all temporal information are smoothed and only steady-state frequency information is retained in $P[n,k]$.

In past time-frequency research, kernels for a number of properties, such as finite-time support and minimizing quadratic interference, have been derived. Although some of these representations may offer advantages in classification of certain types of signals, the goal of sensitive detection or accurate classification has not been explicit. The ability of the aforementioned kernel to reduce time and/or frequency resolution, embodied within

This work was supported by a grant from the Office of Naval Research (N00014-97-1-0082). The authors would like to thank Dr. Victor Chen (Naval Research Laboratory) for his assistance in this work.

the explicit goal of optimal classification (*i.e.* minimum overall probability of error), is the basis for the approach outlined below. When the kernel, $\phi[\eta,\tau]$, is designed with the goal of optimal classification we refer to it as the *signal class-dependent kernel*, or simply *class-dependent kernel*. Furthermore, we refer to the corresponding TFR as the *class-dependent TFR*.

3. OUR APPROACH AND METHODS[1]

Data provided by the Naval Research Lab (NRL) was utilized in this work [5]. This data set contains ten radar pulses from four transmitters. This data comprises three tests from each of the four sources called A2, CCC2, F2, and H2. These will be denoted as class one through four. Each pulse contains 180 complex samples (*i.e.* in-phase and quadrature components). Example time series and TFRs of this data are given in [4].

In order to experimentally study the class-dependent approach, N-fold cross-validation was used [6]. The data were randomly divided into nine training examples and one test example for each of the four classes. Training and testing was performed. This process was repeated and the results averaged, to yield an honest performance estimate of the system.

Our approach is a modification of the signal class-dependent method that has been described in more detail before [1], [2], [3], [4]. The previously described approach finds the single kernel, $\phi[\eta,\tau]$, which maximizes the distance, in a mean-square sense, between the estimated ambiguity functions for each of C different classes. Defining a kernel matrix as $\boldsymbol{\phi}=\phi[\eta,\tau]$ and an ambiguity matrix for class c as $\boldsymbol{A}_c = A_c[\eta,\tau]$, the kernel is selected to satisfy:

$$\arg\max_{\boldsymbol{\phi}} \left\{ \sum_{c'=1}^{C} \sum_{c''=c'+1}^{C} \left\| \boldsymbol{\phi}\circ\boldsymbol{A}_{c'} - \boldsymbol{\phi}\circ\boldsymbol{A}_{c''} \right\|_2^2 \right\}. \tag{4}$$

where $\circ$ represents the Hadamard product (*i.e.* an element-by-element product).

In practice, this maximization is accomplished by rank-ordering the kernel points according to Linear Fisher's Discriminant Ratio:

$$\mathrm{FDR}[\eta,\tau] = \frac{\sum_{c'=1}^{C}\sum_{c''=c'+1}^{C}\left(\mu_{c'}^{\eta,\tau} - \mu_{c''}^{\eta,\tau}\right)^2}{\sum_{c=1}^{C}\left(\sigma_c^{\eta,\tau}\right)^2} \tag{5}$$

where:

$$\begin{aligned} \mu_c^{\eta,\tau} &= \frac{1}{I}\sum_{i=1}^{I} A_c^i[\eta,\tau] \\ \sigma_c^{\eta,\tau} &= \frac{1}{I}\sum_{i=1}^{I}\left|A_c^i[\eta,\tau]\right|^2 - \left|\mu_c^{\eta,\tau}\right|^2 \end{aligned} . \tag{6}$$

$A_c^i[\eta,\tau]$ is an element from the ambiguity function of the i^{th} training example from class c. For actual classification of an unknown time series, the ambiguity function is multiplied, in η and τ, by a binary kernel mask, which is set to "1" at one optimal and, optionally, subsequently lower-ranked kernel points (often required in practice).[2] These kernel points, depending on their locations, effect a smoothing in time and/or frequency of the unknown data. The smoothed version is then compared to a smoothed representative from each class, derived during training. As an added benefit, the class-dependent ambiguity function $(\boldsymbol{\phi}\circ\boldsymbol{A}_c)$ can be transformed into a class-dependent time-frequency representation $CD_c[n,k]$. The implicit time-frequency smoothing can then be viewed.

The optimal number of kernel points (*i.e.* number of points set to "1" in the binary kernel mask) is determined by evaluating the classifier performance using the K best kernel points (*i.e.* the K points with the largest FDR). K_{opt} is selected to be the K for which the probability of overall correct classification is greatest.

We have recently found that performance of this class-dependent technique can be improved by excluding those kernel points that are strongly correlated with higher ranked kernel points. Ideally, full ranking in multiple feature dimensions (*e.g.* Procrustes angle [7], Fisher's Linear Discriminant [8]) would be preferable to the combination of FDR followed by correlation removal. Due to the number of available features (4096 for a 64 by 64 ambiguity function) the ideal approach is computationally prohibitive. Thus, this sub-optimal approach is employed. The incorporation of this correlation reduction differs from our earlier approach [4].

To classify a particular unknown test signal, an ambiguity function is estimated from the signal. After masking with the previously determined binary kernel, the class of the unknown signal is estimated via a maximum likelihood (ML) detector [9] (classifier). To simplify the ML detector, the underlying data is assumed Gaussian. Under this assumption, the ML detector is the multivariate Gaussian classifier (MVG) [10]. The mean and covariance statistics of the selected K points for each class (utilized by MVG) are estimated from the training data.

PREPROCESSING

Before classification, each data segment is preprocessed. First, each radar pulse is individually demeaned and then normalized to a standard deviation of one, in order to prevent classification based on irrelevant or variable features. The second step is necessitated by the particular classification technique employed herein. The center frequency of the transmitter is a variable parameter. Because our method seeks to find a time-frequency representation that maximizes between-class separation, if a particular class in the training set contains a center frequency bias, this will be used as an essential class discriminator. The variability of this parameter makes this unusable as a discriminatory feature. There are three possible solutions to ensure this feature is not incorporated into the classifier.

- Given enough representative data from each transmitter (presumably including variability in the center frequency) the classifier will discard this feature as a possible means of classification. This is equivalent to the classifier "learning" the center frequency is irrelevant.

[1] This approach has been previously proposed in [4], and will serve as the basis to analyze a variety of base distributions here.

[2] This binary mask is selecting, in effect, "features" from the set of points that make up the ambiguity function.

- Only the magnitude of the radar pulse is used for classification. This presumes that there is enough information in the envelope of the radar signature to discriminate classes. We have found that this is not the case.
- The data set is preprocessed to modulate all pulses to the same center frequency. This involves estimation of the center frequency of each pulse and modulation to a new predetermined center frequency.

Due to the size of the data set provided, the latter method is preferred. The large signal to noise ratio of this data makes estimation of the center frequency of the signal relatively easy. It was determined that 34 out of the 40 examples had a center frequency of $0.151(2\pi)$ radians per second while 6 pulses (all from class one) had a center frequency of $0.145(2\pi)$ radians per second.

The selected preprocessing algorithm for this data was to modulate all signals to a center frequency of $0.151(2\pi)$ radians per second. Once this preprocessing algorithm is applied to the data, transmitter identification is implemented as described above.

4. EFFECT OF BASE REPRESENTATION ON OVERALL PERFORMANCE

It is important to note that the kernel for optimal separation (ϕ) maximizes the time-frequency difference *given* the base representation ($\boldsymbol{A}_c$). In our previous work, the Rihaczek ambiguity function has been used as the base representation. The Rihaczek was initially selected because this distribution is the simplest form for the base representation [1]. This intuitively appears to be a desirable property for the base distribution, however no theoretical reason exists to prevent the use of other distributions.

Here, the effect the base representation will be investigated. The performance of the classifier using the Rihaczek ambiguity function as the base representation is compared to the performance obtained with the same classifier using the Wigner-Ville (WV) [11] and short-time Fourier transform (STFT) base representations.

For brevity's sake, the three classifiers studied here will be denoted as:

- **CD-R**: Class-dependent classifier derived from the Rihaczek ambiguity function,
- **CD-WV**: Class-dependent classifier derived from the WV ambiguity function, and
- **CD-STFT**: Class-dependent classifier derived from the STFT ambiguity function.

5. RESULTS

In the provided data set, all examples are time-aligned precisely. Furthermore, the provided data set has approximately infinite signal to noise ratio (SNR). In practice, neither condition can be assured. Therefore, performance of the class-dependent technique using the three proposed base distributions in the presence of noise and timing jitter will be investigated. Four cases presented here are:

Case 1. classification using the original data which contains precise time alignment between all examples,

Case 2. classification using the original data with timing jitter uniformly distributed over the interval ± one sample,

Case 3. classification using the original data with additive white Gaussian noise (AWGN) yielding an SNR of 14 dB, and

Case 4. classification using the original data with timing jitter uniformly distributed over the interval ± one sample and AWGN yielding an SNR of 14 dB.

In all cases, it was found that the 64 by 64 point ambiguity function provided the best performance for CD-R and CD-WV. The CD-STFT performed best using a 64 (τ) by 4 (η) point ambiguity function. For the STFT, a non-overlapping rectangular window was used on the data. A correlation threshold of 0.95 was used to determine what features to exclude. These settings have been used in all results provided in this paper.

CASE 1 — *Ideal data set*.

In the original data set, all the radar pulses are perfectly time-aligned with respect to the envelope of the signal. Using CD-R 100% overall correct classification was achieved. This was achieved using 1 to 6 kernel points, demonstrating the robustness of the Rihaczek base representation. Using CD-WV provided essentially the same classification performance. CD-STFT's performance was still acceptable, achieving 91.12% correct classification on this data set. The rank order curves for each of these classifiers are shown in Figure 1 (a). Using CD-R or CD-WV yields a more robust classifier than CD-STFT on this idealized data set. This result foreshadows the performance over the three remaining cases.

CASE 2 — *Timing jitter*.

Timing jitter uniformly distributed over the interval ± one sample was introduced. With jitter, CD-R still achieved perfect performance: 100% overall correct classification. Again using CD-WV provides essentially equivalent performance. CD-STFT showed significant degradation from Case 1, with a maximum overall correct classification of 86.75%. The rank order curves for each of these is shown in Figure 1 (b). CD-R and CD-STFT again are more robust classifiers than CD-STFT.

In comparison with our earlier approach [4], CD-R performance has increased. This is due to the incorporation of the correlation reduction step discussed earlier.

CASE 3 — *Noise*.

AWGN was introduced to the original data to lower the SNR to 14 dB. Again, CD-R and CD-WV performed significantly better than CD-STFT, providing 95.75% correct classification. Performance using CD-STFT suffered severely, overall correct classification achieved a meager 44.38%. The rank order curve for each of these is shown in Figure 1 (c).

CASE 4 — *Timing jitter and noise*.

With the addition of noise and misalignment, the performance of CD-R and CD-WV drop to approximately 92% correct classification. As expected from prior results, CD-STFT only achieved 42.75% overall correct classification. The rank order curves for each of these is shown in Figure 1 (d).

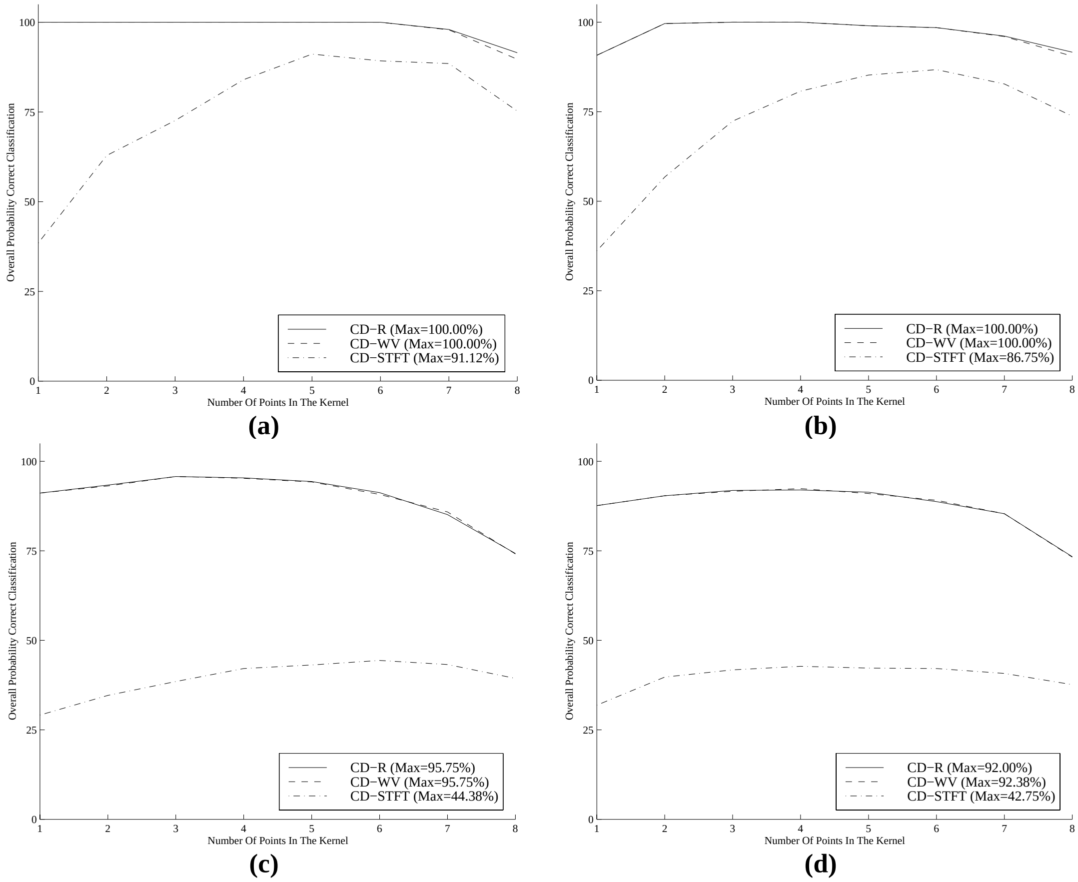

Figure 1. Rank-order curves for (a) Case 1, (b) Case 2, (c) Case 3, and (d) Case 4. In many instances the curves for CD-R and CD-WV overlap.

6. CONCLUSIONS

The class dependent approach has been demonstrated using the Rihaczek, Wigner-Ville and short-time Fourier transform ambiguity functions. The Rihaczek and Wigner-Ville are equivalent projections for classification in the ambiguity plane. The reason the Rihaczek and Wigner-Ville yield identical results is clear once the formulation of both in the ambiguity plane is analyzed. The Rihaczek differs from the Wigner-Ville only by a modulating complex exponential. This modulation has no bearing on separability and hence no bearing on classification performance.

On the other hand, the short-time Fourier transform is clearly an inferior projection. We surmise this is due to the correlation present in this distribution.

7. REFERENCES

[1] J. McLaughlin, J. Droppo, and L. Atlas, “Class-dependent time-frequency distributions via operator theory,*” ICASSP ‘97*, **3**, 1997.

[2] L. Atlas, J. Droppo, and J. McLaughlin, “Optimizing time-frequency distributions via operator theory,” *Proceedings of the SPIE*, **3162**, 1997.

[3] J. McLaughlin and L. Atlas, “Applications of operator theory to time-frequency analysis and classification,” to appear in *IEEE Trans. on Signal Processing*, 1998.

[4] B.W. Gillespie and L.E. Atlas, “Optimization of Time and Frequency Resolution for Radar Transmitter Identification,” To appear in *Proceedings of the SPIE,* 1998.

[5] V.C. Chen, “Time-frequency/time-scale analysis for radar applications,”*http://airborne.nrl.navy.mil/~vchen/tftsa.html*.

[6] L. Breiman, J.H. Freidman, R.A. Olshen, and C.J. Stone, *Classification and Regression Trees,* Wadsworth, Belmont, California, 1984.

[7] G.H. Golub and C.E. Van Loan, *Matrix Computations*, Johns Hopkins, Baltimore, 1989.

[8] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, 1973.

[9] L. L. Scharf, *Statistical Signal Processing*, Addison Wesley, Reading Massachusetts, 1991.

[10] D.H. Kil, and F.B. Shin, *Pattern Recognition and Prediction with Applications to Signal Characterization*, AIP Press, Woodbury, New York, 1996.

[11] M.S. Richman, T.W. Parks, and R.G. Shenoy, “Discrete-Time, Discrete-Frequency, Time-Frequency Analysis,” *IEEE Trans. on Signal Processing*, **46**(6), 1998.