33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

DISCRIMINATIVE TIME-FREQUENCY KERNELS FOR GAIT ANALYSIS FOR AMYOTROPHIC LATERAL SCLEROSIS

Lakshmi Sugavaneswaran, Karthikeyan Umapathy and Sridhar Krishnan

***Abstract*— Many stochastic systems show certain trends which in turn govern their underlying non-stationary time varying behavior. In order to facilitate efficient quantification of such signals, their analysis necessitates the use of robust tools for discerning between different classes of data. Research show that, use of time-frequency techniques offer intelligible representations for non-stationary signals, along with facilitating computation of instantaneous parameters. Further, in order to obtain efficient discrimination machine learning (ML) modules are often used alongside suitable representation techniques. In this work, we exploit the concepts of ML-kernel functions directly by incorporating them in the ambiguity time-frequency (TF) space, thereby obtaining a one-step discrimination between different non-stationary patterns. The proposed technique is evaluated for quantification applications for gait signal analysis. An overall classification accuracy of 93.1% is reported for the neurological gait database consisting of signals from 16-control and 13-amyotrophic lateral sclerosis (ALS) subjects. Results indicate that this scheme offers great potential in designing robust tools for time-varying signal analysis.**

***Index Terms*— Ambiguity Domain, machine learning, kernels, non-stationary**

I. INTRODUCTION

Most real-life signals carry time-varying characteristics, making their analysis relatively complex and difficult. This necessitates the use of efficient schemes, Research in time-frequency analysis is targeted towards two main objectives: (i) to define the time-varying spectrum suitably to facilitate analysis and (ii) to provide efficient computation of the parameters governing the signal characteristics (such as instantaneous features).

Owing to the reasonable performance accuracy, the TF domain is invariably exploited by most of the existing non-stationary signal analysis schemes. One of the available variant of TF representation is obtained using the ambiguity function (AF) which in-turn characterizes the corresponding time-frequency distribution (TFD) (equations 1). Its corresponding weighting factor (or the kernel $\phi(\theta, t)$) controls the tradeoff between cross-term suppression and the maximum achievable TF localization.

Cohen's class of quadratic TFD's are given by [1],

$$C(t, f) = \frac{1}{4\pi^2} \iint M(\theta, t) e^{-tj\theta} e^{-\tau jf} d\theta d\tau \qquad (1)$$

where, $M(\theta, t) = \phi(\theta, t)\ .\ A(\theta, t)$. Here, $M(\theta, t)$ represents the characteristic function that differentiates one TFD from another and is given by the product of the kernel function (ϕ) and the AF. The AF is a two-dimensional function of time delay and Doppler frequency. Due to the specific positioning of the auto- and cross-terms [2], this domain has been successfully employed for applications such as blind source application [3], interference reduction [4], kernel design for frequency-modulated signals [5], biomedical signal representation [6], for detection and estimation applications [4], [7]. The useability of this domain for signal quantification has been effectively demonstrated for non-stationary pathological speech database [8].

The present study addresses the issue associated with non-availability of robust techniques for efficient characterization of time-varying biomedical signals. Here, we investigate the effectiveness of incorporating machine learning principles for design of TF kernels. The rest of the paper is arranged as follows. Section II highlights the significance of the ambiguity domain (AD) and discusses the methodology employed in the design of proposed ML-based TF kernels in detail. In Section III, the performance evaluation of the proposed scheme is elaborated. Section IV concludes the paper emphasizing the main contributions of this study and provides some future direction. Throughout the scope of this article, the TF and the ML-based kernels are represented by k_{tf} and K_{ML} respectively.

II. METHODOLODY

Our investigation to cross-validate the significance of this domain indicates that the auto-terms are concentrated around the origin and the cross-terms are spread away from the origin in AD. Also, AD being a correlative domain, it can be suitably used to capture the short-term and long-term correlations for a time-series signal. Owing to these characteristics, we base our kernel design in the AD-space.

A. Problem Statement

From a signal processing perspective, there can be atleast two distinct kernel design motives:

(i) k_{tf}: kernel(s) that act like a filter and provides a suitable representative space for visual analysis of cross-term free TFDs discriminative feature vectors and

(ii) K_{ML}: kernel(s) that are employed in a classifier in order to achieve higher distance of separability between different classes of signals.

The main objective of this research work is to efficiently design a TF kernel in the AD in an effort to obtain an efficient separation among different signal classes and, in

L. Sugavaneswaran, K. Umapathy and S. Krishnan are with the Department of Electrical and Computer Engineering, Ryerson University, 350 Victoria Street, Toronto, Ontario, M5B 2K3, Canada, E-Mail: {lsugavan,karthi,krishnan}@ryerson.ca

turn, to deduce an optimal feature space with a reduced number of computation steps. A general block diagram of the proposed approach is shown in Figure 1. Here, the signal is first transformed into the AD-space and the mapped coefficients are masked selectively using k_{tf}. The masked coefficients are then used to define the feature vectors. For signal discrimination applications, this feature space is then used in integration with a suitable module (such as pattern classifier). In the next subsection, the proposed ML-based kernels (k_{tfML})are discussed in detail.

B. Proposed Kernel Design

Investigation reveals that TF and most ML-based kernels share certain significant and similar prerequisites in order for them to be ably perform in their allotted domains of action. In this work, some of their common properties are exploited in the kernel design. Firstly, the TF and the ML kernels needs to satisfy non-negativity and Mercer's theorem respectively in most cases. Mercer's theorem states that an n x n matrix (M) is said to be positive definite if and only if for all non-zero vectors 'z' (and z^* represents the conjugate transpose), i.e.

$$Real(z^* M z) > 0. \tag{2}$$

The generated set of cross-(or interfering) terms is placed away from the origin in AD due to the symmetric nature of this correlative domain. Here, the number of non-zero elements depends on the amount of frequency interaction components for each given signal. And owing to the inherent symmetry among the mapped coefficients (in AF) Mercer's condition is satisfied. In addition to Mercer's theorem, the kernel design in this domain can be interpreted as smoothed versions of the Wigner distribution and hence shows interesting peculiarities regarding TF localization. In this work, we have proposed the use of two sets of ML-based kernels: (a) using discriminative kernels and (b) kernel design using generative versions of AD-map.

(a) Discriminative TF kernels: For any given input space, $X \subset R^d$, with input vectors 'x' and 'z', the kernel is defined as,

$$K(x, z) = \langle \phi(x), \phi(z) \rangle \tag{3}$$

where, ϕ is a non-linear (or a linear) map from the input space X to the feature space F, and $\langle ... \rangle$ is an inner product. Usually such a kernel is defined explicitly, thereby defining the map and the feature space. Such a representation is also symmetric and reversible, $K(x, z) = K(z, x)$ and satisfies Cauchy-Schwartz inequality,

$$K^2(x, z) \leq K(x, x) K(z, z) \tag{4}$$

The main idea underlying the choice of a ML kernel is that the derived kernel would incorporate the inherent available non-linearity in feature extraction phase, thereby maximizing the separability between classes. The ML-based kernel is then extended for defining the TF distribution, so that a better representation of time-varying features is achieved. One such kernel that provides maximum separation, commonly referred to as the all-subset or the polynomial kernel is used in this work. The polynomial kernel [9] is represented as,

$$K(x, z) = (\alpha + x.z)^d \tag{5}$$

where α is any scalar quantity and d is the order of the kernel.

Selection of parameters: The choice of 'α' and 'd' parameters are governed by AD-map of the signal. To define these values, the AD-map is divided into bands based on relative energy distribution ratios [8]. This ratio provides the relative energy between auto- and cross-term distributions from the AD-map. From this representation, the region of auto- and cross-term overlap is approximately identified using the energy spread factor and the maximum energy distributed between successive bands. That is, in AD the energy map usually resembles a bell-shaped distribution and in regions where there are more cross-terms, the maximum energy decreases alongside a decrease in the signal spread. Once the region (band number 'n' from the origin) is obtained, the maximum energy for all signals under consideration is computed. thereafter the median value obtained from the ranked maximum energies, is then assigned to 'α' and the parameter 'd' is computed as 2^n.
Having defined the parameters, the kernel function is then incorporated in the characteristic function, thereby defining a new TFD for non-stationary signal analysis. The autocorrelation function or the AF of the TFD is defined as,

$$A(\theta, t) = \int x(t - \frac{\tau}{2}) x^*(t + \frac{\tau}{2}) e^{-j\theta t} dt \tag{6}$$

By substituting (5) and (6) in the characteristics equation,

$$M_{ML}(\theta, t) = \int K \left\{ \left(x(t + \frac{\tau}{2}), x^*(t - \frac{\tau}{2}) \right) \right\} e^{-j\theta t} dt \tag{7}$$

The corresponding TFD can then be obtained by computing the Fourier transform of (7). Similarly, AD-based kernel expressions were deduced for two other widely used SVM kernel functions, namely the linear and the Gaussian kernel.

(b) Generative TF kernels: The second set of kernel is defined by exploiting the different generative models used in ML approaches. Since AD is a correlative energy domain, in order to efficiently control the influence of data lengths and spectral spread, we propose to define a kernel that uses the normalized derivatives of AD-mapped signal components. Such a kernel results in generation of identical and independent distribution datasets and in turn can facilitate discrimination among different signal classes during time-varying signal analysis. The underlying principle is that, the expectation of any AF-transformed input is maximized to obtain better discrimination. The mapped space is transformed such that the new representation is of the form,

$$\phi = \begin{bmatrix} ln(\mathbf{a}_n w') \\ \mathbf{a}_n w' \end{bmatrix} \tag{8}$$

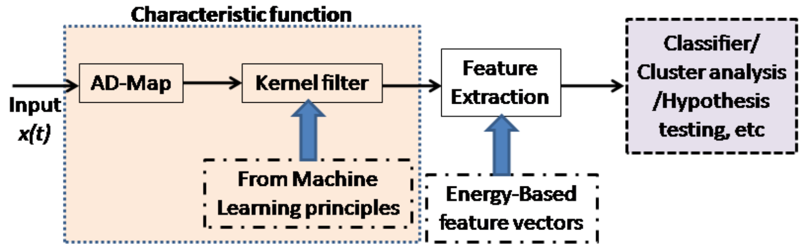

Fig. 1. General Block Diagram of the Proposed Scheme

Where $\mathbf{a}_n$ represents the set of normalized AD-mapped coefficients and w' represents the transpose of the weighting coefficients.

Choice of w***:*** From the AD-map, the maximum intensity in the time-lag axis is computed as the sum of all the energy terms across the entire set of rows for each available index. Similarly, the sum vector along the Doppler frequency axis of AD is calculated. The normalized weighting vector (w) is then defined by computing the dot product of the row and column sum vectors multiplied with the inverse of the maximum energy and an optimal value is calculated across the entire dataset. Such a product usually results in decreased weighting about the origin, since those auto-terms do not contribute towards the overlapping regions with cross-terms.

C. Feature Extraction

After transforming the time-varying signal using a kernel map, three unique feature vectors were derived from the AD-representation plot, given by equations (9), (10) and (11). Owing to the unique representation that the AD-space offers, the extracted feature set can be used to optimally characterize the auto- and cross-term energy spread in the mapped signal.
(a) Total Energy (TE): Computed as sum of all energy terms whose value exceeds a certain set threshold.

$$TE = \sum_{|M_{ML}| \geq M_{th}} |M_{ML}|^2 \tag{9}$$

(b) Total Energy Outside the Origin (TE_o): Calculated as the total energy spread outside the origin (satisfying the threshold criteria).

$$TE_o = TE - |M_{ML}(0,0)|^2_{\geq M_{th}} \tag{10}$$

(c) Sum of Diagonal Terms (SDE): Sum of all the diagonal terms in the AD-SVM coefficient matrix.

$$SDE = \sum diag\left(M_{ML}\right) \tag{11}$$

III. RESULTS

A. Experimental Dataset

In order to evaluate the performance of the proposed scheme, we use the gait dataset obtained from normal and ALS-affected subjects from Physionet [10]. The gait database contains real-life gait signals acquired from both healthy and pathological subjects with ALS. The gait time-series signal is computed by measuring the signals across the force-sensitive insoles that are placed under the subject‘s foot. During the experiments, the subjects were asked to choose their comfortable walking pace along a 77m long hallway. The stride to stride measurements are then recorded over a 5 minute interval. The ALS subjects were not using a wheelchair for mobility and were not diagnosed with any other ailments that might affect the gait stride-to-stride variability. Also, about 80% of the patients shared moderate severity for the disease. Overall, 16 healthy subjects (2 men and 14 women) aged 20-74 years, and 13 ALS subjects (10 men and 3 women) aged 36-70 years contributed to the gait dataset.

B. Performance Assessment

Because accurate characterization, and not visual appearance, is our sole design criterion, the designed kernels does not necessarily provide a visually satisfying representation and owing to which we base our discussion on the obtained performance accuracies. Our initial assessment involved evaluating the proposed discriminative and generative TF kernels for the normal and ALS gait datasets. The results obtained were comparable for both these kernels and hence, we limit our discussion to the effectiveness of the extracted set of novel feature vectors, for signal quantification. The obtained values are compared to certain previously used classification schemes due to the non-availability of benchmark schemes that involved exploiting machine learning concepts for design of time-frequency decomposition techniques.

The input signal is mapped on to the Hilbert space using the k_{tfML} kernel. To verify the discriminative capacity of the proposed kernels, feature vectors are extracted from gait (control and ALS) and is then fed to a Leave-One-Out (LOO) linear discriminant analysis (LDA) classifier. Table I shows the obtained classification results. An overall maximum accuracy of 93.1% is obtained by employing the AD-ML kernel mapping along with the newly deduced feature vectors. The obtained performance measures correspond to a 3% increase over the reported current maximum [11] that can be achieved using any gait classification scheme.

TABLE I

CLASSIFICATION ACCURACIES OBTAINED USING AD-ML KERNELS

Gait Type	Control	ALS	Total
Control	93.8%	6.2%	100%
ALS	7.1%	92.3%	100%

C. Discussion

Figure 2 shows the obtained box plot for the extracted feature vector for control and ALS subjects, respectively. It can be noted that the ALS inter-quartile range (IQR) falls in the mild outlier zone of the control subjects' feature plot. However, the median values and the IQR for both the data set are placed far apart. This in turn implies that the extracted feature set have good discriminative capacity between both the classes ('ALS' and 'Control'). The obtained quantitative results were also compared to two of the published high performance works. Wang et al [12] used the human movement information along with a combination of static and dynamic biometrics and reported a maximum classification accuracy (using nearest-neighbour classifier) of 87.5%. While Wu et al [11] conducted a series of analysis, for classification among control and ALS-affected subjects using an individual's gait and have reported an overall accuracy rate of 82.76% and 89.66% obtained using a LDA-based and non-linear classifier respectively [11]. It is to be noted that the above mentioned works employed non-linear classifiers and in comparison the proposed research direction offers comparable and in certain cases better discrimination. Based on the computed accuracy measures, atleast 3% improvement in signal pattern identification is obtained when the ML-based kernel design is incorporated in AD, in contrast to use of available TF kernels. The above performance metrics indicate that the use of AD-ML kernels offer a new direction in TF kernel design and can facilitate better characterization of biomedical signals.

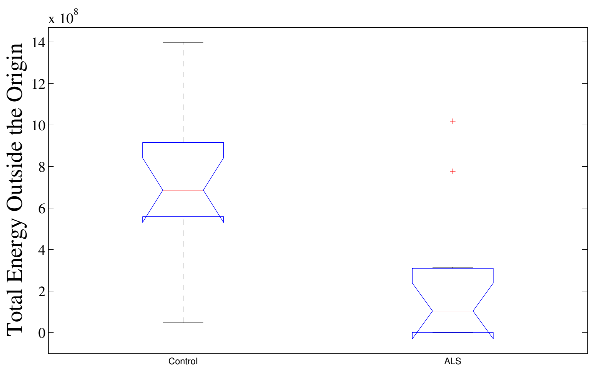

Fig. 2. (left: Control and right: ALS) Box plot of gait TE_os

IV. CONCLUSION

In this paper, the usability of the ML concepts for design of TF kernels that facilitates time-varying signal analysis is investigated in detail. Most existing research is concentrated on developing representation schemes for signal visualization (wherein the cross-terms are filtered out) and limited focus is available for signal quantification works. In this work, a novel method to quantify and visualize non-stationary signals is proposed which takes into consideration the cross-term energy distribution during signal characterization. Such an approach exploits the collective advantages of the AD-based TF schemes and the ML kernels. The newly derived AD-ML kernels offers efficient linear discrimination between different patterns with the aid of minimal number of feature vectors and a maximum overall classification accuracy of 93.1% is obtained. The proposed kernels show a lot of potential in discriminative analysis and further investigation on these kernels on large datasets will facilitate development of nearly-unified robust schemes for non-stationary signal analysis.

ACKNOWLEDGMENT

The authors would like to thank NSERC and Canada Research Chairs' program for funding this research.

REFERENCES

[1] L. Cohen. Time-frequency distributions - a review. *Proceedings of the IEEE*, 77(7):941 –981, July 1989.

[2] J. Sandberg and M. Hansson-Sandsten. A comparison between different discrete ambiguity domain definitions in stochastic time-frequency analysis. *Signal Processing, IEEE Transactions on*, 57(3):868 –877, March 2009.

[3] M. G. Amin and A. Belouchrani. Blind source separation using the spatial ambiguity functions. In *Proceedings of the International Symposium on Time-Frequency and Time-Scale Analysis*, pages 413 –416, 6-9 1998.

[4] P. Wang, J. Yang, and J. Xiong. Kernel design of bilinear tfd for multiple lfm signals with unbalanced amplitudes. In *Signal Processing and Information Technology, 2005. Proceedings of the Fifth IEEE International Symposium on*, pages 523 –527, 2005.

[5] Z. Hong and B. Zheng. An adaptive-kernel design method based on ambiguity domain. In *Time-Frequency and Time-Scale Analysis, 1998. Proceedings of the IEEE-SP International Symposium on*, pages 197 –200, 6-9 1998.

[6] T. Ebrahimi G. Garcia and J. Vesin. Classification of EEG Signals in the Ambiguity Domain for Brain-Computer Interface Applications. Lecture Notes in Computer Science. EUSIPCO, 2002.

[7] L-J. Wang G. Xiong, H-C. Zhao and J-B. Liu. The wide sense time-frequency representation based on the wavelet and its relation with cohen's class. In *Microwave and Millimeter Wave Technology, 4th International Conference on, Proceedings*, pages 677 – 680, 18-21 2004.

[8] L. Sugavaneswaran, K. Umapathy, and S. Krishnan. Exploiting the ambiguity domain for non-stationary biomedical signal classification. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 31 2010.

[9] B. Schlkopf, C.J.C. Burges, and A. J. Smola. *Advances in Kernel Methods: Support Vector Learning*. MIT Press, Cambridge, MA, 1999.

[10] M. E. Cudkowicz A. L. Peterson D. Kaliton J. M. Hausdorff, A. Lertratanakul and A. L. Goldberger. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *J Appl Physiol*, 88:2045–2053, 2000.

[11] Y. Wu and S. Krishnan. Computer-aided analysis of gait rhythm fluctuations in amyotrophic lateral sclerosis. *Medical and Biological Engineering and Computing*, 47(11):1165–1171, 2009.

[12] L. Wang, H. Ning, T. Tan, and W. Hu. Fusion of static and dynamic body biometrics for gait recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(2):149 – 158, 2004.