Automatic Modulation Classification using Information Theoretic Similarity Measures

Aluisio I. R. Fontes*, Leandro A. Pasa*, Vicente A. de Sousa Jr.*, José A. F.Costa* and Luiz F. Q. Silveira*
*Universidade Federal do Rio Grande do Norte (UFRN)
Natal - RN - Brazil
Email:{aluisio.igor,vicente.gppcom,jafcosta}@gmail.com, pasa@utfpr.edu.br, lfelipe@dca.ufrn.br

Fuad M. Abinader Jr.*†
†Instituto Nokia de Tecnologia (INdT)
Manaus - AM - Brazil
Email:fuad.junior@indt.org.br

***Abstract*—Modern wireless systems employ adaptive techniques to provide high throughput while observing desired coverage, Quality of Service (QoS) and capacity. An alternative to further enhance data rate is to apply *cognitive radio* concepts, where a system is able to exploit unused spectrum on existing licensed bands by "sensing" the spectrum and opportunistically access unused portions. Techniques like Automatic Modulation Classification (AMC) could help or be vital for such scenarios. Usually, AMC implementations rely on some form of signal pre-processing, which may introduce a high computational cost or make assumptions about the received signal which may not hold (e.g. Gaussianity of noise). This work proposes a new method to perform AMC which uses a similarity measure from the Information Theoretic Learning (ITL) framework, known as *correntropy coefficient*. It is capable of extracting similarity measurements over a pair of random processes using higher order statistics, yielding in better similarity estimations than by using e.g. correlation coefficient. Experiments with binary modulations show that in the presence of Additive White Gaussian Noise (AWGN), a 97% success rate in classification is achieved at a Signal-to-Noise Rate (SNR) of 5dB without requiring any pre-processing at all.**

***Index Terms*—Modulation Classification, Correntropy**

I. INTRODUCTION

Advances in Software Defined Radio (SDR) and the dissemination of wireless communication systems are motivating the development of "smart" techniques, allowing the automatic reconfiguration of wireless systems as a function of the operation environment. In these scenarios, a clear requirement is that wireless terminals are able to determine useful properties of the modulated wireless signals for demodulation purposes, without explicit exchanging of control information between transceivers. To accomplish this goal, one of the required sub-tasks is the "blind" recognition of the digital modulation scheme applied on the transmitted signal, allowing the receiver to support a wide range of modulation schemes in a dynamic way [1].

Automatic Modulation Classification (AMC) is a class of techniques for recognizing the type of digital modulation scheme used to generate a received modulated signal, with little or even no prior knowledge (such as e.g. its phase, frequency or amplitude) about the modulated signal itself [2]. Performing AMC is a hard task, specially because of the lack of knowledge about the signal. It becomes even harder as the received signal suffers from interferences, noise and channel fading. AMC techniques currently reported in the literature [3]–[11] employ pre-processing modules to aid extracting signal features for classification which, depending on the applied mechanism, may require making assumptions about the received signal which may not hold (e.g.AWGN being the unique source of noise) or may have a high computational cost to be implemented.

This paper proposes using an Information Theoretic Learning (ITL) random processes similarity measure, named *correntropy coefficient* [12], to perform AMC on samples from different binary digital modulation schemes affected by channels with AWGN. Correntropy coefficient uses infinite statistical moments of even order for estimating similarity between sample values from distinct random processes. We claim, from empirical results, that such higher order statistics allows us to extract information from the signals enought for determining which modulation scheme was used on the received signal, in such a way that no pre-processing modules for feature extraction are necessary.

The remainder of this paper is organized as follows. Section II briefly describes state-of-the-art AMC techniques commonly found on literature, discussing their requirements and limitations. Section III presents similarity measures derived from the ITL framework, in special the correntropy coefficient. Section IV presents the proposed AMC technique, based on the correntropy coefficient, while Sections V and VI present, respectively, the evaluation methodology and numerical results obtained from applying the proposed AMC scheme. Finally, Section VII presents conclusions, considerations and future work perspectives.

978-1-4673-1881-5/12/$31.00 ©2012 IEEE

II. AUTOMATIC MODULATION CLASSIFICATION (AMC)

Research on AMC can be classified into two approaches, either (i) *using statistics from the received signal* to define a Maximum Likelihood (ML) function, or (ii)*extracting signal features*, for performing the classification using different Pattern Recognition techniques [3], [13]. Authors in [4] present a survey of AMC approaches (updated in [5]) considering ML-based AMC. In any of those approaches, the classifying system must be capable of correctly determining the type of modulation scheme for a given signal sample among a set of N candidate modulation schemes. An ideal AMC must also satisfy the following requirements [4]:

- Provide high probability of True Positive (TP), and low probability of False Positive (FP) classifications, requiring for that a short observation interval;
- Be able to identify signals from different modulation schemes, and subject to varying channel conditions;
- Be implementable on embedded systems, should work on real-time, and should have low computational cost.

Performing AMC involves two stages [4]: (i) signal pre-processing for feature extraction, and (ii) modulation scheme identification using selected signal features. Figure 1 shows how those stages interact. On the *signal pre-processing* stage, the focus is on estimating system parameters, such as *carrier frequency*, *symbol period*, or *signal power*, or even providing noise reduction and channel equalization. However, common pre-processing techniques for AMC are not restricted to such tasks and include e.g. signal feature extraction.

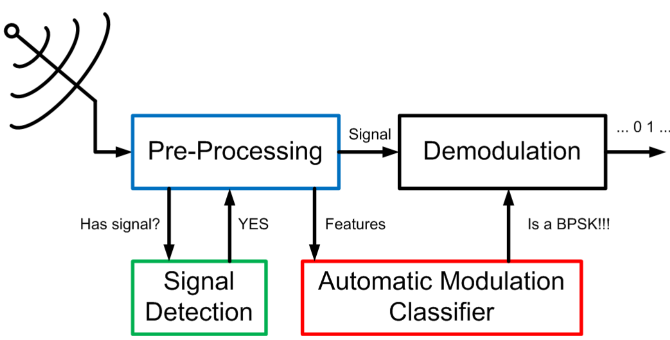


Fig. 1. Automatic Modulation Classification (AMC).

Many techniques have been proposed on literature for signal feature extraction for AMC. Authors in [6] propose estimating the received signal's standard deviation in the pre–processing stage, and using it to train a classifier based on an Artificial Neural Network (ANN). In [7], a *wavelet* transform is used for feature extraction on QAM, PSK and FSK signal samples, while authors in [8] use such features to also train an ANNN for AMC. There are also works proposing using higher order statistics and cyclostationary features [9], Principal Component Analysis (PCA) [10], and ANN with Fuzzy Logic [11]. Some of those pre-processing activities demand a high computational cost, thus making their application on real-life scenarios to be expensive, or even make it not practical for real-time systems with current off-the-shelf technology. This work proposes an AMC technique that uses a similarity measure of reasonable computational complexity in association with a very simple classifier.

III. INFORMATION THEORETIC SIMILARITY MEASURES

A common problem faced by many data processing professionals is how to best extract information contained in data, and to this effect *similarity* is a key concept to quantity temporal signals. Correntropy is a generalized similarity measure between two arbitrary scalar random variables X and Y from the Information Theoretic Learning (ITL) framework, defined by

$$v(X,Y) = E_{XY}[k(X,Y)] = \iint k(x,y)p_{X,Y}(x,y)dxdy, \tag{1}$$

, where the expected value is over the joint space, $p_{X,Y}$ is the joint probability distribution, and $k(\cdot,\,\cdot)$ is any continuous positive definite kernel function. Correntropy is a well-defined function, provided that $k(x,\,y)$ belongs to L_∞ (i.e. its maximal value is finite) [12]. For example, it can be verified that the specific positive definite kernel $k(x,\,y) = xy$ substituted in Equation (1) yields cross-correlation.

In this paper, the adopted correntropy measure is based on the Gaussian kernel, which is symmetric and translation-invariant, and is defined by

$$k(x,\,y) = K_\sigma(x_i - y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x_i-y_i)^2}{2\sigma^2}} \tag{2}$$

, where σ corresponds to the scaling factor of the Gaussian kernel. Substituting $k(x,y)$ in Equation (1) by the Taylor series expansion of the Gaussian kernel function in Equation (2), and assuming that it is valid to interchange the integral with the sum, the correntropy can be expressed by [12]

$$v(X,Y) = \frac{1}{\sqrt{2\pi}\sigma}\sum_{n=0}^{\infty}\frac{(-1)^n}{2^n\sigma^{2n}n!}E[(X-Y)^{2n}]. \tag{3}$$

Equation (3) states that the correntropy is constituted by a summation of all even moments of the difference variable. Thus correntropy keeps the nice bivariate form of correlation, but is still sensitive to the sum of second- and higher-order moments of the random variables. This is a interesting characteristic, because in many applications this sum may be sufficient to quantity better than correlation the relationships of interest, and it is simpler to estimate than the higher-order moments [12]. This property makes correntropy extremely sensitive to higher-order statistical moments, and allows to extract more information from random variables than traditional similarity measures. Besides, as the internal product on $K_\sigma(x_i - y_i)$ tends to zero, correntropy is seen as a robust similarity measure sensitive to time-varying random processes.

In practice, the joint PDF in Equation (1) is unknown and only a finite number of data $\{(x_i,\, y_i)\}_{i=1}^{N}$ are available, leading to the sample correntropy estimator defined by [12]

$$V(\mathbf{X},\mathbf{Y}) = \frac{1}{N}\sum_{i=1}^{N} K_\sigma(x_i - y_i) \tag{4}$$

In Equation (4), the Gaussian kernel is responsible for mapping the random vectors into a *feature space*, named *Reproducing Kernel Hilbert Space* (RKHS) [14]. Therefore, the sample correntropy estimator corresponds to the sample correlation estimator when measured on the RKHS feature space.

As the non-linear mapping performed by the Gaussian kernel does not ensures zero mean even when the original samples are centered, Principe [12] proposes adopting the *centered cross-correntropy*, a centered correlation function measured on the RKHS whose estimator is defined by equation (5):

$$U(\mathbf{X},\mathbf{Y}) = \frac{1}{N}\sum_{i=1}^{N} K_\sigma(x_i - y_i) - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} K_\sigma(x_i - y_j) \tag{5}$$

Santamaria et al. [15] present a new similarity measure, named *correntropy coefficient* and defined in accordance with Equation (6). It corresponds to the cosine of the angle between two random sample vectors transformed on the RKHS, which by using the infinite even moments is capable of extracting more information than the conventional correlation coefficient. Here, $U(\mathbf{X},\mathbf{Y})$ correspond to the centered auto-correntropy of the vectors $\mathbf{X}$ and $\mathbf{Y}$, respectively. While $U(\mathbf{X},\mathbf{X})$ and $U(\mathbf{Y},\mathbf{Y})$ correspond respectively to the centered auto-correntropy of the vectors $\mathbf{X}$ and $\mathbf{Y}$. It can be seen that the correntropy coefficient assumes zero value when the two random variables are independent, and take values near to 1 (or -1) as more similar (or similar but with opposite values) the vectors are.

$$\eta(\mathbf{X},\mathbf{Y}) = \frac{U(\mathbf{X},\mathbf{Y})}{\sqrt{U(\mathbf{X},\mathbf{X})}\sqrt{U(\mathbf{Y},\mathbf{Y})}} \tag{6}$$

Correntropy not only exploits simultaneously spatial and spectral signal characteristics, but it also has additional properties (when compared with e.g. second order statistics) that can be useful in non-Gaussian signal processing. In general, correntropy has benn used on estimation algorithms which employs nonlinearities: temporal principal component analysis (TPCA) corrupted by impulsive noise [16], blind source separation (BSS) [17], image recognition [18], robust signal detection [19], and so on.

IV. Proposed Classifier

This paper proposes a new method for AMC, based on the correntropy coefficient defined by Equation (6). We claim that correntropy coefficient is well-suited to characterize dynamic interdependencies between modulated signal samples, and therefore fits well for AMC purposes. Reasoning is two-fold: (a) received signals can be seen as sums of different random variables, representing both the modulated signals at the transmitter and the different channel effects on the transmitted signals, and (b) correntropy coefficient is sensitive to non-linearities and higher order statistical information.

Also, as correntropy coefficient is able to characterize dynamic interdependencies even when the signals are corrupted by noise [15], we claim that no pre-processing is required, so reducing the computational complexity for AMC as a whole. This simplified design is one of the contributions of this work, and Figure 2 illustrates the general architecture of the classification system in which we evidence its capability of recognizing modulations without pre-processing module.

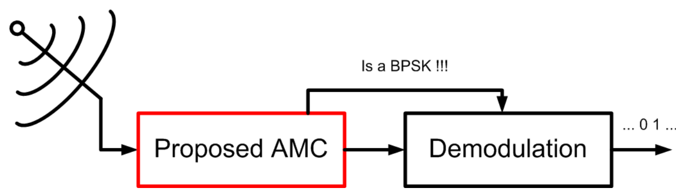


Fig. 2. AMC with Correntropy Coefficient.

In this proposed method, illustrated by Figure 3, the classifier takes a sample from the received signal and, for each supported modulation scheme, it estimates its correntropy coefficient with samples within a set of vectors representing noise-free sample vectors for the supported modulation scheme (herein known as *prototypes*). Decision on how to classify the received signal is given by which modulation provided prototypes returning the highest correntropy coefficient with the received signal sample vector.

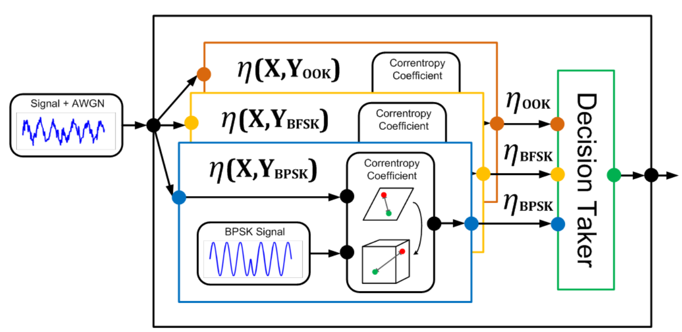


Fig. 3. Correntropy-based Classifier.

V. Evaluation Methodology

BFSK signals are characterized by constant instantaneous amplitude, whereas OOK signals have amplitude fluctuations, and BPSK signals have information in the phase. So, in order to evaluate the performance of the proposed classifier, computational simulations of the proposed AMC classifier were performed on MATLAB®. Samples from binary digital modulation schemes (BFSK, BPSK and OOK) in the presence of Additive White Gaussian Noise (AWGN) of different levels ($1dB$, $5dB$ and $15dB$) were generated and used for correntropy coefficient calculations. Also, prototypes for all the three evaluated modulation schemes were generated and used on such calculations, so the classifier could be able to select the most suitable classification for a given sample by comparing correntropy coefficient values obtained from comparison with prototypes coming from different modulation schemes.

In order to determine the minimal number of signal samples necessary for good classification, numerical results for the correntropy coefficients are calculated for samples sized of 5, 10, 20 and 50 symbols, using a sampling rate of 50 samples per symbol. A minimal of 2000 classification procedures with different signal samples was performed, in order to provide goood statistical confidence to the results. As a metric of efficiency, we adopted the *correct classification rate*, i.e. the rate of signals from a given modulation scheme which were correctly identified by the proposed classifier.

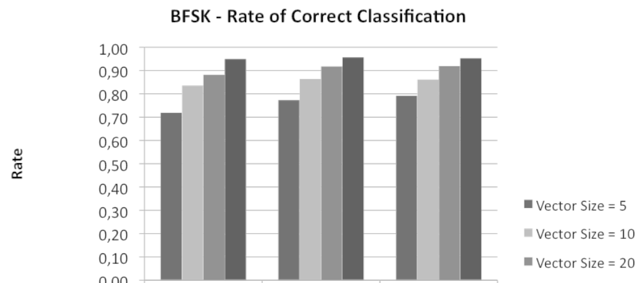
An important parameter to be adjusted when working with ITL similarity measures such as the correntropy coefficient is the σ (variance) used for the Gaussian kernels. Here, σ works as a *scaling factor*, that needs to be selected as a function of both the sample data dynamic range and the number of observed samples. In this work, σ is calculated by applying Silverman's *rule-of-thumb* [12], expressed by

$$\sigma = \sigma_X \left[4N^{-1}\left(2d+1\right)^{-1}\right]^{\frac{1}{(d+4)}} \tag{7}$$

, where d corresponds to the data dimension (in our case, $d = 1$), N corresponds to the sample size and σ_X corresponds to the trace of autocovariance matrix of X.

Analyzing the Equations (5) and (6) used for the calculation of the correntropy coefficient estimator in this proposed AMC classifier, one can conclude that the correntropy coefficient has a computational complexity of the order of $O(n^2)$, mainly due to the double Gaussian summation in Equation (5). With the objective of further reducing the computational cost of the proposed method, an efficient implementation of this summation is accomplished by using the technique known as *Fast Gauss Transform* (FGT) [20]. It allows reducing the final computational complexity to $O(n \log n)$, and it is is widely applied in many applications of pattern recognition.

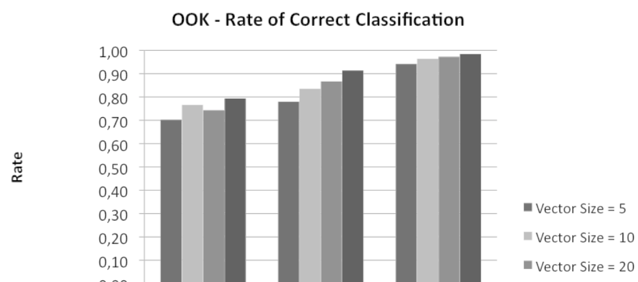
VI. Simulation Results



	1 dB	5 dB	15 dB
Vector Size = 5	0,72	0,77	0,79
Vector Size = 10	0,84	0,86	0,86
Vector Size = 20	0,88	0,92	0,92
Vector Size = 50	0,95	0,96	0,95

SNR [dB]

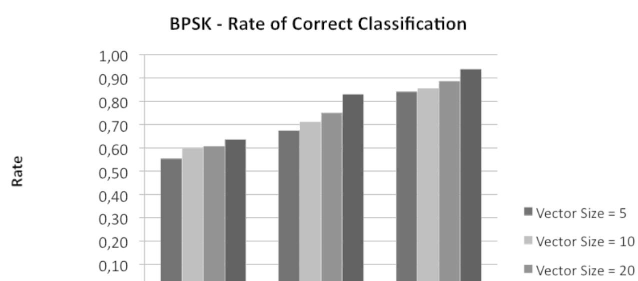
Fig. 4. Numerical results for BFSK with proposed AMC.



	1 dB	5 dB	15 dB
Vector Size = 5	0,70	0,78	0,94
Vector Size = 10	0,77	0,84	0,96
Vector Size = 20	0,74	0,87	0,97
Vector Size = 50	0,79	0,91	0,98

SNR [dB]

Fig. 5. Numerical results for OOK with proposed AMC.



	1 dB	5 dB	15 dB
Vector Size = 5	0,55	0,67	0,84
Vector Size = 10	0,60	0,71	0,86
Vector Size = 20	0,61	0,75	0,89
Vector Size = 50	0,64	0,83	0,94

SNR [dB]

Fig. 6. Numerical results for BPSK with proposed AMC.

In this section, we present numerical results from evaluation study of the proposed AMC classifier, summarized on Figures 4, 5 and 6. The figures show the rate of correct classification as a function of both the tested SNRs and the sample vector size for BFSK, OOK and BPSK, respectively.

One general observation is that these results confirm that the proposed classifier is sensitive to both the sample vector size and the SNR variations. Specifically sspeaking, it can be seen on Figure 4 that BFSK signal classification presented satisfactory results, evidencing that high order statistical moments are able to track frequency changes of modulated signal. For OOK, on the other hand, the rate of correct classification reaches up to 94% with a vector size of only 5 samples, as shown on Figure 5. The OOK signal has one interesting characteristic considering its susceptibility for classification purposes. The absence of energy in one OOK symbol could hide its statistical feature that our classifier method is trying to detect. This is the main reason of poor results for low SNR even compared to BFSK. In this situation, the whole OOK signal is statistically similar to AWGN. Finally, for BPSK results on Figure 6, which does not have neither frequency variation nor information in the amplitude, the hit reaches 94%, indicating a good capacity for the correntropy coefficient to perform a correct classification on the generating random process for the signal.

From the numerical results, we claim that the satisfactory classification results are result mainly of correntropy coefficient's capacity of extracting higher order statistical moments information. Therefore, we advocate the proposed AMC classifier as an AMC classification tool which doesn't demand a high computational implementation cost to provide good classification results.

VII. Conclusion

This work proposes a method for the Automatic Modulation Classification (AMC) of binary digital modulation schemes in the presence of Additive White Gaussian Noise (AWGN), which is based on the *correntropy coefficient*. Numerical results, obtained by computational simulation, indicate that the proposed method can recognize the digital modulation signals effectively. By just using the correntropy coefficient, we can even consider not applying any pre-processing phase at all, which is a differential to other AMC methods commonly found on literature. As future work, we intend to evaluate the performance of the proposed method in other M-ary digital modulation schemes, and under the effect of different channel conditions such as fading and multipath effects.

Acknowledgment

The author Fuad Abinader is supported by FAPEAM Ph.D. scholarship program number 020/2010. The author Aluisio Fontes is supported by the Committee for the Advancement of University Academic Staff - CAPES.

References

[1] O. D. A. A. Y. B.-N. W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," vol. 1, no. 2, pp. 137–156, 2007.

[2] B. Ramkumar, "Automatic modulation classification for cognitive radios using cyclic feature detection," vol. 9, no. 2, pp. 27–45, 2009.

[3] M. Rahman, A. Haniz, M. Kim, and J. Takada, "Automatic modulation classification in wireless disaster area emergency network (W-DAEN)," in *Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), 2011 Sixth International ICST Conference on*, june 2011, pp. 226 –230.

[4] O. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *Communications, IET*, vol. 1, no. 2, pp. 137 –156, april 2007.

[5] J. Xu, W. Su, and M. Zhou, "Likelihood-ratio approaches to automatic modulation classification," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 41, no. 4, pp. 455 –469, july 2011.

[6] I. E.-R. N. El-Madany, "Cognitive digital modulation classifier using artificial neural networks for NGNs." Alexandria,: 7th International Conference On Wireless And Optical Communications Networks (WOCN), 2010, pp. 1–5.

[7] Y. H. K.C. Prokopiw, W. Chan, "Modulation identification of digital signals by the wavelet transform," vol. 147, no. 4, pp. 169–176, 2000.

[8] K. Hassan, I. Dayoub, W. Hamouda, and M. Berbineau, "Automatic modulation recognition using wavelet transform and neural networks in wireless system," vol. 2010, p. 13, 2010.

[9] A. Fehske, J. Gaeddert, and J. Reed, "A new approach to signal classification using spectral correlation and neural networks." Baltimore,: IEEE International Symposium on Dynamic Spectrum Access Networks, 2005, pp. 144–150.

[10] F. He, Y. Yin, and L. Zhou, "Principal component analysis of cyclic spectrum features in automatic modulation recognition." San Jose, CA,: IEEE Military Communications Conference (MILICON), 2010, pp. 1737–1742.

[11] W. Zong, E. Lai, and C. Quek, "Digital modulation classification using fuzzy neural networks," vol. 30, pp. 101–116, 2006.

[12] J. C. Principe, *Information Theoretic Learning - Renyiś Entropy and Kernel Perspectives*, 1st ed. Nova York, EUA: Springer, 2010.

[13] H.-C. Wu, M. Saquib, and Z. Yun, "Novel automatic modulation classification using cumulant features for communications via multipath channels," *Wireless Communications, IEEE Transactions on*, vol. 7, no. 8, pp. 3098 –3105, august 2008.

[14] J. W. Xu, P. P. Pokharel, A. R. C. Paiva, and J. C. Principe, "Nonlinear component analysis based on correntropy." Vancouver,: International Joint Conference on Neural Networks, 2006, pp. 1851–1855.

[15] I. Santamaria, P. P. Pokharel, and J. C. Principe, "Generalized correlation function: Definition, properties, and application to blind equalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2187–2197, 2006.

[16] J. P. Weifeng Liu, P.P. Pokharel, "Correntropy: Properties and applications in non-gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 5286 – 5298, 2006.

[17] R. Li, W. Liu, and J. C. Principe, "A Unifying Criterion for Instantaneous Blind Source Separation Based on Correntropy," *Signal Processing*, vol. 87, pp. 1872–1881, 2007.

[18] K. H. Jeong and J. C. Principe, "The correntropy MACE filter for image recognition," in *Proc. of the 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing (MLSP)*, Sept. 2006, pp. 9–14.

[19] J. C. PRINCIPE and W. LIU, "Robust signal detection using correntropy," Patent Application US 2010/0 197 258 A1, 08 05, 2010. [Online]. Available: http://www.patentlens.net/patentlens/patent/US_2010_0197258_A1/en/

[20] C. Yang, R. Duraiswami, N. A. Gumerov, and L. Davis, "Improved fast gauss transform and efficient kernel density estimation." Ninth IEEE International Conference on Computer Vision (ICCV'03), 2003.