A Feature Weighted Hybrid ICA-SVM Approach to Automatic Modulation Recognition

David Boutte and Balu Santhanam

Department of Electrical and Computer Engineering
MSC01 1100
1 University of New Mexico
Albuquerque, NM: 87131-001
Email: dboutte,bsanthan@ece.unm.edu

***Abstract*—Automatic modulation recognition is a topic of interest in many fields including signal surveillance, multi-user detection and radio frequency spectrum monitoring. A major weakness of conventional modulation recognition algorithms is their reliance on high SNR environments and favorable statistics. In this paper an algorithm is developed using elements of cyclo-spectral analysis, ICA and SVM algorithms to distinguish between different modulation types. By first estimating the cyclic spectrum and then analyzing statistical features of the spectrum using machine learning techniques, the particular modulation type can be determined over a wide range of SNR values. This can further be enhanced by employing ICA algorithms to remove feature redundancy. To demonstrate this; simulations are constructed which illustrate the efficiency of the algorithm using digital phase and amplitude modulation. The algorithm's performance is tested over a wide range of SNR values.**

I. Introduction

Automatic modulation identification plays an important role in signal surveillance and frequency spectrum monitoring. A class of existing modulation recognition approaches make use of instantaneous signal parameters [1] and hierarchical classification methods using higher order moments [2]. However, these algorithms fail to take full advantage of the spectral content of the signal. Most modulation types have been shown to exhibit cyclo-stationary behavior [3]. This additional information can be exploited to classify modulation types in the presence of noise.

Independent component analysis (ICA) is a statistical method for separating a multivariate signal into additive subcomponents supposing the mutual independence of the source signals [4]. This technique is a subclass of source separation algorithms. This algorithm has been widely applied in blind source separation. It has also been directly applied to the modulation recognition problem in [5]. However, in this paper the ICA algorithm will be used to remove statistical dependence between the features by exploiting their non-Gaussianity with higher order statistics.

The field of statistical learning has many discoveries which can be applied to the modulation recognition problem. *Support vector machines* (SVMs) offer a complex decision making process without much of the demands on SNR and signal frame length. SVMs provide an efficient means of classification without placing assumptions on signal statistics or on signal to noise ratios. The standard binary SVM classifier as described in [6] can be simply extended to a multi-classifier for use in this case. Similarly by appropriately weighting the features in importance, problems of misclassifying similar modulations can be resolved.

Research funded through Miratek Corporation, 8201 Lockheed Suite 218, El Paso TX. (915)772-2852 May 2007

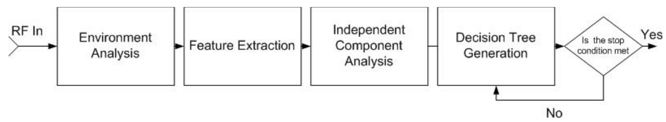


Fig. 1. Block diagram for the feature weighted hybrid ICA-SVM modulation recognition system presented in this paper.

In this paper, we propose the use of cyclostationary statistical analysis combined with ICA-SVM to form a hybrid approach to the problem of modulation recognition. The jointly optimal solution involves optimizing both the ICA and SVM nonlinearities and ICA weights dependent upon changes in the other module. However, due to the number of parameters involved and the coupled nature of the operation the joint optimization problem becomes intractable. A suboptimal solution is presented in this paper which independently optimizes the ICA based on a fixed SVM kernel. By combining these methods together into a hybrid approach, an efficient and robust algorithm is developed that offers superior performance over a wide range of SNR values using a variety of modulation types.

II. Algorithm

A. Cyclic Spectra

Most modulation schemes, both analog and digital, exhibit cyclostationary behavior. One method for recognizing the specific modulation as suggested by [7] which relies directly on the peak locations in the SCD function. In order to exploit the additional information in the cyclic spectrum; it is necessary to estimate of the spectral correlation density function (SCD). The estimation of this function is based on the smoothed cyclic cross periodogram, defined as

$$S_{XY_T}^{\alpha} = \frac{1}{T}\left\langle X_T\left(n, f+\frac{\alpha}{2}\right) Y_T^{*}\left(n, f-\frac{\alpha}{2}\right)\right\rangle, \qquad (1)$$

where $X_T\left(n, f+\frac{\alpha}{2}\right)$ and $Y_T\left(n, f-\frac{\alpha}{2}\right)$ are the complex demodulates of $x(n)$ and $y(n)$ computed as

$$X_T(n, f) = \sum_{-\frac{M}{2}}^{\frac{M}{2}} a(r)x(n-r)e^{-i2\pi f(n-r)T} \qquad (2)$$

$$Y_T(n, f) = \sum_{-\frac{M}{2}}^{\frac{M}{2}} a(r)y(n-r)e^{-i2\pi f(n-r)T}, \qquad (3)$$

where $a(r)$ is a spectral window function.

An algorithm using parameterized features of the signal of interest's SCD is proposed. By parameterizing the SCD a level of noise reduction not seen in [7] is introduced.

B. Feature Selection

Three key statistical features of the SCD that will be utilized are the standard deviation, variance, and kurtosis. The standard deviation of the SCD is computed as

$$\sigma = \sqrt{\frac{1}{M}\sum_{i=\frac{-M}{2}}^{\frac{M}{2}} \left(S_{XX_T}^{i}\right)^2 - \overline{S_{XX_T}}^{2}}, \qquad (4)$$

where M is the number of cycle frequencies the SCD is estimated over. The variance is given by

$$\sigma^2 = \frac{1}{M}\sum_{i=\frac{-M}{2}}^{\frac{M}{2}} \left(\left(S_{XX_T}^{i}\right) - \overline{S_{XX_T}}\right)^2. \qquad (5)$$

The sample kurtosis is

$$g_2 = \frac{\sum_{i=\frac{-M}{2}}^{\frac{M}{2}} \left(\left(S_{XX_T}^{i}\right) - \overline{S_{XX_T}}\right)^4}{\left(\sum_{i=\frac{-M}{2}}^{\frac{M}{2}} \left(\left(S_{XX_T}^{i}\right) - \overline{S_{XX_T}}\right)^2\right)^2} - 3, \qquad (6)$$

where $\overline{S_{XX_T}}$ in (5) denotes the sample mean of the SCD. While the standard deviation and the variance may be slightly redundant in some cases, in others their contrast offers superior performance. Each of these features are dependent on a specific modulation type and by analyzing these features the specific modulation type can be determined. Quantifying performance over changing SNRs can be related to signal separation in communication theory. A similar case applies here; the distance between modulation features measures their resistance to noise. The larger the distance is; the more distinct the features are, and the easier it is to distinguish between them. The ICA directly increases this distance by separating the features into independent components.

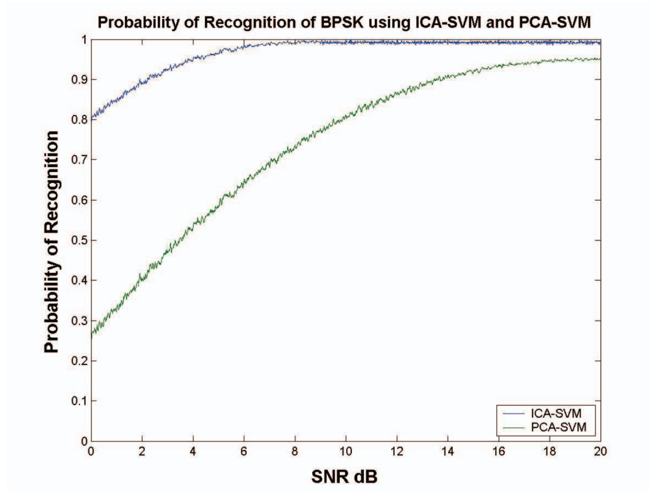


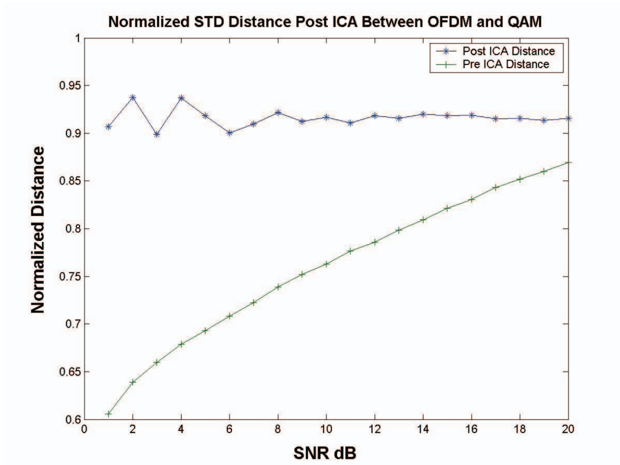
Fig. 2. Probability of Recognition of BPSK using both PCA and ICA processing. The top curve is using ICA processing while the bottom curve uses PCA processing. The improvement in performance with the ICA is indicative of the non-Gaussian nature of the features.

C. Independent Component Analysis

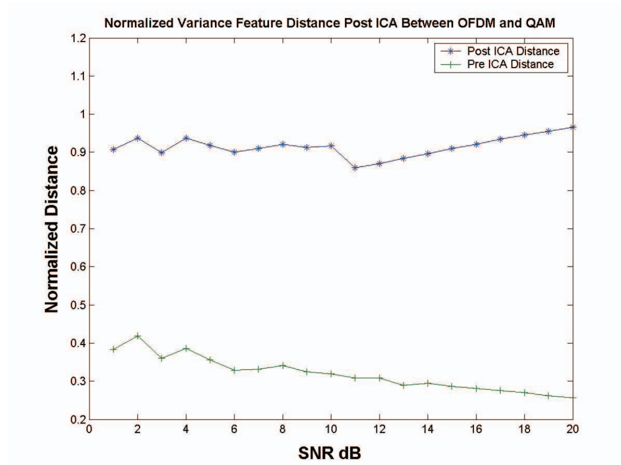
For similar modulation types, there is a large amount of statistical dependence between the feature vectors. To improve recognition probabilities it is desirable to choose a feature set which is statistically independent. ICA has gained popularity in recent years as a topic in blind source separation.

The extracted cyclic features can be viewed as a set of time series. A composite feature matrix is formed by aggregating the cyclic features extracted from the SCD; this composite feature matrix can be viewed as a multivariate signal. By making this assumption; the ICA algorithm can be applied to this composite feature matrix to separate the corresponding features into their independent subcomponents. By removing the redundancy from the features, the distance between them is increased. As an efficient way to to accomplish the feature separation; an ICA algorithm with QR preprocessing [8] was implemented. In addition to this some denoising techniques were applied as described in [9] to further suppress any estimation noise present in the cyclic features. The resulting feature matrix still contains the parameterized SCD information, however, L_2 distance between the features has been increased as seen in Fig(3). This boost in feature distance translates directly to increased recognition in low SNR environments.

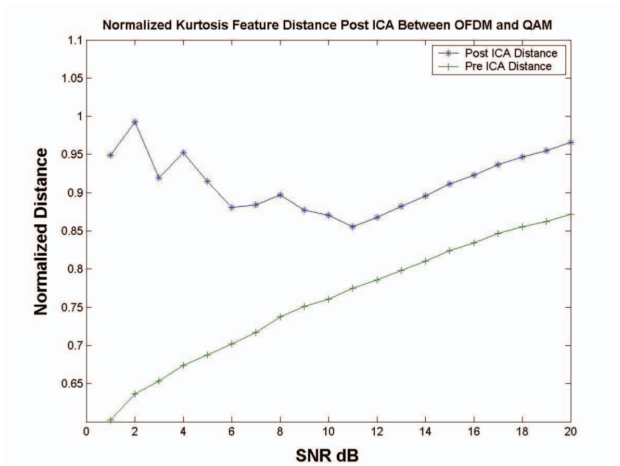
The performance of the ICA-SVM algorithm was compared to a similar *principle component analysis*-SVM (PCA-SVM) algorithm as seen in Fig(5). The ICA-SVM algorithm exhibits a much higher probability of recognition than the PCA-SVM due primarily to the non-Gaussian nature of the feature set. Since the ICA exploits non-Gaussian statistics in the source distributions, this shows up as an increased separation in the features and consequently an increased probability of recognition. As a further qualification, all the modulations under test were to seen to be super-Gaussian.



(a)



(b)



(c)

Fig. 3. Normalized standard deviation, variance and kurtosis L_2 feature distance between QPSK and QAM for cyclic features pre and post ICA processing. Notice the increase in feature distance after application of ICA.

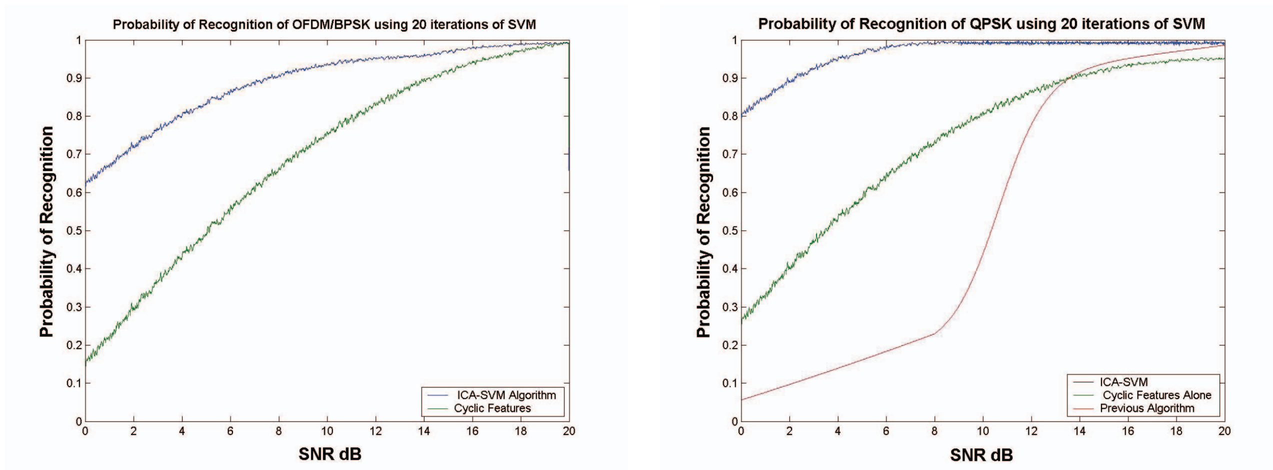
D. Classification

A final component in the algorithm is the classification itself. This is accomplished using a support vector machine architecture [6]. These are a widely used set of linear classifiers that have been applied to many problems. The features are first translated to a higher-order space consisting of all the modulation types of interest. Next maximal separating hyperplanes are constructed dividing the features into separate regions. Depending on the region in which the features fall the classification is made and the modulation reported. This is further refined by repeated use of the feature grouping within the modulation subspace as a constraint. At each iteration features assigned to a specific modulation are grouped in the modulation subspace, the SVM is then rerun on these modified features and new hyperplanes are constructed. This is repeated until the features are grouped in some distance δ. As suggested in [10] the input features to the SVM module can be weighted in importance. Empirical trial suggests that some cyclic features are more important than others during classification, by appropriately weighting these features in the SVM classification error can be reduced. This weighting is of use when there are problems of misclassifying a modulation as another similar type, such as different levels of amplitude or phase shift keying.

It is beneficial to compare the results of this algorithm to previous work. Using Bayesian classifier theory it can be shown that the probability of recognition of some previous algorithms [11] can be bound by

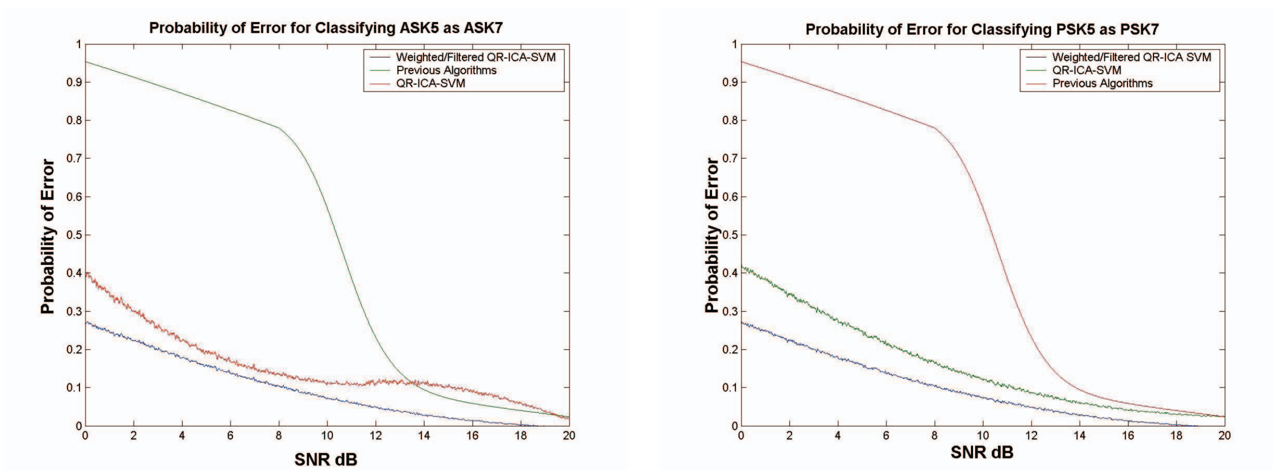
$$1 - P(E) \leq \frac{0.66}{1 + 4000 \exp(-1.83x)} \exp(-0.2x) . \quad (7)$$

This is a fairly loose upper bound on performance, however it does serve to illustrate the difference between prior work in [1] and the proposed algorithm.



(a) (b)

Fig. 4. Probability of recognition using QR-ICA processing, cyclic features and theoretical standard feature performance for both QPSK and OFDM/BPSK signals. The top curve represents the presented algorithm's performance, while the lower curves represent previous algorithms performance. The improvement is most noticeable when applied to QPSK as (b).



(a) Probability of Error for Classifying ASK5 as ASK7 (b) Probability of Error for Classifying PSK5 as PSK7

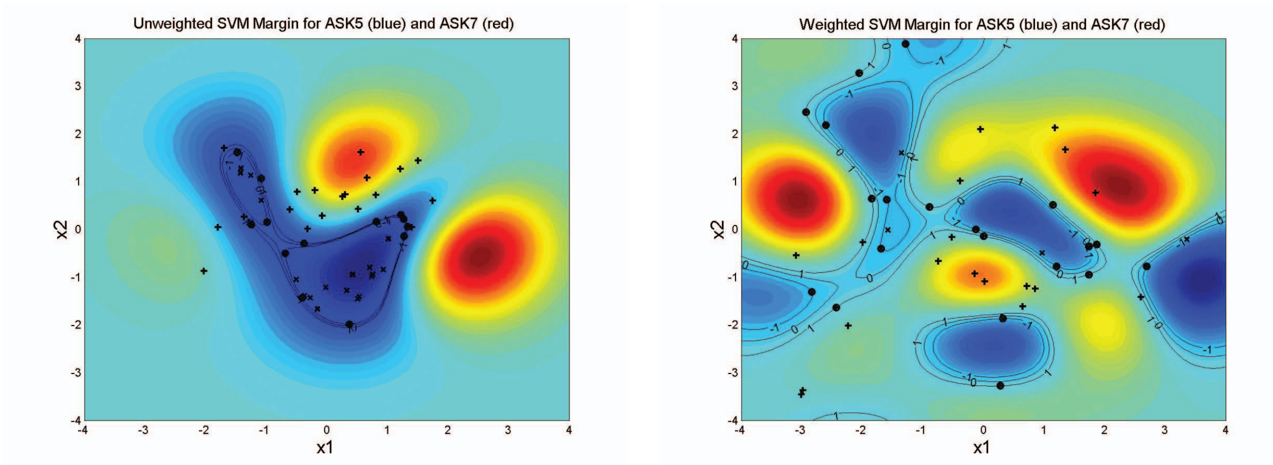
Fig. 5. Probability of error in classifying close modulations, including generic algorithm and weighted and filtered features. The lower curve represents the the presented algorithm's performance using feature importance weighting in the SVM and filtering/denoising in the ICA.

III. Results

An algorithm as outlined in Fig(1) was implemented using MATLAB. A set of modulated signals were generated in additive white Gaussian noise at baseband and statistical dependence is removed from the cyclic feature set using an ICA. The resulting independent feature matrix was then passed to the SVM block to make the classification. This process was repeated for 1000 experiments in a Monte Carlo trial to estimate the probability of recognition.

Fig(4) shows the probability of recognition for QPSK and OFDM/BPSK signals using a QR factorization of the feature matrix to initialize the ICA. A principle component analysis was also performed to contrast with the ICA as seen in Fig(2). The improvement in performance using the ICA-SVM can be attributed partially to the non-Gaussianity present in the features. These probabilities are compared with the theoretical probability of recognition of previous approaches for QPSK and the probability of recognition using cyclic features alone. This probability is increased by using more SVM iterations. However, further experiments have shown that increased iterations above 20 yield no significant improvements in the probability of recognition.

ASK5 and ASK7 have similar cyclic spectra, making them difficult to distinguish between. The most common misclassification is recognizing one signal as the other. This is shown in Fig(5(a)). Even though cyclic spectra of the modulations are close, the proposed algorithm is still able to correctly classify them in the majority of cases. The similar case for



(a)

(b)

Fig. 6. SVM margin of separation between ASK5 and ASK7. Using appropriate feature weighting results in a higher degree of separation, as seen in the left figure, and thus a higher probability of classification.

PSK is shown in Fig(5(b)) with similar results. Preliminary experiments show that appropriately weighting the features as they are input into the SVM block will improve performance in these cases. At the expense of more computational overhead the added feature weighting and denoising is able to drive the probability of error down as shown by the lower curves in Fig(5).

IV. Conclusion

The use of ICA-SVMs has been proposed before by [12]. However there are several differences in the algorithm proposed here. The use of ICA in [12] is a method of extracting a feature set. In this proposed algorithm, the features are extracted from the cyclic spectrum and the ICA is a tool to refine the features and reduce the classifier error present in the SVM.

By looking at statistical features of the spectral correlation density function, increased recognition performance in lower SNR regions can be realized. The chosen cyclic features exhibit a higher L_2 distance than instantaneous signal measurements used in previous modulation recognition algorithms. This increase in feature separation corresponds to an increase in recognition probability. The features can be further separated when an independent component analysis is performed on the composite feature vector. Adjustments to the ICA including denoising and QR initialization allow further refinements in performance. The inclusion of feature importance weighting in the SVM solves some problems associated with misclassification.

References

[1] E. E. Azzouz and A. K. Nandi, *Automatic Modulation Recognition of Communication Signals*. 3300 AA Dordrect, The Netherlands: Kluwer, 1996.

[2] A. Swami and B. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, pp. 416–429, march 2000.

[3] W. A. Gardner and S. Venkataraman, "Performance of optimum and adaptive frequency-shift filters for cochannel interference and fading," Maple Press, 1990.

[4] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, pp. 411–430, Feb. 2000.

[5] L. Hong, "Classification of bpsk and qpsk signals in fading environment using the ica technique," Proceedings of the Thirty-Seventh Southeastern Symposium on System Theory, vol. 1, pp. 471–475, 2005.

[6] N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines and other kernel-based learning methods," Cambridge University Press, 2000.

[7] W. A. Gardner, "Two alternative philosophies for estimation of the parameters of time-series," *IEEE Trans. Info. Theory*, vol. 37, pp. 216–218, Jan. 1991.

[8] M. Gupta and B. Santhanam, "Prior ica based blind multiuser detection in ds–cdma systems," Proceedings of the Thirty-Sixth Asilomar Conference on Signals and Systems, vol. 4, pp. 2155–2159, 2004.

[9] D. Zhang and P. K. Rajan, "Supergaussian data denoising by semi-ica estimation," Proceedings of the Thirty-Seventh Southeastern Symposium on System Theory, vol. 3, pp. 485–490, 2005.

[10] V. N. Vapnik, *Statistical Learning Theory*. New York, New York: John Wiley and Sons, 1998.

[11] M. Ben-Bassat, *Use of distance measures, information measures and error bounds in feature evaluation*. The Netherlands: Handbook of Statistics, 1982.

[12] Y. Qi, D. Doermann, and D. DeMethon, "Hybrid independent component analysis and support vector machinelearning scheme for face detection," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. 1481–1484, 2001.

[13] C. M. Spooner and W. A. Gardner, "Estimation of cyclic polyspectra," Proceedings of the Twenty-fifth Asilomar Conference on Signals and Systems, vol. 2, pp. 370–376, 1991.