Automatic Modulation Classification of Digital Modulation Signals Based on Gaussian Mixture Model

W.H. Ahn, J.W. Choi, C.S. Park, B.S. Seo
Dept. of Electronics Engineering
Chungbuk National University
Cheongju, Korea
{glingi, cjw442, chansp, boseok}@cbnu.ac.kr

M.J. Lee
Agency for Defense Development of Korea
Daejeon, Korea
drdlalswns@dreamwiz.com

***Abstract*—In this paper, we propose an automatic modulation classification scheme for digitally modulated signals, such as MSK, GMSK, BPSK, QPSK, 8-PSK, 16-QAM, 32-QAM, and 64-QAM. As features which characterize the modulation type, higher order cyclic cumulants up to eighth order of the signal are used. For feature classification, a Gaussian mixture model based algorithm is used. Simulation results are demonstrated to evaluate the performance of the proposed scheme under AWGN channels.**

Keywords- automatic modulation classification; Gaussian mixture model; cyclostationary; higher order cyclic cumulants.

I. Introduction

Automatic modulation classification (AMC) is a technique to identify the modulation type of the detected signal as well as to estimate the signal parameters such as carrier frequency and symbol rate, etc [1]. It is widely tried to apply in the field of military and civilian for electronic warfare, spectrum monitoring, surveillance, and cognitive and software defined radios.

There have been many studies on AMC for last two decades. AMC schemes are normally classified into two major categories which are likelihood-based approach [2] and feature-based approach [3]-[10]. The likelihood-based method shows optimal performance in the sense that it maximizes the probability of correct classification. However, it has higher computational complexity for likelihood computation. In addition, it is highly sensitive to modeling mismatch such as timing, phase and frequency offsets, and noise variance.

The feature-based approach attempts to extract a set of features from the received signal. Because the features represent a distinct pattern in a feature space, various pattern recognition algorithms can be applied for classification. Although this approach may not show optimal performance, it is easy to implement and shows nearly optimal performance when appropriate features and a classifier are combined.

In [3], higher order statistics up to sixth order are used to discriminate binary phase-shift keying (BPSK), quadrature phase-shift keying (QPSK), 16-ary quadrature-amplitude modulation (16-QAM), and 64-QAM. As a classifier, a genetic programming combined with the *K*-nearest neighbor (KNN) algorithm is used. It shows good performance even in the presence of noise and frequency offset. In [4]-[6], cyclostationary statistics such as spectral correlation density [4] and cyclic cumulants (CCs) [5, 6] of the signal are used as features. These methods use a simple decision tree [4], the minimum distance metric [5], and the Mahalanobis distance metric [6] for classification. The classification performance is insensitive to model mismatch and independent of any a priori knowledge of signal parameters. In [7], a Gaussian mixture model (GMM) is used for classification of instantaneous amplitudes and phases of binary amplitude-shift keying (BASK), binary frequency-shift keying (BFSK), QPSK, 16-QAM, and 64-QAM signals and it shows better performance compared to [5, 6].

In this paper, we propose a feature-based AMC scheme which uses CCs as features and the GMM for feature classification. We deal with both linear and nonlinear modulations, such as minimum-shift keying (MSK), Gaussian MSK (GMSK), BPSK, QPSK, 8-PSK, 16-QAM, 32-QAM, and 64-QAM.

This paper is organized as follows. In section II, we explain the signal model and entire framework of the proposed scheme. Brief review about CCs used for features is given in section III. The proposed AMC adopting GMM is explained in section IV. Simulation results are demonstrated in section V to evaluate the performance. Finally, the paper concludes in Section VI.

II. System Model

The proposed modulation classification scheme consists of two parts of a feature extraction block and a GMM classifier as shown in Fig. 1. Two signals, $y(n)$ and $r(n)$ are entered to the first block. The received baseband signal $y(n)$ at time *n* is represented as

$$y(n) = \sum_{l=-\infty}^{\infty} a(l)\, p(nT_s - lT)\, e^{j(2\pi \Delta f n T_s + \theta_c)} + w(nT_s) \qquad (1)$$

where $a(l)$ is the *l*-th transmitted symbol and assumed to be independent and identically distributed (IID), which generally holds in digital communications, with unit variance. T_s and T denote the sampling period and symbol

 ISBN: 978-1-61208-353-7

duration, respectively. And Δf is the residual carrier frequency offset and θ_c the residual carrier phase. The oversampling ratio is defined by $\rho = T_s / T$. The pulse $p(t)$ reflects the channel effects and $w(t)$ is zero-mean additive white Gaussian noise (AWGN).

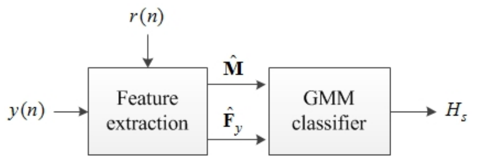

Figure 1. The structure of the proposed modulation classification scheme.

The input $r(n)$ is the signal to generate the reference data for GMM, which has the same modulation type as $y(n)$ with AWGN. The signal is required only for the learning process.

At feature extraction stage, we extract the feature vector $\hat{\mathbf{F}}_y$ and the reference data matrix $\hat{\mathbf{M}}$ from $y(n)$ and $r(n)$, respectively. Then, the GMM classifier is used to predict the modulation type H_s corresponding to the received signal.

III. CYCLIC CUMULANTS (CCS)

Most communication signals made by human represent cyclostationary characteristic, i.e., statistical properties of the signal are varying periodically with respect to time. Up to now, the CCs are widely used for AMC among many cyclostationary statistics of the signal.

The k-th order with q-conjugate (k, q) CCs of $y(n)$ are defined as Fourier coefficients of time-varying k-th order cumulants [8] as follows

$$C_y(\beta, \boldsymbol{\tau})_{k,q} \triangleq \left\langle C_y(n, \boldsymbol{\tau})_{k,q} e^{-j2\pi\beta n} \right\rangle \quad (2)$$

where $< \cdot >$ means sample average, β is the cycle frequency, and $\boldsymbol{\tau} = [\tau_1, \cdots, \tau_n]^T$ is the vector of time lags.

Let $P = \{v_j\}_{j=1}^{p}$ be a set of partitions v_j for the index set $\{1, 2, ..., n\}$, and p is the number of elements in the set. For example, there are five different sets of P for $k = 3$, i.e., {(1, 2, 3)}, {(1), (2, 3)}, {(2), (1, 3)}, {(3), (1, 3)}, and {(1), (2), (3)}, and p is 1, 2, 2, 2, and 3, in order. Then, (2) can be further expressed in terms of cyclic moments according to the relation between moments and cumulants [8] as

$$C_y(\beta, \boldsymbol{\tau})_{k,q} = \sum_P (-1)^{p-1}(p-1)! \left[\sum_{\boldsymbol{\alpha}^T \mathbf{I} = \beta} \prod_{j=1}^{p} M_y(\alpha_j, \boldsymbol{\tau}_{v_j})_{k_j, q_j} \right] \quad (3)$$

where $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_m]^T$ is the vector composed of cycle frequencies, $\mathbf{I}$ is the m-dimensional vector whose all m elements are ones, and $M_y(\alpha_j, \boldsymbol{\tau}_{v_j})_{k_j, q_j}$ is the (k_j, q_j) cyclic moment of $y(n)$ at cycle frequency α_j and delay vector $\boldsymbol{\tau}_{v_j} = [\tau_1, \cdots, \tau_{k_j}]^T$. Then (k_j, q_j) cyclic moment is denoted by

$$M_y(\alpha_j, \boldsymbol{\tau}_{v_j})_{k_j, q_j} = \left\langle \prod_{\eta=1}^{q_j} y^*(n+\tau_\eta) \prod_{\xi=q_j+1}^{k_j} y(n+\tau_\xi)\, e^{-j2\pi\alpha_j n} \right\rangle \quad (4)$$

where * means complex conjugate.

Since the CCs are just Fourier series coefficients of the cumulants as shown in (2), we can notice that the properties of the CCs follow those of the cumulants. Therefore, it is worth examining the properties of the cumulants. Some important properties of the cumulants are as follows [9]:

- If $y(n)$ is Gaussian random process, its cumulants higher than second order are zero.
- The cumulants of the sum of the independent random processes are equal to the sums of their cumulants.
- All odd-order cumulants are equal to zero when the distribution is symmetric.

Cumulants higher than second order are called higher order cumulants. They characterize statistical amplitude features such as the shape of a signal regardless of Gaussian noise. Digitally modulated signal has its own distinctive pulse shape and amplitude distribution, therefore it shows different higher order CCs. On the other hand, the odd-order CCs are zero due its symmetric amplitude distribution. In this paper, we use higher even-order CCs such as fourth, sixth, and eighth order CCs as features to identify the modulation type. This is why the computational complexity greatly increases as the order of CC becomes higher.

Fig. 2 represents the magnitudes of the various (k, q) CCs with the eight modulation signals. These values are calculated from (3) with 10 trials per each modulation type. In this figure, BPSK signal is omitted because its CC values are much greater than others. We notice that each modulation type has distinct values with different kinds of (k, q) CCs. In the case of high level modulation signal (e.g., QPSK and 8-PSK), the overlapping of fourth and sixth order CCs become to split in eighth order CCs. And, except (8, 0) and (8, 4) CCs, all CCs are nearly constant over SNR variation.

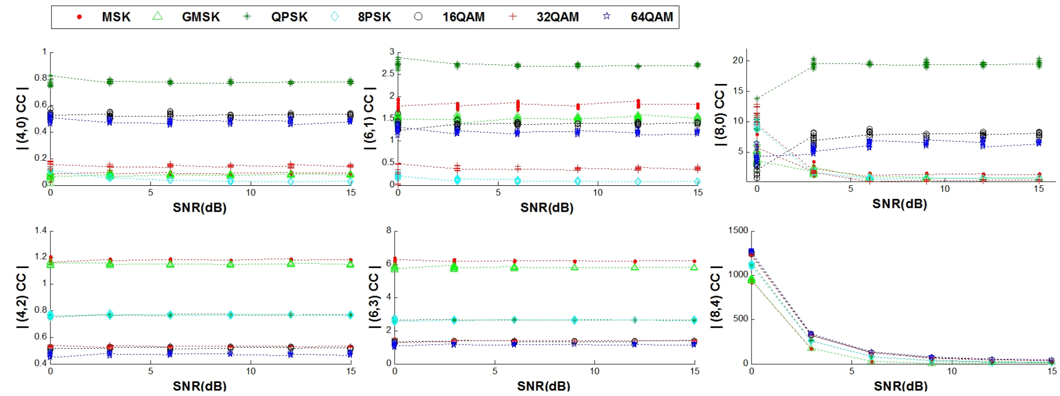

Figure 2. The magnitudes of the *k*-th order with *q*-conjugate (*k*, *q*) cyclic cumulants for eight modulation signals with various SNR.

IV. PROPOSED MODULATION CLASSIFICATION SCHEME

A. Feature extraction

From the results of Figure 2, we choose the three magnitudes of (4, 2), (6, 3), (8, 0) CCs as the features. With the number of elements $N_F = 3$, the feature vector $\mathbf{F} \in \mathbb{C}^{N_F}$ of CC magnitudes at zero-time lag is given by

$$\mathbf{F} = [\ |C_y(\beta,\mathbf{0})_{4,2}|,\ |C_y(\beta,\mathbf{0})_{6,3}|,\ |C_y(\beta,\mathbf{0})_{8,0}|\ ]^T \quad (5)$$

We estimate the CCs from (3) using the magnitude of the maximum value of cyclic moments which are obtained by FFT operation of (4). We select the number of FFT points as follows

$$N_{FFT} = 2^{\lceil \log_2(N\rho) \rceil} \quad (6)$$

where $\lceil x \rceil$ means the smallest integer not less than *x* and N is the number of symbols.

In order to setup reference data, we generate baseband modulation signal $r(n)$ by increasing SNR with N_{snr} steps. Then, estimate the feature vector N_{trial} times for each modulation type. As a result, the total number of training samples is

$$N_{total} = N_{snr} N_{mod} N_{trial} \quad (7)$$

where N_{mod} is the number of candidate modulation signals to be identified.

The feature matrix for the reference data obtained from $r(n)$ is constructed as follows

$$\hat{\mathbf{M}} = \left[\hat{\mathbf{m}}_1, \ \cdots \ , \hat{\mathbf{m}}_{N_{\text{mod}}}\right] \in \mathbb{C}^{N_F \times N_{total}} \quad (8)$$

where $\hat{\mathbf{m}}_i \in \mathbb{C}^{N_F \times (N_{trial} \times N_{snr})}$ corresponds to the *i*-th modulation type, and is given by

$$\hat{\mathbf{m}}_i = [\hat{\mathbf{F}}_1^{\gamma}\ \hat{\mathbf{F}}_2^{\gamma}\ \cdots\ \hat{\mathbf{F}}_{N_{iter}}^{\gamma}, \cdots, \hat{\mathbf{F}}_1^{\gamma+N_{snr}-1}\ \hat{\mathbf{F}}_2^{\gamma+N_{snr}-1}\ \cdots\ \hat{\mathbf{F}}_{N_{iter}}^{\gamma+N_{snr}-1}] \quad (9)$$

where the superscript of the feature vector means SNR of γ to $\gamma + N_{snr} - 1$.

Fig. 3 represents the feature distribution of the reference data in a 3-dimensional feature space spanned by (4, 2), (6, 3), (8, 0) CCs. We observe that there are N_{mod} different feature clusters corresponding to different modulation types and they are not overlapped each other. Since we utilize noisy signals for the reference data, each feature cluster reveals a probability distribution with mean as its center.

Therefore, by using the probability distribution of the reference data, we can improve the performance of modulation classification.

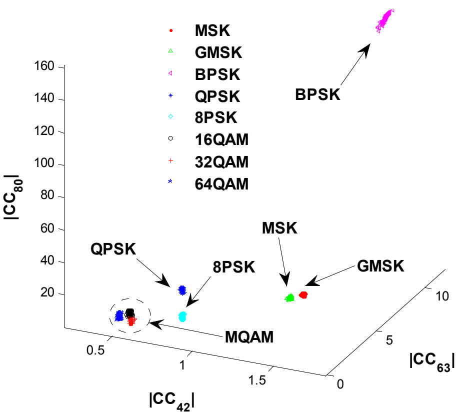

Figure 3. Reference data in a 3-dimentional feature space for the eight modulation signals with SNR of 3 to 10 dB.

B. GMM-based classifier

The modulation classification can be regarded as to find the cluster to which the feature vector $\hat{\mathbf{F}}_y$ of the received signal belongs. For classification, we use a GMM-based method which considers probability distribution of the reference data.

The method is based on the fact that, for the reference data $\hat{\mathbf{M}} = \{\hat{\mathbf{F}}_i\}_{i=1}^{N_{total}}$, it's unknown probability distribution can be represented by a weighted linear combination of multivariate Gaussian density functions, given by [10]

$$p(\hat{\mathbf{M}} \mid \mathbf{\Theta}) = \sum_{s \in S} p(\hat{\mathbf{M}} \mid C_s, \boldsymbol{\theta}_s) P_s \tag{10}$$

where S = {MSK, GMSK, BPSK, QPSK, 8-PSK, 16-QAM, 32-QAM, 64-QAM} is the set of eight modulation types, and $p(\hat{\mathbf{M}} \mid C_s, \boldsymbol{\theta}_s)$ is the N_F-variate Gaussian density function of the cluster C_s with unknown parameter $\boldsymbol{\theta}_s$, which is determined by a mean vector $\boldsymbol{\mu}_s \in \mathbb{C}^{N_F}$ and a covariance matrix $\boldsymbol{\Sigma}_s \in \mathbb{C}^{N_F \times N_F}$. P_s is the prior probability of the Gaussian density function for modulation *s*. The unknown parameter vectors can be collectively represented by

$$\mathbf{\Theta} = \{\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, P_s\} \tag{11}$$

The N_F-variate Gaussian density function is as follows

$$p(\hat{\mathbf{M}} \mid C_s, \boldsymbol{\theta}_s) = \frac{1}{(2\pi)^{N_F/2} \mid \boldsymbol{\Sigma}_s \mid^{1/2}} \cdot \exp\left(-\frac{1}{2} (\hat{\mathbf{M}} - \boldsymbol{\mu}_s)^T \boldsymbol{\Sigma}_s^{-1} (\hat{\mathbf{M}} - \boldsymbol{\mu}_s) \right) \tag{12}$$

The unknown parameter vectors can be obtained by the expectation-maximization (EM) algorithm which maximizes the expectation of the loglikelihood function of the GMM.

To apply the EM algorithm, the initial estimate $\mathbf{\Theta}(0)$ and a termination threshold ε are required for iteration. We estimate the mean vector and covariance matrix from the reference data for the initial estimate $\mathbf{\Theta}(0)$ and assume equal value of prior probabilities for clusters.

The EM algorithm is summarized as follows:

(1) Expectation step: At iteration λ, where $\boldsymbol{\theta}(\lambda)$ is available, compute the expected value of the followings:

$$Q(\mathbf{\Theta}; \mathbf{\Theta}(\lambda)) = \sum_{i=1}^{N_{total}} \sum_{s=1}^{S} P(C_s \mid \hat{\boldsymbol{F}}_i; \mathbf{\Theta}(\lambda)) \cdot \ln(p(\hat{\boldsymbol{F}}_i \mid C_s; \boldsymbol{\theta}_s) P_s \tag{13}$$

(2) Maximization step: Compute the next (λ+1)-th estimate of $\boldsymbol{\theta}$ by maximizing $Q(\mathbf{\Theta}; \mathbf{\Theta}(\lambda))$, that is,

$$\mathbf{\Theta}(\lambda+1) = \mathbf{\Theta}(\lambda) \text{ such that } \frac{\partial Q(\mathbf{\Theta}; \mathbf{\Theta}(\lambda))}{\partial \mathbf{\Theta}} = 0 \tag{14}$$

After some manipulations, the following general forms of parameters are derived as follows

Means:

$$\mu_j(\lambda+1) = \frac{\sum_{i=1}^{N_{total}} P(C_s \mid \hat{\boldsymbol{F}}_i; \mathbf{\Theta}(\lambda)) \hat{\boldsymbol{F}}_i}{\sum_{i=1}^{N_{total}} P(C_s \mid \hat{\boldsymbol{F}}_i; \mathbf{\Theta}(\lambda))} \tag{15}$$

Covariance Matrices:

$$\Sigma_s(\lambda+1) = \frac{\sum_{i=1}^{N_{total}} P(C_s \mid \hat{\boldsymbol{F}}_i; \mathbf{\Theta}(\lambda))(\hat{\boldsymbol{F}}_i - \boldsymbol{\mu}_s(\lambda))(\hat{\boldsymbol{F}}_i - \boldsymbol{\mu}_s(\lambda))^T}{\sum_{i=1}^{N_{total}} P(C_s \mid \hat{\boldsymbol{F}}_i; \mathbf{\Theta}(\lambda))} \tag{16}$$

Prior Probabilities:

$$P_s(\lambda+1) = \frac{1}{N_{total}} \sum_{i=1}^{N_{total}} P(C_s \mid \hat{\boldsymbol{F}}_i; \mathbf{\Theta}(\lambda)) \tag{17}$$

For the *i*-th reference data and cluster C_s, the a posteriori probability for Gaussian density function of C_s is given by

$$p(C_s \mid \hat{\boldsymbol{F}}_i; \mathbf{\Theta}(\lambda)) = \frac{p(\hat{\boldsymbol{F}}_i \mid C_s; \boldsymbol{\theta}_s(\lambda)) P_s(\lambda)}{\sum_{s=1}^{S} P(\hat{\boldsymbol{F}}_i \mid C_s, \boldsymbol{\theta}_s(\lambda)) P_s(\lambda)} \tag{18}$$

If the following termination condition is not satisfied, the iteration (1) and (2) continues.

$$\| \mathbf{\Theta}(\lambda+1) - \mathbf{\Theta}(\lambda) \| < \varepsilon \tag{19}$$

Based on the estimated GMM of the reference data, the modulation type of the received signal can be determined. After computing the a posteriori probability between GMM of each cluster and the feature vector $\hat{\mathbf{F}}_y$ of the received signal, select the cluster representing the maximum value of a posteriori probability, and decide the corresponding modulation type as that of the received signal, as follows

$$H_s = \arg\max_{s \in S} P(C_s \mid \hat{\mathbf{F}}_y; \boldsymbol{\theta}_s) \tag{20}$$

V. SIMULATION RESULTS

We use eight baseband modulation signals of MSK, GMSK, BPSK, QPSK, 8-PSK, 16-QAM, 32-QAM, and 64-QAM to be classified. That is, the number of candidate modulation signals to be identified is $N_{mod} = 8$. The symbol rate and oversampling ratio are $T = 1$ and $\rho = 11$, respectively. A raised-cosine filter with roll-off factor 0.35 is used for pulse shaping in generation of *M*-PSK and *M*-QAM signals. For GMSK, the bandwidth-time product for Gaussian filter is 0.5. We assume that the relative carrier frequency offset Δf is 0 and phase offset $\Delta\theta$ is uniformly distributed over $[-\pi, \pi)$. For the reference data, the modulation signal is generated with 3 to 10 dB of SNR, or $\gamma = 3$ and $N_{snr} = 7$. The initial estimate of the GMM parameter, $\mathbf{\Theta}(0)$, and the termination condition ε are set to be as in Table I and 1e-6, respectively.

TABLE I. PARAMETER ESTIMATE OF REFERENCE DATA

Mod.	Mean			Covariance ($\times 10^{-4}$)			Prior prob.
MSK	1.25	7.0	2.07	0	0	4	0.125
				0	5	-16	
				4	-16	956	
GMSK	1.20	6.46	1.68	0	0	2	
				0	5	-5	
				2	-5	792	
BPSK	1.54	10.63	157.35	0	3	56	
				3	30	599	
				56	599	1.19	
QPSK	0.77	2.63	19.6	0	0	-2	
				0	2	-9	
				-2	-9	809	
8-PSK	0.76	2.63	0.7	0	0	-1	
				0	2	-3	
				-1	-3	287	
16-QAM	0.52	1.34	7.92	0	1	-2	
				1	3	-9	
				-2	-9	753	
32-QAM	0.52	1.35	0.32	0	1	1	
				1	3	1	
				1	1	475	
64-QAM	0.47	1.15	6.46	0	1	1	
				1	2	-2	
				-1	-2	769	

As a performance measure for AMC, we use the probability of correct classification P_c. To obtain the probability, 300 trials are performed in AWGN channels.

Fig. 4 shows the performance with respect to the number of symbols of the received signal at SNR = 10 dB. The performance improves as the number of symbols increases. At low SNR, the bad classification performance of *M*-QAM deteriorates the whole performance of the scheme. Clusters of *M*-QAM are closely located in the feature space as shown in Fig. 3. Therefore, when the number of symbols is small, the variance of the feature vectors is large and results in bad classification of *M*-QAM signal. We observe that the probability of correct classification approximates one for the number of symbols larger than 3000. Therefore, we determine the number of received symbols as 4000 for the next experiments.

Fig. 5 compares the performance of the proposed scheme

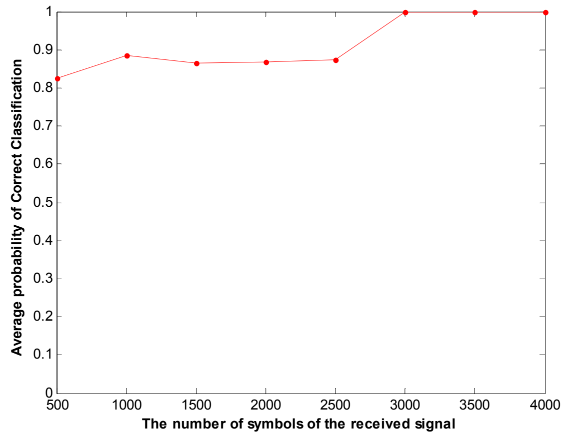

Figure 4. Performance of the proposed scheme with the number of symbols of the received signal at SNR = 10 dB.

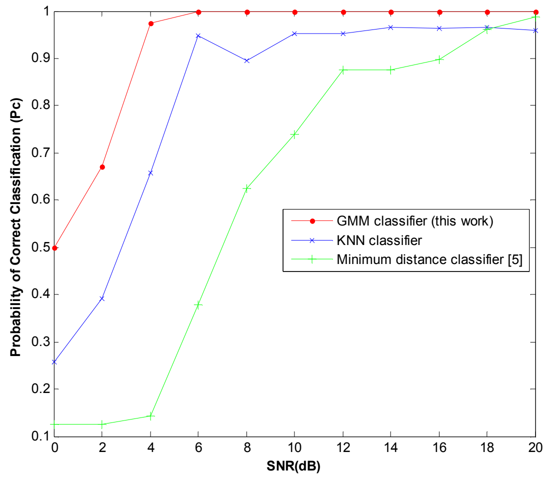

Figure 5. Comparison of the performance of the proposed scheme, KNN classifier, and the minimum distance classifier [5].

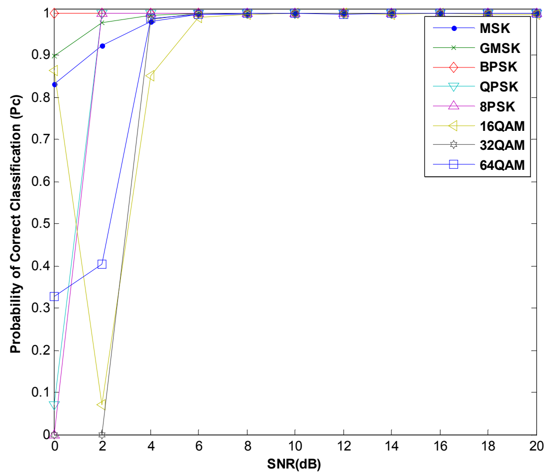

Figure 6. Performance of the proposed scheme for each modulation signal.

with the *K*-nearest neighbor (KNN) classifier and the minimum distance classifier of [5]. In [5], the eighth order cumulant with even number of conjugates is used as a feature. We observed that the performance of the proposed scheme is better than others over the entire range of SNR. When SNR is higher than 8 dB, the performance of the proposed scheme is nearly perfect.

Fig. 6 shows the performance for the individual modulation signal in the same conditions as those of Fig. 5. As shown in Fig. 5, the probabilities of correct classification for all modulation types become 1 with SNR higher than 6 dB. At lower SNR than 4 dB, the performance of the *M*-QAM degrades abruptly because of the reason mentioned above.

VI. Conclusion

In this paper, we proposed an feature-based automatic modulation classification scheme for eight digital modulation signals of MSK, GMSK, BPSK, QPSK, 8-PSK, 16-QAM, 32-QAM, and 64-QAM. We use the magnitude of fourth, sixth, and eighth order cyclic cumulants as features. To approximate the probability distribution of the features, the Gaussian mixture model is used. Based on the probability distribution, we identify the modulation type of the received signal according to the probability of its features. The simulation results show that the classification performance is much better as compared to the conventional K-nearest neighbor classifier and the minimum distance classifier.

Acknowledgment

This work has been supported by the National GNSS Research Center Program of DAPA and ADD.

References

[1] O. A. Dobre and R. Inkol, "Blind signal identification: Achievements, trends, and challenges," 9th International Conference on Communications (COMM), 2012, pp. 349-352)

[2] J. L. Xu, S. Wei, and Z. MengChu, "Likelihood-Ratio Approaches to Automatic Modulation Classification," IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 41, 2011, pp. 455-469.

[3] M. W. Aslam, Z. Zhu, and A. K. Nandi, "Automatic Modulation Classification Using Combination of Genetic Programming and KNN," IEEE Transactions on Wireless Communications, vol. 11, 2012, pp. 2742-2750.

[4] E. Like, V. Chakravarthy, P. Ratazzi, and Z. Wu, "Signal Classification in Fading Channels Using Cyclic Spectral Analysis," EURASIP Journal on Wireless Communications and Networking, 2009, pp. 879-812.

[5] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Cyclostationarity-Based Modulation Classification of Linear Digital Modulations in Flat Fading Channels," Wireless Personal Communications, vol. 54, 2009, pp. 699-717.

[6] O. A. Dobre, M. Oner, S. Rajan, and R. Inkol, "Cyclostationarity-Based Robust Algorithms for QAM Signal Identification," IEEE Communications Letters, vol. 16, 2012, pp. 12-15.

[7] J. G. Liu, W. Xianbin, J. Nadeau, and L. Hai, "Modulation classification based on gaussian mixture models under multipath fading channel," IEEE Global Communications Conference (GLOBECOM), Dec. 2012, pp. 3970-3974.

[8] W. A. Gardner and C. M. Spooner, "The cumulant theory of cyclostationary time-series. I. Foundation," IEEE Transactions on Signal Processing, vol. 42, 1994, pp. 3387-3408.

[9] C. L. Nikias and J. M. Mendel, "Signal processing with higher-order spectra," IEEE Signal Processing Magazine, vol. 10, 1993, pp. 10-37.

[10] S. Theodoridis and K. Koutroumbas, Pattern Recognition. Elsevier Science, 2008.