

Signal Identification Using a New High Efficient Technique

A. Ebrahimzadeh* and S. A. Seyedin**

Abstract: Automatic signal type identification (ASTI) is an important topic for both the civilian and military domains. Most of the proposed identifiers can only recognize a few types of digital signal and usually need high levels of SNRs. This paper presents a new high efficient technique that includes a variety of digital signal types. In this technique, a combination of higher order moments and higher order cumulants (up to eighth) are proposed as the effective features. A hierarchical support vector machine based structure is proposed as the classifier. In order to improve the performance of identifier, a genetic algorithm is used for parameters selection of the classifier. Simulation results show that the proposed technique is able to identify the different types of digital signal (e.g. QAM128, ASK8, and V29) with high accuracy even at low SNRs.

Keywords: Statistical pattern recognition; Signal identification; Support vector machine; Higher order moments, Higher order cumulants.

1 Introduction

ASTI plays an important role in various applications. For example, in military applications, it can be employed for electronic surveillance, monitoring; in civil applications, it can be used for spectrum management, network traffic administration, signal confirmation, software radios, multidrop networks, and intelligent modems. The early researches were concentrated on analogue signals, the recent contributions in the subject focus more on digital types of signals. Primarily, this is due to the increasing usage of such types of signal in novel communication applications.

Generally, ASTI techniques can be categorized in two main principles: the decision theoretic (DT) and the pattern recognition (PR). DT techniques use probabilistic and hypothesis testing arguments to formulate the recognition problem [1-4]. The major drawbacks of DT techniques are their too high computational complexity, lack of robustness to the model mismatch, the difficulties of forming the right hypothesis as well as careful analysis that are required to set the correct threshold values. However,

PR techniques don't need such careful treatment. They are simply implemented. PR techniques can be further divided into two main subsystems: the feature extraction and the classifier [5-18]. The former extracts the features and the latter determines the membership of signal.

Mobasseri [5], proposed an identifier that is based on the constellation shape. He used a Fuzzy-C means clustering method for classification of the PSK4, PSK8 and QAM16. The accuracy rate of the identification exceeded 90% for SNR>5dB. This method is sensitive to the initialization of clustering method. In [9], the authors proposed a technique for identification ASK2, ASK4, PSK2, PSK4, FSK2 and FSK4 signals. The classifier is based on a decision flow. These digital signal types have been identified with success rate about 90% at an SNR of 10 dB. In [10], the authors proposed a digital signal type identification technique based on elementary fourth-order cumulant. The authors were claimed that these features can separate a lot of digital signals. When it is used for identification of the BPSK, PAM4, QAM16 and PSK8, the success rate was about 96% at SNR=10 dB. In [11], the authors proposed a technique to discriminate among ASK, 4DPSK, 16QAM and FSK digital signals. The chosen features are: the kurtosis of the signal, the number of peaks in the phase probability density function (PDF) and the mean of the absolute value signal frequency. It is used a fuzzy classifier. For SNR> 5dB, the identifier worked properly. When SNR is less than 5dB, the performance was worse. In [12], for the first time, Ghani and Lamontagne proposed using a multi-layer perceptron

Iranian Journal of Electrical & Electronic Engineering, 2005.
Paper first received 22th June 2006 and in revised form 21th October 2006.

* The author is with the Department of Electrical and Computer Engineering Noushivani Institute of Technology, University of Mazandaran, Babol, Mazandaran, Iran.
E-mail: ataebrahim@yahoo.com.

** The author is with Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.
E-mail: seyedin@um.ac.ir.

(MLP) neural network with back-propagation (BP) learning algorithm for automatic signal type identification. They showed that neural network classifier outperforms other classifiers such as K-Nearest Neighbor (KNN). In [14], Nandi and Azzouz introduced two classifiers: neural network classifier and fixed threshold classifier, for analog and digital modulation recognition. They showed that the neural network classifier has better performance than the threshold classifier. The overall success rate is over 96% at the SNR of 15 dB. In [15], the authors proposed an identifier for recognition of PSK2 and PSK4. In this identifier a wavelet packet analysis (WPA) as a feature extractor and a multilayer perceptron (MLP) neural network that applies back-propagation (BP) with momentum and adaptive learning rate. This identifier showed a success rate higher than 92% at SNR>5dB. In [16], the authors proposed an identifier for identification of PSK2, PSK4, PSK8, OQPSK, MSK, QAM16, QAM64, FSK2 and FSK4 signal types. The features chosen to characterize the signal types are the mean and the next three moments of the instantaneous characteristics. They used different classifiers and showed that the artificial neural network has better performance than K-Nearest Neighbor (KNN) classifier and the well known binary decision trees. They reported a success rate of 90% with SNR ranges 15-25 dB. However, the performances for lower SNRs are reported to be less than 80%. In [17], the authors proposed an identifier based on cyclic spectral features for identification of AM, USB, LSB, FM, ASK, FSK, BPSK, QPSK and SQPSK. It was claimed that cyclic spectrum poses more advantage than power spectrum in signal type recognition. A full-connected backpropagation neural network is used for classification in this research. The success rate of this identifier is reported around 90% with SNR ranges 5-25 dB.

It can be found from the mentioned explanations that those techniques that use artificial neural networks (ANNs) as the classifier have better performance than others. However, with regard to effectiveness of ANNs, there are some problems. For example ANNs have limitations on generalization ability in low SNRs. In recent years, support vector machines (SVMs), based on statistical learning theory are gaining applications in area of pattern recognition because of excellent generalization capability. In [18], we have proposed an identifier that uses a binary SVM as the classifier. The accuracy of the proposed identifier is exceeded 98% for SNR>4dB. In this paper we propose a new and simple multiclass SVM-based classifier. It has a hierarchical structure.

From the published works it can be found that those identifiers, which use the statistical features, are able to include the digital signal types such as QAM. However, the features that are yet used didn't enable the identifiers to include the types of digital signals like QAM 32,

V29, Star-QAM8, and ASK8 etc. In this paper, we propose a combination of higher order moments and higher order cumulants (up to eighth) as the features. Suitable parameters of SVMs can improve the performance of identifier. We have proposed a genetic algorithm (GA) in order to selection the parameters of SVMs.

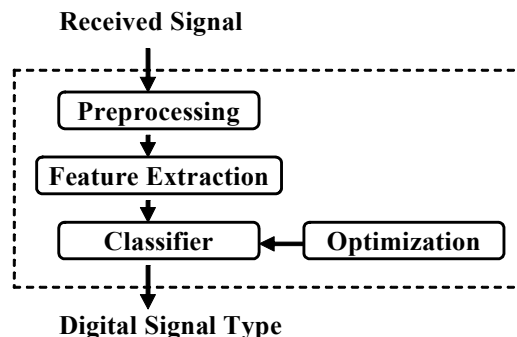


Fig. 1 General scheme of the proposed identifier

Fig.1 shows the general scheme of the proposed identifier. The preprocessing module performs: the rejection of noise outside of signal bandwidth, carrier frequency estimation, recovery of complex envelope, etc. This module is a common stage and we don't more explain. Section 2 presents the feature extraction module as well as the digital signal types that are considered in this paper. Section 3 describes the classifier module. Optimization problem using GA is presented in Section 4. Section 5 shows some simulation results. Finally, Section 6 concludes the paper.

2 Digital Signal Types and Features Extraction

In digital communications, according to the changes in the message parameters, we have four main digital signal types, FSK, ASK, PSK and QAM that most of them are used in M-ary form [19]. Different types of digital signal have different characteristics. Therefore finding the proper features for identification of them, particularly in case of higher order and/or non-square types, is a serious problem. Choosing 'bad' features may make it impossible even for an advanced classifier to perform a simple task, while choosing 'good' features may make it possible for simple classifier to solve complex problems [20].

In this paper the considered digital signal types (CDST) are: ASK2, ASK4, ASK8, PSK2, PSK4, PSK8, Star-QAM8, V29, QAM8, QAM64, QAM128, that for simplifying the indication, we substitute them with P_1 , P_2 , P_3 , P_4 , P_5 , P_6 , P_7 , P_8 , P_9 , P_{10} , and P_{11} respectively. Among the different features that we have computed and experimented, the combination of higher order moments and higher order cumulants (up to eighth) achieve the highest identification performances for CDST. These features can provide a good way to describe the shape of the probability density function.

Following subsections, briefly describe these features and computing method of them.

A. Moments

Probability distribution moments are a generalization of concept of the expected value. Recall that the general expression for the i^{th} moment of a random variable is given by [21]:

$$\mu_i = \int_{-\infty}^{\infty} (s - \mu)^i f(s) ds \quad (1)$$

where μ is the mean of the random variable. The definition for the i^{th} moment for a finite length discrete signal is given by:

$$\mu_i = \sum_{k=1}^N (s_k - \mu)^i f(s_k) \quad (2)$$

where N is the data length. In this study signals are assumed to be zero mean. Thus:

$$\mu_i = \sum_{k=1}^N s_k^i f(s_k) \quad (3)$$

Next, the auto-moment of the random variable may be defined as follows:

$$M_{pq} = E[s^{p-q} (s^*)^q] \quad (4)$$

where p is called the moment order and s^* stands for complex conjugation of s .

Assume a zero-mean discrete based-band signal sequence of the form $s_k = a_k + jb_k$. Using the definition of the auto-moments, the expressions for different orders may be easily derived. For example:

$$M_{41} = E[(a + jb)^3 (a - jb)] = E[a^4 - b^4] \quad (5)$$

B. Cumulants

Consider a scalar zero mean random variable s with characteristic function:

$$\hat{f}(t) = E\{e^{jts}\} \quad (6)$$

Expanding the logarithm of the characteristic function as a Taylor series, one obtains:

$$\log \hat{f}(t) = k_1(jt) + \dots + \frac{k_r(jt)^r}{r!} + \dots \quad (7)$$

The constants k_r in (7) are called the cumulants (of the distribution) of the s . The symbolism for p^{th} order of cumulant is similar to that of the p^{th} order moment. More specially:

$$C_{pq} = \text{Cum}[\underbrace{s, \dots, s}_{(p-q)\text{ terms}}, \underbrace{s^*, \dots, s^*}_{(q)\text{ terms}}] \quad (8)$$

For example:

$$C_{81} = \text{Cum}(s, s, s, s, s, s, s, s, s^*) \quad (9)$$

C. Relation between Moments and Cumulants

The n^{th} order cumulant is a function of the moments of orders up to (and including) n . Moments may be expressed in terms of cumulants as:

$$M[s_1, \dots, s_n] = \sum_{\forall v} \text{Cum}[\{s_j\}_{j \in v_1}] \dots \text{Cum}[\{s_j\}_{j \in v_q}] \quad (10)$$

where the summation index is over all partitions $v = (v_1, \dots, v_q)$ for the set of indices $(1, 2, \dots, n)$, and q is the number of elements in a given partition. Cumulants may be also be derived in terms of moments:

$$\text{Cum}[s_1, \dots, s_n] = \sum_{\forall v} (-1)^{q-1} (q-1)! E[\prod_{j \in v_1} s_j] \dots E[\prod_{j \in v_q} s_j] \quad (11)$$

where the summation is being performed on all partitions $v = (v_1, \dots, v_q)$ for the set of indices $(1, 2, \dots, n)$. For example, assume $n = 3$. In such a case, the available set of indexes is $(1, 2, 3)$, and four different types of partitioning may be obtained for that set: $\{(1, 2, 3)\}$ leading to $q=1$, $\{(1), (2, 3)\}$ leading to $q=2$, $\{2, (1, 3)\}$ leading to $q=2$, $\{3, (1, 2)\}$ leading to $q=2$, $\{(1), (2), (3)\}$ leading to $q=3$. Therefore:

$$\begin{aligned} \text{Cum}[s_1, s_2, s_3] &= (-1)^{1-1} (1-1)! E[s_1, s_2, s_3] + \\ &\quad (-1)^{2-1} (2-1)! E[s_1] E[s_2, s_3] + \\ &\quad (-1)^{2-1} (2-1)! E[s_3] E[s_1, s_2] + \\ &\quad (-1)^{3-1} (3-1)! E[s_1] E[s_2] E[s_3] \\ &= E[s_1, s_2, s_3] - E[s_1] E[s_2, s_3] - \\ &\quad E[s_3] E[s_1, s_2] - E[s_1] E[s_2] E[s_3] \end{aligned} \quad (12)$$

In the same manner cumulants expressions up to eighth order can be computed. For example:

$$C_{80} = M_{80} - 35M_{40}^2 - 630M_{21}^4 + 420M_{20}^2 M_{40}^2 \quad (13)$$

We have computed all of the features for CDST. Table1 shows some of these features for a number of considered digital signal types. These values are computed under the constraint of unit variance in noise free.

Table1 Some of the features for a number of digital signal types

	P_1	P_4	P_5	P_7	P_{10}
M_{41}	2	1	0	0	0
M_{61}	4	1	-1	2.92	-1.3
C_{63}	4	16	4	.160	1.79
C_{80}	78	-244	34	-88.9	-11.5
C_{82}	78	-244	-46	63.31	-27.1

3 Classifier

We have proposed a multiclass SVM based classifier (MCSVMC) that has a hierarchical structure. SVM is a supervised machine learning method that was originally developed for binary classification problems, but it can be used for multiclass classification [22]. SVM uses structural risk minimization (SRM) principle whereas in ANN, empirical risk minimization (ERM) is used [23]. The difference in risk minimization leads to better

generalization performance for SVMs than ANNs. Following subsections describe the binary SVM classifier and multiclass SVM classifier.

A. Binary SVM

SVM performs classification tasks by constructing optimal separating hyperplanes (OSH). OSH maximizes the margin between the two nearest data points belonging to two separate classes.

Suppose the training set, $(x_i, y_i), i=1,2,\dots,l, x \in \mathbb{R}^d, y \in \{-1,+1\}$ can be separated by the hyperplane $\bar{w}^T x + b = 0$, where \bar{w} is weight vector and b is bias. If this hyperplane maximizes the margin, then the following inequality is valid for all input data:

$$y_i(\bar{w}^T x_i + b) \geq 1, \text{ for all } x_i, \quad i=1,2,\dots,l \quad (14)$$

The margin of the hyper-plane is $2/\|\bar{w}\|$. Thus, the problem is the maximizing of the margin by minimizing of $\|\bar{w}\|^2$ subject to (14). This is a convex quadratic programming (QP) problem and Lagrange multipliers $(\alpha_i, i=1,\dots,l; \alpha_i \geq 0)$ are used to solve it:

$$L_p = \frac{1}{2} \|\bar{w}\|^2 - \sum_{i=1}^l \alpha_i [y_i(\bar{w}^T x_i + b) - 1] \quad (15)$$

After minimizing L_p with respect to w and b , the optimal weights are given by:

$$w^* = \sum_{i=1}^l \alpha_i^* y_i x_i \quad (16)$$

The dual of the problem is given by [22]:

$$L_d = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (17)$$

To find the OSH, it must maximize L_d under the constraints of $\sum \alpha_i y_i = 0$. The Lagrange multipliers are only non-zero ($\alpha_i > 0$) when $y_i(\bar{w}^T x_i + b) = 1$. Those training points, for which the equality in (14) holds, are called support vectors (SV) that can satisfy $\alpha_i > 0$. The optimal bias is given by:

$$b^* = y_i - w^{*T} x_i \quad (18)$$

for any support vector x_i . The optimal decision function (ODF) is then given by:

$$f(x) = \text{sgn}(\sum_{i=1}^l y_i \alpha_i^* x^T x_i + b^*) \quad (19)$$

where α_i^* 's are optimal Lagrange multipliers. For input data with a high noise level, SVM uses soft margins can be expressed as follows with the introduction of the non-negative slack variables $\xi_i, i=1,\dots,l$:

$$y_i(\bar{w}^T x_i + b) \geq 1 - \xi_i \quad \text{for } i=1,2,\dots,l \quad (20)$$

To obtain the OSH, it should be minimizing the $\Phi = \|\bar{w}\|^2 / 2 + C \sum \xi_i^k$ subject to (20), where C is the penalty parameter, which controls the tradeoff between the

complexity of the decision function and the number of training examples, misclassified.

In the nonlinearly separable cases, the SVM map the training points, nonlinearly, to a highdimensional feature space using kernel function $K(\bar{x}_i, \bar{x}_j)$, where linear separation may be possible. One of the most famous kernel functions is Gaussian radial basis function (GRBF) given by (21).

$$K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2) \quad (21)$$

where σ is the width of the RBF kernel. After a kernel function is selected, the QP problem is:

$$L_d = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (22)$$

the α_i^* is derived by:

$$\alpha_i^* = \arg \max_{\alpha} L_d \quad (23)$$

$$0 \leq \alpha_i \leq C; i=1,2,\dots,l; \sum_{j=1}^l \alpha_j y_j = 0$$

After training, the following, the decision function, becomes:

$$f(x) = \text{sgn}(\sum_{i=1}^l y_i \alpha_i^* K(x, x_i) + b^*) \quad (24)$$

The performance of SVM depends on several free parameters like penalty parameter(C) and the kernel parameter. These parameters are called hyper-parameters. In this paper we have used the GRBF, because it shows better performance than other kernels. Thus hyper-parameters are: C and σ .

B. Multiclass SVM-based Classifier

There are two widely used methods to extend binary SVMs to multi-class problems [24]. One of them is called the one-against-all (OAA) method. Suppose we have a P-class pattern recognition problem, P independent SVMs are constructed and each of them is trained to separate one class of samples from all others. When testing the system after all the SVMs are trained, a sample is input to all the SVMs. Suppose this sample belongs to class P_1 ideally only the SVM trained to separate class P_1 from the others can have a positive response. Another method is called one-against-one (OAO) method. For a P-class problem, P(P - 1)/2 SVMs are constructed and each of them is trained to separate one class from another class. Again, the decision of a testing sample is based on the voting result of these SVMs.

In this paper we have proposed a novel simple effective MCSVM. It has a hierarchical structure. The received signal is fed to the first SVM (SVM1). SVM1 determines the received signal is belongs to what group. This process will be continued in the same manner until the signal types will be identified by the last SVMs. Fig.2 shows the scheme of hierarchal MCSVM. One of

the advantages this structure is that the number of SVMs is less than in cases of OAO and OAA.

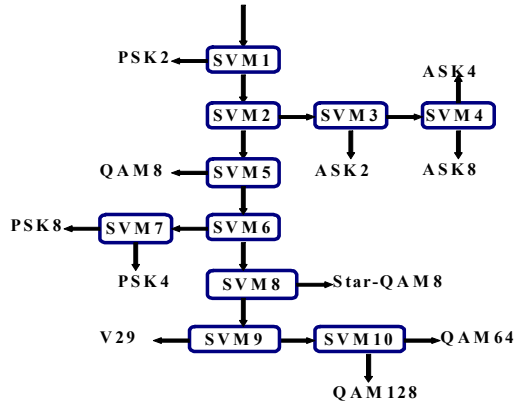


Fig. 2 The proposed Classifier with a hierarchical structure

4 Parameters Selection using GA

Finding the optimal value of hyper-parameters improves the performance of identifier. GAs with their characteristics of high efficiency and global optimization are widely applied in many areas. In this paper we have used GA for parameters selection of SVMs. GA is a stochastic optimization algorithm which adopts Darwin's theory of survival of the fittest [25]. To apply Gas, one has to consider the following issues: the encoding scheme, the methodology to produce the initial population, the fitness function and the genetic operators such as reproduction, crossover and mutation. How to encode a solution of the problem into a chromosome is a key issue for genetic algorithms. According the SVM parameters selection is an optimization problem with constraint, the real coded scheme of variables in a GA and a search operator that respects contiguous regions in the search space may be able to perform better than binary encoded scheme in constrained optimization problems with contiguous search space. Hence, real-encoded scheme is selected as the representation of the parameters in this paper. The research space of these parameters is $C \in [1:5:51]$, $\sigma \in [0.1:2]$.

In a GA, the selection of individuals to produce successive generations plays a vital role. There are several ways to select a new intermediate population. In this paper, we apply the elitism. Elitism is the name of the method that first copies the best chromosome (or few best chromosomes) to the new population. The rest of the population is constructed in ways described. Elitism can rapidly increase the performance of GA, because it prevents a loss of the best-found solution. In our application, we require that the best two strings are always to be included in the new population. This gives a chance to reevaluate their capabilities and improves GA convergence.

Genetic operators are the basic search mechanism of the GA for creating new solutions based on the existing population. The operators are two basic types: mutation

and crossover. Crossover produces two new individuals (offspring) from two existing individuals (parents). Crossover occurs with a crossover probability of P_C . A point is chosen for two strings where their genetic information's are exchanged. In this paper, we use one-point crossover, and typical value of P_C of 0.75. Mutation is intended to prevent falling of all solutions in the population into a local optimum of the solved problem. Mutation operation randomly changes the offspring resulted from crossover. In this paper, mutation occurs with typical mutation probability of 0.1.

GA will rate its own performance around that of the evaluation (fitness) function. The fitness function used in the present work returns the number of correct identification of the test data. The better identification results give rise to higher fitness index. To start the solution process, the GA has to be provided with an initial population. In this paper the random generation of initial solutions for the population is used [26]. The solution process continues from one generation to another selecting and reproducing parents until a termination criterion is satisfied. Convergence of a GA can be defined in several ways. In our application, the maximum number of generation is used as the terminating criterion.

5 Simulation studies

This section presents some simulation results of the proposed technique (identifier). All of the digital signal types are simulated in MATLAB environment. The simulated signals were also band-limited and Gaussian noise was added according to SNRs, -3, 0, 3, 6, 9, 12, and 18 dB. For each modulation type has 1320 realizations of 2048 samples. Among the features that we have computed in Section 2, Table2 shows the chosen features for each SVM. These features achieve the best results for signal type identification.

Table 2 Chosen features for each SVM

SVM's Number	Chosen features
SVM 1	C_{83}
SVM 2	M_{41}
SVM 3	M_{41}
SVM 4	C_{80}
SVM 5	C_{80}, M_{61}
SVM 6	C_{63}
SVM 7	C_{82}
SVM 8	C_{63}
SVM 9	C_{80}, M_{84}
SVM 10	C_{80}, C_{82}

A. Performance without Optimization

Based on some experiments, the values $\sigma=1$ and $C=10$ are selected for all SVMs. Table 3 shows the correct matrix at SNR= 3dB. Table 4 shows the identification results at different SNR values. It can be seen that the performance is generally very good even at low SNRs. Principally, this is due the two facts: chosen novel features and novel classifier. The chosen features have

effective properties in signal representation. On the other hand, the SVM based classifier has high generalization ability.

Table 3 Correct matrix of identifier without optimization at SNR=3dB

	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉	P ₁₀	P ₁₁
P ₁	93.1										
P ₂		94.2									
P ₃			94.3								
P ₄				97.5							
P ₅					96.1						
P ₆						91.2					
P ₇							90				
P ₈								90.2			
P ₉									90.3		
P ₁₀										91.5	
P ₁₁											90.2

Table 4 Performances of identifier (without optimization) at different SNRs (%)

SNR	Training	Testing
-3	85.92	85.28
0	92.12	91.78
3	93.95	93.15
6	97.28	97.14
9	97.52	97.45
18	98.63	98.42

In order to compare the performance of hierarchical SVM-based classifier with another classifier, we have considered a hierarchical MLP-based classifier that SVMs are replaced with MLP neural networks. The simulation setups are the same. We name this technique as TECH2. Fig. 3 shows the performances of two identifiers in different SNR values that term P_C means the percentage of correct classification. It can be seen that our proposed technique (PROTECH) that uses SVM in the structure of classifier has a better performance than of TECH2, particularly for low SNR values. When the SNR is low, TECH2 shows poor performance while in higher SNRs the percentage of correct classification is high. The construction of neural network in low SNRs is not proper, which results in low generalization ability. In higher SNRs the features are proper and closer to the noiseless state and it is easier to construct the neural network and results in high identification percentage.

B. Performance with Applying the Genetic Algorithm

In this subsection we apply the genetic algorithm for model selection of SVMs. Table 5 shows the performances of the optimized identifier for various SNRs. It can be seen that the optimization improves the performances of identifier for all SNRs; especially in lower SNRs. Table 6 shows the optimum parameters of SVMs that are used in hierarchical structure. Table 7 indicates the correct matrix of identifier at SNR= 3dB. Also, we have evaluated the performance of the optimized identifier at a high SNR value. Table 8

indicates the training performance of identifier at SNR= 39dB. The classifier can show up to 100% accuracy.

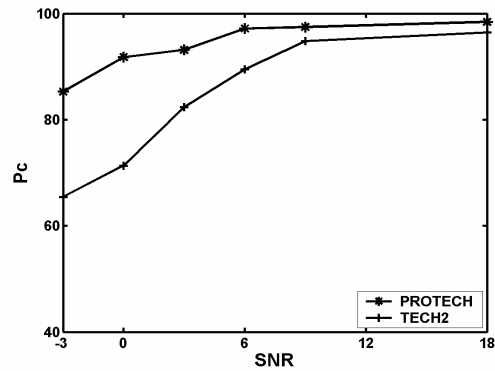


Fig. 3 Performance comparison of PROTECH and TECH2 at different SNRs

To indicate the effectiveness of chosen features, we have used the features that have been introduced in [12]. The structure of the classifier and the simulation setup are the same. We name this technique as TECH3. Fig.4 shows the performances of two identifiers. Results imply that our chosen have highly effective properties in signal representation.

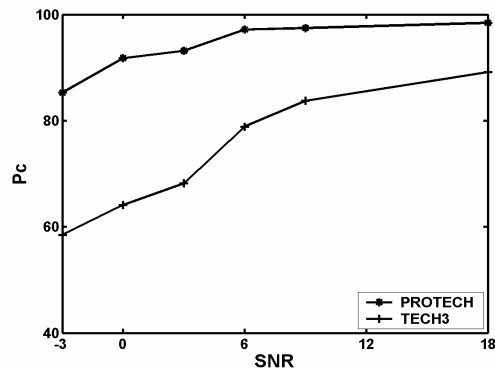


Fig. 4 Performance comparison of PROTECH and TECH3 at different SNRs

Table 5 Performances of identifier with applying the genetic algorithm at different SNRs

SNR (dB)	Training	Testing
-3	92.47	91.93
0	94.56	93.54
3	97.15	96.55
6	98.83	98.51
9	99.02	98.82
18	99.28	99.18

Table 6 Optimum parameters of SVMs

SVM's Number	C	σ
SVM 1	10	1.2
SVM 2	21	1
SVM 3	16	0.8
SVM 4	11	1
SVM 5	16	0.9
SVM 6	21	1.5
SVM 7	16	0.8
SVM 8	26	1.1
SVM 9	46	1.4
SVM 10	6	1.2

Table 7 Confusion matrix of optimized identifier at SNR=3dB

	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉	P ₁₀	P ₁₁
P ₁	97.2										
P ₂		96									
P ₃			96.8								
P ₄				100							
P ₅					98						
P ₆						96.2					
P ₇							95				
P ₈								95.1			
P ₉									95.6		
P ₁₀										95.7	
P ₁₁											95

Table 8 Confusion matrix of optimized identifier at SNR=39dB

	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉	P ₁₀	P ₁₁
P ₁	99.8										
P ₂		100									
P ₃			100								
P ₄				99.6							
P ₅					99.7						
P ₆						100					
P ₇							99.8				
P ₈								99.7			
P ₉									100		
P ₁₀										99.6	
P ₁₁											99.5

As mentioned in [6], direct comparison with other works is difficult in signal type identification. This is mainly because of the fact that there is no single unified data set available. Different setup of digital signal types will lead to different performance. In comparison with other works that we are mentioned, the identifier that we introduced in this paper has many advantages. This identifier has a simple structure and includes a variety of digital signal types and shows great generalization ability for identification of ASK2, ASK4, ASK8, PSK2, PSK4, PSK8, Star-QAM8, V29, QAM8, QAM64 and QAM128 digital signal types: It has a success rate of around 92% at SNR= 0 dB. The performances of the identifier is higher than 98% for SNR>6dB. None of the previous works have such a performance. The proposed identifier is fast in terms of training time. If it were known that changes have occurred, for example, the identifier can easily be trained.

6 Conclusions

ASTI has seen increasing demand in different applications. Most of techniques can only identify a few kinds of digital signal and usually need high SNRs. This paper presents a high efficient technique for identification of digital signal types. In this technique a hierarchical multiclass classifier based on SVMs is proposed. The inputs of this classifier are the combination of higher order moments and higher order cumulants. Each SVM use the features vector and maps the input vectors non-linearity into high dimensional feature space and constructs the optimum separating hyperplane in the space to realize signal recognition.

This technique avoids the overfitting and local minimum. Chosen features of the higher order moments and the higher order cumulants have high ability to signal representation. Optimization using GA, improves the performance of system especially in lower SNRs. The proposed identifier includes different kinds of digital signal and can separate them with high accuracy even at low SNR values.

Acknowledgements

The authors would like to thank Iran Telecommunication Research Center (ITRC) for funding this work.

References

- [1] C. Le Martret and D. Boitea, "A general maximum likelihood classifier for modulation classification," *Proc. ICASSP*, pp. 2165-2168, 1998.
- [2] J. A. Sills, "Maximum-likelihood modulation classification for PSK/QAM," *Proc. MILCOM*, pp. 57-61, 1999.
- [3] W. Wei and J. M. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE Trans. Comm.*, Vol. 48, pp. 189-193, 2000.
- [4] P. Panagotiou and A. Polydoros, "Likelihood ratio tests for modulation classifications," *Proc. MILCOM*, pp. 670-674, 2000.
- [5] K. Assaleh, K. Farrell and R. J. Mammone, "A new method of modulation classification for digitally modulated signals," *Proc. MILCOM*, pp. 712-716, 1992.
- [6] B. G. Mobaseri, "Digital modulation classification using constellation shape," *Signal Processing*, Vol. 80, pp. 251-277, 2000.
- [7] A. Hero and H. Maherem, "Digital modulation classification using power moment matrices," *Proc. ICASSP*, pp. 3285-3288, 1998.
- [8] H. B. Gaan, "Modulation classification based on spectrogram," *Proc. ICMLC*, pp. 3551-3556, 2004.
- [9] E. E. Azzouz and A. K. Nandi, "Automatic identification of digital modulations," *Signal Processing*, Vol. 47, pp. 55-69, 1995.
- [10] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Comm.*, Vol. 48, pp. 416-429, 2000.
- [11] J. Lopatka and P. Macrej, "Automatic modulation classification using statistical moments and a fuzzy classifier," *Proc. ICSP*, pp.121-127, 2000.
- [12] N. Chani and R. Lamontagne, "Neural networks applied to the classification of spectral features for

automatic modulation recognition,” *Proc. MILCOM*, pp. 1494-1498, 1993.

- [13] G. Hagedorn, B. James and C. Miller, “Neural network recognition of signal modulation types,” *Proc. ANNIEC*, pp. 170-175, 1997.
- [14] A. K. Nandi and E. E. Azzouz, “Algorithms for automatic modulation recognition of communication signals,” *IEEE Trans. Comm.*, Vol. 46, pp. 431-436, 1998.
- [15] A. Ebrahimzadeh and S. A. Seyedin, “A new method for automatic digital signal type identification,” *Proc. CIS*, 2005.
- [16] L. Mingquan, X. Xianci and L. Leming, “Cyclic spectral features based modulation recognition,” *Proc. Comm. Tech.*, pp. 792-795, 1998.
- [17] C. L. P. Sehier, “Automatic modulation recognition with a hierarchical neural network,” *Proc. MILCOM*, pp. 111-115, 1993.
- [18] A. Ebrahimzadeh and S. A. Seyedin, “Automatic modulation identification using wpa and svm,” *International Journal of Tomography and Statistics*, Vol. 4, pp. 17-28, 2006.
- [19] J. G. Proakis, *Digital Communications*. New York, USA: McGraw-Hill, 2001.
- [20] D. Stranneby, *Digital Signal Processing: DSP & Applications*. MATLAB Neural Network Toolbox, User's guide, The MathWorks, 2002.
- [21] P. McCullagh, *Tensor Methods in statistics*. New York, USA: Chapman & Hall, 1987.
- [22] C. Cortes and V. Vapnik, “Support Vector Network,” *Machine Learning*, Vol. 20, pp. 1-25, 1995.
- [23] C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” *Data Mining and Knowledge Discovery*, Vol. 2, pp 121-167, 1998.
- [24] B. Scholkopf, C. Burges and V. Vapnik, “Extracting support data for a given task,” *ICKDDM*, pp. 252-257, 1995.
- [25] G. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. New York, USA: Addison-Wesley, 1989.
- [26] Z. Michalewicz, *Genetic algorithms+Data Structures=Evolution Programs*. 3rd Edition. New York, USA: Springer, 1999.