

## Research Article

# Automatic Modulation Classification Exploiting Hybrid Machine Learning Network

Feng Wang , Shanshan Huang, Hao Wang, and Chenlu Yang

Array and Information Processing Laboratory, College of Computer and Information, Hohai University, West Focheng Road No.8, Jiangning District, Nanjing, 211100, China

Correspondence should be addressed to Feng Wang; jihonghope@aliyun.com

Received 2 September 2018; Revised 17 October 2018; Accepted 25 November 2018; Published 4 December 2018

Academic Editor: Ibrahim Zeid

Copyright © 2018 Feng Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is a research hot spot in cognitive electronic warfare systems to classify the electromagnetic signals of a radar or communication system according to their modulation characteristics. We construct a multilayer hybrid machine learning network for the classification of seven types of signals in different modulation. We extract the signal modulation features exploiting a set of algorithms such as time-frequency analysis, discrete Fourier transform, and instantaneous autocorrelation and accomplish automatic modulation classification using naive Bayesian and support vector machine in a hybrid manner. The parameters in the network for classification are determined automatically in the training process. The numerical simulation results indicate that the proposed network accomplishes the classification accurately.

## 1. Introduction

The automatic modulation classification (AMC) of the electromagnetic signals of radar and communication systems is an important function in modern electronic warfare systems [1–4]. The AMC process in a cognitive electronic warfare system for radar and communication signal surveillance is shown in Figure 1. The AMC mainly consists of feature extraction and modulation classification [5–7]. Modulation feature extraction is composed of a series of transform and analysis algorithms in time domain, frequency domain or time-frequency domain, such as time-frequency analysis [8], cyclic cumulant [9–12] and radar ambiguity function [13], and so on. The classification processing consists of various pattern recognition and machine learning algorithms, such as support vector machine (SVM) [14], deep learning [15], and clustering algorithms [16].

Pattern recognition has made a lot of progress and gained extensive applications in the field of computer vision, where SVM, artificial neural network and deep learning are widely used to realize image classification [17–19]. The AMC of the communication and radar signals can be regarded as an important branch of pattern recognition. A series of results have been reported in a range of open literatures.

To recognize the radar emitter signal in [20], the feature is extracted by signal fuzzy function slice and singular value decomposition, and modulation classification is obtained by the utilization of a kind of fuzzy clustering classification approach. In [21], a combination of the “Rihaczek distribution and Hough transform” algorithms is introduced to extract the features of the signals in time-frequency domain, and the AMC of radar signals in quadrature amplitude modulation (QAM) and phased shift keying (PSK) modulation is achieved. In [22], wavelet transform and manifold learning are employed to realize feature extraction and high-dimensional data dimensionality reduction respectively, and the nearest neighbor algorithm is adopted to realize classification of five types of signals (binary amplitude shift keying (2ASK), binary frequency shift keying (2FSK), binary phase shift keying (BPSK), linear frequency modulation (LFM) and Clock Pulse (CP)). A type of feature extraction method using blind channel estimation and cumulant is proposed in [23], and classification of the modulated signals is realized by a type of multiclassification algorithm based on maximum likelihood. [24] accomplishes feature extraction and modulation classification using high-order cumulant and binary tree based SVM, and verifies its performance using various signals such as 2ASK, 4ASK, quadrature phase shift

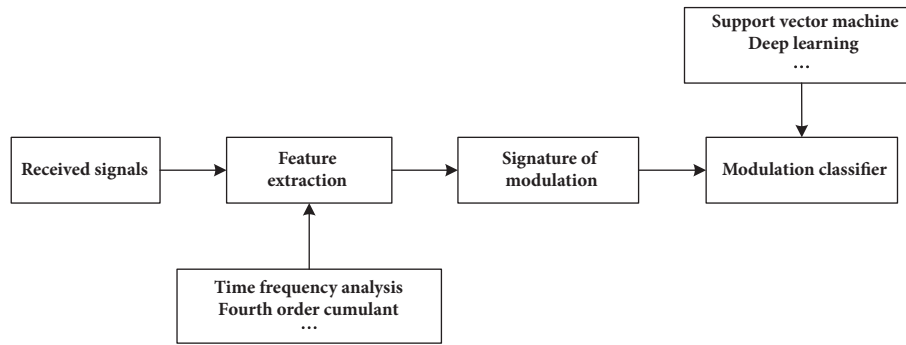


FIGURE 1: The structure of an AMC system for radar and communication signal surveillance.

keying (QPSK), 2FSK and 4-frequency shift keying (4FSK). [25] gives the modulation feature by the exploitation of a type of time-frequency image method based on local binary mode, and identifies the radar signals by SVM algorithm. In [26], genetic programming is carried out for extracting useful features, and the nearest neighbor algorithm is selected for the classification of certain types of radar signals.

It can be seen from the above literatures and several other related literatures [27–32] that there are several common problems in study of this field:

(1) The articles usually focus on signals emitted from either radar or communication systems, while electronic warfare systems in application may receive signals from both of these systems. That is to say, the current studied AMC algorithms lack a universal study on the classification of the radar and communication signals simultaneously. (2) The changes of the signal parameters are not considered in constructing the data sample libraries for training and test, though it has been demonstrated that these changes have a significant influence on the performance of the classification algorithms. (3) Without exploring the different multifeatures of the signal modulations, the AMC approaches proposed in many of the open literatures are based on only one feature, which makes the classification more difficult to implement. (4) Most studies overemphasize the use of advanced classification algorithms, while ignoring the study on signal feature extraction techniques [33]. The characteristics of most types of digital modulation used in modern communication and radar systems have significant differences, which can be obtained according to their definitions using certain feature extraction algorithms. The lack of these features actually increases the difficulties of the subsequent classification algorithms.

Different from the above research works, this paper emphasizes feature extraction and classification of the radar and communication signal simultaneously. Multidimensional modulation features in time-frequency domain, frequency domain, and envelope domain and phase domain are extracted with the utilization of a set of feature extraction algorithms, so as to find the differences of signals in different modulation as much as possible. It aims to reduce the pressure of the classification algorithm effectively and improve the classification accuracy. At the same time, a multilayer hybrid

classification network is constructed for the classification exploiting multiple features, and its effectiveness to improve the classification performance is tested by seven types of radar and communication signals commonly used in practical systems, including BPSK, QPSK, 16-quadrature amplitude modulation (16QAM), LFM, single frequency (SF), 2FSK, and 4FSK.

The rest of this paper is organized as follows. Section 2 gives the structure of the hybrid machine learning algorithm, and describes the principle of the signal modulation feature extraction and classification network. Section 3 analyzes the feature extraction algorithms for the signal set {BPSK, QPSK, 16QAM, LFM, SF, 2FSK, and 4FSK}, including time-frequency analysis, instantaneous auto-correlation (IA), and discrete Fourier transform (DFT). Section 4 analyzes the idea of the dimension reduction of the modulation features when using PCA and SVM for classification. Performance evaluation of the feature extraction and classification network is discussed in Section 5. The Conclusions are drawn in Section 6.

## 2. The Principle of the Hybrid Classification Network

Aiming at classifying the signal set {BPSK, QPSK, 16QAM, LFM, SF, 2FSK, and 4FSK}, this paper proposes a hybrid AMC network consisting of a variety of modulation feature extraction algorithms and machine learning classification methods. The overall framework of the hybrid AMC network is shown in Figure 2.

(1) *Classification of {SF, LFM} and {BPSK, QPSK, 16QAM, 2FSK, and 4FSK}*. According to Figure 2, short-time Fourier transform (STFT) algorithm is used in the first-layer of the network to extract the standard deviation of the first-order difference of the time-frequency spectrum peaks (recorded as  $\zeta_1$ ), identifying {SF, LFM} from {BPSK, QPSK, 16QAM, 2FSK, and 4FSK}.

According to the signal definition of SF and LFM, the first-order difference in time-frequency spectrum of these two signals is a fixed constant. Hence, the values of  $\zeta_1$  for these two modulation types are approximately zeroes, which is quite different from the large  $\zeta_1$  values of the other

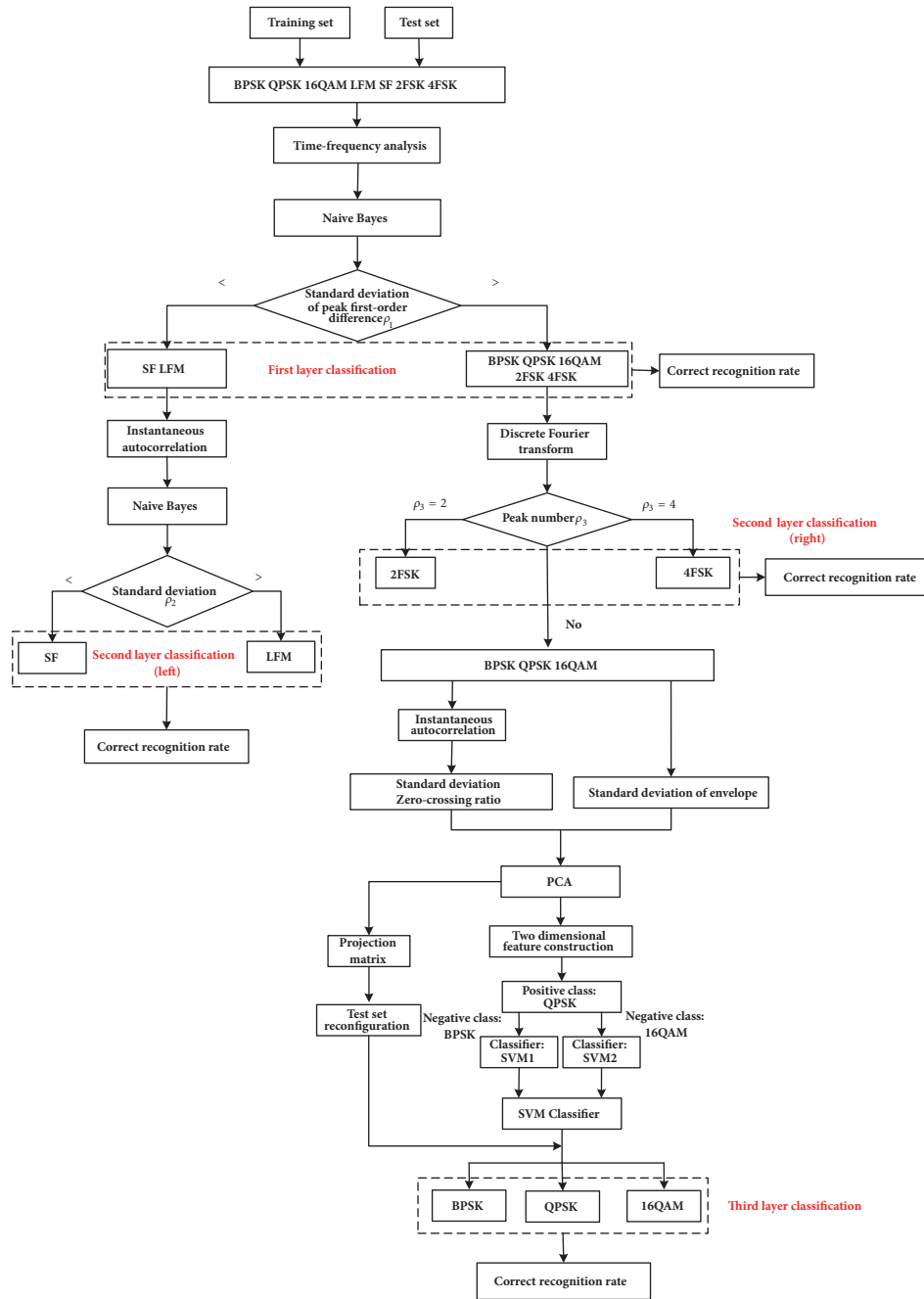


FIGURE 2: Diagram of the classification network.

signals. According to the decision threshold  $\rho_1$ , which can be obtained through naive Bayes [34] during the training period, the signal set {SF, LFM} can be identified from the other signals.

(2) *Classification of {SF} and {LFM}*. The left branch of the second-layer of the network implements the classification of the signal set {SF, LFM}. According to the difference of SF and LFM, the standard deviation feature based on the real part of IA is extracted and recorded as  $\zeta_2$ . LFM and SF can

be classified in terms of the decision threshold  $\rho_2$ , which can be obtained by the naive Bayes training.

(3) *Classification of {2FSK}, {4FSK}, and {BPSK, QPSK, 16QAM}*. The right branch of the second-layer implements the classification of the signal set {BPSK, QPSK, 16QAM, 2FSK, and 4FSK}. The frequency feature based on DFT is extracted for identification. The numbers of frequency peaks of 2FSK and 4FSK are 2 and 4, respectively, while the other three signals have multiple frequency peaks in the bandwidth.

The number of the frequency peaks is extracted and recorded as  $\zeta_3$ . According to the decision threshold  $\rho_3$ , which can be obtained by training, {2FSK4FSK} can be identified from the signal set {BPSK, QPSK, and 16QAM}.

(4) *Classification of {BPSK}, {QPSK}, and {16QAM}*. In the third-layer of the network, the remaining signal set {BPSK, QPSK, 16QAM} of the right branch is classified. Three features, including standard deviation of envelope, zero-crossing ratio, and standard deviation of the real part of IA, are extracted for classification. By determining the principal components with the contribution rate, PCA algorithm is used to reduce the dimensionality of the features from 3-dimension to two-dimension, making it suitable for the application of the one-to-one method of SVM. Regarding QPSK as a positive class, BPSK and 16QAM are sequentially substituted into the SVM classifier as a negative class to find the support vector. Two optimal boundaries are determined in light of the position of the support vector, and the classification of BPSK, QPSK, and 16QAM is accomplished finally.

The classification structure can be regarded as a machine learning network based on sample training and test. It is necessary to construct a large learning sample aggregate for extracting the modulation features of the above seven signals and determine the multiple thresholds for the multilayer classification during the training process. In the test phase, the predetermined thresholds are used for different class, and the correct recognition rate of each class is achieved in the end.

### 3. Feature Extraction Algorithms

The extracted modulation features are mainly based on the differences of radar and communication signals in time-frequency spectrum, frequency spectrum and phase modulation. Different algorithms are used to extract different features for signals of different class. The time-frequency feature is extracted by using STFT, which identifies the signal set {SF, LFM} from the other signals. SF and LFM are discriminated by the feature of the real part of IA. According to the number of frequency peaks which obtained based on DFT, 2FSK and 4FSK are identified from the signal set {BPSK, QPSK, and 16QAM}. The features based on the real part of IA are used to distinguish between BPSK, QPSK, and 16QAM. The following section will discuss the feature extraction algorithms including STFT, IA, and DFT.

**3.1. Feature Extraction Based on STFT.** Time-frequency feature of the signals can be extracted by STFT. For a discrete signal  $s(n)$  at discrete time instant  $n$ , its STFT is given by

$$S(m, k) = \sum_{n=0}^{N_s-1} s(n) h(n-m) e^{-j(2\pi/N_s)nk}, \quad (1)$$

$$k = 0, 1, 2, \dots, N_s - 1,$$

where  $k$  represents the discrete frequency and  $N_s$  is the total frequency number.  $m$  refers to time delay and  $h(n -$

$m)$  denotes the Rectangular window function. In (1), the nonstationary signal can be regarded as the superposition of a series of short-time stationary signals, which highlights the varying characteristics of the original signal frequency with time delay. The peak of the frequency along time delay dimension can be given by

$$P(m) = \max_k \{|S(m, k)|\}, \quad m = 0, 1, 2, \dots, N, \quad (2)$$

where  $N$  denotes the total number of windows and  $P(m)$  represents the maximum frequency peak corresponding to the  $m$ th window. The frequency peak of each time window can be extracted. The difference of frequency peak corresponding to two adjacent time windows can be given by

$$W(m) = P(m+1) - P(m), \quad m = 1, 2, \dots, N-1. \quad (3)$$

The standard deviation of the difference in (3) can be given by

$$\zeta_1 = \sqrt{\frac{1}{N-1} \sum_{m=1}^{N-1} [W(m) - \bar{W}]^2}, \quad (4)$$

where  $\bar{W}$  refers to the mean of  $W$  and  $\bar{W} = (1/(N-1)) \sum_{m=1}^{N-1} W(m)$ . The STFT of the seven waveform types with Signal-to-Noise Ratio (SNR) equal to 20 dB are illustrated in Figure 3. As can be seen from Figure 3(a), the frequency of SF is the same since it has only one frequency. The difference between the adjacent frequencies for SF is constant, so the standard deviation is zero. From Figure 3(b) we can see that the frequency of LFM changes linearly, which leads to a constant difference between the two adjacent frequencies. Hence, the standard deviation is also zero. Figures 3(c) and 3(d) show that the frequencies of 2FSK and 4FSK are variable, which means that the difference between the adjacent frequencies is not constant. Figures 3(e), 3(f), and 3(g) show that when there is a phase variation for BPSK, QPSK, or 16QAM signal, the instantaneous frequency has a large disturb, which leads to fluctuations between the adjacent frequencies. Hence, the standard deviations of the signals such as BPSK, QPSK, 16QAM, 2FSK, and 4FSK will be larger than those of the SF and LFM. In this case, SF and LFM can be identified from the other signals by setting the standard deviation threshold  $\rho_1$ .

**3.2. Feature Extraction Using IA [35].** Features based on IA can be extracted to distinguish between {SF, LFM} and {BPSK, QPSK, 16QAM}, respectively. The IA of a discrete signal  $s(n)$  is of the following form:

$$R(n, m) = s(n) \cdot s^*(n-m), \quad (5)$$

where  $m$  refers to time delay. The difference between the definition of IA and auto-correlation function lies in that there is no time integration in the calculation of IA. The advantage of using IA is that it retains the instantaneous phase information of the signal. The IA expressions of some of the signals are analyzed below.

(1) The signal of SF is of the following form

$$s_1(n) = Ae^{j(2\pi f_0 n + \varphi_0)}, \quad (6)$$



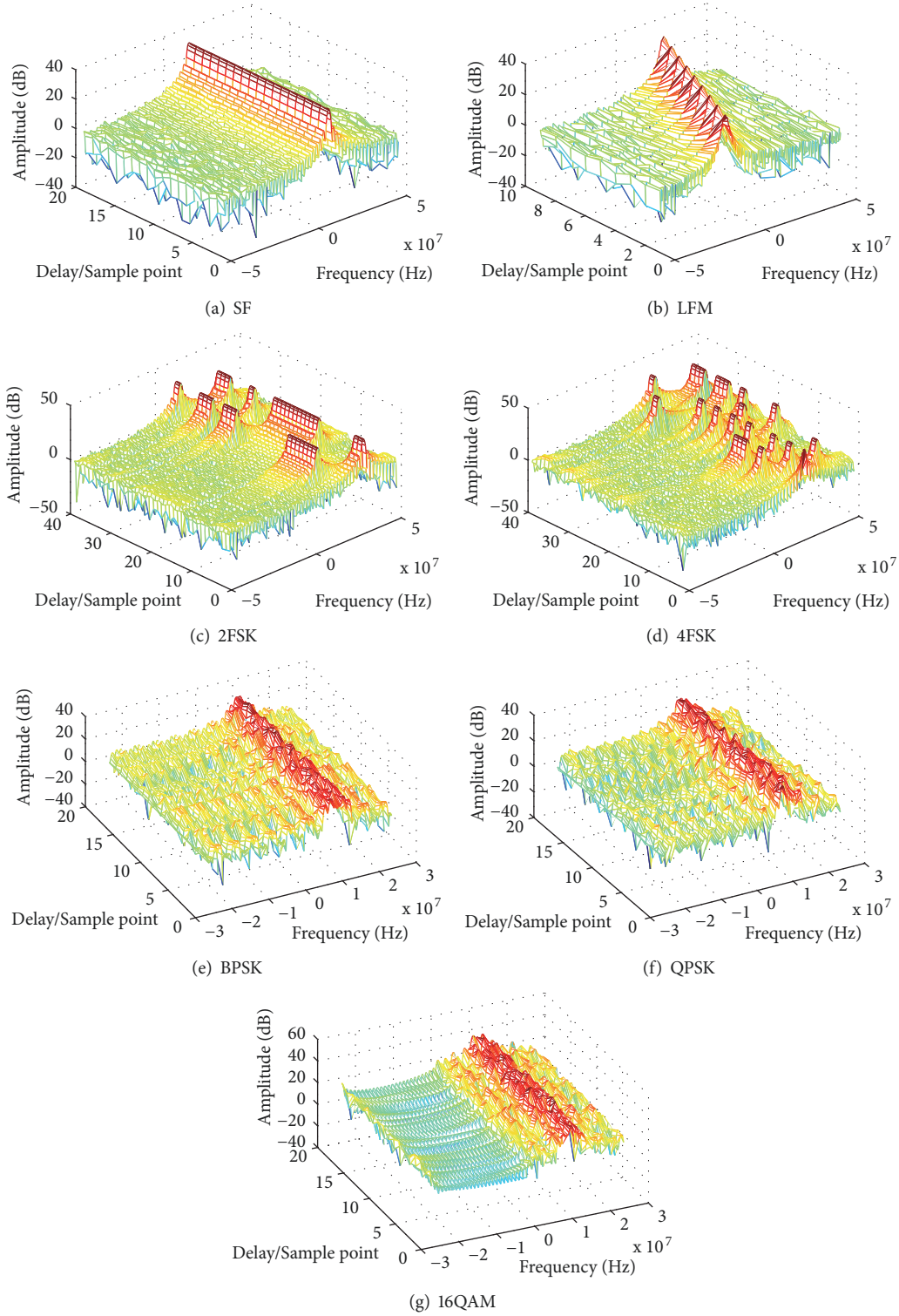


FIGURE 3: Time-frequency spectrum using STFT.

where  $A$  is the amplitude of the signal,  $f_0$  refers to the carrier frequency, and  $\varphi_0$  represents the initial phase of the signal. The real part of IA of SF can be given by

$$R(n, m) = A^2 \cos(2\pi f_0 m), \quad m \leq n \leq q, \quad (7)$$

where  $q$  is the number of samples of the signal. It can be seen from (7) that if  $m$  is certain, the IA of SF is related only to the carrier frequency, which is a constant. Hence, the output of the real part of IA is a direct-current (DC) signal as shown in Figure 4(a).

(2) The signal of LFM is of the following form:

$$s_2(n) = Ae^{j2\pi(f_0n + (1/2)\mu n^2)}, \quad (8)$$

where  $\mu$  is the slope of frequency modulation. The real part of IA of LFM can be given by

$$R(n, m) = A^2 \cos\left(2\pi\left(f_0m - \frac{1}{2}\mu m^2 + \mu mn\right)\right), \quad (9)$$

$$m \leq n \leq q,$$

It can be seen from (9) that the output of the real part of IA is an alternating current (AC) signal of frequency  $\mu m$ , which is shown in Figure 4(b).

(3) The expression of PSK can be given by

$$s_3(n) = Ae^{j(2\pi f_0n + \varphi_i)}, \quad (10)$$

where  $\varphi_i$  denotes the discrete phase of a code group representing BPSK or QPSK. For BPSK, the value of  $\varphi_i$  is 0 or  $\pi$ . For QPSK, the value of  $\varphi_i$  is 0,  $\pi/2$ ,  $\pi$ , or  $3\pi/2$ . The real part of IA of the PSK signal is of the following form:

$$R(n, m) = A^2 \cos(2\pi f_0m), \quad ip + m < n \leq (i+1)p$$

$$R(n, m) = A^2 \cos(2\pi f_0m + \varphi_{i+1} - \varphi_i), \quad (11)$$

$$(i+1)p < n \leq (i+1)p + m,$$

where  $p$  is the number of samples within one code and  $m < p$ . The real part of IA is DC within the same code period. In different code period, it can be divided into two cases: the adjacent code is the same ( $\varphi_{i+1} - \varphi_i = 0$ ) or different ( $\varphi_{i+1} - \varphi_i \neq 0$ ). For BPSK shown in Figure 4(c), the real part of IA is a two-value transition, of which  $\varphi_{i+1} - \varphi_i = 0$  corresponds to a positive transitions and  $\varphi_{i+1} - \varphi_i = \pm\pi$  corresponds to a negative transition. However, there is a status of  $\varphi_{i+1} - \varphi_i = \pm\pi/2$  for QPSK, for which the real part of IA is zero (the projection on the real axis). Therefore, the real part of IA for QPSK is a three-value output, which is illustrated in Figure 4(d).

(4) The signal of 16QAM can be expressed as

$$s_4(n) = A_i e^{j(2\pi f_0n + \varphi_i)}, \quad (12)$$

where  $A_i$  refers to the amplitude of the code group. The real part of IA of 16QAM is of the following form:

$$R(n, m) = A^2 \cos(2\pi f_0m), \quad ip + m < n \leq (i+1)p$$

$$R(n, m) = A_{i+1}A_i \cos(2\pi f_0m + \varphi_{i+1} - \varphi_i), \quad (13)$$

$$(i+1)p < n \leq (i+1)p + m.$$

From (13) we can see that when  $m$  is constant, the output of IA is DC in the same code period. However, it causes a phase jump  $\varphi_{i+1} - \varphi_i$  and amplitude transition  $A_{i+1} \cdot A_i$  between different code period. Hence, the output of the IA for 16QAM is a multivalued transition; see Figure 4(e).

Two features are extracted based on the real part of the IA.

*Feature 1.* Define the standard deviation of IA as

$$\zeta_2 = \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} (a(i) - \bar{a})^2}, \quad (14)$$

where  $a(i)$  is the value of the real part based on IA at time instant  $i$  and  $\bar{a} = (1/N_s) \sum_{i=1}^{N_s} a(i)$  represents the mean of  $a(i)$ . The standard deviation for SF signal will be small since the fluctuation of its IA is small. However, the IA of LFM fluctuates greatly; i.e., the standard deviation is larger. Under this circumstance, SF and LFM can be identified by setting the standard deviation threshold  $\rho_2$  of the real part of IA.

*Feature 2.* Define zero-crossing ratio as

$$\zeta_4 = \text{Num}\{a(i) \in \varepsilon_1\}, \quad (15)$$

where  $\text{Num}\{\cdot\}$  denotes a counter and  $\varepsilon_1$  refers to a small range belonging to zero (such as  $-0.001 < \varepsilon_1 < 0.001$ ). As shown in the Figure 4, a binary jump occurs for the IA of BPSK, meaning that there is no zero in the output. However, the IA of QPSK is of a three-value transition form with a large number of zeroes in the output. The IA of 16QAM is similar to QPSK. Therefore, the difference of zero-crossing ratio between BPSK and {QPSK, 16QAM} signals can be used as a classification feature.

**3.3. Feature Extraction Based on DFT.** For the remaining signal set {BPSK, QPSK, 16QAM, 2FSK, 4FSK}, the frequency spectrum features of the signals are extracted using DFT. According to the signal definitions, the peaks of 2FSK and 4FSK are 2 and 4 within the bandwidth, respectively. However, there are much more peaks for BPSK, QPSK and 16QAM.

Since frequency peaks of the signal set {BPSK, QPSK, 16QAM, 2FSK, and 4FSK} are different, the number of frequency peaks based on DFT can be extracted as a typical feature. For a discrete signal  $s(n)$ , its DFT can be given by

$$f(k) = \sum_{n=0}^{N_s-1} s(n) e^{-j(2\pi/N_s)nk}, \quad k = 0, 1, 2, \dots, N_s - 1, \quad (16)$$

where  $k$  represents the discrete frequency and  $N_s$  is the total frequency number. The number of peak can be defined as

$$\zeta_3 = \text{Num}\{|f(k)| > \varepsilon_2\}, \quad k = 0, 1, 2, \dots, N_s - 1, \quad (17)$$

where  $|\cdot|$  refers to the modulo operation and  $\varepsilon_2$  the threshold of frequency peak, taking 0.7 times of the maximum value. Since the number of peaks of 2FSK and 4FSK is smaller than the other three signals, 2FSK and 4FSK can be identified from other signals by setting the frequency peak threshold  $\rho_3$ .

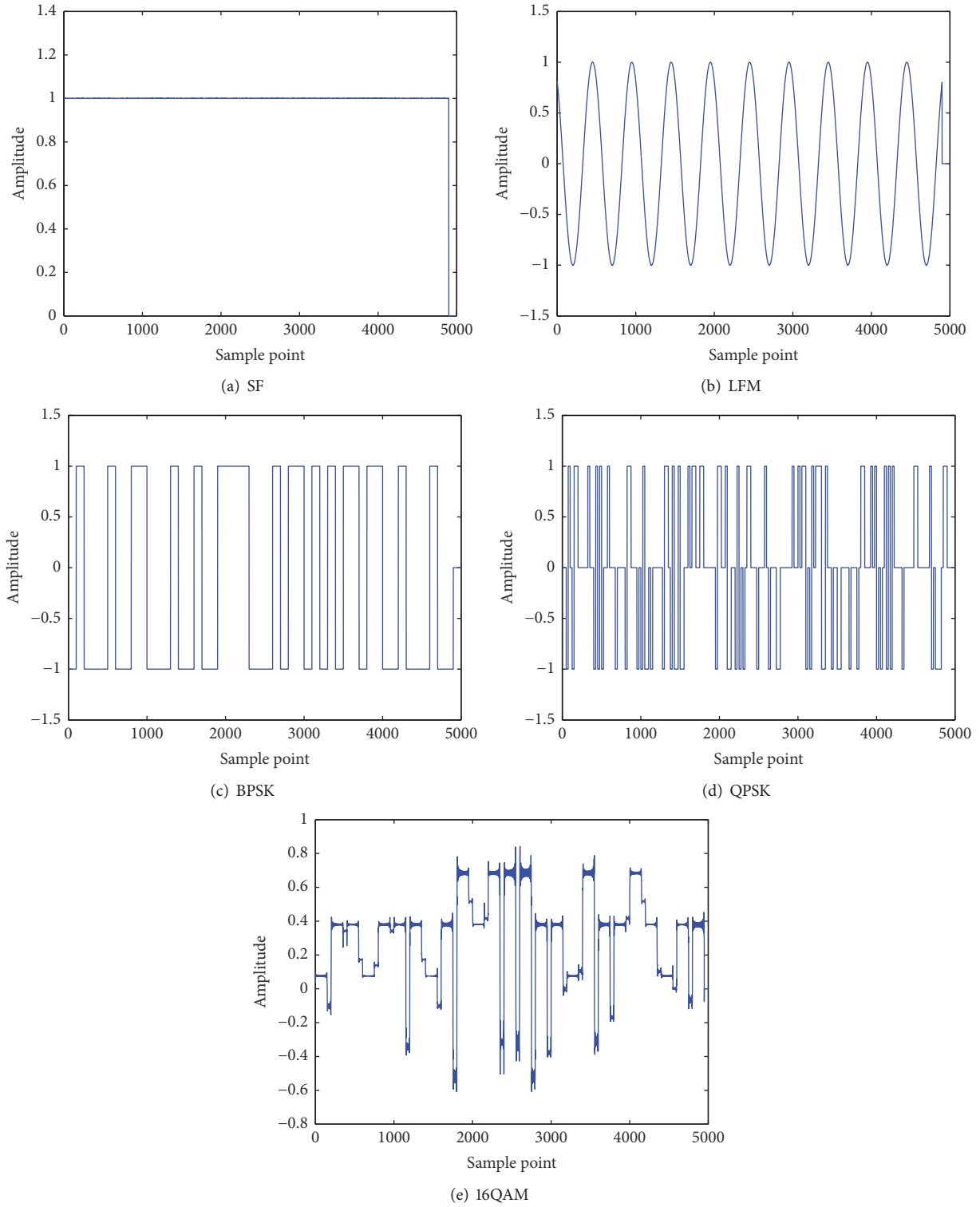


FIGURE 4: The real part of IA of the signals.

**3.4. Feature Extraction Based on Signal Envelope.** The multilevel amplitude of the 16QAM signal is quite different from the constant envelope BPSK and QPSK signal. Hence, envelope features in time domain can be used to classify BPSK, QPSK, and 16QAM. For a discrete signal  $s(n)$ , the

standard deviation of the envelope can be defined as

$$\zeta_5 = \sqrt{\frac{1}{N_s} \sum_{n=1}^{N_s} |s(n) - \bar{s}|^2}, \quad (18)$$

where  $\bar{s} = (1/N_s) \sum_{i=1}^{N_s} |s(i)|$  represents the mean of the instantaneous envelope.

#### 4. Analysis of SVM Based on PCA Dimensionality Reduction

Three features are extracted for the classification of BPSK, QPSK and 16QAM so as to ensure the classification accuracy under various conditions. Due to the large number of features, the classification tends to be complicated. If the three features can be replaced by the two features, SVM can be used to classify the three modulated signals in the Two-dimensional (2D) feature space. Therefore, PCA algorithm is used to perform principal component analysis on the Three-dimensional (3D) features, extracting principal components in features and reducing the dimension of features.

**4.1. PCA Algorithm.** The PCA algorithm transforms the original data with possible correlation into a set of new data with linear independence of each dimension through linear transformation, and it can be used to extract the principal feature components of the data, thereby achieving the purpose of dimensionality reduction [36]. The main idea is to map the  $k_1$  dimensional features to  $k_2$  dimension ( $k_2 < k_1$ ), which is a completely new orthogonal feature called the principal component. It can be easily understood that PCA can be used to find the most useful linear combination, *i.e.*, those new features with relatively large discrimination to achieve the purpose of reducing the dimension.

There are two basic requirements for PCA dimensionality reduction. First of all, the projections of the samples in the principal component direction are required to be as dispersed as possible. The more dispersed projections the larger the variance of the samples; *i.e.*, more useful information is carried in the reduced dimension projections. Secondly, the distances from the sample points to the principal component direction are required to be as small as possible, *i.e.*, the errors can be reduced as much as possible. The steps of the PCA dimensionality reduction algorithm [37] for  $k_1$ -dimensional modulation feature samples are summarized as follows:

- (1) Arrange the modulation feature samples into matrix  $X$  of  $M$  (sample numbers) rows and  $k_1$  columns.
- (2) Process the sample data, recorded as  $\bar{X}$ , including zero-meanization and normalization.
- (3) For the processed sample data, its covariance matrix can be given by

$$R_x = \frac{1}{M} (\bar{X}^T \bar{X}), \quad (19)$$

where  $[\cdot]^T$  refers to transposition operation.

- (4) According to

$$R_x u = \lambda u, \quad (20)$$

calculate the eigenvalue  $\lambda_i$  and the eigenvector  $u_i$  of  $R_x$ . Arrange the eigenvalues from large to small,

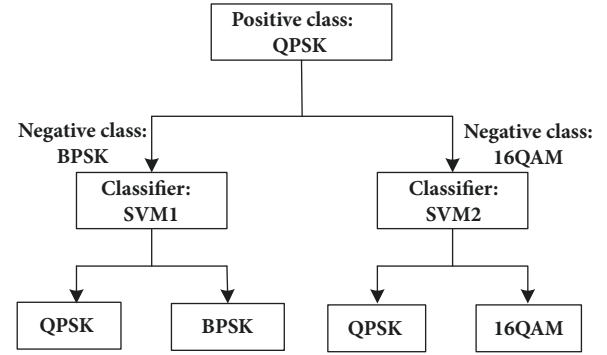


FIGURE 5: Three-class classification of SVM based on one-to-one method.

and the corresponding eigenvectors are also arranged from large to small.

- (5) The contribution rate is defined as

$$\beta = \frac{\sum_{i=1}^{k_2} \lambda_i}{\sum_{i=1}^{k_1} \lambda_i}, \quad (21)$$

where  $k_1$  is the original sample data dimension and  $k_2$  is the sample data dimension after dimensionality reduction. The new matrix  $B$  ( $B = [u_1, \dots, u_{k_2}]$ ) called projection matrix is composed of the feature vectors corresponding to the first  $k_2$  eigenvalues.

- (6) Determine the projection data of the original feature data in the projection matrix, and then its principal component can be given by

$$x = XB. \quad (22)$$

**4.2. One-to-One Multiclassification Method Based on SVM.** SVM is originally an effective binary-class classification method, and its basic model is defined as a linear classifier with largest interval in feature space. For multiclassification problems, SVM can also achieve classification in an one-to-many mode, one-to-one mode, etc. In this paper, the one-to-one mode of SVM is employed due to its simplicity. The flowchart for three-class classification using SVM based on one-to-one method is shown in Figure 5, and it will be used to classify BPSK, QPSK, and 16QAM.

The basic classification principle of SVM is summarized below. The discriminant function of implementing SVM is given by [38]

$$g(x) = y_i (w^T x + b), \quad (23)$$

where  $x$  is the training sample input after dimensionality reduction using PCA,  $w$  refers to a weight vector,  $y_i (\pm 1)$  denotes a category label, and  $b$  is an offset. Its interval is given by

$$\delta_i = y_i (w^T x + b) = |w^T x + b| = |g(x)|. \quad (24)$$



Its geometric interval is given by

$$d = \frac{1}{\|\mathbf{w}\|} |g(x)|. \quad (25)$$

The purpose of SVM is to find the optimal  $w_0$  and  $b_0$ , which is to maximize the geometric interval  $d$ , i.e., to minimize  $w$ . The problem can be transformed into

$$\begin{aligned} \min_w \quad & \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & y_i \left[ (\mathbf{w}^T \mathbf{x}_i + b) \right] \geq 1, \quad i = 1, 2, \dots, M, \end{aligned} \quad (26)$$

where  $\mathbf{x}_i$  is a vector in  $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M]^T$ . Applying Lagrange multiplication into (26), then we get

$$\begin{aligned} L(w, b, \alpha) = & \frac{1}{2} (\mathbf{w}^T \mathbf{w}) \\ & - \sum_{i=1}^M \alpha_i \{ y_i [(\mathbf{w}^T \mathbf{x}_i + b)] - 1 \}, \end{aligned} \quad (27)$$

where  $\alpha_i$  denotes a nonnegative Lagrange multiplier. Calculate partial derivative of  $w$  and  $b$  respectively and make them equal to zero; then we get

$$\mathbf{w} = \sum_{i=1}^M \alpha_i y_i \mathbf{x}_i \quad (28)$$

$$\sum_{i=1}^M \alpha_i y_i = 0.$$

Convert it to a dual problem and the target signal can be given by

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{j=1}^M \alpha_j \\ \text{s.t.} \quad & \sum_{j=1}^M \alpha_j y_j = 0 \\ & \alpha_i \geq 0. \end{aligned} \quad (29)$$

According to (29), the optimal Lagrange multiplier  $\alpha_{0,i}$  can be obtained. Then optimal weight  $\mathbf{w}_0$  can be given by

$$\mathbf{w}_0 = \sum_{i=1}^M \alpha_{0,i} y_i \mathbf{x}_i. \quad (30)$$

Furthermore, according to  $g(x) = \mathbf{w}^T \mathbf{x} + b = \mp 1$ , the optimal bias can be given by

$$b_0 = y_i - \sum_{i=1}^M \alpha_{0,i} y_i \mathbf{x}_i^T \mathbf{x}. \quad (31)$$

Finally, the objective function of the optimal classification can be given by

$$\begin{aligned} g(x) &= \text{sign}(\mathbf{w}^T \mathbf{x} + b_0) \\ &= \text{sign} \left( \sum_{i=1}^M \alpha_{0,i} y_i \mathbf{x}_i^T \mathbf{x} + b_0 \right), \end{aligned} \quad (32)$$

where  $\text{sign}(\cdot)$  is a symbolic function. It can be seen from the above analysis that the determination of the optimal weight vector is determined only by the optimal Lagrange multiplier, the training samples, and their categories. The position of the support vector and the offset are determined through training using the 2D feature data processed by PCA. Finally, the optimal classification boundary is found to achieve the correct classification for the test samples.

The objective of classifying BPSK, QPSK, and 16QAM can be accomplished using the above classification process as depicted in Figure 5. By specifying a signal as a positive class, the rest of the other two signals are treated as negative classes, and finally the one-to-one method is used to classify the multiple signals. Through the above feature analysis, QPSK can be designated as a positive class, and BPSK and 16QAM are sequentially regarded as a negative class. The basic SVM is used for twice to make the two optimal classification boundaries, which can accurately identify the three signals to achieve the classification.

## 5. Simulation Analysis

### 5.1. Performance Analysis without Fading Channel Effect

**5.1.1. Simulation Setup.** In order to verify the performance of the hybrid classifying network, we did the following simulations, including training phase and testing phase. As known to all, bandwidth, code rate and SNR have a much more significant influence on the signal features in comparison with sampling frequency and carrier frequency. Hence, the signal classes for training and testing are simulated by changing BW, CR, and SNR instead of FS and FC for simplicity. In the training phase, the SNRs of the seven types of modulated signals are set to [10dB, 20dB, 30dB], respectively, and the total number of samples is set to 5000. The timing offset is 0.1μs. The parameters for different kinds of signals are shown in Table 1, where TW, BW, CR, FC, and FS stand for time-width, bandwidth, code rate, carrier frequency, and sample frequency, respectively. There are 450 data segments serving as sample data for each type of signal modulation.

**5.1.2. Setting the Thresholds and Optimal Boundary Lines.** In the first-layer of the network, the standard deviation features of the difference of the frequency peaks based on STFT are extracted and shown in Figure 6. Since SF has only one frequency, the difference between adjacent frequencies is approximately zero. Hence, the standard deviation are also nearly zeros. For LFM with linear frequency variation, the difference between the adjacent frequencies is constant, leading to a zero value standard deviation. For 2FSK and 4FSK, the difference between the adjacent frequencies leads to large standard deviations. For the remaining BPSK, QPSK and 16QAM with phase jumps, fluctuations in the difference between adjacent frequencies are the main reasons for large standard deviations.

Through training, the standard deviation threshold  $\rho_1$  is set as 0.4 according to the naive Bayesian algorithm [34]. As

TABLE 1: Parameters for different types of modulated signals.

Signal type	TW/ $\mu$ s	BW/MHz	CR/MHz	FC/MHz	FS/MHz
PSKs	25	10/16/20/40/80	5/8/10/20/40	60	200
16QAM	25	10/16/20/40/80	5/8/10/20/40	60	200
SF	50	0.02	-	10	100
LFM	50	10/20/30/40/50	-	10	100
2FSK	50	12/14/18/22/30	1/2/4/6/10	10/20	100
4FSK	50	32/34/38/42/50	1/2/4/6/10	10/20/30/40	100

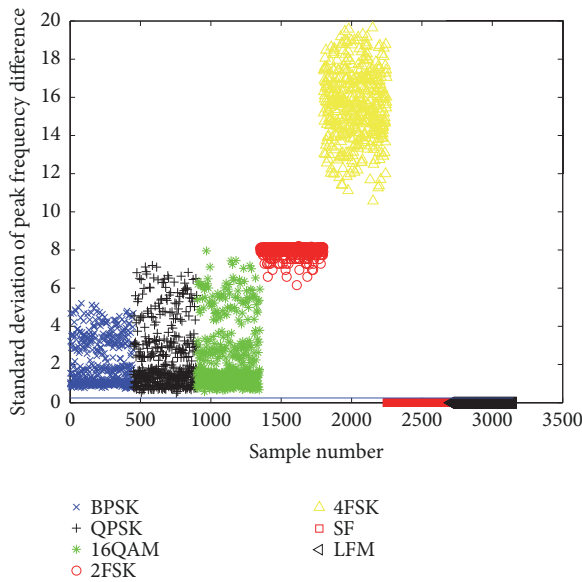


FIGURE 6: Standard deviation of the difference of the STFT peak.

shown in Figure 6, {2FSK, 4FSK, BPSK, QPSK, and 16QAM} is above the boundary line and {SF, LFM} is below the boundary line.

In the left-branch of the second-layer training, standard deviation characteristics based on the real part of IA are extracted to classify SF and LFM. The real part of IA of SF is a DC level, whereas LFM corresponds to an AC signal. Hence, the standard deviation between the two types of modulation is quite different, as shown in Figure 7. Through training, the threshold  $\rho_2$  of standard deviation of the real part of IA can be set to 0.52 according to the naive Bayesian algorithm. As shown in Figure 7, LFM is above the boundary line, whereas SF is below the boundary line.

In the right-branch of the second-layer, the remaining signal set {BPSK, QPSK, 16QAM, 2FSK, and 4FSK} is classified by the features based on DFT. BPSK, QPSK, and 16QAM have multiple peaks within the bandwidth, and the number of peaks increases from 20 to 340 when the BW increases as shown in Figure 8. However, the peak numbers of 2FSK and 4FSK are distributed around 2 and 4, respectively, which means that the thresholds  $\rho_3$  can be set to 2 and 4. In this case, 2FSK and 4FSK can be identified from the signal set {BPSK, QPSK, and QAM}.

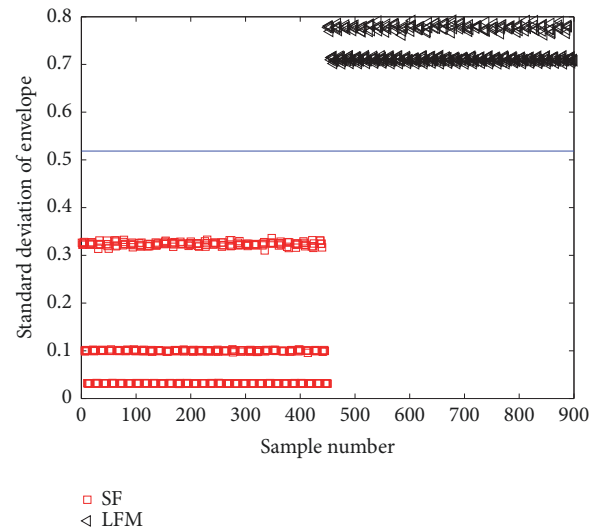


FIGURE 7: Standard deviation of real part of IA of SF and LFM.

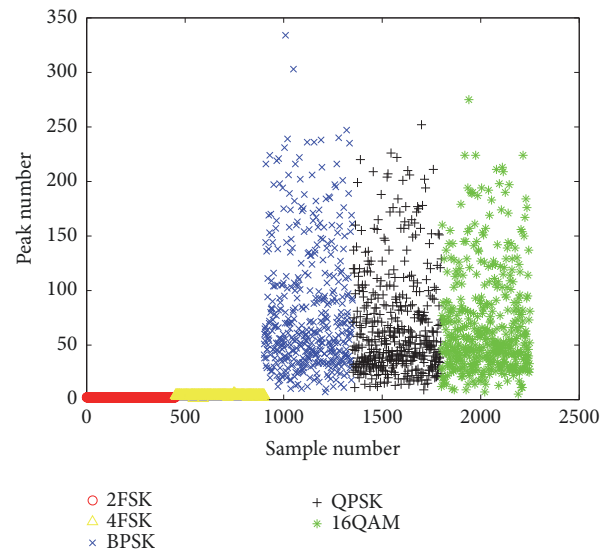


FIGURE 8: Peak numbers of the signal classes.

In the third-layer of the network, the signal set {BPSK, QPSK, and 16QAM} is trained by multiclassification method of SVM based on PCA feature dimension reduction.

The main features employed include standard deviation of the envelope, zero-crossing ratio of the IA, and standard

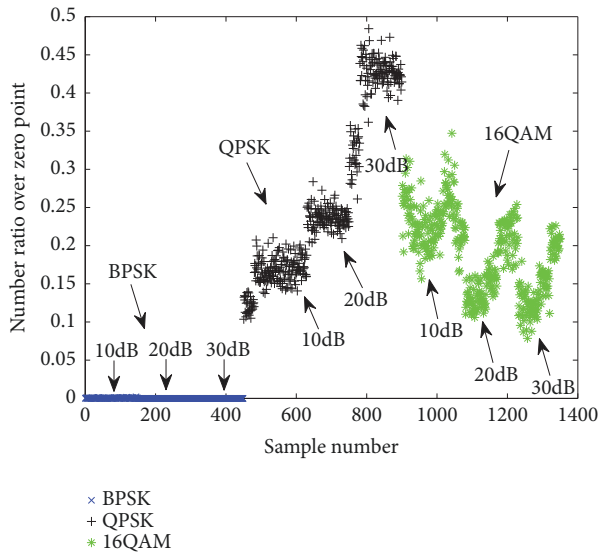


FIGURE 9: Zero-crossing ratio.

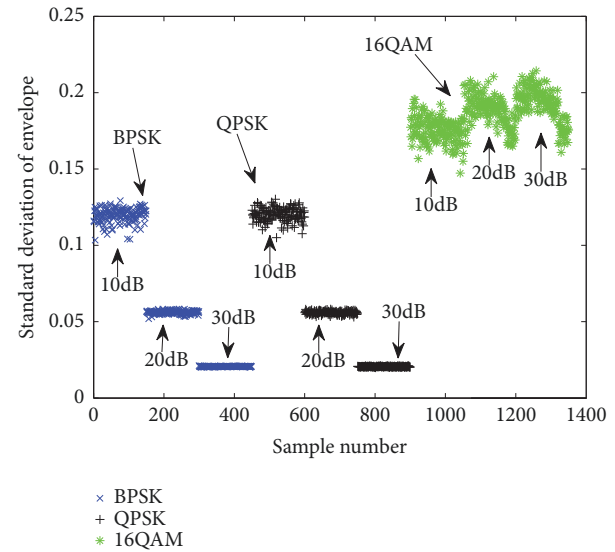


FIGURE 11: Standard deviation of envelope.

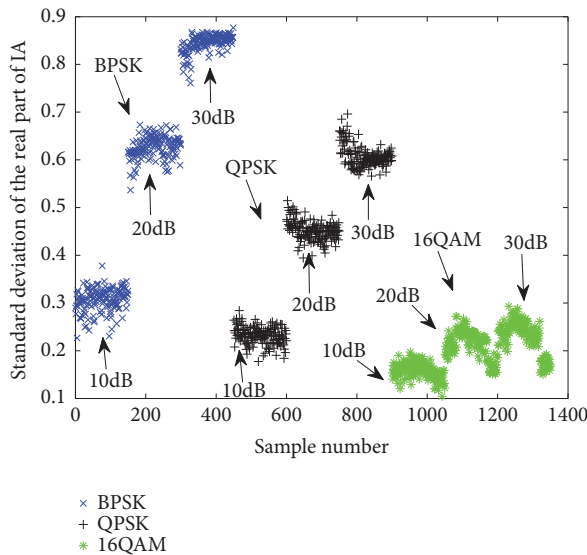


FIGURE 10: Standard deviation of the real part of IA.

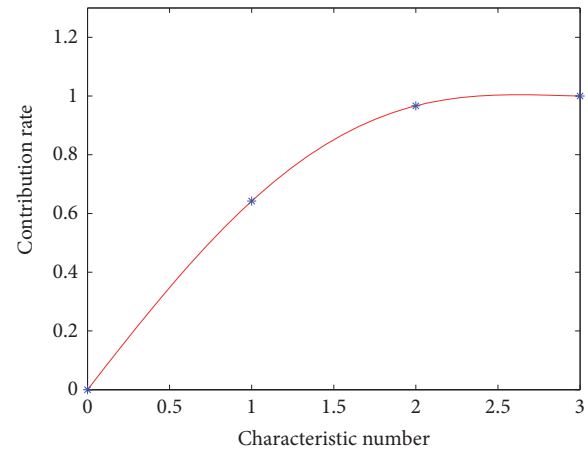


FIGURE 12: Characteristic of the contribution rate.

deviation of the IA. As can be seen from Figures 9–11, the distinguishing characteristics of the signals are much more obvious with the increase of the SNR. The 3D features are analyzed using PCA to make dimension degradation. From Figure 12 we can see that the contribution rate is still over 97% after the dimensions reduces to 2D. It indicates that the new 2D features can reflect more than 97% of the original 3D features. In other words, the new 2D features can replace the original 3D features with little loss.

The new 2D feature data is used as the training set, and the one-to-one method is substituted into the SVM for classification. The first step is to classify BPSK and QPSK. If QPSK is specified as a positive class, then BPSK is used as a negative class. The 2D new features of the two signals are substituted into the basic SVM for training. The positions of

the support vectors (the positions of the circles in Figure 13) are found, thereby determining the optimal boundary 1. According to the optimal boundary 1, the recognition of BPSK and QPSK is attained. The second step is to classify QPSK and 16QAM. If QPSK is specified as a positive class, then 16QAM is used as a negative class. The 2D new features of the two signal classes are substituted into the basic SVM for training. The optimal boundary 2 is determined after the positions of the support vectors are found. According to the optimal boundary 2, the recognition of QPSK and 16QAM are obtained. The classification results are shown in Figure 13, from which we can see that BPSK, QPSK, and 16QAM can be accurately identified by the two optimal boundary lines.

**5.1.3. Performance Analysis.** During the testing phase, the correct recognition rates of the signal set {BPSK, QPSK, 16QAM, LFM, SE, 2FSK, and 4FSK} at different SNRs are shown in Table 2. It can be seen from Table 2 that the correct recognition rate of the signals improves with the increase of

TABLE 2: Correct recognition rate at different SNRs.

Signal type	SNR=10dB	SNR=15dB	SNR=20dB	SNR=25dB	SNR=30dB
SF	95.20%	97.82%	98.40%	99.20%	99.99%
LFM	94.42%	95.80%	96.50%	97.60%	98.50%
2FSK	96.50%	98.20%	98.40%	99.50%	99.67%
4FSK	96.00%	97.50%	98.20%	98.67%	98.80%
BPSK	94.30%	95.20%	96.23%	99.20%	99.47%
QPSK	94.00%	94.78%	95.28%	97.26%	98.65%
16QAM	94.28%	95.12%	96.00%	98.30%	99.26%

TABLE 3: Correct recognition rate at different SNRs with fading channel.

Signal type	SNR=10dB	SNR=15dB	SNR=20dB	SNR=25dB	SNR=30dB
SF	94.10%	96.55%	97.80%	98.20%	98.80%
LFM	93.42%	94.80%	95.50%	96.80%	97.50%
2FSK	93.50%	93.80%	95.34%	96.40%	97.80%
4FSK	92.60%	93.20%	94.80%	95.50%	96.10%
BPSK	89.30%	90.60%	91.80%	93.80%	93.20%
QPSK	89.60%	91.34%	92.15%	92.16%	94.34%
16QAM	90.52%	92.20%	93.85%	94.10%	95.56%

the SNR. Under the scenario of SNR=10dB, the proposed network provides a correct recognition rate of over 94%. The results indicate that the classification performance of the proposed hybrid machine learning network is superior in discriminating between the modulated signal candidates in this paper.

### 5.2. Performance Analysis under Fading Channel Conditions.

Multipath effect of a channel usually leads to serious distortion on the received signal, causing serious degradation on the AMC algorithm. A fading channel is taken into account to analyze the performance of the proposed classification network in this simulation. The received signal model in the fading channel circumstance can be written as

$$z(n) = \sum_{k=0}^{L-1} \mathbf{h}(k) s(n-k) + r(n), \quad (33)$$

where  $s(n)$  is the transmitted signal,  $r(n)$  is the additive white Gaussian noise, and  $\mathbf{h}(k)$ ,  $k = 0, 1, \dots, L-1$  are the  $L$  fading channel coefficients. The channel  $\mathbf{h}(k)$  is considered nonrandom and assumed to be Rayleigh fading. The channel coefficients are randomly generated with variance 0.05 in the simulation except for  $\mathbf{h}(0) = 1$ . Other simulation conditions are the same as the above simulation.

The correct recognition rates of the signal set {BPSK, QPSK, 16QAM, LFM, SF, 2FSK, and 4FSK} at different SNRs under fading channel are shown in Table 3. Compared with Table 2, the correct recognition rate of each signal decreases. SF and LFM go down a bit just about 1%, while 2FSK and 4FSK fall approximately 2%. Especially, the descending value of BPSK, QPSK, and 16QAM can reach about 6%. The result of the comparison indicates that the performance of the classification network in fading channel has a slighter decrease than the scenarios without a fading channel.

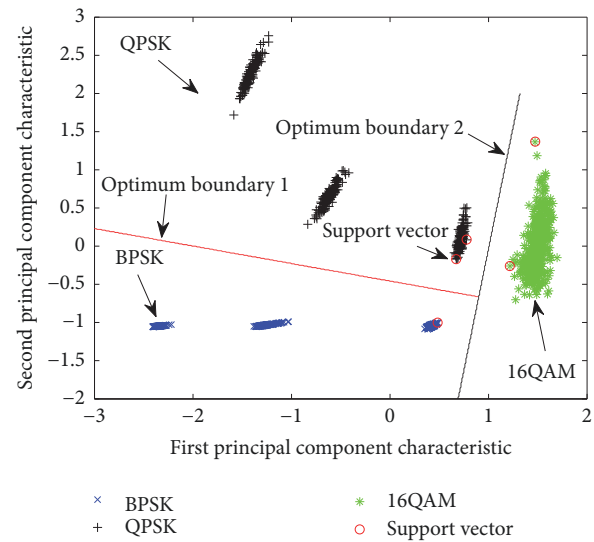


FIGURE 13: Three-class classification based on SVM.

**5.3. Performance Comparison with Algorithm in [9].** The classification of QAM signal in the third layer is an important part in the proposed network, whereas diverse methodologies have been explored in how to classify the QAM signal class. The AMC algorithm based on high-order cyclostationarity proposed in [9] is a classic algorithm for QAM signal classification and has good classification effect and superior performance. This paper applies the second-order instantaneous autocorrelation algorithm to realize AMC, and its performance is compared with the one in [9].

The adopted signals include BPSK, QPSK, and 16QAM. Figure 14 plots the total recognition performance of {BPSK, QPSK, and 16QAM} of the proposed algorithm and that of



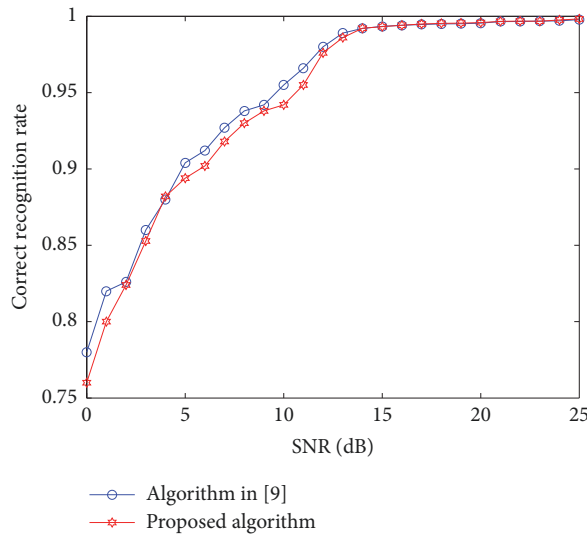


FIGURE 14: Comparison of correct recognition.

the algorithm in [9]. A comparison of these curves shows that the two algorithms have similar performance in classification. The advantage of the instantaneous autocorrelation is less complexity in comparison with that of the high-order cyclostationarity approach.

## 6. Conclusion

This paper proposes an AMC network for the classification of radar and communication signals. In general, a three-layer classification network is employed, consisting of a series of feature extraction and classification methods such as STFT, DFT, IA, PCA, SVM, and naive Bayesian algorithm. Through the training of the large sample data, the setting of the classification thresholds of the machine learning algorithms is automatically realized. During the sample construction process, the comprehensive coverage of signal samples is attained by changing the key parameters such as code rate and bandwidth. The simulation results show that the correct recognition rate of the seven types of modulated signals can reach over 94% at SNR of 10dB and above if channel distortion is not considered. For fading channel scenarios, a degradation of the correct recognition rate of about 6% is observed as a performance comparison study.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was partially supported by the Fundamental Research Funds for the Central Universities (Grant no. 2015B03014) and the Natural Science Foundation of Jiangsu Province (Grant no. BK20151501).

## References

- [1] S. Ayazgok, C. Erdem, M. T. Ozturk, A. Orduyilmaz, and M. Serin, "Automatic antenna scan type classification for next-generation electronic warfare receivers," *IET Radar, Sonar & Navigation*, vol. 12, no. 4, pp. 466–474, 2018.
- [2] C. L. Zhang and X. N. Yang, "Research on the Cognitive Electronic Warfare and Cognitive Electronic Warfare System," *Journal of China Academy of Electronics & Information Technology*, vol. 9, no. 6, pp. 551–555, 2014.
- [3] K. Dabcevic, M. O. Mughal, L. Marcenaro, and C. S. Regazzoni, "Cognitive Radio as the Facilitator for Advanced Communications Electronic Warfare Solutions," *Journal of Signal Processing Systems*, vol. 83, no. 1, pp. 29–44, 2016.
- [4] Z. L. Fan, G. S. Zhu, and H. U. Yuan-Kui, "An Overview of Cognitive Electronic Warfare," *Electronic Information Warfare Technology*, vol. 30, no. 1, pp. 33–38, 2015.
- [5] E. E. Azzouz and A. K. Nandi, *Automatic Modulation Recognition of Communication Signals*, Springer US, Boston, MA, 1996.
- [6] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *IET Communications*, vol. 1, no. 2, pp. 137–156, 2007.
- [7] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Blind modulation classification: a concept whose time has come," in *Proceedings of the IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communication*, pp. 223–228, April 2005.
- [8] D. Zeng, X. Zeng, G. Lu, and B. Tang, "Automatic modulation classification of radar signals using the generalised time-frequency representation of Zhao, Atlas and Marks," *IET Radar, Sonar & Navigation*, vol. 5, no. 4, pp. 507–516, 2011.
- [9] O. A. Dobre, M. Öner, S. Rajan, and R. Inkol, "Cyclostationarity-based robust algorithms for QAM signal identification," *IEEE Communications Letters*, vol. 16, no. 1, pp. 12–15, 2012.
- [10] H. Wang, O. A. Dobre, C. Li, and R. Inkol, "M-FSK signal recognition in fading channels for cognitive radio," in *Proceedings of the 2012 6th IEEE Radio and Wireless Week, RWW 2012 - 2012 IEEE Radio and Wireless Symposium, RWS 2012*, pp. 375–378, USA, January 2012.
- [11] H. Wang, O. A. Dobre, C. Li, and D. C. Popescu, "Blind Cyclostationarity-Based Symbol Period Estimation for FSK Signals," *IEEE Communications Letters*, vol. 19, no. 7, pp. 1149–1152, 2015.
- [12] H. Wu, M. Saquib, and Z. Yun, "Novel automatic modulation classification using cumulant features for communications via multipath channels," *IEEE Transactions on Wireless Communications*, vol. 7, no. 8, pp. 3098–3105, 2008.
- [13] G. Wannberg, A. Pellinen-Wannberg, and A. Westman, "An ambiguity-function-based method for analysis of Doppler decompressed radar signals applied to EISCAT measurements of oblique UHF-VHF meteor echoes," *Radio Science*, vol. 31, no. 3, pp. 497–518, 1996.
- [14] Y. Lin, X.-C. Xu, and Z.-C. Wang, "New individual identification method of radiation source signal based on entropy feature and



- SVM," *Journal of Harbin Institute of Technology (New Series)*, vol. 21, no. 1, pp. 98–101, 2014.
- [15] Z. Luo, L. Liu, J. Yin, Y. Li, and Z. Wu, "Deep learning of graphs with ngram convolutional neural networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 10, pp. 2125–2139, 2017.
  - [16] Z. Jiang, J. Wang, Q. Song, and Z. Zhou, "A Refined Cluster-Analysis-Based Multibaseline Phase-Unwrapping Algorithm," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1565–1569, 2017.
  - [17] S. Hao, W. Wang, Y. Ye, E. Li, and L. Bruzzone, "A Deep Network Architecture for Super-Resolution-Aided Hyperspectral Image Classification With Classwise Loss," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4650–4663, 2018.
  - [18] Y. Wei, W. Xia, M. Lin et al., "HCP: A flexible CNN framework for multi-label image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, pp. 1901–1907, 2016.
  - [19] J. Pei, Y. Huang, W. Huo, Y. Zhang, J. Yang, and T.-S. Yeo, "SAR automatic target recognition based on multiview deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2196–2210, 2018.
  - [20] Q. Guo, P. Nan, X. Zhang, Y. Zhao, and J. Wan, "Recognition of radar emitter signals based on SVD and AF main ridge slice," *Journal of Communications and Networks*, vol. 17, no. 5, pp. 491–498, 2015.
  - [21] D. Zeng, X. Zeng, H. Cheng, and B. Tang, "Automatic modulation classification of radar signals using the Rihaczek distribution and Hough transform," *IET Radar, Sonar & Navigation*, vol. 6, no. 5, pp. 322–331, 2012.
  - [22] B. Feng and Y. Lin, "Radar signal recognition based on manifold learning method," *International Journal of Control and Automation*, vol. 7, no. 12, pp. 399–406, 2014.
  - [23] S. Huang, Y. Yao, Z. Wei, Z. Feng, and P. Zhang, "Automatic Modulation Classification of Overlapped Sources Using Multiple Cumulants," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6089–6101, 2017.
  - [24] L. Wang and Y. Ren, "Recognition of digital modulation signals based on high order cumulants and support vector machines," in *Proceedings of the 2009 ISECS International Colloquium on Computing, Communication, Control, and Management (CCCM)*, pp. 271–274, Sanya, China, August 2009.
  - [25] H. Bai, Y.-J. Zhao, and D.-X. Hu, "Radar signal recognition based on the local binary pattern feature of time-frequency image," *Yuhang Xuebao/Journal of Astronautics*, vol. 34, no. 1, pp. 139–146, 2013.
  - [26] 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, no. 8, pp. 2742–2750, 2012.
  - [27] J. Chorowski and J. M. Zurada, "Learning understandable neural networks with nonnegative weight constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 1, pp. 62–69, 2015.
  - [28] J. L. Xu, W. Su, and M. Zhou, "Likelihood-ratio approaches to automatic modulation classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 41, no. 4, pp. 455–469, 2011.
  - [29] X. Yan, G. Liu, H. Wu, and G. Feng, "New Automatic Modulation Classifier Using Cyclic-Spectrum Graphs With Optimal Training Features," *IEEE Communications Letters*, vol. 22, no. 6, pp. 1204–1207, 2018.
  - [30] J. L. Xu, W. Su, and M. Zhou, "Distributed automatic modulation classification with multiple sensors," *IEEE Sensors Journal*, vol. 10, no. 11, pp. 1779–1785, 2010.
  - [31] H. Abuella and M. K. Ozdemir, "Automatic Modulation Classification Based on Kernel Density Estimation," *Canadian Journal of Electrical and Computer Engineering*, vol. 39, no. 3, pp. 203–209, 2016.
  - [32] F. Wang, O. A. Dobre, C. Chan, and J. Zhang, "Fold-based Kolmogorov-Smirnov Modulation Classifier," *IEEE Signal Processing Letters*, vol. 23, no. 7, pp. 1003–1007, 2016.
  - [33] V. D. Orlic and M. L. Dukic, "Automatic modulation classification algorithm using higher-order cumulants under real-world channel conditions," *IEEE Communications Letters*, vol. 13, no. 12, pp. 917–919, 2009.
  - [34] M. O. Mughal and S. Kim, "Signal Classification and Jamming Detection in Wide-Band Radios Using Naïve Bayes Classifier," *IEEE Communications Letters*, vol. 22, no. 7, pp. 1398–1401, 2018.
  - [35] D. X. Liu and G. Q. Zhao, "Analysis of Pulse Modulation Signals," *Modern Radar*, vol. 25, no. 11, pp. 17–20, 2003.
  - [36] M. S. Mühlhaus, M. Öner, O. A. Dobre, and F. K. Jondral, "A low complexity modulation classification algorithm for MIMO systems," *IEEE Communications Letters*, vol. 17, no. 10, pp. 1881–1884, 2013.
  - [37] R. P. Good, D. Kost, and G. A. Cherry, "Introducing a unified PCA algorithm for model size reduction," *IEEE Transactions on Semiconductor Manufacturing*, vol. 23, no. 2, pp. 201–209, 2010.
  - [38] Ş. Ertekin, L. Bottou, and C. L. Giles, "Nonconvex online support vector machines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 368–381, 2011.