SVM-based Drone Sound Recognition using the Combination of HLA and WPT Techniques in Practical Noisy Environment

Yujing He[1], Ishtiaq Ahmad[1], Lin Shi[1], and KyungHi Chang[1*]
[1] Electronic Engineering Department, Inha University, Incheon, Korea
[e-mail: hyj13640561704@163.com, ishtiaq001@gmail.com, slin0930@gmail.com, and khchang@inha.ac.kr]
*Corresponding author: KyungHi Chang

Abstract

In recent years, the development of drone technologies has promoted the widespread commercial application of drones. However, the ability of drone to carry explosives and other destructive materials may bring serious threats to public safety. In order to reduce these threats from illegal drones, acoustic feature extraction and classification technologies are introduced for drone sound identification. In this paper, we introduce the acoustic feature vector extraction method of harmonic line association (HLA), and subband power feature extraction based on wavelet packet transform (WPT). We propose a feature vector extraction method based on combined HLA and WPT to extract more sophisticated characteristics of sound. Moreover, to identify drone sounds, support vector machine (SVM) classification with the optimized parameter by genetic algorithm (GA) is employed based on the extracted feature vector. Four drones' sounds and other kinds of sounds existing in outdoor environment are used to evaluate the performance of the proposed method. The experimental results show that with the proposed method, identification probability can achieve up to 100 % in trials, and robustness against noise is also significantly improved.

1. Introduction

The development of drones, officially called unmanned air vehicles (UAV), has captured increasing attention from hobbyists and investors in recent years [1]. Such drones have endless commercial applications such as in agriculture, photography, and numerous public services because of their relatively small size and the ability to fly without an on-board pilot. At the same time, they can also carry out chemical, biological, or nuclear attacks, or be employed to help smuggle drugs or illegal immigrants across the border, which may lead to security threats to public safety because they are small in size and fly low enough to elude conventional radar detection.

To cope with threats from illegal drones, an efficient identification method for illegal drone monitoring is required. Those small drones produce typical characteristic sounds by their propellers, which can be the feature to distinguish the sounds of drone from other sounds, such as the sounds of birds and cars, etc. In recent years, many researchers have studied methods of acoustic feature extraction and classification. For feature extraction, Averbuch et al. [2] presents a method for acoustic detection of moving vehicles based on wavelets, and the feature extraction method of HLA for drone identification is introduced by Shi et al. [3]. Mel frequency cepstral coefficients (MFCCs) are extracted for speech recognition by Patel and Rao [4]. In order to identify acoustic features, several classification methods, such as the Bayesian classifier, the hidden Markov models (HMMs) [5], SVM [6], and K-nearest neighbor (KNN) are commonly used. In this paper, we use the HLA and WPT techniques for feature extraction. First, we briefly explain the reason that why we utilize the HLA and WPT techniques for feature extraction and after that Section 3 explains the procedures of these techniques in details.

1.1 Reasons for Selecting the HLA and WPT Techniques to Recognize the Drone Sound

The main reason of choosing the HLA technique in this paper to organize the local spectral peaks which exceed the noise level into families of harmonically related narrow band lines. The existence of a harmonic set shows the presence of a coupled harmonic signal source which is likely to be a flying drone. Only harmonically related peaks are selected in order to distinguish between the different drones sounds. In the spectral domain, the drones generate the acoustic signal waveforms that looks like narrow harmonic components. The feature vector is formed between the relationship of phase and amplitude of the harmonics. So, these harmonics contains some precise information which is related to the specific characteristic features for a drone. In case of WPT, the transform is computationally efficient, and the features extraction is done based on the signals emitted by the specific drone. The energy of the inherent blocks of the wavelet packet coefficients of the signals which is related to a specific frequency band is used to calculate the features set.

1.2 Main Contributions

- In the feature extraction stage, there are two schemes for feature vectors, one is HLA and the other is wavelet based, are applied. We introduce the acoustic feature vector extraction method of HLA, and subband power feature extraction based on WPT.
- Combining the feature vectors extracted by these two procedures (HLA and WPT) is proposed in this paper. Hence, more sophisticated characteristics of sound can be

extracted by jointly feature vector extraction method based on combined HLA and WPT techniques.

- Utilizing the feature vector extracted from sounds, SVM classification with the GA-based parameter optimization is employed for drone identification. Four drones' sounds and other kinds of sounds existing in outdoor environment are used to evaluate the performance of the proposed method.

- The experimental results validate that the proposed methodology show the significant improvement in identification probability and robustness against noise for drone sound recognition in practical noisy environment.

The rest of this paper is organized as follows. In Section 2, we briefly review the existing contributions related to the topic of our interest. Section 3 presents the acoustic feature vector extraction methods to get the characteristics of each sound. The details of SVM classification technology for drone sound identification and the GA-based parameter optimization technology for SVM are provided in Section 4. In Section 5, we perform the experimental results with six trials, and compare the identification probabilities of various techniques. Finally, Section 6 concludes the paper.

2. Related Work

There are plenty of works have been proposed in the literature for UAVs network. This section briefly discusses the ample range of contributions related to UAV network. Many extraordinary advancements have been made in drones technology to widely deploy UAVs network. The 5G infrastructures based on the use of millimeter-wave radio modules, may be efficiently leveraged to offer the much-needed drone detection capabilities. In [7], author presented an approach to recognize the drones via sound emitted by their propellers utilizing MFCC and HMM techniques for feature extraction and classification, respectively. Many other techniques are available in the literature using the acoustic data for feature extraction. In [8], author develop machine learning based drone detector model that employ the SVM technique to recognize the drone. The real-time UAV sound detection and analysis is presented in [9]. Author used the plotted image machine learning (PIL) and KNN methods for transforming the data to detect the drone. In [10], author uses the Gaussian mixture model, convolutional neural network (CNN) and recurrent neural network for drone sound detection. The drone detection system is modelled by multiple acoustic nodes using the machine learning approach in [11]. In this model, Author used the short-term Fourier transform and MFCC technique for training. SVM and CNN were trained with the data collection in person. In [12], author designed a real-time artificial intelligence system for drone detection. The fast Fourier transform (FFT) is performed for sampling the real time data and PIL is utilized with detected audio sample to the server for drone detection. In [13], author presented two-dimentional model to estimate the minimum distance needed to facilitate an avoidance maneuver for spatial configuration. Moreover, the harmonic nature of acoustic generated by propeller-driven aircraft is used to increase the detection distance. In [14], authors evaluated the possibility of using sound analysis as the detection mechanism. The linear predictive scheme is used for drone sound detection. In addition, authors assumed the slope of frequency spectrum to minimize the false alarm and they proved it is useful in terms of false alarm. In [15], author proposes to exploit 5G millimeter-wave deployments to detect violating amateur drones. In [16], author proposes optical flow as benchmark scheme for amateur done motion detection utilizing the monitoring drone camera. The modulation classification and

signal-strength based localization of amateur drones by utilizing surveillance drones is discussed in [17]. In [18], the author discusses the capabilities of the existing detection, tracking, localization, and routing schemes. The author describes that multiple UAVs can work as a cooperative team. This cooperation is not only useful in object detection but also sharing the information for jointly surveillance, sensor data collection, navigation, and collision avoidance, etc.

It should be noted that the safety regulations, bandwidth, and spectral allocation are also important issues for multi-UAVs system and applications. The increasing demand of new devices for wireless and cellular networks are making the problem of spectrum scarcity. To resolve the problem of spectrum scarcity, the author suggested the solution in [19] by harnessing the white spaces of licensed an unsilenced spectrum using the cognitive radio technology. Many UAV applications such as traffic surveillance, crop monitoring, border patrolling, disaster management, and wildfire monitoring can be supported by the cognitive radio technology.

In order to design the aerial networks, the pros and cons of the pre-existing technologies have been addressed in [20]. In [21] and [22], the authors emphasis on the specific application of UAV systems such as disaster response and traffic surveillance, respectively. In [23], author proposed a drone plane for monitoring and targeting the criminals using the real time image processing technique. A visual-based approach is adopted in [24] to detect the UAV by using the template matching and morphological filtering. To maximize the detection range and focus on long range scenarios, the morphological filtering has proven as important component adopted for image processing [25 - 26].

It is important factor that UAV system should be secured against malicious factors. It is necessary to guarantee that the package carrying by the UAV will not be intercepted on the way of client. Hence, safety measures should be taken against the malicious attack as well as traffic interference and communication should be allowed only secure channels.

Multiple UAVs are intended to be part of future air space traffic. Hence, reliable communication and networking is important in allowing successful coordination of aerial. However, still many issues need to be addressed for UAV commercial application. Hence, further contributions are required to fulfil the quality of services demands for multiple UAVs system. Unlike the resource allocation and interference mitigation schemes in [27 - 45], this paper addresses HLA and WPT based drone sound recognition in practical noisy.

3. Techniques for Acoustic Feature Vector Extraction

To distinguish the sound of drone from other sounds, the acoustic features of the sound should be extracted. Therefore, the primary task for drone identification is to extract robust features containing specific information of each sound. The feature extraction is to acquire the most related information from the original data and designate that information in a lower dimensionality space. When the input data to an algorithm is too large to be processed and it is supposed to be redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction [46], [47]. The purpose of feature extraction is not only to reduce the dimensionality but to additionally extract more useful/dominant information hidden in the signals by avoiding unnecessary or redundant information. In relation to signal processing, usually a signal is pre-processed to remove noise, interferences, and artifacts before performing feature extraction. After the features are extracted, classification is performed based on the selected features. The
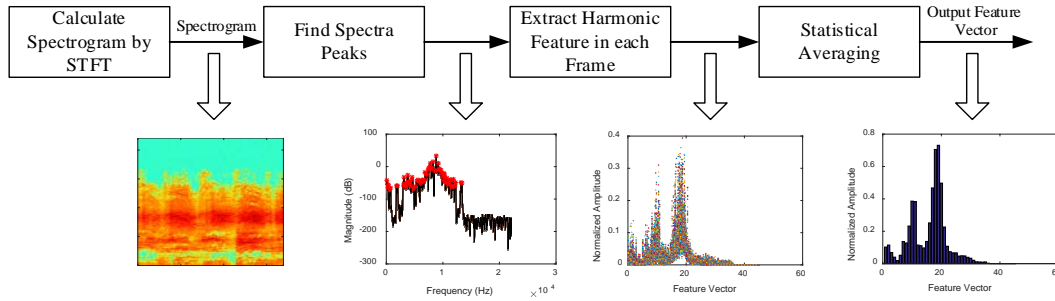
Fig. 1. HLA feature extraction process.

performance of the classifier depends on both how well the signals have been pre-processed and how aptly the features have been extracted. In this section, we explain the procedures of acoustic feature extraction using the HLA, WPT and the combination of HLA & WPT techniques.

3.1 HLA Feature Extraction Procedure based on Spectrogram

Our identification target is the small, low-flying drone, which tends to emit strong harmonic lines produced by its propeller [48]. Therefore, those harmonically related signal components in the spectral domain are used to construct an acoustic feature vector for sound identification via HLA [3].

In practice, most acoustic signals are time-variant and non-stationary. To get the spectrogram in the time-frequency domain, we use short-time Fourier transform (STFT). Given a sampled sound sequence, the spectrogram can be calculated through STFT by dividing a long-time sound sequence into multiple short data frames of equal length by window, and then computing the Fourier transform separately on each short data frame [49]. The STFT can be expressed as

$$\text{STFT}\{x[n]\}(m, w) \equiv \sum_{n=-\infty}^{\infty} x[n]w[n-m]\exp(-jwn) \tag{1}$$

where $x[n]$ and $w[n]$ are the sound sequence and window function, respectively.

To extract the harmonic feature from spectrograms, the local maxima bins in each frame can be located by traversal. If the local maxima are greater than some predefined threshold estimated from local background noise, the bins are labeled as spectra peaks. The harmonic relationships of those peaks are grouped together to form a feature vector set, which is the unique "pattern" of each sound source. To determine the harmonic relationships of the spectra peaks, each detected peak, in turn, is assumed to be the fundamental. After finding the harmonic of the current fundamental from the remaining peaks, the amplitudes of the strong harmonic set can be seen as the harmonic feature vector set. In order to minimize the influence of propagation distance, the feature vector is normalized by the magnitude of the highest harmonics. Finally, the feature vectors from each data frame are statistically averaged to form a feature vector of the sound. A block diagram of the above process is summarized in **Fig. 1**.
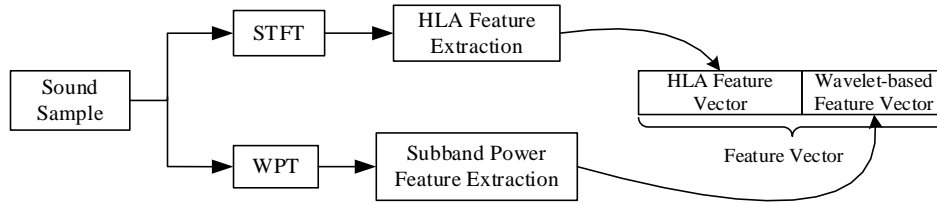
Fig. 2. Combining HLA and Wavelet-based feature vectors.

3.2 Wavelet-based Feature Extraction Procedure

In practice, the acoustic signal recorded from drone is quasi-periodic. In signals from the same source, there exist some dominating frequencies, which may vary in narrow frequency bands, but remain stable to some extent [50]. Hence, the distribution of the energy of acoustic signals over different areas in the frequency domain can provide a reliable characteristic signature for classification.

As the extension of wavelet transform, which provides multi-resolution time-frequency analysis for signal [51], WPT decomposes the detail coefficients (high-frequency components) and approximation coefficients (low-frequency components) of signals by wavelet basis. From N-level WPT, which decomposes signal energy on different time-frequency planes, we can get the wavelet packet coefficient. The integration of the square amplitude of those wavelet packet coefficients is proportional to the power of the frequency band. Therefore, the acoustic feature vector of sound related to subband power can be extracted via WPT:

$$E_j^n = \sum_k \left| x_j^n(k) \right|^2 \tag{2}$$

where $x_j^n(i)$ is the nth wavelet packet coefficient at the jth level, and k is the shift factor in the WPT. Similar to the HLA procedure, after the acoustic feature vector is extracted, it is normalized by the magnitude of the highest power.

3.3 Feature Vector Extraction based on the combined HLA and WPT Technique

To accurately distinguish sounds of drones from other sounds, the feature vector of the sound should contain enough information and characteristics of each sound. Moreover, Wuand Mendel [52] show that several different features extracted from sound can facilitate further classification. Therefore, we propose feature vector extraction based on combined HLA and WPT to acquire not only harmonic features but also the subband power feature based on WPT to increase the discriminability of sounds. The feature vector extracted by this method is

$$\mathbf{c} = [H_1, H_2, ... H_M, E_j^1, E_j^2, ..., E_j^N] \tag{3}$$

where $H_i, i = 1, 2, ..., M$ is the ith element in the harmonic feature vector with length M extracted by HLA, and E_j^n is the subband power calculated by Formula (2).

In this method, the harmonic feature vector of sound is extracted by the HLA procedure after STFT. And the subband power feature vector is also extracted by calculating subband power according to the wavelet packet coefficient of WPT. The final feature vector of sound is formed by combining the feature vectors of HLA and subband power based on WPT, as shown in **Fig. 2**. Complexities for the optimization of feature extraction and training in the next section may not be problematic, because the optimization procedure will be performed prior to real-time identification.

4. Identification of Drone Sound by SVM Classifier

To classify sounds through the feature vector extracted from sound, we need to construct a classifier, trained by the feature vectors and the class labels of known sounds. After training, the classifier formed with some classification criterion can be used to predict the class of sound without class label. As a widely used method for data classification, SVMs map feature vectors in low-dimensional space into a defined high-dimensional space through a kernel function, and find the best hyperplane by learning from experimental data to separate given points into two predefined classes [6].

4.1 Reasons for Choosing the SVM Classifier for Drone Sound Recognition

The main reason for selecting the SVM classifier is the discriminative approach that depends on two assumptions. First, the high-dimensional transforming data can be easily converted into the simple problem utilizing the linear discriminant function. Second, SVM use only such training patterns which are closely to the decision surface that deliver valuable information. Moreover, the real strength of SVM is the kernel trick which is very helpful to solve any complex problem. Hence, SVM woks well even there is no idea on the data and it scales relatively well to high dimensional data.

4.2 Support Vector Machine Classification

Given a training set $(\mathbf{x}_i, y_i)$, $i = 1, 2, ..., l$, where $\mathbf{x}_i \in R^n$ is the feature vector, and $y_i \in \{1, -1\}$ is the label associated with $\mathbf{x}_i$, for the two-class linearly separable case, the data points will be correctly classified by a hyperplane, $\mathbf{w}^T \mathbf{x} + b = 0$. All the data points satisfy the classification criterion:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \tag{4}$$

The SVM finds an optimal separating hyperplane with the maximum margin by solving the following optimization problem:

$$\underset{\mathbf{w}, \mathbf{b}}{\text{minimize}} \quad \Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum \xi_i \tag{5}$$
$$\text{subject to: } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \ \xi_i \geq 0$$

where the non-negative slack variables ξ_i are introduced to deal with the error caused by noise or misclassification of the training set, and C is the penalty parameter, which is applied to control overfitting.

An efficient method to solve this optimization problem is constructing a dual problem by introducing Lagrange multiplier α_i:

$$L(w,b,\alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{n}\zeta_i - \sum_{i=1}^{n}\alpha_i(y_i(\mathbf{w}^T\mathbf{x}_i+b)-1+\zeta_i) - \sum_{i=1}^{n}r_i\zeta_i \qquad (6)$$

By the saddle point of the Lagrange function, formula (5) can be transformed to:

$$\underset{\alpha}{\text{minimize}} \quad \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j)$$
$$\text{subject to: } \sum_{i=1}^{l}\alpha_i y_i = 0, \ 0 \le \alpha_i \le C \qquad (7)$$
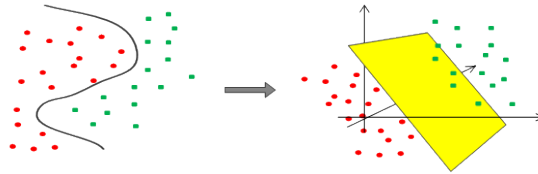

Fig. 3. Non-linear separable data become linear separable data in higher-dimensional space.

The solution α_i of formula (7) determines the parameters $\mathbf{w}$ and $\mathbf{b}$ of the hyperplane. Thus, we get the mathematical expression of the classifier as

$$f(\mathbf{x}) = \sum_{i=1}^{l}\alpha_i y_i K(\mathbf{x}_i,\mathbf{x}) + b \qquad (8)$$

where $K(\mathbf{x}_i,\mathbf{x}_j) \equiv \Phi(\mathbf{x}_i)^T\Phi(\mathbf{x}_j)$ is a kernel function for mapping the training feature vectors into some higher-dimensional feature space. In this way, the complex non-linear separable problem in the feature space becomes a simple linear separable problem in the higher-dimensional feature space, as shown in **Fig. 3**. The widely used kernel function in many classification problems is a radial-basis function (RBF), which always provides good classification performance. Therefore, we choose the RBF kernel function, which is shown in formula (9) for SVM:

$$K(\mathbf{x}_i,\mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2) \qquad (9)$$

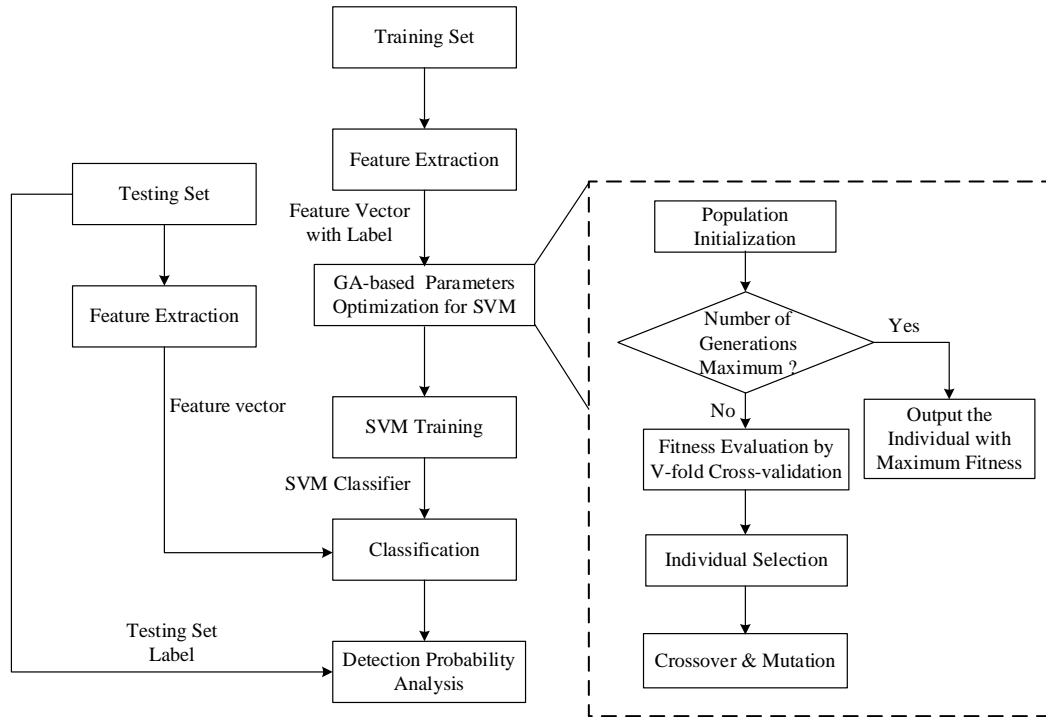## 4.3 GA-based Parameter Optimization for SVM

Fig. 4. Overall flow chart for feature extraction and classification.

In SVM, there are two parameters that will influence classification performance, and that should be determined before training the classifier; one is the penalty coefficient, C, in Formula (7), and the other is the RBF kernel function coefficient, γ, in Formula (9). C controls the tradeoff between maximizing the geometric margin of two classes and minimizing the data point deviation, whereas γ implicitly determines the distribution of the data mapped to the new feature space.

To optimize these two parameters, we introduce a method of GA-based parameter optimization [53]. GA is an evolutionary algorithm based on the mechanics of natural selection and genetics for optimization. According to GA, each pair of values for these two parameters is considered individual. Initial individuals are generated randomly, and then for each generation, the existing individuals are selected by fitness function. The fitter individuals are selected and modified (through crossover and mutation) to form a new generation. At the end of the evolution, the individual with the highest fitness is chosen as the SVM parameter. To judge the fitness of each individual (the value of two parameters), we use v-fold cross-validation as the fitness function. In this method, the training set is randomly divided into v subsets of equal size. Among them, one subset is tested using the classifier trained on the remaining $v-1$ subsets, in turn, until all subsets are tested once. The fitness of each individual is the percentage of correctly classified data in cross-validation out of the total amount of data in the training set. The overall flow chart for feature extraction and classification is given in **Fig. 4**.

5. Experimental Results and Performance Evaluation

5.1 Description on Experiment Setup

In the following experiments, recorded sounds from four drones and several other sounds downloaded from a public audio database are utilized to evaluate the performance of drone identification. The four drones we use are a SYMA X5SW (Drone 1), a BYROBOT Drone Fighter (Drone 2), a WLtoys Skywalker (Drone 3), and a DJI Phantom 3 Professional (Drone4). Other sound sources, which could be interference with drone identification in the outdoor environment, are cars, birds, rain, and planes. We recorded sounds three times for

Table 2. Training set in each trial.

	Training set		Training set
Trial 1	Drone 1_1, Drone 1_2, Drone 1_3 Drone 2_1, Drone 2_2, Drone 2_3 Drone 3_1, Drone 3_2, Drone 3_3 Car 1, Car 2, Car 3, Car 4, Car 5 Bird 1, Bird 2, Bird 3 Rain 1, Rain 2, Rain 3 Plane 1, Plane 2, Plane 3	Trial 2	Drone 1_1 Drone 2_1 Drone 3_1 Car 1, Car 2, Car 3, Car 4, Car 5 Bird 1, Bird 2, Bird 3 Rain 1, Rain 2, Rain 3 Plane 1, Plane 2, Plane 3
Trial 3	Drone 1_1, Drone 1_2, Drone 1_3 Drone 2_1, Drone 2_2, Drone 2_3 Car 1, Car 2, Car 3, Car 4, Car 5 Bird 1, Bird 2, Bird 3 Rain 1, Rain 2, Rain 3 Plane 1, Plane 2, Plane 3	Trial 4	Drone 1_1, Drone 1_2, Drone 1_3 Drone 2_1, Drone 2_2, Drone 2_3 Car 1, Car 2, Car 3 Bird 1, Bird 2 Rain 1, Rain 2 Plane 1, Plane 2
Trial 5	Drone 1_1, Drone 1_2, Drone 1_3 Car 1, Car 2, Car 3, Car 4, Car 5 Bird 1, Bird 2, Bird 3 Rain 1, Rain 2, Rain 3 Plane 1, Plane 2, Plane 3	Trial 6	Drone 1_1, Drone 1_2, Drone 1_3 Car 1 Bird 1 Rain 1 Plane 1

Table 1. Five methods of feature extraction and classification.

	Feature vector extraction method	Classification method
HLA + SVM	HLA	SVM with default parameter
HLA + SVM with GA	HLA	SVM with optimized parameter by GA
Wavelet + SVM	Wavelet-based feature extraction	SVM with default parameter
Wavelet + SVM with GA	Wavelet-based feature extraction	SVM with optimized parameter by GA
(HLA + Wavelet) + SVM with GA	Combine HLA and Wavelet-based feature vector	SVM with optimized parameter by GA

each drone, labeling them Drone i_1, Drone i_2, and Drone i_3, where i is the index of the drone. We choose six sounds for cars and four sounds for birds, rain, and planes, named Car 1, Car 2, etc. Among them, the four sounds for planes are Helicopter (Plane 1), Boeing 737 (Plane 2), Boeing 747 (Plane 3), and Airbus A320 (Plane 4).

The length of each sound we use is 6 s, with 44.1 KHz sampling frequency. Each sound is divided into 50 samples with the same length (0.12 s). To extract the HLA feature vector, each sample is divided into short data frames with a length of 0.02 s and the overlap factor between adjacent frames is 0.5. Then, the 1024-point fast Fourier transform (FFT) is computed at each data frame. By using the HLA procedure, a 45-component harmonic feature vector is extracted.

At the same time, a 32-component subband power feature vector is extracted, based on the wavelet packet coefficient of a five-level WPT with a Daubechies wavelet.
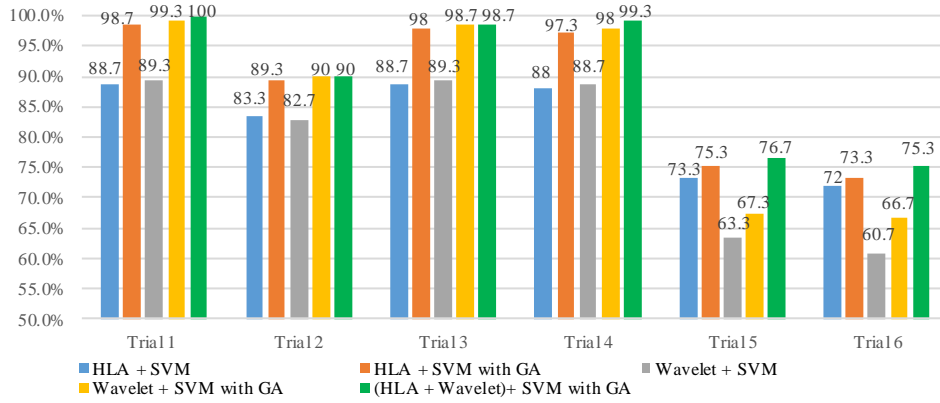

Fig. 5. Identification probability by various feature extraction and classification methods.

With the feature vector extracted from the sounds, six trials are performed with different training sets in order to train the SVM classifier. The training set for each trial is shown in **Table 1**. To identify the sound of the drone, we use two-class SVM classification, where class 1 contains drone sounds labeled 1, and class 2 contains other sounds (including all the sounds of cars, birds, rain, and planes, labeled -1). To compare the classification performance, the testing sets for all the trials are same, including Drone 4_1, Drone 4_2, Drone 4_3, Car 6, Bird 4, Rain 4, and Plane 4.

Our experiments are carried out in the Matlab R2015a development environment by extending the Libsvm toolbox, which is originally designed by Chang and Lin (2001). In each trial, five different methods are used, as shown in **Table 2**. SVM classification parameters C and γ in the first and third methods use the default values in Libsvm, i.e., $C=1$ and $\gamma=1/k$, where k is the dimension of the feature vector. In other methods using SVM with GA, the population of the GA in each generation is 30, and the total evolutional generation is 100.

5.2 Recognition Results and Performance Evaluation

The experimental results from six trials with five methods are shown in **Fig. 5**. The vertical coordinate is the identification probability, which is the proportion of the number of correctly predicted data by classifier out of all the testing data. Here, we only show the identification probability for drone sounds for two reasons. First, in our experiment, the sounds of the drone belong to class 1, which has the specific feature of drones, while all the other sounds belong to class 2, which occupies the majority of the feature space, because it contains many features of different sound sources. Under this background, the proportion of other sounds misclassified as class 1 (false alarm) is much lower than the probability of drone sounds misclassified as class 2 (missed identification). Another, and also the dominating reason is that, in the drone identification system, false alarm is negligible, because the further drone identification procedure can easily eliminate those errors, but the missed identification can leave a threat from an illegal drone.

As shown in **Fig. 5**, the identification probability in Trial 1 is the highest, even up to 100 %, because more training data are input in Trial 1 to train the classifier, which is a more accurate hyperplane from which to distinguish the sounds of drone from other sounds.
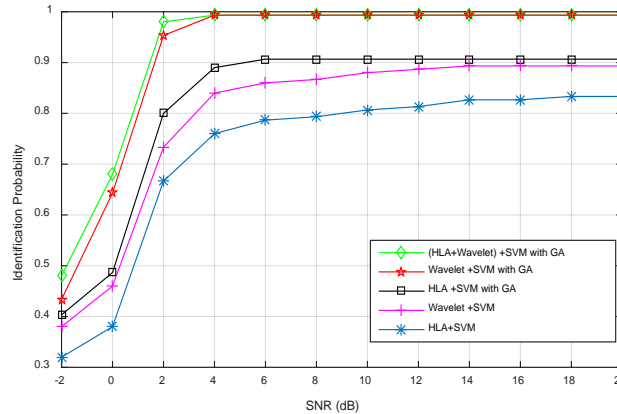

Fig. 6. SNR versus identification probability.

Comparing Trial 3 with Trial 1, there is one less type of drone sounds, and the performance decreases by only about 1 %, whereas comparing Trial 5 with Trial 3, there is also one less type of drone sounds, but the performance decrease by around 20 %. We see that when we use more kinds of sounds for drones, the identification performance will be better, especially when the kinds of sounds in the training set are fewer. Comparing Trial 2 with Trial 1, the kinds of sounds are the same, and there are more sounds for each drone in Trial 1, which gets the approximate 10 % improvement in identification probability. In addition, comparing Trial 3 and Trial 4, or Trial 5 and Trial 6, we find that with the same drone sounds for training, if we add other sounds, the identification performance can also be improved.

For all the trials, the (HLA + Wavelet) + SVM with GA method get the best identification performance, which even reached 100 % in Trial 1. With same feature vector method, the identification performance of the classification method using SVM with GA is much better than using SVM without GA. And with the same classification method, the method of combining HLA and wavelet-based feature extraction is better than separate HLA or wavelet-based feature extraction.

In addition, we also test the robustness of each method against additive white Gaussian noise (AWGN) using the test sounds with different signal-to-noise ratio (SNR), which add AWGN to the original sound of the drone recorded. The identification probability of each method in Trial 1 by varying SNR is shown in **Fig. 6**. The identification probabilities for all methods are enhanced sharply when SNR is less than 4 dB, and growth slowly, or level off, when SNR is greater than 4 dB. Similar to the results without AWGN, the identification probability of (HLA + Wavelet) + SVM is the best under various noise environments (greater than 95 % when SNR equals 2 dB).

6. Conclusions

To reduce threats from illegal drones, capable acoustic feature extraction and classification techniques are used for drone sound identification. The acoustic feature extraction method of HLA and wavelet-based subband power based on WPT is introduced in this paper. In addition,

to improve identification performance, a feature extraction method based on the combined HLA and WPT is proposed. Furthermore, utilizing the feature vector extracted from sounds, SVM classification technology and GA-based parameter optimization for SVM are employed for drone sound identification. The experimental results show that as we add more data to train the classifier, identification performance can be improved to 100 % from 75.3 %. Compared with other methods, the method we propose, which uses combined HLA and WPT for feature extraction and SVM with GA for classification, achieves the highest identification probability, even up to 100 %. In addition, we also test the robustness of each method by varying SNR, and the experimental results show that under AWGN environment, the method we propose achieves adequate identification performance of more than 95 % at 2 dB SNR.

Acknowledgements

Conflicts of Interest

The authors declare that they have no conflict of interest.

Abbreviations

HLA	Harmonic Line Association
WPT	Wavelet Packet Transform
SVM	Support Vector Machine
GA	Genetic Algorithm
UAV	Unmanned Air Vehicles
MFCCs	Mel Frequency Cepstral Coefficients
HMMs	Hidden Markov Models
KNN	K-nearest Neighbor
PIL	Plotted Image Machine Learning
CNN	Convolutional Neural Network
STFT	Short-time Fourier Transform
RBF	Radial-basis Function
FFT	Fast Fourier Transform
AWGN	Additive White Gaussian Noise
SNR	Signal-to-noise Ratio

References

[1] I. Bekmezci, O. K. Sahingoz, and S. Temel, "Flying ad-hoc networks (FANET): a survey," *Ad Hoc Networks*, vol. 11, no. 3, pp. 1254–1270, Jan. 2013. Article (CrossRef Link)

[2] A. Averbuch, V.A. Zheludev, and N. Rabin, "Wavelet-based acoustic detection of moving vehicles," *Multidimensional Systems and Signal Processing*, vol. 20, no. 1, pp. 55-80, Mar. 2009. Article (CrossRef Link)

[3] W. Shi, B. Bishop, and G. Arabadjis, *Sensor Fusion - Foundation and Applications*, InTech, Jun. 2011.

[4] I. Patel and Y. S. Rao, "Speech recognition using hidden Markov model with MFCC-subband technique," in *Proc. of International Conference on Recent Trends in Information, Telecommunication and Computing (ITC)*, pp.168-172, March 2010.

[5] X. Zhang and Y. Wang, "A hybrid speech recognition training method for HMM based on genetic algorithm and Baum Welch algorithm," in *Proc. of IEEE Second International Conference on Innovative Computing, Information and Control*, pp. 572-576, Sept. 2007.

[6] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778-1790, Aug. 2004. Article (CrossRef Link)

[7] L. Shi, I. Ahmad, Y. He, and K. H. Chang, "Hidden Markov model based drone sound recognition using MFCC technique in practical noisy environments," *Journal of Communication and Networks*, Vol. 20, No. 5, pp. 509-518, Oct. 2018. Article (CrossRef Link)

[8] A. Bernardini, F. Mangiatordi, E. Pallotti, and F. U. Bordoni, "Drone detection by acoustic signature identification," in *Proc. of IS&T International Symposium on Electronic Imaging*, pp. 60-64, Jan, 2017, Article (CrossRef Link)

[9] J. Kim, C. Park, J. Ahn, Y. Ko, J. Park, and J. C. Gallagher, "Real-time UAV sound detection and analysis System," in *Proc. of IEEE Sensors Applications Symp.*, pp. 1–5, Mar. 2017. Article (CrossRef Link)

[10] S. Jeon, J. W. Shin, Y. J. Lee, Y. H. Kwon, and H. Y. Yang, "Empirical study of drone sound detection in real-life environment with deep neural networks," in *Proc. of European Signal Processing Conference*, pp. 1858–1862, 2017.

[11] B. Yang, E. T. Matson, and J. C. Gallagher, "UAV detection system with multiple acoustic nodes using machine learning models," in *Proc. of IEEE International Conference on Robotic Computing*, 2 pp. 493-498, 2019.

[12] J. Kim and D. Kim, "Neural network based real-time UAV detection and analysis by sound," *Journal of Advanced Information Technology and Convergence*, Vol. 8, No.1, pp. 43-52, Jul. 2018. Article (CrossRef Link)

[13] B. Hrvey and S. O. Young, "Acoustic detection of a fixed-wing UAV," *Drones*, Vol. 2 No. 1, pp. 1-18, 2018. Article (CrossRef Link)

[14] L. Hauzenberger and E. Holmberg Ohlsson, *Drone Detection Using Audio Analysis*, Master's thesis, 2015.

[15] D. Solomitckii, M. Gapeyenko, V. Semkin, S. Andreev, and Y. K- oucheryavy, "Technologies for efficient amateur drone detection in 5G millimeter-wave cellular infrastructure," *IEEE Wireless Communications Magazine*, vol. 56, no. 1, pp. 43-50, Jan. 2018. Article (CrossRef Link)

[16] G. R. Rodrguez-Canosa et al., "A real-time method to detect and track moving objects (datmo) from unmanned aerial vehicles (UAVs) using a single camera," *Remote Sensing*, vol. 4, no. 4, pp. 1090–1111, 2012. Article (CrossRef Link)

[17] M. M. Azari, H. Sallouha, A. Chiumento, S. Rajendran, E. Vinogradov, and S. Pollin, "Key technologies and system trade-offs for detection and localization of amateur drones," *IEEE Wireless Communications Magazine*, vol. 56, no. 1, pp. 51-57, Jan. 2018. Article (CrossRef Link)

[18] Z. Kaleem and M. Rehmani, "Amateur drone monitoring: state-of-the-art architectures, key enabling technologies, and future research directions," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 150-159, May. 2018. Article (CrossRef Link)

[19] Y. Saleem, M. H. Rehmani, and S. Zeadally, "Integration of cognitive radio technology with unmanned aerial vehicles: Issues, opportunities, and future research challenges," *J. Netw. Comput. Appl.*, vol. 50, pp. 15–31, Apr. 2015. Article (CrossRef Link)

[20] L. Gupta, R. Jain, and G. Vaszkun, "Survey of important issues in UAV communication networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1123–1152, Nov. 2015. Article (CrossRef Link)

[21] A. Puri, "A survey of unmanned aerial vehicles (UAV) for traffic surveillance," *Dept. Comput. Sci. Eng., Univ. of South Florida*, Tampa, FL, USA, pp. 1-29, Jan. 2005.

[22] S. Ghafoor, P. D. Sutton, C. J. Sreenan, and K. N. Brown, "Cognitive radio for disaster response networks: Survey, potential, and challenges," *IEEE Wireless Commun.*, vol. 21, no. 5, pp. 70–80, Oct. 2014. Article (CrossRef Link)

[23] S. Karim, A. A. Laghari, and M. R. Asif, "Image processing based proposed drone for detecting and controlling street crimes," in *Proc. of IEEE International Conference on Communication Technology*, pp. 1725- 1730, 2017.

[24] R. Opromolla, G. Fasano, and D. Accardo, "A vision-based approach to UAV detection and tracking in cooperative applications," *Sensors*, vol. 18, pp. 1-26, 2018. Article (CrossRef Link)

[25] G. Fasano, D. Accardo, A. E. Tirri, E. and D. Lelli, "Morphological filtering and target tracking for vision-based UAS sense and avoid," in *Proc. of International Conference on Unmanned Aircraft Systems*, pp. 430–440, May 2014.

[26] J. Lai, J. J. Ford, L. Mejias, and P. O. Shea, "Characterization of Sky-region Morphological-temporal Airborne Collision Detection," *Journal of Field Robot*, vol. 30, no. 2, pp. 171-193, 2013. Article (CrossRef Link)

[27] I. Ahmad, W. Chen, and K. H. Chang, "Co-channel interference analysis using cooperative communication schemes for the coexistence of PS-LTE and LTE-R networks," in *Proc. of IEEE Communication and Electronics Special Session on LTE Technologies and Services*, pp. 181-182, Jul. 2016.

[28] I. Ahmad, Z. Kaleem, and K. H. Chang, "Block error rate and UE through- put performance evaluation using LLS and SLS in 3GPP LTE downlink," in *Proc. of Korean Institute of Communication and Information Sciences*, pp. 512-516, Feb. 2013.

[29] I. Ahmad, W. Chen, and K. H. Chang, "LTE-railway user priority-based cooperative resource allocations schemes for coexisting public safety and railway networks," *IEEE Access*, vol. 5, pp, 7985-8000, May 2017. Article (CrossRef Link)

[30] Z. Kaleem, M. Z. Khaliq, A. Khan, and T. Q. Duong, "PS- CARA: context-aware resource allocation scheme for mobile public safety networks," *Journal of Sensors*, vol. 18, no. 5, pp. 1-17, May 2018. Article (CrossRef Link)

[31] Z. Kaleem, Y. Li, and K. H. Chang, "Public safety users priority-based energy and time-efficient device discovery scheme with contention resolution for ProSe in 3GPP LTE-A systems," *IET Communications*, vol. 10, no. 15, pp. 1873-1883, Jul. 2016. Article (CrossRef Link)

[32] I. Ahmad, Z. Kaleem, and K. H. Chang, "Uplink power control for interference mitigation based on users priority in two-tier femtocell network," in *Proc. of IEEE International Conference on ICT Convergence*, pp. 474-475, Oct. 2013.

[33] I. Ahmad, K. H. Chang, "Analysis on MIMO transmit diversity and multiplexing techniques for ship ad-hoc networks under a maritime channel model in coastline areas," in *Proc. of IEEE International Conference on ICT Convergence*, pp. 18-20, Oct. 2017.

[34] I. Ahmad, K. H. Chang, "Analysis on MIMO transmit diversity techniques for ship ad-hoc network under a maritime channel model in coastline areas," *Journal of Korean Institute of Communications and Information Sciences*, vol. 42, no. 2, pp. 383-385, Feb. 2017. Article (CrossRef Link)

[35] I. Ahmad, Z. Kaleem, K. H. Chang, "QoS priority based femtocell user power control for interference mitigation in 3GPP LTE-A HetNet," *Journal of Korean Institute of Communications and Information Sciences*, vol. 39, no. 2, pp. 61-74, Feb. 2014. Article (CrossRef Link)

[36] W. Chen, I. Ahmad, and K. H. Chang, "Co-channel interference management using eICIC/FeICIC with coordinated scheduling for the coexistence of PS-LTE and LTE-R networks," *EURASIP Journal on Wireless Communications*, vol. 2017, no. 34, pp. 1-14, Dec. 2017. Article (CrossRef Link)

[37] I. Ahmad, Z. Kaleem, R. Narmeen, L. D. Nguyen, and D. B. Ha, "Quality-of-service aware game theory-based uplink power control for 5G heterogeneous networks," *Mobile Networks and Applications*, Vol. 24, No. 2, pp 556–563, Apr. 2019. Article (CrossRef Link)

[38] I. Ahmad, K. H. Chang, "Effective SNR mapping and link adaptation strategy for next-generation underwater acoustic communications networks: a cross-layer approach," *IEEE Access*, Vol. 7, pp. 44150-44164, Apr. 2019. Article (CrossRef Link)

[39] Z. Kaleem, M. Yousaf, A. Qamar, A. Ahmad, Trung Q. Duong, W. Choi, A. Jamalipour, "UAV-Empowered Disaster-Resilient Edge Architecture for Delay-Sensitive Communication," *IEEE Network*, Vol. 99, pp. 1-9, Jan. 2019. Article (CrossRef Link)

[40] I. Ahmad and K. H. Chang, "Design of system-level simulator architecture for underwater acoustic communications and networking," in *Proc. of ICTC*, pp. 384-386, Oct. 2016.

[41] Z. Kaleem, I. Ahmad, and C. Lee, "Smart and energy efficient LED street light control system using zigbee network," in *Proc. FIT*, Islamabad, Pakistan, pp. 361-365, Dec. 2014.

[42] W. Chen, I. Ahmad and K. H. Chang, "Analysis on the co-channel interference for the coexistence of PS-LTE and LTE-R networks," in *Proc. of Conference of Korean Institute of Communications and Information Sciences (KICS)*, Jeju, Korea, pp. 202-203, June 2016.

[43] Alamgir, I. Ahmad and K. H. Chang, "On the underwater channel model and network layout," in *Proc. of Conference of Korean Institute of Communications and Information Sciences (KICS)*, Korea, pp, 202-203, Jan. 2018.

[44] J. Xiao, I. Ahmad and K. H. Chang, "eMBMS and V2V communications for vehicle platooning in eV2X system," in *Proc. Conference of Korean Institute of Communications and Information Sciences (KICS)*, Jun. 2018.

[45] U. A. Mughal, I. Ahmad and K. H. Chang, "Virtual cells operation for 5G V2X communications," in *Proc. Conference of Korean Institute of Communications and Information Sciences (KICS)*, Korea, pp, 1-2, Jan. 2019.

[46] G. Kumar and P. K. Bhatia, "A detailed review of feature extraction in image processing system," in *Proc. of International Conference on Advanced Computing and Communication Technologies*, pp. 5-12, Apr. 2014.

[47] N. Anandhi and R. Avudaiammal, "Vehicle detection and classification from acoustic signal using ANN and KNN," in *Proc. of International Conference on Communication and Signal Processing (ICCSP)*, pp. 0066-0069, Apr. 2017.

[48] J. A. Robertson, J. C. Mossing, and B. Weber, "Artificial neural network for acoustic target recognition," in *Proc. of SPIE*, pp. 939-950, Apr. 1995.

[49] X. Ouyang and M.G. Amin, "Short-time Fourier transform receiver for nonstationary interference excision in direct sequence spread spectrum communications," *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 851-863, Apr 2001. Article (CrossRef Link)

[50] A. Averbuch, E. Hulata and V. Zheludev, "A wavelet packet algorithm for classification and detection of moving vehicles," *Multidimensional Systems and Signal Processing*, vol. 12, no. 1, pp. 9-31, Jan. 2001. Article (CrossRef Link)

[51] C. C. Lin, S. H. Chen, and T. K. Truong, "Audio classification and categorization based on wavelets and support vector machine," *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 5, pp. 644-651, Sept. 2005. Article (CrossRef Link)

[52] H. Wu and J. M. Mendel, "Classification of battlefield ground vehicles using acoustic features and fuzzy logic rule-based classifiers," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, pp, 56-72, Feb. 2007. Article (CrossRef Link)

[53] C. L. Huang and C. J. Wang, "A GA-based feature selection and parameters optimization for support vector machines," *Expert Systems with applications*, vol. 31, no. 2, pp. 231-240, Aug. 2006. Article (CrossRef Link)

Yujing He received her B.S. degree in Communication Engineering from the Chongqing University of Posts and Telecommunications, China, in 2014. She has completed her M.S. degree in the Key Lab of Mobile Communication Technology, Chongqing University of Posts and Telecommunications, China. During the period of Sept. 2015 to Aug. 2016, she has been an Exchange Student in Inha University, Korea. Her research focus includes flying ad-hoc networks and random access for MTC.

Ishtiaq Ahmad received the B.S. degree in electrical engineering from the University of Engineering and Technology (UET), Peshawar, Pakistan, in 2007, and the M.S. degree in electronic engineering from Inha University, South Korea, in 2014, where he is currently pursuing the Ph.D. degree with the Electronic Engineering Department.
From 2007 to 2008, he was a BSS Engineer with O&M Department, Zong Pakistan. Since 2009, he has been a Lecturer with the Faculty of Engineering and Technology, Gomal University, Pakistan. He has authored several international journals and IEEE conference papers, and also holds US and Korean patents. His research interests include interference management in 3GPP LTE-A & 5G systems, cross-layer design, public safety & mobile ad-hoc networks (for AUV & Ship), maritime/underwater communications, and connected-autonomous vehicle (C-AV). Mr. Ahmad was a recipient of the Jungseok International Scholarship to pursue his M.S. and Ph.D. degrees at Inha University, due to his excellent academic career. He received the Outstanding Research Award from the Inha University, in 2019, for his Excellence of Journal Publication.

Lin Shi received her B.S. degree in Communication Engineering from the Harbin Engineering University, China, in 2014. She has completed her M.S. degree in the Key Lab of Mobile Communication Technology, Chongqing University of Posts and Telecommunications, China. During the period of Sept. 2015 to Aug. 2016, she has been an Exchange Student in Inha University, Korea. Her research focus includes flying ad-hoc network and 5G new waveform technology.

KyungHi Chang received the B.S. and M.S. degrees in electronics engineering from Yonsei University, Seoul, South Korea, in 1985 and 1987, respectively, and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, TX, USA, in 1992. From 1989 to 1990, he was with the Samsung Advanced Institute of Technology (SAIT) as a member of the research staff and was involved in digital signal processing system design. From 1992 to 2003, he was with the Electronics and Telecommunications Research Institute (ETRI) as a Principal Member of the technical staff, where he led the design teams involved in the WCDMAUE modem and 4G radio transmission technology (RTT). He is currently with the Electronic Engineering Department, Inha University. His research interests include 3GPP LTE-A&5G New RAT, HetNet, cross-layer design, public safety and mobile ad-hoc networks (for AUV & Ship), maritime/underwater communications, and connected autonomous vehicle (C-AV). Hewas a recipient of the LG Academic Awards, in 2006, the Haedong Best Paper Awards, in 2007, the IEEE ComSoc Best Paper Awards, in 2008, the Haedong Academic Awards, in 2010, and the SKT SafeNet Best Idea Awards, in 2015. He is currently a Chairman of the expert committee in SafeNet Forum, and mobile and automotive convergence committee in 5G Forum. He has served as an Editor-in-Chief and an Executive Director for the Journal of Korean Institute of Communications and Information Sciences (KICS), from 2010 to 2012 and in 2013, respectively. He has also served as an Editor of ITU-R TG8/1 IMT.MOD.