An Acoustic-Based Surveillance System for Amateur Drones Detection and Localization

Zhiguo Shi, *Senior Member, IEEE*, Xianyu Chang, Chaoqun Yang, *Student Member, IEEE*, Zexian Wu, and Junfeng Wu

***Abstract*—Due to cost reduction and device miniaturization, amateur drones are now widely used in numerous civilian and commercial applications. However, the abuse of amateur drones has resulted in emerging threats to personal privacy and public security. To alleviate these threats, we design an acoustic-based surveillance system, which can achieve the capacity of amateur drones detection and localization with 24/7 (24 hours per day and 7 days per week under normal circumstances) availability. In the designed system, a detection fusion algorithm and a TDOA estimation algorithm based on the Bayesian filter are applied to improve the performance of drone detection and localization. Field experiments are carried out, and the results demonstrate that the designed system can detect and locate an amateur drone in real time with high accuracy and 24/7 availability.**

***Index Terms*—Acoustic arrays, drone surveillance, detection fusion, drone localization, TDOA estimation.**

I. Introduction

AMATEUR drones, also called unmanned aerial vehicles, are recently enjoying great popularity due to low cost and various applications [1]–[3]. The popularity of amateur drones is a double sword. On the one hand, it facilitates us on account of boundless applications in communication, agriculture and numerous public services [4]–[9]. However, on the other hand, it also leads to social concerns due to the illegal abuse of amateur drones in privacy disclosure, drug trafficking and terrorist attacks, etc [10]–[12]. Therefore, there is an urgent need for amateur drones surveillance systems.

Recently, the increasing threats of amateur drones have motivated researchers to investigate drones surveillance systems. However, it is challenging to build a real-time drones surveillance system due to the small size, low speed and low flying altitude of amateur drones [13]–[15]. Generally speaking, a drone surveillance system should have the functions of detecting and locating drones in real time with high accuracy and 24/7 availability. According to [16], the technologies utilized for amateur drones surveillance mainly include acoustic-based, radar-based, video-based and radio-frequency(RF)-based technologies. Radar is an effective sensor for detecting and tracking flying objects. However, due to low altitude, small size and slow speed of drones, it is difficult for radars to achieve efficient drone surveillance. Optical sensors can display the appearances and traces of drones. However, it is hard to accurately obtain the actual spatial velocities and positions of drones. By capturing and analyzing the communicating signals between drones and their controllers, RF-based surveillance approaches can achieve drone detection and tracking. However, it is challenging to achieve drone surveillance considering pre-set flight path, ambient RF noise and multi-path effect. Recently, with the popularity of electronic devices, there are some literatures investigating drones detection and localization using crowdsensing [17]–[19]. Though drones detection and localization based on crowdsensing is a quite good idea, it is far from real application because of several limitations, such as the willing of participants. Based on the technical analysis in [16], it follows that acoustic-based technologies are pretty promising and feasible for amateur drones surveillance at short range (about no more than 200 meters).

However, there exists few preliminary work on acoustic-based amateur drones surveillance systems [20]–[22]. Christnacher *et al.* [20] deployed a tetrahedral shape acoustic array with four microphones to estimate the direction of arrival (DOA) of an amateur drone. Nevertheless, it poses high errors and low accuracy on the estimation of the drone's azimuth and the elevation angles. Tong *et al.* [21] applied a cross acoustic array to track an amateur drone with the trajectory of a straight line. But only the drone with special trajectories such as a straight line can be tracked. Busset *et al.* [22] introduced an advanced acoustic camera composed of a 120-element microphone array and a video camera to detect amateur drones, however, it fails to locate the drone and the cost of the system is expensive. In summary, the research on acoustic-based amateur drones surveillance systems is still in infancy, and the recent development of acoustic-based drones surveillance systems is far from meeting the demand of real-time surveillance with high accuracy and 24/7 availability.

Motivated by this, in this paper we build an acoustic-based surveillance system designed to be real-time, effective and 24/7 available for amateur drones detection and localization. A brief introduction of the acoustic-based surveillance system has been

Manuscript received July 12, 2019; revised October 28, 2019 and December 23, 2019; accepted December 26, 2019. Date of publication January 6, 2020; date of current version March 12, 2020. This work was supported by National Natural Science Foundation of China under Grants 61772467, 61731004, and 61901413. The review of this article was coordinated by Dr. L. Zhao. *(Corresponding author: Junfeng Wu.)*

Z. Shi is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China, and also with the Alibaba-Zhejiang University Joint Institute of Frontier Technologies, Hangzhou 310027, China (e-mail: shizg@zju.edu.cn).

X. Chang, Z. Wu, and J. Wu are with the State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China (e-mail: 363093247@qq.com; 339822315@163.com; jfwu@zju.edu.cn).

C. Yang is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: chaoqunyang@zju.edu.cn).

Digital Object Identifier 10.1109/TVT.2020.2964110

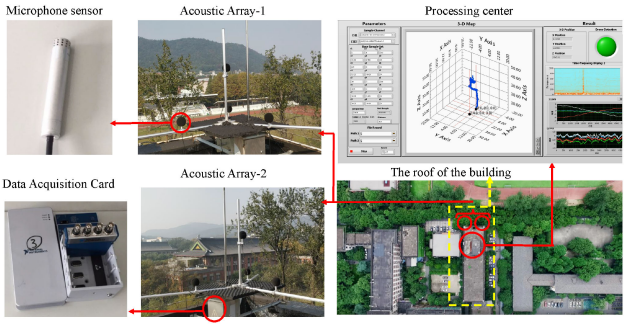


Fig. 1. System deployment. This system includes two tetrahedron-shape acoustic arrays, two data acquisition cards, and one processing center. A GUI to display results is shown at the top right corner, where the center is used to display the real-time localization results, and the circle at the top right corner is used to display the real-time detection results.

presented in the conference [23]. Compared with the conference paper which focuses on drone localization, in this paper we introduce the uni-microphone drone detection algorithm and multi-microphone detection fusion algorithm of our system. Moreover, we propose a novel TDOA estimation algorithm based on Bayesian framework which shows great advantages. In summary, this paper presents a more complete study of the designed system and the associated algorithms for amateur drones detection and localization. The main contributions of this paper are listed as follows:

1) We design an acoustic-based surveillance system, which has the functions of drones detection and localization with high accuracy and 24/7 availability.
2) We present an effective feature extraction method to extract the features of drones' acoustic signals based on which SVM is applied for uni-microphone drone detection. Besides, a multi-microphone detection fusion algorithm is adopted to improve the accuracy of drone detection.
3) We propose a new time delay of arrival (TDOA) estimation algorithm based on the Bayesian filter and the Gaussian mixture model, which effectively eliminates the effect of multiple paths in the process of drones localization.

The rest of this paper is organized as follows. Section II presents the architecture of the designed acoustic-based surveillance system. The algorithms for drones detection and localization are presented in Section III and Section IV, respectively. The results of field experiments are shown in Section V. Finally, some remarks are given in Section IV.

II. System Architecture

As shown in Fig. 1, the surveillance system, deployed on the rooftop of the Administration building at Zhejiang University, consists of a data acquisition part and a processing center.

The data acquisition part is composed by two tetrahedron-shape arrays and two data acquisition cards. As shown in Fig. 2, each array consists of four CHZ-213 1/2-inch metrological microphones with waterproof design whose parameters are shown in Table I. For simplicity, we label the eight microphones as Mic-1, Mic-2, . . ., Mic-8. The distance between each microphone and the bottom center is 1 m in each tetrahedron-shape array, and

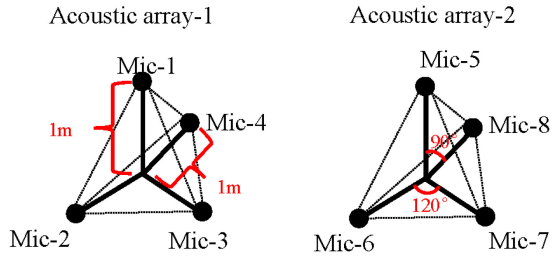


Fig. 2. Acoustic array model.

TABLE I
Specifications of the Microphones

NAME	CHZ-213
Frequency response range	10 Hz ~20 kHz
Sound field	free field
Upper limit of dynamic range (3% distortion)	>146 dB
Background noise	<16 dBA
Pressure coefficient (250 Hz)	-0.010 dB/kPa
Temperature coefficient (250 Hz)	-0.005 dB/℃ (-20~+60 ℃)

the distance between the two tetrahedron-shape arrays is 14 m. The type of the two data acquisition cards is NI-9234 with four channels. Their sample rate is set to 25600 Hz, and their output digitized signals will be sent to the processing center through optical fiber.

The functions of processing center include configurating the surveillance system, operating detection and localization algorithms, and displaying the results of drones surveillance. The hardware of the processing center is a computer with an Intel Core i7-6700 CPU and an AMD Radeon HD 7700 series graphics card. The software of the processing center is programmed by Labview and Matlab. Therein, the procedure programmed with Labveiw is to deal with data access, data pre-processing, and displaying results, while the procedure programmed with Matlab is to deal with data processing including detection and localization.

III. Drones Detection

The detection procedure of the proposed surveillance system includes two steps: feature extraction of drones' acoustic signals and SVM-based detection fusion.

A. Feature Extraction of Drones' Acoustic Signals

The acoustic signals of an amateur drone contain multiple components, including propeller blades sound, engine noise, wind noise, etc, of which propeller blades sound, a signal composed of a series of harmonics, takes the superiority in terms of amplitude [24]. Therefore, the acoustic signals of the amateur drone captured by Mic-i at time step t can be approximately expressed as

$$x_i(k) = s\left(k-\tau_i\right) + n_i(k) = \sum_{h=1}^{H} a_h \sin[2\pi f_h(k-\tau_i)+\psi_h] + n_i(k), \quad (1)$$

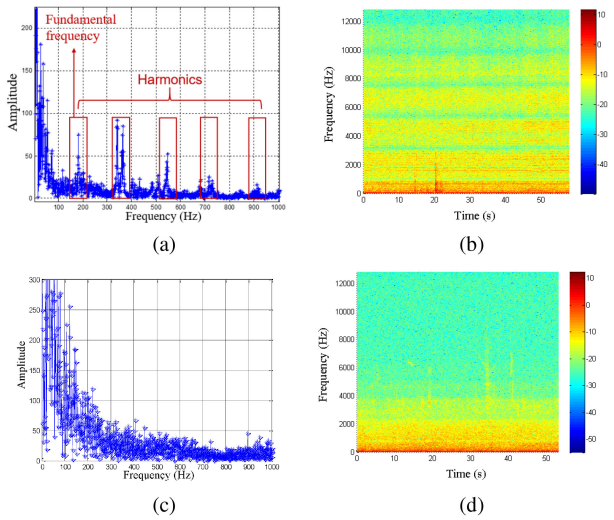


Fig. 3. Time-Frequency analysis of the drone's signals and background noise. (a) Frequency spectrum of the drone's signals. (b) Time-frequency spectrum of the drone's signals. (c) Frequency spectrum of background noise. (d) Time-frequency spectrum of background noise.

where $i = 1, 2 \ldots, 8$, h is the number of harmonics, a_h and ψ_h are the amplitude and the phase of h-th harmonic, respectively, τ_i is the sound propagation time delay from the drone to Mic-i, and $n_i(k)$ is the noise.

Time-Frequency analysis such as short-time Fourier transform (STFT) is used to analyze the feature contained in acoustic signals. Fig. 3 presents the frequency spectrum and time-frequency spectrum of the drone's signals and the background noise. Thereinto, in Fig. 3(b) and Fig. 3(d), different colors represent different spectrum amplitudes of the acoustic signals. From Fig. 3(a) and Fig. 3(b), it follows that the main difference between the acoustic signals of the drone and background noise is the existence of harmonics, which implies that harmonics' features can be utilized for drone detection.

In the designed surveillance system, the harmonics' feature is captured by extracting cepstral coefficients (CC) of the drone's acoustic signals. In order to capture the temporal features of drone's acoustic signals, apart from CC, the first-order and second-order frame-to-frame spectrums difference of CC are also extracted. By concatenating these features, we obtain a higher dimensional feature vector.

B. SVM-Based Detection Fusion

In the designed surveillance system, we detect the drone by distinguishing whether the extracted feature belongs to the drone or the background noise. In other words, the problem of drone detection is modeled as a two-class classification problem. If the extracted feature is classified as the drone's feature, the drone is detected. If the extracted feature is classified as the background noise's feature, the drone is not detected.

Support vector machine (SVM) is a typical supervised learning model used for classification analysis [25]. We use SVM to classify the feature vectors of the drone's acoustic signals and background noise. In other words, the drone is detected if the extracted feature vectors of the received acoustic signals are classified as the feature vectors of the drone's acoustic signals.

In the designed surveillance system, eight microphones are deployed to capture the drones' acoustic signals. Therefore, there exists eight detection results at each time step. Let $\boldsymbol{y}$ denote the detection results of all microphones, i.e.,

$$\boldsymbol{y} = [y_1, y_2, \ldots, y_8]^T, \tag{2}$$

where y_i obtained via SVM is the detection result of Mic-i, taking values from $\{-1, 1\}$. Specifically, $y_i = -1$ means that no drone is detected, while $y_i = 1$ represents that the drone is detected.

To improve the detection performance, detection fusion is considered in our system, i.e., given $\boldsymbol{y}$, finding a mapping function $f(\boldsymbol{y})$ to fuse all individual detection results. From the viewpoint of minimizing the overall probability of errors, according to [26], the optimal fusion strategy is $f(\boldsymbol{y}) = \boldsymbol{w}\boldsymbol{y}$, where $\boldsymbol{w}$ denotes the fusion weight vector, which is given as

$$\boldsymbol{w} = [w_1, w_2, \ldots, w_8], \tag{3}$$

and w_i represents the fusion weight of Mic-i. From the view of minimizing the overall probability of errors, the optimal fusion weight can be given as [26]

$$w_i = \begin{cases} \log \frac{P_i}{F_i}, & \text{if } y_i = 1 \\ \log \frac{1-F_i}{1-P_i}, & \text{if } y_i = -1, \end{cases} \tag{4}$$

where P_i and F_i denote the detection rate and false alarm rate of the Mic-i, which largely depend on the signal-to-noise ratio (SNR). However the SNR is hard to estimate in real time since it is only from the God's perspective that one can know whether there is a drone. Here we estimate them by extensive Monte-Carlo experiments. Specifically, we record ten acoustic episodes and each of them lasts ten minutes. Among them, five episodes consist of only noise signal, and the other five episodes contain the drone signal where the drone is flying randomly between 10m and 20m. Then the detection and false alarm rate are calculated as the mean of samples respectively.

After the estimation of P_i and F_i, the detection fusion rule given as follows:

$$x_f = \begin{cases} 1, & \text{if } \sum_{i=1}^{8} \boldsymbol{w}\boldsymbol{y} > 0 \\ -1, & \text{otherwise.} \end{cases} \tag{5}$$

IV. Drone Localization

Target localization has been studied extensively in the last decade [27]–[29]. The TDOA-based approach is one of the most popular localization approaches for acoustic source localization [30]–[32]. An accurate TDOA estimate is the premise of the TDOA-based approach. However, it is challenging to estimate TDOA with high accuracy in the designed surveillance system in view of multipath effects. In the rest of this section, we will first present the effects caused by multiple paths, and then propose an algorithm to overcome the multipath effects.

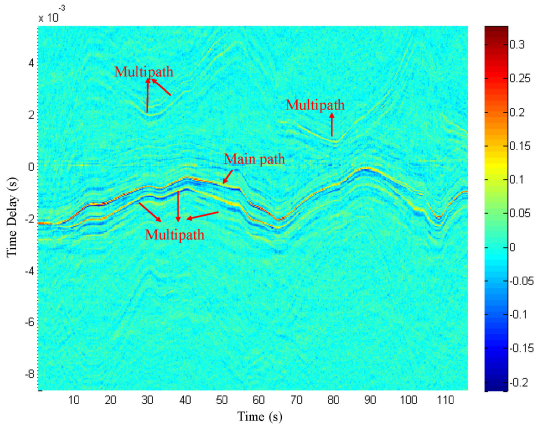


Fig. 4. $R_{x_m x_n}(\tau, k)$ with different τ and different k (The position with peak values are marked in deep colors, and there are mainly two lines that go in similar trends).

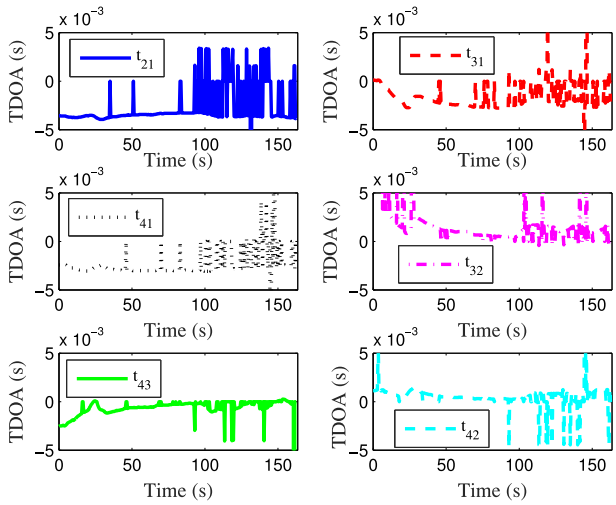


Fig. 5. The curves composed by the τ_{max} of each independent cross correlation $R_{x_m x_n}(\tau, k)$ with different k, there exit numerous peaks due to multipath effects.

A. Multipath Effects

For the Mic-m and Mic-n of acoustic array-1 in Fig. 2, where $m, n = 1, 2, 3, 4$ and $m \neq n$, $x_m(t)$ and $x_n(t)$ are used to denote the acoustic signals received by the two microphone sensors, and $G_{x_m x_n}(f)$ is their cross power spectrum. Then the cross correlation of the two received signals can be calculated by using the following generalized cross correlation with phase transform (GCC-PHAT) method as

$$R_{x_m x_n}(\tau, k) = \int_{-\infty}^{\infty} \frac{G_{x_m x_n}(f)}{|G_{x_m x_n}(f)|} e^{-j2\pi f \tau} df, \tag{6}$$

where $R_{x_m x_n}(\tau, k)$ is the cross correlation at time step k. Generally speaking, the τ associated with the maximum value of $R_{x_m x_n}(\tau, k)$, τ_{max}, is the estimated TDOA between the two signals. However, as mentioned above, due to multipath effects, it is challenging to estimate TDOA with high accuracy in the designed surveillance system.

Fig. 4 and Fig. 5 present the effects caused by multiple paths on the calculated TDOA. Fig. 4 presents $R_{x_m x_n}(\tau, k)$ with different τ and different k. As the figure shows, there exists at least two similar curves composed by the maximum values of $R_{x_m x_n}(\tau, k)$ at different k, which means that the maximum values may be contaminated by multipath effects.

According to the acoustic array model in Fig. 2, the number of independent cross correlation for acoustic array-1 is six. For each independent cross correlation, Fig. 5 presents the curves composed by the τ_{max} of each independent cross correlation $R_{x_m x_n}(\tau, k)$ with different k. Due to the limitation of the drone's velocity, it follows that the TDOA will not change dramatically in the two successive time steps. However, there exits numerous peaks in Fig. 5, which will lead to the wrong localization results. Therefore, due to too numerous peaks presented in Fig. 5, the TDOAs are nearly invalid after 90 s.

B. TDOA Estimation Based on Bayesian Framework

Due to the limitation of the drone's velocity, it is reasonable to assume that the TDOA will not change dramatically in the two successive time steps. Accordingly, historical information can be utilized to optimize the estimation of TDOA. We consider the Bayesian framework which iteratively estimates the current state of a parameter based on historical estimated states and current observation in our system [33]. The Bayesian framework includes two steps: prediction and updating.

Prediction. Suppose that the estimated TDOA at time step $k-1$ is $\widehat{\tau}_{mn}(k-1)$, which is the time delay corresponding to the maximum value of the estimated normalized GCC function $\widehat{R}_{mn}(\tau, k-1)$, and $\widehat{R}_{mn}(\tau, k-1)$ is represented by the blue curve with circle in Fig. 6.

$\widehat{R}_{mn}(\tau, k|k-1)$ can be approximated by a Gaussian mixture model (GMM), i.e.,

$$\widehat{R}_{mn}(\tau, k|k-1) \approx \sum_{i=1}^{I} g_i \mathcal{N}(u(k|k-1), \sigma^2(k|k-1)), \tag{7}$$

where $\mathcal{N}(u(k|k-1), \sigma^2(k|k-1))$ denotes a Gaussian distribution with mean $u(k|k-1)$ and covariance $\sigma^2(k|k-1)$. The GMM is constructed as follows [34]:

1) Determine the I peaks of $\widehat{R}_{mn}(\tau, k|k-1)$;
2) Determine the I blocks $\{B_1, \ldots, B_K\}$ associated with the different peaks. Each block means the peak interval, which starts at its left foot and ends at the right foot. As the red curve with triangle in Fig. 6 shows, the $\widehat{R}_{mn}(\tau, k|k-1)$ is approximated by a GMM with four Gaussian distributions. Meanwhile, $\{B_1, \ldots, B_4\}$ associated with the four peaks are labeled in Fig. 6.
3) Calculate the Gaussian distribution corresponding to each block;
4) Normalize the weights $\{g_1, \cdots g_I\}$.

The Gaussian distribution $\mathcal{N}(u(k|k-1), \sigma^2(k|k-1))$ associated with the i-th block B_i is given by [34]:

$$u(k|k-1) = \tau_{mn}^i(k-1), \tag{8}$$

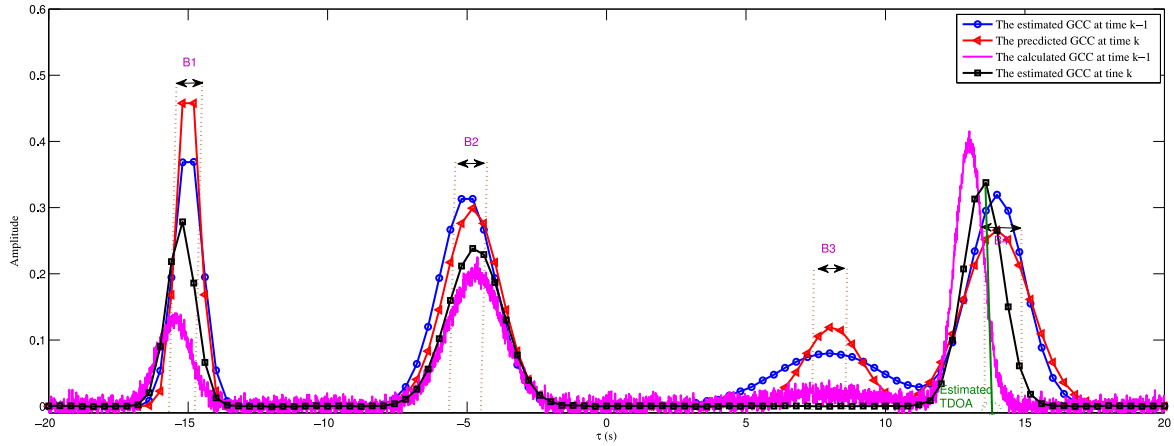


Fig. 6. A sketch of the TDOA estimation based on Bayesian framework.

$$\sigma^2(k|k-1)=\frac{\sum\limits_{i,\tau_j\in B_i}(\tau_j-u(k|k-1))\widehat{R}_{mn}\left(\tau_j,k-1\right)}{\sum\limits_{i,\tau_j\in B_i}\widehat{R}_{mn}\left(\tau_j,k-1\right)}, \tag{9}$$

where $\tau^i_{mn}(k-1)$ is the time delay corresponding to the maximum value of in the i-th block B_i. Suppose that there is only one drone, then I is set to 1 and g_1=1, which give

$$u(k|k-1)=\widehat{\tau}_{mn}(k-1), \tag{10}$$

$$\sigma^2(k|k-1)=\frac{\sum_{\tau\in B_1}(\tau-u(k|k-1))\widehat{R}_{mn}\left(\tau,k-1\right)}{\sum_{\tau\in B_1}\widehat{R}_{mn}\left(\tau,k-1\right)}. \tag{11}$$

Update. When receiving the new normalized GCC function calculated via Eq. (6), which is shown by the pink curve with no symbol in Fig. 6, the estimated normalized GCC function $\widehat{R}_{mn}(\tau,k)$ can be derived as

$$\widehat{R}_{mn}\left(\tau,k\right)=\widehat{R}_{mn}(\tau,k|k-1)R_{mn}(\tau,k). \tag{12}$$

In Fig. 6, $\widehat{R}_{mn}(\tau,k)$ is shown by the black curve with square. Accordingly, the estimated TDOA at time step k can be given as

$$\widehat{\tau}_{mn}(k)=\operatorname{argmax}\widehat{R}_{mn}\left(\tau,k\right). \tag{13}$$

Based on the measured data, for simplicity, Fig. 7 presents an example to denote the relationship among the predicted, calculated and estimated GCC functions, where the predicted GCC function is approximated by a GMM with one Gaussian distribution.

C. Localization Based on Estimated TDOAs

We use $\boldsymbol{S}_m$, $\boldsymbol{S}_n$ and $\boldsymbol{S}_0(k)$ to denote the spatial locations of Mic-m, Mic-n and the drone at time step k, respectively. Then for the pair of Mic-m and Mic-n in array-1, we have

$$d_{mn}(k)=\left\|\boldsymbol{S}_m-\boldsymbol{S}_0(k)\right\|-\left\|\boldsymbol{S}_n-\boldsymbol{S}_0(k)\right\|, \tag{14}$$

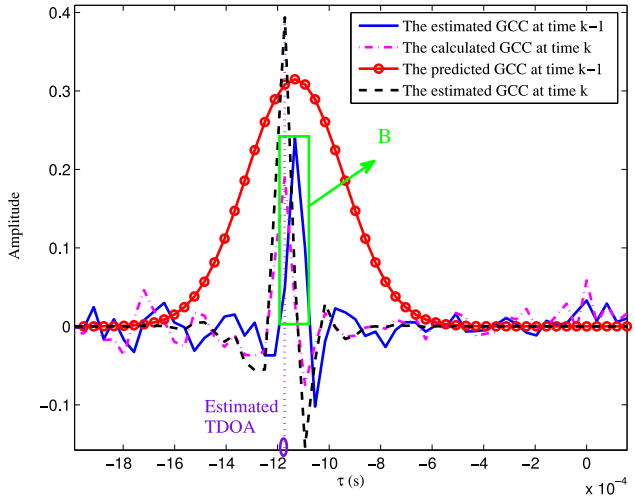


Fig. 7. The calculated and estimated GCC functions.

where $d_{mn}(k)=c\widehat{\tau}_{mn}(k)$, is the propagation difference of arrival between the two microphones at time step k, c is the sound velocity, and $\widehat{\tau}_{mn}(k)$ has been estimated using Eq. (13).

Squaring the Eq. (14) in both sides, it follows that

$$(\boldsymbol{S}_m-\boldsymbol{S}_n)\boldsymbol{S}_0^T(k)+d_{mn}(k)d_{n0}(k)=h_{mn}(k), \tag{15}$$

where

$$d_{n0}(k)=\left\|\boldsymbol{S}_n-\boldsymbol{S}_0(k)\right\|, \tag{16}$$

$$h_{mn}(k)=(\boldsymbol{S}_m\boldsymbol{S}_m^T-\boldsymbol{S}_n\boldsymbol{S}_n^T-d^2_{mn}(k))/2, \tag{17}$$

which yields the following form:

$$\left[\left(\boldsymbol{S}_m-\boldsymbol{S}_n\right)\,d_{mn}(k)\right]\begin{bmatrix}\boldsymbol{S}_0^T(k)\\ d_{n0}(k)\end{bmatrix}=h_{mn}(k). \tag{18}$$

Similarly, for array-2, it follows that

$$\left[2(\boldsymbol{S}_{\tilde{m}}-\boldsymbol{S}_{\tilde{n}})\;2d_{\tilde{m}\tilde{n}}(k)\right]\begin{bmatrix}\boldsymbol{S}_0^T\\ d_{\tilde{n}0}(k)\end{bmatrix}=h_{\tilde{m}\tilde{n}}, \tag{19}$$

where $\tilde{m}$ and $\tilde{n}$ denote Mic-$\tilde{m}$ and Mic-$\tilde{n}$ in array-2, respectively.

To make full use of all the estimated TDOAs, we choose $n=1$ as the reference node, then we can write Eq. (18) for $m=2,3,4$

in acoustic array-1. Similarly, choosing $\tilde{n} = 5$ as the reference node, we can also write the similar equations for $\tilde{m} = 6, 7, 8$ in acoustic array-2. Then we have:

$$\boldsymbol{A}(k)\boldsymbol{X}(k) = \boldsymbol{H}(k), \tag{20}$$

where

$$\boldsymbol{A}(k) = \begin{bmatrix} \boldsymbol{S}_2 - \boldsymbol{S}_n & d_{2n}(k) & 0 \\ \boldsymbol{S}_3 - \boldsymbol{S}_n & d_{3n}(k) & 0 \\ \boldsymbol{S}_4 - \boldsymbol{S}_n & d_{4n}(k) & 0 \\ \boldsymbol{S}_6 - \boldsymbol{S}_{\tilde{n}}(k) & 0 & d_{6\tilde{n}}(k) \\ \boldsymbol{S}_7 - \boldsymbol{S}_{\tilde{n}}(k) & 0 & d_{7\tilde{n}}(k) \\ \boldsymbol{S}_8 - \boldsymbol{S}_{\tilde{n}}(k) & 0 & d_{8\tilde{n}}(k) \end{bmatrix}, \tag{21}$$

$$\boldsymbol{H}(k) = [h_{2n}(k) \quad h_{3n}(k) \quad h_{4n}(k) \quad h_{6\tilde{n}}(k) \quad h_{7\tilde{n}}(k) \quad h_{8\tilde{n}}(k)]^T, \tag{22}$$

and

$$\boldsymbol{X}(k) = [\boldsymbol{S}_0(k) \quad d_{n0}(k) \quad d_{\tilde{n}0}(k)]^T. \tag{23}$$

Eq. (20) holds true only approximately in practice because of measurement errors and microphone calibration errors [35], [36]. Consequently, a reasonable method to estimate $\boldsymbol{X}(k)$ based on Eq. (20) is via the minimization of the following least-squares (LS) criterion [37]:

$$\gamma = \|\boldsymbol{A}(k)\boldsymbol{X}(k) - \boldsymbol{H}(k)\|. \tag{24}$$

Regardless of the dependence of $\boldsymbol{X}_1(k) = \boldsymbol{S}_0^T(k)$ on the remaining elements of $\boldsymbol{X}(k)$, it is well known that this unconstrained minimization of Eq. (24) is [38]

$$\widehat{\boldsymbol{X}}(k) = \left(\boldsymbol{A}^T(k)\boldsymbol{A}(k)\right)^{-1}\boldsymbol{A}^T(k)\boldsymbol{H}(k). \tag{25}$$

If $S_0(k) \in \mathbb{R}^3$, then the corresponding unconstrained LS estimate $\widehat{\boldsymbol{S}}_0(k)$ of $\boldsymbol{S}_0(k)$ is given by

$$\widehat{\boldsymbol{S}}_0^T(k) = [I_3 \quad 0_{3\times 2}]\widehat{\boldsymbol{X}}(k), \tag{26}$$

where I_n denotes the identity matrix of size n and $0_{i\times j} \in \mathbb{R}^{i\times j}$ denotes the matrix with its elements all being 0.

This combination of the above equations can obtain one result of drone's location. Observing the above formulas, we can find that there are six equations but five unknown parameters. Even if one of the TDOA calculations is inaccurate, the finally result can be estimated properly. In reality, the computational redundancy provided by the two arrays can effectively improve the accuracy of localization.

Similarly, let $n = 2, 3, 4$ and $\tilde{n} = 2, 3, 4$ as reference points respectively, and let $m \neq n$ and $\tilde{m} \neq \tilde{n}$, we can obtain numerous different combinations of equations for acoustic array-1 and array-2. Generally speaking, suppose that there are M microphone sensors in each acoustic array, and there are N acoustic arrays, then we can write M^N kinds of combined equations totally. For the designed system, $M = 4$, $N = 2$, so we can get 16 kinds of combined equations and 16 kinds of location results.

Further, the obtained 16 location results are clustered, and the center of the largest cluster is taken as as the position of the drone. Fig. 8 shows the clustered result at one time step, where the 16 locations are assigned into two clusters. It is easy to see that the cluster 1 is most likely to be the true values.

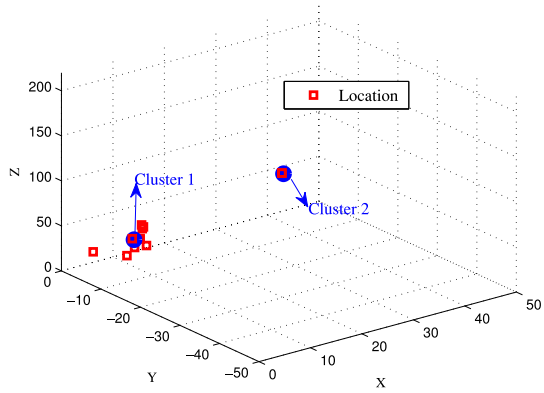


Fig. 8. The cluster result in one time step.

Therefore we regard the center of cluster 1 as the position of the drone at this time step.

V. Performance Evaluation

In this section, field experiments are conducted to verify the designed system's performance of drone detection and localization. It is notable that our system can run in real time with high accuracy and 24/7 availability, i.e., 24 hours per day and 7 days per week under normal circumstances. Abnormal circumstances mainly include some bad weather conditions, such as rainy and snowy day. In fact, the drone can not fly under there circumstances, so we don't need to care about. Considering the environment in the daytime is more noisy and complex than that at night, we only present the experiment results in the daytime. It is sure that the performance of drone surveillance at night will be better. DJI Phantom 3 is used in our experiments. In fact, for most amateur drones, their acoustic signal frequency bands are close to 200 HZ, which means that different types of drones do not affect our experimental results.

A. Drone Detection Evaluation

The detection rate is influenced by numerous factors, including the sensitivities of microphones, surrounding noise, the distances between the drone and the arrays, etc. In essential, these factors can be attributed to the SNR. To measure the SNR, we collect the surrounding noise for a long time when there is no drone in the surveillance region, and we divide the collected noise into L segments, then we obtain the average spectrum of the stationary noise as:

$$\left|\bar{X}_{\text{noi}}(\omega)\right| = \sum_{q=1}^{L} |X_{\text{noi}}^q(\omega)|/L, \tag{27}$$

where $X_{\text{noi}}^q(\omega)$ is the Fourier transform of each segment noise. Another stationary noise processing method can be referred to [39]. If the noise is non-stationary, we could estimate noise by minimum statistic algorithm proposed in [40]. Further, we evaluate the SNR of the drone's acoustic signals as:

$$R_{\text{SNR}} = 10\log\frac{|X_{\text{dro}}(\omega)|^2}{\left|\bar{X}_{\text{noi}}(\omega)\right|^2}, \tag{28}$$

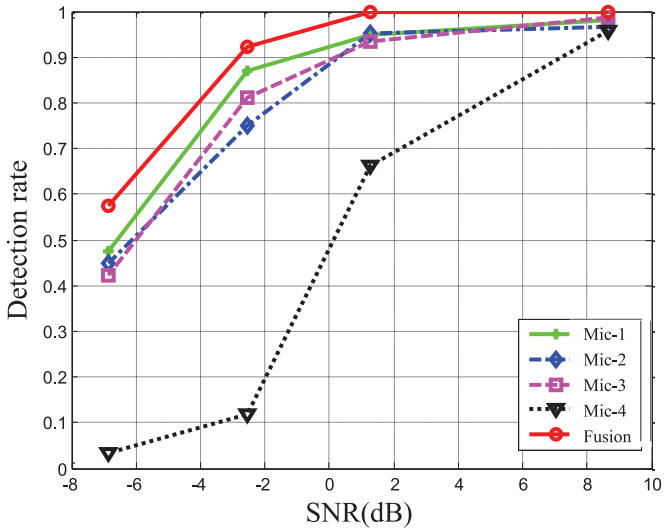


Fig. 9. The detection results of acoustic array-1.

TABLE II
THE RESULTS OF THE FALSE ALARM RATE

	Mic-1	Mic-2	Mic-3	Mic-4	Fusion	Average
Test 1	6.17%	5.20%	9,83%	1.11%	5.06%	5.58%
Test 2	3.26%	1.63%	5.95%	1.13%	1.91%	2.99%
Test 3	11.93%	5.97%	6.21%	1.67%	6.44%	6.45%

where $X_{\mathrm{dro}}(\omega)$ is the Fourier transform of the drone's acoustic signal polluted by noise.

The feature vectors of the drone's acoustic signals are extracted every 0.5 s. The number of positive samples is 10044 and the number of negative samples is 7936. All the samples are used for SVM to train the classification model, and totally 1423 support vectors are obtained.

We let an amateur drone hover in different distances away from the acoustic arrays in order to obtain different SNRs. The detection results are calculated every 0.5 s, and the final detection rate is given as the probability that the drone is detected in this period of time.

The detection results are shown in Fig. 9. From this figure, we find that the detection rate increases as the SNR improves, and the performance is improved after detection fusion. The results of the false alarm rate are listed in TABLE II. Three groups of experiments have been carried out. The comparison of the false alarm rate reveals that the results of the proposed fusion algorithm is better than that of the average of the whole microphones. We note that the curve of Mic-4 is obviously different from others, the reason perhaps is that the characteristics of Mic-4, such as sensitivity have some problem and the SVM classification model is not such suitable for Mic-4.

B. Drone Localization Evaluation

The TDOA estimation results of acoustic array-1 by applying the proposed algorithm in Section IV-B are shown in Fig. 10. Comparing Fig. 10 with Fig. 5, it follows that the false TDOA peaks caused by multipath effects are eliminated and a smooth TDOA curve is abstracted for each microphone pair.

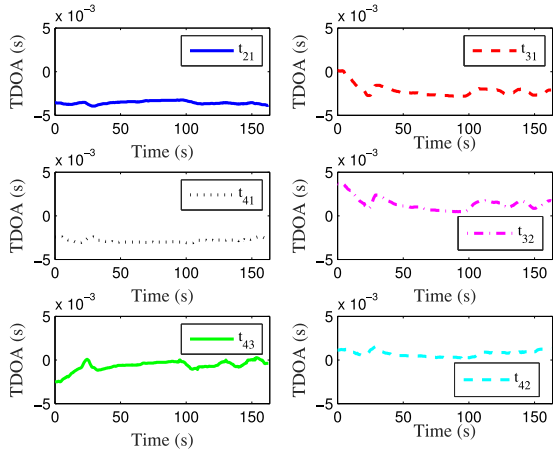


Fig. 10. The TDOA estimation results by using the proposed TDOA estimation algorithm.

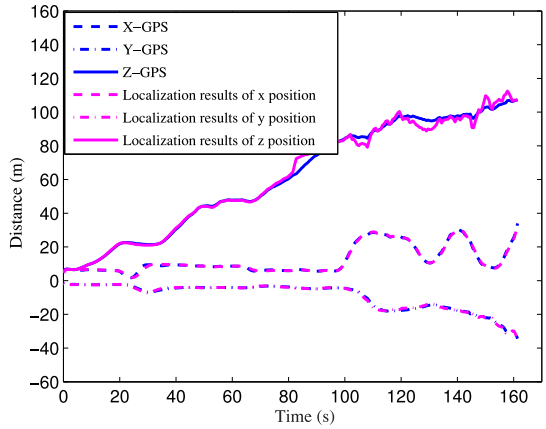


Fig. 11. The GPS ground truth and localization results.

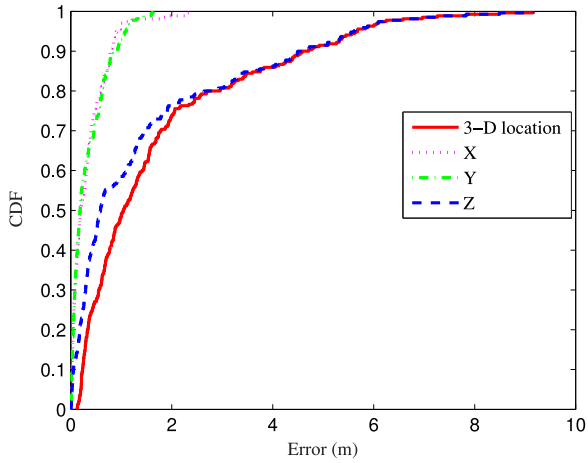


Fig. 12. Estimation errors of the drone's location.

We use the algorithm presented in Section IV-C to calculate the locations of the flying drone. The GPS data and the localization results are shown in Fig. 11. Fig. 12 shows the localization errors in Fig. 11 where CDF means the cumulative density function of the localization errors in our experiment results, for

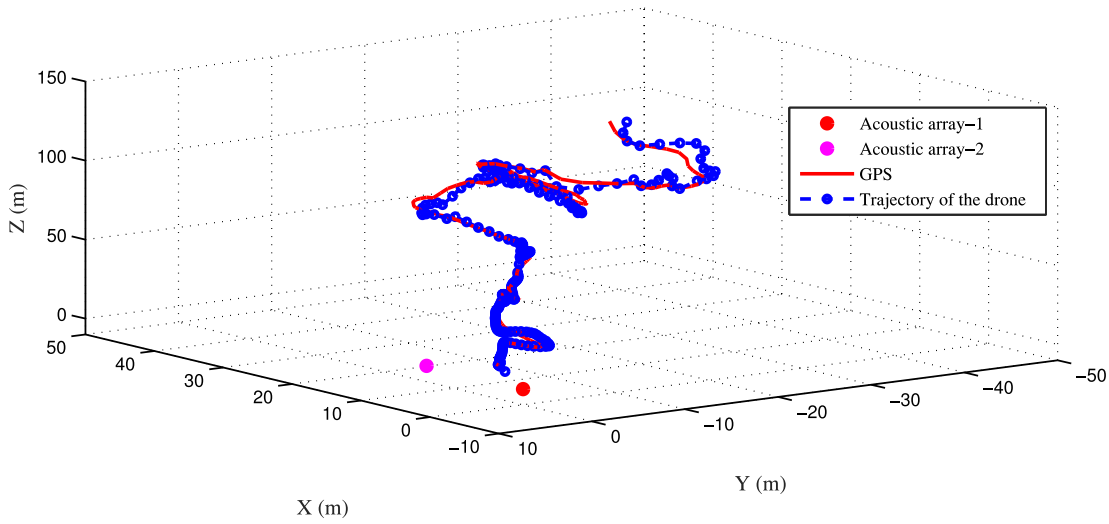


Fig. 13. The real-time trajectory of the drone when compared with GPS in 3-D coordinate axis (night-time).

example, we can see that 90% of the estimation errors are less than 5 m.

The 3-D trajectory of GPS data and the localization results are shown in Fig. 13. It follows that our system achieves good performance for drone localization. Moreover, more results of offline and online drone localization are shown in a demo.[1] The longest distance has reached more than 100 m under the condition that the SNR is lower than −5 dB.

VI. Conclusion

In this paper, we build up an acoustic-based drones surveillance system for drones detection and localization. We presents the whole drone detection and localization procedures of this system. A detection fusion algorithm and a TDOA estimation algorithm based on Bayesian framework are proposed to improve the performances of detection and localization respectively. Extensive field experiments are conducted to verify the real-time effectiveness of our system for drones detection and localization.

References

[1] G. Ding, Q. Wu, L. Zhang, Y. Lin, T. A. Tsiftsis, and Y. Yao, "An amateur drone surveillance system based on the cognitive Internet of Things," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 29–35, Jan. 2018.

[2] 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 Communi. Mag.*, vol. 56, no. 1, pp. 51–57, Apr. 2018.

[3] X. Zhou, Q. Wu, S. Yan, F. Shu, and J. Li, "UAV-Enabled secure communications: Joint trajectory and transmit power optimization," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4069–4073, Apr. 2019.

[4] N. H. Motlagh, T. Taleb, and O. Arouk, "Low-altitude unmanned aerial vehicles-based Internet of Things services: Comprehensive survey and future perspectives," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 899–922, Dec. 2016.

[5] C. Yin, Z. Xiao, X. Cao, X. Xi, P. Yang, and D. Wu, "Offline and online search: UAV multiobjective path planning under dynamic urban environment," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 546–558, Apr. 2018.

[1]https://youtu.be/FiJwWfiNv6g, https://youtu.be/rfpdOyaeqJY, https://youtu.be/ZZIgk5P-z5s.

[6] C. Yang, Z. Wu, X. Chang, X. Shi, J. Wu, and Z. Shi, "DOA estimation using amateur drones harmonic acoustic signals," in *Proc. IEEE SAM*, Sheffield, UK, Jul. 2018, pp. 587–591.

[7] M. Liu, J. Yang, and G. Gui, "DSF-noma: UAV-assisted emergency communication technology in a heterogeneous Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5508–5519, Jun. 2019.

[8] M. Li, N. Cheng, J. Gao, Y. Wang, L. Zhao, and X. Shen, "Energy-efficient UAV-assisted mobile edge computing: Resource allocation and trajectory optimization," *IEEE Trans. Veh. Technol.*, accepted.

[9] C. Zhou, Y. Gu, S. He, and Z. Shi, "A robust and efficient algorithm for coprime array adaptive beamforming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1099–1112, Feb. 2018.

[10] "Drone flying over SW China airport affects flights," Accessed: May 2, 2017. [Online]. Available: http://en.people.cn/n3/2017/0502/c90000-9209828.html

[11] W. Ripley, "Drone with radioactive material found on Japanese prime minister's roof," Accessed: Apr. 22, 2015. [Online]. Available: https://edition.cnn.com/2015/04/22/asia/japan-prime-minister-rooftop-drone/index.html

[12] S. Dinan, "Drones become latest tool drug cartels use to smuggle drugs into U.S.," Accessed: Aug. 20, 2017. [Online]. Available: https://www.washingtontimes.com/news/2017/aug/20/mexican-drug-cartels-using-drones-to-smuggle-heroi/

[13] D. He *et al.*, "A friendly and low-cost technique for capturing non-cooperative civilian unmanned aerial vehicles," *IEEE Netw.*, vol. 33, no. 2, pp. 146–151, Apr. 2019.

[14] C. Yang, L. Feng, Z. Shi, R. Lu, and K. R. Choo, "A crowdsensing-based cyber-physical system for drone surveillance using random finite set theory," *ACM Trans. Cyber Phys. Syst.*, vol. 4, no. 3, 2019.

[15] X. Yuan *et al.*, "Capacity analysis of UAV communications: Cases of random trajectories," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7564–7576, Aug. 2018.

[16] X. Shi, C. Yang, W. Xie, C. Liang, Z. Shi, and J. Chen, "Anti-drone system with multiple surveillance technologies: Architecture, implementation and challenges," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 68–174, Apr. 2018.

[17] G. Yang, S. He, and Z. Shi, "Leveraging crowdsourcing for efficient malicious users detection in large-scale social networks," *IEEE Internet Things J.*, vol. 4, no. 2, pp. 330–339, Apr. 2017.

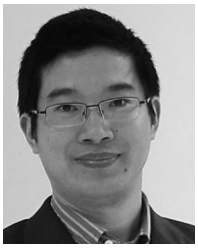
[18] G. Yang, S. He, Z. Shi, and J. Chen, "Promoting cooperation by the social incentive mechanism in mobile crowdsensing," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 86–92, Mar. 2017.

[19] G. Yang, X. Shi, L. Feng, S. He, Z. Shi, and J. Chen, "CEDAR: A cost-effective crowdsensing system for detecting and localizing drones," *IEEE Trans. Mob. Comput.*, to be published, doi: 10.1109/tmc.2019.2921962.

[20] F. Christnacher *et al.*, "Optical and acoustical UAV detection," in *Proc. SPIE*, Oct. 2016, pp. 1–13.

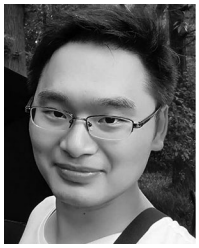
[21] J. Tong, W. Xie, Y. Hu, M. Bao, X. Li, and W. He, "Estimation of low-altitude moving target trajectory using single acoustic array," *J. Acous. Soc. Amer.*, vol. 139, no. 4, pp. 1848–1858, 2016.

[22] J. Busset, F. Perrodin, P. Wellig, B. Ott, K. Heutschi, and T. Nussbaumer, "Detection and tracking of drones using advanced acoustic cameras," in *Proc. SPIE Conf. Secur. Defence*, 2015, pp. 1–6.
[23] X. Chang, C. Yang, J. Wu, X. Shi, and Z. Shi, "A surveillance system for drone localization and tracking using acoustic arrays," in *Proc. IEEE SAM*, Sheffield, U.K., Jul. 2018, pp. 573–577.
[24] A. Sedunov, H. Salloum, A. Sutin, and N. Sedunov, "UAV passive acoustic detection," in *Proc. IEEE HST*, Woburn, USA, Oct. 2018, pp. 1–6.
[25] J. Dong, A. Krzyzak, and C. Suen, "Fast SVM training algorithm with decomposition on very large data sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 603–618, Apr. 2005.
[26] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, pp. 98–101, Jan. 1986.
[27] B. Zhou, Q. Chen, P. Xiao, and L. Zhao, "On the spatial error propagation characteristics of cooperative localization in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1647–1658, Feb. 2017.
[28] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and doa estimation based massive mimo system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.
[29] J. Wang, R. Ghosh, and S. Das, "A survey on sensor localization," *J. Contr. Theory Appl.*, vol. 8, no. 1, pp. 2–11, 2010.
[30] T. Kundu, "Acoustic source localization," *Ultrasonics*, vol. 54, no. 1, pp. 25–38, Jan. 2014.
[31] A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 2, pp. 439–443, Feb. 2013.
[32] J. Zhang and H. Liu, "Robust acoustic localization via time-delay compensation and interaural matching filter," *IEEE Trans. Signals Process.*, vol. 63, no. 18, pp. 4771–4783, Sep. 2015.
[33] C. Yang, Z. Shi, K. Han, J. J. Zhang, Y. Gu, and Z. Qin, "Optimization of particle cbmember filters for hardware implementation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 9027–9031, Sep. 2018.
[34] O. Youssef, F. Friedrich, D. Magimai, and K. Dietrich, "A TDOA gaussian mixture model for improving acoustic source tracking," in *Proc. EUSIPCO*, Bucharest, Romania, Aug. 2012, pp. 1339–1343.
[35] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoustic. Speech Signal Process.*, vol. 12, pp. 1661–1669, Dec. 1987.
[36] S. Fu, L. Zhao, Z. Su, and X. Jian, "UAV based relay for wireless sensor networks in 5G systems," *Sensors*, vol. 18, no. 8, pp. 2413, 2018.
[37] A. Beck, P. Stoica, and J. Li, "Exact and approximate solutions of source localization problems," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1770–1778, May 2008.
[38] P. Stoica and J. Li, "Source localization from range-difference measurements," *IEEE Signal Process. Mag.*, vol. 23, no. 6, pp. 63–66, Nov. 2006.
[39] J. Zhang, S. Chepuri, R. Hendriks, and R. Heusdens, "Microphone subset selection for MVDR beamformer based noise reduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 3, pp. 550–563, Mar. 2018.
[40] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001.



Zhiguo Shi (M'10–SM'15) received the B.S. and Ph.D. degrees in electronic engineering from Zhejiang University, Hangzhou, China, in 2001 and 2006, respectively. Since 2006, he has been a Faculty Member with the Department of Information and Electronic Engineering, Zhejiang University, where he is currently a Full Professor. From 2011 to 2013, he visited the Broadband Communications Research Group, University of Waterloo, Waterloo, ON, Canada. His current research interests include signal and data processing, and smart grid communication and network. He is an Editor for the IEEE NETWORK, *KSII Transactions on Internet and Information Systems*, and *IET Communications*. He was a recipient of the Best Paper Award from the IEEE Wireless Communications and Networking Conference Shanghai, China, in 2013, the IEEE/CIC International Conference on Communications in China Xi'an, China, in 2013, and the IEEE Wireless Communications and Signal Processing Huangshan, China, in 2012, and the Scientific and Technological Award of Zhejiang Province, China, in 2012.

Xianyu Chang received the B.Sc. degree from the College of Control Science and Engineering, China University of Petroleum, Qingdao, China, in 2016, and the M.E. degree from the College of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2019. His current research interests include array signal processing and deep learning model acceleration.



Chaoqun Yang (S'14) received the B.Sc. degree in marine technology from Xiamen University, Xiamen, China, in 2015, and the Ph.D. degree in information science and electronic engineering from Zhejiang University, Hangzhou, China, in 2019. He is currently an Engineer with the Nanjing Research Institute of Electronics, Nanjing, China. His current research interests include radar signal and data processing.

Zexian Wu received the B.Sc. degree in automatic from North China Electric Power University, Beijing, China, in 2017. He is currently working toward the master's degree with the Department of Control Science and Engineering, Zhejing University, Hanzhou, China. His current research interests include speech enhancement and speech localization.

Junfeng Wu received the B.Eng. degree from Zhejiang University, Hangzhou, China, and the Ph.D. degree in electrical and computer engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2009 and 2013, respectively. From September to December 2013, he was a Research Associate with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. From January 2014 to June 2017, he was a Postdoctoral Researcher with the Autonomic Complex Communication Networks, Signals and Systems, Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden. He is currently with the College of Control Science and Engineering, Zhejiang University. His research interests include networked control systems, state estimation, and wireless sensor networks, multiagent systems. He received the Guan Zhao-Zhi Best Paper Award at the 34th Chinese Control Conference in 2015.