CONTEMPORARY FEATURES EXTRACTION TECHNIQUES FOR DETECTING MALICIOUS DRONES

Ali Y. Al-Zahrani

Department of Electrical and Electronic Engineering, University of Jeddah, Jeddah21493, Saudi Arabia

ABSTRACT

Today, drone-based attacks represent serious threats to the security and safety of public infrastructures. For successfully detecting a malicious drone in a given zone, there are three phases: signal collection (sensing), features extraction and classifications. Signal collection can be performed using available sensing technologies such as radar, acoustics sensors and electro-optic technologies, among others. The classification phase is often achieved using general-purpose algorithms such as Naive Bayes and support vector machine (SVM). On the other hand, the features extraction phase is very problem-specific, and its performance depends on several factors such as the used sensory technology, environment, and the drone characteristics. Features engineering is a designing stage that aims at identifying the most distinctive information carriers which capture the drone's discriminative characteristics. In this paper, we present effective drones' features extraction techniques for the most popular sensory technologies available which are radar, RF analyzers, acoustic sensors, and electro-optic sensors. We focus on identifying the most distinctive features of drones and show how to extract them out of the collected signals.

KEYWORDS

Drone, Detection, Anti-Drone System, Features Engineering, Radar, Radio Frequency, Acoustics

1. INTRODUCTION

The emergence of drones has created many useful applications in many fields such as rescue, security, communications, environmental, among others [1]–[3]. However, drones pose significant safety and security concerns, with a notable increase in terrorist drone-based attacks observed recently. Today's drones are widely different and can be classified in many ways according to certain characteristics such as size, weight, flight range, application, aerodynamic technique, navigation and control method [3]. There are general classifications in the literature, for example, drones can be categorized as either low-flying, small, and slow (LSS) or low-flying, small and fast (LSF) [4]. Both types have their place in the airborne attacks, and both have the following capabilities:

- Ability to accomplish its task autonomously.
- Ability to form a swarm that can attack a target.
- Recognizing a victim's face among a crowd.
- Possessing more remote channels, and control options.
- Ability to paralyze civilian air traffic control.

To secure a given zone, an anti-drone system which encompasses drone detection, tracking and neutralization capabilities must be in place. Developing a drone detection capability is evidently a complex task because of the significant resemblance between drones and the background clutter.

Drones are similar to many natural objects such as birds in many aspects; physical size, flying altitude and velocity domain. This situation can even be worse when combined with the environmental, climatic and atmospheric abnormality.

For simplification, the detection process can be divided into three phases: signal collection, features engineering and classification. Signal collection phase can be performed using existing sensing technologies, and the classification phase can be achieved using standard schemes such as support vector machine (SVM) classifier. However, features engineering is very problem specific. It extracts from the collected signals relevant and distinctive information which uniquely captures drone's characteristics.

The performance of features engineering depends on several factors such as drone type, environment, and the type of technology being used for signal collection. Extracting appropriate features, which carry rich and relevant information, ensures useful model and ultimately produces an efficient classifier [5].

Nevertheless, there are cases where explicit features extraction is not required. When a deep learning (DL)-based classifier is selected, the algorithm implicitly learns the features from the input data/signal. However, this advantage of DL causes long delays, and requires more data and higher computation & memory resources.

In this paper, we present various engineering methods for drones' features extraction. We focus on the main sensory technologies including radar signal processing, acoustics signal processing, radio frequency (RF) analyzers, electro-optics (EO) technologies. We investigate various underlying signal processing techniques potentially employed in these technologies. The contributions of this paper can be summarized as follows:

- An architecture framework for anti-drone system is introduced.
- For each sensory technology, the signal processing required to prepare the received signal for features extraction, is demonstrated.
- The most important features, which carry unique characteristics related to drones, are introduced and mathematically described.

In Reference [3], we compared the aforementioned sensory technologies through a pros-and-cons analysis. Furthermore, we discussed the methodology for evaluating their detection performance. The rest of this paper is organized as follows. Section 2 briefly introduces the anti-drone system. Then, a thorough presentation of drone features engineering along with the underlying signal processing is placed in Section 3. Finally, Section 4 concludes this paper, summarizes the main challenges in this area and presents potential future directions.

2. ARCHITECTURE OF ANTI-DRONE SYSTEM

In general, offensive drones may perform two types of attacks: cyber-attack or physical attack. In the former type, the drone flies above a target zone and promotes itself as a free WiFi access point or a legitimate radio base station to be able to hack or infect the target network. In the latter type, the drone flies above a target zone to perform a physical task such as surveying or bombing. In both cases, drone detection, tracking and neutralization are the main functions of an effective anti-drone system (ADS). Figure 1 shows the basic mechanism of an anti- drone system. When a flying object is sensed by the ground apparatus, the received signal is processed, and the object features are extracted and sent to a well-trained classifier. If the features represent the characteristics of a drone with certain degree of confidence, the system declares a danger event

and hence invoke the threat mitigation measures. The neutralization functions aim at isolating the drone from its remote controller, taking control over, and capturing it or falling it down. This step is usually followed by comprehensive forensic investigations [6]. In the sequel, the detection technology, which is the first vital step in any ADS, is thoroughly discussed.

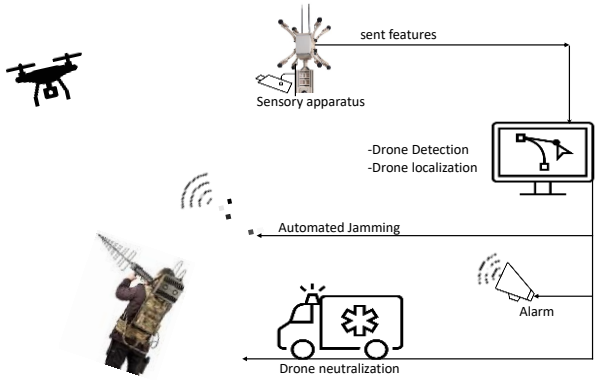

Figure 1. Illustration of the main components in anti-drone systems.

3. DRONE DETECTION

An anti-drone system possesses a detection function when it is capable of classifying objects into desired and non-desired objects. Thus, distinguishing drones from other objects such as balloons and birds is the core task of detection function. Figure 2 shows the three phases of drones detection. There is a set of sensory technologies which can broadly be divided into passive or active. The contemporary detection techniques depend on either radar signals, acoustic signals, RF signals, electro-optic sensors or a fused combination of them. In the following subsections, we focus on drone features extraction methods when the above-mentioned sensory technologies are used.

Figure 2. The three phases of drone's detection.

3.1. Features Extraction from Radar Signals

Radio detection and ranging (Radar) is an active sensory technology which transmits a burst of electromagnetic waves of certain frequency and receives their echoes to detects surrounding objects and estimate their locations and speed. Most currently used radars operate on short range (S band) and long range (L band). These bands are relatively low frequency bands [7], and thus, the resulted sensing resolution is relatively low for detecting small drones which are characterized by small radar cross section (RCS). Therefore, anti-drone radar systems must operate on higher frequency bands such as X band (8-12 GHz) to effectively detect small size objects [8]. There are two important broad features for drones, shape and size, and pattern of movement. Conventional radar analyzes the reflected echo signal to create the detected object RCS profile as shown in Figure 3. RCS represents the signature of the drone shape and size. The RCS profile is passed to a classifier to evaluate whether the created RCS profile matches a drone

RCS predefined profile [4]. However, there are various objects that feature similar shape and size as the drones, and thus, false alarms may increase.

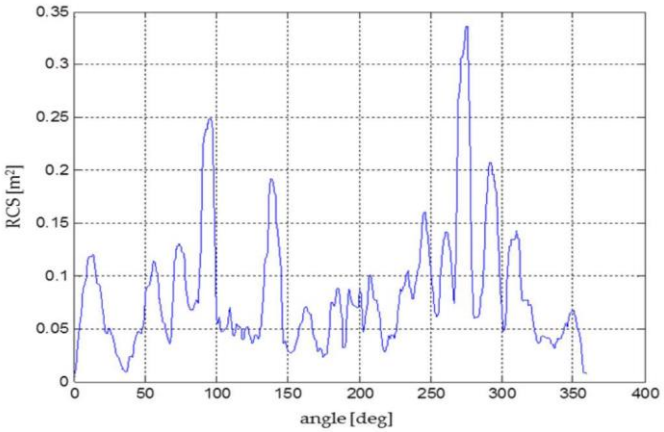

Figure 3. RCS profile of a quadcopter drone [4].

Movement patterns of drones can be derived through Doppler signature (DS) analysis [9]–[12]. Radar systems process echo signals that exhibit frequency shifts caused by the target's motion, forming what is termed the Doppler signature. Drones, with their distinct rotational and translational movements, generate unique Doppler signatures that enable their identification and classification.

Special class of drone Doppler signature is the micro- Doppler signature (m-DS) which captures drone small parts movements such as the blades, wings and propellers [10]– [16]. In these references, the m-DS has been proven to be exceptionally functional for detecting drones, especially, the flapping-wing and rotary-wing drones.

There are mainly two types of radars: continuous wave (CW) radar and pulse radar. The CW radar, which is more effective in terms of bandwidth, power and cost, can further be divided into two sub-categories, namely: unmodulated CW radar and frequency modulated CW (FMCW) radar.

Considering FMCW, the transmitted RF signal out of the radar is of the following form:

$$s(t) = A\cos\left(2\pi\left(f_c + \frac{B}{2T}t\right)t\right), \qquad (1)$$

where A is the signal amplitude, B is the range of the chirp and T is the duration of one chirp. While the carrier frequency is f_c, the operating frequency is $f_c + \frac{B}{2T}t$ which is a linear function of time[1]. The reflected signal (the echo) due to the existence of some target in radar sight is:

$$r_{rf}(t) = A_R\cos\left(2\pi\left(f_c + \frac{B}{2T}(t-\tau)\right)(t-\tau)\right) + n(t) \qquad (2)$$

[1]Hence, FMCW signal is also known as linear FMCW (LFMCW)

where A_R is the amplitude of the reflected RF signal, τ is the round-trip time and $n(t)$ is the noise component. The distance d between the radar and the target can be found from the round-trip time: $d = \frac{c\tau}{2}$, where c is the speed of light. For the signal to be digitally processed, and hence the features to be extracted, the reflected RF signal should be first converted into a complex baseband signal. This can be done by passing the RF signal through an I/Q demodulation and then a low pass filter $\mathcal{LF}\{\,.\,\}$. Thus, the complex baseband signal is characterized by equation (3).

$$r_{bb}(t) = \mathcal{LF}\{r_{rf}(t)\, cos(2\pi(f_c + \tfrac{B}{2T}t)t)\} + j\mathcal{LF}\{r_{rf}(t)\, sin(2\pi(f_c + \tfrac{B}{2T}t)t)\}. \quad (3)$$

Converting the desired signal component in (3) from Cartesian coordinate to polar coordinate yields,

$$r_{bb}(t) = A_r e^{j\theta(t)} + n_I(t) + jn_Q(t) \qquad (4)$$

where A_r and $\theta(t)$ are the magnitude and phase of the complex baseband signal. $n_I(t)$ and $n_Q(t)$ are the in-phase and quadrature components of the low frequency equivalent noise.

A crucial part of the feature extraction is an efficient representation of the considered signal (i.e., $r_{bb}(t)$). In the following, we list the most popular signal representation.

- **Short Time Fourier Transform (STFT)** applies Fourier transform consecutively on short portions of $r_{bb}(t)$. These short portions are defined by a sliding windowing function $\omega(t)$. Mathematically, STFT can be found by:

$$R_{bb}(f, \xi) = \int_{-\infty}^{\infty} r_{bb}(t)\omega(t - \xi)e^{-j2\pi ft}dt \qquad (5)$$

  where $R_{bb}(f, \xi)$ is the complex valued STFT as a function of f and ξ which are Doppler frequency and time, respec- tively. Clearly, STFT is the traditional Fourier transform of the product $r_{bb}(t)\omega(t - \xi)$, and as the window function $\omega(t)$ slides by ξ the Fourier transform changes. Hence, STFT is a complex function (magnitude and phase) of both Doppler frequency and time. It essentially describes the dynamic of the spectra of a signal. STFT is the most successful and popular radar signal analysis techniques [10]. Further, many other techniques are actually based on STFT as will be shown in the sequel.

  Note that the Doppler frequency directly defines the velocity v of the target via: $v\cos(\theta) = \frac{\lambda f}{2}$, where λ is the wavelength of the radar signal, and θ is the aspect angle between the target direction and the radar line-of-sight [17]. Therefore, STFT can be displayed as a function of v and ξ.

- **Spectrogram (SG)** is the squired magnitude of the STFT. In other words, Spectrogram is the power spectral density of $r_{bb}(t)$ over time-frequency grid. Mathematically,

$$SG_{bb}(f, \xi) = |R_{bb}(f, \xi)|^2 \qquad (6)$$

  Figure 4 shows an example of a quadcopter spectrogram.

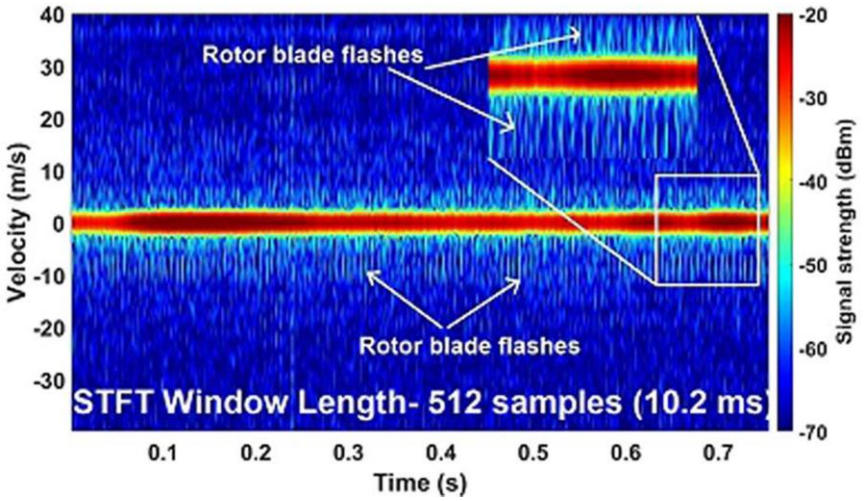

Figure 4. Spectrogram of a quadcopter at 45 m range detected by a 94 GHz radar [13]. The fast repetitive spikes are due to drone blade flashing which is lacked by other flying objects such as birds. Hence, these spikes represents a distinctive characteristic for drones.

- **Cadence Velocity Diagram (CVD)** is the magnitude of the Fourier transform of the STFT magnitude ($|R_{bb}(f,\xi)|$), with respect to time ξ. Mathematically,

$$CVD_{bb}(f,f_k) = |\mathcal{F}\{|R_{bb}(f,\xi)|\}|, \qquad (7)$$

where f_k is called cadence frequency. CVD describes the repetition rate of different velocities. In other words, It is a metric of how often different velocities repeat over time. Moreover, it characterizes the size and frequency of the STFT components which carry information about the moving parts of the target. It can even detect a swarm of drones [18].

- **Cepstrogram (CG)**applies cepstral analysis consecutively on short portions of $r_{bb}(t)$ [19]. These short portions are defined by a sliding windowing function $\omega(t)$.

In the following few lines, we explain the cepstral analysis which is an important area of signal processing. Cepstral analysis (a variant of spectral analysis) of the signal $r_{bb}(t)$ is defined as the inverse Fourier transform taken over the natural logarithm of the signal magnitude spectrum. Mathematically, the real cepstrum[2]$\mathrm{C}(q)$ can be found by:

$$C(q) = \mathcal{F}^{-1}\{\ln(|R_{bb}(f)|)\}, \qquad (8)$$

where the independent variable q is called quefrency (lag time) measured in second. $R_{bb}(f)$ is the complex spectrum of the signal $r_{bb}(t)$, and $\mathcal{F}^{-1}\{\cdot\}$ is the inverse Fourier transform which can be computationally performed using inverse fast Fourier transform (iFFT). Cepstral is a tool used to identify periodic structures within a signal's spectrum. Specifically, it isolates periodic patterns in the spectral magnitude, such as harmonic frequencies. For instance, while the spectrum $R_{bb}(f)$reveals peaks at harmonic frequencies of a fundamental frequency, the cepstrum transforms these harmonic spectral peaks into a single distinct peak at a corresponding quefrency. This property makes cepstral analysis particularly effective for

[2]In fact, there are three related cepstrums: real cepstrum which is shown in equation (8), power cepstrum $C_p(q) = 4C^2(q)$, and complex cepstrum $C_c(q) = \mathcal{F}^{-1}\{\ln(R_{bb}(f))\}$ [19].

detecting signal echoes and distinguishing multiple overlapping targets. Cepstral can be used to determine the micro-Doppler periodicity, which corresponds to the angular velocity of the propellers or rotors. In good conditions, it is proven to be valuable in estimating the number of rotors and their individual angular velocity [12].

Since CG is a result of applying cepstral analysis on short portions of the signal using widowing function, CG appears as a function of quefrency q and time ξ which are both measured in seconds. Mathematically, cepstrogram can be found by:

$$CG_{bb}(q,\xi) = \mathcal{F}^{-1}\{\ln(|R_{bb}(f,\xi)|)\}, \qquad (9)$$

where $R_{bb}(f,\xi)$ is the STFT, and the inverse Fourier transform is taken with respect to the Doppler frequency f. Figure 5 shows an example of a cepstrogram.

- **Wigner-Ville Distribution (WVD)**is a quadratic time-frequency representation. WVD is used to analyse non-stationary signals to provide a high-resolution joint time-frequency energy density of a signalmaking it particularly useful for studying transient or rapidly varying signals.

Unlike the previous signal analysis,WVD does not depend on STFT. In fact, WVD is exploited to resolve the problem of low resolution that STFT suffers from [18]. WVD of the signal $r_{bb}(t)$ is defined as the Fourier transform of the product: $r_{bb}\left(t+\frac{s}{2}\right)\cdot r_{bb}\left(t-\frac{s}{2}\right)$, with respect to the shifting variable s, that is:

$$W_r(t,f) = \int_{-\infty}^{\infty} r_{bb}\left(t+\frac{s}{2}\right)\cdot r_{bb}\left(t-\frac{s}{2}\right)e^{-j2\pi fs}ds. \qquad (10)$$

WVD can be easily computed by applying the fast Fourier transform (FFT). WVD possesses several interesting properties and has a very close connection with the ambiguity function [20].

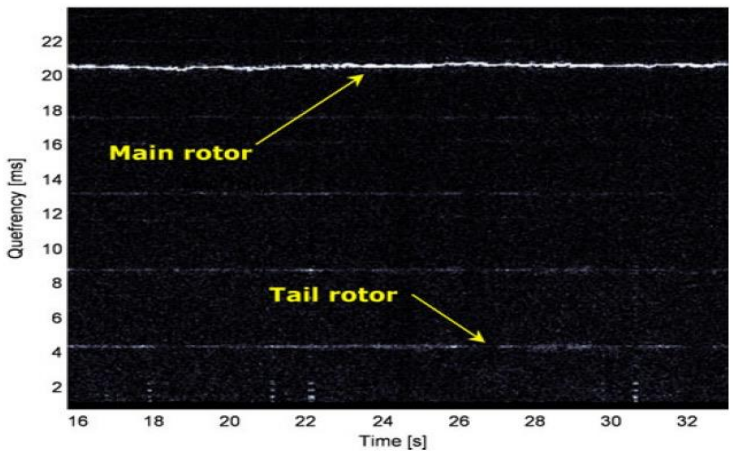

Figure 5. Cepstrogram of a MAV-size helicopter [12].

- **Empirical Mode Decomposition (EMD)** decomposes a signal into intrinsic mode functions (IMF) and a residue [21]. Assuming $r_{bb}(t) \in L_p[0,T]$[3], it can be decomposed using EMD such that

$$r_{bb}(t) = \sum_{j=1}^{J} m_j(t) + q(t), \qquad (11)$$

where $m_j(t) \in L_p[0,T]$ is the j^{th} intrinsic mode function, and $q_j(t) \in L_p[0,T]$ is a residue. The basic idea of EMD is considering the signal as slow oscillations superimposed by fast oscillations. Clearly, EMD breaks down the signal without leaving the time domain. Figure 6 shows the first four IMFs of a fixed-wing drone echo signal. The decomposition is according to the time-scale of the oscillations. The first IMFs contain the highest oscillating components while the last IMFs have the lowest frequency content. Every IMF must satisfy two conditions. First, the average of the envelop must be zero. Second, the number of zero-crossings differs from the number of local extrema by at most one. Further, IMFs possess orthogonality feature, that's, they are mutually orthogonal.

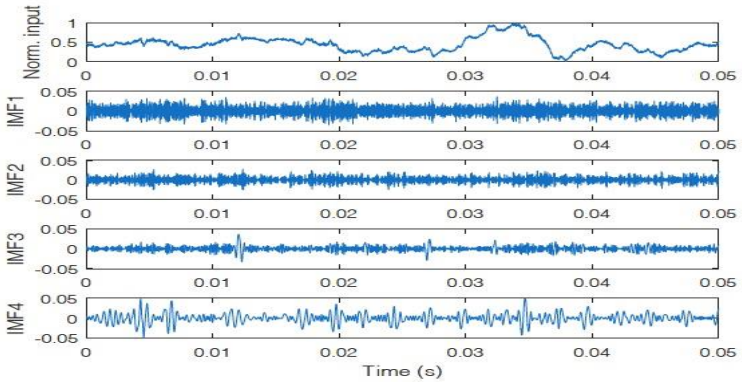

Figure 6. A micro-Doppler radar echo and its first four IMFs due to a fixed- wing drone [15].

There are many other techniques for representing radar signals such as weighted spectrum [22], Malvar wavelets, the S-transform and various types of wavelet transform [23]. Once the considered signal is well represented, features such as height, max. height, radial velocity, spectrogram frequency profile (SFP), CVD frequency profile, Cepstrum coefficients, spectral correlation function and many others can be extracted and fed to a standard classifier such as SVM, Decision tree, K-nearest neighbour (KNN), naive Bayes (NB), linear discriminant analysis (LDA) and many others.

3.2. Features Extraction from Radio Frequency (RF) Signals

Remotely controlled drones usually perform a two-way RF communication with the ground station in order to exchange control and surveillance information. This wireless signalling occurs at least 30 times per second [24], and has special pattern called RF-signature (or RF-fingerprint).

[3] $L_p[0,T]$ denotes the set of complex-valued signals defined on the interval $t \in [0,T]$such that $\int_0^T \|x(t)\|^p dt < \infty$, and equipped with the metric d_p, where$1 \leq p \leq \infty$.

RF analysers detect drones RF-signature by processing the electromagnetic (EM) emissions in the protected zone. In other words, they receive RF signals available in the surrounding space and process them aiming at detecting any suspicious signalling activity, which may represent a communication between a drone and its ground station [25].

The commercial drones have special signalling schemes (i.e., protocol signature) which is distinct from other types of communications over the same frequency band. Therefore, the RF analyser algorithms compare the captured signalling with a library of predefined signals. Once a certain level of matching is attained, the ADS declares danger.
The most descriptive characteristics of drone communications which can be processed to generate useful features are as follows:

- **Packet sizes**transmitted from the remote controller (RC) to the drone and from the drone to the RC [26] and their means.

- **Packets Inter-arrival time**measured in both links. These two features require identifying with high precision the start and end points of packets.

- **Start points of drone transmissions**based on calculated threshold τ using expectation maximization (EM) algorithm [27].

- **RF hash fingerprint**is generated by extracting attributes of RF signal preamble waveforms [28]. The fundamental attributes are the distance between adjacent peaks of the preamble signal and their respective locations. By comparing the mean of the distances with each (weighted) distance, hash fingerprints are generated according to the following rule:

$$h_i = \begin{cases} 1 \; if \bar{d} < wd_i \\ -1 \; if \bar{d} > wd_i \\ 0 \; for \bar{d} = wd_i \end{cases} \qquad (12)$$

  where h_i denotes the RF hash fingerprint, $\bar{d}$ is the average distances between the adjacent peaks and w is some weight.

- **Magnitude Spectrum**$X[m]$ of the raw RF signal $x[n]$ which is calculated using the discrete Fourier transform: $X[m] = \left|\sum_{n=0}^{N-1} x[n] e^{-j2\pi \frac{mn}{N}}\right|$, where N is the total number of time samples of $x[n]$, and M is the total number of frequency bins in $X[m]$. This kind of features should be taken in segments such that each segment consists of N samples.

The first four features of the above list are categorized as time-domain techniques which are mostly rely on the existence of an abrupt change at the start point of the signal. Below, we list several discriminative energy-time-frequency based features. In this category, the raw RF time-domain signal is first transformed into the energy-time-frequency domain using spectrogram. Then, the energy trajectory, which is a function of time, is computed from the spectrogram[4]. After that, the energy transient[5] $f_E(n)$ is estimated by searching for the most abrupt change in the mean or variance of the normalized energy trajectory. Finally, a set of statistical features are extracted from the energy transient [25].

[4]Samples values of the energy trajectory function is computed by taking the maximum value across all frequencies in the spectrogram.

[5]The energy transient defines the transient characteristics of the signal in energy domain.

- **Skewness**γ is a metric of the asymmetry of the energy distribution around the value of the mean μ:

$$\gamma = \frac{1}{N\sigma^3}\sum_{n=1}^{N}(f_E(n) - \mu)^3 \qquad (13)$$

- **Variance**σ^2 is a metric of the spread of the energy distribution around the value of the mean value:

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(f_E(n) - \mu)^2 \qquad (14)$$

- **Entropy** H is a metric of Shannon entropy which is a measure of uncertainty:

$$H = \sum_{n=1}^{N} f_E(n)\log_2 f_E(n) \qquad (15)$$

- **Kurtosis**k is a metric of the flatness or sharpness of the energy transient:

$$k = \frac{1}{N\sigma^4}\sum_{n=1}^{N}(f_E(n) - \mu)^4 \qquad (16)$$

Thus, the above RF fingerprint features can be used to train and test a machine learning classifier such as kNN, LDA, SVM which can be used to classify any new RF communication observed in the protected zone.

3.3. Features Extraction from Acoustics Signals

Drones often produce remarkable sound waves due to its oscillating objects such as propellers and engines. Acoustics sensing technology is a set of microphones that are installed in a carefully selected points to detect the sound waves produced by drones. Any captured sound is processed, then its features are extracted and compared to a library of drone acoustic signatures [29]–[33].

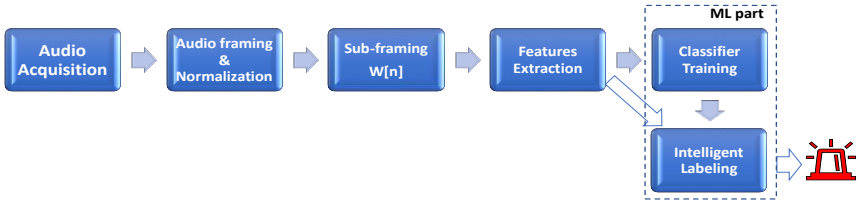

Figure 7. A block diagram of detection by acoustics-based technology.

Figure 7 illustrates the detection process using acoustics signals. The process starts with audio acquisition which includes picking the sounds from the surrounding environment followed by analogue to digital conversion. Then, the collected digitized signal is broken into a sequence of normalized frames[6] $x[n]$ each of 5 seconds duration. Each frame $x[n]$ is further broken into sub-frames $u[n]$ using a moving Hamming window $w[n]$ of length L samples with overlapping shifts of s samples ($s < L$). Thus, the l^{th} sub-frame is found by:

$$u_l[n] = x[n] \cdot w[n - ls], \qquad (17)$$

where

[6] These frames are usually normalized in the range $[-1, ,1]$.

$$w[n] = 0.54 - 0.46\cos(\frac{2\pi n}{L-1}), \quad 0 \leq n < L \qquad (18)$$

In the sequel, we describe the most popular drone acoustic features, and how they are extracted from a typical sub-frame $u[n]$.

- **Temporal Centroid** is the balancing point of the signal amplitude over time. In other words, Temporal centroid is the weighted mean of the samples' indices, with their samples values as the weights. Mathematically,

$$C_t = \frac{\sum_{k=1}^{L-1} k \cdot u[k]}{\sum_{k=1}^{L-1} u[k]} \qquad (19)$$

- **Spectral Centroid (CS)** is the balancing point of the signal's spectrum U(f). CS enables to specify whether a given frequency is a higher or lower frequency with respect to $u[n]$. Since CS is a good predictor of the brightness of a sound, it serves as an indicator of drone's existence as most drones produce similar sound brightness. CS can mathematically be found by,

$$C_s = \frac{\sum_{m=0}^{L-1} f(m) \cdot U[m]}{\sum_{m=0}^{L-1} U[m]} \qquad (20)$$

where $f(m)$ is the centre frequency of the m^{th} bin in the frequency domain, and $U[m] = \left|\sum_{k=0}^{L-1} u[k] \cdot e^{-j2\pi m\frac{k}{L}}\right|$, that is the DFT magnitude of $u[n]$.

- **Zero-Crossing rate (ZCR)** is the average number of times where the signal changes sign within a given time window. Mathematically,

$$ZCR = \frac{1}{(L-1)}\sum_{k=1}^{L-1} \frac{|sgn(u[k]) - sgn(u[k-1])|}{2}, \qquad (21)$$

where

$$sgn(u[n]) = \begin{cases} -1 & for u[n] < 0 \\ +1 & for u[n] > 0 \\ sgn(u[n-1]) & for u[n] = 0 \end{cases}.$$

ZCR is a very useful feature in identifying voiced audio.

- **Short Time Energy** measures the energy variations of the environmental sound over time. It is computed as follows.

$$E_u = \frac{1}{L}\sum_{k=0}^{L-1} |u[k]|^2 \qquad (22)$$

- **Spectral roll-off (SRO)** is the highest frequency below which a certain fraction β of the total energy resides. SRO can be found by solving the following equation,

$$\sum_{m=0}^{SRO} |U[m]|^2 = \beta \sum_{m=0}^{L-1} |U[m]|^2 \qquad (23)$$

- **Linear predictive coding (LPC)** is a signal analysis which provides coefficients that carry the characteristics of the audio sub-frame $u[n]$. The idea of LPC is that the current sample of the audio sub-frame can be estimated by a linear combination of p previous samples. That is,

$$u[n] \approx \sum_{i=1}^{p} \alpha_i u[n-i] \qquad (24)$$

where $\{\alpha_i\}_{i=1}^{p}$ is the set of LPC coefficients. The values of LPC coefficients can be determined by minimizing the mean-squired error (MMSE) over one sub-frame. In general, the result of MMSE[7] is

$$Ra = r \Rightarrow a = R^{-1}r \qquad (25)$$

where a is a vector, whose elements are LPC coefficients. The vector $r = [r(1)\, r(2) \ldots r(p)]$, where $r(i) = \sum_{k=0}^{L-1-i} u[k]u[k+i]$ is the sub-frame autocorrelation of delay i. In addition, $R$is a p x p Toeplitz and symmetric matrix[8] which can be formed by the vector $[r(0)\, r(1) \ldots r(p-1)]$.

- **Linear Predictive Cepstral Coefficients (LPCC)** is a very useful techniques for estimating the parameters of a sound signal such as its pitch. LPCC are computed from the linear predictive coding coefficients $\{\alpha_i\}_{i=1}^{p}$ as follows:

$$C_q = \sum_{i=1}^{q-1} {}^{i}_{q}{-}C_i \alpha_{q-i} + \begin{cases} \alpha_q for\ 1 \le q \le p \\ 0 \quad for q > p \end{cases}, \qquad (26)$$

where C_q is the cepstral coefficients. LPCC is considered a time-domain approach for extracting the cepstral coefficients which describe the overall shape of the spectral envelop.

- **Mel-frequency Cepstrum Coefficients (MFCC)** is the discrete cosine transform (DCT) of the logarithm of the Mel-scaled spectrum of the signal $u[n]$. Unlike LPCC, MFCC is a frequency-domain approach for extracting the cepstral coefficients with a major difference that the frequency axis is converted into Mel-scale[9], and the magnitude spectrum is converted into Mel-spectrum. Thus, converting the frequency axis into Mel-scale is performed by the following relation [36],

$$f_{mel} = 2595 \log_{10}(1 + \frac{f}{700}) \qquad (27)$$

where f_{mel} is the frequency in Mel scale. Then, the Mel-spectrum is obtained by multiplying the magnitude spectrum by a bank of Ψ triangular Mel weighing filters whose outputs are $Y(\psi)$. Such a process is a spectrum smoothing where more perceptually meaningful frequencies are emphasized while the less meaningful frequencies are wrapped up into a small number of Melfrequency bins. The outputs of the bank of filters are then taken into logarithmic scale. Finally, The MFCC can be found via DCT as follows:

$$c_j = \sum_{\psi=1}^{\Psi} \log_{10}(Y(\psi)) \cos\left(\frac{\pi j(\psi-0.5)}{\Psi}\right) \text{for } j = 1,2,\cdots,J \qquad (28)$$

where c_j represents the j^{th} MFCC coefficient.

---

[7] For solving this particular MMSE problem, there are two equivalent methods: the autocorrelation and covariance methods [34]. The result in (25) is based on the autocorrelation method.

[8] Due to this special structure of R, (25) can be solved with a complexity of $\mathcal{O}(p^2)$using Levinson-Durbin's recursive method instead of using the traditional Gaussian elimination method whose complexity is $\mathcal{O}(p^3)$.

[9] The relation of Mel-scale to Hz-scale is linear at a more perceptually meaningful bands and logarithmic at less perceptually important bands. For example, in modeling human auditory, mel-scale is linear at low frequencies (below 1kHz) and logarithmic at higher frequencies (above 1kHz) because it has been found that for speech, the higher frequencies are perceptually less important than the lower frequencies [35].

There are many other acoustic features such as Gammtone Cepstral coefficients [37], slop of the frequency spectrum [30] and harmonic features [31] which can be extracted and fed to a suitable ML classifier to detect rogue drones. Further, time-frequency representations of the acoustic signal such as STFT can be used as well to train deep neural networks such as CNN or RNN [29], [38]. Authors in [32] investigated the effectiveness of different popular deep learning models. They compared the performance of CNN, RNN and Gaussian Mixture Model (GMM) when all are fed with MFCC and mel-spectrogram of drone's sound as the input feature vector. They found that RNN model demonstrated the best performance with CNN and GMM where it recorded the best F-Score of 0.8009 and at a shortest processing time.

3.4. Features Extraction from Electro-Optic (EO) Sensors

EO technology employs terahertz frequencies in order to detect drones [7], [39]. Therefore, it requires line of sight link between the sensory element and the target. There are three categories of EO technology: visible light optics, infrared thermal imaging and laser detection and ranging (LADAR). The first category uses high-definition cameras to detect the visible light reflected from the rogue drones whereas the infrared thermal imaging uses the infrared band to gauge the heat differences in the protected sky (Figure 8). Then, specific algorithms run for spotting a drone image or heat differences caused by a drone. The LADAR, on the other hand, illuminates the protected space with a laser light, then collects the reflected light using optical sensors to produce an image of the protected zone. Unlike the first two technologies, LADAR is an active technology, thus, it can provide precise images and distance measurements over relatively long ranges.

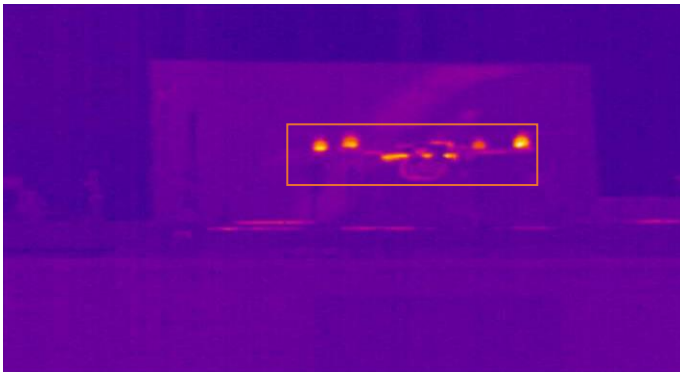

Figure 8. Quadcopter thermal image using Infrared technique [4].

It is noteworthy that the signals captured by EO technologies take a form of two- or three-dimensional signals (i.e., images or videos). Accordingly, image processing, computer vision and/or pattern recognition techniques must be employed during the signal processing and feature extraction stage. Generally, there are two main approaches for extracting targets' features out of the EO signals.

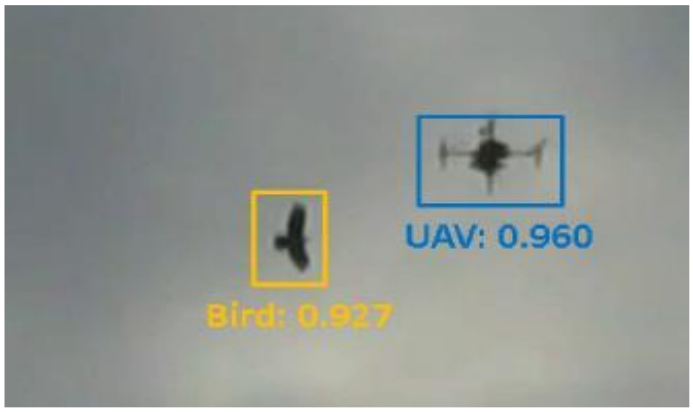

Figure 9.Image keypoint-based feature extraction approach [40]

- **Keypoint-based approach** is a class of methods that produce object bounding boxes by detecting and grouping their keypoints. In this method, salient points (keypoints) are detected since they are likely to represent important content information (Figure 9). Examples of this approach are as follows:

    - Median Background Subtraction:

      Images are first gone through median background subtraction (MBS) specially if the cameras are static [40], [41]. In the design stage, images are taken for the zone to be protected. These images are called background images $I_{bg}(x, y)$. During the system operation, the difference image between the captured image and the background image is computed which results in foreground image (flying objects). Thus,

      $$I_{fg}(x, y) = \left|I(x, y) - I_{bg}(x, y)\right| \qquad (29)$$

      where $I(x, y)$ is the value of the pixel (x, y) at the captured image, and I_{fg} is the resulted foreground image. Then, a threshold value is used to discriminate between pixels in the background image and foreground image. The threshold value must be chosen carefully since too high value results in miss detection while very low value results in a high false alarm rate. Finally, the foreground is gone through special image processing modules such as connected component analysis for pixels clustering and then sent to a ML algorithm for classification. The learning phase requires enough image data without drones to build up an appropriate background reference with which the images taken during the operational phase can be compared.

    - Region Proposal Network:

      In case of moving cameras, the regional proposal network (RPN) or some variant of it is used for detecting rogue drones. RPN was first proposed by Ren et al. in [42]. It produces a set of potential regions that are likely to contain a target. Then, a faster regional-based CNN deep learning algorithm is applied as described in [43].

- **Hierarchical and cascaded approach** is based on the fact that images consist of levels of visual features such as edges, gradients, corners, colors...etc. In this approach, low-level features (such as edges) are computed at low level of processing hierarchy. As the process moves up to a higher level, features of lower levels are combined forming higher-order

features such as corners. Gradually, this process will converge to correspond to small parts and to objects. Examples of this approach are Histogram of Gradient (HOG) [44] and Local Binary Pattern (LBP) [44].

Nevertheless, images or video clips may be processed directly by computer vision and DL algorithms specially CNN where features are implicitly extracted by the neural network itself. However, this comes at the expense of more computational power and longer delay.

4. CONCLUSIONS

This paper investigates the detection of low-flying, small, and slow (LSS) drones using four key different modalities: radar, acoustic sensors, electromagnetic (EM) emissions, and electro-optical (EO) systems. For each modality, we presented recent advancements in signal processing techniques aimed at extracting distinctive features from drone-reflected or emitted signals. While each technology exhibits unique strengths and limitations, our analysis suggests that strategic integration of complementary detection systems could enhance detection performance.

There are many fundamental challenges which ought to be addressed by the research community. For example, identifying the most information-carrying features and how much distinctive information each feature carries play a central role in the detection success. In addition, designers need to know the fundamental design limit. In other words, they need to know the minimum number of features required to achieve a certain detection performance level. Also, the features selection strategy is better to be studied. Given set of features, an analytical method that identify the best suited classifier is required. Furthermore, for a catalogue of features required by a given classifier, what is the best time-frequency representation of the collected signal (e.g., the m-DS) for extracting the required features? Finally, there must be some trade-off between the detection performance metrics such as the trade-off between the sensitivity and specificity. It's important to identify a strategy for striking a balance between these metrics. This may depend on the type of the protected zone which may lead to establishing zones classification system.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Anas Hashmi for thorough review and valuable suggestions.

REFERENCES

[1] A. Hashmi, A novel drone-based search and rescue system using bluetooth low energy technology," *Engineering, Technology & Applied Science Research*, vol. 11, pp. 7018–7022, Apr, 2021.

[2] A. Sazaly, M. Mohd Ariff, and A. F. Razali, "3d indoor crime scene reconstruction from micro uav photogrammetry technique," *Engineering, Technology & Applied Science Research*, vol. 13, pp. 12020–12025, Dec, 2023.

[3] A. AlZahrani, "Unauthorized drones: Classifications, detection and neutralization techniques," *in the 11th International Conference on Systems and Control (ICSC)*, Tunisia, Dec. 2023

[4] J. Farlik, M. Kratky, J. Casar, and V. Stary, "Multispectral detection of commercial unmanned aerial vehicles," *Sensors*, vol. 19, no. 7, 2019.

[5] S. Bjo¨rklund, H. Petersson, and G. Hendeby, "Features for micro-doppler based activity classification," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1181–1187, 2015.

[6] F. Alotaibi, A. Al-Dhaqm, and Y. D. Al-Otaibi, "A conceptual digital forensic investigation model applicable to the drone forensics field," *Engineering, Technology & Applied Science Research*, vol. 13, pp. 11608–11615, Oct. 2023.

[7] G. Birch, J. Griffin, and M. Erdman, "Uas detection, classification, and neutralization: Market survey," *technical report, Sandia National Laboratories*, 2015.

[8] J. Patel, F. Fioranelli, and D. Anderson, "Review of radar classification and rcs characterisation techniques for small uavs or drones," *IET Radar, Sonar & Navigation*, vol. 12, no. 9, pp. 911–919, 2018.

[9] V. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-doppler effect in radar: phenomenon, model, and simulation study," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 2–21, 2006.

[10] A. Coluccia, G. Parisi, and A. Fascista, "Detection and classification of multirotor drones in radar sensor networks: A review," *Sensors*, vol. 20, no. 15, 2020.

[11] R. Harmanny, J. de Wit, and G. Cabic, "Radar micro-doppler feature extraction using the spectrogram and the cepstrogram," in *11th European Radar Conference*, pp. 165–168, 2014.

[12] R. Harmanny, J. de Wit, and G. Cabic, "Radar micro-doppler miniuav classification using spectrograms and cepstrograms," *International Journal of Microwave and Wireless Technologies*, vol. 7, no. 3-4, pp. 469–477, 2015.

[13] S. Rahman and D. Robertson, "Radar micro-doppler signatures of drones and birds at k-band and w-band," *Scientific Reports*, vol. 8, pp. 17396, Nov, 2018.

[14] B. Oh, X. Guo, F. Wan, K. Toh, and Z. Lin, "An emd-based micro- doppler signature analysis for mini-uav blade flash reconstruction," in *2017 22nd International Conference on Digital Signal Processing (DSP)*, 2017, pp. 1–5.

[15] B. Oh, X. Guo, F. Wan, K. Toh, and Z. Lin, "Micro-doppler mini-uav classification using empirical-mode decomposition features," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp. 227–231, 2018.

[16] B. Oh, X. Guo, and Z. Lin, "A uav classification system based on fmcw radar micro-doppler signature analysis," *Expert Systems with Applications*, vol. 132, pp. 239–255, 2019.

[17] M. Richard, *Fundamentals Radar Signal Processing*, 2nd ed. New York, U.S.: *McGraw-Hill Education*, 2014.

[18] B. Taha and A. Shoufan, "Machine learning-based drone detection and classification: State-of-the-art in research," *IEEE Access*, vol. 7, pp. 138669–138682, 2019.

[19] J. Benesty, M. Sondhi, and Y. Huang, "Handbook of Speech Processing", *Springer*, 2008.

[20] B. Torre sani, "Time-frequency and time-scale analysis," *Signal Process- ing for Multimedia*, pp. 55–70, Feb, 1999.

[21] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, Royal Society of London: London, UK, 1998.

[22] J. Ge´rard, J. Tomasik, C. Morisseau, A. Rimmel, and G. Vieillard, "Micro-doppler signal representation for drone classification by deep learning," in *28th European Signal Processing Conference (EU- SIPCO)*, 2021, pp. 1561–1565.

[23] L. Samantaray, M. Dash, and R. Panda, "A review on time-frequency, time-scale and scale-frequency domain signal analysis," *IETE Journal of Research*, vol. 51, no. 4, pp. 287–293, 2005.

[24] P. Nguyen, M. Ravindranatha, A. Nguyen, R. Han, and T. Vu, "In- vestigating cost-effective rf-based detection of drones," in *Proceedings of the 2nd Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use*, DroNet '16, (New York, USA), 2016, pp. 17-22.

[25] M. Ezuma, F. Erden, C. K. Anjinappa, O. Ozdemir, and I. Guvenc, "Micro-uav detection and classification from rf fingerprints using ma- chine learning techniques," in *2019 IEEE Aerospace Conference*, 2019, pp. 1–13.

[26] A. Alipour-Fanid, M. Dabaghchian, N. Wang, P. Wang, L. Zhao, and K. Zeng, "Machine learning-based delay-aware uav detection and operation mode identification over encrypted wi-fi traffic," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2346–2360, 2020.

[27] C. Zhao, M. Shi, Z. Cai, and C. Chen, "Detection of unmanned aerial vehicle signal based on gaussian mixture model," in *2017 12th International Conference on Computer Science and Education (ICCSE)*, 2017, pp. 289–293.

[28] Z. Shi, M. Huang, C. Zhao, L. Huang, X. Du, and Y. Zhao, "Detection of lssuav using hash fingerprint based svdd," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–5.

[29] J. Busset, F. Perrodin, P. Wellig, B. Ott, K. Heutschi, T. Ruhl, and T. Nussbaumer, "Detection and tracking of drones using advanced acoustic cameras," in *The SPIE, Unmanned/Unattended Sensors and Sensor Networks XI; and Advanced Free-Space Optical Communication Techniques and Applications*, vol. 9647, 2015, pp. 53–60.

[30] L. Hauzenberger and E. Ohlsson, "Drone detection using audio analysis," M.S. thesis, Lund Univ., Lund, Sweden, 2015.

[31] B. Harvey and S. O'Young, "Acoustic detection of a fixed-wing uav," *Drones*, vol. 2, no. 1, 2018.
[32] S. Jeon, J. Shin, Y. Lee, W. Kim, Y. Kwon, and H. Yang, "Empirical study of drone sound detection in real-life environment with deep neural networks," in *25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 1858–1862.
[33] Z. Kaleem and M. H. 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, 2018.
[34] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, 2nd ed. Prentice Hall, 2000.
[35] B. Logan, "Mel frequency cepstral coefficients for music modeling," in *In International Symposium on Music Information Retrieval (ISMIR)*, 2000.
[36] M. Z. Anwar, Z. Kaleem, and A. Jamalipour, "Machine learning inspired sound-based amateur drone detection for public safety applications," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2526– 2534, 2019.
[37] S. Salman, J. Mir, M. T. Farooq, A. N. Malik, and R. Haleemdeen, "Machine learning inspired efficient audio drone detection using acoustic features," in *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*, 2021, pp. 335–339.
[38] Y. Seo, B. Jang, and S. Im, "Drone detection using convolutional neural networks with acoustic stft features," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2018, pp. 1–6.
[39] D. Mototolea, "A study on the methods and technologies used for detection, localization, and tracking of LSSUASs," *Journal of Military Technology*, Dec. 2018.
[40] A. Schumann, L. Sommer, J. Klatte, T. Schuchert, and J. Beyerer, "Deep cross-domain flying object classification for robust uav detection," in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2017, pp. 1–6.
[41] L. Sommer, A. Schumann, T. Mu ̈ller, T. Schuchert, and J. Beyerer, "Flying object detection for automatic uav recognition," in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2017, pp. 1–6.
[42] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real- time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, vol. 28, 2015.
[43] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137– 1149, 2017.
[44] F. Gokce, G. Ucoluk, E. Sahin, and S. Kalkan, "Vision-based detection and distance estimation of micro unmanned aerial vehicles," *Sensors*, vol. 15, no. 9, pp. 23805–23846, 2015.

AUTHOR

Ali Y. Al-Zahrani received his B.S. in Electrical Engineering with honour from King Fahad University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 2002. He Further received his M.S. and PhD in Electrical and Computer Engineering from Carleton University, Ottawa, Canada in 2010 and 2015, respectively. Ali is currently an associate professor in the Department of Electrical and Electronic Engineering at the University of Jeddah, Saudi Arabia. He worked as an engineer at the Saudi Basic Industries Corporation (SABIC) from 2002 to 2007. His research interests include RF sensing, Open RAN, radio resource allocation, massive MIMO and interference management in wireless communication systems. He published several papers in these topics.