Drone Detection based on An Audio-assisted Camera Array

Hao Liu[1] Zhiqiang Wei[1] Yitong Chen[2,*] Jie Pan[1] Le Lin[1] Yunfang Ren[1]
[1]Department of Computer Science, [2]Law & Politics School
Ocean University of China
Qingdao, China
liu.hao@ouc.edu.cn, weizhiqiangouc@126.com, chenyitong@outlook.com

***Abstract*—In recent years, small, inexpensive UAV - also known as drone - has made great progress, which brings privacy and security issues. Drone detection is one of the important methods to solve these problems. The main challenges of drone detection are: (1) A drone is very confusing with other flying objects such as a bird; (2) the low-flying UAV occlusion happens very frequent, and (3) the existing surveillance coverage is limited. Camera array can be used for large-scale airspace observation. In this paper, we propose a modular camera array system with audio assistance, which consists of dozens of high-definition cameras and multiple microphones, to monitor drones. It can be mounted on a carrier after reducing the size. The system captures the image and audio information of the surroundings in various directions. The system can integrate the information of multiple sensors, and identify the characteristics of the drone to achieve higher efficiency of drone monitoring.**

Keywords- drone detection; camera array; sensor fusion

I. INTRODUCTION

Drones are aircraft that have no onboard human pilot. In recent years, civilian drone market is developing rapidly. In variety types of drone, small inexpensive drones such as quadcopters for the general consumer have been proliferating, which multiplies the potential for benefits from drones, and exacerbates the risks. For public security or privacy considerations, unauthorized drones are prohibited on airports, private fields, and other similar areas. However, due to the growing number of small drones, especially quadcopter-like drones, there is an urgent need to set the low-altitude airspace monitoring equipment in no-fly zones to timely detect unlawful flying drones.

However, despite some commercial products, the technologies and products for drone detection still need to be improved. Traditional military radar is limited in applicability, for the general civilian areas, visual or other passive detection method is more appropriate than active methods. There are some unique challenges for small drone detection:

- The appearance of a drone is difficult to distinguish from that of other flying objects such as birds, especially when a drone is still far away.
- Due to battery and communication limitations, consumer-grade drones typically operate at low altitudes, creating complex and variable backgrounds, while objects such as trees, houses occlude the drone very often.
- Drones may appear in all directions, and monitoring equipment should be able to monitor multiple directions at the same time.

Obviously, because the visual features of drone are more complex, we can improve the accuracy of drone detection by introducing other auxiliary features. For two reasons, we believe that audio is an effective type of feature for drone detection. First of all, the sound monitoring equipment works in a passive way, the intrusion to the environment is acceptable; Secondly, drones high-speed motor in operation will emit significant noise, that can be detected at a relatively long distance.

In this paper, we propose a camera array with audio assistance for drone detection. The device contains multiple cameras and microphones that record videos (image sequences) and sounds in all directions. Feature extraction is performed on images and audio respectively, and then a variety of features are merged to realize drone detection.

- Hardware Design: We present a novel design of a massively multiview capture system consisting of thirty high-definition cameras and three microphones. The system's capturing, transmission and processing components are all mounted on a lightweight frame.
- Drone flight data set: We collected a series of data sets that various drone models in a variety of conditions. Each set of data contains synchronized image sequences, audio, flight trajectories and other information such as weather conditions.
- Drone detection algorithm: We have adopted a drone detection algorithm that applied a fusion of video and audio features. According to our experimental results, this method has a great improvement in accuracy compared with the single-feature method.

The system described in this paper provides unprecedented resolution at a low cost with the promise of facilitating the civilian drone monitoring technology and regulate drone flight behavior.

II. RELATED WORK

In recent years, the capabilities of small drones have greatly increased, and their manufacturing costs have significantly reduced. This multiplies the potential for benefits from drones and exacerbates the privacy and security risks[1]–[3].

As early as in the twentieth century, there are some studies on the detection of military aircraft[4]–[9]. Most

978-1-5090-6549-3/17 $31.00 © 2017 IEEE
DOI 10.1109/BigMM.2017.57

recently, with the rapid proliferation of consumer drones, some commercial drone detection products were presented[10], [11]. However, as the trend has just begun, drone detection technology research is just emerging. Most existing products utilize traditional radar-based methods[10]. However, radar has a hard time picking up these small, plastic, electric-powered drones because that is not what they were created to do. Some of the deployed radar-based products do not play an effective role. Moreover, as an active detection technology, radar is potentially invasive for the environment. Some products use radio frequency (RF) detection technology. For a particular type of drones, RF detection technology has an excellent detection results. However, such devices are costly and susceptible to interference from other signals such as Wi-Fi.

Other drone detection methods include visual-based detection[5], [12], [13], audio-based detection[4], [14], [15], and thermal-based detection. The principle of vision-based drone detection is similar to that of pedestrian detection or vehicle detection[16]. These methods use machine learning to detect the appearance or motion feature of the object and achieved great successes in many areas[17], [18]. Visual detection method has good counter-jamming performance, but because of the complex feature of small drone, visual-based detection method is less robust. The audio-based drone detection method has some limitations in urban environment, but it is a good assistance to other detection methods with improved equipment and feature extraction. The thermal-based detection method is limited by the resolution of the device and can not effectively detect the small drone.

To locate the drone, it is usually necessary to cover the monitoring area with multiple detection devices. Radars often use scanning to achieve this coverage and target location. The audio or visual detection method achieves this through a camera array or a microphone array. The camera array has many applications in human motion capture, 3D reconstruction, light field camera and so on. The purposes of these kinds of application are different from that of drone detection. So there is a large overlap between the FOV of each camera. The microphone array is used for noise reduction and acoustic camera applications, and it will be useful for drone detection.

III. Modular Audio-assisted Camera Array

We present a camera array system with microphone, which is designed to continuously monitor drone throughout the sky ball. The complexity of drone detection and uncertainty flight position often lead to detection failure. To handle these challenges, our system uses thirty high-definition cameras and three microphones mounted on a hemispherical frame, providing complete coverage of the sky. Higher camera resolution and frame rate enhance the visual feature of drone detection. Three directional microphones provide additional audio feature and improve the accuracy of drone detection. The system generates approximately 11.5 Gbps, and to handle this we propose a modularized architecture to parallel and distributed capture and processing. In this section, we describe the structure and architecture of the system, including acquisition, transmission and processing. As shown in Figure 1 and Figure 2.

A. Structure Design

The physical structure of the system is a set of concentric discs placed in multiple layers. This physical structure is selected because it can take advantage of the space inside the system to mount devices as many as possible. To reduce the weight of the system, the frame is made of aluminum alloy. The structure has a total height of 483 mm and an external ball radius 500 mm. There are four layers of the discs. The height and size of each disc take into account a variety of factors, including the placement of the device and the camera perspective.

Our design provides a complete coverage of the sky. To determine the placement of cameras, we calculate the FOV of the camera based on the focal length of the lens and the aspect ratio of the image. Using this FOV, we calculate the horizontal placement of one layer based on the horizontal viewing angle and the desired coverage. Then we derive the vertical placement of each layer from the vertical viewing angle. In particular,

$$l_x = \frac{2d}{\left(\frac{1}{\tan\frac{\pi}{n}} - \frac{1}{\tan\frac{\theta_x}{2}}\right)} \tag{1}$$

$$l_y = 2\tan\frac{\pi}{n}\left(\frac{d}{\sin\frac{2\pi}{n}} - \frac{\tan\theta_y}{\frac{l_y}{2}}\right) \tag{2}$$

where the d is radius between a camera and the system center. The l_x is the distance between horizontal view intersection point and the system center while the l_y is the that for the vertical view intersection point. The n is the

Figure. 1. The structure. The view of the system with the device mounted on the frame

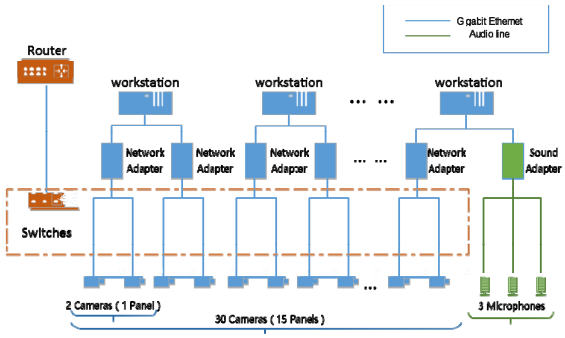

Figure. 2. Modularized system architecture. The equipment consists of thirty HD cameras, three microphones and 6 workstations. All the devices are connected through Gigabit Ethernet. The workstations are controlled by a master client.

number of cameras. The θ_x and θ_y are horizontal view angle and vertical view angle.

We use 3ds Max to achieve a scripted structural design simulation. Given a camera FOV and the desired coverage, the script automatically generates a visual simulation of the system structure. According to the above initialization, we use the simulation results to fine-tune the final system structure design.

B. *System Architecture*

Figure 3 shows the architecture of our system, which currently consists of 30 cameras, 3 microphones, 8 workstation nodes, and some network devices. The cameras are mounted on the outer of the discs, the microphone is mounted on the top disc, and other devices are mounted on the inner part of the discs. Every two cameras are grouped together, while each node controlling two groups through two separate network cards. In addition, the node controlling two cameras controls three microphones. Table 1 shows the specifications of the primary devices.

The devices in the system make up a local area network. Two gigabit switches connect each camera group and the corresponding workstation. Moreover, the switches are connected in cascade to the routers. A single camera produces an uncompressed video stream at 414 Mbps and, thus, for the entire set of 30 cameras the data rate is approximately 12.4 Gbps. Meanwhile, the data rate of the three microphones is approximately 4.6 Mbps.

TABLE I. DEVICE SPECIFICATION

Device	Att		
	Quantity	*Specification*	*Value*
Camera	30	Resolution	UXGA, 1920x1080
		Sensor	Global Shutter CMOS
		Lens focal length	7.6
		FPS	25
Microphone	3	Impedance	2000 ohms
		Sensitivity	-32dB
		Frequency response	50Hz-16kHz
Workstation	8	Model	ZOTAC EN1060
		Details	CPU:Intel i5-6400T; RAM:32GB; STORAGE:Dual 1TB SSD; Network: Dual gigabit; Graphics: NVIDIA GTX 1060

To handle this enormous stream, the visual system pipeline has been designed with a modular communication and control structure. Each subsystem consists of four cameras and a workstation node. Each node has dual purposes: it serves as a distributed storage unit and participates in a cluster as a multicore computing node. The system is controlled via a master node that the system operator can use to control all functions of the system.

IV. DRONE DETECTION

Our detection pipeline is illustrated in Figure 4 and contains the following steps:

- According to the initialization signal , synchronize multiple video streams and audio streams.
- Divide the video streams into a series of image frames, and divide the audio streams into a series of audio segments.
- Extract the feature of each frame and the feature of each the audio segment respectively.

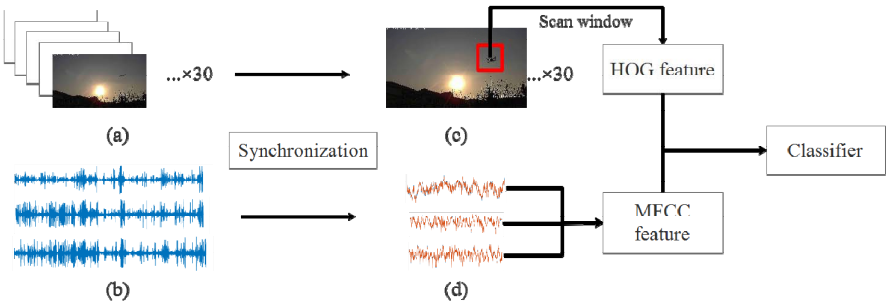

Figure. 3. Processing flow of drone detection. (a) Original image sequences. (b) Original audio samples. (c)(d) Synchronized image and audio frame.

- Classify each audio segment as containing drone noise or not. A similar classification is also performed for each image frame.
- The result of drone detection is obtained by the fusion of audio and image classification results.
- Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. *Data Acquiring and Synchronization*

Compared to 3D reconstruction, drone detection does not require accurate frame synchronization, so we do not use a special clock generator. To overcome the frame synchronization problem caused by the network delay, we developed a mobile phone application. This APP uses Bluetooth to control multiple handsets that emit both audible and visual signals as a start marker for video and audio capture.

Cameras are set to a fixed 25 FPS. Starting from the start marker, the audio stream is divided into multiple 1/25 second segments. Then we correspond the thirty video frames and three audio segments together for further procedures.

B. *Feature Extraction*

To achieve real-time drone detection, the current system uses relatively simple features. We use the HOG feature as the image feature while using MFCC(Mel Frequency Cepstral Coefficents) as the audio feature.

Due to the occlusion of Drone and the frequent disturbance of birds, we introduced MFCC audio feature for classification. MFCC is a non-linear mapping of the original frequency according to the auditory mechanism of the human ear. It is the most commonly used audio feature in current recognition tasks. The audio sections from the three audio streams are divided into sub-frames, and each sub-frame produces a 24-dimensional vector. Three audio sections generate a 72-dimensional vector. We then use the 40x40 image window to get the HOG feature and combine it with the audio feature. Finally, we train an SVM classifier to detect the drone in the scene. Meanwhile, we train another SVM classifier with audio feature to detect the noise of drones for early warning in our system.

V. RESULTS

In this section, we evaluate the performance of our system on a challenging dataset that includes multiple types of drones flying under various condition. It contains many real-world challenges, such as complex backgrounds and occlusions. The audio in the data set is synchronized with the image sequence. We divide the data sets into two groups, one for the training set and one for the test set.

A. *Data Set*

We use our system to collect a dataset. The dataset contains 20 collections of audio and image sequences taken under a variety of conditions. Each collection of data contains 30 image streams and 3 audio streams. The image sequence has a resolution of 1920x1080 and a frame rate of 25FPS. The audio sampling rate is 48kHz, and sampling precision is 16bit. The image sequences and audio streams are synchronized. The varieties of the dataset are as follows:

- Drone Type: Currently, the dataset contains 4 different types of drone. Including DJI Phantom4 (white, quadcopter), DJI Mavic (black, folding quadcopter), custom blue hexacopter (large), custom framed quadcopter (multiple colors). These models of the drone can represent the majority of consumer multirotor aircraft.
- Flight Range: The data set contains drones flight data in large areas. The flight altitudes range from 0 meters to 100 meters and the horizontal distance between a drone and the system range from 3meters to 200 meters.
- Background: The image background of the dataset includes the sky, buildings, mountains, trees and so on. Audio backgrounds include moving vehicles, human conversations, and wind noise.

Currently, our dataset contains approximately 120GB of data. Figure 4 shows a sample from dataset from all 30 cameras and the corresponding three audio clips.

B. *Drone Detection*

We use a randomly selected 90% of the data to train the classifier and use the remaining 10% of the data as the test set. We manually supplied 9680 bounding boxes centered on a drone. For a steadily flying drone video, we manually provide the bounding box in start and end frames and then interpolate to generate bounding boxes in intermediate frames. In additional, we pick a bounding box's upper, lower, left and right adjacent bounding box as a negative sample. The size of each bounding box is 40x40 pixels. Negative audio samples are recorded separately. We use the audio frames of size 40ms that equal to image frames.

To demonstrate the effectiveness of our approach, we compare it against the visual-feature-based drone detection method. Table 2 shows that our approach is more accurate than the one based on a single visual feature. In the case of

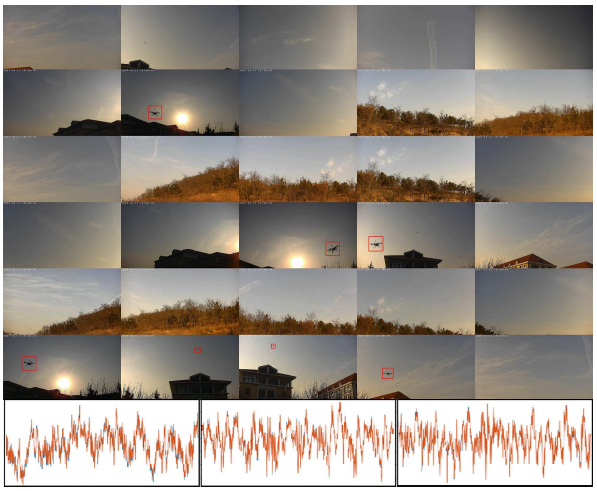

Figure. 4. Frames of 30 cameras and 3 microphones from the dataset. Drones are marked with red square. We keep the original exposures to demonstrate the robustness of our method.

obscured drone(positive samples), our approach can identify the presence of drone in the scene; in the case of interferences in the environment (negative samples), our approach can effectively reduce false detection. Figure 5 shows an occluded drone, and our system detects the presence of the drone.

TABLE II. PRECISION OF DETECTION METHODS

Samples	Precision	
	Visual Based	***Sound Assistant***
Positive	79.54%	95.74%
Negative	47.67%	82.61%

VI. CONCLUSION

We showed that audio features play a vital role in drone detection. We, therefore, developed an audio-assisted camera array that can monitor drone throughout the sky. At the same time, we proposed a novel drone detection method, which uses fused visual and audio features to detect drone in the scene. In our experiments, this method is more accurate than the visual-feature-based method. Besides, we have collected a dataset containing 20 sets of data. This data set can be used for drone detection of the bechmark.

In the future work, we will improve the system by two aspects. Primary, we will introduce motion information to achieve stable tracking of drone; secondary, we will introduce the noise reduction and localization algorithm based on microphone array to improve the accuracy of drone detection. Besides, we are considering other detection methods based on deep learning.

ACKNOWLEDGMENT

This research is sponsored by Shandong Province key research and development plan(2016ZDJS09A01) and Shandong Province Science development plan (2014GGE29069). The authors thanks all anonymous reviewers for the valuable comments and suggestions.

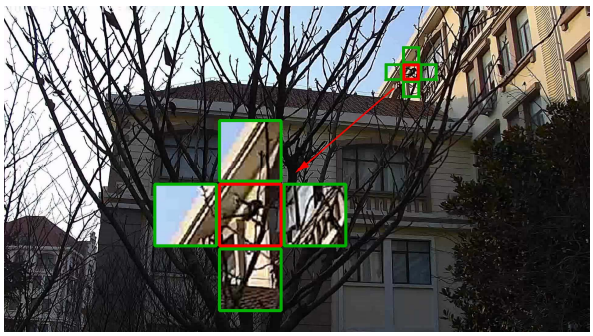

Figure. 5. An example of a occluded drone with complex background. Detection result and sampleing areas are enlarged to the original size. The detection result is labelled with red square and the negative samples are labelled with green squares.

REFERENCES

[1] R. Clarke and L. Bennett Moses, "The regulation of civilian drones' impacts on public safety," *Comput. Law Secur. Rev.*, vol. 30, no. 3, pp. 263–285, 2014.

[2] R. Clarke, "Understanding the drone epidemic," *Comput. Law Secur. Rev.*, vol. 30, no. 3, pp. 230–246, 2014.

[3] M. Brooks, "Welcome to the personal drone revolution," *New Sci.*, vol. 216, no. 2894, pp. 42–45, 2012.

[4] D. Chiang, W. Fishbein, and D. Sheppard, "Acoustic aircraft detection sensor," in *1993 Proceedings of IEEE International Carnahan Conference on Security Technology*, 1993, pp. 127–133.

[5] K. Dimitropoulos, N. Grammalidis, D. Simitopoulos, N. Pavlidou, and M. Strintzis, "Aircraft detection and tracking using intelligent cameras," in *IEEE International Conference on Image Processing 2005*, 2005, vol. 2, p. II-594-7.

[6] B. Kamgar-Parsi, A. K. Jain, and J. E. Dayhoff, "Aircraft detection: a case study in using human similarity measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 12, pp. 1404–1414, 2001.

[7] R. H. Khan and D. Power, "Aircraft detection and tracking with high frequency radar," in *Proceedings International Radar Conference*, 1995, pp. 44–48.

[8] J. R. Vasquez, K. M. Tarplee, E. E. Case, A. M. Zelnio, and B. D. Rigling, "Multisensor 3D tracking for counter small unmanned air vehicles (CSUAV)," in *Proc. SPIE*, 2008, vol. 6971, pp. 697107–697111.

[9] W. Yi and S. Marshall, "A novel approach for automatic aircraft detection," in *2000 10th European Signal Processing Conference*, 2000, pp. 1–4.

[10] S. Stapleford, "Drone aircraft detector," US 9337889 B1, 10-May-2016.

[11] B. Hearing and J. Franklin, "Drone detection and classification methods and apparatus," US 9275645 B2, 01-Mar-2016.

[12] A. Rozantsev, V. Lepetit, and P. Fua, "Detecting Flying Objects using a Single Moving Camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PP, no. 99, p. 1, 2016.

[13] A. Rozantsev, V. Lepetit, and P. Fua, "Flying objects detection from a single moving camera," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, vol. 07–12–June, pp. 4128–4136.

[14] A. Aljaafreh and A. Al-Fuqaha, "Multi-target classification using acoustic signatures in wireless sensor networks: A survey," *Signal Process. Int. J.*, vol. 4, no. 4, p. 175, 2010.

[15] J. Busset, F. Perrodin, P. Wellig, B. Ott, K. Heutschi, T. Rühl, and T. Nussbaumer, "Detection and tracking of drones using advanced acoustic cameras," in *Proc. SPIE*, 2015, vol. 9647, p. 96470F–96470F–8.

[16] "Mobileye Technology." [Online]. Available: http://us.mobileye.com/technology/. [Accessed: 07-Jan-2017].

[17] L. Huang, S. Tang, Y. Zhang, S. Lian, and S. Lin, "Robust human body segmentation based on part appearance and spatial constraint," *Neurocomputing*, vol. 118, pp. 191–202, 2013.

[18] W. Liu, T. Mei, and Y. Zhang, "Instant Mobile Video Search With Layered Audio-Video Indexing and Progressive Transmission," *IEEE Trans. Multimed.*, vol. 16, no. 8, pp. 2242–2255, 2014.