Poster: Audio-based Drone Ranging and Localization using Deep Learning

Gunhoo Park
Chung-Ang University
School of Computer Science and Engineering
Seoul, Republic of Korea
gunhoo0216@cau.ac.kr

Jeongyeup Paek
Chung-Ang University
School of Computer Science and Engineering
Seoul, Republic of Korea
jpaek@cau.ac.kr

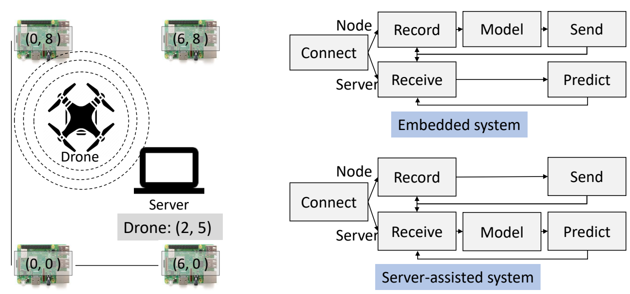
CCS CONCEPTS

• **Human-centered computing** → *Ubiquitous and mobile computing*; • **Computer systems organization** → *Embedded software*; • **Computing methodologies** → Machine learning.

KEYWORDS

UAV; Drone; Ranging; Localization; Deep Learning

ACM Reference Format:
Gunhoo Park and Jeongyeup Paek. 2019. Poster: Audio-based Drone Ranging and Localization using Deep Learning. In *The 17th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '19), June 17–21, 2019, Seoul, Republic of Korea.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3307334.3328613



(a) Scenario (b) System architecture

Figure 1: System scenario and architecture design

1 INTRODUCTION

Advancements in Micro-UAV (a.k.a drone) technology over the recent years have allowed drones to be utilized in various areas ranging from military applications to civilian usages [1, 2]. As such, accurate distance measurement and localization of drones became critical not only for its mission but also for detecting and identifying malicious usages [8]. Although TOA-based acoustic ranging method [5], acoustic signature based localization [6], vision-based search [7], and deep learning based drone detection techniques [3] exist, their usability and scope have yet been limited.

To this end, we propose a *real-time* audio-based system that uses *deep learning* for not only detecting but also *ranging* and *localization* of a drone. To explore the design space and investigate the feasibility of real-time acoustic ranging using deep learning, we first measure the drone detection accuracy and processing latency using two deep learning models on both an embedded and a server-class device. We then analyze the relationship between detection probability and distance measurement, and compare between binary- and multi-class classification approaches. This model will provide accurate distance estimate to a drone in real-time, thus allowing localization.

2 SYSTEM DESIGN

Figure 1(a) illustrates our scenario; Our system consists of several embedded sensor nodes deployed in the target area of interest for detecting drones, and also a server connected to those sensor nodes. Each embedded sensor node records audio sound, (pre-)processes it, and sends the processed data to the server. The server receives those data from each node, and computes the final location of the drone using the collected information in *real-time*. The key idea is to use the detection (classification) probability output of audio-based deep learning model as a proxy for estimating the distance to the drone. Whether this hypothesis holds, and if so, what would be the model for the relationship is the question. Once we obtain accurate distances from multiple sensor nodes, localization using trilateralization technique is relatively straight forward.

To investigate the feasibility of our idea, we first collected ∼1100 seconds of drone and background sounds in various environments. Then, we adopted Mel-frequency Cepstral Coefficients (MFCC) as the audio pre-processing method. The deep learning models that we have experimented with in this study are CNN and DNN, the two widely used techniques in the deep learning field. However, in general, DNN was performing better than CNN in several aspects. For example, using multiclass classification method with our collected data, classification accuracy was 90.8% for CNN and 97.1% for DNN with similar processing latency. For this reason, we decided to focus on DNN.

In order to design a system that can process in real-time, we first needed to investigate whether this is feasible through in-depth analysis of latency components of the system. This includes sound recording, pre-processing, network latency, classification processing, etc. For this purpose, we implemented two prototype systems as shown in Figure 1(b): an embedded system and a server-assisted system. The key difference is on 'where' the deep learning classification is done. In the embedded system, each embedded sensor node

ACM ISBN 978-1-4503-6661-8/19/06.
https://doi.org/10.1145/3307334.3328613

records the sound, pre-process and processes it, and sends the classification results of the deep learning model to the server. The server only calculates the distance to and position of the drone through ranging and localization algorithms based on the information received from each node. One the other hand, in the server-assisted system, each embedded sensor node only performs sound recording, and transmit the audio file to the server. The server receives the raw audio file from each node, pre/processes it, computes the distance to the drone through the deep learning classification and ranging algorithm, and calculates the final location by compiling the drone information of each node. To compare these two systems, we recorded sound in one second units, and compared the processing latencies at every steps. The total time required to process one second audio was only ~0.86 second for the server-assisted system, whereas embedded approach took ~6.22 second. Based on the result, we concluded that the server-assisted approach is suitable for real-time processing whereas the embedded approach is incapable.

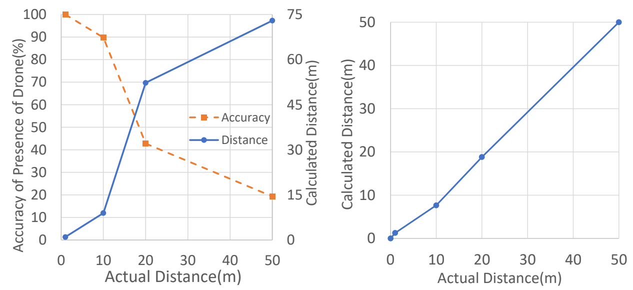
For the ranging algorithm, our approach is to build a model that can estimate the distance based on the deep learning classification result. In other words, our idea is to use the detection (classification) probability output of the audio-based deep learning algorithm as an input to a model that can estimate the distance to a drone. For this purpose, we have first explored a binary classification method in which the output of the deep learning model is the accuracy of the presence or absence of the drone. Training data set for this method has two simple classes: background noise (no drone) vs. drone sound within 10 meters distance. The second method is to have multiple classes in the training data according to the distance of the drone (e.g. 1m, 10m, 20m, 50m, no drone), and generating classification probability of each class for the test input. Then, the estimated distance will be a weighted average of those class distances and the classification probability.

3 PRELIMINARY EXPERIMENT RESULTS

We implemented a prototype of our system using RaspberryPi3 with an audio card [1] as a sensor node, and used a Linux laptop as the server. Tensorflow[2] is used as the deep learning library. Figure 2(a) plots the drone detection accuracy output from the DNN-based binary classification (left y-axis), as well as the estimated distance to the drone (right y-axis) based on the binary classification result. As can be seen from the figure, the probability of drone presence (detection) is not linearly correlated with the distance. This is not surprising; we were not necessarily expecting a linear relationship, but were interested in finding out 'some' relationship on which we can build a model between the two. Nevertheless, however, the estimated distance calculated using simple weighted averaging was accurate with the ground truth up till a certain point (20 meters in this result) after which it started to diverge.

In contrast, Figure 2(b) plots the estimated distance result from multiclass classification (drone detection classified by 4 different ranges), and shows that the estimated distance is well matched with the ground truth distance within acceptable error range. Based on these results, our ranging and localization algorithms will be based on multiclass classification model. The question is how may classes,

[1] http://www.audioinjector.net/rpi-hat
[2] https://www.tensorflow.org/



(a) Binary-classification (b) Multi-classification

Figure 2: Comparison of ranging accuracy for binary- and multi-class classification methods.

and we are still in the process of designing the optimal number of classes needed for accurate ranging while minimizing processing overhead subject to real-time processing.

Once the server obtains 3 or more ranging results, it combines them to compute the drone's location using the trilateralization technique. While doing so, the three distance estimates may not be accurate enough to pin-point the location of the drone (i.e. three circles not meeting at a single point, a circle enclosed in another, etc.). In such case, we can use geometric adjustment scheme [4] for estimation. Finally, we have implemented a prototype of proposed system, and have shown that it is feasible to update the drone's distance and location information in real-time. Our future work includes further investigation on the multi-class classification, devising an accurate model for estimating distance based on classification result, and implementing a large-scale practical system that can be used in real-world.

ACKNOWLEDGMENTS

This research work is currently supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2017R1D1A1B03031348.

REFERENCES

[1] Reg Austin. 2010. *Unmanned aircraft systems: UAVs design, development and deployment.* American Institute of Aeronautics and Astronautics.
[2] Bill Canis. 2015. *Unmanned aircraft systems (UAS): Commercial outlook for a new industry.* Congressional Research Service Washington.
[3] Sungho Jeon, Jong-Woo Shin, Young-Jun Lee, Woong-Hee Kim, YoungHyoun Kwon, and Hae-Yong Yang. 2017. Empirical Study of Drone Sound Detection in Real-Life Environment with Deep Neural Networks. *CoRR* abs/1701.05779 (2017). http://arxiv.org/abs/1701.05779
[4] Jeongyeup Paek, JeongGil Ko, and Hyungsik Shin. 2016. A Measurement Study of BLE iBeacon and Geometric Adjustment Scheme for Indoor Location-Based Mobile Applications. *Mobile Information Systems* 2016, 8367638 (Oct. 2016).
[5] Chunyi Peng, Guobin Shen, and Yongguang Zhang. 2012. BeepBeep: A high-accuracy acoustic-based system for ranging and localization using COTS devices. *ACM Transactions on Embedded Computing Systems (TECS)* 11, 1 (2012), 4.
[6] S. Sadasivan, M. Gurubasavaraj, and S. R. Sekar. 2001. Acoustic signature of an unmanned air vehicle exploitation for aircraft localisation and parameter estimation. *Defence Science Journal* 51, 3 (May 2001), 279.
[7] J. Tisdale, A. Ryan, , D. Tornqvist, and J. K. Hedrick. 2008. A multiple UAV system for vision-based search and localization. In *2008 American Control Conference.* 1985–1990.
[8] Ãřlker Bekmezci, Ozgur Koray Sahingoz, and Åđamil Temel. 2013. Flying Ad-Hoc Networks (FANETs): A survey. *Ad Hoc Networks* 11, 3 (2013), 1254 – 1270.