3D Motion Capture of an Unmodified Drone with Single-chip Millimeter Wave Radar

Peijun Zhao[†], Chris Xiaoxuan Lu[§], Bing Wang[†], Niki Trigoni[†], and Andrew Markham[†]

[†]Department of Computer Science, University of Oxford, United Kingdom
[§]School of Informatics, University of Edinburgh, United Kingdom

Abstract— Accurate motion capture of aerial robots in 3D is a key enabler for autonomous operation in indoor environments such as warehouses or factories, as well as driving forward research in these areas. The most commonly used solutions at present are optical motion capture (e.g. VICON) and Ultrawideband (UWB), but these are costly and cumbersome to deploy, due to their requirement of multiple cameras/anchors spaced around the tracking area. They also require the drone to be modified to carry an active or passive marker. In this work, we present an inexpensive system that can be rapidly installed, based on single-chip millimeter wave (mmWave) radar. Importantly, the drone does not need to be modified or equipped with any markers, as we exploit the Doppler signals from the rotating propellers. Furthermore, 3D tracking is possible from a single point, greatly simplifying deployment. We develop a novel deep neural network and demonstrate decimeter level 3D tracking at 10Hz, achieving better performance than classical baselines. Our hope is that this low-cost system will act to catalyse inexpensive drone research and increased autonomy.

I. Introduction

Motion capture of dynamic aerial robots (e.g. a small UAV or quadcopter drone) is a key enabling capability for navigation, path planning and autonomy. Motion capture in this context means the estimation of precise 3D position in real-time, with an external sensor. Unlike on-board self-localization with techniques such as Visual-Inertial Odometry, localization with an external sensor can give an absolute 3D location, and does not suffer from accumulated drifting. In outdoor areas with a clear sky view, the use of RTK GPS, often combined with secondary sensors such as IMU, can provide excellent absolute positioning accuracy.

Indoors or in GPS denied areas such as tunnels, alternative tracking approaches are typically employed. Broadly, these can be divided into optical or RF based techniques. VICON [1] is one of the most widely used optical motion capture platform. These systems provide high precision 3D tracking of a reflective marker placed on a drone, provided that there is a good view from two or more spatially distributed cameras. These systems however are expensive ($10k upwards), time-consuming to set up and calibrate, and require modifying the drone to be able to track it. To prevent dead-zones, a large number of cameras (e.g. 10) need to be placed around the tracking volume and typically are mounted at height (e.g. 2-3 m) to provide a wide field of view. A large amount of research has considered the use of UWB ranging as a multi-lateration approach [2], [3]. These systems provide reasonable accuracy (10cm), but require the drone to carry a UWB transceiver to achieve pair-wise distance estimation. They also need a network of UWB transceivers distributed around the convex hull of the area to be tracked, leading to a time-consuming setup and calibration phase. It's typically not cheap as well for setting up a UWB localization system (around $300 for an anchor and $30 for a tag).
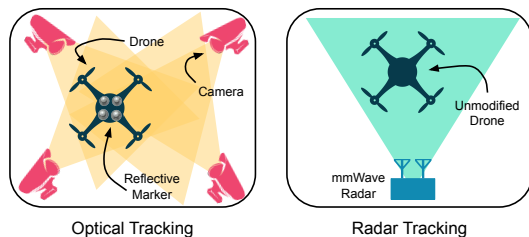

Fig. 1: Compared with the optical tracking method, our proposed system does not require any marker modification on the drone and only needs a low-cost mmWave radar.

In this paper we present a new approach for drone motion capture that is able to operate from a single point (i.e. it does not require multiple anchors to be carefully arranged around the tracking volume). As a result, it is easy to setup and deploy in new environments. mDrone is able to track an unmodified drone in 3D with precision comparable to UWB ($< 10cm$). Based on a single chip mmWave integrated radar transceiver, mDrone is low-cost (<$300) and low-power, and has the capability of operating in the dark and other visually challenging conditions.

Our primary innovation is the use of single chip mmWave radar which brings its own set of unique challenges. Although radar has a long history of being used to track aeroplanes at extremely long ranges (hundreds of kms), the majority of solutions are military/aviation grade, large, costly and power-hungry. Recent advances in microchip fabrication have led to the availability of multi-channel, wideband radar transceivers that have been successfully used in diverse applications such as SLAM [4]–[7], industrial automation [8], and human-computer interactions [9]–[12]. However, a number of challenges arise from the problem of drone tracking. Firstly, the drone is unmodified, i.e. it does not carry any bright target e.g. a retroreflector. This makes it challenging to be detected, but we demonstrate how to exploit measurements of the propeller's doppler velocity to refine the detection. Secondly, the drone is physically large (e.g. 60 cm across), so accurately tracking its centre is difficult as

the drone appears as a non-uniform blob in the radar returns. Thirdly, existing techniques are computationally expensive or inaccurate, precluding the ability for real-time motion capture. To address these challenges, we introduce mDrone, a new deep learning pipeline and demonstrate its performance and robustness in comparison to five conventional signal processing baselines. We believe that this low-cost motion capture platform will not only be useful as a low-cost research and development tool, but also enable a number of downstream applications such as precision maintenance and inspection, robotic interaction, and increased autonomy in warehouses.

Our main contributions include:

- We propose a deep learning method for the estimation of drone position based on mmWave radar
- We collected 100 sequences of data and evaluated our proposed algorithm against multiple conventional signal processing algorithms. The dataset will be released to the community together with the code.
- Our system achieves a mean error of less than 10 cm in 3D whilst operating at 10 Hz.

II. RELATED WORK

A. Cooperative Localization

In cooperative tracking, modifications are made to the drone in order to track it. To achieve this, either markers or transceivers are installed on the drone itself. One of the most accurate 6-DOF cooperative tracking methods is based on Optical Tracking systems (e.g. Vicon). Hemispherical markers are attached to the body of the drone to form a rigid and spatially unique pattern, which are modeled and tracked by the system [1], [13]. Optical tracking systems are expensive (tens of thousands of dollars to equip a room with an optical tracking system) and time consuming to install and calibrate. Santos et. al. propose an optical tracking replacement system, for the estimation of 6-DOF drone pose, with data from multiple sensors, including RGB-D camera and on-board IMU, based on computer vision and sensor fusion approaches [14].

UWB localization systems based on RF ranging and multilateration are widely used for their lower cost, whilst still providing accurate positioning (errors typically around $10cm$ or higher for outdoor scenarios) [2], [15], [16]. Ledergerber et. al. designed a system using one-way UWB for drone self-localization [17], and Angle of Arrival estimation [18]. There are also many works that fuse UWB with other sensors for a better localization accuracy, such as UWB/IMU/Vision [19], [20], UWB/IMU [21], and UWB/Radar [22]. UWB techniques are also used for multi-UAV localization [23]–[25].

B. Non-cooperative Localization

Despite the fact that computer vision algorithms for object recognition and tracking with cameras are developing rapidly [26], cameras do not work robustly under extreme light conditions. Researchers have been exploring other sensors for passive drone localization, including Light

Detection And Ranging (LiDAR) [27] and LAser Detection And Ranging (LADAR) [28]. Audio sensor arrays can be used for drone localization [29], which are based on TDOA principles, but they are sensitive to ambient noise and have limited ranges [30].

Different types of radars are also widely used for drone detection and localization [31]–[33], amongst which, FMCW radars have attracted a great deal of research interest [34]. In our work, we used a commercial-off-the-shelf mmWave FMCW radar, and focus on exploring the combination of signal processing and deep learning techniques to greatly improve short-range drone localization accuracy.

III. CONVENTIONAL DRONE TRACKING PIPELINE
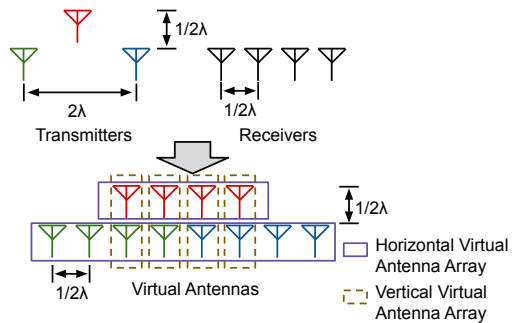
A. FMCW Radar Background

Fig. 2: Texas Instruments IWR6843 mmWave Radar MIMO Virtual Antenna Array

We first introduce the principles and limitations behind conventional tracking pipelines. We consider a Frequency-Modulated Continuous-Wave (FMCW) Radar, which detects objects with electromagnetic wave chirps whose carrier frequency increases linearly with time. To demodulate, the reflected signal is mixed with the transmitting signal and produces an intermediate frequency (IF) signal, which is the difference between the transmitting and receiving signal. As the transmitting frequency is linearly increasing, the frequency shift of the IF signal is thus proportional to the distance of the object. The amplitude of the signal varies as a function of strength of the reflectivity. The distance to an object can be calculated by extracting peak frequency components of the IF signal. This is typically accomplished using a signal processing technique such as a Fast Fourier Transform (FFT). As the radar emits chirps continuously, the phase shift between chirps at a particular frequency component of the IF signal can be used to estimate the speed of the object, caused by a Doppler shift.

In this work, we consider the use of a Texas Instruments(TI) IWR6843 single chip radar, but the concepts presented here are generalizable to other platforms. The TI IWR6843 radar features a 3Tx/4Rx MIMO antenna array, which forms 12 virtual antennas (VA) based on the combination of Tx-Rx pairs [35]. The layout of the VA array is shown in Fig. 2. The phase differences between the antennas of the linear receiver arrays (horizontal and vertical) can be used to estimate the azimuth and elevation angle of arrival

of different objects in the scene. For each frame, the radar is able to generate a data cube of complex numbers, with the axes respective to chirp index, samples, and virtual antennas.

B. Sensitivity to Moving Objects

FMCW radars are particularly good at detecting moving objects, because a small movement of the target leads to a large phase shift of the receiving signals. A UAV's rotating propellers can easily be detected and separated from the background static clutter. This is performed by a clutter removal algorithm that removes stationary objects from the FFT, as stationary objects have no phase shift.

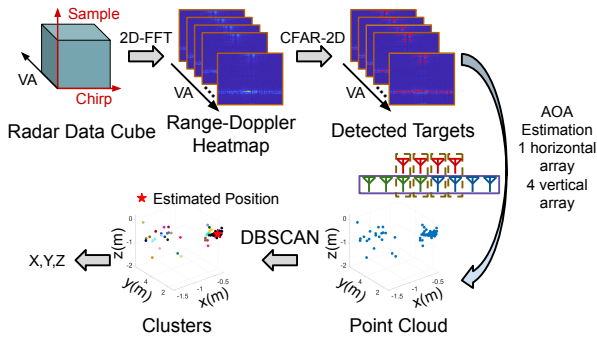### C. Tracking the Drone with a Radar Point Cloud

Fig. 3: Radar Point Cloud pipeline

The mmWave radar is able to generate a 3D point cloud of the scene based on the principles detailed above, which can be used to localize the drone. As input, the radar data cube is first processed along the chirp-sample dimension for 2D object detection, by forming a range-Doppler heatmap. The 2D-CFAR algorithm is used to extract dominant points. The phase differences between the virtual antennas corresponding to a detected point are then used to estimate the 3D bearing angle of each point. These are then combined to produce a sparse 3D point cloud. The full pipeline is shown in Fig. 3. Using the 3D point cloud, a clustering algorithm such as DBScan is used to segment the point cloud into clusters. As the background clutter has been removed, the largest cluster corresponds to the drone, under the assumption that the scene is relatively stationary. By taking the centroid of the largest cluster, we can estimate the position of the drone.

D. Tracking the Drone with Distance and Angle Estimation

An alternative approach to estimate the location of the UAV can be performed by omitting the chirp axis of the radar cube, and only considering the Sample-VA plane. Conventional signal processing algorithms, such as the Fast Fourier Transform, or the super-resolution based MUSIC (MUltiple SIgnal Classification) algorithm [36], can be used to estimate the distance as well as both the azimuth and elevation angle of arrival (AoA) of the object. Given the distance and angle from multiple virtual antennas we are able to obtain the location of the drone. By repeating this process over the chirp axis of the radar cube, we can estimate the trajectory of the drone.
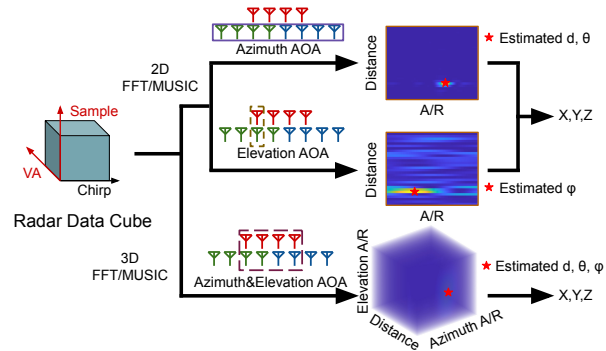

Fig. 4: 2D (top) and 3D (bottom) FFT and MUSIC pipelines.

As a refinement, we can perform 3D FFT or MUSIC directly on the cubic data formed by the Sample axis and the VA matrix shown in bottom pipeline of Fig. 4, for a globally optimal estimation. The complexities of these algorithms are however exponential in the number of dimensions. 2D FFT/MUSIC based pipelines run significantly faster than 3D FFT/MUSIC based pipelines, but are more sensitive to noise and hence lack robustness and accuracy.

E. Problems with Conventional Drone Tracking Pipeline

Although tracking a drone based on conventional signal processing algorithms is straightforward and easy to implement, these methods suffer from a number of problems which lead to low accuracy in practice. First of all, the drone is not a single point, but a complex object with multiple rotating points. Depending on the orientation of the drone relative to the antenna array, the signals from one or more propellers can be blocked or occluded by the drone body itself. This typically leads to a systematic localization error. In addition, there are many empirical parameters in conventional algorithms and it can be difficult to fine-tune and assign optimal values to those parameters. The localization results produced by conventional algorithms are not stable, and contain many outliers. This is because in some frames the noise or multipath scattering overwhelms the primary signal and is wrongly identified as the target.

IV. PROPOSED METHOD

Our method for drone localization contains two parts. Firstly, the radar cube data is preprocessed in range-elevation and range-azimuth heatmaps. Secondly, a deep neural network, which consumes the heatmaps as input data and produces a 3D localization estimation. Fig. 5 shows the pipeline of the proposed method and the two parts are discussed below.

A. Data Preprocessing

The data processing part is similar to a conventional 2D FFT pipeline, except for that it produces 6 heatmaps in total (2 azimuth heatmaps and 4 elevation heatmaps), rather than just one azimuth heatmap and one elevation heatmap for target localization as in the conventional 2D FFT pipeline. In addition, heatmaps of multiple chirps with the data cube are extracted and sent into the neural network.
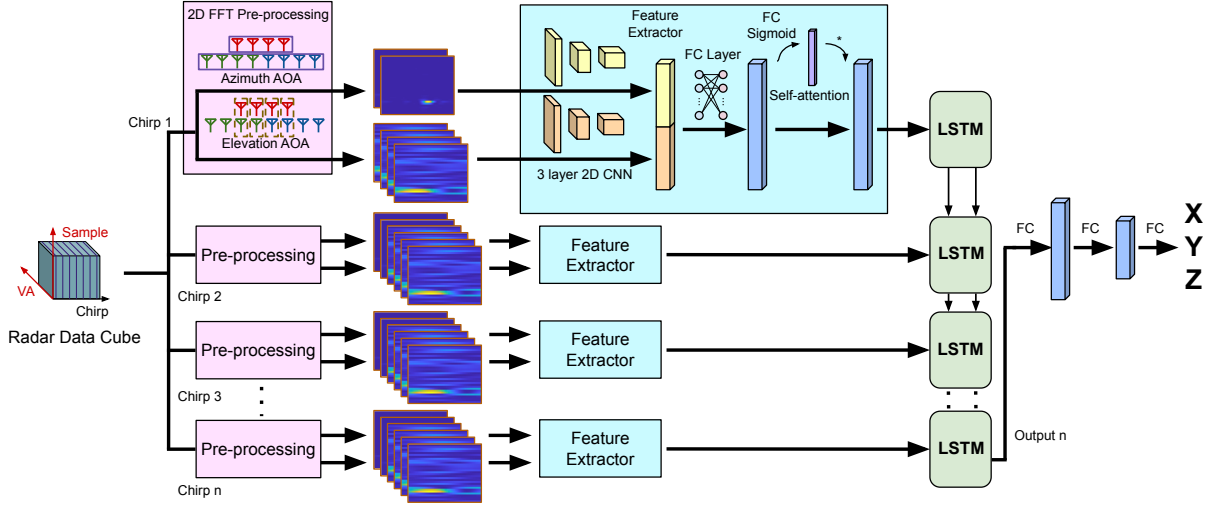
Fig. 5: Proposed Method. For each chirp, the pre-processing module generates two range-azimuth heatmaps and four range-elevation heatmaps, which are sent into the Feature Extractor respectively. The feature vectors of multiple chirps in a frame are input into a LSTM network and the last output is used to estimate the 3D location.

B. Network Structure

The neural network we propose contains two parts. The first part is a feature extractor, which takes range-azimuth heatmaps and range-elevation heatmaps as input. Two 3layer 2D CNNs are used to extract an azimuth feature vector and an elevation feature vector. The two feature vectors are concatenated together and fed into a fully-connected layer to form an overall chirp feature vector which combines both azimuth and elevation information. A self-attention mechanism [37] is applied to the chirp feature, to assist in rejecting background noise by focusing on relevant features. The feature vectors of each chirp are sent into a LSTM network. Three fully connected layers are then used to estimate the final output from the hidden layer of LSTM at the last timestamp. Note that though our proposed method features a LSTM which takes a time-series of input, it uses the same amount of information (the data in a single frame) as the baseline methods.
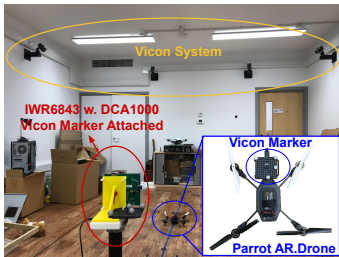
V. IMPLEMENTATION

A. Data Collection

Fig. 6: Experiment Settings

1) Experiment Setup: We collected 100 1-minute sequences of a drone flying in a room equipped with a Vicon system to provide accurate ground truth, as shown in Fig. 6. For each sequence, the data collection begins after the drone is launched into the air so the 100 sequences purely contain the data of the drone flying. The drone was manually controlled to fly randomly in the experiment space. We use a TI IWR6843ISK together with DCA1000EVM for raw data streaming. A desktop computer with Windows operating system is used for configuring the device and recording raw radar data, with mmWave Studio software. A laptop with Ubuntu operating system and ROS is used for Vicon data recording, and the Vicon data is later parsed to extract ground truth.

2) Coordinate System Definition: The mmWave Radar is a right hand system, with y axis pointing forward, x axis pointing right, and z axis pointing upward.

3) Time Synchronization: Since the raw data streaming and Vicon data recording happen on two platforms respectively, we need to synchronize the timestamps of each frame in the mmWave data to the corresponding Vicon ground truth. The time of the two platforms are synchronized using NTP server. mmWave Studio is not able to provide timestamp logging for each frame, so we record the time when the radar starts and extrapolate the timestamp of each frame based on frame period.

B. Implementation

We have 5 baselines in total, as introduced in Section. III, including Point Cloud based method, 2D FFT based method, 3D FFT based method, 2D MUSIC based method, and 3D MUSIC based method. All the conventional signal processing baselines are implemented in MATLAB. The size of the range FFT is 256, and the size of the Angle FFT is 180; in MUSIC, the range swept is 1m to 4m, with an interval of 0.1m, and the azimuth angle swept is 30°to 150°, with interval of 1°, and elevation angle -15°to 15°, with interval of 1°.

For our proposed method, the data preprocessing is done with MATLAB and the neural network is implemented in

Python with PyTorch library.

VI. EVALUATION

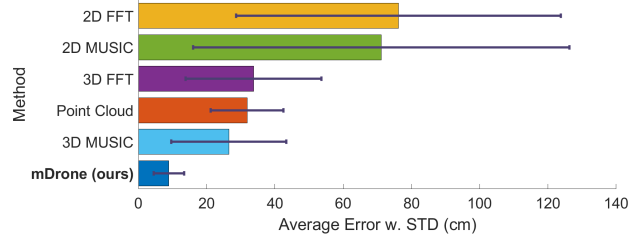## A. Localization Evaluation

Fig. 7: Localization Error with Standard Deviation (cm).

As our proposed methods involves deep learning techniques, we randomly picked 10 sequences as testing sequences, 10 sequences as validation sequences and the remaining 80 sequences as training sequences. All competing conventional methods are tested on the testing sequences. The deep neural network in our proposed pipeline is trained on the training sequences for 20 epochs and the parameters which produce the best validation result are saved and tested on the testing sequences. Fig. 7 summarizes the result of the localization evaluation.

From the result we can see that our proposed method significantly outperforms the competing signal processing methods. The average error of our proposed method is ≈3 times smaller the best competing baseline, which is 3D MUSIC. Besides lower average error, our proposed method is also much more robust than competing methods, for it has a Standard Deviation less than half of the Point Cloud based method, which is the most stable baseline. This proves that our proposed method can produce excellent results for UAV localization.
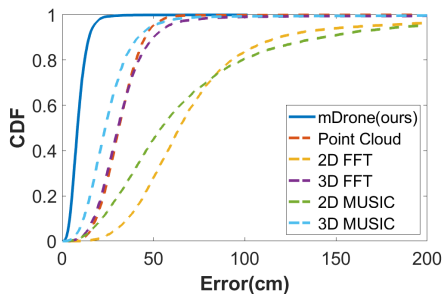

Fig. 8: Cumulative Distribution Function (CDF) of the errors produced by different methods.

To further investigate the localization error of different methods, we plot the Cumulative Distribution Function (CDF) of the errors of different methods, as shown in Fig. 8. The CDF shows that over 90% of time, the error is less than 15cm, which is significantly lower than the next best competitor (3D-MUSIC: 46cm).

A sample of the comparison of different techniques is shown in Fig. 9. 2D FFT and 2D MUSIC are not stable as they search for a single global maximum on the 2D heatmap, and often produce high errors. This is because a wrong prediction in either azimuth heatmap or elevation
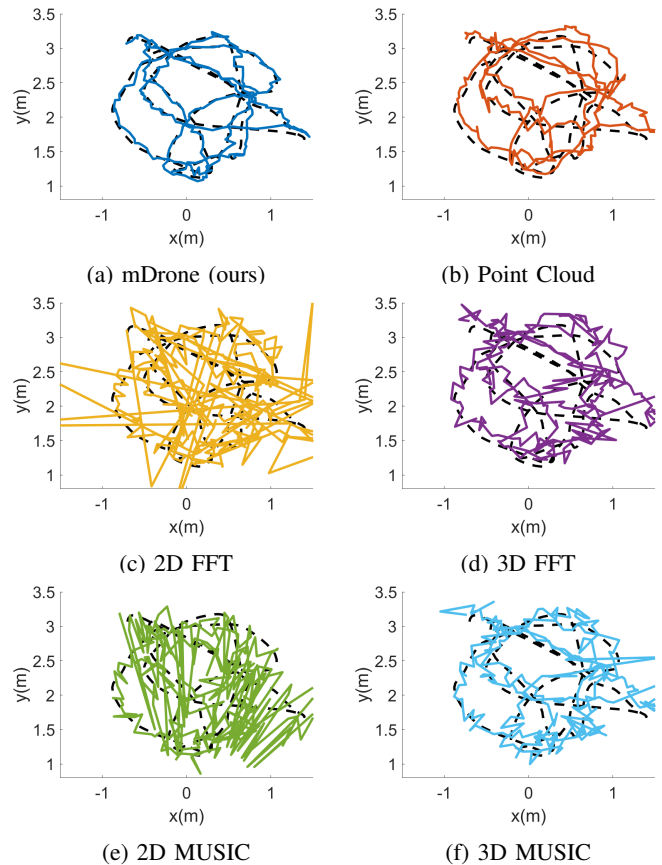

Fig. 9: Qualitative localization comparison on a sample trajectory. Black dashed lines represent ground truth.

Method	Time per Frame (ms)
Point Cloud	101
2D FFT	**5**
3D FFT	123
2D MUSIC	16477
2D MUSIC (parallel)	15426
3D MUSIC	282430
3D MUSIC (parallel)	220755
mDrone (Ours)	43

TABLE I: Running Time Comparison.

heatmap would result in a large 3D overall error. 3D FFT and 3D MUSIC search for the peak in a 3D heatmap and hence more robust, with 3D MUSIC producing a better result thanks to the super-resolution feature. However, the localization accuracy is still worse than mDrone. The Point Cloud technique is more stable than 3D FFT and 3D MUSIC, but still significantly inferior compared to mDrone. Adding other techniques like sliding window average or Kalman Filter/Smoother on the single radar cube localization results may further improve the stability and accuracy of mDrone, but it is beyond the topic of this paper and is not discussed here.

B. Running Time Comparison

We compare the average running time of the conventional baselines as well as our proposed method. The MUSIC pipelines are implemented with both single thread and 16 thread parallel for speeding up. The experiment is performed

with a AMD Ryzen 3800X CPU with 64GB main memory and a NVIDIA RTX2080Ti graphics card. The results are shown in TABLE. I. mDrone is the second fastest among all the competing methods. The MUSIC algorithm based methods are shown to be extremely resource consuming as it takes more than 15 seconds for 2D music to process a radar data cube, and 220 seconds for 3D music, even in parallel mode. The Point cloud method and 3D FFT method take around 100 ms for a radar data cube, and are more than 2 times slower than mDrone. The only method that is faster than ours, which is the 2D FFT method, produces much worse positioning results. mDrone takes around 43ms for a single frame on average. Note that the frame (radar data cube) period is 100ms in this work, which means that with proper implementation, mDrone is capable of running in real time at 10Hz.

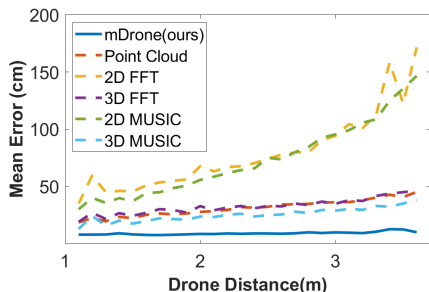## C. Impact of Distance on Tracking Error

Fig. 10: Mean Error v.s. Distance. Mean errors of conventional baselines increase with distance, while mean error of mDrone is relatively constant.

We calculated the mean error at different distances for each method, with the step of 0.1m. Results are shown in Fig. 10. The errors of conventional baseline methods, especially 2D FFT and 2D MUSIC, increase significantly with distance, as angular error results in a larger overall position estimation error when the range is further. However, mDrone is able to achieve a relatively stable performance at different target distances, as the neural network takes the whole 2D range-angle heatmaps as inputs, rather than peak points, which could produce more reliable predictions.

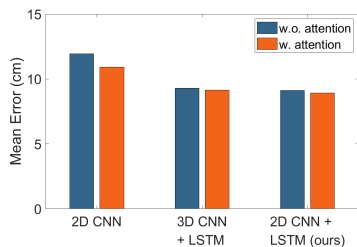## D. Ablation Study: Impact of Network Structure

Fig. 11: Comparison between Different Network Structures.

To show the functionality of different parts in our proposed neural network, we do the following three experiments:

- To prove that 2D CNN is a good choice for feature extraction, we replace the 2D CNN feature extractor with a 3D CNN feature extractor. (3D CNN + LSTM in Fig. 11)
- To show the effect of LSTM, we remove the LSTM and only use the features extracted by the feature extractor from the first chirp to estimate the localization result. (2D CNN in Fig. 11)
- To prove the effect of self-attention, we remove the self-attention module in each model. (Blue bars in Fig. 11)

Each of the above-mentioned models is trained with the same 80 training sequences for 20 epochs. The accuracy comparison is shown in Fig. 11. As we can see from the result, the neural network produces better results when using a LSTM layer after the feature extractor than directly using FC layers to produce the final output by around 20%. Using either 3D CNN or 2D CNN as feature extractor would produces similar results, while a 2D CNN is slightly better than 3D CNN by around 2.5%. For all three structures, adding self-attention on the extracted feature increases the model performance by around 1.5% to 8%. The results as a whole though demonstrate the utility of data-driven techniques to track the moving drone, being relatively insensitive to the precise network structure.

VII. DISCUSSION

As mDrone is able to process one frame in less than 100ms, it could be used in real time in a 10Hz system with proper implementation. The localization performance of mDrone is relatively stable at different target distances, which implies that the tracking distance could be potentially extended further. Our future works would include studying the impact of drone speed on localization accuracy, improving tracking range, and the ability to generalize to completely different drone models (e.g., helicopter, hexacopter, mini-drones, etc., as different drones have different radar cross sections [38]), different sensor setups, as well as simultaneously tracking multiple drones.

VIII. CONCLUSION

In this work, we propose mDrone, a method of non-cooperative drone localization with a grounded millimeter wave radar, as an alternative for optical tracking and UWB. mmWave radar is very suitable for localizing drones as it is very sensitive to moving objects so the drone can be easily detected by it's propellers even when hovering. mDrone consists of a two-step tracking pipeline and is compared to several conventional signal processing baselines. Results show that our method could achieve an average localization error of 8.92cm, which is more than 3 times better compared to the best baseline method. mDrone is also more stable than conventional signal processing methods, and is efficient enough to be used in real time.

REFERENCES

[1] S. Al Habsi, M. Shehada, M. Abdoon, A. Mashood, and H. Noura, "Integration of a vicon camera system for indoor flight of a parrot ar drone," in *2015 10th International Symposium on Mechatronics and its Applications (ISMA)*. IEEE, 2015, pp. 1–6.

[2] K. Guo, Z. Qiu, C. Miao, A. H. Zaini, C.-L. Chen, W. Meng, and L. Xie, "Ultra-wideband-based localization for quadcopter navigation," *Unmanned Systems*, vol. 4, no. 01, pp. 23–34, 2016.

[3] N. Macoir, J. Bauwens, B. Jooris, B. Van Herbruggen, J. Rossey, J. Hoebeke, and E. De Poorter, "Uwb localization with battery-powered wireless backbone for drone-based inventory management," *Sensors*, vol. 19, no. 3, p. 467, 2019.

[4] C. X. Lu, S. Rosa, P. Zhao, B. Wang, C. Chen, J. A. Stankovic, N. Trigoni, and A. Markham, "See through smoke: robust indoor mapping with low-cost mmwave radar." in *MobiSys*, 2020, pp. 14–27.

[5] C. X. Lu, M. R. U. Saputra, P. Zhao, Y. Almalioglu, P. P. B. de Gusmao, C. Chen, K. Sun, N. Trigoni, and A. Markham, "milliego: Single-chip mmwave radar aided egomotion estimation via deep sensor fusion," in *ACM Conference on Embedded Networked Sensor Systems (SenSys)*, 2020.

[6] A. Yassin, Y. Nasser, A. Y. Al-Dubai, and M. Awad, "Mosaic: Simultaneous localization and environment mapping using mmwave without a-priori knowledge," *IEEE Access*, vol. 6, pp. 68 932–68 947, 2018.

[7] M. Aladsani, A. Alkhateeb, and G. C. Trichopoulos, "Leveraging mmwave imaging and communications for simultaneous localization and mapping," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 4539–4543.

[8] C. Jiang, J. Guo, Y. He, M. Jin, S. Li, and Y. Liu, "mmvib: micrometer-level vibration measurement with mmwave radar," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–13.

[9] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–19, 2016.

[10] H. Liu, Y. Wang, A. Zhou, H. He, W. Wang, K. Wang, P. Pan, Y. Lu, L. Liu, and H. Ma, "Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4, pp. 1–28, 2020.

[11] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, 2019, pp. 51–56.

[12] Z. Li, Z. Lei, A. Yan, E. Solovey, and K. Pahlavan, "Thumouse: A micro-gesture cursor input through mmwave radar-based interaction," in *2020 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2020, pp. 1–9.

[13] J. Donald, "Indoor flight of ar drone 2.0 uav in vicon motion capture system," *The UNSW Canberra at ADFA Journal of Undergraduate Engineering Research*, vol. 10, no. 1, 2018.

[14] M. C. Santos, L. V. Santana, A. S. Brandão, M. Sarcinelli-Filho, and R. Carelli, "Indoor low-cost localization system for controlling aerial robots," *Control Engineering Practice*, vol. 61, pp. 93–111, 2017.

[15] J. Tiemann, F. Schweikowski, and C. Wietfeld, "Design of an UWB indoor-positioning system for UAV navigation in GNSS-denied environments," *2015 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2015*, 2015.

[16] A. Masiero, F. Fissore, R. Antonello, A. Cenedese, and A. Vettore, "A comparison of uwb and motion capture uav indoor positioning." *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2019.

[17] A. Ledergerber, M. Hamer, and R. D'Andrea, "A robot self-localization system using one-way ultra-wideband communication," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 3131–3137.

[18] A. Ledergerber, M. Hamer, and R. D'Andrea, "Angle of arrival estimation based on channel impulse response measurements," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 6686–6692.

[19] A. Benini, A. Mancini, and S. Longhi, "An imu/uwb/vision-based extended kalman filter for mini-uav localization in indoor environment using 802.15. 4a wireless sensor network," *Journal of Intelligent & Robotic Systems*, vol. 70, no. 1-4, pp. 461–476, 2013.

[20] D. Hoeller, A. Ledergerber, M. Hamer, and R. D'Andrea, "Augmenting ultra-wideband localization with computer vision for accurate flight," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 12 734–12 740, 2017.

[21] W. You, F. Li, L. Liao, and M. Huang, "Data Fusion of UWB and IMU Based on Unscented Kalman Filter for Indoor Localization of Quadrotor UAV," *IEEE Access*, vol. 8, pp. 64 971–64 981, 2020.

[22] S. Zahran, M. M. Mostafa, A. Masiero, A. M. Moussa, A. Vettore, and N. El-Sheimy, "Micro-radar and UWB aided UAV navigation in GNSS denied environment," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 42, no. 1, pp. 469–476, 2018.

[23] J. Tiemann and C. Wietfeld, "Scalable and precise multi-UAV indoor navigation using TDOA-based UWB localization," *2017 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2017*, vol. 2017-Janua, no. September, pp. 1–7, 2017.

[24] W. Shule, C. M. Almansa, J. P. Queralta, Z. Zou, and T. Westerlund, "UWB-Based Localization for Multi-UAV Systems and Collaborative Heterogeneous Multi-Robot Systems," *Procedia Computer Science*, vol. 175, no. 2019, pp. 357–364, 2020. [Online]. Available: https://doi.org/10.1016/j.procs.2020.07.051

[25] Y. Qi, Y. Zhong, and Z. Shi, "Cooperative 3-D relative localization for UAV swarm by fusing UWB with IMU and GPS," *Journal of Physics: Conference Series*, vol. 1642, p. 012028, 2020.

[26] C. Aker and S. Kalkan, "Using deep networks for drone detection," in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2017, pp. 1–6.

[27] K. Mirceta, C. Bohak, B. H. Kim, M. Y. Kim, and M. Marolt, "Drone segmentation and tracking in grounded sensor scanned lidar datasets," in *Proceedings of the Zbornik sedemindvajsete mednarodne Elektrotehniške in racunalniške konference ERK*, 2018, pp. 384–387.

[28] B. H. Kim, D. Khan, C. Bohak, W. Choi, H. J. Lee, and M. Y. Kim, "V-rbnn based small drone detection in augmented datasets for 3d ladar system," *Sensors*, vol. 18, no. 11, p. 3825, 2018.

[29] X. Chang, C. Yang, J. Wu, X. Shi, and Z. Shi, "A surveillance system for drone localization and tracking using acoustic arrays," in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2018, pp. 573–577.

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

[31] İ. Güvenç, O. Ozdemir, Y. Yapici, H. Mehrpouyan, and D. Matolak, "Detection, localization, and tracking of unauthorized uas and jammers," in *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. IEEE, 2017, pp. 1–10.

[32] M. Caris, W. Johannes, S. Sieger, V. Port, and S. Stanko, "Detection of small uas with w-band radar," in *2017 18th International Radar Symposium (IRS)*. IEEE, 2017, pp. 1–6.

[33] G. Wei, Y. Zhou, and S. Wu, "Detection and localization of high speed moving targets using a short-range uwb impulse radar," in *2008 IEEE Radar Conference*. IEEE, 2008, pp. 1–4.

[34] J. Drozdowicz, M. Wielgo, P. Samczynski, K. Kulpa, J. Krzonkalla, M. Mordzonek, M. Bryl, and Z. Jakielaszek, "35 ghz fmcw drone detection system," in *2016 17th International Radar Symposium (IRS)*. IEEE, 2016, pp. 1–4.

[35] Texas Instruments, "Mimo radar," [Online], https://www.ti.com/lit/an/swra554a/swra554a.pdf.

[36] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[38] V. Semkin, J. Haarla, T. Pairon, C. Slezak, S. Rangan, V. Viikari, and C. Oestges, "Analyzing radar cross section signatures of diverse drone models at mmwave frequencies," *IEEE Access*, vol. 8, pp. 48 958–48 969, 2020.