Weiguo Wang, Yuan He *Tsinghua University, Beijing, China*
Luca Mottola *Politecnico di Milano, Milan, Italy; RI.SE, Sweden; and Uppsala University, Uppsala, Sweden*
Shuai Li, Yimiao Sun, Jinming Li *Tsinghua University*
Hua Jing, Ting Wang, Yulei Wang *Meituan, Beijing, China*

Editors: Nicholas D. Lane and Xia Zhou

ACOUSTIC LOCALIZATION OF DRONES IN PRECISE LANDING:

The Research and Practice with MicNest

Excerpted from "Acoustic Localization of Drones in Precise Landing: The Research and Practice with MicNest," from SenSys '22: *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems* with permission. https://dl.acm.org/doi/10.1145/3560905.3568515 ©ACM 2022

Delivery drones have the potential to revolutionize transportation and distribution of goods. With their autonomous navigation capabilities, they offer an efficient solution to bypass the challenges posed by complex urban traffic and enable instant package delivery.

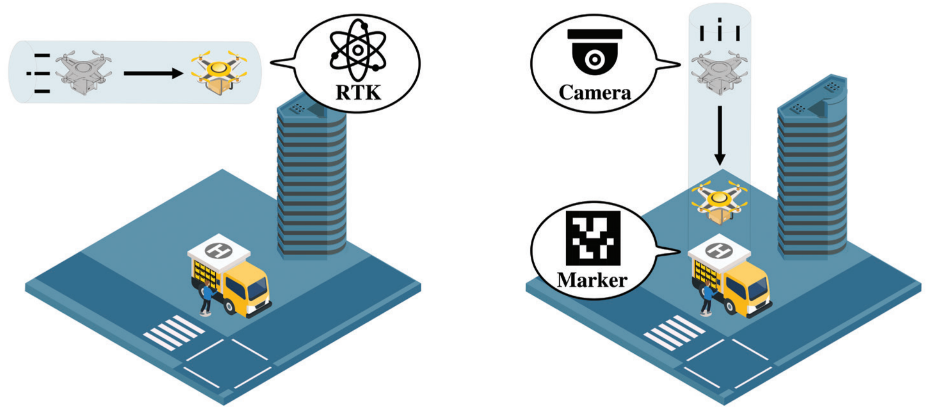
Many industrial firms are actively exploring the commercial feasibility of instant drone deliveries. Meituan, one of the largest companies in this area, devised a systematic approach to instant delivery using drones. The process begins with loading the package onto the drone, which then takes off, ascends to cruising altitude, and sets a direct course towards the designated destination. Typically, this destination is a Meituan-operated self-collection station located near the customer.

As the drone approaches the destination, a precise landing procedure is initiated, as illustrated in Figure 1. Initially, the drone adopts a horizontal approach to achieve vertical alignment with the landing platform, as shown in Figure 1a. Subsequently, it begins the descent, as depicted in Figure 1b. Upon safely docking onto the landing platform, the package is carefully deposited into the self-collection station, allowing the customer to retrieve it at his or her convenience.

Precise landing is extremely important. The consequences of not landing precisely can only be underestimated. Loss of the transported good is obvious but not the most serious result. Should the drone miss the landing platform even partly, it will quickly lose control and crash, possibly damaging surrounding objects or hurting people nearby.

In the traditional approach of instant drone delivery, precise landing relies on two localization techniques: Real-Time Kinematic Positioning (RTK) and visual markers. During the horizontal approach, the drone primarily utilizes RTK for navigation towards the self-collection station. However, in urban areas with high buildings, the GPS signal can be obstructed or reflected, leading to degraded RTK performance, especially within urban canyons. For the vertical descent, a visual marker is placed on the ground, and the drone uses a downward-facing camera to detect the marker and then localize itself [1]. However, this method has multiple limitations: it is sensitive to lighting variations, making marker detection challenging during foggy or nighttime conditions; the camera's field of view may not fully capture the marker, resulting in constrained horizontal coverage; and, as line-of-sight is required between the camera and the marker, the system through-put is limited, preventing concurrent landings of multiple drones.

In summary, the existing techniques cannot provide robust and round-the-clock localization services for delivery drones, which motivates us to explore new technical solutions.



(a) Horizontal approach. (b) Vertical descent.

FIGURE 1. Two phases of drone landing.

LOCALIZING DRONES WITH ACOUSTIC SIGNALS

We present *MicNest*, an acoustic localization system to assist drones in precise landing. As illustrated in Figure 2, a speaker carried by a drone broadcast purposefully designed acoustic pulses. Multiple microphones are deployed on the landing platform in carefully devised configurations. By localizing the speaker from the microphone signals, we localize the drone during the descent. The localization results are transmitted to the drone via WiFi and used as an input for navigation, closing the control loop that drives the drone onto the landing platform.

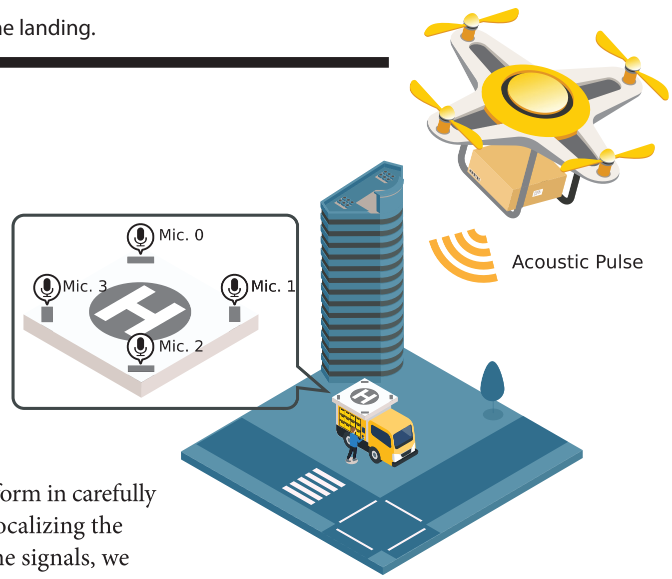


FIGURE 2. MicNest overview.

Why acoustics? MicNest is rooted in the unique features of acoustic signals. The spatial resolution of a signal is proportional to its speed and inversely proportional to its bandwidth. Thus, the slower a signal is, the finer spatial resolution it can provide. Acoustic signals with a limited bandwidth, say 24 KHz, can provide fine-grained spatial resolution, around 0.71 cm when the sampling rate is 48 kHz. In comparison, the spatial resolutions of RF signals like UWB with a 1.3 GHz bandwidth or mmWave with a 4 GHz bandwidth are 10 cm and 3.75 cm, respectively. However, localizing drones via acoustics must tackle three fundamental challenges.

1. The SNR of acoustic pulses is inherently low. The transmission power of the speaker must be limited to avoid acoustic discomfort. MicNest needs to achieve long-range localization, so acoustic pulses experience significant attenuation. Further, background noise in many cities is intrinsically strong (around 40-75 dB SPL) and when airborne, drone propellers generate strong acoustic interference, possibly up to 104 dB SPL.

2. Acoustic signals experience non-linear signal distortions due to Doppler effect. The severity of this effect is inversely proportional to signal speed. Compared to RF signals, the sound speed is much lower. It can be expected that acoustic pulses experience serious distortion when drones are airborne. Modern flight control loops take flight decisions at 400+ Hz, rapidly changing the drone velocity. An acoustic pulse thus experiences various degrees of Doppler effect, ultimately undergoing non-linear distortions.

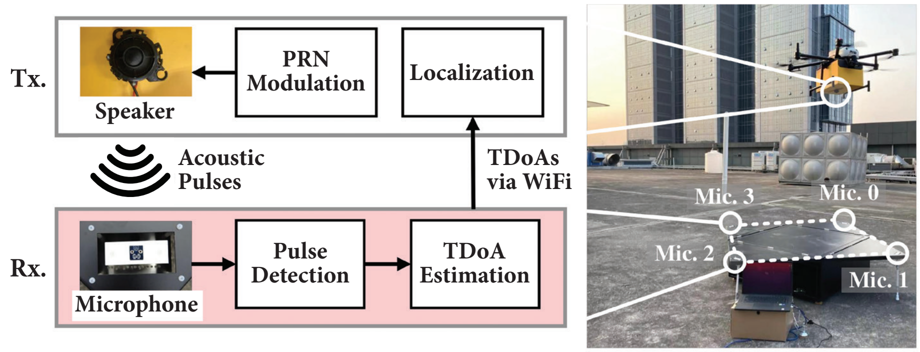


FIGURE 3. MicNest architecture and data pipeline.

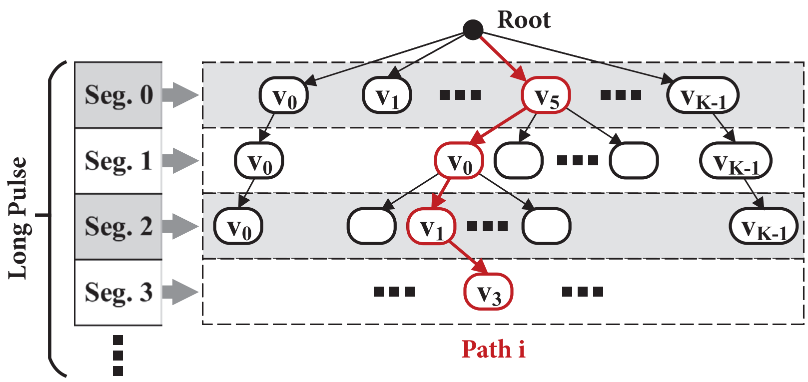


FIGURE 4. Matched Filter Tree.

3. Signal processing must withstand the latency constraint imposed by the real-time nature of flight control loops. The latter consumes location information as one of their most critical inputs. Evidence shows that increasing the latency of location updates may represent a source of system instability. To make things even more challenging, MicNest must provide low-latency location information at a time when the system dependability is most important: during landing.

In short, the technical challenge we face is *how to detect low-SNR pulses with non-linear distortion at an extremely low latency.*

MicNest

In this section, we first introduce the overview of MicNest. Then, we explain how to tackle the above challenges. Figure 3 shows the system architecture of MicNest. Full details are nonetheless available [3, 4].

Overview

PRN modulation. We adopt Pseudo-Random Noise (PRN) modulation to generate the pulses. Specifically, we take a drone's identifier as a random seed and generate a sequence of Gaussian random variables as the codes of the pulse. In MicNest, the code rate equals the sampling rate of the microphones, that is, 48 kHz. The corresponding frequency band of pulses is 0-24 kHz.

The PRN modulation offers three benefits.
1. MicNest can provide concurrent detection and localization of drones. Because PRN pulses are orthogonal if they are statistically independent of each other, we can detect these pulses separately from the collided signal and identify each drone.
2. The use of PRN pulses makes MicNest friendly to the human ear. As drones may operate in populated areas, pulses should not cause acoustic discomfort; but PRN has the same acoustic characteristics as white noise and is almost imperceptible to human ears.
3. MicNest is resistant to impersonation attack, because pulses are (pseudo) randomly generated and it is difficult for third parties to generate the same pulse and impersonate a drone.

Pulse detection. First, we need to detect the pulses from the signals recorded by the microphones. Matched filters are a standard method to detect acoustic pulses. To detect the pulse, the matched filter correlates the transmitted signal with the received one. Ideally, by feeding a stream of the received signal into the matched filter, we obtain a stream of correlations results. Upon observing a correlation peak in the output, we consider a pulse to be detected.

In fact, pulse detection is the most challenging part in MicNest due to low SNR, non-linear distortion, and strict latency requirement. We will elaborate on our solutions next.

TDoA estimation. Once an acoustic pulse is detected, we calculate the times when the pulse arrived at microphones (ToAs) and time difference of arrivals (TDoAs), that is, the differences in ToAs. In MicNest, we compute TDoAs between opposite microphones on the landing platform, that is, <Mic. 0, Mic. 2> and <Mic. 1, Mic. 3> in Fig. 3. This is because opposite microphones have the largest inter-microphone distance and thus yield the largest aperture.

Localization. The TDoAs, are transmitted to the drone, e.g., via WiFi. Based on this information, the drone can establish two hyperboloid equations. Based on the altitude estimated by the on-broad sensors, such as barometers, ultrasound sensors, and downward-facing LIDARs, the drone coordinates can be solved.

TACKLING PULSE DETECTION

As illustrated in Figure 4, the key enabling technology behind MicNest is a novel pulse detector called Matched Filter Tree (MFT), which can detect a low-SNR pulse subject to non-linear distortion.

The key idea we exploit is to model pulse detection as a tree search. Figure 4 illustrates the idea. We split one pulse into M equal segments denoted as Seg. 0, 1, ... M-1. The segments are short enough that the drone velocity can be considered constant within the duration of a segment. Therefore, each segment experiences a linear Doppler distortion.

We build a search tree where the nodes at each level correspond to the possible drone velocities during the transmission of a segment. For each segment, we consider several possible velocities to compensate the Doppler shift it experiences. Ideally, if all segments are compensated with the correct velocities, the new pulse spliced by the compensated segments restores its code synchronization with the received pulse, eliminating the non-linear distortion.

Let velocities $< v^{(0)}, v^{(1)}, ..., v^{(M-1)} >$ be one possible combination of candidate velocities, for example, corresponding to path *i* in Figure 4, where $v^{(m)}$ denotes the candidate velocity for Seg. m. If velocities $< v^{(0)}, v^{(1)}, ..., v^{(M-1)} >$ along path i match the actual drone velocities, the new pulse compensated with these velocities is again synchronized with the received pulse; thus, it has maximum correlation with the received signal because the non-linear distortion is minimized. Therefore, the problem of detecting pulses corresponds to finding a solution path in the search tree that can minimize the non-linear distortion.

TACKLING LATENCY

To enable low-latency localization, we present three techniques to accelerate tree search.

First, we note that the drone velocity does not change abruptly: the velocity of the next segment is unlikely to considerably deviate from that of the current segment. This observation allows us to reduce the branching factor of the tree, that is, the number of candidate velocities, thus abating the processing overhead.

Second, we observe that the pulse-audio correlation can be decomposed into multiple segment-audio correlations. This means that if segment-audio correlations are available, the pulse-audio correlation can be calculated by adding segment-audio correlations. To implement this, a two-step process is applied. In Stage 1, performed before the tree search, all possible segment-audio correlations are calculated and cached. In Stage 2, performed during the tree search, the required segment-audio correlations are retrieved from the lookup table. These correlations are shifted and then added element-wisely. The vector add operation in Stage 2 can be efficiently parallelized. We take advantage of native NVIDIA CUDA kernel to further accelerate Stage 2. In our implementation, searching one tree path takes only 5.3 μs on the NVIDIA RTX 3070, on average.

Third, instead of visiting all tree paths in a brute-force way, we adopt a heuristic method to reduce the total visit count. Our method is like Monte-Carlo Tree Search (MCTS). Our insight is: for each search path, its corresponding maximum correlation values contain the useful information, which can be exploited to guide towards the solution path in the search tree. The path that has a larger correlation value is more likely to be closer to the solution path, and thus the nodes along this path are more promising to be the nodes of the solution path. Therefore, we pay more attention to nodes that look promising to avoid traversing the search tree exhaustively.

EVALUATION

Our experiments show that MicNest can localize a drone 120 m away with 0.53% relative localization error at 20 Hz location update frequency. When navigating drones during landing, MicNest can achieve a success rate of 94%, with an average landing error corresponding to the distance between landing point and target point, of only 4.3 cm. The rest of this section provides experimental evidence.

Implementation

We use drone equipment and a deployment setting that closely mimic actual applications.

Drone. As shown in Figure 5, we use a custom drone manufactured by Meituan. The drone is equipped with six propellers, each hooked to a brushless TMotor and steered by the PX4 flight controller. The drone has a payload capacity up to 2.6 kg at liftoff, which is the maximum load that local regulations allow. The only additional equipment is the speaker, which is attached to the bottom of the drone. We use a VISTEON speaker weighing a mere 47 g. The speaker volume is empirically set to 70-75 dB SPL (measured at 1 m distance), which is arguably moderate.

Landing platform and software. We build a squared foldable landing platform, shown in Figure 6, measuring 1 x 1 m and 1.41 x 1.41 m when folded or unfolded, respectively. Four omni-directional digital microphones are installed at the corners. The distance between two diagonal microphones is 1.86 m.

We use an XMOS XU216 data acquisition board to drive and sample the microphones, so that the four signals are synchronized. The sampling rate is 48 KHz. The board then streams the audio signals to a laptop via USB UAC 2.0 with a latency lower than 0.5 ms.

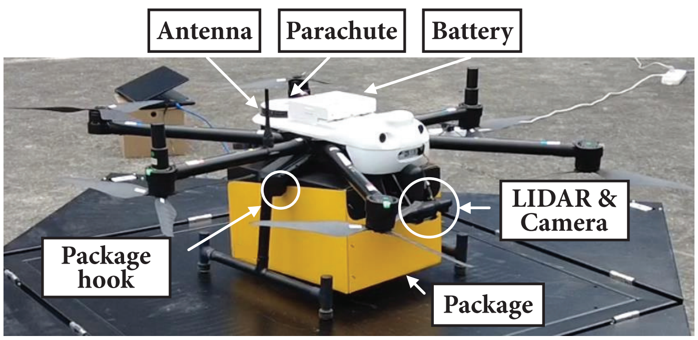


FIGURE 5. The delivery drone made by Meituan.

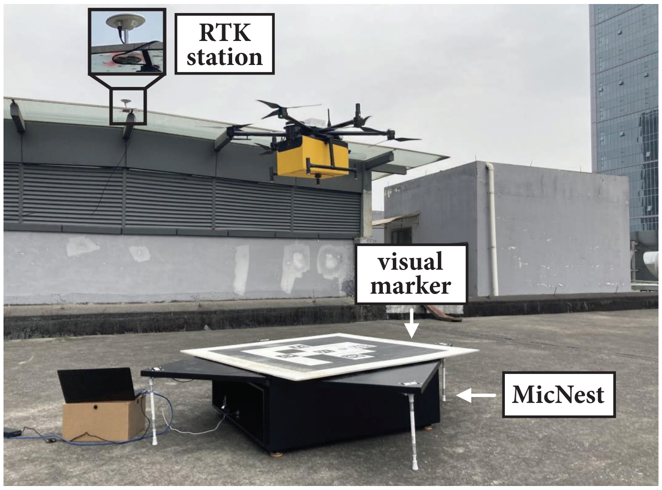
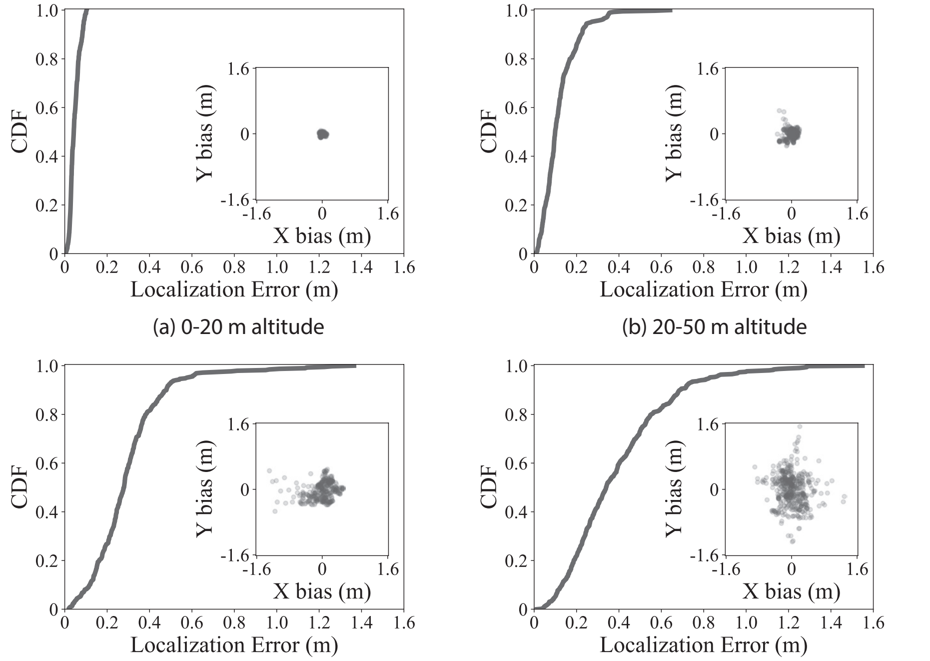


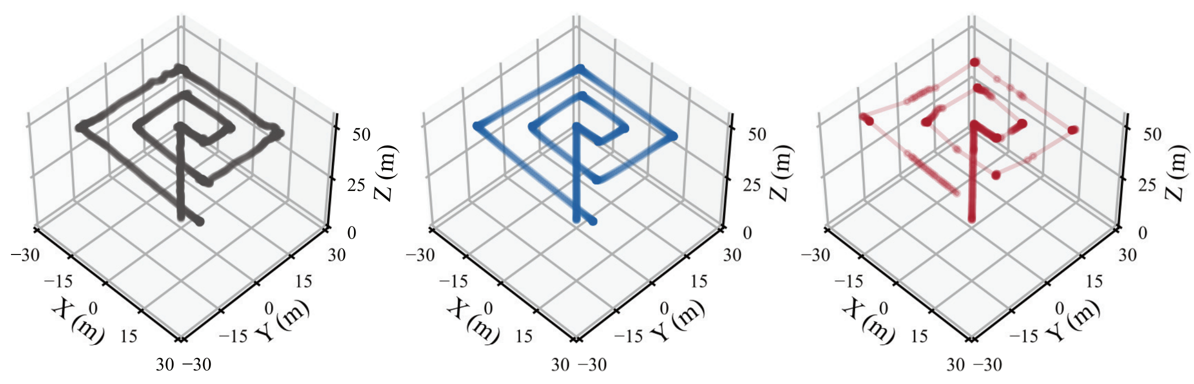
FIGURE 6. Experimental setting.



(a) 0-20 m altitude
(b) 20-50 m altitude
(c) 50-80 m altitude
(d) 80-120 m altitude

FIGURE 7. Localization error compared to altitude.

MICNEST IS CURRENTLY UNDERGOING ACTIVE DEVELOPMENT AND MAKING SIGNIFICANT STRIDES. IT HAS SUCCESSFULLY BEEN INTEGRATED INTO THE FLIGHT CONTROL SYSTEM OF MEITUAN DRONES, MARKING A MAJOR MILESTONE



(a) MicNest
(b) RTK
(c) Visual Marker

FIGURE 8. Drone trajectories localized by MicNest, RTK, and using the visual marker. (We provide an illustrative video of this experiment and of the results at other altitudes on our website) [2].

We use a high-pass filter with a cutoff frequency of 500 Hz to pre-process the audio signals. MFT is implemented in C++ with the CUDA 11.0 library, running on a machine with an Intel i9-11900H CPU, 32 GB memory, and an NVIDIA RTX 3070 GPU.

Key Results

Localization accuracy. We program the drone to perform a vertical flight up to 120 m. We plot the cumulative distribution function (CDF) of the localization error, compared to RTK that represents ground truth, at different altitude intervals in Figure 7a-d. Note that MicNest calculates the horizontal coordinates of the drone at a given altitude. The scattered plot in each figure shows the localization biases of MicNest.

When the altitude is below (above) 20 m (80 m), the median error is 0.043 m (0.339 m). On average, the relative error, that is, the absolute localization error over the distance to the platform, is only 0.53%.

Localization trajectory. As an example, Figure 8a-c show the localization results as the drone flies a squared spiral at 50 m altitude. An illustrative video is available [2]. The plots demonstrate how MicNest and RTK successfully localize the drone throughout the whole flight. In contrast, the visual marker works intermittently because it is difficult for the camera on the drone to capture the visual marker, especially at higher altitudes. Results at different altitudes are nonetheless available on our website [2].

Drone landing. Guiding the drone onto the landing platform accurately and robustly is the ultimate design goal of MicNest. In this experiment, we feed the localization results of MicNest to the flight controller of the drone. For each experiment, the drone first flies to an altitude of 120 m and begins to land. During drone landing, the flight controller is forced to rely solely on the localization results of MicNest to navigate the drone onto the center of the landing platform.

We consider the landing operation as successful whenever the drone hits the center of the landing platform with a maximum error of 10 cm. This is smaller than the frame size of the drone and an acceptable margin of error for applications, such as drone deliveries, where the drone drops a packet after landing or performs actions that require aligning the drone with the landing platform. We repeat this experiment 50 times to gather statistically relevant metrics.

Figure 9 reports the results. MicNest navigates the drone onto the landing platform with a success rate of 94%. Only 3 cases of failure exist, and all of them are not caused by MicNest. Two of them are caused by the WiFi connection loss and one is caused by an error independent of MicNest, as the object avoidance module was mistakenly triggered.

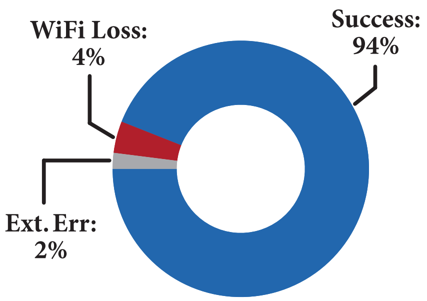


FIGURE 9. Statistics of landing results.

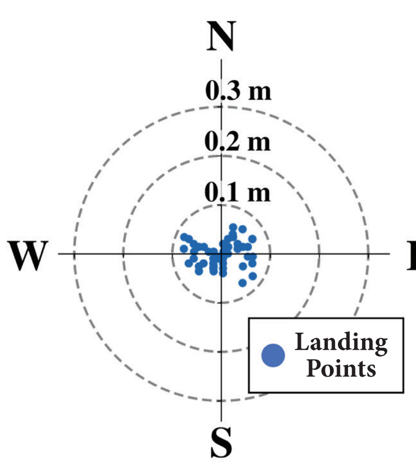


FIGURE 10. Distribution of landing points.

Figure 10 further illustrates the spatial distribution of landing points. The average landing error is only 4.3 cm.

A demonstration video that shows the process of navigating drone landing is available [2]. Refer to our paper [3,4] for detailed evaluation results.

ONGOING AND FUTURE WORK

MicNest is currently undergoing active development and making significant strides. It has successfully been integrated into the flight control system of Meituan drones, marking a major milestone. We remain dedicated to accumulating further test results to verify that MicNest is ready to deliver reliable and dependable performance for commercial use. Extensive flight tests, totaling over 200 hours, have already been conducted in various scenarios, including challenging conditions like night, rain, and fog.

In terms of the on-board components, we are currently designing and implementing a speaker box, which supports plug-and-play usage for the drones. Additionally, for the ground part of the system, our plan is to incorporate the distributed microphones and the computing module into Meituan self-collection station by the end of 2023. ■

Weiguo Wang received his PhD from Tsinghua University and his BE from University of Electronic Science and Technology of China (UESTC). His research interests include acoustic sensing and mobile computing. wwg18@mails.tsinghua.edu.cn

Yuan He is an associate professor in the School of Software and BNRist of Tsinghua University. He received his PhD from Hong Kong University of Science and Technology. His research interests include wireless networks, Internet of Things, pervasive and mobile computing. heyuan@tsinghua.edu.cn

Luca Mottola is a full professor at Politecnico di Milano (Italy) and a Senior Researcher at RI.SE Sweden. His lab develops new technologies at the frontier of Internet of Things, including mobile embedded computing, intermittent computing, nanosatellites, and Internet-connected robotics. luca.mottola@polimi.it

Shuai Li works as an algorithm engineer at Meituan. He received his BE and MS degree from Tsinghua University. His research interests include Internet of Things and wireless sensing lis20@mails.tsinghua.edu.cn

Yimiao Sun is currently a PhD student at Tsinghua University. He received his BE from the University of Electronic Science and Technology of China (UESTC) in 2021. His research interests include Internet of Things and wireless sensing. sym21@mails.tsinghua.edu.cn

Jinming Li is currently a PhD student at Tsinghua University. He received his BE degree from the University of Electronic Science and Technology of China (UESTC) in 2021. His research interests include Internet of Things and wireless sensing. li-jm19@mails.tsinghua.edu.cn

Hua Jing works as a software director at Meituan. He received his ME and BE from Beihang University. His research interests include heterogeneous computing and realtime communication. jinghua03@meituan.com

Ting Wang works as a system developer at Meituan. She received her BE degree from Sichuan University and her ME degree from Tsinghua University. She works in the fields of UAV and RFID. wangting63@meituan.com

Yulei Wang works as a software architecture manager at Meituan. He received his BE and ME from Beihang University. He works in the fields of UAV and robot. wangyulei02@meituan.com

REFERENCES

[1] OpenCV, "Detection of aruco markers," https://docs.opencv.org/4.x/d5/dae/tutorial_aruco_detection.html, 2022.

[2] MicNest, Landing page of micnest. https://micnest.github.io, 2023.

[3] W. Wang, L. Mottola, Y. He, J. Li, Y. Sun, S. Li, H. Jing, and Y. Wang. 2022. Micnest: Long-range instant acoustic localization of drones in precise landing. *Proceedings of ACM SenSys*, Boston, CT, USA.

[4] Y. He, W. Wang, L. Mottola, S. Li, Y. Sun, J. Li, H. Jing, T. Wang, and Y. Wang. 2023. Acoustic localization system for precise drone landing. *IEEE TMC.* DOI: 10.1109/TMC.2023.3288299