2023 International Conference on Unmanned Aircraft Systems (ICUAS)
June 6-9, 2023 | Lazarski University, Warsaw, Poland

2023 International Conference on Unmanned Aircraft Systems (ICUAS) | 979-8-3503-1037-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICUAS57906.2023.10155866

Deep Learning–based Reassembling of an Aerial & Legged Marsupial Robotic System–of–Systems

Prateek Arora*, Tolga Karakurt, Eleni Avlonitis, Stephen J. Carlson, Brandon Moore, David Feil-Seifer, Christos Papachristos

* *Consider for Best Student Paper Award*

***Abstract*— In this work we address the System-of-Systems reassembling operation of a marsupial team comprising a hybrid Unmanned Aerial Vehicle and a Legged Locomotion robot, relying solely on vision-based systems and assisted by Deep Learning. The target application domain is that of large-scale field surveying operations under the presence of wireless communication disruptions. While most real-world field deployments of multi-robot systems assume some degree of wireless communication to coordinate key tasks such as multi-agent rendezvous, a desirable feature against unrecoverable communication failures or radio degradation due to jamming cyber-attacks is the ability for autonomous systems to robustly execute their mission with onboard perception. This is especially true for marsupial air / ground teams, wherein landing onboard the ground robot is required. We propose a pipeline that relies on Deep Neural Network-based Vehicle-to-Vehicle detection based on aerial views acquired by flying at typical altitudes for Micro Aerial Vehicle-based real-world surveying operations, such as near the border of the 400ft Above Ground Level window. We present the minimal computing and sensing suite that supports its execution onboard a fully autonomous micro-Tiltrotor aircraft which detects, approaches, and lands onboard a Boston Dynamics Spot legged robot. We present extensive experimental studies that validate this marsupial aerial / ground robot's capacity to safely reassemble while in the airborne scouting phase without the need for wireless communication.**

I. INTRODUCTION

Autonomous robots have become a driving force in research and in industry over the recent decades, with numerous application domains including personal use, but more importantly search and rescue [1, 2], industrial inspection of civilian infrastructure [3–7], and exploration of both terrestrial challenging environments [8–13] and of extra-planetary worlds as well [14, 15]. Interestingly, the potential for collaborative heterogeneous System-of-Systems deployments that has long been investigated in multiagent research, has recently started to show its advanced capabilities [16–19] when dealing with real-world environments and challenges. At the same time, small-scale robotic solutions continue to widen the operational envelope of their missions, by taking on tasks such as airborne surveillance over large-scale field locations in-the-wild ([20–23]) and presenting capabilities such as long-term multi-day autonomy through self-sustained landing and recharging cycles [24].

Fig. 1. Field demonstration of the Marsupial System-of-Systems autonomous reassembling operation for a micro-sized Hybrid Aerial & Legged Locomotion robot.

In this context, seeking to facilitate wide-field robotic missions with marsupial Aerial & Legged (ground) System-of-Systems that can be executed with long-term resilience, we identify the need for reliable air-to-ground reassembling that can be executed without relying on persistent Vehicle-to-Vehicle wireless communication. The addressed scenarios are these of performing reassembling at *predetermined-but-approximate rendezvous* locations under communications failure due to unforeseen subsystem malfunctions, cyber-warfare frequency jamming, or the need to operate "silently" w.r.t. to the systems' wireless spectrum footprint. Thus, this paper proposes and experimentally validates an approach which relies on a vision-only paradigm assisted by an appropriately trained Deep-Learned framework and a custom-designed Fiducial-Marker setup and proper sensor selection and marshalling on the aerial robot side. The goal is to achieve airborne detection of the Legged system at various altitude scales and possible backgrounds corresponding to different field environments, as well as systematic reassembling approach, landing, and docking, with zero communication taking place between both systems. The proposed visual detection system is studied w.r.t. its operating performance in varying real-world field conditions and scales, and is finally experimentally demonstrated w.r.t. its effectiveness to systematically achieve the air-to-ground reassembling operation.

The remainder of this paper is structured as follows: Section II discusses relevant prior work in the field. Section III overviews our proposed approach for the communication–less reassembling of the Aerial & Legged System-of-Systems in large-scale field deployments. Actual system performance

This material is based upon work supported by the NSF Awards: *2148788: EPSCoR RII Track-1: Harnessing the Data Revolution for Fire Science*, and *IIS-2150394: Research Experiences for Undergraduates Site: Collaborative Human-Robot Interaction for Robots in the Field*. The presented content and ideas are solely those of the authors.

The authors are with the University of Nevada, Reno, 1664 N. Virginia, 89557, Reno, NV, USA prateeka@nevada.unr.edu

specifics are elaborated in Section IV. Experimental results presenting the autonomous reassembling operation are shown in Section V, and our conclusions are drawn in Section VI.

II. RELATED WORK

In the field of robotics, the rendezvous problem –and more specifically the close-proximity coordination with possible collaborative physical attachment (i.e. System-of-Systems assembling)– has been extensively studied in multi-robot deployments of Unmanned Ground Vehicles (UGVs) and Unmanned Air Vehicles (UAVs) [2, 25–32]. The works in [33, 34] focus on generating pre-designed ground vehicle waypoints and aerial vehicle routing with specific endurance optimization objectives (e.g., recharging), and [35] presents strategies for collaborative battery swapping scheduling, while others [36] focus on system development to achieve UGV-to-UAV marsupial powering. HALE UAV [32] demonstrated GPS-assisted fixed-wing touchdown on a wheeled mobile platform where the UGV has to be manually accelerated and aligned with the aircraft for successful landing. Even if the limiting factor of human intervention is disregarded, most approaches to some extent rely on the requirement for –at least– intermittent wireless communication between the air vehicle and the ground platform.

To achieve airborne tracking with the purpose of reassembling an Aerial & Ground robot team, [31] proposes a strictly vision-based rendezvous cone-guidance scheme using a monocular camera. Another established method is to employ visual servoing relying on Fiducial markers for vision-based landing [37, 38]. Such approaches remain limited in terms of operating altitude ranges, due to the reduced accuracy of the derived distance estimates, or the complete failure at detecting any realistically-sized marker that could be placed on a ground system.

Moreover, the ground system type of choice –wheeled robots–, is limited to traversing relatively flat terrains; Legged Locomotion systems have demonstrated their vast superiority in negotiating unstructured real-world terrains [39] in the wild. Even though such systems have been combined with multicopters [40], it is commonly assumed that networking connectivity between agents is a constantly present facilitating factor for the considered tasks.

In case of unforeseen communication outage (subsystems failure, cyberattacks, requirement for "bandwidth-quiet" operation" e.g., to prevent mission hijacking [41]), our approach can still facilitate the air-to-ground reassembling by relying on strict visual and onboard perception and state estimation [42, 43], given an approximately know rendezvous region at realistic flight altitudes for wide-area surveying micro-sized UAVs. This also surpasses the capacity of similar concepts [44] which employ multicopters tailored to constrained environments, due to the hybrid flight envelope offered by a VTOL / Fixed-Wing micro aerial robot, with the additional capacity to be recharged in-the-field while docked onto the ferrying legged system [24].

III. PROPOSED APPROACH

This section details the primary components of the proposed communication-less air-to-ground Reassembling.

A. Autonomous Hybrid Micro Aerial System

The flying platform used is the MiniHawk–VTOL [45, 46], a rapidly–prototyped fixed–wing VTOL aircaft designed with the focus on adaptability for research and ease of manufacture. The aircraft has a 800mm wingspan, wing area of 17dm^2, an all–up–weight of 1100g to 1400g, and can sponsor a variety of sensory devices and compute elements. Here we use the Intel©T265 VIO sensor, a USB Webcam, a Benewake©TFMini Plus micro 1D LiDAR sensor, and the mRo PixRacer Pro flightstack with its accompanying Magnetometer and IMU suite. These systems are used by the Khadas VIM3 single–board computer for GPS–denied navigation in an unstructured outdoor environment.

B. Legged Locomotion System

The other key unit of our marsupial system–of–systems is the Boston Dynamics Spot®, a 12–DoF quadrupedal robot with 14kg payload capacity that offers a runtime of 90 minutes. Spot carries an in–house designed docking and recharging backpack (DRB) that is rigidly mounted to the rails on its back with a passive self–centering design that allows the MiniHawk–VTOL to align itself by centering and sliding towards the edge once it has landed on the backpack. Additionally, the backpack houses three sets of actuated claws that engage to latch onto the skids of the MiniHawk–VTOL to hold it firmly in place during any dynamic motion of the Spot locomotion as well as provide electrical contact for rapid charging to replenish the energy reserves of the MAV. Affixed at the center of the backpack is a small Fiducial-Tag of dimensions $5.6\text{cm}\times5.6\text{cm}$ allowing for visual localization of the backpack for precision landing.

C. Air-to-Ground System Reassembling

In this work, we address the problem of reassembling an aerial and a ground robotic unit in the context of deployment of the marsupial system–of–systems performing a large–scale surveillance mission that allows for a significant extension of the MAV's operational capacity via repeated docked–recharging offered by the legged robot. More specifically, we aim to perform the reassembling by leveraging onboard perception without the need for any wireless communication between the team's agents in an environment possibly affected by radio degradation or wireless signal jammers. We consider an autonomously surveying MAV flying in a fixed–wing configuration following an arbitrary path over unmapped and unstructured terrain as well as the legged robot present in the vicinity over the ground terrain executing its independent mission objective. Our proposed pipeline executes as soon as the surveying MAV's energy reserves decline past a threshold, at which point, the MAV begins to find the Spot legged robot via the aerial imagery captured by the onboard color-camera and subsequently land on it to leverage fast charging.

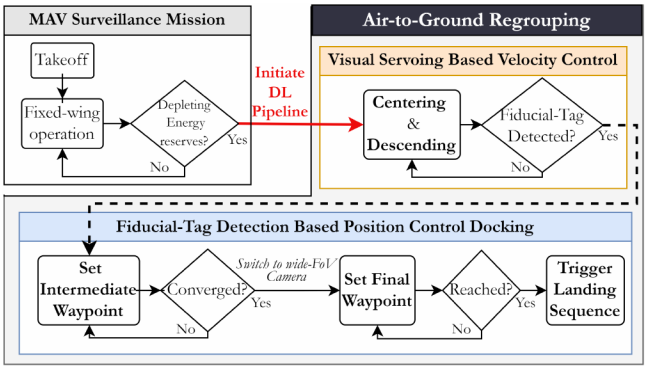

Fig. 2. Overview of the Air-to-Ground Reassembling algorithm based on velocity and position control components.

We present a vision–based radio communication–less approach to address the problem, comprising of two phases, namely a) *visual servoing–based velocity centering*, that aims to align the geometric center of detected Spot robot with the center of the image and descending towards it, and b) *Fiducial-Tag detection–based position control docking* comprising the last leg of docking with the objective of performing precise landing on the backpack. The former phase continues to center the MAV over the Spot and descent until the first instance of the Fiducial-Tag detection occurs, and subsequently switches over to the later phase to complete the landing. Figure 2 presents the flow chart of the algorithmic components of our approach which are elaborated followingly.

i) *Visual Servoing Based Velocity Control*

The first component of our work's contribution is a pipeline that leverages Deep Learning framework to detect and localize Spot by regressing four corner points to obtain a bounding box around it in the image plane as well as provide a confidence score for each detection. Occasionally, the network may provide multiple candidate detections of Spot and we select the best candidate with the most superior detection confidence in addition to a lower bound threshold of 60%. The network provides us with an accurate 2D location, however, we lack the scale information required to get a full 3D pose of the target object, essentially rendering it impossible to employ a full position control policy. Instead, we leverage visual servoing policy with the objective of aligning the bounding box center $p = \{x_{bb}, y_{bb}\}$ with the image center $p_T = \{x_{img}, y_{img}\}$ making it perfectly suitable to our use case. We then define an error in pixels and compute a velocity vector based on a proportional–derivative controller in the color-camera coordinate frame, $\mathcal{F}_C$ as follows:

$$e_p = p - p_T \tag{1}$$

$$V = K_p e_p + K_d \dot{e_p} \tag{2}$$

$$V_C = min(max(V, -V_{max}), V_{max}) \tag{3}$$

where V_{max} and $-V_{max}$ are the upper and lower bounds for the resultant velocity vector respectively. Given a body frame of reference, $\mathcal{F}_B$ rigidly attached to the center of the aircraft, we transform the computed velocity vector (V_C) in the body frame of reference to get a resulting velocity vector (V_B).

With the above policy, an aggressive velocity vector may lead the vehicle to excessively roll or pitch causing the narrow-FoV color-camera to lose sight of the Spot. To overcome this, we further condition the velocity vector V_B to compensate negatively for greater roll and pitch angles.

$$V^i_{des} = \begin{cases} cos(\alpha)V^i_B & if\ V^i_B > 0\ \&\ \alpha > 0\ or \\ & V^i_B < 0\ \&\ \alpha < 0 \\ V^i_B & otherwise \end{cases} \tag{4}$$

with $\alpha = \phi, i = y$ or $\alpha = \theta, i = x$, where ϕ & θ are roll and pitch angles, and the superscript i represents the x or y component of the velocity vector. Ideally, the system converges to a steady state when the error becomes zero, however, achieving zero pixel error becomes a near impossible task given the non–linearity of the system, especially when the detection occurs at high altitudes producing relatively small bounding boxes. Therefore we consider that the convergence is achieved when the error in pixels e_p reduces below a threshold. Furthermore, since the size of the bounding box varies largely with the altitude at which detection takes place there is no single threshold value that defines convergence. We proceed to discretize the bounding box into seven bins $[< 50, < 120, < 200, < 300, < 400, < 600, \& < 900]$ and get a corresponding convergence threshold τ_i, derived empirically. For a discretized bounding box, we determine that convergence is achieved if $e_p < \tau_i$, essentially marking the velocity centering process as complete and proceeding to the vertically descend in altitude.

The velocity centering policy aligns the image center with the center of Spot detection in the color-camera frame of reference, essentially making the vehicle hover above the Spot at a certain altitude. The next reasonable step is to simply descend while maintaining the same position in the horizontal plane. However, given the constraints of the vehicle, designed to hover with a positive pitch of 10 degrees with the nose pointing slightly, it hovers with an offset in the horizontal x–y plane rather than hovering perfectly above the Spot, rendering the policy of straight–forward vertical descent ineffective, essentially driving the Spot away from the color-camera's FoV. We instead proceed with an iterative policy that continuously performs centering, and descends vertically once the centering is achieved, followed by the centering again once the pixel error e_p grows beyond the convergence threshold τ_i and so on. This pipeline guides the MAV to align and descend itself closer to the Spot and stops executing as soon as the first instance of the Fiducial-Tag is detected in the same narrow-FoV color-camera, and proceeds to the next phase for the final descent.

ii) *Fiducial-Tag Detection Based Position Control Docking*

This section presents the final phase of docking that allows for precision landing by leveraging real–time 3D pose estimates obtained from detecting the Fiducial-Tag present on the backpack. We define a state vector $\xi^W_B = [x_B, y_B, z_B, \psi_B]^T$ representing the 3D position of the

center–of–mass of the aircraft and its heading vector as well as another vector $\xi_T^W = [x_T, y_T, z_T, \psi_T]^T$ representing the estimated Fiducial-Tag 3D position and yaw angle, both expressed in an inertially aligned world frame of reference $\mathcal{F}_W$. Ideally, the estimated Fiducial-Tag state can be forwarded as a reference command to the low-level position controller to achieve docking, however, large positional tracking error can cause aggressive overshoot which entails the need for a high–level position commander. We employ the carrot–chasing algorithm that essentially conditions a predefined trajectory by following an incremental virtual waypoint rather than a distant one to avoid overshooting and obtain a smooth system response. Mathematically, we define the virtual carrot waypoint as $\mathcal{W} = \xi_B^W + \hat{n}l$, with $\hat{n}$ being a unit vector in the direction of the vector $v = \xi_T^W - \xi_B^W$ and l being a tunable parameter that controls the spacing between the current and the virtual waypoint.

At the same time, the pose of the Fiducial-Tag is continuously estimated from the small-FoV color-camera. Evidently, the detection accuracy improves with greater scale, i.e. the pose estimates get more accurate as the tag gets closer. This requires us to follow a particular trajectory profile such that the tag remains in the camera view even with the roll and pitch motion of the vehicle along the path. Given the slight nose-up hover stance of the vehicle, we achieve the profiling by selecting two waypoints, a) an intermediate waypoint ξ_{Ti}^W at an offset of a meter above the tag, and b) a final waypoint ξ_{Tf}^W with an offset of a 0.25m above the tag. Overall, the vehicle is first commanded to reach the intermediate waypoint ξ_{Ti}^W and after converging, subsequently aims to reach the final waypoint ξ_{Tf}^W. Both ξ_{Ti}^W and ξ_{Tf}^W are considered reached when the following convergence criteria are met:

$$|e_x| < \tau_x, |e_y| < \tau_y, (|e_z| < \tau_z \;||\; z < z_{ref}), |e_\psi| < \tau_\psi, \quad (5)$$
$$|v_x| < \tau_{v_x}, |v_y| < \tau_{v_y}, |v_z| < \tau_{v_z}, |v_\psi| < \tau_{v_\psi},$$

The convergence threshold values τ_i for the final waypoint ξ_{Tf}^W, are stricter than those for the intermediate waypoint ξ_{Ti}^W, i.e. the MAV follows a smooth trajectory profile as it passes through ξ_{Ti}^W and as it reaches ξ_{Tf}^W the vehicle is constrained to maintain small positional and velocity errors. This ensures that the MAV is positioned to hover directly above the tag, followed by triggering a landing command on the authority of the onboard computer allowing it to perform a vertical descent until the skids come in contact with the DRB and finally disarms automatically. Lastly, the rear switch on the DRB is activated by the impact force of the MAV sliding onto it, engaging the claws to grip the vehicle and hold it in place.

It is highlighted here, that we use both the small-FoV color camera and the wide-angle fisheye camera for estimating the Fiducial-Tag pose. For real-time deployment and computational efficiency, we switch between the two camera image streams instead of detecting the tag with both at the same time, significantly speeding up the system performance. The switch happens as soon as the intermediate waypoint ξ_{Ti}^W is reached. We leverage the complementary benefits provided by the two cameras, i.e. the narrow-FoV camera with a larger focal length provides better pose estimates from greater distances while the fisheye camera provides a wider view of the scene from near range. This behavior is extremely desirable to get continuous estimates of the tag detection, especially near the last leg of the landing, as the vehicle has the tendency to drift due to external disturbances such as ground effect and/or wind gusts.

iii) *Deep Learning Network Training and Implementation*

For the purposes of Spot detection we leverage the YOLOv3 [47] Deep Learning framework, which is capable of discerning multiple instances of object classes in aerial images with densely packed, distributed with large–scale variation. In this work, we train the network on our custom dataset containing aerial images of the Spot robot by leveraging transfer learning approach. The dataset was collected over a period of a few months to incorporate images with different environments, seasons, multiple altitudes as well as varying lighting conditions, containing over 3000 images. In order to deploy the trained model on the NPU on-board Khadas VIM3, it is required to be quantized, which refers to the techniques of converting the floating point weights to lower bandwidths such as integers. This allows for a more compact representation of the model, without compromising inference time accuracy while significantly reducing the computational cost.

IV. SYSTEM PERFORMANCE

In this section, we discuss the performance of the trained Deep Learning network across various image scales and backgrounds, essentially demonstrating the efficacy of the deployed network for our envisioned task.

A. Deep-Learned Detection Statistics I

Table I provides the precision, recall, mean-Average-Precision (mAP), and f1 score for detections across different altitudes measured in meters above the ground level (AGL) by the proposed Deep Learning architecture, essentially demonstrating its performance at various scales. The network exhibits superior performance with high precision as well as recall scores, indicating its ability to generalize across various scales. It can be observed that the accuracy of detections decreases with the rise in altitude, which is an expected behavior as very few pixels belonging to the relevant class category are observed. Figure 3 depicts a few instances of Spot detections at 20m, 40m, and 60m AGL altitudes. It is noted here that the network is able to get accurate results at the 60m altitude mark which is almost imperceptible even to the human eye.

B. Deep-Learned Classification Statistics II

In this sub-section we discuss the performance of the network in detecting Spot in different environments, and against different backgrounds. Table II provides precision, recall, mAP, and f1 scores for Spot detection across different backgrounds encountered in the field deployment such as asphalt, snow, mud, and bushes, and Figure 4 depicts few

TABLE I

PERFORMANCE OF THE DEEP LEARNING DETECTION NETWORK AT DIFFERENT ALTITUDES ABOVE GROUND LEVEL

Altitude	precision	recall	mAP	f1
20 m	0.893	0.898	0.893	0.895
30 m	0.865	0.896	0.861	0.880
40 m	0.884	0.893	0.880	0.888
50 m	0.813	0.844	0.813	0.828
60 m	0.796	0.787	0.796	0.791

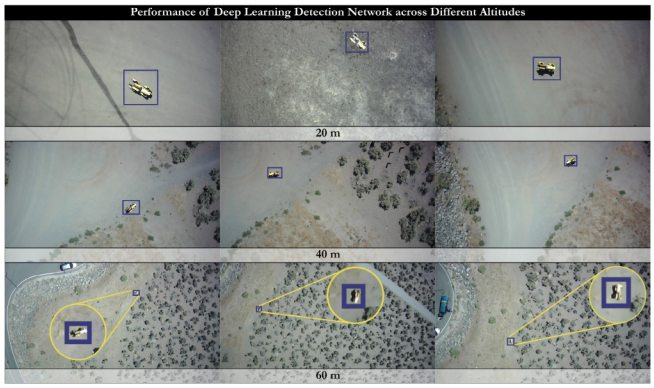

Fig. 3. The figure depicts a few instances of Spot detection across different altitudes of 20m, 40m, and 60m. The detection remains accurate across different scales with high confidence.

such instances. Overall, the network consistently performs with high accuracy, with the lowest f1 score of 0.864 for asphalt and the highest f1 score of 0.946 for the environment with bushes.

TABLE II

PERFORMANCE OF THE DEEP LEARNING DETECTION NETWORK ACROSS DIFFERENT ENVIRONMENTS

Background	precision	recall	mAP	f1
asphalt	0.887	0.843	0.8875	0.864
snow	0.894	0.869	0.879	0.881
bushes	0.905	0.992	0.904	0.946
mud	0.878	0.925	0.897	0.901

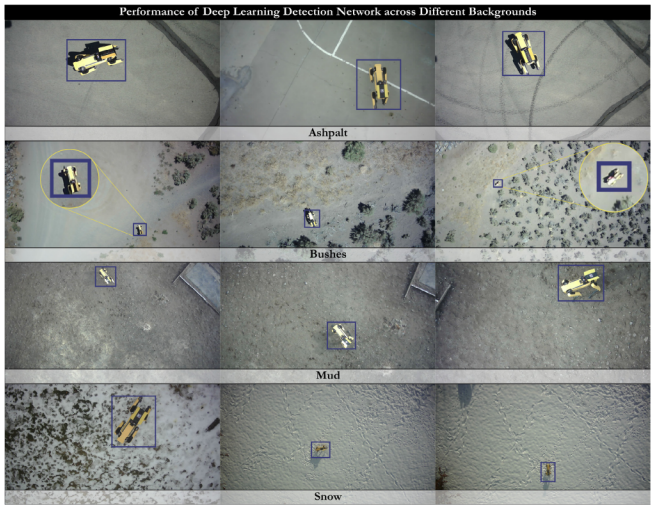

Fig. 4. Instances of Spot detection across different backgrounds such as asphalt, presence of bushes, mud, and snow. It is noted here that the network detects accurately against partly snowy/icy (bottom left) as well as completely snow-covered (bottom right) backgrounds.

V. EXPERIMENTAL STUDY

We experimentally tested the proposed system-of-systems' autonomous air-to-ground reassembling functionality in real-world field experiments. The experiments were conducted in freezing-cold temperatures, and it is noted that the initiating height in the demonstrated sequences is intentionally limited (although higher altitudes can still yield good detections as shown in the system performance results shown of the previous Section), due to flight-time safety considerations for the particular operating area. The corresponding results are illustrated in Figure 5.

A. Nominal Reassembling

The first set of rows in Figure 5 demonstrates the nominal "smooth" operation envisioned for the proposed system-of-systems.

In the upper *Yellow box* row, a visualization of the mission progress is given, showcasing: i) Transitioning out of the initial visually-servoed guidance phase (sequence of small transform axes shows the vehicle trajectory history) as soon as the Fiducial-Tag is initially detected (large transform axis in the ground) in the short-FoV color-camera. Also here the sequence of approach waypoints (yellow arrows) leading to the intermediate waypoint is shown. ii) The landing approach history until the intermediate waypoint (determined by continuously estimating the Fiducial-Tag pose via the color-camera) is achieved. iii) Arriving at the final waypoint (determined by estimating the Fiducial-Tag pose via the fisheye-camera), right above the Docking & Recharging Backpack. iv) Touchdown onto the pad. v) Indicative illustration of the final pose of the aerial vehicle after touchdown w.r.t. the (last-estimated) Fiducial-Tag pose. The well-aligned outcome is achieved due to the DRB passive docking design.

In the following row, the *Blue box* illustrates the views observed by the short-FoV color-camera, and the ongoing detection of the Legged Robot from the air. At the same time, the Fiducial-Tag attached on the Docking & Recharging Backpack, which is used to determine the intermediate approach waypoint, is also shown. The *Red box* shows the corresponding views from the fisheye camera. It is evident how its utility is at close-up distances, where it excels at acquiring consistent views of the Fiducial-Tag while hovering over and around the pad, due to its wide Field-of-View.

The plots illustrate the time evolution of the: i) z-altitude state, the ii) x, y-displacement, and finally the iii) ψ orientation during the sequence. The illustrated $x^{ref}, y^{ref}, z^{ref}, \psi^{ref}$ values correspond to the intermediate (or the final) waypoint above the pad, as determined by the Fiducial-Tag pose estimation. These start to appear as soon as the Fiducial-Tag is initially detected after visual servoing. It can be observed that initially the pose is unstable (due to using the short-FoV color-camera from a larger distance), while once the switchover to the fisheye-camera and the final waypoint occurs (at which point there is a "jump" particularly observable in the z^{ref} value) it becomes significantly more consistent. The time of the touchdown command is when the $x^{ref}, y^{ref}, z^{ref}, \psi^{ref}$ values disappear.

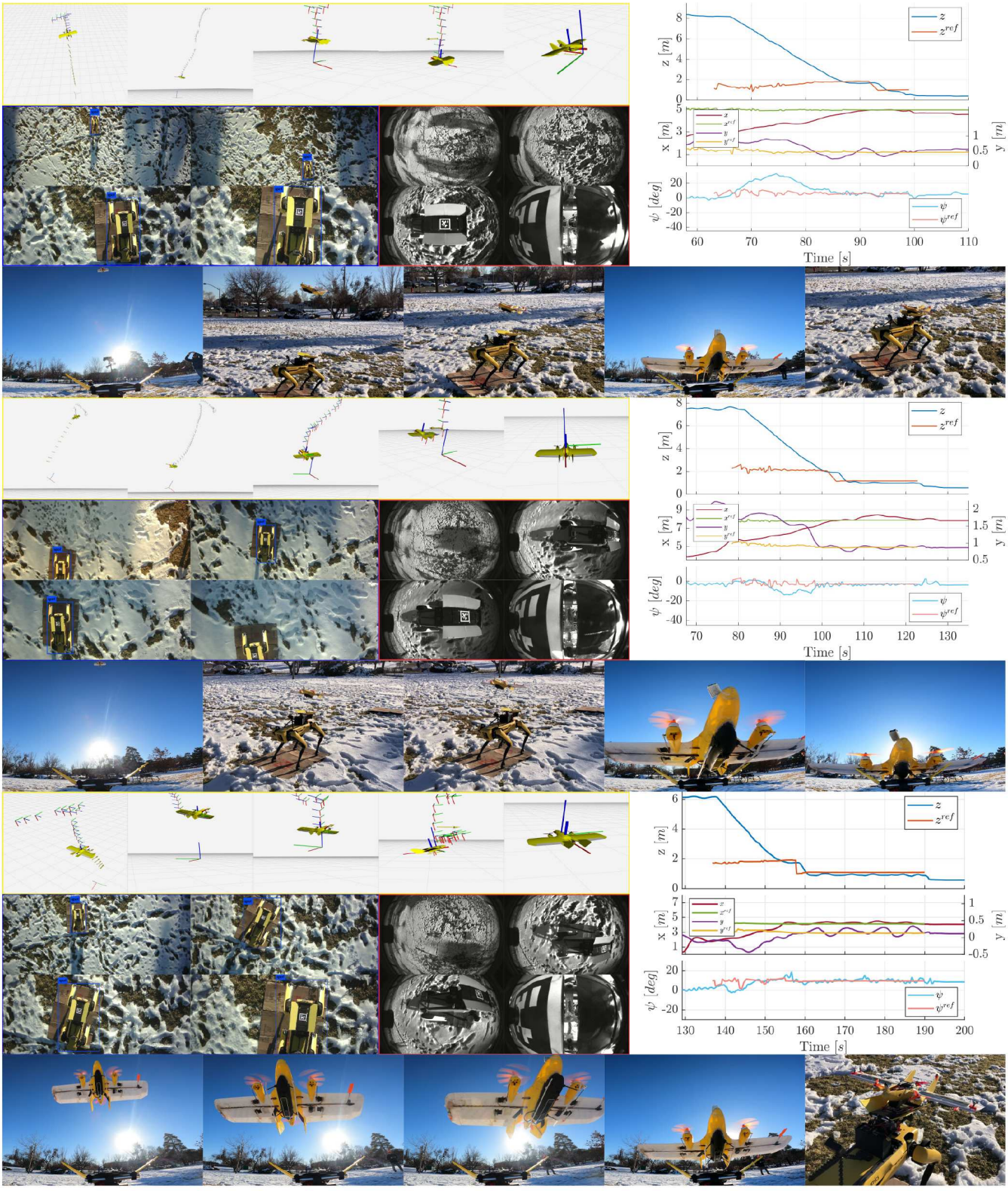

Fig. 5. Experimental Validation. Each set of rows comprises: *a) - Yellow box*: Estimated pose of Docking & Recharging Backpack (large transform axis, derived from Fiducial-Tag detection), Vehicle flight path (sequence of small transform axes), and Approach and landing waypoints sequence (yellow arrows). *b) - Blue Box*: Indicative views of the color-camera and Spot detection instances. *c) - Red Box*: Indicative views of the fisheye-camera from initial height to final touchdown. *d) - Plots*: Time evolution of the approach and touchdown sequence, x, y, z, ψ states and waypoint values $x^{ref}, y^{ref}, z^{ref}, \psi^{ref}$. *d) - Video Feed*: Characteristic instances captured in the field by a Spot-mounted camera, and a handheld one.
1st Set of Rows: A nominal reassembling sequence with favorable flight conditions. *2nd Set of Rows*: A reassembling sequence with a sudden wind-gust experienced during the final approach. *3rd Set of Rows*: A reassembling sequence with intermittent wind perturbations experienced before the final touchdown.

The following row shows indicative imagery captured by video footage in the field, using a camera mounted on top of the Legged robot and pointed at the Docking & Recharging Backpack, as well as a handheld camera. More specifically in this sequence, we demonstrate: i) the view from the ground robot while the aerial vehicle is approaching, ii) the hovering pose corresponding to the intermediate waypoint, at which point the fisheye-camera is close enough to establish detection of the Fiducial-Tag, iii) the hovering pose corresponding to the final waypoint, which is the last step before the final landing command is issued, iv) the moment right before touchdown into the pad, and v) the eventual reassembled system-of-systems.

B. Reassembling under Wind-Gust

The middle set of rows in Figure 5 illustrates a case where a sudden wind gust is experienced at the last phase of the air-to-ground reassembling operation.

More specifically, we point out the fourth subfigure in the *Yellow box* sequence that illustrates the instantaneous forward deviation w.r.t. the final waypoint (yellow arrow). A video instance of this deviation can also be observed in the lower row, where the second subfigure shows the vehicle approaching the final waypoint, but in the third subfigure it has been "blown away" from it. The plots also clearly show this gust's effect on the vehicle's x state around the 110 s mark, but also additionally illustrate the vehicle's general struggle against wind perturbations (also in y state). Overall the fourth subfigure of the video instances row shows how the final touchdown occurs with the vehicle off-center, but eventually ends up aligned and safely landed due to the DRB's passive docking design.

It is also noted in the *Blue box* how it is highly realistic to assume that Deep-Learned visual servoing is insufficient for the proposed operation, as the Legged robot constantly comes comes out of view during the approach. The Fiducial-Tag based method gives consistent guidance into the pad, due to its ability to provide a full relative-pose estimate. The *Red box* also shows the significant utility of the wide-FoV fisheye-camera in shorter distances, as the Fiducial-Tag can still be observed even during the deviations caused by wind gusts, ensuring relative reference consistency during the final position-hold operation.

C. Reassembling under Perturbation

The last set of rows in Figure 5 illustrates a case where wind ripples are experienced during the last phase.

Similarly to what was described before, the fourth subfigure in the *Yellow box* shows the vehicle after touchdown has been achieved, but it can clearly be seen that from its pose history (sequence of smaller transform axes) that it struggled with repeated perturbations "blowing" it around. Again, especially the y state plots indicate the time evolution of this phenomenon around 160-190 s, and second and third video instances show the magnitude of this side-to-side relative motion that inhibits the pipeline from commanding the final landing. The final touchdown moment in given in the subsequent image, as well as a closeup of the eventually reassembled marsupial system-of-systems.

Finally, indicative instances of airborne detection of the Legged robot even under relative rotation are given in the *Blue box*, and additional instances of the utility of the wide-FoV fisheye-camera during short-range relative motion (as expected due to wind perturbations experienced in real-world field missions) are given *Red box*, where in the Fiducial-Tag's visibility is maintained.

VI. CONCLUSIONS

In this work, we presented a pipeline for the air-to-ground reassembling of an Aerial & Legged marsupial System-of-Systems operating within the context of wide-field collaborative surveying missions, which does not depend on Vehicle-to-Vehicle communication. The proposed system was studied w.r.t. its operating performance in real-world field conditions against different altitude scales and operating environments, and its effectiveness was demonstrated in field experiments of the air-to-ground reassembling operation.

REFERENCES

[1] T. Tomic, K. Schmid, P. Lutz, A. Domel, M. Kassecker, E. Mair, I. L. Grixa, F. Ruess, M. Suppa, and D. Burschka, "Toward a fully autonomous uav: Research platform for indoor and outdoor urban search and rescue," *IEEE robotics & automation magazine*, vol. 19, no. 3, pp. 46–56, 2012.

[2] P. Arora and C. Papachristos, "Mobile manipulation–based deployment of micro aerial robot scouts through constricted aperture-like ingress points," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 6716–6723.

[3] A. Bircher, K. Alexis, M. Burri, P. Oettershagen, S. Omari, T. Mantel, and R. Siegwart, "Structural inspection path planning via iterative viewpoint resampling with application to aerial robotics," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 6423–6430.

[4] C. Papachristos, K. Alexis, L. R. G. Carrillo, and A. Tzes, "Distributed infrastructure inspection path planning for aerial robotics subject to time constraints," in *2016 international conference on unmanned aircraft systems (ICUAS)*. IEEE, 2016, pp. 406–412.

[5] F. Mascarich, T. Wilson, C. Papachristos, and K. Alexis, "Radiation source localization in gps-denied environments using aerial robots," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 6537–6544.

[6] C. Papachristos and K. Alexis, "Augmented reality-enhanced structural inspection using aerial robots," in *2016 IEEE international symposium on intelligent control (ISIC)*. IEEE, 2016, pp. 1–6.

[7] G. Paul, S. Webb, D. Liu, and G. Dissanayake, "Autonomous robot manipulator-based exploration and mapping system for bridge maintenance," *Robotics and Autonomous Systems*, vol. 59, no. 7-8, pp. 543–554, 2011.

[8] S. Khattak, C. Papachristos, and K. Alexis, "Keyframe-based direct thermal–inertial odometry," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 3563–3569.

[9] S. Khattak, F. Mascarich, T. Dang, C. Papachristos, and K. Alexis, "Robust thermal-inertial localization for aerial robots: A case for direct methods," in *2019 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2019, pp. 1061–1068.

[10] C. Papachristos, S. Khattak, and K. Alexis, "Uncertainty-aware receding horizon exploration and mapping using aerial robots," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 4568–4575.

[11] C. Papachristos, M. Kamel, M. Popović, S. Khattak, A. Bircher, H. Oleynikova, T. Dang, F. Mascarich, K. Alexis, and R. Siegwart, "Autonomous exploration and inspection path planning for aerial robots using the robot operating system," in *Robot Operating System (ROS)*. Springer, Cham, 2019, pp. 67–111.

[12] C. Papachristos, F. Mascarich, S. Khattak, T. Dang, and K. Alexis, "Localization uncertainty-aware autonomous exploration and mapping with aerial robots using receding horizon path-planning," *Autonomous Robots*, vol. 43, no. 8, pp. 2131–2161, 2019.

[13] C. Papachristos, S. Khattak, F. Mascarich, T. Dang, and K. Alexis, "Autonomous aerial robotic exploration of subterranean environments relying on morphology–aware path planning," in *2019 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2019, pp. 299–305.

[14] W. Johnson, S. Withrow-Maser, L. Young, C. Malpica, W. Koning, K. WJF, M. Fehler, A. Tuano, A. Chan, A. Datta *et al.*, *Mars Science Helicopter Conceptual Design*. National Aeronautics and Space Administration, Ames Research Center, 2020.

[15] R. D. Lorenz, E. P. Turtle, J. W. Barnes, M. G. Trainer, D. S. Adams, K. E. Hibbard, C. Z. Sheldon, K. Zacny, P. N. Peplowski, D. J. Lawrence *et al.*, "Dragonfly: A rotorcraft lander concept for scientific exploration at titan," *Johns Hopkins APL Technical Digest*, vol. 34, no. 3, p. 14, 2018.

[16] M. Kulkarni, M. Dharmadhikari, M. Tranzatto, S. Zimmermann, V. Reijgwart, P. De Petris, H. Nguyen, N. Khedekar, C. Papachristos, L. Ott *et al.*, "Autonomous teamed exploration of subterranean environments using legged and aerial robots," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 3306–3313.

[17] N. Michael, S. Shen, K. Mohta, V. Kumar, K. Nagatani, Y. Okada, S. Kiribayashi, K. Otake, K. Yoshida, K. Ohno *et al.*, "Collaborative mapping of an earthquake damaged building via ground and aerial robots," in *Field and service robotics*. Springer, 2014, pp. 33–47.

[18] P. Arora and C. Papachristos, "Mobile manipulation–based deployment of micro aerial robot scouts through constricted aperture-like ingress points," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 6716–6723.

[19] P. De Petris, S. Khattak, M. Dharmadhikari, G. Waibel, H. Nguyen, M. Montenegro, N. Khedekar, K. Alexis, and M. Hutter, "Marsupial walking-and-flying robotic deployment for collaborative exploration of unknown environments," *arXiv preprint arXiv:2205.05477*, 2022.

[20] S. J. Carlson, T. Karakurt, P. Arora, and C. Papachristos, "Integrated solar power harvesting and hibernation for a recurrent-mission vtol micro aerial vehicle," in *2022 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2022, pp. 237–244.

[21] P. Arora, S. J. Carlson, T. Karakurt, and C. Papachristos, "Deep-learned autonomous landing site discovery for a tiltrotor micro aerial vehicle," in *2022 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2022, pp. 255–262.

[22] S. J. Carlson and C. Papachristos, "Solar energy harvesting for a land-to-recharge tiltrotor micro aerial vehicle," in *2022 IEEE Aerospace Conference (AERO)*. IEEE, 2022, pp. 1–8.

[23] S. J. Carlson, P. Arora, and C. Papachristos, "A multi-vtol modular aspect ratio reconfigurable aerial robot," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 8–15.

[24] S. J. Carlson, P. Arora, T. Karakurt, B. Moore, and C. Papachristos, "Towards multi-day field deployment autonomy: A long-term self-sustainable micro aerial vehicle robot," in *2023 International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, p. to appear.

[25] Y. Jang, C. Oh, Y. Lee, and H. J. Kim, "Multirobot collaborative monocular slam utilizing rendezvous," *IEEE Transactions on Robotics*, vol. 37, no. 5, pp. 1469–1486, 2021.

[26] G. B. Haberfeld, A. Gahlawat, and N. Hovakimyan, "Risk-sensitive rendezvous algorithm for heterogeneous agents in urban environments," in *2021 American Control Conference (ACC)*, 2021, pp. 3455–3460.

[27] S. G. Manyam, D. W. Casbeer, I. E. Weintraub, and C. Taylor, "Trajectory optimization for rendezvous planning using quadratic bézier curves," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 1405–1412.

[28] J. Balaram, M. Aung, and M. P. Golombek, "The ingenuity helicopter on the perseverance rover," *Space Science Reviews*, vol. 217, no. 4, pp. 1–11, 2021.

[29] P. Arora and C. Papachristos, "Launching a micro–scout uav from a mobile robotic manipulator arm," in *2021 IEEE Aerospace Conference (50100)*. IEEE, 2021, pp. 1–8.

[30] F. M. Zegers, P. Deptula, H.-Y. Chen, A. Isaly, and W. E. Dixon, "A switched systems approach to multiagent system consensus: A relay–explorer perspective," *IEEE Transactions on Robotics*, pp. 1–20, 2023.

[31] P. Karmokar, K. Dhal, W. J. Beksi, and A. Chakravarthy, "Vision-based guidance for tracking dynamic objects," in *2021 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2021, pp. 1106–1115.

[32] T. Muskardin, G. Balmer, L. Persson, S. Wlach, M. Laiacker, A. Ollero, and K. Kondak, "A novel landing system to increase payload capacity and operational availability of high altitude long endurance uav," in *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2016, pp. 495–504.

[33] S. Ramasamy, J.-P. F. Reddinger, J. M. Dotterweich, M. A. Childers, and P. A. Bhounsule, "Coordinated route planning of multiple fuel-constrained unmanned aerial systems with recharging on an unmanned ground vehicle for mission coverage," *Journal of Intelligent & Robotic Systems*, vol. 106, p. 81, 2022.

[34] P. Maini and P. B. Sujit, "On cooperation between a fuel constrained uav and a refueling ugv for large scale mapping applications," in *2015 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2015, pp. 1370–1377.

[35] A. B. Asghar, G. Shi, N. Karapetyan, J. Humann, J.-P. Reddinger, J. Dotterweich, and P. Tokekar, "Risk-aware resource allocation for multiple uavs-ugvs recharging rendezvous," 2022. [Online]. Available: https://arxiv.org/abs/2209.06308

[36] S. Martínez-Rozas, D. Alejo, F. Caballero, and L. Merino, "Path and trajectory planning of a tethered uav-ugv marsupial robotics system," 2022. [Online]. Available: https://arxiv.org/abs/2204.01828

[37] A. Rodriguez-Ramos, C. Sampedro, H. Bavle, I. G. Moreno, and P. Campoy, "A deep reinforcement learning technique for vision-based autonomous multirotor landing on a moving platform," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1010–1017.

[38] D. Lee, T. Ryan, and H. J. Kim, "Autonomous landing of a vtol uav on a moving platform using image-based visual servoing," in *2012 IEEE International Conference on Robotics and Automation*, 2012, pp. 971–976.

[39] A. Bouman, M. F. Ginting, N. Alatur, M. Palieri, D. D. Fan, T. Touma, T. Pailevanian, S.-K. Kim, K. Otsu, J. Burdick, and A.-a. Agha-Mohammadi, "Autonomous spot: Long-range autonomous exploration of extreme environments with legged locomotion," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 2518–2525.

[40] P. Fankhauser, M. Bloesch, P. Krüsi, R. Diethelm, M. Wermelinger, T. Schneider, M. Dymczyk, M. Hutter, and R. Siegwart, "Collaborative navigation for flying and walking robots," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 2859–2866.

[41] E. Vattapparamban, I. Guvenc, A. I. Yurekli, K. Akkaya, and S. Uluagac, "Drones for smart cities: Issues in cybersecurity, privacy, and public safety," in *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2016, pp. 216–221.

[42] J.-C. Lee, C.-C. Chen, C.-T. Shen, and Y.-C. Lai, "Landmark-based scale estimation and correction of visual inertial odometry for vtol uavs in a gps-denied environment," *Sensors*, vol. 22, no. 24, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/24/9654

[43] Z. Ding, T. Yang, K. Zhang, C. Xu, and F. Gao, "Vid-fusion: Robust visual-inertial-dynamics odometry for accurate external force estimation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 14 469–14 475.

[44] P. De Petris, S. Khattak, M. Dharmadhikari, G. Waibel, H. Nguyen, M. Montenegro, N. Khedekar, K. Alexis, and M. Hutter, "Marsupial walking-and-flying robotic deployment for collaborative exploration of unknown environments," 2022. [Online]. Available: https://arxiv.org/abs/2205.05477

[45] S. J. Carlson and C. Papachristos, "The MiniHawk-VTOL: Design, modeling, and experiments of a rapidly-prototyped tiltrotor uav," in *2021 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2021, pp. 777–786.

[46] S. J. Carlson and C. Papachristos, "Migratory behaviors, design principles, and experiments of a vtol uav for long-term autonomy," *ICRA 2021 Aerial Robotics Workshop on "Resilient and Long-Term Autonomy for Aerial Robotic Systems"*, 2021. [Online]. Available: https://www.aerial-robotics-workshop.com/uploads/5/8/4/4/58449511/icra2021-aerial_paper_8.pdf

[47] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018. [Online]. Available: https://arxiv.org/abs/1804.02767