2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

WAVN: Wide Area Visual Navigation for Large-scale, GPS-denied Environments

Damian M. Lyons, Mohamed Rahouti
Computer & Information Science
Fordham University
New York, USA
{dlyons,mrahouti}@fordham.edu

***Abstract*—This paper introduces a novel approach to GPS-denied visual navigation of a robot team over a wide (i.e., out of line of sight) area which we call WAVN (Wide Area Visual Navigation). Application domains include small-scale precision agriculture as well as exploration and surveillance. The proposed approach requires no exploration or map generation, merging, and updating, some of the most computationally intensive aspects of multi-robot navigation, especially in dynamic environments and for long-term deployments. In contrast, we extend the visual homing paradigm to leverage visual information from the entire team to allow a robot to home to a distant location. Since it only employs the latest imagery, the approach can be resilient to the current state of the environment. WAVN requires three components: identification of common landmarks between robots, a communication infrastructure, and an algorithm to find a sequence of common landmarks to navigate to a goal. The principal contribution of this paper is the navigation algorithm in addition to simulation and physical robot results characterizing performance. The approach is also compared to more traditional map-based approaches.**

I. Introduction

Successful and efficient navigation of a team of robots operating in a large-scale, long time-frame environment poses many challenges [11] including collaborative, long-term SLAM [3] and dynamic route planning [29]. The situation becomes even more complicated if GPS cannot be assumed [24]. We address the problem of the wide-area place-to-place navigation for robots that are part of a team deployed in a large outdoor space over a time scale in which there may be significant visual and structural changes. One example is a team of robots deployed for low-cost precision agriculture [5]; High precision, high-reliability GPS can be an entrance barrier for small family farms. A successful solution to this challenge would open the way to revolutionizing small farming to compete with big agribusiness [10]. A second example is a team of military robots deployed in a GPS-denied environment. McManus [20] underlines the challenge of military operation in areas where GPS cannot be guaranteed due to occlusion, indoor situations, GPS jamming, or is otherwise insufficient for navigation tasks.

The principal contributions of this paper are:

- An approach to wide-area, visual navigation (WAVN) for moving a robot, or group of robots, from one place to another distant location whose current appearance is known.
- The approach leverages instantaneous information from the robot team rather than a map representation developed by collaborative exploration and mapping and operates without GPS.
- Simulation and physical robot experimental results to quantify WAVN performance and help understand how it compares to other approaches.

Section II discusses the state of the art. Section III introduces our approach: Wide Area Visual Homing (WAVN) and its components. An experiment to compare WAVN performance with more traditional map-building approaches is presented in Section IV. Simulation and Physical robot results are presented in Sections V and VI. Lastly, Section VII discusses these results and areas of future work.

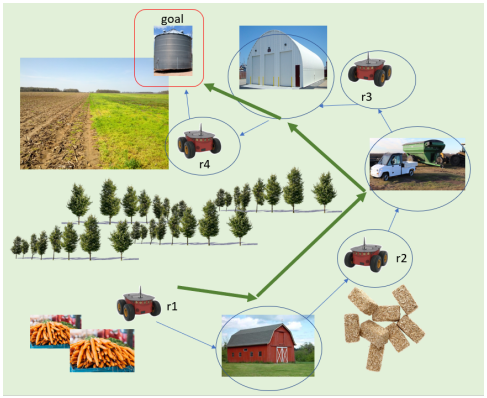

Fig. 1. Agricultural example of Wide Area Visual Navigation for robot r1 to deliver produce to distant goal. The chain of robots with common visual landmarks leading to the goal is shown as blue circles/ellipses. The final navigation path is shown as thicker green line.

II. Prior Work

Prior approaches to GPS-denied navigation typically leverage a robot's Inertial Motion Unit (IMU) and often other sensors such as laser ranging or visual imagery [2], [7], [24].

Scaramuzza's [24] SFLY project is an impressive example of this, addressing mapping, navigation, and coverage for a team of UAVs. Onboard camera information is combined with IMU data to stabilize each UAV and to navigate locally. Each UAV constructs a 3D map, and these are integrated into a single global map in an off-board, centralized global mapping and navigation ground station. Each UAV streams information from three cameras to the ground station, which matches the imagery to a map using an image vocabulary search along with a geometric verification. Stereo processing is then used to construct a 3D dense, merged map. Multi-robot mapping can use a centralized ground station like this or a decentralized design, where each robot merges map information [23]; in either case, map merging can impose a high computational load.

To build the map, some UAVs must be sent on exploratory mapping missions composed of multiple loops around the working environment. The advantage of this approach is that it then enables future map-based navigation. Of course, the map must also be kept up to date, an additional computational demand [3] if the environment is dynamic and especially an issue for long-term deployments, as we might expect for an agricultural application.

Visual Homing (VH) [15], [21], [25], [26] is a popular, map/GPS-independent approach to navigation. Inspired by insect navigation, VH compares a goal image with the current image and determines a motion vector to bring the current image closer to the goal image. Approaches include holistic methods such as warping [8], or feature-based methods such as Average Landmark Vector [15], and feature correspondence methods [26]. VH performance can be improved further by adding depth information [21] and integrating with local obstacle avoidance [9]. However, VH requires that the goal image is in the view from the start. Stelzer [25] uses a topological map data structure (LT-Map) to support VH to a distant location as a sequence of VH steps, each starting with the goal in view. This still requires a separate set of explorations to generate the map, though the computational load is lighter than for 3D dense mapping.

Mao's [19] UAV topological map is based on aerial landmarks selected for reliable landmark recognition but is still constructed a-priori and would need to be kept updated for long-term deployment. Wu [28] employs a deep learning approach to transfer image goals to point navigation on a map in a cluttered space. However, they use a conventional mapping module to build the map. Yang [29] addresses constructing a polygonal map as a computationally efficient, visibility-graph-based navigation proceeds, guaranteeing it always uses the latest information. However, the end-point of navigation is still known, so even for unknown terrain, planning knows where to plan (though not what is there).

Our work differs from this prior work in several key regards. We have a team of robots, each carrying out their own activities distributed over an outdoor space. Our goal is to keep the cost and computational footprint low, and for this reason: we adopt VH as our primary navigation approach; we will not conduct global mapping or map-merging; and, we will not rely on GPS. Because we do not assume a separate map-construction, exploration, or map-maintenance activity, a robot needing to travel to a distant, image-goal location has no local information even to begin planning.

III. Wide Area Visual Homing

We assume a (potentially heterogeneous) team of n robots $r_1, \ldots, r_n$ distributed over an area with no constraint that any robot sees any other robot but that some robots have a partially overlapping field of view. We assume that individual robots or task teams of robots may be engaged in another primary task distinct from WAVN, such as weeding, collecting produce, inspecting crops, and so forth. WAVN is invoked when one such robot or task team needs to navigate to a distant, out of view location, for example, to deliver collected produce to a remote storage location or transport vehicle.

We make no assumption of GPS or previously constructed and shared map information to guide the robot to the remote location. However, we assume that team members can communicate with each other, at least intermittently, and use a communication infrastructure consisting of a distributed blockchain algorithm [4] with each robot having a copy of the blockchain ledger. Robots update the ledger with a copy of their panoramic field of view on a regular basis as they work. Although other communication frameworks could be adopted here [13], blockchain provides a convenient, lightweight, secure, and distributed communication channel upon which we implement the primary contribution of this paper, the WAVN algorithm.

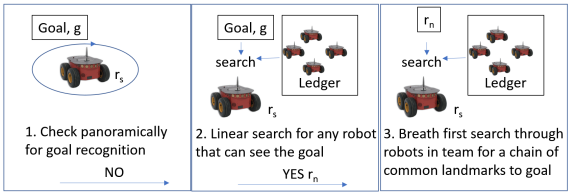

Fig. 2. Steps for path selection in the WAVN algorithm.

The remote location is communicated to the robot r_s as an image goal g. Fig. 2 illustrates the steps in path selection. If r_s can see g in its current field of view $visible(r_s, g)$, using, e.g., panoramic camera resources [12] or by simply rotating a single camera panoramically and applying a target recognition strategy [6], [17], then our previously developed correspondence-based visual homing algorithm is used to navigate to the target [9], [21]. If $visible(r_s, g)$ fails, then for each robot r_i with a field of view $fov(r_i)$ detailed in the ledger, $visible(r_i, g)$ is checked.

In the case that $visible(r_n, g)$ is found, the next step is *WAVN search*. Fig. 1 shows an example navigation problem. Edges indicate common visibility between robots from the robot to an ellipse-enclosed landmark. The goal for r_1 is a landmark on the upper left (in red). Once $visible(r_n, g)$ is established, not one, but rather a chain of common landmarks, needs to be constructed, as shown in the figure. r_1 should visually home to each of the common landmarks in turn, and

after the last one, home to the final goal, as shown by the heavier, green arrows on the figure.

We identify three crucial components of this approach: landmark modeling, WAVN search, and communication architecture, addressed in subsections III-A, III-B, III-C respectively. The contribution of this paper is the WAVN search, and we will leverage prior work for the other two components.

A. Common Landmark Detection

Place recognition [16] is often used in topological mapping approaches as well as in SLAM loop closure. Our requirements differ from place recognition in that we expect two robots not to see the same place but to see some scene components in common. The objective is to use one such commonly seen scene component as an intermediate image goal for VH. This also differs from regular image recognition in that the objective is not to identify what is seen, only to confirm that both robots can see it.

In prior work, we [18] evaluated three approaches to visual landmark models for what we called the 'Do you see what I see' (DYSWIS) problem: two robots, r_k and r_j, inspecting all their surrounding scenery and determining if they see a common visual landmark. A feature-based approach to matching images (using Scale Invariant Feature Transform - SIFT - features) from both robots is compared to an approach using a Convolutional Neural Network (CNN) classification of objects followed by feature matching of any regions classified as the same object model. Input panoramic imagery is six $\frac{\pi}{3}$-FOV component images collected by rotating each robot $r_{i\in\{k,j\}}$ in six increments of $\frac{\pi}{3}$, generating $I_{i,a}$ for $a \in \Theta$, and $\Theta = \{0, \ldots, \frac{5\pi}{3}\}$.

For feature-matching alone, each $I_{k,a}$ is matched with each $I_{j,b}$ for $a, b \in \Theta$. For the CNN approach, Yolov3 *pre-trained on the OpenImages dataset*, which has 600 commonly seen object classes, is applied to each $I_{i,a}, i \in \{k, j\}, a \in \Theta$. For each $I_{i,a}$, the highest confidence object bounding box is identified $BB_{i,a}$. Finally, each $BB_{k,a}$ is feature-matched with each $BB_{j,b}$ of the same class. Using a ROS/Gazebo implementation and by relating identified landmarks to the existence (or not) of Gazebo models at those locations, we showed that the Yolo/SIFT approach is dramatically better at identifying actual common landmarks. This landmark model approach has the additional advantage for WAVN that any commonly matched $BB_{k,a}$ can now *also* be used as a navigation image goal.

B. WAVN Search Algorithm

WAVN search is a breath-first (BF) search through the space of landmarks seen in common between pairs of robots to identify a sequence of landmarks as shown in Fig. 1. Robot r_s is tasked to reach g, an out of sight destination. We define a WAVN search node as:

$$(s, p), s \in S, p \in L^*.$$

where $S = R \times R \times L$ for the set of robots R and landmarks L, where $(r_1, r_2, cm) \in S$ indicates that cm has been identified as a common landmark between r_1 and r_2, and where $p \in L^*$ is the sequence of landmarks identified so far in the search. Algorithm 1 implements the BF search on S. By its BF properties, this algorithm is guaranteed to find a path if one exists. That path will have the shortest number of intermediate landmarks supported by current visibility in the environment.

Algorithm 1 WAVN Algorithm

```
1: procedure WAVN(r_s, g)
2:     fringe = [(0, r_s, 0), ()]
3:     while len(fringe) > 0 do
4:         (r_1, r_2, cm), p = pop(fringe)
5:         if visible(r_2, g) then
6:             p' = add final homing step to p
7:             return p'
8:         end if
9:         lmks = ⋃_{r_j ∈ T − r_2} clm(r_2, r_j) × {r_j}
10:        for each (lm, r) ∈ lmks do
11:            s' = (r_2, r, lm)
12:            p' = add lm homing step to p
13:            add s', p' to fringe
14:        end for
15:    end while
16: end procedure
```

Analysis. If at most b robots see any common landmark, the worst-case search complexity for common landmarks is $O(b^d)$ for goal path (landmark sequence) length d. While greater b means that alternate paths to the goal can be explored and could be used opportunistically, it also increases the computational complexity. If there are n robots evenly distributed in the workspace, then the maximum path length is $\ell \approx log_b(n+1)$, logarithmic in the size of the robot team. If we assume that each robot sees landmarks just in its vicinity, then a worst-case path length would be $O(n)$.

Search Policies. In the case where depth information is available (e.g., in [21] we include stereo depth in homing), then the *closest common landmark* can be chosen from $clm(r_1, r_2)$:

$$closestR(r_1, r_2) = \underset{c \in clm(r_1, r_2)}{\mathbf{argmin}}\ d(r_1, c) + d(r_2, c)$$

Line 11 of Algorithm 1 is modified to append $d(r, lm) + d(r_2, lm)$ to s', and line 4 is modified to $pop(sorted(fringe))$, yielding a uniform cost search. The resultant path has some advantages: it will traverse most closely to all the robots whose common landmarks are in the path and, in this sense, is the most resilient available path. It remains in territory easily viewed by other robots and is potentially responsive to path blockages.

A second definition of closest is to select c closest to the previous landmark in the sequence, so

$$closestL(s, p) = \underset{c \in clm(r_1, r_2)}{\mathbf{argmin}}\ \triangle(d(r_1, cp), d(r_1, c))$$

for $cp = last(p)$ and $\triangle$ the length of the third side of the triangle r_1, cp, cp. By minimizing the length of each section of the path, this has the advantage of being the shortest of the available paths.

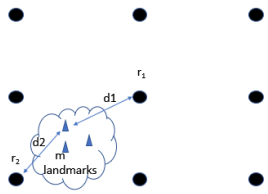

Fig. 3. Section of the grid of $n \times n$ robots with at most m landmarks, each located a distance $d1, d2$ between each pair r_1, r_2.

C. Communication Infrastructure

Blockchain technology [1] has been integrated into a broad range of modern applications, including, but not limited to, connected and autonomous vehicles (CAVs) and the Internet of Things (IoT). It is used here to provide modular decentralized communication among visual homing robots. This approach makes available to each robot the imagery information necessary to implement the WAVN search algorithm locally.

Beyond the communication needs of the WAVN algorithm, blockchain is consensus-based and trustworthy in terms of data integrity (immutability), auditability, and transparency [1], [22]. It potentially provides a tamper-proof record of robot transactions (i.e., individual robot's location record) and the rapid retrieval of imagery [27]. Arguably, this would be more crucial for a military application of WAVN than for the agricultural application that has been the primary focus here, and we postpone any investigation of this to future work.

IV. The Crooked Way

One way in which we can compare WAVN with more traditional multi-robot GPS-denied SLAM is via the paths generated. A key advantage of generating a map from an a-prior exploration phase, either 3D such as [23], [24] or topological as in [19], is that as long as the endpoint is physically reachable on the map, the shortest path can be generated. WAVN does not require computation or exploration to generate or merge maps; however, its ability to find a path is limited by how the team of robots is physically distributed, and the path itself could be quite 'crooked.' To quantify this comparison, we conduct a simulation experiment.

The locations of a team of n^2 robots were simulated as a square grid of $n \times n$ points, Fig. 3. Each robot shared m landmarks with each adjacent robot, and each landmark was assigned a random distance from that robot. Selecting a random robot and landmark, the WAVN algorithm can be used to find a sequence of common landmarks terminating in the goal landmark. Some percentage e of the common landmarks can be randomly removed to model different distributions of the team of robots over the environment - the factor that influences WAVN path availability and quality in comparison with map-based approaches.

Two series of 500 random searches were simulated for each value of $n \in \{2, \ldots, 10\}$ and $e \in \{0, 5, 10, \ldots, 95\}$. In the first series, an arbitrary choice is made whenever there is more than one common landmark. In the second case, the landmark with the smallest sum of distances to each robot

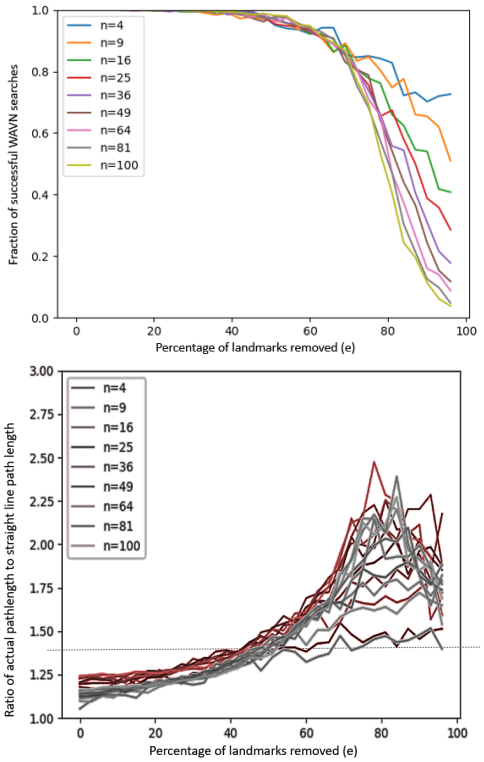

Fig. 4. (a) Successful fraction of WAVN searches, and (b) Path length ratio graphs for both series overlayed: closest landmark metric (black) versus arbitrary (red), dashed line shows $\sqrt{2}$

is selected (the resilient path policy). The path length of each path is accumulated into two averages for each value of n, e: an average path length and an average ratio of the path length to the straight line distance from the initial robot to the final landmark. This second measure yields insight into how 'crooked' the common landmark path is as opposed to a straight line. We use the Manhattan distance ratio $\sqrt{2}$ as a threshold: below this, we would consider that the WAVN path is not especially 'crooked.'

Fig. 4(a) shows a graph of the number of successful WAVN searches for each value of n and e. Most searches are successful until e exceeds 80%; at that point, the team has become disconnected, and an increasing number of targets are no longer accessible.

Fig. 4(b) shows the graph of the average path ratio for all n and e using closest landmarks ($closestL$)in black and using arbitrary landmarks in red. For all n and for values of e below approx. 50%, the path length ratios are less than the Manhattan ratio. As landmark erosion reaches 80%, the ratio increases, closing in on 2.5 times as long as the straight line path. The fall-off after that is likely due to the partitioning of the team and resulting in smaller paths.

Almost uniformly, the resilient path policy (black) yields smaller ratios. For small values of e, the path can be as much as 60% shorter, with this benefit reducing strongly as e increases.

V. Simulation Experiments

The WAVN algorithm was evaluated in two experiments, one in simulation and one using physical robots in the lab.

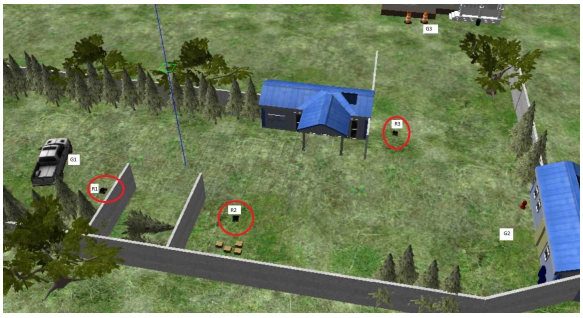

Fig. 5. Three robot WAVN ROS/Gazebo simulation (initial robot locations outlined in red, (R_1-R_3) are robots and (G_1-G_3) goal locations).

A. WAVN Search in ROS/Gazebo

The WAVN approach was evaluated using a ROS/Gazebo simulation with a team of three Pioneer 3-AT robots in an environment where the three robots had mutually shared and mutually occluded portions of their field of view, Fig. 5. The environment includes terrain features such as walls, trees, and houses that can function as common landmarks. Each robot is equipped with a Hokuyo laser ranging for short-range obstacle avoidance and PT-mounted cameras for landmark detection. The single YOLO/SIFT method of [18] was used for landmark detection and recognition. However, ORB was substituted for SIFT as its performance on graphical imagery was slightly better. CUDA-enabled YOLOv3[1] was used, running at 30 fps on an NVIDIA GeForce RTX 3080, pre-trained on the openImages dataset [14]. The homing methods of [9], [21] were implemented, but without stereo depth. Once homing concluded, the robots then carried out a guarded move to within $3m$ of the landmark. This strategy was implemented because getting close to the landmark means most of it will no longer be in view.

A number of trials ($n = 36$) were generated for the three robots that required one to navigate to a goal landmark initially out of its view but in view of one of the other robots and involving a single common landmark. This restricted paths to R_1 to G_3, R_2 to G_1 and G_3.

A trial is successful if a robot identifies a common landmark with another robot that can see the final homing goal image and successfully navigates to the common landmark and thence to the goal. Failure can occur because a common landmark is not correctly identified or because the final goal cannot be recognized from the common landmark.

B. Simulation Results

The mission success results are summarized in Table V-B. Wide-area navigation succeeded in 32 of the 36 missions (88.8%). All trial paths are shown in Fig. 6, color-coded based on the three missions (failures are shown in black). All paths show a 'crooked' shape as the robot travels to the common landmark and then to the destination. This is quite severe in R_2 to G_1 and unavoidable since there is a wall in the way. The path length ratio (actual length to straight line distance) is

[1]https://pjreddie.com/darknet/yolo/

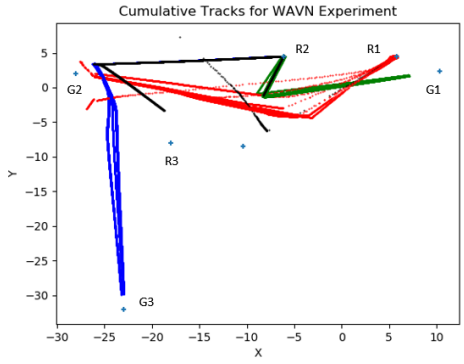

Fig. 6. Paths for all successful WAVN trials overlayed with robot locations (R_1-R_3) and goal locations (G_1-G_3).

TABLE I
SUCCESS/FAILURE RESULTS OF THREE-ROBOT ROS/GAZEBO WAVN EVALUATION EXPERIMENT.

Metric	Mission			Average
	R_1/G_2	R_2/G_1	R_2/G_3	
Nbr. of Missions	10	13	13	
Average lengths	34.52	21.84	46.94	34.44
Straight line ratios	1.14	2.05	1.56	1.62
Nbr. of common landmarks seen	2.8	2.8	2.75	2.78
Nbr. of successes	8	13	11	0.88
Nbr. of failures to find common landmark	1	0	1	0.05
Nbr. of failures to complete final homing	1	0	1	0.05

best for G_2 (below $\sqrt{2}$) and worst for G_1, again, because of the wall. There were, on average, between 2 and 3 common landmarks for each trial. The four mission failures, 2 on G_2 and 2 on G_3, were equally divided between failing to choose a common landmark correctly and failing to reacquire the homing target once at the common landmark.

Table II shows the final homing accuracy for the experiments. Homing performance was poor on some $G3$ missions, compromised by the final guarded move stopping at an object other than the landmark. Omitting these three trials would result in an average accuracy of $2.71m$ (the goal is to move within $3m$ of the landmark) with a standard deviation of $0.33m$. The selection of a terminal guarded move strategy proved a flawed strategy.

VI. ROBOT EXPERIMENTS

Two Turtlebot3 (Burger) mobile robots equipped with LDS-01 laser ranger and Raspberry Pi camera module V2.1 were

TABLE II
NAVIGATION ACCURACY RESULTS. Final average ($\overline{acc}$) and standard deviation (SD) of accuracy for a subset of the missions.

Mission	Nbr.	Avg. length	Straight line ratio	$\overline{acc}$	SD
R_1/G_2	8	34.52	1.14	2.36	1.15
R_2/G_1	13	21.84	2.05	3.26	0.02
R_2/G_3	8	53.36	1.39	2.15	0.03
Average		34.03	1.62	2.71	0.33

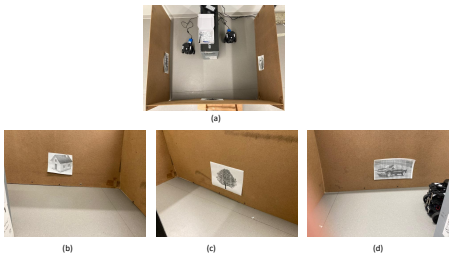

Fig. 7. Environment for physical robot experiments. (a) top view showing robots at example locations; (b)-(d) side views showing landmarks.

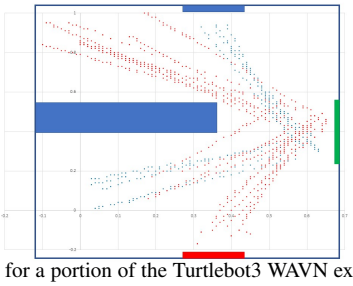

Fig. 8. Tracks for a portion of the Turtlebot3 WAVN experiments. Blue tracks start on the bottom and red tracks on the top.

used for the physical robot experiments. Each robot was running ROS Kinetic under Raspian on a Raspberry Pi 3B+. All WAVN calculations were carried out on a central server.

A. Robot Experimental Setup

The robots were placed in a square enclosure separated by a dividing wall, Figure 7. Each robot was thus prevented from seeing one side wall of the enclosure, but both could see the front wall. Natural landmark imagery of houses, trees, and vehicles was attached to the hidden and non-hidden walls. Figure 7 shows the overhead (a) and side views (b)-(d), including nominal initial positions for the two robots.

For $n = 30$ trials, each robot was placed in a location within a radius of $6cm$ of the position shown with offsets selected randomly. The same set of three landmarks was used for all trials. Each robot was directed in turn for half the trials to conduct wide-area visual homing to the hidden landmark on the opposite side of the enclosure. Data was collected on all landmark analyses and WAVN searches, as well as on the trajectories of the robots.

B. Robot Results

Fig. 8 is a scatter-plot of 15 of the 30 tracks with superimposed experimental enclosure and landmark positions. (Including all 30 tracks presents too cluttered of a graph to see individual motions.) Blue tracks start on the right of the barrier (bottom of the figure) and navigate to the hidden landmark (a house) on the left via the common landmark shown in green (a tree). The red point navigates to the hidden landmark on the left (a truck) via the same common landmark.

Of the $n = 30$ trials, 29 or 97% successfully completed the navigation. However, 4 of the 30 trials experienced some landmark identification failure during the navigation:

- 1 (3%) failed to find a common landmark on the first attempt but succeeded on the second;
- 2 (6%) failed on the first attempt to identify the final home location after homing to the intermediate landmark, but succeeded on the second attempt; and,
- 1 failed (3%) to find the final location in two attempts after homing to the common landmark, at which point navigation failed.

The terminal guarded move strategy introduced no addition issues for these trials.

VII. Conclusions

We have introduced a novel approach to GPS-denied Wide Area Visual Navigation (WAVN) for a team of robots. A lightweight visual homing approach is used to navigate to image goals. Distant, out of view, goals are handled by leveraging the instantaneous panoramic imagery from the entire network to find a sequence of intermediate landmarks, image goals, leading to the final homing step. This paper has focused on just the search algorithm to generate the sequence of intermediate landmarks, and its contributions include the algorithm, complexity analysis, and a discussion of two key search policy variants to select resilient paths versus shortest paths. Both ROS/Gazebo and physical robot results are presented, showing the algorithm's performance.

Comparison with more traditional map-based approaches to this problem centers on what WAVN incurs for not having computationally intensive map-building, merging and updating activities. Simulation studies show that despite no map, WAVN searches can be successful over a wide range of team distributions in the environment, and the paths generated only become especially crooked once the team has few landmarks in common.

The WAVN failures in both simulation and physical experiments could be improved with an improved landmark model. [18] went on to evaluate clusters of two and three CNN-classified objects extracted from each robot image and matched using Kuhn-Munkres and found that clusters of three objects perform significantly better. They theorize that clusters reduce aliasing issues (one tree is very similar to any other tree) and propose this could be improved by looking for object diversity and additional semantic content in clusters. Future work will focus on evaluating this and conducting outdoor experiments.

References

[1] Anwaar Ali, Mohamed Rahouti, Siddique Latif, Salil Kanhere, Jatinder Singh, Umar Janjua, Adnan Noor Mian, Junaid Qadir, Jon Crowcroft, et al. Blockchain and the future of the internet: A comprehensive review. *arXiv preprint arXiv:1904.00733*, 2019.

[2] Abraham Bachrach et al. Range - robust autonomous navigation in gps-denied environments. *IEEE/RSJ Int. Conf. on Int. Rob. & Sys. (IROS)*, 2010.

[3] Nicholas Carlevaris-Bianco and Ryan M. Eustice. Long-term simultaneous localization and mapping with generic linear constraint node removal. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013.

[4] Eduardo Castelló Ferrer. The blockchain: a new framework for robotic swarm systems. In *Proceedings of the future technologies conference*, pages 1037–1058. Springer, 2018.

[5] Josse DeBaerdemaeker. Precision agriculture technology and robotics for good agricultural practices. *IFAC Proceedings*, 46(4), 2013.

[6] Alberto Dionigi, Alessandro Devo, Leonardo Guiducci, and Gabriele Costante. E-vat: An asymmetric end-to-end approach to visual active exploration and tracking. *IEEE International Conference on Robotics and Automation*, 2022.

[7] S. Druen. Robotic navigation in gps-denied environments using the strapdown navigation algorithm with zero-velocity updates. *M.S. Thesis, Naval Postgraduate School*, 2020.

[8] M. Franz, B. Scholkopf, M. Mallot, and H. Bulthoff. Where did i take that snapshot? scene-based homing by image matching. *Biological Cybernetics(79)*, pages 191–202, 1998.

[9] Fuqianq Fu and Damian Lyons. An approach to robust homing with stereovision. *SPIE Defense & Security 2017 Conference on Unmanned Systems Technology XX*, April 2018.

[10] Benjamin Graeuba, M. Chappell, Hannah Wittm, Samuel Ledermanne, Rachel Bezner-Kerrf, and Barbara Gemmill-Herrena. "the state of family farms in the world. *World Development*, 87, 2016.

[11] Jose Guivant, Juan Nieto, Favio Masson, and Eduardo Nebot. Navigation and mapping in large unstructured environments. *The International Journal of Robotics Research*, page 449–472.

[12] Vitor Guizilini, Igor Vasiljevic, Rares Ambrus, Gregory Shakhnarovich, and Adrien Gaidon. Full surround monodepth from multiple cameras. *IEEE International Conference on Robotics and Automation*, 2022.

[13] Yiannis Kantaros and Michael M. Zavlanos. Distributed intermittent communication control of mobile robot networks under time-critical dynamic tasks. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5028–5033, 2018.

[14] Ivan Krasin et al. Openimages: A public dataset for large-scale multi-label and multi-class image classification. *Dataset available from https://github.com/openimages*, 2016.

[15] D. Lambrinos et al. Mobile robot employing insect strategies for navigation. *Robotics and Autonomous Systems*, 30, 2000.

[16] S. Lowry, S. Sunderhauf, N. Newman, and J.J. Leonard. Visual place recognition: a survey. *IEEE Trans. Robot. 1–19*, 2015.

[17] Damian Lyons, Ben Barriage, and Luca Del Signore. The effect of horizontal field of view on stereovision-based visual homing. *Robotica*, 38(5), 2020.

[18] Damian Lyons and Noah Petzinger. Visual homing for robot teams: Do you see what i see? *SPIE Conference on Unmanned Systems Technology (UST)*, April 2022.

[19] J. Mao, X. Hu, L. Zhang, and M. Milford. A bio-inspired goal-directed visual navigation model for aerial mobile robots. *Journal of Intelligent and Robotics Systems*, (100):289–310, 2020.

[20] Colin Mcmanus et al. Autonomous navigation and mapping in gps-denied environments at defence r&d canada. *Proceedings of NATO Symposium SET 168: Navigation Sensor and Systems in GNSS Denied Environments*, 2012.

[21] P. Nirmal and D. Lyons. Homing with stereovision. *Robotica*, 34(12), 2015.

[22] Mohamed Rahouti, Kaiqi Xiong, and Nasir Ghani. Bitcoin concepts, threats, and machine-learning security solutions. *IEEE Access*, 6:67189–67205, 2018.

[23] R. Reid and T. Braunl. Large-scale multi-robot mapping in magic 2010. *IEEE 5th International Conference on Robotics, Automation and Mechatronics (RAM)*, 2011.

[24] Davide Scaramuzza et al. Vision controlled micro flying robots. *IEEE/ Rob. & Aut. Magazine*, Sept. 2014.

[25] Annett Stelzer, Mallikarjuna Vayugundla, Elmar Mair, Michael Suppa, and Wolfram Burgard. Towards efficient and scalable visual homing. *Int. J. Robotics Research*, 37(2-3), 2018.

[26] A. Vardy. Long-range visual homing. *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, 2006.

[27] Viktor Vasylkovskyi, Sérgio Guerreiro, and João Silva Sequeira. Blockrobot: Increasing privacy in human robot interaction by using blockchain. In *2020 IEEE International Conference on Blockchain (Blockchain)*, pages 106–115. IEEE, 2020.

[28] Q. Wu et al. Image-goal navigation in complex environments via modular learning. *Arxiv*, 2022.

[29] Fan Yang, Chao Cao, Hongbiao Zhu, Jean Oh, and Ji Zhang. Far planner: Fast, attemptable route planner using dynamic visibility update. *arXiv:2110.09460 [cs.RO]*, 2022.