Wait, That Feels Familiar: Learning to Extrapolate Human Preferences for Preference-Aligned Path Planning

Elvin Yang[2], Haresh Karnan[1], Garrett Warnell[2,4], Peter Stone[2,5], Joydeep Biswas[2]

***Abstract*—Autonomous mobility tasks such as last-mile delivery require reasoning about operator-indicated preferences over different types of terrain to ensure robot safety and mission success. However, coping with *out of distribution* data, such as encountering novel or visually distinct terrains due to lighting variations, remains a fundamental problem in visual navigation. Existing solutions either require labor-intensive manual data recollection and labeling or use hand-coded reward functions that may not align with operator preferences. In this work, we posit that in many situations, operator preferences over novel terrains can be inferred by relating *inertial-tactile* observations of novel terrains to known terrains experienced by the robot. Leveraging this insight, we introduce *Preference Adaptation for Terrain-awarE Robot Navigation* (PATERN), a novel framework for extrapolating operator terrain preferences for visual navigation. PATERN learns an inertial-tactile representation space from the robot's experience and uses nearest-neighbor search in this space to estimate operator preferences over novel terrains. Through physical robot experiments in off-road environments, we evaluate PATERN's adaptability to novel terrains and challenging lighting conditions, and in comparison to baseline approaches, we find that PATERN successfully generalizes to novel terrains and varied lighting conditions while being aligned with operator preferences.**

***Index Terms*—Vision-Based Navigation, Learning from Experience.**

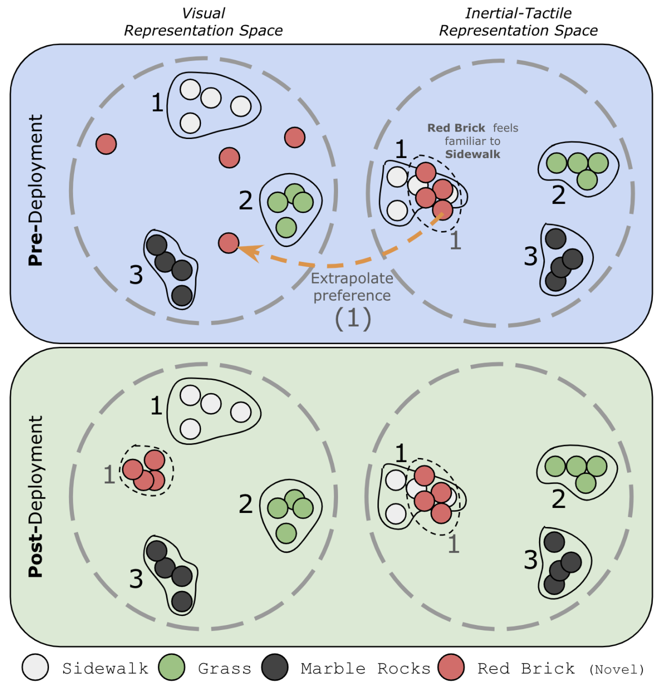

Fig. 1. An illustration of the intuition behind preference extrapolation in PATERN. Shown here are the visual and inertial-tactile representation spaces containing three known and one novel terrain, across two deployment stages. Operator preferences of the three known terrains are marked numerically, with 1 being the most preferred and 3 being the least preferred. In the pre-deployment stage, a novel terrain (`red brick`) is encountered and the preference order of its nearest neighbor (`sidewalk`) inferred from inertial-tactile representations is transferred (extrapolated) to the corresponding samples in the visual representation space. The extrapolated preference order is used within a learning procedure to update both the visual representations and the visual preference function. The post-deployment stage shows extrapolated preferences in both the inertial-tactile and updated visual representation spaces for the novel terrain samples. Note that the inertial-tactile representations are not updated and are used only to supervise the preference ordering of novel terrains in relation to known terrains.

I. INTRODUCTION

Autonomous mobile robots traversing off-road environments need to visually identify distinct terrain features in order to navigate in an operator-preference-aligned manner, to ensure their safety and mission success. However, during autonomous deployment, robots may encounter novel terrains [1], [2] and dynamic real-world conditions such as varied lighting, which may lie outside the known training distribution, posing a significant challenge for vision-based off-road navigation [3].

Equipping robots with the capability to handle novel terrain conditions for preference-aligned path planning is a challenging problem in off-road navigation. Prior approaches to address this problem include collecting more expert demonstrations [4], [5], [6], labeling additional data [7], [8], [9], and utilizing hand-coded reward functions to assign traversability costs [10], [11], [12]. While these approaches have been successful at off-road navigation, collecting more expert demonstration data and labeling may be labor-intensive and expensive, and utilizing hand-coded reward functions may not always align with operator preferences. We posit that in certain cases, while the terrain may look visually distinct in comparison to prior experience, similarities in the inertial-tactile space may be leveraged to extrapolate human preferences over such terrains. For instance, assuming a robot has experienced `cement sidewalk` and `marble rocks`, and prefers the former over the latter (as expressed by the operator), when the robot experiences a visually novel terrain such as `red bricks`, which feels inertially similar to traversing over `sidewalk`, it is more likely that the robot should also prefer `red bricks` over `marble rocks`. While it is not possible to know the operator's true preferences without query-

[1]Walker Department of Mechanical Engineering, The University of Texas at Austin `haresh.miriyala@utexas.edu`

[2]Department of Computer Science, The University of Texas at Austin `{eyang, joydeepb, pstone}@cs.utexas.edu`

[4]Army Research Laboratory `garrett.a.warnell.civ@army.mil`

[5]Sony AI, North America

ing them, we hypothesize that in cases where the operator is unavailable, extrapolating preferences from the inertial-tactile space is a plausible way to estimate traversability preferences for novel terrains.

Leveraging the intuition of extrapolating operator preferences for visually distinct terrains that are familiar in the inertial-tactile space, we introduce *Preference Adaptation for Terrain-awarE Robot Navigation* (PATERN), a novel framework for extrapolating operator terrain preferences for visual navigation. PATERN learns an inertial-tactile latent representation space from the robot's prior experience and uses nearest-neighbor search in this space to estimate operator preferences for novel terrains. Fig. 1 provides an illustration of the intuition behind preference extrapolation in PATERN. We conduct extensive physical robot experiments, evaluating PATERN against state-of-the-art off-road navigation approaches, and find that PATERN is empirically successful with respect to preference alignment and in adapting to novel terrains and lighting conditions seen in the real world.

II. RELATED WORK

In this section, we review related work on visual off-road navigation, with a focus on preference-aligned path planning.

A. Supervised Methods

To learn terrain-aware navigation behaviors, several prior methods have been proposed that used supervised learning from large curated datasets [8], [9] as supervision to pixel-wise segment terrains [7]. Guan et al. [7] propose a transformer-based architecture (GANav) to segment terrains, and manually assign traversability costs for planning. While successful at preference-aligned navigation, fully-supervised methods suffer from domain shift on novel terrains and may require additional labeling.

B. Self-Supervised Methods

To alleviate the need for large-scale datasets for off-road navigation, several self-supervised learning methods have been proposed that learn from data collected on the robot. Specifically, prior methods in this category have explored using inertial Fourier features [10], odometry errors [12], future predictive models [11], and trajectory features [13] to learn traversability costs for off-road navigation. While successful in several off-road navigation tasks such as *comfort-aware navigation* [10], such methods use a hand-coded reward/cost model to solve a specific task and do not reason about operator preferences over terrains. In contrast with prior methods, PATERN utilizes the prior experience of the robot and extrapolates operator preferences to novel terrains.

Sikand et al. propose VRL-PAP [6] in which both a visual representation and a visual preference cost are learned for preference-aligned navigation. However, to handle novel terrains, VRL-PAP requires additional human demonstrations which may not always be available. Orthogonal to VRL-PAP, PATERN focuses on extrapolating operator preferences from known to visually novel terrains.

III. APPROACH

In this section, we formulate the problem of preference-aligned path planning and introduce our proposed approach PATERN, to extrapolate operator preferences to novel terrains.

A. Preference-Aligned Path Planning

We formulate the problem of preference-aligned path planning as a local planning problem. We assume the robot has a state space $\mathcal{S}$, action space $\mathcal{A}$, and deterministic transition function $\mathcal{T} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$. The state space of the robot is composed of states $s = [x, y, \theta, \phi_i, \phi_v]$, in which $[x, y, \theta] \in SE(2)$ denote the robot's position and $[\phi_i, \phi_v]$ are inertial and visual properties of the ground, experienced by the robot, respectively. Given a goal location G, the local planning problem is to use model predictive control to reach G such that the resulting trajectory does not violate operator-defined terrain preferences.

We use a receding horizon motion planner based on constant-curvature arcs $\{\Gamma_1, \Gamma_2, \ldots, \Gamma_k\}, \Gamma \in \mathcal{S}^N$ each comprised of a fixed number of equally-spaced intermediate states. At each time step, the planner truncates the arcs such that they do not intersect with obstacles in the environment and solves for the optimal arc $\Gamma^* = \underset{\Gamma}{\arg\min} J(\Gamma, G)$ minimizing an objective function $J(\Gamma, G)$, $J : (\mathcal{S}^N, \mathcal{S}) \rightarrow \mathbb{R}^+$. For terrain preference-based planning, we define J as

$$J(\Gamma, G) = J_g(\Gamma(N), G) + J_t(\Gamma), \tag{1}$$

where J_g measures distance-based cost to the goal G, and J_t evaluates a preference-based cost of the terrain traversed by Γ. As in prior work [6] [7], J_t is based on a learned function that takes some form of visual observation as input and outputs a real-valued scalar. Given a constant curvature arc Γ, $J_t(\Gamma)$ can be evaluated as:

$$J_t(\Gamma) = \frac{1}{N} \sum_{t=1}^{N} \lambda^i u_{vis}(f_{vis}(\Gamma(i))), \tag{2}$$

where $f_{vis}(\Gamma(i))$ denotes a mapping from RGB space of an image patch of terrain at intermediate state i along Γ to a learned visual representation ϕ_{vis}. The function $u_{vis}(\cdot)$ maps from the visual representation space to a real-valued non-negative scalar.

B. Extrapolating to Visually Novel Terrains

In contrast to previous work, PATERN utilizes two parallel pipelines with disjoint sensing modalities. The first uses visual terrain data that a robot is able to observe before traversing terrain. The second uses inertial-tactile data that a robot must drive over terrain to collect. Each pipeline is comprised of an encoder ($f_{vis} : \mathbb{V} \rightarrow \Phi_{vis}$ and $f_{it} : \mathbb{I} \rightarrow \Phi_{it}$) and a preference-aligned utility function ($u_{vis} : \Phi_{vis} \rightarrow \mathbb{R}^+$ and $u_{it} : \Phi_{it} \rightarrow \mathbb{R}^+$).

Representation Learning: The goal of the encoders is to learn intermediate normalized representation spaces that simultaneously map similar terrain features to the same regions while separating dissimilar terrain features. We use a triplet

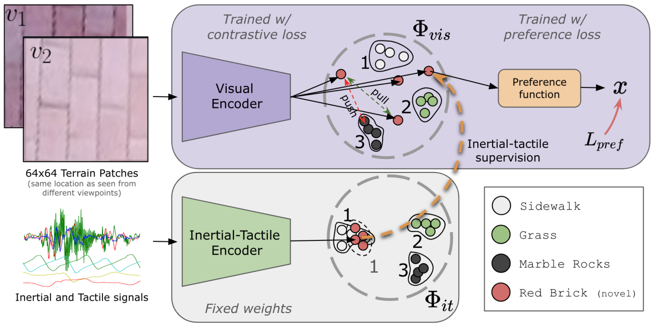

Fig. 2. An illustration of the training setup utilized in PATERN. We utilize two encoders to map visual and inertial samples observed at the same location, to Φ_{vis} and Φ_{it} respectively.

loss-based [14] representation learning approach to learn parameters for each encoder such that

$$\forall a, p, n \in f(\cdot).\ d(a, n) - d(a, p) \geq \delta,$$

where d is a distance metric, and δ is a margin of separation between different terrain features. We use angular distance $d_\theta(x, y) = \arccos \frac{x \cdot y}{\|x\|\|y\|}$ and margin $\delta = 1$.

Learning Preference-Aligned Utility Functions: u_{vis} and u_{it} are learned from a partial order obtained through self-supervision or initial specification by an operator. We use a modified margin ranking loss [14] with an additional term for equivalent preferences $y = 0$.

$$L_{pref}(x_1, x_2, y) = \begin{cases} |x_1 - x_2| & y = 0 \\ \max(0, -y * (x_1 - x_2) + \delta) & y \in \{-1, 1\} \end{cases} \tag{3}$$

Inferring Preferences for Novel Terrains: During robot deployments, divergences between the evaluations of the predictive visual utility function and the retroactive inertial-tactile utility function for the same geolocation indicate that the robot has traversed a novel terrain T_{k+1}. In the case that the novel terrain exhibits a similar inertial-tactile response to a known terrain, we posit that (1) the divergence occurred due to visual properties of the novel terrain and (2) the novel terrain can inherit operator preferences assigned to the known terrain.

To determine whether the novel terrain exhibits similar inertial-tactile responses to a known terrain, we calculate the centroid C_{k+1} of its intermediate inertial-tactile representations and compare it to the centroids $C_1, C_2, \ldots C_k$ of the pre-deployment terrain categories in the inertial-tactile representation space. In addition, we calculate the mean angular deviation of the inertial-tactile representations of T_{k+1}:

$$\varepsilon_{k+1} = \frac{1}{n} \sum_{i=1}^{n} \arccos \frac{C_{k+1} \cdot p_i}{\|C_{k+1}\|\|p_i\|}$$

When the angular difference between the nearest pre-deployment centroid $C_\star = \underset{i \in [1,k]}{\arg\min}\ d(C_i, C_{k+1})$ and C_{k+1} is less than the margin δ and $\varepsilon_{k+1} < \delta$, we infer a preference $T_{k+1} = T_\star$ to self-supervise an updated visual utility function.

Fig. 2 shows the training setup and architecture utilized in PATERN.

IV. EXPERIMENTS

We seek to empirically answer the following questions:

(Q_1) Is PATERN able to accurately infer preferences to rectify visual prediction errors caused by visually novel terrains and varying lighting conditions?

(Q_2) How well does PATERN adhere to operator preferences as compared to baseline methods in short-scale navigation tasks in environments with many visually distinct terrains?

A. Setup and Implementation

The robot platform used for experiments is a Boston Dynamics Spot equipped with a Velodyne VLP-16 LiDAR, Vectornav VN-100 IMU, Microsoft Azure Kinect DK camera, and NVIDIA GeForce RTX 3060 mobile GPU.

We collect a dataset of trajectories over different types of terrain around the UT Austin campus during the day by teleoperating the robot with varying linear and angular velocities. The final dataset consists of approximately 10 minutes of training data and 5 minutes of holdout evaluation data for the following terrains: `cement pavement`, `grass`, `rocks`, `pebble pavement`, and `bushes`. We define the operator preferences as `cement pavement` = `pebble pavement` $\succ$ `grass` $\succ$ `rocks` $\succ$ `bushes`.

We compare PATERN to the following baselines:

Obstacle-Avoidance Only: A purely geometric obstacle-avoidant planner.

RCA: Our implementation of the self-supervised RCA [10] algorithm.

GANav: A semantic segmentation framework trained on RUGD [8] using the implementation provided by the authors.

Fully-Supervised: A visual terrain cost function learned end-to-end using supervised costs derived from operator preferences.

Human Reference: A preference-aligned reference trajectory of the robot, teleoperated by a human.

The RCA and FULLY-SUPERVISED baselines are trained using the entirety of our collected dataset.

TABLE I
HAUSDORFF DISTANCE OF TRAJECTORIES TRACED BY DIFFERENT APPROACHES TO A HUMAN REFERENCE TRAJECTORY.

Approach	Environment				
	1	2-D	2-N	3-D	3-N
Obstacle-Avoidance Only	2.87	2.34	2.34	3.44	3.69
RCA[10]	0.84	0.91	6.061	2.57	7.37
GANav[7]	1.47	2.98	3.07	0.898	1.42
Fully-Supervised (Day)	0.58	0.44	2.735	0.763	6.747
PATERN Pre-Deployment Model	0.54	2.31	2.29	2.305	5.76
PATERN Post-Deployment Model	-	0.56	1.097	0.86	0.763

B. Short-Scale Experiments

We evaluate PATERN in several environments within the UT Austin campus containing a variety of different terrains as shown in Fig. 3. In each environment, we run repeated trails of each baseline method and a "pre-deployment" instance of

Fig. 3. Trajectories traced by PATERN and other baseline approaches across three different environments and two varied lighting conditions within the UT Austin campus. Note the drastic changes in the appearance of the terrain between day and night, which causes a significant challenge for visual off-road navigation.

PATERN trained on a restricted subset of our dataset with only the `cement`, `grass`, and `rocks` terrains in overcast lighting conditions. We then run trials of a self-supervised "post-deployment" instance of PATERN that uses its new experience in these environments to extrapolate preferences to the novel terrains or novel lighting conditions. Table I shows quantitative results using the mean undirected Hausdorff distance between a human reference trajectory and evaluation trajectories of each method.

In Env. 1, the initial PATERN model successfully navigates in an operator-preference-aligned manner. However, in Envs. 2 and 3, during the day, we see that the pre-deployment model fails to adhere to operator preferences when encountering novel terrains such as `rocks` and `bush` as well as new lighting conditions on cement pavement. Through preference-extrapolation in the post-deployment stage using PATERN, we see that the updated model is able to successfully navigate in an operator-preference-aligned manner. While the fully supervised baseline has slightly lower Hausdorff distances in these two settings compared to the updated PATERN model, neither of the two approaches traversed on unpreferable terrain. In Envs. 2 and 3 during the night we see that both the pre-deployment model and many baseline approaches fail due to novel lighting conditions on terrains such as `marble rock` even though the same baseline models were able to navigate successfully during the day. However, through extrapolation using PATERN on night-time data, the robot is able to successfully traverse in an operator-preference-aligned manner.

V. LIMITATIONS

PATERN uses similarities between novel terrains and known terrains in its learned inertial-tactile representation space to infer that the novel terrain is equally preferable to the known terrain. Thus, it must have prior knowledge of a terrain with similar inertial-tactile characteristics to make this association correctly. Additionally, PATERN utilizes inertial-tactile observations that require a robot to physically drive over terrains, which may not always be feasible.

VI. CONCLUSIONS

In this work, we present *Preference Adaptation for Terrain-awarE Robot Navigation* (PATERN), a novel framework for extrapolating operator terrain preferences for visual navigation. PATERN learns an inertial-tactile representation space to detect similarities between visually novel terrains and its set of known terrains and self-supervise its visual utility function to be aligned with operator preferences. We demonstrate that PATERN successfully extrapolates terrain preferences through experience in environments with novel terrains and lighting conditions to improve its ability to navigate according to operator terrain preferences.

ACKNOWLEDGMENT

This work has taken place in the Autonomous Mobile Robotics Laboratory (AMRL) at UT Austin. AMRL research is supported in part by NSF (CAREER-2046955, IIS-1954778, SHF-2006404), ARO (W911NF-19-2-0333), DARPA (HR001120C0031), Amazon, JP Morgan, and Northrop Grumman Mission Systems. The views and conclusions contained in this document are those of the authors alone.

References

[1] H. Karnan, K. S. Sikand, P. Atreya, S. Rabiee, X. Xiao, G. Warnell, P. Stone, and J. Biswas, "Vi-ikd: High-speed accurate off-road navigation using learned visual-inertial inverse kinodynamics," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 3294–3301.

[2] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning robust perceptive locomotion for quadrupedal robots in the wild," *Science Robotics*, vol. 7, no. 62, p. eabk2822, 2022.

[3] X. Xiao, B. Liu, G. Warnell, and P. Stone, "Motion planning and control for mobile robot navigation using machine learning: a survey," 2020.

[4] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to end learning for self-driving cars," 2016.

[5] Y. LeCun, U. Muller, J. Ben, E. Cosatto, and B. Flepp, "Off-road obstacle avoidance through end-to-end learning," in *Proceedings of the 18th International Conference on Neural Information Processing Systems*, ser. NIPS'05. Cambridge, MA, USA: MIT Press, 2005, p. 739–746.

[6] K. S. Sikand, S. Rabiee, A. Uccello, X. Xiao, G. Warnell, and J. Biswas, "Visual representation learning for preference-aware path planning," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 11 303–11 309.

[7] T. Guan, D. Kothandaraman, R. Chandra, A. J. Sathyamoorthy, K. Weerakoon, and D. Manocha, "Ga-nav: Efficient terrain segmentation for robot navigation in unstructured outdoor environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 8138–8145, 2022.

[8] M. Wigness, S. Eum, J. G. Rogers, D. Han, and H. Kwon, "A rugd dataset for autonomous navigation and visual perception in unstructured outdoor environments," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5000–5007.

[9] P. Jiang, P. Osteen, M. Wigness, and S. Saripalli, "Rellis-3d dataset: Data, benchmarks and analysis," 2020.

[10] X. Yao, J. Zhang, and J. Oh, "Rca: Ride comfort-aware visual navigation via self-supervised learning," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 7847–7852.

[11] G. Kahn, P. Abbeel, and S. Levine, "Badgr: An autonomous self-supervised learning-based navigation system," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1312–1319, 2021.

[12] A. J. Sathyamoorthy, K. Weerakoon, T. Guan, J. Liang, and D. Manocha, "Terrapn: Unstructured terrain navigation using online self-supervised learning," 2022.

[13] L. Wellhausen, A. Dosovitskiy, R. Ranftl, K. Walas, C. Cadena, and M. Hutter, "Where should i walk? predicting terrain properties from images via self-supervised learning," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1509–1516, 2019.

[14] D. P. Vassileios Balntas, Edgar Riba and K. Mikolajczyk, "Learning local feature descriptors with triplets and shallow convolutional neural networks," in *Proceedings of the British Machine Vision Conference (BMVC)*, E. R. H. Richard C. Wilson and W. A. P. Smith, Eds. BMVA Press, September 2016, pp. 119.1–119.11. [Online]. Available: https://dx.doi.org/10.5244/C.30.119