2024 IEEE International Conference on Robotics and Automation (ICRA)
May 13-17, 2024. Yokohama, Japan

2024 IEEE International Conference on Robotics and Automation (ICRA) | 979-8-3503-8457-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICRA57147.2024.10611475

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

Haresh Karnan[1*], Elvin Yang[2*†], Garrett Warnell[3,4], Joydeep Biswas[3], Peter Stone[3,5]

***Abstract*— Autonomous mobility tasks such as last-mile delivery require reasoning about operator-indicated preferences over terrains on which the robot should navigate to ensure both robot safety and mission success. However, coping with *out of distribution* data from novel terrains or appearance changes due to lighting variations remains a fundamental problem in visual terrain-adaptive navigation. Existing solutions either require labor-intensive manual data re-collection and labeling or use hand-coded reward functions that may not align with operator preferences. In this work, we posit that operator preferences for visually novel terrains, which the robot should adhere to, can often be extrapolated from established terrain preferences within the *inertial-proprioceptive-tactile* domain. Leveraging this insight, we introduce *Preference extrApolation for Terrain-awarE Robot Navigation* (PATERN), a novel framework for extrapolating operator terrain preferences for visual navigation. PATERN learns to map inertial-proprioceptive-tactile measurements from the robot's observations to a representation space and performs nearest-neighbor search in this space to estimate operator preferences over novel terrains. Through physical robot experiments in outdoor environments, we assess PATERN's capability to extrapolate preferences and generalize to novel terrains and challenging lighting conditions. Compared to baseline approaches, our findings indicate that PATERN [1] robustly generalizes to diverse terrains and varied lighting conditions, while navigating in a preference-aligned manner.**

I. INTRODUCTION

To ensure the safety, mission success, and efficiency of autonomous mobile robots in outdoor settings, the ability to visually discern distinct terrain features is paramount. This necessity stems not only from direct implications for robot functionality but also from the operator-indicated terrain preferences that the robot must adhere to. Often, these preferences are motivated by the desire to protect delicate landscapes, such as flower beds, or to mitigate potential wear and tear on the robot by avoiding hazardous surfaces. However, during autonomous operations, ground robots frequently face unfamiliar terrains [1], [2] and dynamic real-world conditions such as varied lighting, that lie outside the distribution of visually recognized terrains where operator preferences have been pre-defined. This mismatch presents significant challenges for vision-based outdoor navigation [3].

[1]Walker Department of Mechanical Engineering, The University of Texas at Austin, haresh.miriyala@utexas.edu
[2]University of Michigan, Ann Arbor, eyy@umich.edu
[3]Department of Computer Science, The University of Texas at Austin, {joydeepb, pstone}@cs.utexas.edu
[4]Army Research Laboratory, garrett.a.warnell.civ@army.mil
[5]Sony AI, North America
*Equal Contribution, sorted alphabetically by last name
† Work done while at UT Austin
[1]Project website: https://elvout.github.io/patern/

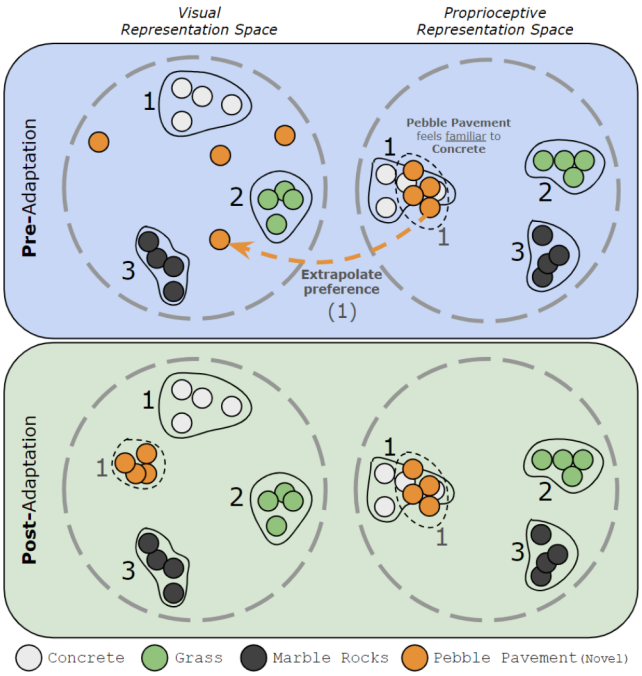

Fig. 1. An illustration of the intuition behind preference extrapolation in PATERN. Operator preferences of the three known terrains are marked numerically, with 1 being the most preferred and 3 being the least preferred. In the pre-adaptation stage, a novel terrain (`pebble pavement`) is encountered and the preference order of its nearest neighbor (`concrete`) inferred from proprioceptive representations is transferred (extrapolated) to the corresponding samples in the visual representation space. The extrapolated preference order is used to update both the visual representations and the visual preference function. The post-adaptation stage shows extrapolated preferences in the updated visual representation space for the novel terrain.

Equipping robots with the capability to handle novel terrain conditions for preference-aligned path planning is a challenging problem in visual 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 visual 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-proprioceptive-tactile space may be leveraged to extrapolate operator preferences over such terrains, that the robot must adhere to. For instance, assuming a robot has experienced `concrete pavement` 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 `pebble pavement` which feels inertially similar to traversing over `concrete pavement`, it is more likely that the operator might also prefer `pebble pavement` over `marble rocks`. While it is not possible to know the operator's true preferences without querying them, we submit that in cases where the operator is unavailable, hypothesizing preferences through extrapolation from the inertial-proprioceptive-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-proprioceptive-tactile space (collectively known as *proprioceptive* for brevity), we introduce *Preference extrApolation for Terrain-awarE Robot Navigation* (PATERN) [2], a novel framework for extrapolating operator terrain preferences for visual navigation. PATERN learns a proprioceptive latent representation space from the robot's prior experience and uses nearest-neighbor search in this space to estimate operator preferences for visually novel terrains. Fig. 1 provides an illustration of the intuition behind preference extrapolation in PATERN. We conduct extensive physical robot experiments on the task of preference-aligned off-road navigation, evaluating PATERN against state-of-the-art 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 in 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 use supervised learning on large curated datasets [8], [9], [14] 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 visual navigation, several self-supervised learning methods have been proposed that learn from data collected on the robot [15]. Specifically, prior methods in this category have explored using inertial Fourier features [10], reinforcement learning [16], contact vibrations [17], [18], proprioceptive feedback [19], odometry errors [12], future predictive models [11], acoustic features [20], and trajectory features [21] to learn traversability costs for visual navigation. While successful in several visual 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. Similarly, STERLING[22] introduces a self-supervised representation learning approach for visual representation learning. However, a limitation for both VRL-PAP and STERLING is their dependence on additional human feedback when dealing with novel terrains, which might not be immediately available during deployment. Distinct from VRL-PAP and STERLING, PATERN focuses on extrapolating operator preferences from known terrains to visually novel terrains.

III. PRELIMINARIES

We formulate preference-aligned planning as a local path-planning problem in a state space $\mathcal{S}$, with an associated action space $\mathcal{A}$. The forward kino-dynamic transition function is denoted as $\mathcal{T} : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ and we assume that the robot has a reasonable model of $\mathcal{T}$ (*e.g.,* using parametric system identification [23] or a learned kino-dynamic model [24], [1], [25]), and that the robot can execute actions in $\mathcal{A}$ with reasonable precision. For ground vehicles, a common choice for $\mathcal{S}$ is $\mathrm{SE}(2)$, which represents the robot's x and y position on the ground plane, as well as its orientation θ.

The objective of the path-planning problem can be expressed as finding the optimal trajectory $\Gamma^* = \arg\min_{\Gamma} J(\Gamma, G)$ to the goal G, using any planner (e.g. a sampling-based motion planner like DWA [26]) while minimizing an objective function $J(\Gamma, G)$, $J : (\mathcal{S}^N, \mathcal{S}) \to \mathbb{R}^+$. Here, $\Gamma = \{s_1, s_2, \ldots, s_N\}$ denotes a sequence of states. The sequence of states in the optimal trajectory Γ^* is then translated into a sequence of actions, using a 1-D time-optimal controller, to be played on the robot. For operator preference-aligned planning, the objective function J is articulated as,

$$J(\Gamma, G) = J_G(\Gamma(N), G) + J_P(\Gamma), \tag{1}$$

Here, J_G denotes geometric costs based on the proximity of the robot's state to the goal G and obstacle avoidance, while J_P imparts a cost based on terrain preference. Crucially, J_P is designed to capture operator preferences over different terrains; less preferred terrains incur a higher cost. Though earlier studies leverage human feedback to ascertain J_P for unfamiliar terrains [6], [22], in this work, we hypothesize that in certain situations, operator preferences for novel terrains can be extrapolated from known terrains, obviating operator dependency during real-world deployment. Thus, our novel contribution is a self-supervised framework for extrapolating J_P from known terrains to visually novel terrains by leveraging inertial-proprioceptive-tactile observations of a robot, without inquiring additional human feedback.

[2] A preliminary version of this work was presented at the PT4R workshop at ICRA 2023 [13]

IV. APPROACH

In this section, we present *Preference extrApolation for Terrain-awarE Robot Navigation* (PATERN), a novel framework for extrapolating operator preferences for preference-aligned navigation. We first detail an existing framework for terrain-preference-aligned visual navigation. We then introduce PATERN for self-supervised extrapolation of operator preferences from known terrains to visually novel terrains by leveraging proprioceptive feedback.

A. A Two-Step Framework for Preference-Aligned Planning

For real-time preference-aligned planning, inspired from earlier studies [6], [22], we postulate that $J_P(\Gamma)$ can be estimated in a two-step approach from visual observations of patches of terrain at $s \in \mathcal{S}$ along Γ. Let $O \in \mathcal{O}$ represent these observations. We denote Π as a projection operator that extracts visual observation O of terrain at s by yielding image patches from homography-transformed bird's eye view images [6], [22]. First, a visual encoder, denoted as f_{vis}, maps O from the RGB space to a latent vector $\phi_{vis} \in \Phi_{vis}$ such that observations from identical terrains cluster closely in Φ_{vis} and are distinct from those of differing terrains. Next, a real-valued preference utility is estimated from ϕ_{vis} using a learned preference utility function $u_{vis} : \Phi_{vis} \rightarrow \mathbb{R}^+$ trained with ranked preferences of terrains, derived either from demonstrations [6], or by active querying [22]. Adopting the popular formulation of Zucker et al. [27], we train the utility function with the margin-based ranking loss [28]. To estimate $J_P(\Gamma)$ during planning, we employ an exponential cost formulation given by, $J_P(\Gamma) = \sum_{s \in \Gamma} e^{-u_{vis}[f_{vis}(\Pi(s))]}$, constraining the costs to be non-negative, which we find works well in practice [27]. This two-step framework for estimating $J_P(\Gamma)$ has been utilized successfully in recent works [6], [22] for operator preference-aligned navigation. Training details of the visual encoder and the utility function are provided in Section V.

While the above two-step framework effectively handles known terrains with pre-defined preferences, it faces challenges when the robot encounters visually novel terrains that lie beyond the training distribution of f_{vis} and u_{vis}. Towards addressing this problem, the primary contribution of our work is a self-supervised framework to extrapolate operator preferences to novel terrains and adapting f_{vis} and u_{vis} to ensure successful preference alignment.

B. Extrapolating Preferences for Visually Novel Terrains

Leveraging the intuition that in addition to visual appearance, operator preferences over terrains are likely also based on the "feel" of the underlying terrain such as bumpiness, stability, or traction, we posit that in many situations, operator preferences for novel terrains can be deduced by relating the proprioceptive modality to known terrains. Utilizing these rich, alternate data sources offers deeper insight into terrain properties, enabling us to extrapolate terrain preferences when direct operator feedback is unavailable. Upon initially encountering a novel terrain, before undergoing any adaptation, we designate this stage as the *pre-adaptation* phase. During this phase, the visual encoder and utility function operate based on previously known operator preferences. However, once preferences are extrapolated and the visual encoder and utility functions are subsequently retrained to adapt, the system progresses to the *post-adaptation* phase, as shown in Fig. 1.

Inertial-Proprioceptive-Tactile Encoder and Utility Function: In PATERN, in addition to the visual encoder, we introduce a non-visual encoder that independently processes the inertial, proprioceptive (joint angles and velocities), and tactile feet data—collectively referred to as *proprioception* for brevity—observed by the robot as it traverses a terrain. This encoder maps proprioception observations into a proprioceptive representation space Φ_{pro}, such that representations $\phi_{pro} \in \Phi_{pro}$ of the same terrain are closely clustered whereas those of distinct terrains are farther apart. Additionally, a utility function $u_{pro} : \Phi_{pro} \rightarrow \mathbb{R}^+$ maps the proprioceptive representation vector $\phi_{pro} \in \Phi_{pro}$ to a real-valued preference utility, similar to the visual utility function. Note that, to estimate $J_P(\Gamma)$ during deployment, we only use u_{vis} and not u_{pro} since we cannot observe the proprioceptive components of a future state without traversing the terrain first.

Pre-Adaptation Phase: While traversing known terrains that are in-distribution, the visual and proprioceptive utility values tend to align closely. However, for visually novel terrains, discrepancies often emerge between the utility values predicted from the visual and proprioceptive modalities. In PATERN, we utilize the mean-squared error between the predicted utilities as a signal to detect visually novel, out-of-distribution terrains. Although any novelty detection mechanism can be integrated within PATERN, such as the unimodal approach by Burda et al. [29], our primary focus is on a framework that extrapolates operator preferences for novel terrains. Moreover, any foundational approach employing the two-step framework for preference cost estimation [6], [22], [30], as elaborated in Subsection IV-A, can be utilized in the pre-adaptation phase. For clarity, we use the notation PATERN$^-$ to represent the baseline algorithm in its unadapted state and PATERN$^+$ to indicate the updated model in the post-adaptation phase.

Extrapolating Operator Preferences: Given a novel terrain segment for which operator preferences are unknown, we propose to self-supervise preference assignment by first clustering its proprioceptive representations ϕ_{pro} and then associating it with the closest known existing cluster in Φ_{pro}, assigning the same operator preference as the known-cluster, as illustrated in Fig. 1. Following this self-supervised preference assignment, the visual encoder and visual utility function for novel terrain segments are finetuned by aggregating newly gathered experience with existing data.

V. IMPLEMENTATION DETAILS

In this section, we describe the implementation details of PATERN. We first describe data pre-processing, followed by training details in the pre-adaptation phase. Finally, we describe adapting the visual encoder and utility function using the extrapolated preference in PATERN.

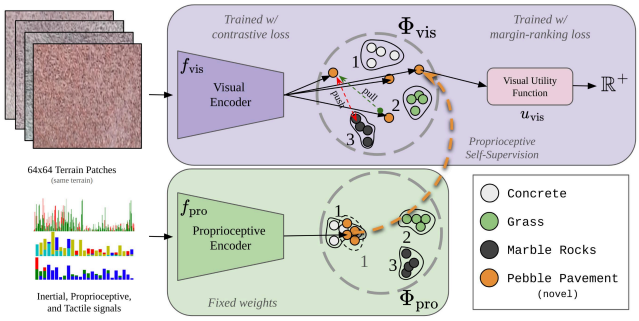

Fig. 2. An illustration of the training setup for preference extrapolation proposed in PATERN. We utilize two encoders to map visual and inertial-proprioceptive-tactile samples to Φ_{vis} and Φ_{pro} respectively. For a visually novel terrain, a preference is hypothesized and extrapolated from Φ_{pro}, following which the visual encoder and utility function are retrained.

A. Data Pre-Processing

In tandem with the visual patch extraction process used in the projection operator Π as in prior methods [6], [22], for every state s_t, we also extract a 2-second history of time-series inertial (angular velocities along the x and y-axes and linear acceleration in the z-axis), proprioceptive (joint angles and velocities), and tactile (feet depth penetration estimates) data. To ensure the resulting input data representation for training is independent of the length and phase of the signals, we compute statistical measures of center and spread as well as the power spectral density, and maintain that as the input. All the visual patches extracted with the projection operator Π and the non-visual data for each state s are then tagged with their corresponding terrain name, given that each trajectory uniquely contains a particular terrain type. In addition to processing the recorded data in the pre-adaptation phase, a human operator is queried for a full-order ranking of terrain preference labels.

B. Pre-Adaptation Training

We use a supervised contrastive learning formulation inspired by Sikand et al. [6] to train the baseline functions f_{vis} and u_{vis}, represented as neural networks.

Training the Encoders: Given labeled visual patches and proprioception data, we generate triplets for contrastive learning such that for any anchor, the positive pair is chosen from the same label and the negative pair is sourced from another label. Given such triplets, we use triplet loss [31] with a margin of 1.0 to independently train the visual and proprioception encoders through mini-batch gradient descent using the AdamW optimizer. For the visual encoder, we use a 3-layer CNN of 5×5 kernels, each followed by ReLU activations. This model, containing approximately 250k parameters, transforms 64×64 size RGB image patches into an 8-dimensional vector representation ϕ_{vis}. Similarly, our inertial encoder consists of a 3-layer MLP with ReLU activations, encompassing around 4k parameters, and maps proprioceptive inputs to an 8-dimensional vector ϕ_{pro}. To mitigate the risk of overfitting, data is partitioned in a 75-25 split for training and validation, respectively.

Training the Utility Functions: In our setup, the utility function is represented as a two-layer MLP with ReLU non-linearity and output activation that maps an 8-dimensional vector into a singular non-negative real value. Given ranked operator preferences of the terrains, we follow Zucker et al. [27] and train the visual utility function u_{vis} using a margin-based ranking loss [28]. Furthermore, to ensure consistent predictions from u_{vis} and u_{pro} for both visual and non-visual observations at identical locations, we update parameters of u_{pro} using the loss $\mathcal{L}_{\text{MSE}}(u_{\text{pro}}) = \frac{1}{N}\sum_{i=1}^{N}\left(sg(u_{\text{vis}}(\phi_{vis})) - u_{\text{pro}}(\phi_{pro})\right)^2$. Here, $sg(\cdot)$ denotes the stop-gradient operation, and ϕ_{vis} and ϕ_{pro} are the terrain representations from paired visual and non-visual data, respectively, at the same location.

The functions f_{vis} and u_{vis} prior to adaptation are collectively termed as PATERN$^-$, signifying their non-adapted state with respect to visually novel terrains. In our implementation, although we use supervised contrastive learning, in instances where explicit terrain labels might be absent, one can resort to self-supervised representation learning techniques, such as STERLING [22], to derive f_{vis} and u_{vis}. PATERN can be applied regardless of the specific representation learning approach used.

C. Preference Extrapolation Training

During deployment, if the robot encounters a visually novel terrain, both visual and inertial-proprioceptive-tactile data is recorded to be used in the adaptation phase in PATERN, aiding in preference extrapolation and subsequent model adaptation. We refer to this collected data as the *adaptation-set*. We extract paired visual and non-visual observations at identical locations from the *adaptation-set* and use f_{pro} to extract proprioceptive representations ϕ_{pro}. We cluster samples of ϕ_{pro} and perform a nearest-neighbor search against existing terrain clusters from the pre-adaptation dataset that is within a threshold μ. We set this threshold to be the same as the triplet margin value of 1.0 which we find to work well in practice. This procedure seeks a known terrain that "feels" similar to the novel terrain which then inherits the preference of its closest match. Following this self-supervised preference extrapolation framework, the adaptation-set is aggregated with the pre-adaptation training set, and the visual encoder f_{vis} is retrained using the procedure described in V-B. Additionally, the visual utility function u_{vis} is retrained with the extrapolated preference for the novel terrain. The updated functions f_{vis} and u_{vis} are collectively referred to as PATERN$^+$. Figure 2 illustrates retraining and preference extrapolation as described above.

VI. EXPERIMENTS

In this section, we describe the physical robot experiments conducted to evaluate PATERN against other state-of-the-art visual off-road navigation algorithms. Specifically, our experiments are designed to explore the following questions:

(Q_1) Is PATERN capable of extrapolating operator preferences accurately to novel terrains?

(Q_2) How effectively does PATERN navigate under challenging lighting scenarios such as nighttime conditions?

(Q_3) How well does PATERN perform in large-scale real-world off-road conditions?

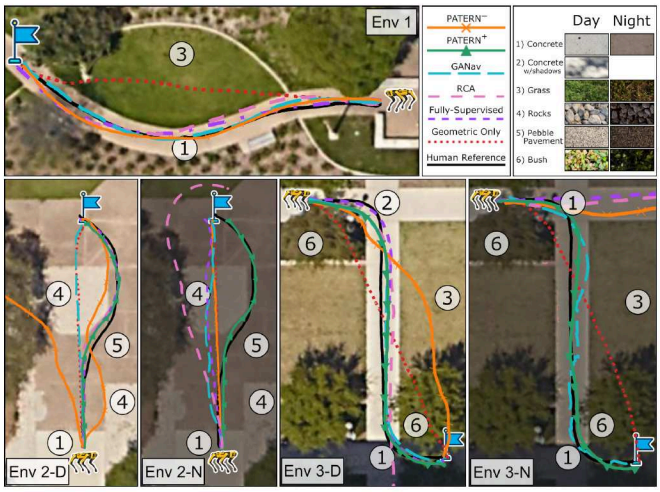

Fig. 3. Trajectories traced by PATERN and baseline approaches across three different environments and varied lighting conditions within the UT Austin campus. Note the drastic changes in the appearance of the terrain between day and night, which pose a significant challenge for visual navigation. In environments where PATERN$^-$ fails to generalize, PATERN$^+$ successfully extrapolates and reaches the goal in a preference-aligned manner.

To study Q_1 and Q_2, we execute a series of experiments consisting of small-scale outdoor navigation tasks. We then conduct a large-scale autonomous robot deployment along a 3-mile outdoor trail to qualitatively evaluate Q_3. We use the legged Boston Dynamics Spot robot for all experiments, equipped with a VectorNav VN100 IMU, front-facing Kinect RGB camera, Velodyne VLP-16 LiDAR for geometric obstacle detection, and a GeForce RTX 3060 mobile GPU. For local planning, we use an open-source sampling-based motion planner called GRAPHNAV [32] and augment its sample evaluation function with the inferred preference cost $L_P(\Gamma)$. For real-time planning, we run f_{vis} and u_{vis} on the onboard GPU.

TABLE I

MEAN HAUSDORFF DISTANCE RELATIVE TO A HUMAN REFERENCE TRAJECTORY.

Approach	Environment				
	1	**2-D**	**2-N**	**3-D**	**3-N**
Geometric-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	0.58	**0.44**	2.735	**0.763**	6.747
PATERN$^-$	**0.54**	2.31	2.29	2.305	5.76
PATERN$^+$	-	0.56	**1.097**	0.86	**0.763**

A. Data Collection

To collect labeled data for training PATERN$^-$, we manually teleoperated the robot across the UT Austin campus, gathering 8 distinct trajectories each across 3 terrains: `concrete`, `grass`, and `marble rocks`. Each trajectory, five minutes long, is exclusive to a single terrain type for ease of labeling and evaluation. The pre-adaptation training data was collected in the daytime under sunny conditions. Our evaluations then centered on two preference extrapolation scenarios: one, extending to new terrains such as `pebble-pavement`, `concrete-with-shadows`, and `bushes`, all experienced under varying daylight conditions ranging from bright sunlight to overcast skies, and two, adapting to nighttime illumination for familiar terrains that appear visually different. In our experiments, we use the preference order `concrete` $\succ$ `grass` $\succ$ `marble rocks`.

TABLE II

MEAN TRAJECTORY PERCENTAGE ALIGNED WITH OPERATOR PREFERENCES.

Approach	Environment				
	1	**2-D**	**2-N**	**3-D**	**3-N**
Geometric-Only	44.0%	68.8%	68.8%	43.6%	43.6%
RCA[10]	100%	97.3%	67.4%	100%	99.4%
GANav[7]	93.9%	71.6%	71.4%	100%	100%
Fully-Supervised	100%	100%	71.7%	100%	93.6%
PATERN$^-$	**100%**	94.1%	71.6%	81.3%	100%
PATERN$^+$	-	**100%**	**98.2%**	**100%**	**100%**

B. Quantitative Small-Scale Experiments

We evaluate PATERN in three environments with a variety of terrains within the UT Austin campus. We also test under two different lighting conditions, as shown in Fig. 3. The primary task for evaluation is preference-aligned visual off-road navigation, in which the robot is tasked with reaching a goal, while adhering to operator preferences over terrains.

To evaluate the effectiveness of PATERN, we compare it against five state-of-the-art baseline and reference approaches: **Geometric-Only** [32], a purely geometric obstacle-avoidant planner; **RCA** [10], a self-supervised traversability estimation algorithm based on ride-comfort; **GANav**, a semantic segmentation method [3] trained on RUGD dataset [8]; **Fully-Supervised**, an approach that utilizes a visual terrain cost function comprehensively learned using supervised costs drawn from operator preferences; and lastly, **Human Reference**, which offers a preference-aligned reference trajectory where the robot is teleoperated by a human expert. To train the RCA and Fully-Supervised baselines, in addition to the entirety of the pre-adaptation dataset on the 3 known terrains, we additionally collect 8 trajectories each on the novel terrains during daytime.

In each environment, we perform five trials of each method to ensure consistency in our evaluation. For each trial, the robot is relocalized in the environment, and the same goal location G is fed to the robot. In the environments where PATERN$^-$ fails to navigate in a preference-aligned manner, we run five trials of the self-supervised PATERN$^+$ instance that uses experience gathered in these environments to extrapolate preferences to the novel terrains or novel lighting conditions. Fig. 3 shows the qualitative results of trajectories traced by each method in the outdoor experiments. Only one trajectory is shown for each method unless there is significant variation between trials. Table I shows quantitative results

[3]https://github.com/rayguan97/GANav-offroad

using the mean Hausdorff distance between a human reference trajectory and evaluation trajectories of each method. Table II shows quantitative results for the mean percentage of preference-aligned distance traversed in each trajectory. Note that both the reported metrics may be high if a method does not reach the goal but stays on operator-preferred terrain.

From the quantitative results, we see that, as expected, the PATERN$^-$ approach is able to successfully navigate in an operator-preference-aligned manner in Env. 1, which did not contain any novel terrain types. However, PATERN$^-$ fails to consistently reach the goal and/or navigate in alignment with operator preferences in the remaining environments. In the daytime experiments, Env. 2 contains a novel terrain (`pebble pavement`) absent from training data for PATERN$^-$, while Env. 3 contains both a novel terrain type (`bush`) and novel visual terrain appearances caused by tree shadows. In the nighttime experiments, all terrains contain novel visual appearances. Following deployments in Envs. 2 and 3, PATERN extrapolates terrain preferences for new visual data using the corresponding proprioceptive data to retrain environment-specific PATERN$^+$ instances. In each environment that the PATERN$^-$ model fails, the self-supervised PATERN$^+$ model is able to successfully navigate to the respective goal in a preference-aligned manner, without requiring any additional operator feedback during deployment, addressing Q_1 and Q_2. While the fully-supervised baseline more closely resembles the human reference trajectory compared to PATERN$^+$ during the day in Envs. 2 and 3, unlike the fully-supervised approach, PATERN$^+$ does not require operator preferences over all terrains and is capable of extrapolating to visually novel terrains.

C. Qualitative Large-Scale Experiment

Fig. 4. Trajectory trace of a large-scale qualitative deployment of PATERN$^+$ along the 3-mile trail. With only five minutes of supplementary data, PATERN required only one manual intervention to stay on the trail and successfully completes the hike, demonstrating robustness and adaptability to real-world off-road conditions.

To investigate Q_3, we execute a large-scale autonomous deployment of PATERN along a challenging 3-mile off-road trail [4]. The robot's objective is to navigate in a terrain-aware manner on the trail by preferring `dirt`, `gravel` and `concrete` over `bush`, `mulch`, and `rocks`. Failure to navigate in a preference-aligned manner may cause catastrophic effects such as falling into the river next to the trail. An operator is allowed to temporarily take manual control of the robot only to prevent such catastrophic effects, adjust the robot's heading for forks in the trail, or yield to pedestrians and cyclists. The PATERN$^-$ model used for small-scale experiments is augmented with approximately five minutes of combined additional data for `dirt`, `bush`, and `mulch` terrains commonly seen in the trail. Following this preference extrapolation, the PATERN$^+$ model is able to successfully navigate the 3-mile trail, while only requiring one human intervention. Fig. 4 shows the trajectory of the robot and a number of settings along the trail, including the single unexpected terrain-related intervention in the lower right corner, for the hour-long deployment. Additionally, we attach a video recording of the robot deployment[5]. This large-scale study addresses Q_3 by qualitatively demonstrating the effectiveness of PATERN in scaling to real-world off-road conditions.

VII. LIMITATIONS AND FUTURE WORK

PATERN uses similarities between novel and known terrains in its learned proprioception representation space to extrapolate preferences. Thus, PATERN needs to have had experiences with terrains bearing close inertial-proprioceptive-tactile resemblances for successful extrapolation. A noticeable limitation is that if a terrain with similar proprioceptive features has not been observed, PATERN might be unable to extrapolate operator preferences. Additionally, PATERN utilizes non-visual observations that require a robot to physically drive over terrains, which may be unsafe or infeasible. Extending PATERN with depth sensors to handle non-flat terrains is a promising direction for future work.

VIII. CONCLUSION

In this work, we present *Preference extrApolation for Terrain-awarE Robot Navigation* (PATERN), a novel approach to extrapolate human preferences for novel terrains in visual off-road navigation. PATERN learns inertial-proprioceptive-tactile representations to detect similarities between visually novel terrains and the set of known terrains. Through this self-supervision, PATERN successfully extrapolates operator preferences for visually novel terrain segments, without requiring additional human feedback. Through extensive physical robot experiments in challenging outdoor environments in varied lighting conditions, we find that PATERN successfully extrapolates preferences for visually novel terrains and is scalable to real-world off-road conditions.

ACKNOWLEDGEMENT

This work has taken place in the Learning Agents Research Group (LARG) and Autonomous Mobile Robotics Laboratory (AMRL) at UT Austin. LARG research is supported in part by NSF, ONR, FLI, ARO, DARPA, Lockheed Martin, GM, and Bosch. AMRL research is supported in part by NSF, ARO, DARPA, Amazon, JP Morgan, Nissan, and Northrop Grumman Mission Systems. Peter Stone serves as the Executive Director of Sony AI America and receives financial compensation for this work. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

[4]Ann and Roy Butler trail, Austin, TX, USA

[5]PATERN deployed in the trail: https://youtu.be/j7159pE0u6s

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] E. Yang, H. Karnan, G. Warnell, P. Stone, and J. Biswas, "Wait, that feels familiar: Learning to extrapolate human preferences for preference-aligned path planning," in *ICRA2023 Workshop on Pretraining for Robotics (PT4R)*, 2023.

[14] A. Zhang, C. Eranki, C. Zhang, J.-H. Park, R. Hong, P. Kalyani, L. Kalyanaraman, A. Gamare, A. Bagad, M. Esteva, and J. Biswas, "Towards robust robot 3d perception in urban environments: The ut campus object dataset," 2023.

[15] J. Frey, M. Mattamala, N. Chebrolu, C. Cadena, M. Fallon, and M. Hutter, "Fast traversability estimation for wild visual navigation," 2023.

[16] H. Karnan, G. Warnell, X. Xiao, and P. Stone, "Voila: Visual-observation-only imitation learning for autonomous navigation," 2021.

[17] C. A. Brooks and K. D. Iagnemma, "Self-supervised classification for planetary rover terrain sensing," in *2007 IEEE aerospace conference*. IEEE, 2007, pp. 1–9.

[18] A. Pokhrel, A. Datar, M. Nazeri, and X. Xiao, "Cahsor: Competence-aware high-speed off-road ground navigation in se(3)," 2024.

[19] A. Loquercio, A. Kumar, and J. Malik, "Learning visual locomotion with cross-modal supervision," 2022.

[20] J. Zürn, W. Burgard, and A. Valada, "Self-supervised visual terrain classification from unsupervised acoustic feature learning," *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 466–481, 2020.

[21] 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.

[22] H. Karnan, E. Yang, D. Farkash, G. Warnell, J. Biswas, and P. Stone, "Self-supervised terrain representation learning from unconstrained robot experience," in *ICRA2023 Workshop on Pretraining for Robotics (PT4R)*, 2023.

[23] N. Seegmiller, F. Rogers-Marcovitz, G. Miller, and A. Kelly, "Vehicle model identification by integrated prediction error minimization," *The International Journal of Robotics Research*, vol. 32, no. 8, pp. 912–931, 2013.

[24] X. Xiao, J. Biswas, and P. Stone, "Learning Inverse Kinodynamics for Accurate High-Speed Off-Road Navigation on Unstructured Terrain," *arXiv e-prints*, p. arXiv:2102.12667, Feb. 2021.

[25] P. Atreya, H. Karnan, K. S. Sikand, X. Xiao, S. Rabiee, and J. Biswas, "High-speed accurate robot control using learned forward kinodynamics and non-linear least squares optimization," in *Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on*. IEEE, 2022, pp. 11 789–11 795.

[26] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics & Automation Magazine*, vol. 4, no. 1, pp. 23–33, 1997.

[27] M. Zucker, N. Ratliff, M. Stolle, J. Chestnutt, J. A. Bagnell, C. G. Atkeson, and J. Kuffner, "Optimization and learning for rough terrain legged locomotion," *The International Journal of Robotics Research*, vol. 30, no. 2, pp. 175–191, 2011.

[28] 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

[29] Y. Burda, H. Edwards, A. Storkey, and O. Klimov, "Exploration by random network distillation," 2018.

[30] D. S. Brown, W. Goo, and S. Niekum, "Better-than-demonstrator imitation learning via automatically-ranked demonstrations," in *Conference on robot learning*. PMLR, 2020, pp. 330–359.

[31] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[32] J. Biswas, "Amrl autonomy stack," https://github.com/ut-amrl/graph_navigation, 2013.