IG-MCTS: Human-in-the-Loop Cooperative Navigation under Incomplete Information

Extended Abstract

Shenghui Chen*
The University of Texas at Austin
Austin, Texas, USA
shenghui.chen@utexas.edu

Ruihan Zhao*
The University of Texas at Austin
Austin, Texas, USA
ruihan.zhao@utexas.edu

Sandeep Chichali
The University of Texas at Austin
Austin, Texas, USA
sandeepc@utexas.edu

Ufuk Topcu
The University of Texas at Austin
Austin, Texas, USA
utopcu@utexas.edu

ABSTRACT

We study human–robot navigation under incomplete information, where a robot with accurate local observations assists a human with an inaccurate global map. We introduce CoNav-Maze and propose Information Gain Monte Carlo Tree Search (IG-MCTS), which plans actions to jointly optimize task reward and human information gain using a learned human perception model. A 14-participant eye-tracking study shows IG-MCTS cuts communication by over 97% and reduces cognitive load relative to teleoperation and instruction-following, while maintaining comparable navigation performance.

KEYWORDS

Human-robot interaction; MCTS; Incomplete information

ACM Reference Format:
Shenghui Chen, Ruihan Zhao, Sandeep Chichali, and Ufuk Topcu. 2026. IG-MCTS: Human-in-the-Loop Cooperative Navigation under Incomplete Information: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/OBKI3637

1 INTRODUCTION

Effective collaboration between humans and robots in complex environments requires efficient information exchange under uncertainty. In many real-world scenarios—such as search and rescue, exploration, or remote inspection—robots operate in partially observed spaces where human operators possess complementary task understanding but face limited communication bandwidth. Teleoperation requires a high cognitive load [9]; instruction-following allows robots to execute high-level human plans [1] but requires substantial communication of accurate instructions. To address these limitations, how can an autonomous agent selectively communicate or act to maximize shared situational awareness and task performance without continuous supervision?

*Equal contribution.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/OBKI3637

This paper studies a cooperative navigation problem where a human and a robot must jointly reach target locations under asymmetric and incomplete information. The human operator possesses a global but inaccurate map, while the robot acquires accurate local observations. We formalize this setting in *CoNav-Maze*, a simulated grid-based environment adapted from MemoryMaze [10]. The robot navigates adjacent cells while the human provides high-level trajectory suggestions based on an evolving internal perception state, refined through robot communication. The environment is modeled as an MDP $(\mathcal{S}, \mathcal{A}, T, R_{\text{env}}, \gamma)$, where states encode the robot's position and remaining goals, actions include movement and image transmission, transitions are deterministic, R_{env} defines task rewards, and $\gamma \in [0, 1)$ is the discount factor. At each step, the robot observes its local surroundings up to a distance d, optionally receives a human-specified trajectory ζ_t, and selects an action a_t to move or transmit an image. The human refines their initially inaccurate perception state $x_t \in \mathcal{X}$ over time based on the robot's actions and communications to provide more accurate guidance.

Problem Statement. Formally, the robot plans over a finite horizon H to maximize the expected cumulative reward that jointly accounts for task progress and information gain:

$$\max_{a_{0:H-1}} \mathbb{E}\left[\sum_{t=0}^{H-1} \gamma^t \left(R_{\text{task}}(\tau_{t+1}, \zeta) + \|x_{t+1} - x_t\|\right)\right], \tag{1}$$

where τ_t denotes the robot's history. The task reward encourages goal reaching, while the information reward captures the degree to which robot actions refine the human's perception.

2 NEURAL HUMAN PERCEPTION MODEL

We model how the human's perceptual state evolves in response to robot communication. As the robot moves and transmits images, the human revises their map by editing the walls. Formally, We define the *Neural Human Perception Model* (NHPM) as

$$x_{t+1} = F_\theta(x_t, \tau_t, o_t),$$

where x_t is the human perception state, τ_t is the robot's recent trajectory, o_t is the transmitted image, and θ are trainable parameters. NHPM is implemented as a fully convolutional network that predicts probabilistic perception updates. Inputs are converted into aligned 2D representations and concatenated channel-wise. The

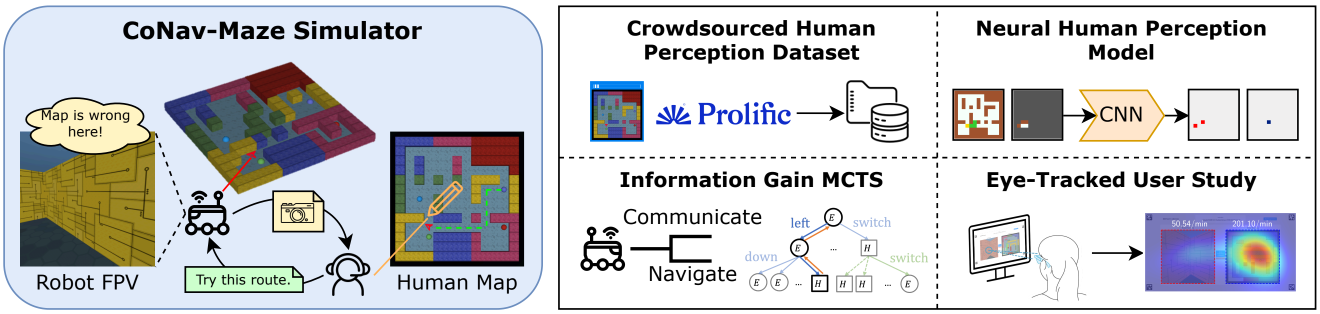

Figure 1: Overview for efficient human–robot collaborative navigation: Left: CoNav-Maze; Right: NHPM trained from crowd-sourced human perceptual updates (Section 2), IG-MCTS (Section 3), and eye-tracked user study evaluation (Section 4).

network outputs two probability maps corresponding to adding and removing walls. We train NHPM using a crowdsourced dataset of 113 human mapping episodes in which participants updated a maze map based on robot trajectories and images [4, 11, 13].

3 INFORMATION GAIN MONTE CARLO TREE SEARCH

To solve Equation (1), we propose *Information Gain Monte Carlo Tree Search* (IG-MCTS), which plans robot actions by jointly optimizing task completion and informative communication [3]. IG-MCTS integrates the following components:

Reward Augmentation. IG-MCTS maximizes $R_{\text{aug}} = R_{\text{task}} + R_{\text{info}}$, where R_{task} promotes efficient navigation toward goals, and $R_{\text{info}} = \|x_{t+1} - x_t\|_1$ measures the magnitude of changes in human perception. Since the robot does not observe x_t directly, R_{info} is estimated using NHPM. Under the model's probabilistic outputs $(p_{\text{add}}, p_{\text{remove}})$, the expected information reward reduces to the sum of predicted edit probabilities across all cells:

$$\mathbb{E}[\|x_{t+1} - x_t\|_1] = \|p_{\text{add}}\|_1 + \|p_{\text{remove}}\|_1.$$

Uncertainty-Aware Planning. IG-MCTS follows standard MCTS phases but incorporates uncertainty during simulation. Each tree node represents a state (τ, x), consisting of the robot's trajectory history and a sampled human perception state. Two action types receive separate treatments:

(1) **Movement actions** follow the maze dynamics with partial observability. Cells beyond the robot's explored area assume a 50% probability of containing a wall. During expansion, they are assigned a feasibility score $\delta = 0.5$; during rollout, feasibility is resolved via sampling.
(2) **Communication actions** follow the learned human perception model.

During backpropagation, node values are updated using feasibility-weighted returns. For a transition with feasibility δ' and child return q', the propagated return is $q = r + \gamma \left[\delta' q' + (1 - \delta') \frac{Q(v)}{N(v)}\right]$, where $\frac{Q(v)}{N(v)}$ provides a fallback estimate when transitions fail. This mechanism enables robust planning under both environmental and human-model uncertainty.

4 EXPERIMENTS

We evaluate (i) the accuracy of the Neural Human Perception Model (NHPM) in predicting human map updates and (ii) the effectiveness of IG-MCTS in reducing human cognitive load and communication while maintaining task performance.

4.1 Human Perception Model Evaluation

We assess NHPM's ability to predict how humans update their internal maps given robot trajectories and shared images. As a baseline, we adapt a Logistic Psychometric Function (LPF) to operate at the grid-cell level, modeling update probability as a function of distance-based stimulus intensity[8, 12, 14]. This Grid-Based LPF (GLPF) treats each cell independently and cannot capture spatial context. Across held-out test environments, NHPM substantially outperforms GLPF in both prediction loss and accuracy, even when GLPF is fit directly on test data. These results demonstrate that modeling spatial structure and contextual dependencies is critical for accurately predicting human perceptual updates.

4.2 User Study

We conducted a within-subject user study comparing IG-MCTS to teleoperation and instruction-following with 14 graduate students (78.6% male, 21.4% female; age 26.8 ± 3.17 years). Participants navigated multiple maze layouts while wearing a Pupil Labs Core eye tracker [7]. We evaluated task performance (robot steps and communication) and human cognitive load using established eye-tracking metrics, including pupil dilation, blink rate, and fixation shifts between the robot's egocentric view and a global map.

Across all tasks, IG-MCTS reduced indicators of cognitive load relative to teleoperation [2]. Participants showed smaller pupil dilation [5] and fewer shifts of visual attention between the robot view and the global map [6], indicating reduced visual and cognitive effort when low-level control was automated. Blink behavior further reflected lower mental workload under IG-MCTS compared to teleoperation [15, 16]. At the same time, IG-MCTS dramatically reduces communication—by over an order of magnitude—relative to both baselines, while achieving comparable or fewer robot steps than instruction-following. These findings support our hypotheses that selective, information-driven communication enables efficient collaboration without increasing human burden.

REFERENCES

[1] Peter Anderson, Qi Wu, Damien Teney, Jake Bruce, Mark Johnson, Niko Sünderhauf, Ian Reid, Stephen Gould, and Anton Van Den Hengel. 2018. Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*. 3674–3683.

[2] Jackson Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin* 91, 2 (1982), 276.

[3] Shenghui Chen, Ruihan Zhao, Sandeep Chinchali, and Ufuk Topcu. 2025. Human-agent coordination in games under incomplete information via multi-step intent. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*.

[4] Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. 2019. Learning latent dynamics for planning from pixels. In *Proceedings of the International Conference on Machine Learning*. 2555–2565.

[5] Eckhard H Hess and James M Polt. 1964. Pupil size in relation to mental activity during simple problem-solving. *Science* 143, 3611 (1964), 1190–1192.

[6] Antony William Joseph and Ramaswamy Murugesh. 2020. Potential eye tracking metrics and indicators to measure cognitive load in human-computer interaction research. *Journal of Scientific Research* 64, 1 (2020), 168–175.

[7] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. 1151–1160.

[8] Stanley A Klein. 2001. Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & psychophysics* 63, 8 (2001), 1421–1455.

[9] MD Moniruzzaman, Alexander Rassau, Douglas Chai, and Syed Mohammed Shamsul Islam. 2022. Teleoperation methods and enhancement techniques for mobile robots: A comprehensive survey. *Robotics and Autonomous Systems* 150 (2022), 103973.

[10] Jurgis Pasukonis, Timothy Lillicrap, and Danijar Hafner. 2022. Evaluating long-term memory in 3d mazes. *arXiv preprint arXiv:2210.13383* (2022).

[11] Nicolaas Prins et al. 2016. *Psychophysics: A practical introduction*. Academic Press.

[12] Bernhard Treutwein and Hans Strasburger. 1999. Fitting the psychometric function. *Perception & psychophysics* 61, 1 (1999), 87–106.

[13] Manuel Watter, Jost Springenberg, Joschka Boedecker, and Martin Riedmiller. 2015. Embed to control: A locally linear latent dynamics model for control from raw images. *Proceedings of the International Conference on Neural Information Processing Systems* 28 (2015).

[14] Felix A Wichmann and N Jeremy Hill. 2001. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & psychophysics* 63, 8 (2001), 1293–1313.

[15] Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. 2016. Measuring cognitive load using eye tracking technology in visual computing. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*. 78–85.

[16] Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. 2018. Studying eye movements as a basis for measuring cognitive load. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–6.