Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ28246

Enabling Robots to Determine the Most Energy-Saving Path Through Visual Sensors

Luohan Chen[1*†], Yutong Hou[2†], Jiaze Hong[3†], Jiawei Zhao[4†], Yichen Sun[5†]

[1]Elite Knowit College, Shenzhen, China
[2]Nanjing Jinling High School, Hexi Campus, Nanjing, China
[3]Nanjing Jinling High School, Nanjing, China
[4]Rosedale Global High School, Shenzhen, China
[5]International Department, High School Affiliated to South China Normal University, Guangzhou, China
**Corresponding Author. Email: n4463433@gmail.com*
[†]All the authors contributed equally to this work and should be considered as co-first author.

Abstract. In this research, the general goal is to design an intelligent algorithm to make the robot decide which path consumes the least energy by itself. With progress in deep learning, robotics, and SLAM tech, mobile robots can now work outdoors. Tracked ones use things like how soft the ground is to avoid obstacles, and adaptive robots with good path planning can handle different places. But there are some problems: their energy use isn't well monitored, sometimes humans still need to help carry out plans, and high-precision models need lots of computing power, which makes them not last long outside. Thus, our project aims at making the robot find the most energy-saving route and follow it.

Keywords: Energy-saving, Robotics, Path-planning, Large Language Model.

1. Introduction

Current robot navigation systems primarily optimize for the shortest path or the fastest travel time. In complex environments, these systems typically rely on environmental modeling and path search algorithms to calculate candidate routes from the starting point to the destination, selecting the one that is short- est in distance or quickest to traverse. While this approach offers advantages in terms of transportation efficiency and task completion speed, its optimization objectives are largely restricted to the spatial and temporal dimensions.

Nevertheless, existing navigation systems do not adequately account for en- ergy consumption. Under varying terrain conditions, modes of movement, or payloads, the shortest or fastest path does not necessarily correspond to the most energy-efficient one. For instance, a steep uphill path may be shorter than a flat detour but consume significantly more energy. Ignoring energy considera- tions often results in low energy efficiency during task execution, which shortens battery life and reduces overall mission success. This reveals a critical limitation of conventional navigation: its reliance on geometric optimality (shortest path) or temporal optimality (fastest path) without consideration of energy optimality (energy-efficient path).

DOI: 10.54254/2755-2721/2026.TJ28246

To address this limitation, we propose the concept of energy-efficient path planning. Unlike traditional approaches, this method prioritizes energy effi- ciency as the primary optimization objective rather than distance or time alone. Within this framework, robots can integrate multimodal information—such as visual perception and environmental feature recognition—to estimate the energy cost of different routes. As a result, the robot can select a path that minimizes energy consumption, emphasizing not only "arriving at the destination" but also "arriving with less energy expenditure."

Several key research questions must be addressed to realize this approach.

First, how should energy consumption be quantified and compared? Energy us- age cannot be represented by a single value, and thus should be transformed into comparable indicators such as battery discharge rate or motor rotational speed. Second, how can energy savings be balanced against efficiency? If the computa- tional cost of intelligent path planning outweighs the benefits of reduced energy consumption, then such planning is not meaningful. Robots, therefore, need the capability to estimate whether performing such calculations is worthwhile. Third, how can the gap between simulation and reality be minimized? Simula- tion environments are typically ideal and controllable, whereas real-world condi- tions involve unpredictable disturbances and system errors, requiring extensive experiments and real-world data validation. Finally, how can deep learning support path planning? By leveraging deep learning, robots can enhance their computational capabilities, enabling faster estimation, greater decision-making flexibility, and improved adaptability in planning routes.

To validate this idea, we designed an experimental framework consisting of three modules. First, a 360.camera periodically captures road images to record features such as road type and width. Second, these images are processed by a GPT-based system, which evaluates road characteristics, assigns scores to can- didate routes, and compares them to suggest the most suitable path. Finally, Docker Desktop is employed as the control platform to execute the robot's move- ment. This framework provides the structural foundation for our experiments, while specific methodological details will be discussed in later sections.

In conclusion, energy-efficient path planning not only addresses the limi- tations of current navigation systems but also significantly improves mission completion rates and environmental adaptability. This study expands the opti- mization target of robot navigation from traditional "geometric optimality" to "energy optimality" and, by integrating multimodal information with intelligent algorithms, lays the groundwork for future energy-aware navigation systems.

2. Related work

Traditional Navigation With the progress of deep learning, robot mobility, and simultaneous localization and mapping (SLAM) technology, mobile robots have moved from laboratories to real-world environments. The physical features of the terrain have the most direct impact on the mobility of tracked robots, such as whether the surface is soft or slippery, which helps the robots avoid both geometric obstacles and non-geometric obstacles like mud holes or loose stones.

Additionally, as mobile robots are applied more widely, their ability to adapt to the environment can be improved through interactions with the world, making it easier for them to be trained to adjust to different conditions [1]. Moreover, robot path planning and decision-making functions have achieved relatively ma- ture results, allowing them to operate effectively in various situations.

However, monitoring robot energy consumption remains a challenge, and sometimes path planning and decision-making processes may require human in- tervention. Furthermore, achieving high precision demands significant resources for training, which can reduce the endurance of robots

DOI: 10.54254/2755-2721/2026.TJ28246

in practical scenarios. Thus, our project aims to develop a robot that can identify the most energy-efficient route and follow it accurately.

AGR Navigation 1. Advancements in Visual-Language Navigation for AGR Research on visual-language navigation has mainly focused on simula- tions. However, these methods often struggle to perform well in real-world settings. A new approach has been introduced, combining vision and language with the knowledge from pre-trained models to create a multimodal and seman- tic bridge between image and text concepts, thereby improving the supervision of key-point navigation and enhancing durability [2]. This enables robots to navigate around obstacles without major accidents. 2. Core System: AGR Nav The AGR Nav system is specifically designed for air-ground robot navigation, integrating a lightweight SCO Net, which uses a self-attention mechanism to precisely identify obstacles by focusing on contextual and hidden information, along with fast map updating and step-by-step path planning to achieve safe and energy-efficient navigation routes. These routes have been tested in both simulations and real-world environments, demonstrating superior performance compared to previous methods [3]. 3. Technical Improvements and HE-Nav Currently, AGR navigation systems have made significant advances in handling mild obstacles, such as in densely built areas, with complex layouts serving as typical testing environments. These systems benefit from the use of 3D Semantic Scene Completion networks, which help enhance robot navigation planning by predicting semantic features and physical geometry of the scene. By predicting the occupied positions in each scene at a voxel level, Euclidean Signed Distance Field (ESDF) maps are constructed to improve collision avoidance paths.

HE-Nav, for example, is considered the first effective solution for AGR nav- igation in highly cluttered spaces. A series of experiments have shown that HE-Nav performs better than other systems in various tasks, including those involving large-scale obstacles and tight turns [4].

BEV Navigation A deep and reliable analysis of maps has been conducted, leading to the proposal of a new navigation method called BEV Nav, which uses deep reinforcement learning to learn a Bird's-Eye View (BEV) representation for reliable decision-making. This method demonstrates strong robustness against crowds, making it the state-of-the-art solution across different challenging tasks and benchmark tests, including in environments with dense human crowds and vehicle traffic. This allows navigation systems to handle a variety of difficult situations, making robot operations in crowded areas more beneficial and safer [5].

DOI: 10.54254/2755-2721/2026.TJ28246

3. Methodology

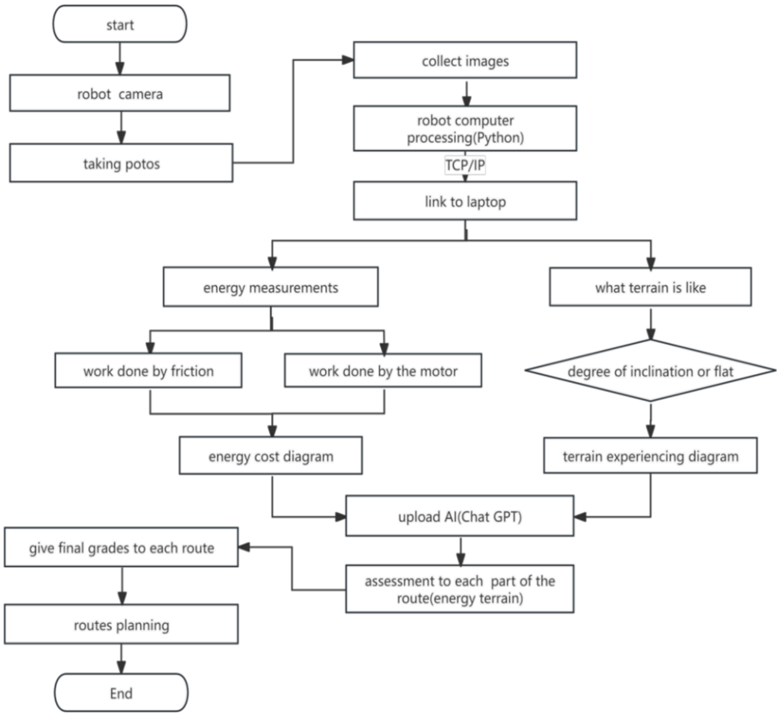


Figure 1. Enter caption

Robot Control We have installed an RGB-D camera directly above the robot to capture synchronized color and depth image data at consistent 10- second intervals. This setup allows us to continuously monitor a variety of road conditions, including weather-related factors such as rain, snow, or fog, as well as surface properties like wetness.

Data collection In addition to environmental conditions, the camera sys- tem records key physical attributes of the road, such as the type of pavement (e.g., asphalt, concrete, gravel), the width of the lanes, and the length of visi- ble segments. All image and depth data are streamed to an onboard computer within the robot for initial processing and temporary storage.

AI Module These collected images are subsequently transmitted to a cen- tral laptop for further analysis. From there, the image data will be uploaded via the GPT API to an algorithmic system designed for detailed scene understand- ing and decision-making support. Once the GPT-4o reads the images from a specific folder, it will gain data in two ways: analyze directly and score different paths, or use semantic segmentation, then score all the areas on the picture. We used the Llava and Deepseek model at first instead of GPT. However, due to their limitation, these two models can't run the prompt as required. Specif- ically, Llava wasn't able to answer the question about energy consumption, while the Deepseek model couldn't receive images. As a result, we turned to try ChatGPT.Using the GPT API, the system will analyze the visual and depth information to compare two or more forked paths. The model will assess which route offers the greatest energy efficiency based on factors such as road incline, surface smoothness, obstacles, and historical energy usage patterns. It will also provide explanatory insights justifying the recommendation.

Comparison Finally, the output from GPT will be structured into a YAML- formatted report. This report will include a comparative energy score for each road option, along with clear reasoning for the selection of the optimal path aimed at minimizing energy consumption.

DOI: 10.54254/2755-2721/2026.TJ28246

4. Experiment results

4.1. Quantitative results

To achieve our goal of minimizing energy consumption in path planning, the Large Language Model (GPT-4o) is used to analyze the terrain and the energy saving levels of each road, and output YAML files that contain the score of the energy saving levels of each investigated fork road.

The approximate scoring for each terrain is present in the table below, noting that a higher score means higher energy efficiency.

Table 1. Approximate score for different terrains

terrains	grass	dirt	wet	stairs	stone	bricks	tile	slopes
scores	60	75	25	10-25	45	85	95	20-70

4.2. Analyze

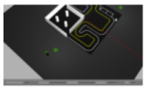
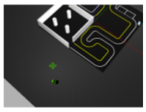
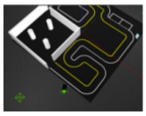
The following prompt is used to obtain scores from the GPT-4o. To address the challenge of handling large numbers of 3D assets, we cluster related assets into groups using the following prompt. Prompt:

You are an assistant who is required to look at a single image from a robot's camera .
The left half of the image is Road 1, and the right half is Road 2 .
Score the energy efficiency level of road 1 and road 2 from 0 to 100, considering the same mobile robot traveling for same distance on them.
Specifically, identify the terrains and road conditions of the roads on each side . Given that, it is assumed that there are no potential
obstacles that are not depicted in the images .
For each road, first give a score from 1 to 50 to indicate the
feasibility of the terrain (0 means the most complex terrain, which has less energy efficiency and is least feasible to travel on) .
Then give a score from 1 to 50 about the road condition (noting that a lower score indicates worse condition!) .
The final score is equal to the sum of scores for terrain feasibility and road condition .
The higher total is better . Analyze visual clues (e.g . , grass, pavement,
wetness, unevenness) to estimate reasonable integer scores . "
Please output a YAML File! .
\YAML file should be only in this exact format:\n final_scores:\n
road_1: <int>\n
road_2: <int>\n
Do not include any extra explanation or markdown .

In the experiment, we have included 5 sets of data to test the feasibility of our approach in 5 fork roads with different terrain. The robot turns right or left in the simulator to indicate which path it chooses.

DOI: 10.54254/2755-2721/2026.TJ28246

Table 2. Robotic path choice analysis across five sets

Sampl e	Images Input	Model Output	Robot Control
1		final_scores: road_1: 55 road_2: 85	 
2		final_scores: road_1: 60 road_2: 80	 
3		final_scores: road_1: 65 road_2: 75	 
4		final_scores: road_1: 80 road_2: 88	 
5		final_scores: road_1: 85 road_2: 78	 

The analysis for the movements ofthe robot in each of the 5 trials is presented below, noting that the road on the left-hand side is road 1, and the road on the right-hand side is road 2.

• Sample 1: The robot in the simulator spins clockwise, choosing the right-hand side path. Via manual judgment, the robot has successfully chosen the optimal path.

• Sample 2: The robot in the simulator spins clockwise, choosing the right- hand side path. Via manual judgment, the robot has successfully chosen the optimal path.

• Sample 3: The robot in the simulator spins clockwise, choosing the right- hand side path. Via manual judgment, the robot has successfully chosen the optimal path.

• Sample 4: The robot in the simulator spins clockwise, choosing the right- hand side path. Via manual judgment, the robot has successfully chosen the optimal path.

• Sample 5: The robot in the simulator spins clockwise, choosing the left- hand side path. Via manual judgment, the robot has successfully chosen the optimal path.

Overall, the robot shows a high accuracy in choosing paths in the simulator.

In all five trials conducted, it successfully chose the optimal path.

DOI: 10.54254/2755-2721/2026.TJ28246

5. Conclusions

In our work, we used a camera to collect photos of paths under different condi- tions, with these paths corresponding to different ground roughness, humidity, and morphologies. When importing the photos into programming software, initial parsing issues arose due to errors like incorrect punctuation. After cor- rections, we drew preliminary conclusions based on the robot's walkability scores and energy consumption data. We compiled the data into a table and found that tile-paved roads are the most energy-efficient, with an accuracy rate of about 89 percent, relatively accurate compared to the baseline.

However, there are areas for improvement: the number of experiments and photos is insufficient; robot latency may cause photos to be blurry and their sequence to be disordered; the manufacturing cost of the robot is high; testing takes too much time; and running too many programs can easily cause the robot and devices to freeze or lag.

Acknowledgement

Luohan Chen, Yutong Hou, Jiaze Hong, Jiawei Zhao and Yichen Sun con- tributed equally to this work and should be considered co-first authors.

References

[1] Peng Xu, Liang Ding, et al. Learning physical characteristics like animals for legged robots. National Science Review, 10(5), 2023.

[2] Hui Yuan, Yan Huang, et al. Multimodal pretrained knowledge for real-world object navigation. Machine Intelligence Research, 22(4): 713–729, 2025.

[3] Junming Wang, Zekai Sun, Xiuxian Guan, Tianxiang Shen, Zongyuan Zhang, Tianyang Duan, Dong Huang, Shixiong Zhao, and Heming Cui. Agrnav: Efficient and energy-saving autonomous navigation for air-ground robots in occlusion-prone environments. arXiv preprint arXiv: 2403.11607v1 , 2024.

[4] Junming Wang, Zekai Sun, Xiuxian Guan, Tianxiang Shen, Dong Huang, Zongyuan Zhang, Tianyang Duan, Fangming Liu, and Heming Cui. He- nav: A high-performance and efficient navigation system for aerial-ground robots in cluttered environments. IEEE Robotics and Automation Letters, 9(11): 10383–10390, 2024.

[5] Jiahao Jiang, Yuxiang Yang, Yingqi Deng, Chenlong Ma, and Jing Zhang. Bevnav: Robot autonomous navigation via spatial-temporal contrastive learning in bird's-eye view. arXiv preprint arXiv: 2409.01646v1, 2024.