Egocentric Vision Module for Adaptive Gait Planning in Lower-Limb Exoskeletons

Miroljub Mihailovic, Matteo Terreran, Stefano Ghidoni, Alberto Pretto, Luca Tonin, Emanuele Menegatti, Stefano Tortora
Department of Information Engineering, University of Padova, Italy
miroljub.mihailovic@phd.unipd.it, {matteo.terreran, stefano.ghidoni, alberto.pretto, luca.tonin, emanuele.menegatti, stefano.tortora}@unipd.it

***Abstract*—This work presents an egocentric vision module for adaptive gait planning in lower-limb exoskeletons. By combining exoskeleton-mounted and head-mounted cameras, the system builds a unified 3D map of the environment, enabling early obstacle detection and robust terrain classification. An Environment-Adaptive Gait Planning (EAGP) system computes safe footpaths, while augmented reality feedback provides intuitive visualization for the user. A dedicated dataset with step and corridor walking scenarios was collected to support development and evaluation.**

***Keywords*—Lower-limb exoskeletons, egocentric vision, adaptive gait planning, environment perception, RGB-D camera, augmented reality**

I. Introduction

Lower-limb exoskeletons (LLEs) have shown strong potential in rehabilitation, assistance, and mobility augmentation. However, despite recent progress, major challenges remain, particularly in collision avoidance, context recognition, and autonomous operation. To address these limitations, robust perception systems are needed to accurately sense the environment ahead of the user and adapt locomotion accordingly.

This work focuses on enhancing the vision module by combining exoskeleton-mounted cameras with an egocentric vision system based on an AR visor. The visor integrates cameras and a display, providing an additional viewpoint from the user's head, which extends the field of view and improves perception of distant or partially occluded obstacles. Fusing these complementary perspectives enables more reliable scene understanding, especially in unstructured environments.

Beyond perception, the AR visor also enables real-time visualization and user feedback. Augmented reality overlays can project critical information, such as safe footholds, predicted trajectories, or obstacle warnings, directly into the user's field of view, thereby enhancing safety and confidence. The same framework can be extended to virtual reality (VR) for immersive training and rehabilitation without physical risk.

II. System Overview

Figure 1 illustrates the proposed control framework, which integrates the Egocentric Vision Module with Environment-Adaptive Gait Planning (EAGP). The primary objective is to perceive the environment in front of the exoskeleton, interpret traversable surfaces and obstacles, and generate collision-free trajectories that enable safe and fluid locomotion.

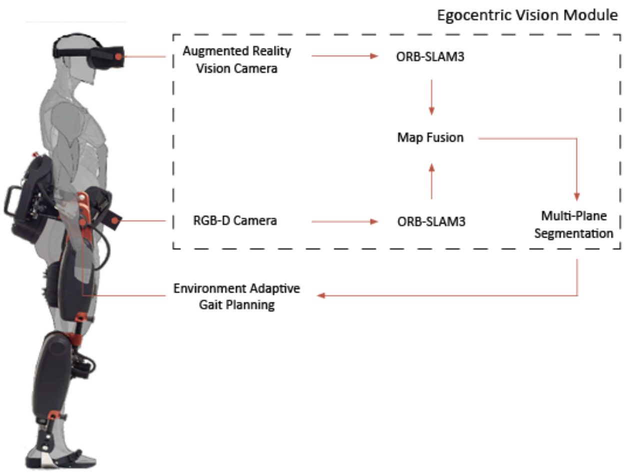

Fig. 1. System overview: the exoskeleton control framework integrates the Egocentric Vision Module for environment perception and Environment Adaptive Gait Planning.

A. Egocentric Vision Module

The robotic vision module aims to extract the geometric characteristics of the walking environment using information provided by vision systems such as RGB-D cameras.

Existing exoskeleton perception systems typically rely on cameras mounted on the device itself to monitor the area immediately surrounding the user's feet. While effective for detecting nearby obstacles, this setup only reveals objects when they are already very close and often fails to capture their full shape or extent — making it difficult to determine whether the obstacle is small and passable or large and blocking the path. In contrast, humans naturally look ahead while walking to anticipate upcoming obstacles and plan safe foot placement. Inspired by this behavior, we combine an exoskeleton-mounted RGB-D camera (*exo-camera*) with a forward-facing AR visor camera (*AR-camera*), enabling the system to perceive both the near-field environment and the wider space in front of the user, as shown in Figure 2. This complementary sensing approach supports earlier obstacle detection and richer scene understanding, ultimately improving navigation safety and fluidity.

The proposed dual-camera perception system relies on

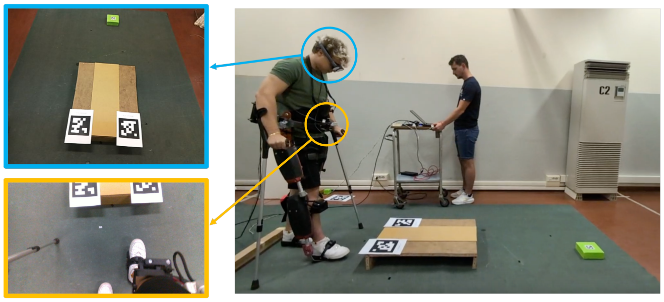

Fig. 2. Proposed perception system: a downward-facing RGB-D camera on the exoskeleton (orange) captures the nearby floor region, while an AR visor camera (blue) extends the observable space in the wearer's forward direction.

images from the *exo-camera* and the *AR-camera* to build a unified, metrically scaled 3D map of the environment. Each camera is processed independently using ORB-SLAM3 [1] with IMU fusion, which provides metric trajectories and sparse 3D landmarks. To continuously estimate the time-varying relative pose between the two sensors, we employ DUSt3R [2] to extract image correspondences across cameras, even under little or no field-of-view overlap. These correspondences are lifted to 3D using the per-camera reconstructions and used to compute the rigid transformation aligning the two maps, with outlier rejection and temporal smoothing to ensure stability. The resulting transformation allows us to fuse observations from both sensors into a single global representation, combining the local floor detail captured by the exoskeleton camera with the wider field-of-view information from the visor, enabling a richer and more complete perception of the user's surroundings.

This rich spatial representation serves as the input to a perception pipeline designed to identify traversable surfaces and obstacles relevant for locomotion. Multi-plane segmentation is performed via *MLESAC* [3], combining normal estimation with likelihood-based plane fitting to detect planes such as ground, ramps, and stairs. At a higher level, scene interpretation classifies environments into flat ground, staircases (parallel planes with constant vertical offset), or ramps (inclined planes). Non-plane points are clustered using Euclidean distance: compact clusters close to traversable planes are labeled as obstacles, while sparse points are discarded as noise. This layered perception pipeline enhances robustness and generalization for locomotion and navigation in unstructured environments.

B. Environment-Adaptive Gait Planning

The Environment-Adaptive Gait Planning (EAGP) framework integrates Egocentric Vision, pose estimation, and adaptive motion generation. The perception module, described in Section II-A, provides traversable surfaces and obstacle information, which are then used by the *foothold selection* module to identify safe step locations. Candidate regions are evaluated according to distance, obstacle clearance, and gait feasibility. Once the foothold is determined, the a collision-free trajectory is computed that avoids obstacles. The trajectory is represented as time-parameterized polynomials in the sagittal plane and refined iteratively to satisfy kinematic constraints and safety margins. Through inverse kinematics, the swing footpath is converted into joint-level commands for execution by the low-level controller. This adaptive pipeline allows the exoskeleton to dynamically adjust step length, height, and timing, ensuring collision-free locomotion even in unstructured environments [4].

III. Dataset

A dedicated dataset was collected to evaluate the proposed perception and planning framework in representative real-world conditions. Two experimental setups were designed:

- **Case 1 – Step with April Tags:** A sequence of walking trials including step-up and step-down over a small platform. April Tags were placed on the ground for ground-truth localization. During each trial, synchronized data were recorded from the exoskeleton sensors (joint angles, proprioceptive signals), the exo-camera, and the AR-camera. Six repetitions were performed to capture variability.
- **Case 2 – Corridor walking:** A 30-meter walk along a corridor, including different conditions such as slope-up and slope-down segments. All traversable terrains were included, and synchronized multi-sensor data were recorded to evaluate the robustness of perception and planning under continuous locomotion.

This dataset enables both quantitative evaluation of trajectory generation and qualitative assessment of AR-based feedback in dynamic scenarios.

IV. Conclusion

We introduced an egocentric vision framework and a dedicated dataset for environment-adaptive gait planning in lower-limb exoskeletons. This work is part of an ongoing development, where perception and planning modules are being integrated and tested in real-time scenarios.

Acknowledgment

This work was developed within the project funded by Next Generation EU - "Age-It - Ageing well in an ageing society" project (PE0000015), National Recovery and Resilience Plan (NRRP) - PE8 - Mission 4, C2, Intervention 1.3".

References

[1] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE transactions on robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[2] S. Wang, V. Leroy, Y. Cabon, B. Chidlovskii, and J. Revaud, "Dust3r: Geometric 3d vision made easy," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 20 697–20 709.

[3] P. Torr and A. Zisserman, "Mlesac: A new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314299908329

[4] E. Trombin, S. Tortora, E. Menegatti, and L. Tonin, "Environment-adaptive gait planning for obstacle avoidance in lower-limb robotic exoskeletons," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024, pp. 13 640–13 647.