A Wearable Computer Vision System With Gimbal Enables Position-, Speed-, and Phase-Independent Terrain Classification for Lower Limb Prostheses

Linrong Li, Xiaoming Wang, Qiaoling Meng, and Hongliu Yu, *Member, IEEE*

***Abstract*— Computer vision can provide upcoming walking environment information for lower limb-assisted robots, thereby enabling more accurate and robust decisions for high-level control. However, current computer vision systems in lower extremity devices are still constrained by the disruptions that occur in the interaction between human, machine, and the environment, which hinder optimal performance. In this paper, we propose a gimbal-based terrain classification system that can be adapted to different lower limb movements, different walking speeds, and gait phases. We use a linear active disturbance rejection controller to realize fast response and anti-disturbance control of the gimbal, which allows computer vision to continuously and stably focus on the desired field of view angle during lower limb motion interaction. We also deployed a lightweight MobileNetV2 model in an embedded vision module for real-time and highly accurate inference performance. By using the proposed terrain classification system, it can provide the ability to classify and predict terrain independent of mounting position (thighs and shanks), gait phase, and walking speed. This also makes our system applicable to subjects with different physical conditions (e.g., non-disabled subjects and individuals with transfemoral amputation) without tuning the parameters, which will contribute to the plug-and-play functionality of terrain classification. Finally, our approach is promising to improve the adaptability of lower limb assisted robots in complex terrain, allowing the wearer to walk more safely.**

***Index Terms*— Computer vision, terrain classification, continuous prediction, phase-independent, gimbal control.**

I. Introduction

LOWER limb robotics such as exoskeletons and prostheses, can improve the mobility level and quality of life of individuals with disabilities [1], [2]. Controlling these assistive robotics requires coordination with the user's motion and intent, failing which can lead to abnormal movements and even falls. However, conventional on-board mechanical sensors have difficulty predicting the future motion of the user, which makes coordinated control of the lower limb robots challenging. One effective solution is to fuse information from the external walking environment to provide more accurate and robust high-level decisions.

The use of proprioceptive sensors such as surface electromyography (sEMG) electrodes [3], [4], [5], [6], [7], [8], capacitive sensors [9], [10], [11], and inertial measurement units (IMUs) [6], [12], [13] can indirectly infer information about the terrain. This is because the human gait will adjust in advance to the upcoming terrain [14]. The fused features extracted from these proprioceptive sensors are typically fed into machine learning algorithms for terrain classification [5] or the estimation of parameters like stair height or ramp incline [6]. In recent years, some researchers have utilized foot trajectory signals estimated by IMU sensors to accomplish terrain classification through heuristic algorithms prior to heel contact [12], [13]. Although these methods have achieved satisfactory results, the predictive performance of proprioceptive sensors for the external environment can only be realized in one step (i.e., less than 1s) [15]. The predictive performance within one step, while sufficient to allow the high-level controller of the prosthesis to switch smoothly between modes, is available only when the prosthetic side is leading [16]. Moreover, since proprioceptive sensors collect kinematic, kinetic, or physiological signals from the user, the performance of the pre-trained model depends on the specific user [17], as well as the location where the sensor is worn [6].

Unlike proprioceptive sensors, exteroceptive sensors are capable of receiving information from the external environment of the body, with powerful predictive performance (usually equivalent to a few steps ahead of time) and user-independent potential. In some previous studies, laser

Manuscript received 14 June 2023; revised 29 August 2023 and 1 October 2023; accepted 4 November 2023. Date of publication 8 November 2023; date of current version 21 November 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62073224 and in part by the National Key Research and Development Program of China under Grant 2018YFB1307303. *(Corresponding author: Hongliu Yu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Committee on Ethics of Medicine of Shanghai Gongli Hospital.

The authors are with the Institute of Rehabilitation Engineering and Technology, University of Shanghai for Science and Technology, Shanghai 200093, China, also with the Shanghai Engineering Research Center of Assistive Devices, Shanghai 200093, China, and also with the Key Laboratory of Neural-Functional Information and Rehabilitation Engineering of the Ministry of Civil Affairs, Shanghai 200093, China (e-mail: yhl98@hotmail.com).

Digital Object Identifier 10.1109/TNSRE.2023.3331273

sensors [5], ranged sensors in arrays [18], and radar sensors [19] have been used to achieve terrain classification and recognition. While these range sensors show a feasible predictive range, they all share the common problem of being limited in capturing a wide range of environmental features. To tackle this issue, certain researchers have employed depth sensors for terrain edge detection [20], [21], [22] or 3D cloud classification [23], [24], [25] since they provide an accurate mapping of the environment and are robust to a variety of textures and lighting conditions. Recently, there have also been researchers who directly classify the RGB images of the captured external environment with promising results [26], [27].

A lateral challenge with exteroceptive sensors is that they move with the body of the wearer. Some researchers installed the camera on the user's waist [28] or chest [23], [29], [30] to improve stability, but this compromised the system's compactness and the user's comfort. Wearing the sensor on the head [26], [31] is a solution that provides a wider field of view; however, it leads to more uncorrelated data and can diminish the predictive performance of nearby terrain (i.e., 1 or 2 steps away). Another common option is to install the sensors on the lower extremity device [18], [25], [26], [27] to enhance the integration of the system. However, movements of the lower limbs (such as heel strike and swing flexion) often cause blurring of the captured images and a narrower field of view. These factors negatively impact the accuracy of recognition and the predictive performance of the environment. Therefore, the vast majority of studies choose to perform recognition during mid-stance, as it offers optimal predictive performance and camera stability. Recently, Zhong et al. proposed an uncertainty-aware frame selection strategy that can dynamically select reliable and critically predicted frames based on lower limb motion and environmental context [27]. The real-time performance of the system is improved without compromising the accuracy of inference. While existing methods can achieve satisfactory results, they are all phase-based (or phase-dependent) and lack continuous, highly accurate predictive ability throughout the gait cycle. This ability is crucial when dealing with complex environments or cases involving terrain transitions. Therefore, it remains an open question how to minimize the negative impact of the human-machine-environment interaction on computer vision systems installed in the lower extremity.

The gimbal can maintain the position of the object unchanged while in motion. This system has been widely used in stabilization systems [32] and tracking systems [33], [34]. Inspired by this technology, we propose a terrain classification system that comprises a wearable gimbal and a compact, low-power machine vision module. Compared to the vision fixed (VF) solutions used in other papers (i.e., where the camera angle is passively changed due to lower limb movements), we utilize the maneuverability of the gimbal to create a terrain classification system with vision tracking (VT) capability. This simple yet effective approach can mitigate the disturbance of the human-machine-environment on the recognition system, thereby enhancing the continuous prediction and classification performance during locomotion (see Fig. 1). The major contributions of this article are as follows.

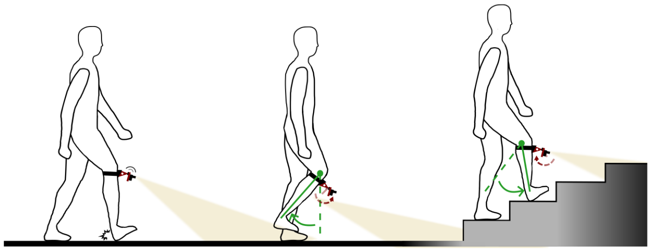


Fig. 1. Continuous terrain classification based on the gimbal during walking. The wearable gimbal system reduces external disturbances for computer vision and enables terrain classification independent of gait phase and cadence by detecting the user's movements.

1) To ensure stable capture performance for the camera worn on the lower limb, we implemented an extended state observer-based active disturbance rejection controller in the control system of the wearable gimbal. This controller allows for rapid response to lower limb motion and external disturbances. By employing this approach, it is possible to achieve continuous terrain classification performance that is independent of sensor position, gait speed, and gait phase.
2) In terms of terrain classification algorithms, we evaluate the real-time performance of several advanced lightweight CNN models on a low-cost computer vision embedded device, providing reference implications for pursuing plug-and-play machine vision applications.
3) We compared the performance of the terrain classification system at different sensor mounting positions (thigh and shank), different speeds, and gait phases, evaluating the impact of both VT and VF solutions on the prediction and classification performance of the terrain classification system.

Our proposed method does not require specific parameter tuning for individuals with varying physical conditions, which can enhance the plug-and-play capability of wearable terrain classification systems. The proposed terrain classification system also provides more accurate, robust, and seamless mode switching capabilities for lower limb assistive devices. This is crucial in enabling disabled patients to walk safely on various terrains, including complex terrain situations.

II. Materials and Methods

A. Embedded System Design

Fig. 2 shows the main components of the terrain classification system: the computer vision system and the gimbal system. The computer vision system uses a compact, low-power machine vision module (OpenMV4 H7 Plus) equipped with a high-performance processor (STM32H743, 480 MHz) and a 5-megapixel sensor (OV5640). A WiFi module connects to the OpenMV CAM through an expansion interface, allowing for wireless transmission of video streaming data. We use a 3.7V LIPO battery to power the OpenMV CAM separately. TThe entire computer vision system is mounted on the gimbal's head, which is made using 3D printing. This gimbal can be rotated in the sagittal and coronal planes within a range of 0-180°. In order to minimize current interference

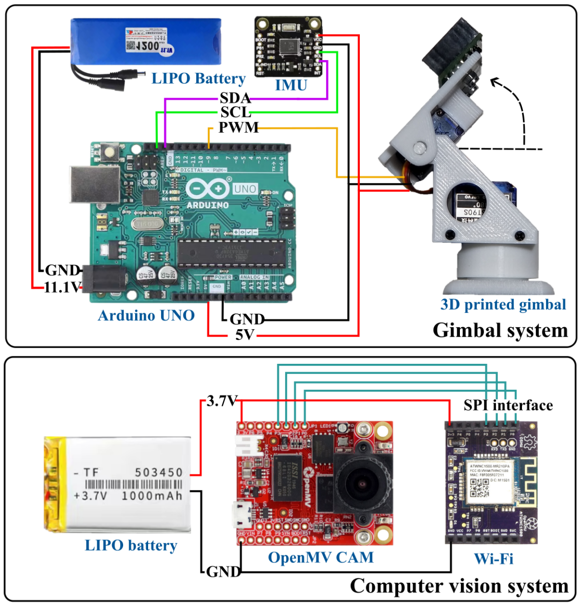
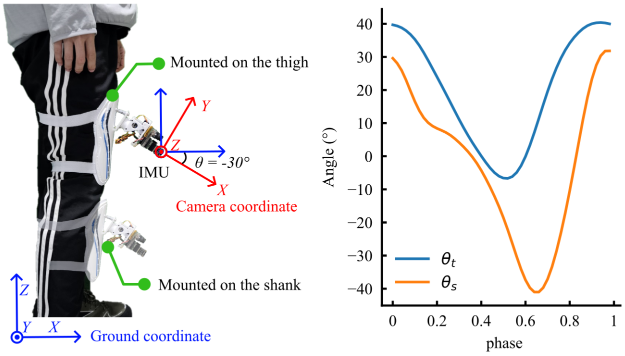


Fig. 2. Embedded hardware component of the terrain classification system.



(a) (b)

Fig. 3. (a) Sagittal-plane view of the human leg. The terrain classification system can be worn on the subject's thigh and shank, where the camera having a target tracking angle of -30°. (b) The shank angle (θ_s) and thigh angle (θ_t) are defined from the global vertical while walking on level ground.

and optimize the performance of hardware resources, we use an additional control board (Arduino UNO, 16 MHz) to execute the control of the gimbal. The gimbal's motor utilizes a user-friendly servo motor that is driven by pulse width modulation (PWM) signals. We also installed an IMU sensor (BNO055) on the gimbal to measure the rotation angle of the gimbal in the sagittal plane. The overall size of the terrain classification system (including the battery and circuit board) is 102 mm × 65 mm × 40 mm and it weights 363.6 grams. By utilizing a wearable leg guard, the system can be comfortably and conveniently attached to the thigh or shank of non-disabled subjects (see Fig. 3 (a)).

B. Control Approach

1) Control Target Setting: In this article, the aim of controlling the gimbal is to minimize the impact of lower limb motion on the performance of the terrain classification system. To accomplish this task, implementing the target tracking capability of the gimbal is an option. We rotate the camera 30° counterclockwise around the Z-axis of the IMU (at this point, the angle between the camera and the horizontal plane is $-30°$), and this position provides both near and distant terrain information well when the subject is standing (Fig. 3(a)). Since the camera's view is influenced by the movement of the lower limbs (Fig. 3(b)), we set the tracking target for the low-level controller at -30°. This allows the computer vision system to maintain the same predictive performance as in the stance state while the subjects are walking.

2) Design of Controller: In order to achieve robust VT of lower limb movements for different users, the control system must have high response sensitivity and anti-disturbance performance. Active disturbance rejection control (ADRC) utilizes an extended state observer (ESO) to achieve real-time tracking and compensation of disturbances, which can handle disturbance that cannot be accurately modeled and achieve fast response [35]. This controller has been used in the control of gimbal and has achieved improved results compared to traditional PID controllers [33]. In this paper, we design a linear ADRC (LADRC) controller and implement it in the motion interaction control between the gimbal and the lower limb. Compared to ADRC, LADRC reduces the parameter configuration, making it more suitable for engineering applications [36].

The general framework of the control system is shown in Fig. 4. The LADRC mainly consists of a nonlinear error feedback controller (NLEFC) and a linear ESO (LESO). During operation, the NLEFC non-linearly adjusts the control gain, LESO estimates the disturbance in real time, then compensates for the observed disturbance with feedforward. The output of LADRC is eventually used as an input to PI control to achieve closed-loop regulation of the motor speed. Next, we will mainly introduce the design of NLEFC and LESO.

3) Linear Extended State Observer: In this study, LESO is used to estimate the disturbances and uncertainties that arise in the wearable gimbal system during walking, and feedforward compensation is given to counteract the effects of these disturbances on the follower system. First, the controlled object is reduced to the following second-order system:

$$\ddot{y} + a_1\dot{y} + a_2 y = \omega + bu, \tag{1}$$

where y is the feedback position input, a_1 and a_2 are the system parameters, ω is the external disturbance, b is the control gain, and u is the output of the position controller. For (1), it can be rewritten in the following form:

$$\begin{aligned}\ddot{y} &= -a_1\dot{y} - a_2 y + \omega + (b - b_0)u + b_0 u\\ &= f(y, \dot{y}, \omega) + b_0 u,\end{aligned} \tag{2}$$

where $f(y, \dot{y}, \omega)$ contains the total external and internal disturbances, the extended state variables of the system are $x_1 = y$, $x_2 = \dot{y}$, $x_3 = f(y, \dot{y}, \omega)$, and the state equation of

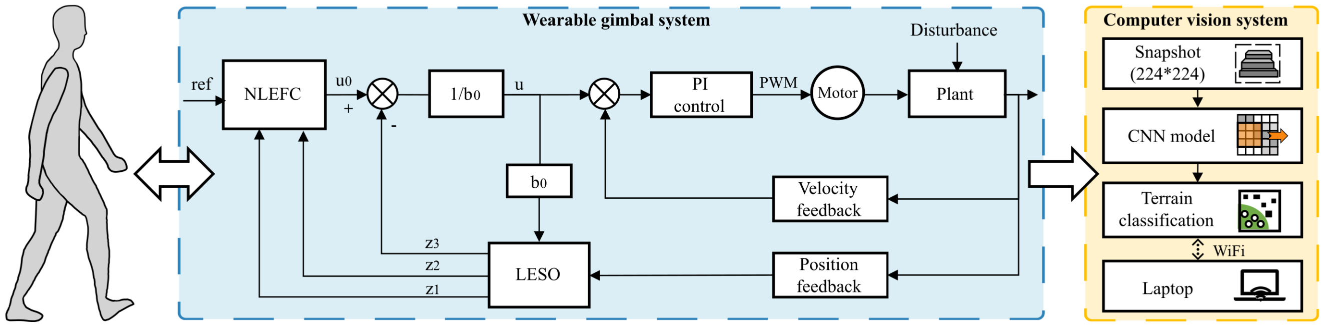


Fig. 4. The control architecture of wearable gimbal system. The LADRC controller uses LESO to observe disturbances in the system and the NLEFC to achieve a nonlinear combination of errors. The output of the LADRC is fed into the PI controller to achieve speed control of the gimbal.

the system is shown as follows:

$$\begin{cases} x_1 = & y \\ \dot{x_1} = & x_2 \\ \dot{x_2} = & x_3 + b_0 u \\ \dot{x_3} = & \dot{f}(y, \dot{y}, \omega) \end{cases} \tag{3}$$

Then, LESO can be established in the following way:

$$\begin{cases} \varepsilon_1 = & z_1 - y \\ \dot{z_1} = & z_2 - \beta_1 \varepsilon_1 \\ \dot{z_2} = & z_3 - \beta_1 \varepsilon_1 + b_0 u \\ \dot{z_3} = & -\beta_3 \varepsilon_1, \end{cases} \tag{4}$$

where ε is the error signal of the state observer, z_1 is the position observation signal, z_2 is the observed angle difference signal, and z_3 is the estimation of the disturbance, β_1, β_2, and β_3 are the gains of the LESO, which are the parameters to be set.

By appropriately selecting parameters β_1, β_2, and β_3, it is possible to achieve the tracking of each state variable of the system by the observer. Configuring the three poles of the observer to the -ω_0 of the left half of the real axis of the s-plane enables the adjustment of gain parameters by selecting the observer's bandwidth. This simplifies the parameter design of LESO. According to the stability requirements of NLESO, the ideal parameter tuning equations are $\beta_1 = 3\omega_0$, $\beta_2 = 3\omega_0^2/5$ and $\beta_3 = \omega_0^3/10$. However, if the value of β_3 is too large, it can easily lead to oscillation in the controller output. Therefore, we appropriately lower the value of β_3, and finally set $\beta_1 = 3\omega_0$, $\beta_2 = 3\omega_0^2/5$, and $\beta_3 = \omega_0^3/20$. Here, we set the value of ω_0 to 16.

4) Nonlinear Error Feedback Controller: Traditional PID control sums each error linearly weighted, which can often result in system overshoot or instability when the error is significant. This means that if the subject's gait speed is too fast, it can negatively impact the gimbal system's target tracking. NLEFC changes the error combination pattern and adaptively adjusts the PD control gain based on the magnitude of the error. This enhancement improves the accuracy and response speed of the control system. As shown in Fig. 4, NLEFC combines the errors between the input values and the z_1 and z_2 estimated by LESO nonlinearly to calculate the control parameter u_0. The nonlinear feedback is combined in the following form:

$$e_1 = s_1(t) - z_1(t), \tag{5}$$

$$e_2 = s_2(t) - z_2(t), \tag{6}$$

$$u_0 = \kappa_2 fal(e_1, \alpha_1, \delta) + \kappa_2 fal(e_2, \alpha_2, \delta), \tag{7}$$

where e_1 is the position error, e_2 is the differential signal of the position error, u_0 is the output of the NLEFC, k_1 and k_2 as the control gain can be equated to the values of k_p and k_d in the PD controller, and $fal()$ is the nonlinear parameter adjustment function.

The expression of the $fal()$ function in (7) is as follows:

$$fal(\varepsilon, \alpha, \delta) = \begin{cases} |\varepsilon|^a sgn(\varepsilon), |\varepsilon| > \delta \\ \varepsilon/\delta^{1-a}, |\varepsilon| \le \delta \end{cases} \quad \delta > 0, \tag{8}$$

where α is a nonlinear factor, the smaller α is, the greater the output gain of the function. δ is a filtering factor, which affects the limit value of the control gain. The smaller δ is, the greater the limit value of the function output when the error is small. NLEFC adaptively adjusts the PD control gain based on this error combination approach to achieve large error but small gain control and small error but large gain control.

After the nonlinear combination, feedforward compensation for disturbances is performed:

$$u(t) = (u_0 - z_3(t))/b_0, \tag{9}$$

where $z_3(t)/b_0$ is the feedforward compensation for model errors and disturbances, and the level of b_0 reflects the compensation strength. Then, a linear combination of feedforward compensation $u(t)$ and the angular velocity from the gyroscope is fed to the PI controller to achieve closed-loop control of the motor's speed. After parameter tuning, we chose the following parameters for the PI controller: $k_p = 1.5$, $k_i = 0.5$.

C. Construction of Training Datasets

Compared to depth cameras, RGB cameras are less expensive and easier to use; however, they lack sensitivity to geometric information about the terrain. Therefore, to use RGB cameras for terrain classification, the dataset must be large enough to train a sufficiently robust model. In this paper, we use ExoNet - one of the largest and most diverse open-source datasets of wearable camera walking environment images available - to facilitate the development of CNN-based terrain classification systems [30]. ExoNet has over 922, 000 images that contain a variety of everyday indoor and outdoor walking environments. To conduct a more targeted study, we reclassified and refined these images through manual annotation and extraction, and the final dataset has three main categories: level ground (LG), stair ascent (SA), and stair descent (SD). The dataset has a total of about 197 000 RGB images, including 77 000 of LG, 75 000 of SA, and 45 000 of SD.

D. Classification Algorithms

After preparing the dataset, we utilize convolution-based lightweight deep learning algorithms to train the terrain classifier. Some of the models in Keras' functional API, such as MobileNetV1, MobileNetV2, EfficientNet B0, NASNetMobile, and DenseNet121, have shown high classification accuracy and low computational cost on the ImageNet database [30], [37]. This characteristic allows these models to be applied on devices with limited computing resources, such as mobile or embedded devices. Therefore, we consider these five lightweight deep learning models as candidates. Our goal is to select the best model that can be deployed on-board devices for real-time recognition based on the offline evaluation results.

III. EXPERIMENTS

A. Offline Experiment

1) Model Training: We divided the dataset presented in Section II-C into three parts: the training set (80%), the validation set (10%) and the test set (10%). Additionally, the images were resized to 224 × 224 size as input for the models. We trained the five lightweight CNN models mentioned in Section II-D on Google Cloud using the Keras functional API. We fixed the batch size, epochs, dropout rates, and initial learning rates at 32, 20, 0.1, and 0.0001, respectively, based on the results of multiple training runs. To reduce the model's parameters and prevent overfitting, we employ a single fully connected layer (8 neurons, 0.1 dropout, ReLu) connected to the output layer (3 neurons, Softmax). All models utilize transfer learning to expedite the training process.

After completing the training, we conducted merit-seeking work on these models based on multi-scale metrics. Similar to the previous study [30], we use multiply-accumulates (MAC) expressed in billions (B), inference time (IT) and accuracy (%) as evaluation metrics to assess the relationship between model complexity and classification performance. The inference time is estimated based on the OpenMV4 H7 Plus (STM32H743, 480 MHz). We also consider Flash and RAM, two key metrics for deep learning models in embedded deployments. This enables us to compare model sizes in a more intuitive manner and evaluate the memory usage of these models in resource-constrained embedded devices. Before deployment to the device, all models were optimized using the Edge Impulse EON^{TM} compiler to improve on-device performance.

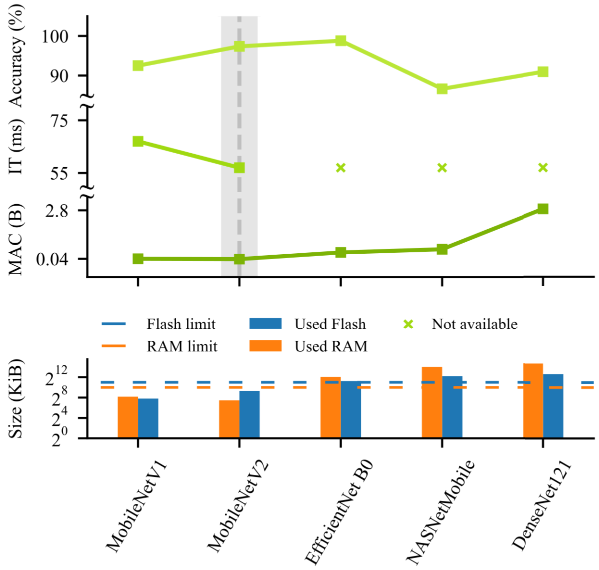


Fig. 5. Comparison of offline evaluation results of five lightweight CNN models. We evaluate the CNN models based on four metrics: embedded memory usage size (size), multiply-accumulates (MAC), inference time (IT), and accuracy. The memory usage of EfficientNet B0, NASNetMobile and DensNet121 exceeds the capacity of the OpenMV CAM, so their IT are not available. We chose MobileNetV2 as the model for real-time classification since it offer higher accuracy and lower model complexity compared to MobileNetV1.

2) Offline Results: Fig. 5 shows a comparison of the results for offline validation. For classification performance, EfficientNet B0 (98.8%) had the highest accuracy, outperforming MobileNetV2 (97.37%), MobileNetV1 (89.52%), NASNetMobile (86.97%), and DenseNet121 (90.91%). However, except for MobileNetV1 and MobileNetV2, the model sizes of EfficientNet B0, NASNetMobile, and DenseNet121 all exceed the performance capabilities of the controller chip (Flash: 2048 KiB, RAM: 1024 KiB), so the IT of these three models is non-testable. When comparing MobileNetV2 with MobileNetV1, the former has a lower MAC (0.0231 B compared to 0.0407 B) while still having a higher accuracy and lower inference time (67 ms compared to 57 ms). Based on the comparison of the offline experimental results mentioned above, we have chosen MobileNetV2 for conducting real-time classification experiments.

B. Real-Time Experiment

1) Experiment Setup: We invited five non-disabled subjects (ABs) and two transfemoral amputees (TFAs) to participate in real-time experiments in two scenarios: indoor and outdoor. The basic information of the subjects is shown in Table I. The study protocol was approved by the Committee on Ethics of Medicine of Shanghai Gongli Hospital, and each participant provided written informed consent before the experiment.

The experimental procedures for indoor and outdoor settings are shown in Table II. The indoor experimental environment is primarily utilized to validate the predictive performance of our method during continuous terrain transitions. As shown in Fig. 6, a stairway terrain is placed in front of the treadmill to simulate the situation of terrain transition during walking. In this scenario, the distance between the subject and the stairs

TABLE I
RELEVANT INFORMATION OF THE SUBJECTS

Subjects	AB1	AB2	AB3	AB4	AB5	TF1	TF2
Height (m)	1.80	1.91	1.75	1.60	1.82	1.68	1.71
Weight (kg)	70	80	60	55	72	70	68
Age (years)	25	23	26	26	23	32	29
Gender	Male	Male	Male	Female	Male	Male	Male
Amputation side	-	-	-	-	-	Right	Left
Amputation time	-	-	-	-	-	2013	2016
Prosthetic knee	-	-	-	-	-	Total knee	3R60

TABLE II
EXPERIMENTAL PROTOCOL SETTING

Subeject	Environment	Speed (km/h)	Position	Gimbal
5 ABs, 2 TFAs	Indoor (treadmill)	1.2	Shank	✓
				×
			Thigh	✓
				×
		1.6	Shank	✓
				×
			Thigh	✓
				×
		2.0	Shank	✓
				×
			Thigh	✓
				×
5 ABs	Outdoor	Self-selected	Shank	✓
				×
			Thigh	✓
				×

✓ The gimbal is activated (VT)
× The gimbal does not work (VF)

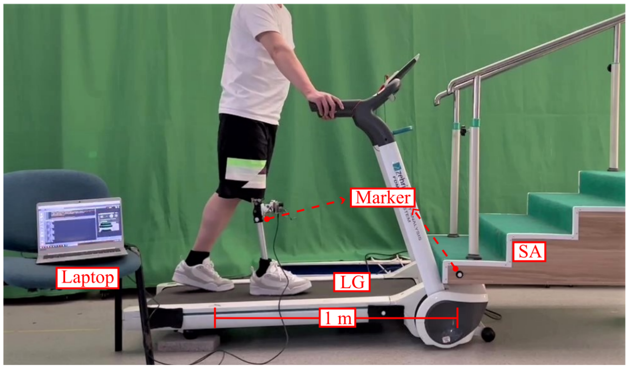


Fig. 6. Indoor experiment environment. A stairway terrain is overlaid with the front of the treadmill to simulate the continuous terrain transition scenario (LG⟶SA). An amputee wearing a passive prosthesis (Total Knee 2000, Össur, Iceland) walked on this terrain with the VF state.

while walking on the treadmill varied approximately between 0.5 to 1 meter, and the subject was considered to be in a terrain transition situation with each step. Each subject was asked to walk on the treadmill at 1.2 km/h, 1.6 km/h, and 2.0 km/h, with no fewer than 60 times of terrain transitions (i.e., no fewer than 60 steps) in each case. Two comparison experiments are included for walking at each speed, i.e. the comparison of the effect of sensors mounted position, and the comparison of performance of LADRC-based VT and VF.

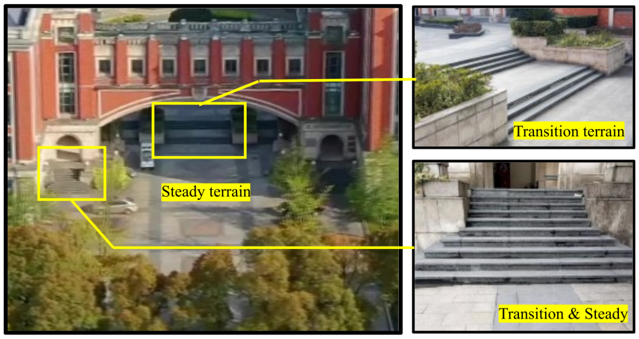


Fig. 7. Outdoor experiment environment.

During walking, our terrain classification system is connected to the laptop via a USB cable for video data acquisition. We also attached markers on the stairs and on the subject's lower limb to calculate the horizontal distance s_d between the two in the sagittal plane by myoVIDEO (Noraxon, USA). The prediction time T_{pred} can be obtained using the following equation:

$$T_{pred} = s_d^{max}/v, \tag{10}$$

where s_d^{max} is the maximum value of s_d for consecutive correctly detected terrain transitions within one gait cycle, and v denotes the walking speed of the subject. We also recorded transitional accuracy, which were derived by dividing the number of accurate predictions by the total number of predictions. In this scenario, the system's prediction of the terrain transition is considered accurate if the result is SA. Note that subjects only walk on the treadmill without actual terrain transitions (i.e., LG⟶SA) in indoor experiment, but the terrain classification system continuously detects the transition terrain.

The outdoor experimental environment is primarily utilized to validate the real-time classification performance of our system. The walking route for the outdoor experiment is shown in Fig. 7, which includes three terrains: LG, SA, and SD. There are a total of 12 terrain transition times (three times each for LG⟶SA, SA⟶LG, LG⟶SD, and SD⟶LG). Each subject was asked to complete the walking task five times. Similar to the procedure in the indoor experiment, each non-disabled subject was asked to wear the terrain classification system on their shank and thigh while walking outdoors at a self-selected pace (see Table II). We also installed an additional IMU on the subject's lower limb to capture the kinematic signal. During the experiment, the terrain classi-

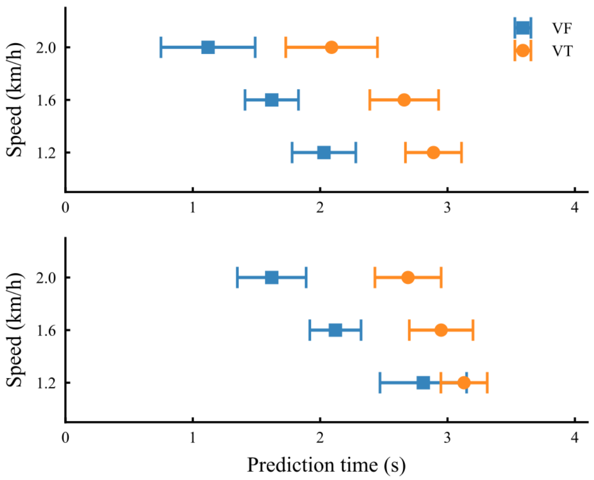


Fig. 8. Comparison of prediction time in the case of terrain transition. From top to bottom is a comparison of the prediction time when the camera is mounted on the shank and thigh for continuous walking in the transition state. Error bars indicate ±1 standard error of the mean across relevant trials.

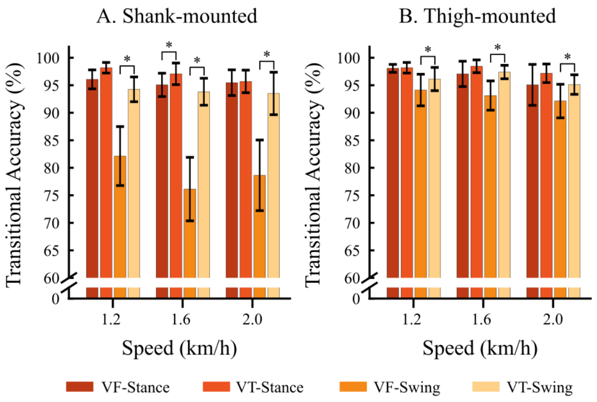


Fig. 9. Comparison of transition accuracy of indoor experiment. A and B indicate the average transitional accuracy across all participants (n=7) at different gait phases and walking speeds when the camera is mounted on the shank and thigh, respectively. VF-Stance represents the classification performance of the terrain classification system in the VF state during the stance phase, and the others are similar. The square bracket shows the result of a paired two-sided t-test (degrees of freedom (df)=6, $*p < 0.05$). Bars indicate ±1 standard error of the mean over relevant trials.

fication system saves the real-time classification results and kinematic data to the memory card. By comparing the video frames, we further extracted the classification accuracy of different sensor mounting locations during terrain transitions and steady states. The steady state accuracy corresponds the correct classification when the current state and the next state are identical, whereas the transitional accuracy accounts for non-identical cases. The overall accuracy combines the steady state and transition cases.

2) Indoor Experiment Results: The prediction time of the terrain classification system varied at different walking speeds (see Fig. 8). When walking at speeds of 1.2 km/h, 1.6 km/h, and 2.0 km/h, the average prediction time of VT-Shank was 2.54 ± 0.28 s, which was better than the 1.53 ± 0.27 s of the VF-Shank used for comparison. As expected, the proposed method has a positive impact on the prediction time of the terrain classification system. This improvement also applies when the sensor is mounted on the thigh, and its prediction time in the case of VT is 2.92 ± 0.23 s, which is better than that of VF with 2.18 ± 0.27 s.

The level of prediction performance directly leads to differences in the accuracy of the terrain classification system for transition situations (see Fig. 9). We conducted a paired t-test to comparison tests of transitional accuracy in the VF and VT states (including stance and swing phases, respectively) to determine statistical differences between the pairs of interest ($p = 0.05$). The interaction between all pairs was significant during the swing phase of the gait. In the stance phase, a significant difference was observed only when the device was mounted on the shank and at a speed of 1.6 km/h ($p = 0.02$). These results demonstrate the enhanced accuracy of our method for terrain recognition systems across various gait phases.

We compared the variation in camera angle between the VF state and the LADRC-based VT state during continuous walking (see Fig. 10(a-f)). We also calculated the root mean square error (RMSE) of the camera angle in the VF and VT states to evaluate the stability performance of the proposed method (see Fig. 10 (g)). The RMSE of VT-Shank was 2.52, 3.16, and 4.53 at 1.2 km/h, 1.6 km/h and 2.0 km/h, respectively, much lower than the 19.59, 20.54, and 22.44 in the VF-Shank state. On the contrary, the RMSE of VT-Thigh is slightly better than VT-Shank with 1.73, 2.68, and 2.92, respectively.

3) Outdoor Experiment Results: The results of the outdoor experiments are shown in Fig. 11. The overall accuracy of the terrain classification system using VT when the camera is mounted on the shank can reach 96.58 ± 1.76% (transition: 95.13 ± 2.23%, steady: 97.44 ± 1.47%), outperforming the 90.07 ± 3.62% in the case of VF (transition: 84.33 ± 5.50%, steady: 93.72 ± 2.71%) (see Fig. 11(A)); When mounted on the thigh, the system achieves an overall accuracy of 98.03 ± 0.91% using VF (transition: 97.27 ± 1.06%, steady: 98.29 ± 0.95%), again better than the 96.18 ± 1.70% in the VF state (transition: 92.22 ± 3.18%, steady: 97.53 ± 1.12%).

IV. Discussion

A. Performance of Offline Classification

In the offline experiments, we compared five lightweight CNN models, and the results showed that MobileNetV2 is potentially more suitable for our current real-time experiments. EfficientNet B0 achieved the highest classification accuracy (98.8%), but like NASNetMobile and DenseNet121, it was excluded as a candidate due to its large size, which would hinder embedded deployment. Although the model size of MobileNetV2 is larger than that of MobileNetV1, the former has a lower inference time (67 ms), lower MAC (0.0231 B), and higher accuracy (97.37%). These results show that MobileNetV2 provides a better balance between classification accuracy, IT, and model complexity, facilitating plug-and-play effects on devices. After completing the embedded deployment of the model, we also measured the program's execution time and the video's frame rate, which were 67.02 ± 1.36 ms and 14.91 ± 0.15 FPS, respectively. Simon et al. showed that users were not affected by the 90 ms latency of the prosthetic mode transition [38], which indicates that the

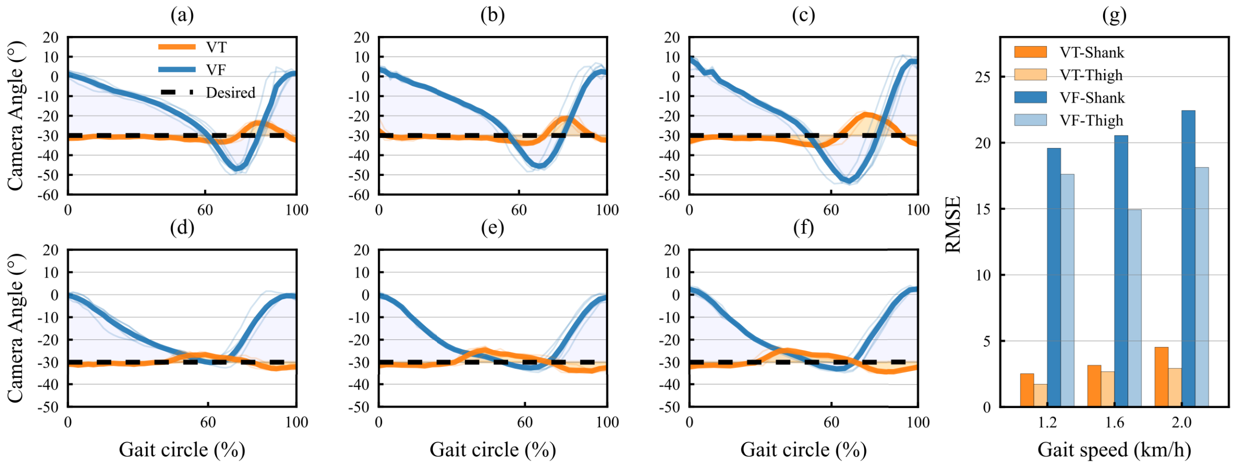


Fig. 10. Comparison of camera angle variation in VF and VT states during walking. (a)-(c) Comparison of camera (shank-mounted) angle at 1.2 km/h, 1.6 km/h, and 2.0 km/h. (d)-(f) Comparison of camera (thigh-mounted) angle at 1.2 km/h, 1.6 km/h, and 2.0 km/h. (g) Comparison of the RMSE between the camera angle and the ideal VT angle. VT-Shank indicates the tracking performance of the VT when the terrain classification system is mounted on the shank, with lower values of RMSE indicating less camera oscillation.

real-time performance of our classification system is sufficient for everyday use.

B. Performance of Indoor Experiment

1) Effect of Walking Speed: Since the T_{pred} is inversely proportional to the speed and the s_d is fixed within a certain range in this experiment, the T_{pred} of the terrain classification system decreases as the walking speed increases in both the VT and VF cases (see Fig. 8). The walking speed also affects the quality of the camera images, which in turn affects the classification accuracy of the terrain classification system. For example, the transition accuracy of VF-Swing at 1.2 km/h is 82.13%, while at 2.0 km/h the accuracy drops by 3.5% (see Fig. 9(A)). As a comparison, it is shown that by using our method, the VT-Swing achieves more than 93.51% transition accuracy at all three walking speeds. Such improvement in transition accuracy allows the terrain classification system to achieve a T_{pred} of no less than 2.09 ± 0.36 seconds at 2.0 km/h, which can leave enough time for high-level control of the wearable robot. In addition, the average T_{pred} for VT was 2.74 seconds, which was a 44.9% improvement over VF (see Fig. 8). All of these results demonstrate that the proposed method effectively mitigates the impact of speed on terrain classification and prediction performance.

2) Effect of Gait Phase: We note that the average transitional accuracy of the stance phase is 96.79%, which is higher than that of the swing phase in both the VF and VT cases (see Fig. 9). This is because the motion of the stance phase is always smaller than that of the swing phase, regardless of the walking speed. This provides a more stable capture performance for the camera. We also observed a significant improvement in the transition accuracy of the terrain classification system during the swing phase compared to the VF-Swing when using our method. This is primarily because the gimbal system offers a superior field of view to the camera during the swing phase, which is not possible in the VF solution (see Fig. 12). When the camera is mounted on the shank, the average improvement for VT-Swing is 14.9% and 3.42% for the thigh, and the average transition accuracy reaches 93.86% and 96.22% for the two mounting positions, which is close to the classification performance of the standing phase (see Fig. 9). These results show that our method can achieve terrain classification that is independent of gait.

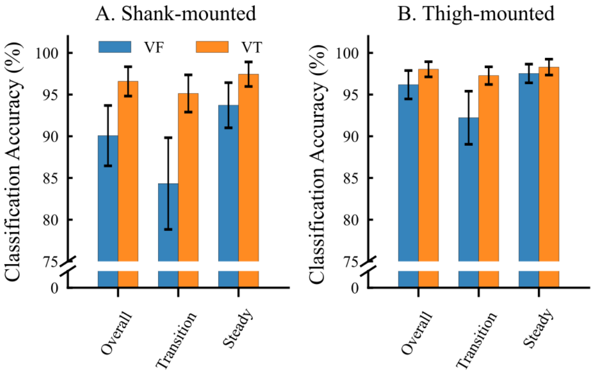


Fig. 11. Outdoor experimental results. A and B represent the outdoor classification accuracy of the terrain classification system mounted on the shank and thigh, respectively. Error bars indicate ±1 standard error of the mean across relevant trials.

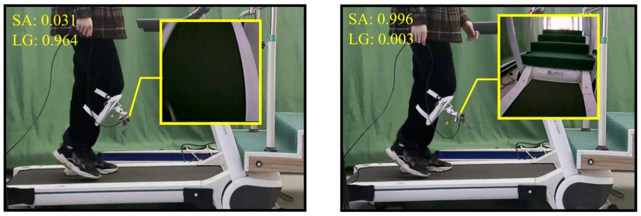


Fig. 12. Comparison of prediction performance of terrain classification systems with VF (left) and VT (right) states. Our method enables the terrain classification system to obtain better vision and higher transition accuracy (SA: 99.6%) in swing flexion phase.

3) Effect of Camera Locations: Since the ideal tracking angle of the gimbal in this paper is not parallel to the ground, mounting the camera higher will result in a wider predictive range. As expected, we found that the T_{pred} of the terrain classification system mounted on the thigh was better than on the shank. However, this gap is reduced when using our method (Fig. 8). This result suggests that our method can potentially improve the variance in predicted performance due to mounting location. Another advantage of thigh-mounted is that the movement of the thigh is smaller compared to the shank, which allows the camera to be more stable during the swing phase. As shown in Fig. 9, the average transition accuracy in the VT-Swing case was 93.86% (shank-mounted) and 96.22% (thigh-mounted), respectively. The above results demonstrate that our method can achieve position-independent terrain classification.

4) Performance of LADRC-Based VT: Unlike the VF used in other papers, this paper uses LADRC-based VT control, which allows the view of the camera to be adjusted autonomously. We compared the angle changes of these two solutions during walking as a way to evaluate the predictive stability performance of the camera. As shown in Fig. 10(a-f), the tracking errors in VT mainly arise during the swing phase, while stable visual tracking can be achieved in the stance phase (including heel strike, mid-stance, and toe-off). This can be used to explain the results of Fig. 9, where the stance phase has a higher accuracy than the swing phase. The LADRC-based VT-Shank control has a maximum RMSE value of 4.53 during walking, which is much lower than the 21.16 of the VF-Shank (see Fig. 10(g)). The results show that the LADRC-based VT control algorithm is able to keep the camera's field of view continuously focused on the terrain ahead to minimize the impact of gait phase and walking speed on the performance of the terrain classification system.

C. Performance of Outdoor Classification

In outdoor environment, factors such as walking uncertainty and collisions between people and the environment are more likely to affect the image quality of the camera. By using our method, the overall accuracy of outdoor terrain classification was 96.58 ± 1.76% (shank-mounted) and 98.03 ± 0.91% (thigh-mounted), an improvement of 6.51% and 1.85% over the VF scheme, respectively (see Fig. 11). This shows that the proposed terrain classification system is suitable for unfamiliar outdoor environments. It also demonstrates that the LADRC-based VT control provides resistance to disturbance. In the case of terrain transition, our method is able to achieve a classification accuracy of no less than 95.13% (shank-mounted), which is comparable to the accuracy achieved indoors (see Fig. 11). This is because subjects walking outdoors typically adjust their walking pattern to the stance phase as the initial state to negotiate transitional terrain, which usually facilitates correct classification. Future work could fuse visual and neural signals (e.g., EMG) to further improve the accuracy of real-time classification. To realize this work, selecting the optimal mounting position of the gimbal based on the position of the electrodes is a necessary endeavor, as this can effectively reduce the interference of artifacts.

D. Impacts on Users

The proposed terrain classification system has an overall size and mass of 102 mm × 65 mm × 40 mm and 363.6 grams respectively. These dimensions are much smaller than those of the adult leg (the 50th percentile of calf circumference for adult males and females is 392 mm and 375 mm, respectively [39]). Although some non-disabled subjects indicated that the wearing side felt weight-bearing compared to the opposite leg, no obvious effect of weight on lower limb kinematic data was found during the experiment. Moreover, the non-disabled subjects generally perceived a change in the center of mass of the device during walking, which was opposite to the trend of leg movements (Fig. 1). Reducing the weight of the load end of the gimbal (i.e., the computer vision module) could effectively alleviate this sensation. On the contrary, one amputee subject reported that he did not feel the relative motion of the gimbal when the device was mounted on the prosthetic shank. Instead, he felt that the prosthesis and the gimbal system were integrated. Therefore, integrating the gimbal system with the prosthesis may improve the overall comfort of the user's gait. The size of the system compared to its weight may have a more direct impact on the wearer's lower extremity movement, or even their hand movement. For example, some subjects expressed concern about devices mounted on the proximal thigh (or socket) due to the risk of the device being obscured by clothing or the collision of the device with the hand. Mounting the device on the distal end of the thigh (or socket) is a solution that not only greatly reduces the risks described above but also reduces the cognitive burden on the user, even though it will sacrifice a certain amount of predictive performance.

E. Limitations

AA limitation of the proposed terrain classification system is that the size and weight need further optimization to enhance the user experience. For size, being too large is not conducive to the operation of the terrain classification system in crowded environments and can increase the mental burden on the user. Optimizing the size of the gimbal (60 mm × 65 mm) is an effective solution because it dominates the proportions of the entire system. This article uses a two-axis gimbal, which might be halved in size if a single-axis gimbal were used. However, a single-axis gimbal is less stable than a multi-axis one. Future work will weigh the impact of both size and stability on user experience and system performance. For weight, the lighter the weight of the wearable device, the more beneficial it will be for the user's comfort in long time wearing. The total weight of the system's battery and main control board is 150 grams, so if the embedded system of the intelligent lower limb prosthesis is used as the main control of the gimbal, it is promising to realize a weight reduction of more than 40 %.

V. Conclusion

This paper presents a terrain classification system with a gimbal for high-level control of lower limb prostheses. The LADRC-based controller can effectively suppress the effects of external disturbances on the wearable gimbal control system during human-machine-environment interaction and provide fast response performance. Offline experiments demonstrate that MobileNetV2 can achieve feasible real-time inference performance on resource-constrained embedded devices, which can make our proposed system much more practical. Real-time indoor and outdoor experiments show that the proposed system can provide stable and consistently high predictive performance during walking or motion pattern transitions, independent of walking speed, gait phase, and installation position. This will greatly simplify the operationalization of the device and help to achieve plug-and-play functionality for terrain classification, even if subjects have varying physical conditions. The proposed method can also lead to more accurate and robust high-level control decisions for lower limb robotics, thus ensuring the user's walking safety in everyday terrain or even complex terrain.

References

[1] G. S. Sawicki, O. N. Beck, I. Kang, and A. J. Young, "The exoskeleton expansion: Improving walking and running economy," *J. NeuroEng. Rehabil.*, vol. 17, no. 1, pp. 1–9, Feb. 2020.

[2] D. Farina et al., "Toward higher-performance bionic limbs for wider clinical use," *Nature Biomed. Eng.*, vol. 7, no. 4, pp. 473–485, May 2021.

[3] J. D. Miller, M. S. Beazer, and M. E. Hahn, "Myoelectric walking mode classification for transtibial amputees," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2745–2750, Oct. 2013.

[4] H. Huang, F. Zhang, L. J. Hargrove, Z. Dou, D. R. Rogers, and K. B. Englehart, "Continuous locomotion-mode identification for prosthetic legs based on neuromuscular-mechanical fusion," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 10, pp. 2867–2875, Oct. 2011.

[5] M. Liu, D. Wang, and H. Huang, "Development of an environment-aware locomotion mode recognition system for powered lower limb prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 4, pp. 434–443, Apr. 2016.

[6] J. Camargo, W. Flanagan, N. Csomay-Shanklin, B. Kanwar, and A. Young, "A machine learning strategy for locomotion classification and parameter estimation using fusion of wearable sensors," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 5, pp. 1569–1578, May 2021.

[7] R. Gupta and R. Agarwal, "Single channel EMG-based continuous terrain identification with simple classifier for lower limb prosthesis," *Biocybernetics Biomed. Eng.*, vol. 39, no. 3, pp. 775–788, Jul. 2019.

[8] D. Joshi and M. E. Hahn, "Terrain and direction classification of locomotion transitions using neuromuscular and mechanical input," *Ann. Biomed. Eng.*, vol. 44, no. 4, pp. 1275–1284, Apr. 2016.

[9] B. Chen et al., "Locomotion mode classification using a wearable capacitive sensing system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 5, pp. 744–755, Sep. 2013.

[10] E. Zheng, L. Wang, K. Wei, and Q. Wang, "A noncontact capacitive sensing system for recognizing locomotion modes of transtibial amputees," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 12, pp. 2911–2920, Dec. 2014.

[11] E. Zheng and Q. Wang, "Noncontact capacitive sensing-based locomotion transition recognition for amputees with robotic transtibial prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 2, pp. 161–170, Feb. 2017.

[12] F. Gao, G. Liu, F. Liang, and W.-H. Liao, "IMU-based locomotion mode identification for transtibial prostheses, orthoses, and exoskeletons," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1334–1343, Jun. 2020.

[13] R. Stolyarov, M. Carney, and H. Herr, "Accurate heuristic terrain prediction in powered lower-limb prostheses using onboard sensors," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 2, pp. 384–392, Feb. 2021.

[14] J. S. Matthis, J. L. Yates, and M. M. Hayhoe, "Gaze and the control of foot placement when walking in natural terrain," *Current Biol.*, vol. 28, no. 8, pp. 1224–1233, Apr. 2018.

[15] A. H. A. Al-dabbagh and R. Ronsse, "A review of terrain detection systems for applications in locomotion assistance," *Robot. Auto. Syst.*, vol. 133, Nov. 2020, Art. no. 103628.

[16] D. Xu, Y. Feng, J. Mai, and Q. Wang, "Real-time on-board recognition of continuous locomotion modes for amputees with robotic transtibial prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 10, pp. 2015–2025, Oct. 2018.

[17] A. J. Young, A. M. Simon, N. P. Fey, and L. J. Hargrove, "Classifying the intent of novel users during human locomotion using powered lower limb prostheses," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Nov. 2013, pp. 311–314.

[18] S. Sahoo, M. Maheshwari, D. K. Pratihar, and S. Mukhopadhyay, "A geometry recognition-based strategy for locomotion transitions early prediction of prosthetic devices," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1259–1267, Apr. 2020.

[19] B. Kleiner, N. Ziegenspeck, R. Stolyarov, H. Herr, U. Schneider, and A. Verl, "A radar-based terrain mapping approach for stair detection towards enhanced prosthetic foot control," in *Proc. 7th IEEE Int. Conf. Biomed. Robot. Biomechatronics*, Aug. 2018, pp. 105–110.

[20] N. E. Krausz, T. Lenzi, and L. J. Hargrove, "Depth sensing for improved control of lower limb prostheses," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2576–2587, Nov. 2015.

[21] R. Munoz, X. Rong, and Y. Tian, "Depth-aware indoor staircase detection and recognition for the visually impaired," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2016, pp. 1–6.

[22] B. H. Hu, N. E. Krausz, and L. J. Hargrove, "A novel method for bilateral gait segmentation using a single thigh-mounted depth sensor and IMU," in *Proc. 7th IEEE Int. Conf. Biomed. Robot. Biomechatronics (Biorob)*, Aug. 2018, pp. 807–812.

[23] A. H. A. Al-Dabbagh and R. Ronsse, "Depth vision-based terrain detection algorithm during human locomotion," *IEEE Trans. Med. Robot. Bionics*, vol. 4, no. 4, pp. 1010–1021, Nov. 2022.

[24] K. Zhang et al., "Environmental features recognition for lower limb prostheses toward predictive walking," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 465–476, Mar. 2019.

[25] K. Zhang et al., "A subvision system for enhancing the environmental adaptability of the powered transfemoral prosthesis," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3285–3297, Jun. 2021.

[26] B. Zhong, R. L. D. Silva, M. Li, H. Huang, and E. Lobaton, "Environmental context prediction for lower limb prostheses with uncertainty quantification," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 458–470, Apr. 2021.

[27] B. Zhong, R. L. D. Silva, M. Tran, H. Huang, and E. Lobaton, "Efficient environmental context prediction for lower limb prostheses," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 6, pp. 3980–3994, Jun. 2022.

[28] S. Carvalho, J. Figueiredo, and C. P. Santos, "Environment-aware locomotion mode transition prediction system," in *Proc. IEEE Int. Conf. Auto. Robot Syst. Competitions (ICARSC)*, Apr. 2019, pp. 1–6.

[29] A. G. Kurbis, B. Laschowski, and A. Mihailidis, "Stair recognition for robotic exoskeleton control using computer vision and deep learning," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2022, pp. 1–6.

[30] B. Laschowski, W. McNally, A. Wong, and J. McPhee, "Environment classification for robotic leg prostheses and exoskeletons using deep convolutional neural networks," *Frontiers Neurorobotics*, vol. 15, Feb. 2022, Art. no. 730965.

[31] K. Zhang et al., "Foot placement prediction for assistive walking by fusing sequential 3D gaze and environmental context," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2509–2516, Apr. 2021.

[32] R. J. Rajesh and P. Kavitha, "Camera gimbal stabilization using conventional PID controller and evolutionary algorithms," in *Proc. Int. Conf. Comput., Commun. Control (IC)*, Sep. 2015, pp. 1–6.

[33] B. Ahi and A. Nobakhti, "Hardware implementation of an ADRC controller on a gimbal mechanism," *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 6, pp. 2268–2275, Nov. 2018.

[34] A. Altan and R. Hacioglu, "Model predictive control of three-axis gimbal system mounted on uav for real-time target tracking under external disturbances," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106548.

[35] J. Q. Han, "From PID technique to active disturbances rejection control technique," *Control Eng. China*, vol. 9, no. 3, pp. 13–18, Jun. 2002.

[36] G. Herbst, "A simulative study on active disturbance rejection control (ADRC) as a control tool for practitioners," *Electronics*, vol. 2, pp. 246–279, Sep. 2013.

[37] A. Wong, "NetScore: Towards universal metrics for large-scale performance analysis of deep neural networks for practical on-device edge usage," in *Proc. Int. Conf. Image Anal. Recognit.*, vol. 11663, 2019, pp. 15–26.

[38] A. M. Simon et al., "Delaying ambulation mode transition decisions improves accuracy of a flexible control system for powered knee-ankle prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 8, pp. 1164–1171, Aug. 2017.

[39] M. A. McDowell, C. D. Fryar, C. L. Ogden, and K. M. Flegal, "Anthropometric reference data for children and adults: United States, 2003–2006," *Nat. Health Stat. Rep.*, vol. 10, pp. 1–48, Oct. 2008.