Environment-based Assistance Modulation for a Hip Exosuit via Computer Vision

Enrica Tricomi[1], Mirko Mossini[2], Francesco Missiroli[1], Nicola Lotti[1], Michele Xiloyannis[3], Loris Roveda[4], and Lorenzo Masia[1]

arXiv:2211.15346v1 [cs.RO] 28 Nov 2022

Abstract—Just like in humans vision plays a fundamental role in guiding adaptive locomotion, when designing the control strategy for a walking assistive technology, Computer Vision may bring substantial improvements when performing an environment-based assistance modulation. In this work, we developed a hip exosuit controller able to distinguish among three different walking terrains through the use of an RGB camera and to adapt the assistance accordingly. The system was tested with seven healthy participants walking throughout an overground path comprising of staircases and level ground. Subjects performed the task with the exosuit disabled (*Exo Off*), constant assistance profile (*Vision Off*), and with assistance modulation (*Vision On*). Our results showed that the controller was able to promptly classify in real-time the path in front of the user with an overall accuracy per class above the 85%, and to perform assistance modulation accordingly. Evaluation related to the effects on the user showed that *Vision On* was able to outperform the other two conditions: we obtained significantly higher metabolic savings than *Exo Off*, with a peak of $\approx -20\%$ when climbing up the staircase and $\approx -16\%$ in the overall path, and than *Vision Off* when ascending or descending stairs. Such advancements in the field may yield to a step forward for the exploitation of lightweight walking assistive technologies in real-life scenarios.

Index Terms—Exosuits; Computer Vision; Adaptive Walking Assistance; Assistive Robotics.

I. INTRODUCTION

Vision plays a fundamental role in bipedal walking, as it contributes in modulating navigation according to the surrounding space [1]. In humans, environment scanning is a major actor in providing feedback loop to the central nervous system for cognitive-to-motor transformations, thereby endorsing adaptive locomotion [2].

Within the context of human walking assistance through wearable devices, unlike humans, robot control strategies have been largely based on gait kinetics and kinematics, and investigations have often excluded the users' surrounding environment by studying locomotion in restricted cases (e.g., treadmill or straight overground walking) [1]. However, in the real world, it is very likely that users need to continuously

[1] E. Tricomi, F. Missiroli, N. Lotti, and L. Masia are with the Institut für Technische Informatik (ZITI), Heidelberg University, 69120 Heidelberg, Deutschland.

[2] M. Mossini is with Department of Mechanical Engineering, Politecnico di Milano, via La Masa 1, 20156 Milan, Italy.

[3] M. Xiloyannis is with the Institute of Robotics and Intelligent Systems, ETH Zürich, Zürich, 8092, Switzerland.

[4] L. Roveda is with Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA), Scuola Universitaria Professionale della Svizzera Italiana (SUPSI), Università della Svizzera italiana (USI), via la Santa 1, 6962, Lugano, Switzerland

∗ corresponding author: enrica.tricomi@ziti.uni-heidelberg.de

and dynamically modify their natural gait pattern to adapt to stairs, ramps, or curved path [3]. To account for the change in kinematics when walking on dissimilar terrains, accurate modulation of the assistance delivered to the targeted joints, in terms of timing and magnitude, has to be provided to minimize walking energy expenditure and, therefore, to maximize users' benefits [4], [5], [6], [7].

In the current state of the art, diverse methods addressed this issue implementing manual hand switching strategies or automatic techniques based on the user's biomechanics, yet identification and smooth transitions across different locomotion modalities remains one of the major challenges in assistive robotics [8], [9], [10]. Additional analyses about the user's extra-personal space should complement gait recognition strategies in order to have proper and tailored assistance modulation [11].

As humans rely on vision and the underlying neural feedback to achieve this goal, a feasible approach to embed the aforementioned information in the control loop of robotic systems is to perform environment scanning and classification through computer vision. The approach has been already extensively tested in the context of wheeled and legged robots [12], [13], where camera-based algorithms are developed in order to guide the system navigation on various terrains and path planning. Nonetheless, a very small niche in literature is taken up by the inclusion of computer vision in the control of wearable assistive devices. The idea is recently getting a foothold among researchers, but very few works exist so far [14], [11], [6], [15]. Moreover, the contribution coming from these works is confined to a conceptual level, i.e., authors verified the performance of the vision algorithm, but the device was never tested in an active modality on the users [8].

In this framework, we developed and tested a novel control strategy for a hip exosuit incorporating computer vision to perform assistance modulation based on classification of the environment in front of the wearer recorded through an RGB camera. Our controller is able to distinguish if the user is about to climb up or down a staircase, or to walk on level ground, and to adapt the assistance profile accordingly in order to comply with the change in the gait kinematic pattern [16]. The underlying hypothesis is that an adaptive controller would improve the metabolic benefit of receiving assistance from the exosuit than a static controller (i.e., assistance magnitude does not change depending on the task). To prove the point, we present here a comparison between the cases in which computer vision is active versus when it is not, to verify the enhanced performance of the system by using the former. As far as we are aware of, this is the first work in which a lower-
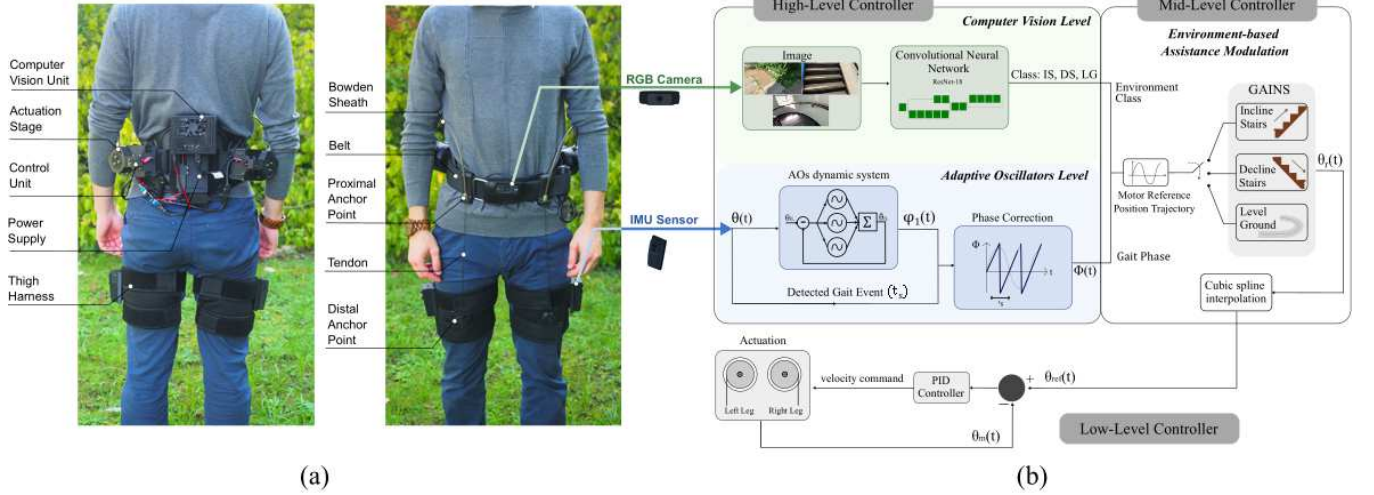
Fig. 1. *Hip exosuit design and real-time control framework* (a) The exosuit is a fully-actuated tendon-driven system that supports hip flexion during walking. It is made of a lightweight belt hosting the actuation stages, control unit, and batteries, and two thigh harnesses. Two inertial sensors capture the user's kinematics and an RGB camera scans the environment around the user (b) The Real-Time control framework is designed to distinguish among three kinds of terrains: Incline Stairs, Level Ground, or Decline Stairs. In the High-Level controller, images from the camera are processed in the *Computer Vision level* in parallel to the users' gait phase estimation via Adaptive Oscillators (*Adaptive Oscillators level*). The combination of these output is processed at the Mid-Level to generate an appropriate motor reference trajectory delivered to the actuator at the Low-Level.

limb exosuit incorporating computer vision is tested with active assistance complemented by an analysis of its effects on study participants. Within the context of the present work, our main goal was to develop an assistive technology able to produce environment-based smooth and flexible assistance, such that users could take full advantage from its use and real-world deployment can be promoted in the long term.

II. EXOSUIT DESIGN

The control strategy based on computer vision was tested on a fully-actuated exosuit to support hip flexion during natural human locomotion. The assistive device (Fig.1-a) comprises of (i) two actuation stages, one for each leg, and main core of the active support system, (ii) a pair of textile harnesses wrapped around the users' thighs, (iii) and a belt (RDX Sport, Stafford, Texas, USA) used to hold the actuation and electronics. Assistive forces are transmitted from the actuation stages to the user via external tendons (Black Braided Kevlar Fiber, KT5703-06, $2.2\,$kN max load, Loma Linda CA, USA). The weight of the device is $3.6\,$kg, and it is powered by a battery (Tattu, $14.8\,$V, $3700\,$mAh, 45C) allowing the user to receive active support for ≈ 10h of continuous operation.

Each actuation stage comprises of a flat brushless motor (T-Motor, AK60-6, $24\,$V, 6:1 planetary gear-head reduction, Cube Mars actuator, TMOTOR, Nanchang, Jiangxi, China) driving a pulley with $\varnothing 35\,$mm diameter wounding the artificial tendon of the corresponding leg. Two Bowden cables (Shimano SLR, $\varnothing 5\,$mm, Sakai, Ōsaka, Japan) connect the actuation units to two proximal anchor points, placed at the level of the waist and anchored to the belt, and are needed to cover the artificial tendons in charge of transferring the mechanical power of the motors to the user. The artificial tendons are connected to the subject's thighs via 3D printed distal anchor points sewed on the two soft fabric harnesses.

The sensors embedded in the suit comprise of two Inertial Measurement Units (IMU, Bosch, BNO055, Gerlingen, Germany), placed laterally on each thigh harness and used to detect the hip joint kinematics, each communicating with one control unit via Bluetooth Low Energy communication protocol (BLE, Feather nRF52 Bluefruit, Adafruit).

The control unit is in charge of running the real-time control algorithm: the *High-Level Controller* incorporating the computer vision framework runs on an embedded processor (NVIDIA Jetson Nano, Santa Clara, CA, USA) at $30\,$Hz using an RGB camera (Logitech C920s PRO HD WEBCAM, Newark, CA, USA) placed in the frontal part of the belt as acquisition unit, while the *Mid-Level* and *Low-Level* run on microcontrollers (Arduino MKR 1010 WiFi, Arduino, Ivrea, Italy) at $100\,$Hz. The boards communicate with each other via Universal Serial Bus protocol.

The control framework was implemented in MAT-LAB/Simulink (MathWorks, Natick, Massachusetts MA, USA), with the exception of the Computer Vision algorithm written in Python.

III. REAL-TIME CONTROL FRAMEWORK

We implemented a real-time Control Framework for our hip exosuit based on computer vision to determine the path in front of the user and to discriminate among three different walking conditions (i.e., incline stairs (IS), level ground (LG), and decline stairs (DS)) with the aim to perform an environment-based assistance modulation.

The control strategy (see Figure 1-b) is composed of three layers: a *High-Level Controller* running the computer vision algorithm in parallel to a gait phase estimation approach, a *Mid-Level Controller* generating the actuator reference motion based on the environment and user's motion predicted at the High-Level, and a *Low-Level Controller* implemented as a classical PID.
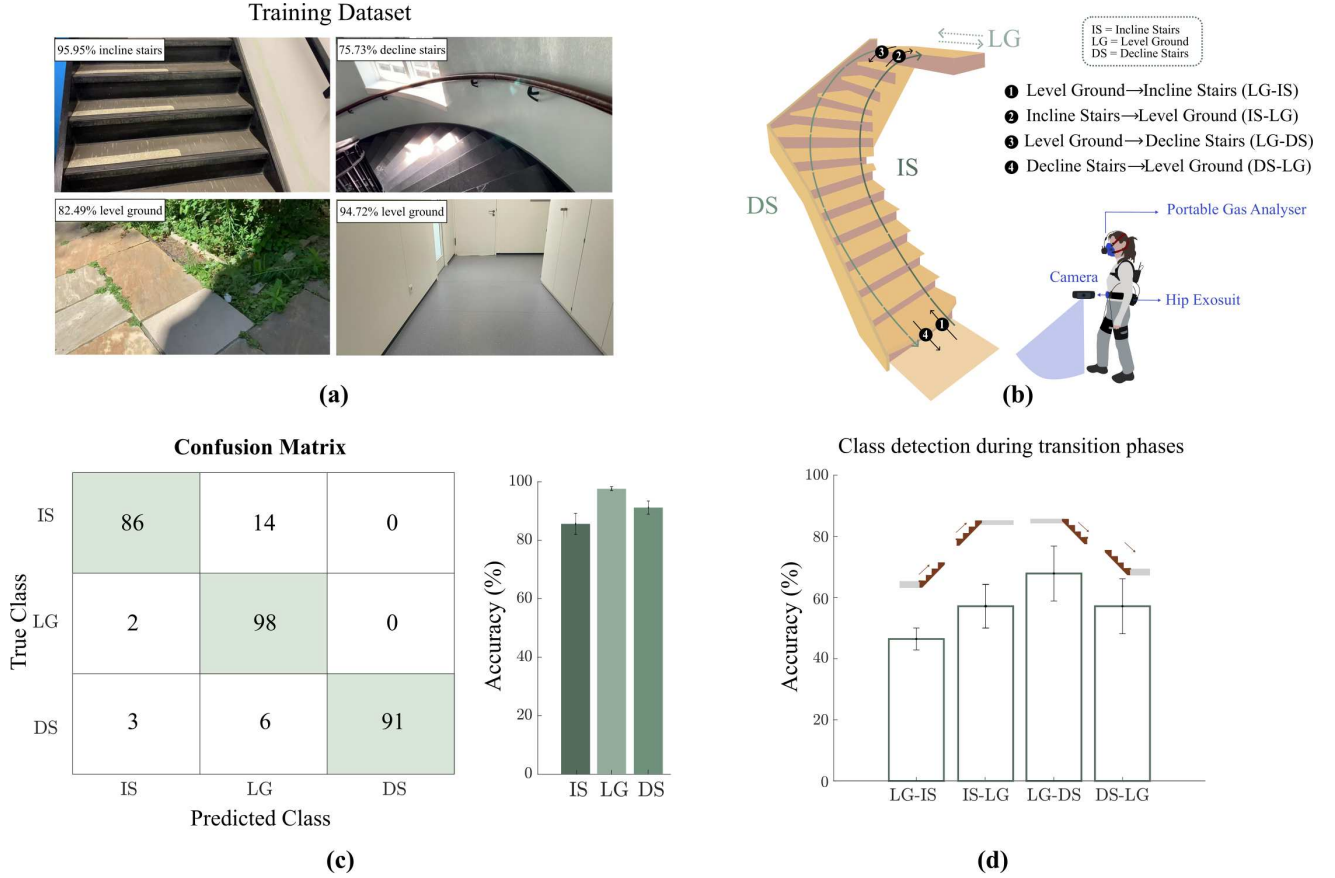
(a)

(b)

(c)

(d)

Fig. 2. *Computer Vision* (a) Examples of images contained in the training dataset for the ResNet-18 neural network. (b) Experimental Protocol: participants walked throughout a path transitioning from level ground - stair climbing - level ground and viceversa. The task was repeated with the exosuit disabled, without Computer Vision enabled, and with Computer Vision. (c) Classification accuracy per class: each class reported a classification accuracy above 85%. The confusion matrix is shown to report the distribution of miclassifications across subjects. (d) Accuracy across transition phases: each bar represents the mean accuracy computed considering the step before and after each path transition. Here classification performance drops down to values $< 80\%$.

A. High-Level Controller: Environment classification through computer vision and gait phase estimation

The High-Level Controller operates on two separate levels running in parallel to perform: (1) environment classification and (2) user's gait phase recognition through Adaptive Oscillators. We will refer to them as the *Computer Vision level* and the *Adaptive Oscillators level*.

1) Computer Vision level: The implemented Computer Vision level is based on a Convolutional Neural Network (CNN) model running at 30 Hz, able to predict the environment in the near space of the user through an RGB camera. When it comes to image classification, CNNs have shown capabilities to outperform other kinds of classifiers, such as support vector machines [11]. This is the case because multi-layer neural networks are able to optimally learn image features from training data, thus performing a more effective classification [14]. However, they require a vast dataset for network training in order to promote generalization.

In our application, we adopted a CNN model composed of 18 layers (ResNet-18 [17]) that was pre-trained on 10.500 images (see Section III-D). This net has been widely used in image classification applications for its low computational costs, high accuracy and simplicity to train [17]. The net takes

as input the environmental images recorded in real-time by the camera placed on the belt of the user. These are processed with a resolution of $400 \text{x} 400$ pixels and classified according to three classes: Incline Stairs (IS), Level Ground (LG), and Decline Stairs (DS). Finally, the outputs of this level are the predicted class (i.e., IS, LG, and DS) and its confidence, which are sent, together with gait phase data, to the *Mid-Level Controller*, as explained in a dedicated section (III-B).

2) Adaptive Oscillators level: The user's gait phase during locomotion is estimated in real-time from the hip joint kinematic (obtained by signals recorded by two IMUs) through the algorithm based on Adaptive Oscillators (*AOs*) presented in our previous works [18], [4]. The hip flexion angle of each leg $\theta(t)$ is reconstructed as sum of 3 sinusoids and indicated as $\hat{\theta}(t)$:

$$\hat{\theta}(t) = \alpha_0(t) + \sum_{n=1}^{3} \alpha_n(t) \sin(\varphi_n(t)) \tag{1}$$

whose amplitude $\alpha_n(t)$, phase $\varphi_n(t)$, frequency $\omega(t)$, and offset $\alpha_0(t)$ are continuously estimated via an error-driven approach via the function $F(t) = \theta(t) - \hat{\theta}(t)$, as follows:

$$\dot{\alpha}_0(t) = \eta F(t) \tag{2}$$

$$\dot{\alpha}_n(t) = \eta F(t) \sin(\varphi_n(t)) \tag{3}$$

$$\dot{\varphi}_n(t) = \omega(t) \cdot n + \nu_\varphi \frac{F(t)}{\sum \alpha_n} \cos(\varphi_n(t)) \tag{4}$$

$$\dot{\omega}(t) = \nu_\omega \frac{F(t)}{\sum \alpha_n} \cos(\varphi_1(t)) \tag{5}$$

Gains η, ν_φ, and ν_ω were set respectively equal to 5, 20 and 20 from preliminary trials aimed at studying the convergence performance of the algorithm.

Among the AOs extracted quantities, the fundamental phase $\varphi_1(t)$ (i.e., $\varphi_n(t)$ with $n = 1$) is the one guiding the estimation of the user's gait phase: $\varphi_1(t)$ is normalized in the range $[0\ 2\pi)$ rad (i.e., $\varphi'_1(t) = \mathrm{mod}(\varphi_1(t), 2\pi)$), and then shifted according to a corrective factor $\psi(t)$ in charge of aligning the beginning of each step with a value of 0 rad:

$$\psi(t) = \int E_\psi \omega(t) e^{-\omega(t)(t-t_s)} \tag{6}$$

with t_s being the instant in time indicating the beginning of each step and E_ψ defined as below:

$$E_\psi = \begin{cases} -\varphi'_1(t_s) - \psi(t_s), & \text{if } 0 \le \varphi'_1(t_s) < \pi \\ 2\pi - \varphi'_1(t_s) - \psi(t_s), & \text{if } \pi \le \varphi'_1(t_s) < 2\pi \end{cases} \tag{7}$$

The final gait phase variable, here named $\phi(t)$, is an indication of the progression along the gait cycle at each step, and it is eventually obtained as:

$$\phi(t) = \mathrm{mod}(\varphi'_1(t) + \psi(t), 2\pi) \tag{8}$$

B. Mid-Level Controller: Environment-based assistance modulation

The Mid-Level Controller adjusts the level of assistance delivered to each leg according to the classification of the environment in front of the user performed by the *Computer Vision level* and the estimated wearer's motion.

The reference motor position trajectory, $\theta_r(t)$, is generated from the gait phase data obtained from the *Adaptive Oscillators level* as below:

$$\theta_r(t) = \begin{cases} M_{IS} \sin(\phi(t) - \pi) & \text{if Incline Stairs} \\ M_{LG} \sin(\phi(t) - \pi) & \text{if Level Ground} \\ M_{DS} \sin(\phi(t) - \pi) & \text{if Decline Stairs} \end{cases} \tag{9}$$

being M_{IS}, M_{LG}, and M_{DS} gains that modulates the amplitude of walking assistance depending on the environment classified at each step (IS, LG, and DS). These three gains were tuned from preliminary trials according to kinematic considerations: we took into account the relative change between hip flexion profiles during swing according to the work of Riener et al. [16].

Considering that environment classification is continuously performed throughout the walking according to the image acquisition frequency, in order to guarantee a smooth transition between assistive profiles, we implemented a "ranked-voting" procedure to decide the final class within each step performed by the user. In detail, gait phase data are used to identify the times in which the subject reaches the maximum hip extension: this is, indeed, an indication of the beginning of the swing phase of walking and the time instant at which the device should apply zero assistive torque. In such a way, any change in the assistance amplitude does not cause users's movement disruptions. At this instant, which of the three equations in (9) should take place is decided according to the overall classification confidence reached along the previous step:

$$\begin{cases} C_{IS} = \sum_{n=1}^{N_{IS}} conf_{IS} & \forall\ t_n \text{ with class IS} \\ C_{LG} = \sum_{n=1}^{N_{LG}} conf_{LG} & \forall\ t_n \text{ with class LG} \\ C_{DS} = \sum_{n=1}^{N_{DS}} conf_{DS} & \forall\ t_n \text{ with class DS} \end{cases} \tag{10}$$

where C_{IS}, C_{LG}, and C_{DS} are counters indicating the goodness of prediction and are defined as the confidence values, $conf$, summed over the times that the respective class was predicted (N_{IS}, N_{DS}, N_{LG}) at each time stamp (t_n) along a step. The class with the highest counter value is then chosen and the three counters are subsequently reset to zero. The described procedure was implemented in order to enhance the robustness of the control algorithm to misclassifications occurring in the sporadic case of scarce lightening or confounding visual factors.

At last, the final actuator reference motion, $\theta_{ref}(t)$, is obtained from $\theta_r(t)$ after cubic spline interpolation and sent to the *Low-Level Controller*.

C. Low-Level Controller

The Low-Level Controller compares the measured device state, i.e., actual position of the motor, with the desired one, i.e., the reference motion $\theta_{ref}(t)$, to minimize the error through reference tracking via closed-loop feedback position control. The position error is converted into motor angular velocity through a PID controller.

D. Computer Vision Training and Parameters Optimization

The CNN chosen model (i.e., ResNet-18) was trained with a dataset composed of 5000 manually selected images contained in the open-source dataset ExoNet [11], complemented by 5500 extra images acquired by the authors. Figure 2-a presents some examples contained in the training dataset. The total 10500 images were divided among the three classes and then scaled down to 400x400 pixels.

The dataset was split such that the 80% was used as training set, the 10% as validation set, and the 10% as test set. The network was trained for 60 epochs, using the Adam algorithm as optimizer, a batch size of 16 and a learning rate of 0.1, the latter decreased of ten times every 20 epochs.

IV. STUDY DESIGN

Seven healthy subjects were enrolled to test the system (3 males/4 females, age 25.0 ± 2.2 years, mean$\pm$SD, weight 60.9 ± 13.7kg, height 169.0 ± 9.9cm). Informed consent forms were signed before the beginning of experiments. Research procedures were performed according to the Declaration of Helsinki and were approved by the Ethical Committee of Heidelberg University (resolution S-313/2020).
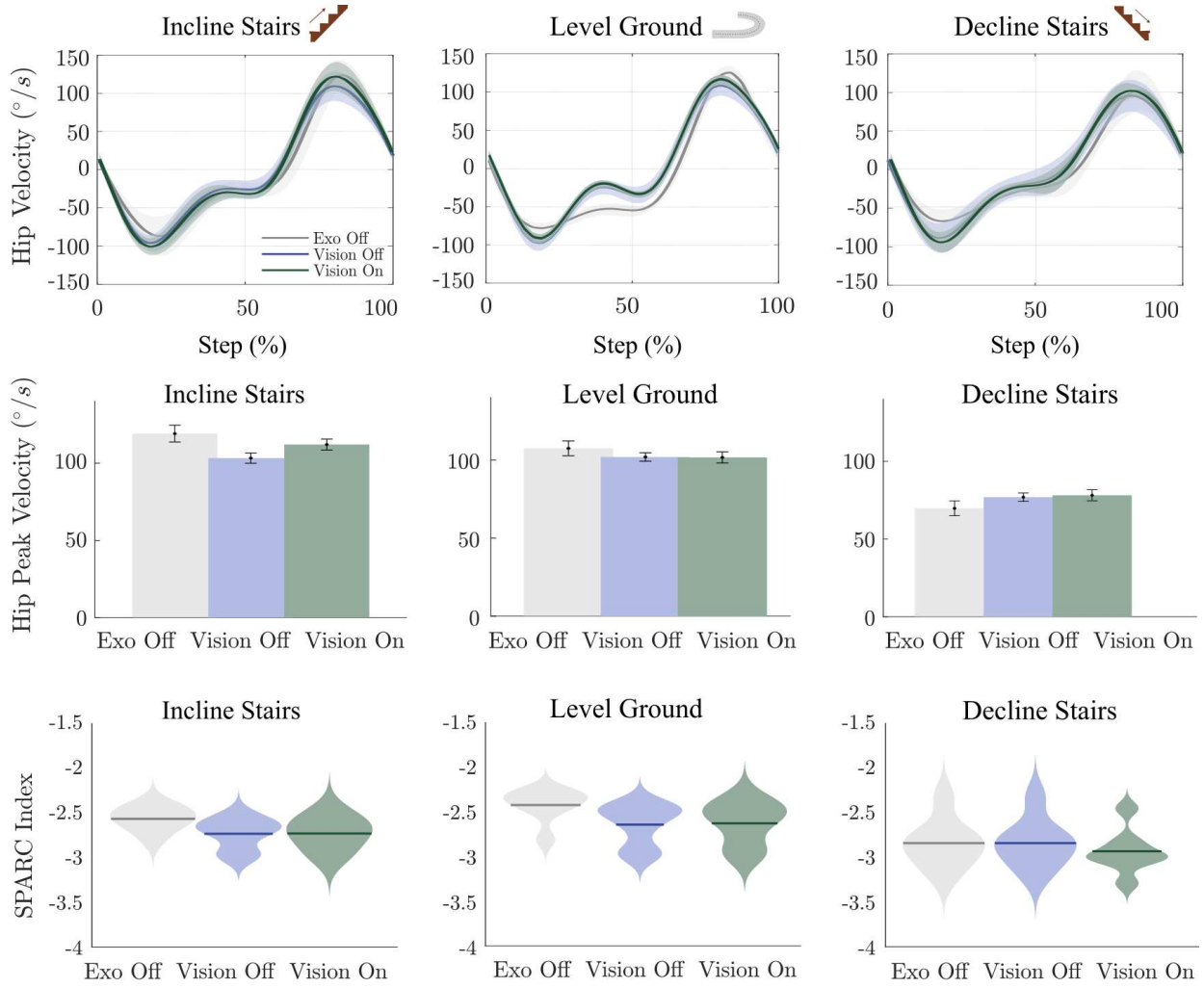
Fig. 3. *Kinematic Assessments*. Hip velocity profiles as percentage of the step duration for a typical subject are shown in the top row for the three path segments. Solid lines indicate the mean across steps and legs (grey for *Exo Off*, blue for *Vision Off*, and green for *Vision On*); shaded areas represent the standard deviation. The second and third row report the peak velocity during swing as average across steps and subjects, and the movement smoothness expressed in terms of SPectral ARC length (SPARC) index.

A. Experimental Protocol and Data Acquisition

The study participants were asked to walk at their preferred speed along a path with total length equal to $250\,\mathrm{m}$ arranged over different floors, comprising eight consecutive ramps of stairs (for a total amount of 88 stairs, each 120x18x26 cm) and $150\,\mathrm{m}$ of level ground walking. All subjects walked through the path starting from the bottom climbing up the full stairway from a level ground condition; once at the top they proceeded along the level ground path, and, finally, downstairs until the starting point (Figure 2-b).

The path described above was repeated three times in three different conditions: (a) exosuit disabled (*Exo Off*); (b) exosuit enabled without environment-based modulation (*Vision Off*); (c) exosuit enabled with environment-based modulation (*Vision On*). Condition (b) - *Vision Off* was implemented as described in section III by imposing the same gain to the three equations in (9) (i.e., M_{LG}). Participants rested for at least $20\,\mathrm{min}$ in between trials to limit the effects of fatigue and conditions were randomized across subjects to avoid order effects.

We collected oxygen and carbon dioxide consumption data through a portable gas analyser (K5, COSMED, Rome, Italy) to assess the metabolic energy expenditure and the effects of receiving assistance from the device with environment-based modulation compared to the remaining two conditions. At the beginning of the experiment, subjects were asked to breath normally for $4\,\mathrm{min}$ in a standing resting condition. The average metabolic cost at quite standing was subtracted from the metabolic data in later analysis to extract the cost of walking for each repetition of the task.

Classification data from the *Computer Vision level*, the hip kinematics recorded from IMU sensors, and the actuator torques were also recorded for all subjects across repetitions.

B. Data Analysis

Computer Vision We analysed the performance of the *Computer Vision level* of the controller in terms of real-time class prediction accuracy. For each class (IS, LG, and DS)
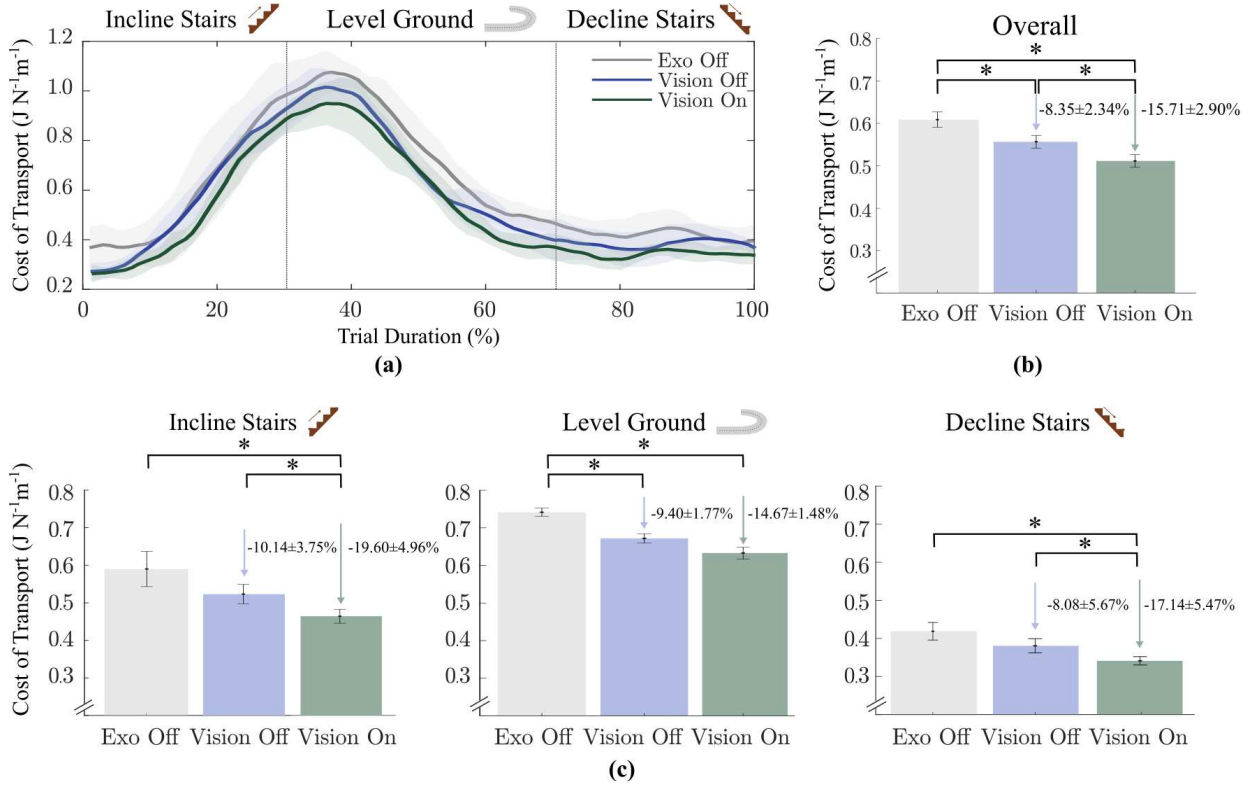
Fig. 4. *Cost of Transport*. (a) CoT profiles for the *Exo Off* (grey), *Vision Off* (blue), and *Vision On* (green) conditions averaged across subjects. Solid lines represent the mean curve, while shaded areas are the standard deviations. For visual representation, timeseries are presented as percentage of trial duration. (b) Mean CoT computed on the overall path. (c) Mean CoT computed separately in the three path segments (IS, LG, DS). Savings for the *Vision Off* and *Vision On* conditions are computed in relation to *Exo Off*. Statistical significance in pairwise comparisons is indicated with the symbol *.

we assessed the number of correct predictions, True Positives (TP), as a percentage of the total number of predicted gait cycles belonging to a class, i.e., the sum of the TP and False Positives (FP):

$$Accuracy(\%) = \frac{TP}{TP + FP} \tag{11}$$

Moreover, to assess how this metric varies across transition phases, as the change from level ground to stairs and vice versa, we evaluated the accuracy of prediction corresponding to the step before and the one after each transition.

Kinematic Assessments Kinematic evaluations were performed on hip angular velocity profiles after segmentation into steps. We evaluated the effects of the *Vision On* and *Vision Off* conditions with respect to *Exo Off* in terms of swing peak velocity and movement smoothness. The latter was computed as Spectral ARC length (SPARC) index on the norm of the hip angular velocity as defined in [19].

Cost of Transport Oxygen and carbon dioxide data were analysed to estimate the energy efficiency of walking for the *Exo Off*, *Vision Off*, and *Vision On* conditions for the overall path and the three path segments (IS, LG, and DS). Since subjects were instructed to complete the path at self-selected speed, to account for the different walking patterns across conditions, we analysed the metabolic data in terms of Cost

of Transport (CoT) according to the following definition:

$$CoT = \frac{P}{wgv} \tag{12}$$

being P the net metabolic cost of walking estimated using the Péronnet and Massicotte's equation [20] subtracted by the average cost at quite standing, w the subject's weight, g the gravitational acceleration, and v the average walking speed computed as covered distance over time.

C. Statistical Analysis

We assessed the data normality distribution with the Shapiro-Wilk test ($\alpha = 0.05$). All the analysed metrics resulted normally distributed. We evaluated the three conditions presented in the study (*Exo Off* vs. *Vision Off* vs. *Vision On*) for kinematics and CoT data via an one-way repeated measurements ANalysis Of VAriance (ANOVA). Sphericity condition was evaluated using the Mauchly's test. When the ANOVA results were significant, we performed a post-hoc analysis using paired t-tests with Bonferroni correction to evaluate pairwise differences between conditions. Reported measurements are presented as mean $\pm$ standard error (SE).

V. RESULTS

Computer Vision The performance of the *Computer Vision level* are presented in Figure 2-c and -d in terms of accuracy per class and for transition phases as a mean across subjects

and conditions. For each class (Figure 2-c), we obtained a prediction accuracy above 85%: $85.56 \pm 3.58\%$ for IS, $97.57 \pm 0.69\%$ for LG, and $91.14 \pm 2.25\%$ for DS. Results are graphically presented also in terms of confusion matrix to report the distribution of misclassifications. Instead, lower values were obtained when analysing the classifier accuracy in correspondence of transition phases (Figure 2-d): $46.43 \pm 3.57\%$ for the transition from LG to IS, $57.14 \pm 7.14\%$ from IS to LG, $67.86 \pm 8.99\%$ from LG to DS, and $57.14 \pm 8.99\%$ from DS to LG.

Kinematic Assessments Results of the kinematic assessments are reported in Figure 3 for the three path segments. For the sake of a simpler result presentation, data are reported as mean between legs, since the one-way ANOVA showed no side-related difference. The top row of Figure 3 shows the time series of the hip angular velocity of a representative subject averaged across steps for the three conditions and divided by path segments. Peak velocities reached during the swing phase and the movement smoothness are reported in the second and third row, respectively. Statistical analysis did not show significant differences across conditions in any of the three path segments, as a sign of device transparency towards the users' physiological kinematic pattern.

Cost of Transport Results about metabolic evaluations are presented in Figure 4. Exclusively for this analysis, one of the seven subjects was left out because of fatigue onset in subsequent tests. Figure 4-a shows the Cost of Transport (CoT) timeseries as a percentage of trial duration averaged across subjects for the three conditions. Considering the overall CoT along the path, results showed that subjects experienced the highest energy expenditure when walking in the *Exo Off* condition, while the energy cost of walking decreased progressively during *Vision Off* and *Vision On*, with the latter reaching the lowest value (see Figure 4-a and b). Here, CoT resulted significantly different both during *Vision Off* (saving of $-8.35 \pm 2.34\%$, $p < 0.05$) and *Vision On* (saving of $-15.71 \pm 2.90\%$, $p < 0.05$) with respect to *Exo Off*. The decreasing trend across conditions was confirmed by analysing the CoT separately in the three walking segments: we reached savings of $-10.14 \pm 3.75\%$ and $-19.60 \pm 4.96\%$ during IS for *Vision Off* and *Vision On* respectively, $-9.40 \pm 1.77\%$ and $-14.67 \pm 1.48\%$ during LG, and $-8.08 \pm 5.67\%$ and $-17.14 \pm 5.47\%$ during DS. Considering the three walking segments, the *Vision Off* condition showed a statistically significant difference than *Exo Off* only for the LG walking ($p = 0.004$), while *Vision On* resulted always significantly different from *Exo Off* ($p < 0.05$). Additionally, in each presented case we found a statistically significant difference between *Vision Off* and *Vision On*, with the exception of Level Ground walking, where *Vision On* showed slightly higher values than *Vision Off*, yet not enough to achieve significance ($p = 0.09$).

VI. DISCUSSION AND CONCLUSION

By taking inspiration from the role that the visuomotor control plays in humans to promote adaptive locomotion, we developed a control strategy for a hip exosuit based on Computer Vision to scan the environment surrounding the user and to adapt the device assistance accordingly.

To the best of our knowledge, in this context, no other existing works in literature have reached our same advanced stage so far. Works from Laschowski et al. [11], [14] are actively contributing to this research stream by providing open source databases of various walking environments in order to encourage the development of vision-based assistive systems for the lower-limbs. Whereas, the study from Qian et al. [8] is currently one of the most advanced in the field: along similar lines, the authors proposed an High-Level controller based on the combination of computer vision for terrains classification and an AOs-based approach for gait phase estimation. Yet, the Mid- and Low-Level controllers were described just at a conceptual level, and subjects never walked with the device with active assistance mode.

Instead, we propose here a very first example of a fully developed controller for human locomotion assistance via hip exosuit in combination with an experimental evaluation to assess its efficacy. In our system, information from wearable sensors are coupled with the ones coming from environment classification for the purpose of performing assistance modulation based on the near extra-personal space of the user. Compared to our previous work (Zhang et al. [4]), through the incorporation of computer vision running in parallel to the wearer's kinematic estimation, we were able to exploit our system on diverse terrains, such as climbing up and down stairs, while achieving enhanced metabolic energy savings.

Our results proved that the *Computer Vision level* was able to properly classify the three chosen scenarios (IS, LG, and DS) reporting an overall accuracy above 85% for each class. However, class detection during transition phases, i.e., class prediction corresponding to the step before and after each environment change, showed a higher rate of misclassifications and prediction accuracy dropped down to values $< 80\%$. Whether this could depend to sensitivity of camera position or other confounding factors is not clear and future studies should try to enhance robustness across gait transition phases. In our experiments, these misclassifications did not cause any loss of balance or stability nor affected metabolic results in the overall evaluation, nonetheless sort of perfect accuracy in distinguish among different classes is preferred to avoid inconveniences of such kind [21].

Kinematic assessments revealed that users could benefit from the exosuit assistance without experiencing an alteration of the physiological motion. This is a good indicator that our device, being soft and lightweight, does not hinder subjects' movements, as expected from literature [22].

Analysing the users' energy expenditure along the path, the inclusion of computer vision information, namely *Vision On* condition, clearly outperformed *Vision Off*, fully complying with our initial hypothesis. Both in the overall path and in the analysis of path segments, in the *Vision On* condition subjects were able to achieve significantly higher metabolic savings than *Exo Off*, with a peak of $\approx -20\%$ when climbing up the staircase and $\approx -16\%$ considering the overall path. Furthermore, CoT in *Vision On* demonstrated to provide a significantly higher benefit also than *Vision Off*, except during

level ground walking. This is indeed an expected result, given that for this class the controller was set to behave the same. Even so, a trend toward reduction is still noticeable. We can speculate that, since the metabolic rate is largely influenced by the preceding actions, a higher reduction in the *Vision On* condition in LG may result from the larger saving obtained in the previous segment (i.e., IS). To further prove the benefit provided by the *Computer Vision level*, one may notice that the absence of assistance modulation (*Vision Off* condition) did not bring to substantial benefits while climbing up or down the stairs.

Possible limitations of our work include the recognition of solely three cases of terrains and poorer classification performances during path transition phases. Next studies should assess the proposed control strategy in more diverse outdoor evaluations with the inclusion of an extended pool of recruited subjects and a higher number of terrains cases, for instance ascending or descending a ramp, or walk along a curved path. Additionally, further efforts should focus on understanding how to properly detect and assist gait at the edge of environment change.

As of now, the promising results bode well in view of an extended assessment and a possible exploitation of our assistive wearable exosuit in the real world. Just like in humans, the integration of a visual feedback seems to encourage higher human-robot symbiosis, thus bringing to lower energy expenditures. In the long term, such metabolic savings could be beneficial in the case of wellness applications in older adults populations by promoting higher walking cadences or prolonged walking sessions.

ACKNOWLEDGEMENTS

REFERENCES

[1] G. Courtine and M. Schieppati, "Human walking along a curved path. i. body trajectory, segment orientation and the effect of vision," *European Journal of Neuroscience*, vol. 18, no. 1, pp. 177–190, 2003.

[2] Y. Koren, R. Mairon, I. Sofer, Y. Parmet, O. Ben-Shahar, and S. Bar-Haim, "Vision, cognition, and walking stability in young adults," *Scientific reports*, vol. 12, no. 1, pp. 1–10, 2022.

[3] I. Kang, D. D. Molinaro, S. Duggal, Y. Chen, P. Kunapuli, and A. J. Young, "Real-time gait phase estimation for robotic hip exoskeleton control during multimodal locomotion," *IEEE robotics and automation letters*, vol. 6, no. 2, pp. 3491–3497, 2021.

[4] X. Zhang, E. Tricomi, F. Missiroli, N. Lotti, C. Bokranz, D. Nicklas, and L. Masia, "Enhancing gait assistance control robustness of a hip exosuit by means of machine learning," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7566–7573, 2022.

[5] K. Seo, J. Lee, Y. Lee, T. Ha, and Y. Shim, "Fully autonomous hip exoskeleton saves metabolic cost of walking," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 4628–4635.

[6] B. T. Quinlivan, S. Lee, P. Malcolm, D. M. Rossi, M. Grimmer, C. Siviy, N. Karavas, D. Wagner, A. Asbeck, I. Galiana *et al.*, "Assistance magnitude versus metabolic cost reductions for a tethered multiarticular soft exosuit," *Science robotics*, vol. 2, no. 2, p. eaah4416, 2017.

[7] J. Kim, B. T. Quinlivan, L.-A. Deprey, D. Arumukhom Revi, A. Eckert-Erdheim, P. Murphy, D. Orzel, and C. J. Walsh, "Reducing the energy cost of walking with low assistance levels through optimized hip flexion assistance from a soft exosuit," *Scientific reports*, vol. 12, no. 1, pp. 1–13, 2022.

[8] Y. Qian, Y. Wang, C. Chen, J. Xiong, Y. Leng, H. Yu, and C. Fu, "Predictive locomotion mode recognition and accurate gait phase estimation for hip exoskeleton on various terrains," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6439–6446, 2022.

[9] M. R. Tucker, J. Olivier, A. Pagel, H. Bleuler, M. Bouri, O. Lambercy, J. d. R. Millán, R. Riener, H. Vallery, and R. Gassert, "Control strategies for active lower extremity prosthetics and orthotics: a review," *Journal of neuroengineering and rehabilitation*, vol. 12, no. 1, pp. 1–30, 2015.

[10] A. J. Young and D. P. Ferris, "State of the art and future directions for lower limb robotic exoskeletons," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 2, pp. 171–182, 2016.

[11] B. Laschowski, W. McNally, A. Wong, and J. McPhee, "Computer vision and deep learning for environment-adaptive control of robotic lower-limb exoskeletons," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 4631–4635.

[12] Z. Fu, A. Kumar, A. Agarwal, H. Qi, J. Malik, and D. Pathak, "Coupling vision and proprioception for navigation of legged robots," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 273–17 283.

[13] D. Kim, D. Carballo, J. Di Carlo, B. Katz, G. Bledt, B. Lim, and S. Kim, "Vision aided dynamic exploration of unstructured terrain with a small-scale quadruped robot," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2464–2470.

[14] B. Laschowski, W. McNally, A. Wong, and J. McPhee, "Environment classification for robotic leg prostheses and exoskeletons using deep convolutional neural networks," *Frontiers in Neurorobotics*, vol. 15, p. 730965, 2022.

[15] T. Li, X. Wu, H. Dong, and H. Yu, "Estimation of upper limb kinematics with a magnetometer-free egocentric visual-inertial system," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 1668–1674.

[16] R. Riener, M. Rabuffetti, and C. Frigo, "Stair ascent and descent at different inclinations," *Gait & posture*, vol. 15, no. 1, pp. 32–44, 2002.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[18] E. Tricomi, N. Lotti, F. Missiroli, X. Zhang, M. Xiloyannis, T. Muller, S. Crea, E. Papp, J. Krzywinski, N. Vitiello *et al.*, "Underactuated soft hip exosuit based on adaptive oscillators to assist human locomotion," *IEEE Robotics and Automation Letters*, 2021.

[19] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, and E. Burdet, "On the analysis of movement smoothness," *Journal of neuroengineering and rehabilitation*, vol. 12, no. 1, pp. 1–11, 2015.

[20] F. Peronnet, D. Massicotte *et al.*, "Table of nonprotein respiratory quotient: an update," *Can J Sport Sci*, vol. 16, no. 1, pp. 23–29, 1991.

[21] K. A. Inkol and J. McPhee, "Assessing control of fixed-support balance recovery in wearable lower-limb exoskeletons using multibody dynamic modelling," in *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2020, pp. 54–60.

[22] M. Xiloyannis, R. Alicea, A.-M. Georgarakis, F. L. Haufe, P. Wolf, L. Masia, and R. Riener, "Soft robotic suits: State of the art, core technologies, and open challenges," *IEEE Transactions on Robotics*, 2021.