This focus article has been published in Science Robotics, Vol. 9, Issue 90, May 22, 2024:
DOI: https://www.science.org/doi/10.1126/scirobotics.adj8812

Wearable robots for the real world need vision

Letizia Gionfrida[1,2,*,†], Daekyum Kim[3,4,*], Davide Scaramuzza[5], Dario Farina[6], and Robert D. Howe[2,†]

Abstract

To enhance wearable robots, understanding user intent and environmental perception with novel vision approaches is needed.

I. Main Text

We can now imagine a future where many of the one billion people in the world who live with disabilities can go about their daily lives without impairment, thanks to wearable robots [1]. These devices, including exoskeletons and prosthetics, have the potential to revolutionize the way we assist individuals with impairments. For the upper-limb, wearables can provide grip strength and grasp stability in manipulation tasks, and for the lower-limb, they can improve gait patterns and reduce energy expenditure. These systems have seen a surge in development, with initial work largely focused on mechanical design, the interface to the human body, and sensing the user's limbs. This has produced effective systems to aid in basic grasping tasks and locomotion on level terrain [2].

Extension to more sophisticated tasks and higher levels of assistance will require inferring the user's intent. For example, an assistance glove needs to know that the user wants to grasp a particular object to perform specific tasks, and then adapt the grasp type and finger span for that object and task. For leg exoskeletons or prosthetics, the system needs to detect that the user plans to ascend stairs or traverse a slippery walkway, so joint torques can be adjusted to maximize assistance and stability.

At present, the most popular method of inferring user intention for the lower-limb is based on inertial sensors for kinematic information from the user. For example, heel strikes can be estimated using an inertial measurement unit on the foot. A control strategy based on previous gait cycles can predict the current gait cycle by assuming that the user intended a similar locomotion mode. Another method of inferring user intention is by leveraging neuromuscular interfaces, such as electromyography (EMG). This approach measures muscle electrical signals to infer motor activation. For example, EMG signals from body parts proximal to a limb amputation can be used to infer the intentional actions of the missing limb to control active upper-limb prostheses. Interfaces based on these biological signals and the user's behavior provide an estimate of the user's internal states, but the amount of information that can be decoded is limited to simple inferences like detecting changes in walking speed through joint angle sensing or triggering prosthetic hand closure with EMG pulses [3]. This limits wearables to a small number of tasks, and control is often perceived as complex and unnatural by the user [4]. This is one of the reasons for the relatively large abandonment rate of powered upper-limb prostheses.

To expand the range of tasks and quality of assistance, wearable robots must use information about the context where motor actions occur. For example, with extensive machine learning, an EMG sensor on leg muscles can detect changes in muscle activity associated with a transition between level locomotion and ascending stairs. Exclusively based on EMG, the classification error during transitions is four times higher than the classification error during steady-state [5]. On the other hand, the knowledge of the context (the location of the stairs and the direction of walking) would allow a similar prediction several steps ahead and with greater accuracy.

Computer vision can play a central role in obtaining information about the environment and task context. Vision provides rich, immediate, and interpretable information about a user and their surroundings, as demonstrated by human visual capabilities. Recent vision-based technologies for human pose estimation and action classification can provide extensive information about human behaviors [6]. Driver and pedestrian intent prediction can be one good example for benchmark. Sensing the surrounding environment is a well-explored robotics problem that can be achieved by technologies such as object/scene recognition and simultaneous localization and mapping [7].

Merging visual behaviors with contextual information to infer people's intentions is still in its earliest stages [8], and presents unsolved challenges. A general approach could train intention estimation systems end-to-end, using data that includes

[1] Department of Informatics, Faculty of Natural Mathematics and Engineering Sciences, King's College London, Bush House, 30 Aldwych, London WC2B 4BG, United Kingdom.

[2] John A. Paulson School of Engineering and Applied Sciences and Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA.

[3] School of Smart Mobility, Korea University, Seoul 02841, South Korea.

[4] School of Mechanical Engineering, Korea University, Seoul 02841, South Korea.

[5] Robotics and Perception Group, Department of Informatics, University of Zurich, Andreasstrasse 15, 8050 Zurich, Switzerland.

[6] Department of Bioengineering, Faculty of Engineering, Imperial College London, Exhibition Rd, South Kensington, London SW7 2BX, United Kingdom.

[*] These authors contributed equally to this work.

[†] Corresponding author. Email: letizia.gionfrida@kcl.ac.uk (L.G.); howe@seas.harvard.edu (R.D.H.)

video and wearable signals as well as task outcomes. This, however, requires addressing problems such as the representation of appropriate user intentions and labeling of assistance levels. Collecting large amount of data to train a machine learning model could also be problematic for wearable robots that target people with disabilities such as spinal cord injury or stroke, where large numbers of trials may be difficult to collect and motor capabilities will vary greatly between users. It is underexplored whether data collected from healthy individuals can be successfully applied to assist people with disabilities.

Vision-based intent detection will require new control strategies that rely on visual inputs and incorporate vision-in-the-loop for end-to-end control. A camera sensor coupled to a machine learning algorithm can pose a large computational load for an embedded device in a wearable robot. The controller must be able to react quickly in accordance with human movements, and therefore, slow visual data and long inference times will limit utility. Vision can also be combined with wearable sensors that provide information on internal states of the user such as EMG or ultrasound sensors.

Ensuring safety, privacy, and acceptability of these systems will be a major issue; for example, visual occlusions could cause misclassification of intent, which could result in injury. While privacy remains a concern, vision-based home rehabilitation systems that adhere to the Health Insurance Portability and Accountability Act of 1996 (HIPAA) can offer a precedent for tailored assistance and augmentation both within and beyond clinical settings. Conducting market research is also imperative to gain an enhanced understanding of customer perceptions and draw lessons from prior setbacks and challenges that beset initial attempts.

Maximizing the range and complexity of tasks will ultimately require information from the user's body and visual information about the environment (Fig. 1). The final control strategy will be semi-autonomous, requiring shared control, which needs to be transparent and acceptable for the user. Event-based cameras [9] enable new embedded vision applications addressing frame-based camera bandwidth-latency tradeoffs, vital for wearable robots requiring high update rates with low bandwidth. The combination promises to empower wearable robots, contributing to ongoing efforts aimed at promoting a higher quality of life for those in need.

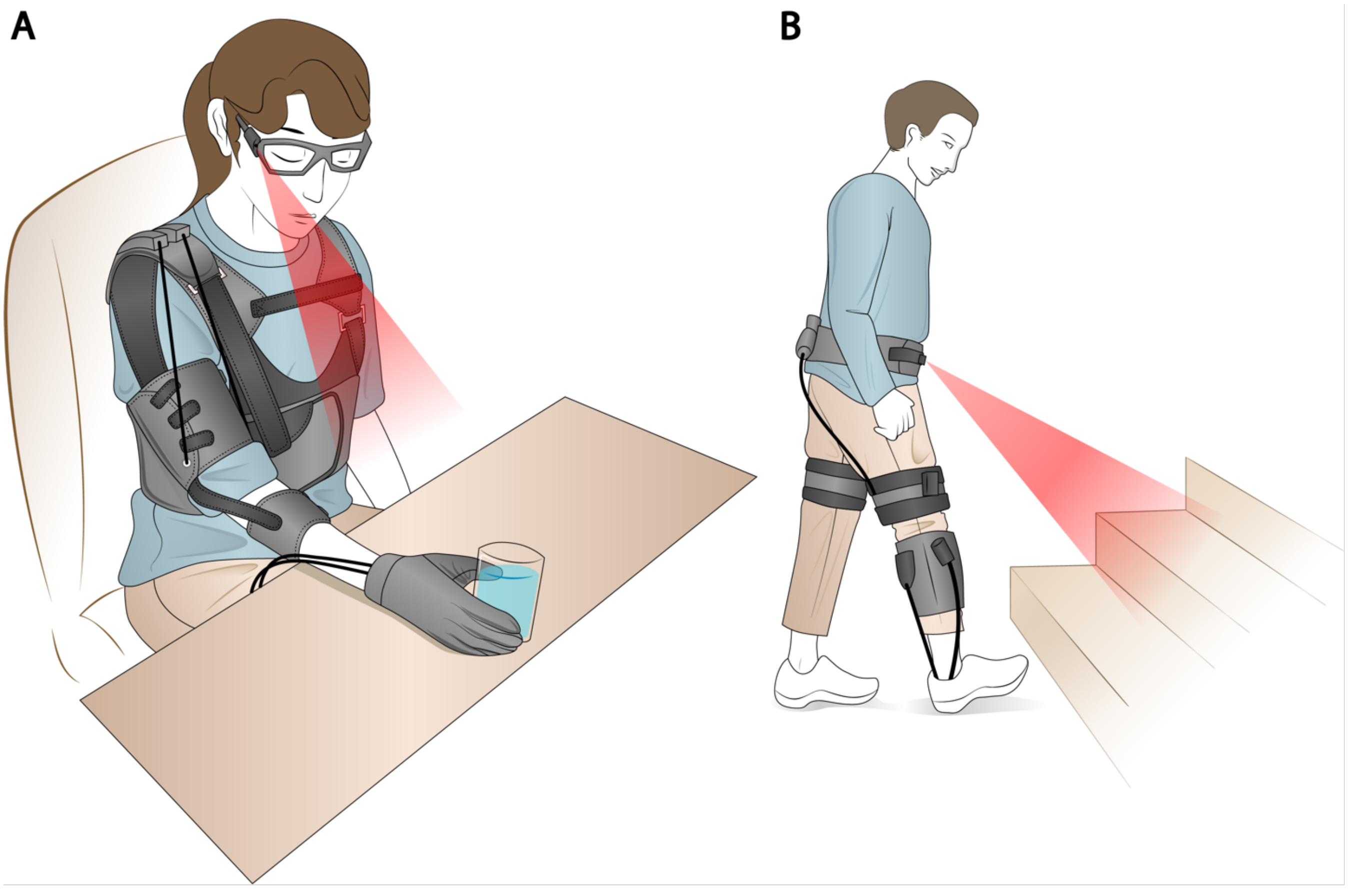

Fig. 1. **Examples of Environmental Awareness in Wearable Robots (A)**: In a vision-based grasp assistance system, the user might wear glasses with a camera and a robotic glove that augments grasp forces. The system can use machine learning-based image processing to classify the target object and infer the likely task the user wants to accomplish. In the example shown here, the system recognizes a full glass of water and infers that the user intends to take a drink. The system then selects a wrap grasp tailored to the size of the glass and closes the hand when vision indicates that the fingers surround the glass [10].

(B): A lower-limb assistance system can integrate wearable sensors and vision to expand the range of assistance that can be provided. In this representative example, a vision system detects a staircase in the user's path. The system uses inertial measurement units to detect heel strikes and estimates which footfall will be the first on a raised step. The wearable robot controller then triggers extra assistance torque to help raise the user's center of gravity, with precise timing of the assistance adjusted by electromyography signals indicating the user's leg muscle activation.

References

[1] UN World Health Organization (WHO), "World Report on Disability: Summary," 2011. WHO/ NMH/VIP/11.01.

[2] 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*, vol. 38, no. 3, pp. 1343–1362, 2021.

[3] D. P. Losey, C. G. McDonald, E. Battaglia, and M. K. O'Malley, "A review of intent detection, arbitration, and communication aspects of shared control for physical human–robot interaction," *Applied Mechanics Reviews*, vol. 70, no. 1, p. 010804, 2018.

[4] D. Farina, I. Vujaklija, R. Brånemark, A. M. Bull, H. Dietl, B. Graimann, L. J. Hargrove, K.-P. Hoffmann, H. Huang, T. Ingvarsson, *et al.*, "Toward higher-performance bionic limbs for wider clinical use," *Nature biomedical engineering*, vol. 7, no. 4, pp. 473–485, 2023.

[5] M. Yip, S. Salcudean, K. Goldberg, K. Althoefer, A. Menciassi, J. D. Opfermann, A. Krieger, K. Swaminathan, C. J. Walsh, H. Huang, *et al.*, "Artificial intelligence meets medical robotics," *Science*, vol. 381, no. 6654, pp. 141–146, 2023.

[6] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognition*, vol. 108, p. 107561, 2020.

[7] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[8] C. Shi, D. Yang, J. Zhao, and H. Liu, "Computer vision-based grasp pattern recognition with application to myoelectric control of dexterous hand prosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 9, pp. 2090–2099, 2020.

[9] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, *et al.*, "Event-based vision: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 1, pp. 154–180, 2020.

[10] D. Kim, B. B. Kang, K. B. Kim, H. Choi, J. Ha, K.-J. Cho, and S. Jo, "Eyes are faster than hands: A soft wearable robot learns user intention from the egocentric view," *Science Robotics*, vol. 4, no. 26, p. eaav2949, 2019.