ARTIFICIAL INTELLIGENCE

Eyes are faster than hands: A soft wearable robot learns user intention from the egocentric view

Daekyum Kim[1,2]*, Brian Byunghyun Kang[1,3,4]*, Kyu Bum Kim[1,3,4], Hyungmin Choi[1,3,4], Jeesoo Ha[1,2], Kyu-Jin Cho[1,3,4†], Sungho Jo[1,2†]

To perceive user intentions for wearable robots, we present a learning-based intention detection methodology using a first-person-view camera.

Intention dictates the future actions of a person: People attain their desired goals through a series of executed intentional actions, known as behaviors (*1*, *2*). However, in the case of people who suffer from quadriplegia (*3*, *4*), their behaviors are a poor reflection of their desired intentions. This is because their intentions are not delivered to their muscles, resulting in failure to perform expected actions. Spinal cord injury (SCI) prevents individuals from performing activities of daily living (ADL), such as holding a bottle. To assist them, wearable robots need to identify users' grasping or releasing intentions, allowing the users to properly manipulate the objects around them.

Studies have suggested the use of intention detection methodologies based on biosignal sensors and mechanical sensors. Intention detection methods using biosignal sensors, such as electroencephalography (EEG) (*5*, *6*) and electromyography (EMG) (*7*), generate intentions by detecting signals produced by the user's body parts. For methods using mechanical sensors—such as pressure sensors (*8*), bending sensors (*9*), and button switches (*10*)—intentions are created by user-generated signals from additional body motions. These intention detection methods identify user-generated signals that aim to trigger grasping or releasing motions. Because they look for trigger signals, these methods require calibrations to accurately identify the biosignals or the additional actions by the user.

This article presents a paradigm for learning-based intention detection methodologies that perceives user intentions for wearable hand robots by using a first-person-view (egocentric) camera. This is validated by verifying the following hypothesis: User intentions can be inferred through the collection of spatial and temporal information. Spatial information is characterized by the relationship between the hand and the target object. Temporal information is characterized as the history of user's arm behaviors from past observable moments to the present observed moment. The machine learning model used in this study, Vision-based Intention Detection network from an EgOcentric view (VIDEO-Net), was designed on the basis of this hypothesis, comprising spatial and temporal subnetworks as depicted in fig. S1. The suggested method is advantageous in that it detects user intentions without requiring any person-to-person calibrations and additional actions. In other words, the robot is able to interact with humans seamlessly.

The wearable hand robot used in this study is a modified version of Exo-Glove Poly, a tendon-driven soft wearable hand robot for SCI (*10*). The robot is composed of a glove, an actuation system, a first-person-view camera, and a computing device. Figure 1 depicts the system overview.

To verify the accordance of grasp and release intentions between the interpreted intentions and the user's intentions, the intentions obtained from the suggested method were compared with the intentions interpreted from the measured EMG signals. The activated intentions in our system led the intentions from the EMG signals for grasping and releasing by 0.3 and 0.8 s at most, respectively, as shown in fig. S2B. Figure S2C depicts the heat maps activated from the last convolutional layers during each task. The results show that identified grasping and releasing intentions, from the suggested method, followed a similar trend when compared to the output signals from the measured EMG signals.

Both healthy individuals and a patient with SCI performed practical experiments through a series of pick-and-place tasks. Healthy user performance was characterized by the average grasping, lifting, and release time of different objects using this model and a button as a reference measure, as shown in fig. S3. In the experiment with the SCI user, the user was instructed to reach toward a series of target objects without any additional actions (movie S1).

In conclusion, this paper suggests a learning-based paradigm that aims to detect user intentions based on the principle that people's intentions can be inferred through spatial and temporal information. Unlike existing intention detection methods where users command a robot to realize intentional actions, this paradigm detects user intentions based on user arm behaviors and hand-object interactions through obtained visual information. The experiments in this paper show that VIDEO-Net offers a practical solution for grasping and releasing various objects seamlessly without calibrations or additional actions.

VIDEO-Net can be limited in complex environments where the target object, the hand, and the arm behavior are not properly identified by the vision system. Multiple objects in the camera view and occlusions made by the objects will hinder the correct capturing of the spatial and the temporal information necessary to identify user intentions. This limitation can be overcome by incorporating other sensors to supplement the lacking information or by combining other intention detection methods. Sensors, such as bending sensors and inertial measurement units (IMUs), can supplement insufficient information about arm behaviors when the arm is not visible in the monocular camera scenes. An eye-tracking interface

[1]Soft Robotics Research Center, Seoul National University, Seoul 08826, Korea. [2]Neuro-Machine Augmented Intelligence Laboratory, School of Computing, KAIST, Daejeon 34141, Korea. [3]Biorobotics Laboratory, Department of Mechanical Engineering, Seoul National University, Seoul 08826, Korea. [4]Institute of Advanced Machines and Design, Seoul National University, Seoul 08826, Korea.
*These authors contributed equally to this work.
†Corresponding author. Email: kjcho@snu.ac.kr (K.-J.C.); shjo@kaist.ac.kr (S.J.)

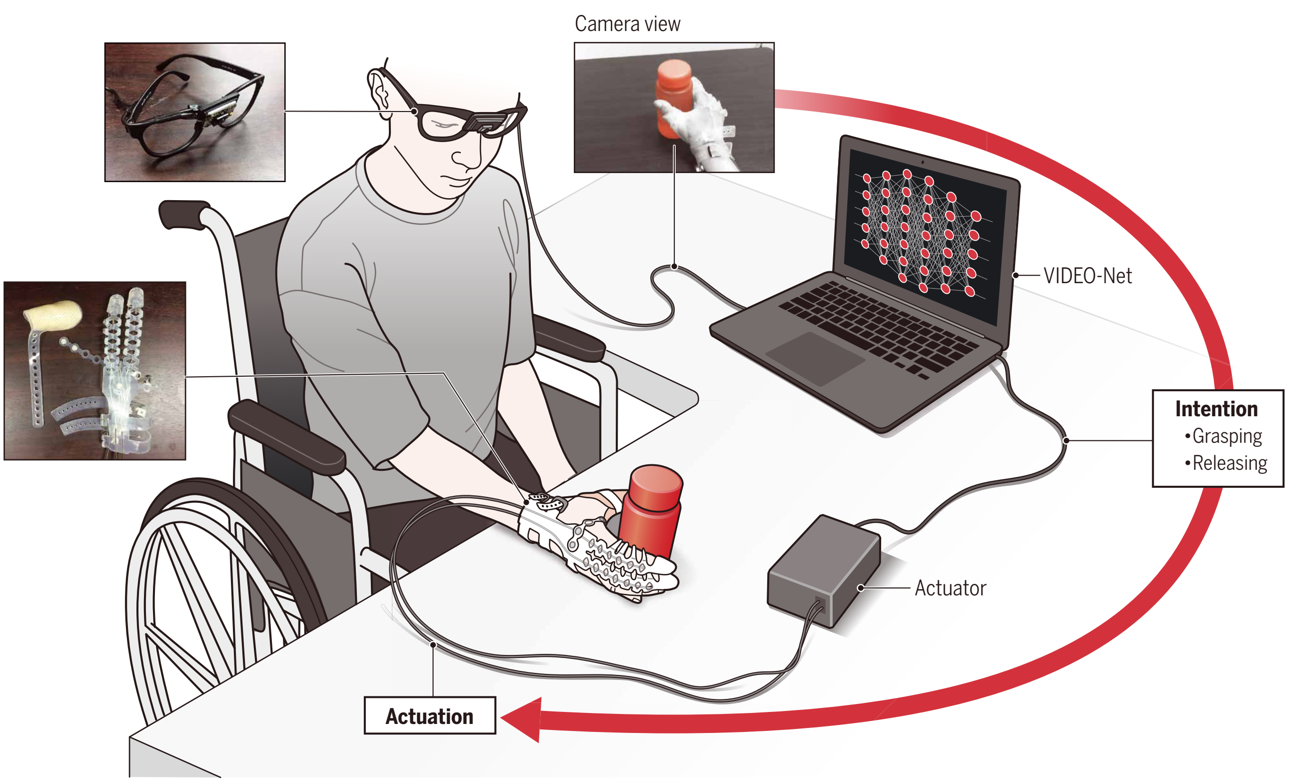

Fig. 1. System overview. Our system comprises the wearable hand robot with an actuation system and a computing device, which enumerates "intention" outputs from VIDEO-Net. Exo-Glove Poly is used for the wearable hand robot, and a first-person-view camera is mounted on the glasses. The glove is manipulated by tendon-driven actuators in the actuation system.

can be integrated to provide additional spatial information by distinguishing the target object from surrounding objects. A multimodal system that incorporates our system with other intention detection methodologies using EEG, EMG, and electrooculography (EOG) could greatly increase its performance of detecting user intentions.

The proposed paradigm could be extended to more complex user intentions, not limited to grasping and releasing, by incorporating information about the object, the environment, and the human-object relationship. This approach, using the spatial and temporal information of the environment and the human limbs, may be able to predict people's complex intentions more accurately and to overcome the limitations of other intention detection methods that capture intention directly from the user.

SUPPLEMENTARY MATERIALS

robotics.sciencemag.org/cgi/content/full/4/26/eaav2949/DC1
Section S1. Data processing
Section S2. EMG signal collection
Fig. S1. VIDEO-Net.
Fig. S2. Hypothesis verification through comparison with EMG signal.
Fig. S3. Pick-and-place results for individual objects.
Movie S1. Demonstrations for practical usage by a patient with SCI.
References (*11–13*)

REFERENCES AND NOTES

1. M. Bratman, *Intention, Plans, and Practical Reason* (Harvard Univ. Press, Cambridge, Massachusetts, 1987).
2. B. F. Malle, J. Knobe, The folk concept of intentionality. *J. Exp. Soc. Psychol.* **33**, 101–121 (1997).
3. G. J. Snoek, M. J. IJzerman, H. J. Hermens, D. Maxwell, F. Biering-Sorensen, Survey of the needs of patients with spinal cord injury: Impact and priority for improvement in hand function in tetraplegics. *Spinal Cord* **42**, 526–532 (2004).
4. A. Singh, L. Tetreault, S. Kalsi-Ryan, A. Nouri, M. G. Fehlings, Global prevalence and incidence of traumatic spinal cord injury. *Clin. Epidemiol.* **6**, 309–331 (2014).
5. S. R. Soekadar, M. Witkowski, C. Gómez, E. Opisso, J. Medina, M. Cortese, M. Cempini, M. C. Carrozza, L. G. Cohen, N. Birbaumer, N. Vitiello, Hybrid EEG/EOG-based brain/neural hand exoskeleton restores fully independent daily living activities after quadriplegia. *Sci. Robot.* **1**, eaag3296 (2016).
6. C. I. Penaloza, S. Nishio, BMI control of a third arm for multitasking. *Sci. Robot.* **3**, eaat1228 (2018).
7. K. O. Thielbar, K. Triandafilou, H. C. Fischer, J. M. O'Toole, M. L. Corrigan, J. M. Ochoa, M. E. Stoykov, D. G. Kamper, Benefits of using a voice and EMG-driven actuated glove to support occupational therapy for stroke survivors. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 297–305 (2017).
8. J. Huang, W. Huo, W. Xu, S. Mohammed, Y. Amirat, Control of upper-limb power-assist exoskeleton using a human-robot interface based on motion intention recognition. *IEEE Trans. Autom. Sci. Eng.* **12**, 1257–1270 (2015).
9. H. In, B. B. Kang, M. Sin, K.-J. Cho, Exo-Glove: A wearable robot for the hand with a soft tendon routing system. *IEEE Robot. Autom. Mag.* **22**, 97–105 (2015).
10. B. B. Kang, H. Lee, H. In, U. Jeong, J. Chung, K. J. Cho, Development of a polymer-based tendon-driven wearable robotic hand, in *Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA)* (2017), pp. 3750–3755.
11. J. J. Woods, B. Bigland-Ritchie, Linear and non-linear surface EMG/force relationships in human muscles. An anatomical/functional argument for the existence of both. *Am. J. Phys. Med.* **62**, 287–299 (1983).
12. T. Y. Fukuda, J. O. Echeimberg, J. E. Pompeu, P. R. G. Lucareli, S. Garbelotti, R. O. Gimenes, A. Apolinário, Root mean square value of the electromyographic signal in the isometric torque of the quadriceps, hamstrings and brachial biceps muscles in female subjects. *J. Appl. Res.* **10**, 32–39 (2010).
13. A. J. Fridlund, J. T. Cacioppo, Guidelines for human electromyographic research. *Psychophysiology* **23**, 567–589 (1986).

Acknowledgments: We thank C. W. Bundschu for revising the manuscript and J. J.-R. Song for working on graphical presentation of figures. **Funding**: This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) under grant NRF2016R1A5A1938472. **Author contributions**: D.K. and

B.B.K. contributed to the study design, interpretation of the data, and writing of the findings. D.K. designed and constructed associated software. B.B.K. designed and constructed the hardware. K.B.K., H.C., and J.H. performed the experiments and data analysis. K.-J.C. and S.J. directed the project and edited the manuscript. **Competing interests:** K.-J.C., S.J., B.B.K., D.K., H.C., K.B.K. are inventors on patent application (10-2018-0133652) submitted by Seoul National University and KAIST that covers intention detection methodology. J.H. has no competing interests. **Data and materials availability**: K.-J.C. and S.J. may be contacted for additional information.

10.1126/scirobotics.aav2949

Citation: D. Kim, B. B. Kang, K. B. Kim, H. Choi, J. Ha, K.-J. Cho, S. Jo, Eyes are faster than hands: A soft wearable robot learns user intention from the egocentric view. *Sci. Robot.* **4**, eaav2949 (2019).

Eyes are faster than hands: A soft wearable robot learns user intention from the egocentric view

Daekyum Kim, Brian Byunghyun Kang, Kyu Bum Kim, Hyungmin Choi, Jeesoo Ha, Kyu-Jin Cho and Sungho Jo

Sci. Robotics **4**, eaav2949.
DOI: 10.1126/scirobotics.aav2949