2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Visually-guided Grip Selection for Soft-Hand Exoskeleton

Xingying Chen†, Simone Löhlein†, John Nassour*, Stefan K. Ehrlich, Nicolas Berberich, and Gordon Cheng

***Abstract*— This paper presents a visually-guided grip selection based on the combination of object recognition and tactile feedback of a soft-hand exoskeleton intended for hand rehabilitation. A pre-trained neural network is used to recognize the object in front of the hand exoskeleton, which is then mapped to a suitable grip type. With the object cue, it actively assists users in performing different grip movements without calibration. In a pilot experiment, one healthy user completed four different grasp-and-move tasks repeatedly. All trials were completed within 25 seconds and only one out of 20 trials failed. This shows that automated movement training can be achieved by visual guidance even without biomedical sensors. In particular, in the private setting at home without clinical supervision, it is a powerful tool for repetitive training of daily-living activities.**

I. Introduction

The loss of hand functionality due to stroke and other neurological conditions hinders effective occupational performance and independent living ability. In addition to the rehabilitation, patients also need assistant in daily living tasks. Recently, research and applications at the intersection of robotics and neuroscience are gaining attention and progressively start making positive impacts on healthcare [2]. Using robots supports patients' movement and allows for more intense and repetitive training additionally to supervised therapy [1]. To enable more flexible use, the large and expensive machines can also be replaced by soft textile-based exoskeletons, which are particularly suitable for hand and finger therapy because of their lightweight and portability. To better assist the user, hand exoskeletons should be able to predict user's intention. Exoskeleton control through intention detection can be achieved through different techniques, such as force sensing [17], motion sensing [15], [10], electroencephalography (EEG) [3], and electromyography (EMG) [13]. However, neurotechnology based methods such as EEG and EMG are highly subject-dependent and need regular calibration, which is not feasible for independent training applications at home. Vision-based intention decoding, e.g. with a mobile eye tracker, shifts the source of information used for intention detection to the object. Using external motion cue is especially beneficial for patients with severe impairment since the EMG signals they produce are less distinguishable than those from patients with moderate impairment [6]. This paper introduces

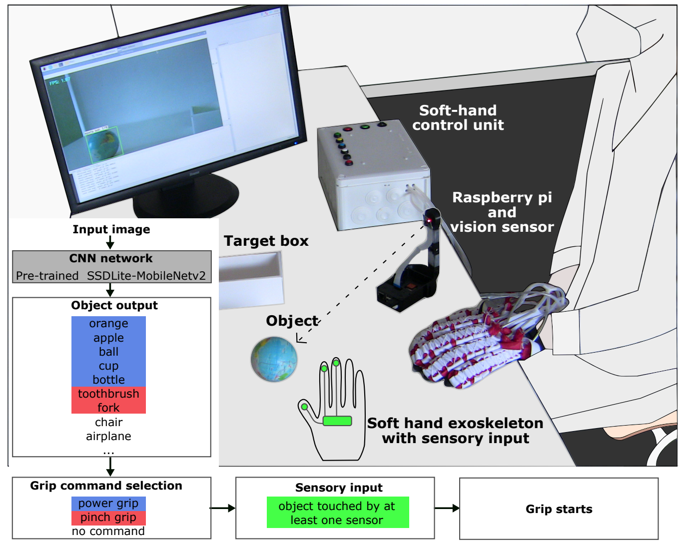

Fig. 1: An overview of the system setup. A camera is placed at $0.15\,\mathrm{m}$ height in front of the user to capture the object. A Raspberry Pi is connected to the camera, the screen, and the soft hand control unit. The sensory soft-hand exoskeleton is connected to the control unit. As an experiment setup, the user has to move the ball into the box.

a wearable soft-hand exoskeleton for stroke rehabilitation which uses visual object detection to select different object-dependent grip types that are essential for activities of daily living (ADL). Thereby, it allows users to perform additional training sessions without assistance and gives them back the feeling of independence in everyday life. Section II presents the soft-hand exoskeleton and the object recognition system. In Sec. III, we demonstrate the experiments and the evaluation of the object recognition system. The results are shown and discussed in Sec. IV, then the conclusion in Sec. V.

II. System description

A. Visually-guided intention recognition

In previous studies, the possibility of using video data to infer activities of daily living, for instance, recorded by wearable sensors embedded in conventional reading glasses, has been demonstrated by [16]. Particular challenges are posed by the image quality due to distance and motion blurring as well as the actual object recognition. In the setup chosen here the former is circumvented by the stationary rehabilitation setting in which the user is asked to perform various tasks while remaining seated. The latter is constrained in our experimental design by the selection of objects relevant to everyday life, such as targeted movement of objects, food

† Equal Contribution
*Corresponding author nassour@tum.de
Authors are with the Chair for Cognitive Systems, Department of Electrical and Computer Engineering, Technische Universität München, 80333 Munich, Germany. This work was supported by the Elite Master Program in Neuroengineering at the Technische Universität München, funded by the Elite Network Bavaria (ENB)

intake and daily hygiene, which are also common tasks in the evaluation and rehabilitation process [9], [5]. The complete pipeline is demonstrated in Figure 1.

B. Wearable soft hand exoskeleton

For our setup we used an exoskeleton with soft textile-based actuators proposed in [11], [12]. In this previous work, the potential of this soft exoskeleton to assist movements was validated by a reduced *flexor digitorum sublimis* muscle activity when the subject grasps and releases the object. On the exoskeleton, each finger is supported by a pair of flexor and extensor actuators. All actuators are housed and glued on a glove. The design of the housing determines the behavior of the inflatable textile tube. Four servo motors control the pneumatic inflation of the actuators by an air pump. These motors are controlled by an ESP32 microcontroller to actuate the extensor and flexor actuators of two groups of fingers. The first group is the thumb and the index fingers, while the second contains middle, ring and little fingers. Therefore, different grip types can be realized by activating a subset of the actuators. In order to start the grip movement at a correct time, we attached four force sensors to the thumb, the index finger, the middle finger, and the palm.

C. Implementation of grip movement

In Lee et al.'s work [6], grip movements are divided into pinch grip and power grip. The former is used for fine-grained interactions, the latter for heavier spherical objects. To perform pinch grip, both extension and flexor actuators for the thumb and the index finger, as well as the flexor actuators of the other three fingers, are activated. As the flexor actuator is stronger than the extensor actuator (design properties), activating both actuators (flexor/extensor) simultaneously for the thumb and index fingers results in closer fingertips, enabling the user to grasp small objects such as a toothbrush (Fig. 2 left). If only extensor actuators are activated, thumb and index fingers stretch along their own direction without touching each other. While the activation of only flexor actuators is more appropriate for power grasp. In power grasp, we exclusively activate the flexor actuators for all fingers to grasp a ball-shaped object (Fig. 2 right). After the system recognizes the intended grip movement, tactile sensory information from the force sensors on the hand exoskeleton's finger tips is used to trigger the motion. The force sensors' readings have a sampling rate of $400\,\mathrm{Hz}$. The moving average with a window length of 10 was taken to smooth the sensory input. After the motion is triggered by the sensory interaction with an object, air is pumped into the corresponding actuators for about $3.5\,\mathrm{s}$, then it is held for four seconds allowing the user to move his arm, and finally let out to release the grip.

III. Experiments

To evaluate the system's suitability for extra in-home training sessions, we first tested the object recognition system, as well as different parameter combinations and conducted a pilot experiment for a full study approved by the TU Munich

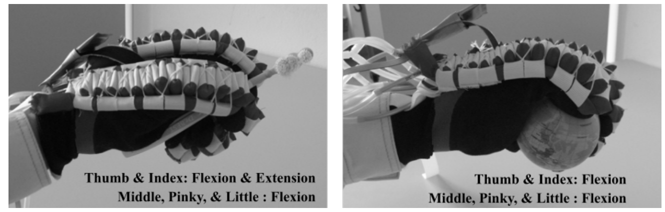

Fig. 2: Pinch grip (left) and power grip (right) with the soft exoskeleton.

institutional ethics review board under the reference number 167/21 S.

A. Object recognition system

To process the visual input, we connected a five-megapixel camera to a Raspberry Pi 3 (Model B+) with Raspbian GNU OS via a CSI-camera connection. A pre-trained Google SSDLite-MobileNet-v2 algorithm was used to classify objects [14]. The algorithm was trained on COCO dataset [7] which contains about 80 object types. We selected a subset of objects and mapped them to a grip command based on their shape. Each object was placed in front of the camera with distance $Dist$ representing the reaching area of a subject. The network outputs several objects with probabilities. The object prediction is adopted if its probability reaches a threshold $theta_{obj}$. To prevent sending a wrong grip command, we introduced another buffer with the last ten object detections. Only if a certain percentage $theta_{buffer}$ of them indicate the same object, the grip command is sent to the exoskeleton. To test the system behavior in response to variations of θ_{buffer}, θ_{obj} and $Dist$, we measured the following data: the accuracy of object recognition $AccRec$ and grip initiation $AccGrip$, the delay until the first correct recognition $DelayRec$ (positively influenced by /proportional to θ_{obj}), and the delay till sending the grip command $DelayGrip$ (positively influenced by to θ_{buffer}). The experiment was restricted to two commodity items (apple and knife) that cover both grip types. Each object was recorded three times for $30\,\mathrm{s}$. Given the parameters from the previous experiment, we included more objects to test the recognition accuracy and recorded each three times for $30\,\mathrm{s}$. Here we selected objects relevant to daily tasks: apple, bottle, cup (power grip), toothbrush, and knife (pinch grip). Other objects in the dataset do not have a corresponding grip command.

B. Rehabilitation exercises and ADL tasks

Our experiment scenarios were adapted from the Task-Specific Training [4] and the evaluation tests of motor function [8], [9]. We defined two rehabilitation tasks (Rehab) and two ADL tasks. In Rehab 1 and 2, the subject is asked to grasp a hard and a soft ball and drop it into the target box placed $0.3\,\mathrm{m}$ right to the ball. In ADL 1, the subject needed to pick up a toothbrush and circle it around the target plate three times before releasing it. In ADL 2, the subject needed to pick up the bottle and pretend to pour water into a cup. Both the plate and the cup were $0.3\,\mathrm{m}$ right to the toothbrush or bottle. These tasks were chosen

TABLE I: Evaluation of recognition performance: apple and knife.

Object	Apple	Knife	Apple	Knife	Apple	Knife
Dist [cm]	30	30	50	50	30	30
θ_{buffer} *[%]*	60	60	60	60	30	30
AccRec [%]	100 $\pm$0	61.1 $\pm$47.1	95.6 $\pm$6.1	0	100 $\pm$0	53.5 $\pm$8.7
AccGrip[%]	100 $\pm$0	100 $\pm$0	100 $\pm$0	0	100 $\pm$0	66.7 $\pm$47.1
DelayRec [s]	0.5 $\pm$0	0.5 $\pm$0	0.85 $\pm$0.47	∞	0.5 $\pm$0	1.8 $\pm$1.89
DelayGrip [s]	4.17 $\pm$0.47	4.0 $\pm$0.5	4.5 $\pm$0.82	∞	2.5 $\pm$0	7.5 $\pm$2.83

because repeating grasp and release movements can exercise hand muscles, and interaction with these objects is common in daily life. The experiment is organized in subsequent trials. At the beginning of each trial, the subject wearing the exoskeleton placed the hand in a relaxed position with the wrist at the table edge. The object recognition algorithm was started at t_1. The command for the grip type was sent to the exoskeleton at t_2, and the subject started to reach the object placed at $0.3\,\mathrm{m}$ in front of the resting position of the left arm. The subject touched the object at t_3 and the sensory information triggered the actuation of the exoskeleton at t_4. Finally, the subject performed the required task, and the trial ended when the object was released at t_5. The opening of the exoskeleton happens automatically after a given time frame. For each task, the subject performed five trials. We denoted a trial as successful if the object did not fall during the trial. To evaluate the performance, we calculated delay to send grip command ($T_1 = t_2$-t_1), reaching duration ($T_2 = t_3$-t_2), delay in motion ($T_3 = t_4$-t_3), and task completion time ($T_4 = t_5$-t_1). These metrics are relevant for evaluating the system because the delays influence user experience and the total time of a single trial should be acceptable to repetitively conduct the exercise.

IV. Results and Discussion

A. Object Recognition

Table I shows three different combinations of θ_{buffer}, and $Dist$ for apple and knife. θ_{obj} is set to its default value 50% because reducing it lead to more wrong detections. $AccRec$ depends on the object shape and the distance. Apple has a higher $AccRec$ than knife. Smaller elongated objects, such as the knife were no longer detected in a distance of $0.5\,\mathrm{m}$ due to the low image quality of the camera. Increasing θ_{buffer} from 30% to 60% increased $AccGrip$ from 0.67 to 1 for the knife. In principle, increasing θ_{buffer} incurs larger $DelayGrip$. However, this is not observed with the knife. Also, the changing image quality over time leads to the variation in $DelayRec$ and $DelayGrip$. The trade-off between $DelayGrip$ and $AccGrip$ can be circumvented by increasing the computational power to process the input more efficiently to reduce the delay between sending each object prediction.

In the remaining experiments, we use the following combination θ_{obj}: 50%, θ_{buffer}: 60%, $Dist$: $0.3\,\mathrm{m}$. Figure 3 shows the confusion matrix for each category. For the power

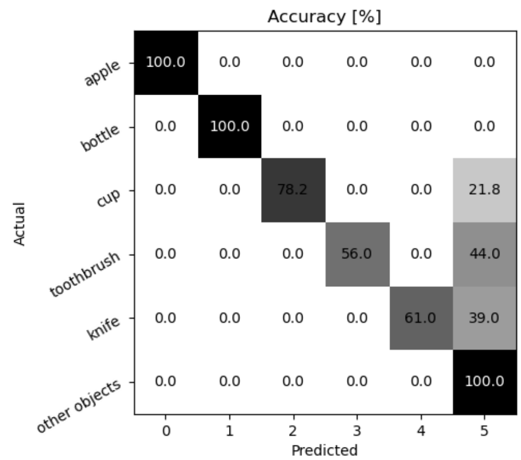

Fig. 3: Confusion matrix of object recognition system. ”Unknown”: The object (game control) is not in the training data. ”Other objects” refers to objects not associated with a certain grip type.

grip objects with a high spherical volume, the accuracy exceeds 75%. For the elongated pinch grip objects, the system reached a performance of 56-61%. Furthermore, all false detections belonged to ”other objects” and does not trigger a wrong grip command.

B. Task performance

Here we present the experiment results of a healthy male subject in Figure 4 and 5.

***Time to send grip command** (T_1)* is roughly 4 to 5 s for all four objects where 3 s are predetermined by θ_{buffer} at a rate of only 2 fps due to the computational cost of running the model on a Raspberry Pi. Additional factors such as sensitivity to orientation were excluded in the scope of the experiments. Although this delay is relatively high and might be perceived as unpleasant to the user, it ensured that no wrong command was sent.

***Time to reach object** (T_2)* has a smaller variance compared to T_3. There is no significant difference between tasks, which also corresponds to the fact that the distance of the movement is the same in all tasks. As stated previously, the input area can easily be increased with a higher resolution input. For flexibility in practical use, it is envisaged to replace the camera with a wearable glasses.

***Time to trigger motion** (T_3)* shows no significant difference between the power and pinch grip (p=0.30). However, it is influenced by the material of the object. The mean delay is much larger for the soft compared to the hard ball. This could be explained by the difficulty to reach the threshold of the force sensor to trigger the motion if the object yields. Since a smaller relative threshold of sensory input increases the risk of random initiation, these delays can not be completely avoided.

***Task completion time** (T_4)* Figure 5 shows that roughly 60% of the trials were completed within $20\,\mathrm{s}$ and 80% finished within $22\,\mathrm{s}$. Even with maximum delay due to problems triggering the movement, a maximum time of $25\,\mathrm{s}$ was not exceeded, which corresponds to a difference of 9 s compared to the ideal case. Since the focus of the study is the grasping, we used a fixed value for the duration until the grip release, but it could later be replaced by other algorithms.

Trial success Only during the experiment with the toothbrush, the subject dropped the object in one out of five trials.

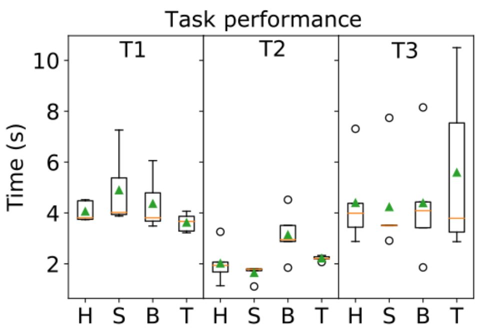

Fig. 4: Evaluation of task performance with hard ball “H”, soft ball “S”, bottle “B”, and toothbrush ”T”. T1, T2, and T3 are the delay in sending grip command, the time to reach object, and the delay in triggering motion respectively. Median: orange; mean: green; outlier: dot.

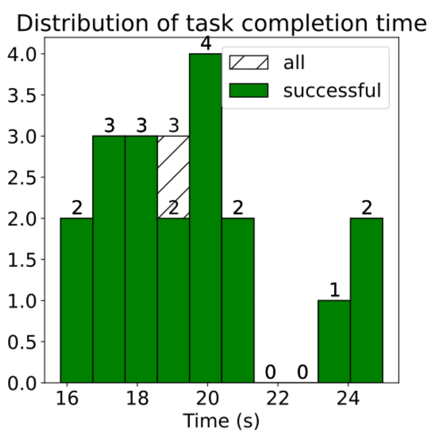

Fig. 5: Time to complete a trial. Green: successful trial. Dashed: All trials.

V. Conclusion

In this work, we presented a visually-guided soft-hand exoskeleton to assist user’s grasping movement. A pilot performed four grasp-and-move tasks repeatedly and completed 19 out of 20 successfully, without dropping the object. The delay we observed in dynamics was caused by object detection and the limited air flow of the pump, which can be avoided in future by optimizing the algorithms and improving the hardware. The device offers a promising option for an independent add-on training additional to physio-therapeutical sessions. It is important to emphasize that our system can assign different grip types to objected regardless of the level of impairment and does not require additional calibration since we dispense the usage of biomedical sensors. Moreover, it is possible to interact directly with different objects of daily life, and thus achieving an independence that represents an additional motivation for the user. In the future, the trade-off between accuracy and delay can be adjusted according to the user’s preference and the suggestion of therapists.

References

[1] M. Babaiasl, S.H. Mahdioun, P. Jaryani, and M. Yazdani. A review of technological and clinical aspects of robot-aided rehabilitation of upper-extremity after stroke. *Disability and Rehabilitation: Assistive Technology*, 11(4):263–280, 2016.

[2] G. Cheng, S.K. Ehrlich, M. Lebedev, and M.AL. Nicolelis. Neuroengineering challenges of fusing robotics and neuroscience. *Science Robotics*, 5(49):7–10.

[3] N. Cheng, K.S. Phua, H.S. Lai, P.K. Tam, K.Y. Tang, K.K. Cheng, et al. Brain-computer interface-based soft robotic glove rehabilitation for stroke. *IEEE Transactions on Biomedical Engineering*, 67(12):3339–3351, 2020.

[4] C-Y. Chu and R.M. Patterson. Soft robotic devices for hand rehabilitation and assistance: a narrative review. *Journal of neuroengineering and rehabilitation*, 15(1):1–14, 2018.

[5] H.S. Jørgensen, H. Nakayama, H.O. Raaschou, J. Vive-Larsen, M. Støier, and T.S. Olsen. Outcome and time course of recovery in stroke. part i: Outcome. the copenhagen stroke study. *Archives of physical medicine and rehabilitation*, 76(5):399–405, 1995.

[6] S.W. Lee, K.M. Wilson, B.A. Lock, and D.G. Kamper. Subject-specific myoelectric pattern classification of functional hand movements for stroke survivors. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(5):558–566, 2010.

[7] T-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C.L. Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[8] F.I. Mahoney and D.W. Barthel. Functional evaluation: the barthel index: a simple index of independence useful in scoring improvement in the rehabilitation of the chronically ill. *Maryland state medical journal*, 1965.

[9] D.M. Morris, G. Uswatte, J.E. Crago, E.W. Cook, and E. Taub. The reliability of the wolf motor function test for assessing upper extremity function after stroke. *Archives of physical medicine and rehabilitation*, 82(6):750–755, 2001.

[10] J. Nassour, H.G. Amirabadi, S. Weheabby, A.A. Ali, H. Lang, and F. Hamker. A robust data-driven soft sensory glove for human hand motions identification and replication. *IEEE Sensors Journal*, 20(21):12972–12979, 2020.

[11] J. Nassour and F. Hamker. Enfolded textile actuator for soft wearable robots. In *2019 IEEE International Conference on Cyborg and Bionic Systems*, pages 60–65, 2019.

[12] J. Nassour, F. Hamker, and G. Cheng. High-performance perpendicularly-enfolded-textile actuators for soft wearable robots: Design and realization. *IEEE Transactions on Medical Robotics and Bionics*, 2(3):309–319, 2020.

[13] P. Polygerinos, K.C. Galloway, S. Sanan, M. Herman, and C.J. Walsh. Emg controlled soft robotic glove for assistance during activities of daily living. In *2015 IEEE international conference on rehabilitation robotics (ICORR)*, pages 55–60. IEEE, 2015.

[14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and LC. Chen. Mobilenetv2: The next generation of on-device computer vision networks. In *CVPR*, 2018.

[15] P. Weber, E Rueckert, R. Calandra, J. Peters, and P. Beckerle. A low-cost sensor glove with vibrotactile feedback and multiple finger joint and hand motion sensing for human-robot interaction. In *IEEE International Symposium on Robot and Human Interactive Communication*, pages 99–104, 2016.

[16] K. Zhan, S. Faux, and F. Ramos. Multi-scale conditional random fields for first-person activity recognition on elders and disabled patients. 16(PB), 2015.

[17] YM. Zhou, D. Wagner, K. Nuckols, R. Heimgartner, C. Correia, M. Clarke, D. Orzel, C. O’Neill, R. Solinsky, S. Paganoni, et al. Soft robotic glove with integrated sensing for intuitive grasping assistance post spinal cord injury. In *2019 International conference on robotics and automation (ICRA)*, pages 9059–9065. IEEE, 2019.