AI-based Semiautonomous Control Strategy for upper-limb prostheses

Gianmarco Cirelli[1,*], Christian Tamantini[1,2,*], Luigi Pietro Cordella[3] and Francesca Cordella[1]

[1]*Unit of Advanced Robotics and Human-Centered Technologies, Università Campus Bio-Medico di Roma, Via Alvaro del Portillo 21, Rome, Italy*

[2]*Institute of Cognitive Sciences and Technologies, National Research Council of Italy, Via Giandomenico Romagnosi 18a, Rome, Italy.*

[3]*University of Naples Federico II, Via Claudio 21, 80125 Naples, Italy*

Abstract

Artificial Intelligence-based Semiautonomous control strategies (AI-based SCS) could significantly improve the reliability and naturalness of prosthetic hand control.

The integration of a computer vision system (CVS) and the user motion intention allows the prosthetic hand to autonomously recognize the object to be grasped and to select the appropriate hand posture, with the user in charge of initiating the execution of the grasp.

In this work, an AI-based SCS integrating EMG signals with a CVS is presented. The control strategy was assessed both in laboratory settings and in more complex real life scenarios. The results show strong performance of the proposed SCS with an Accuracy in Object and Grasp Classification above 97%, a Mean Time of Execution of 0.483s, and a Mean Angular Error and Estimation Stability in wrist orientation of $16.26 \pm 8.62°$ and 0.2, respectively. The control strategy effectively managed also complex situations, achieving a 100% Success Rate in the conducted tests.

Keywords

Upper Limb Prosthesis, Artificial Intelligence, Computer Vision, Semiautonomous Control Strategy

1. Introduction

The loss of an upper limb greatly affects an individual quality of life, and existing prosthetic devices that rely solely on electromyographic (EMG) signals have limitations in terms of adaptability and control [1, 2]. Traditional threshold-based EMG control is sequential and lacks intuitiveness, allowing users to manage only a limited number of hand gestures, while muscle pattern recognition techniques are often inconsistent [3], leading to usability issues and increased abandonment of prostheses. Semiautonomous control strategies (SCS) based on Artificial Intelligence (AI) [4] have been proposed to alleviate cognitive demands and improve the accuracy, and intuitiveness of prosthetic control. More specifically, over the past decade, the use of visual information from a Computer Vision System (CVS) to assist prosthesis users in selecting the most appropriate configuration has gained increasing interest within the scientific community [5].

Current methods in the literature use a combination of RGB cameras and ultrasonic sensors to classify hand gestures and determine wrist orientation [6], or they incorporate RGB-D cameras [7]. However, the size and weight of these setups hinder their practical integration into prosthetic devices. To optimize grasp selection and reduce computational overhead, convolutional neural networks (CNNs) have been applied to classify RGB images [8], but this often comes at the expense of wrist orientation estimation.

This study introduces a novel AI-based SCS that takes into account both the user intended motion and the configuration of the hand and wrist simultaneously [9]. Unlike previous approaches where the

11th Italian Workshop on Artificial Intelligence and Robotics (AIRO 2024)

*Corresponding author.

gianmarco.cirelli@unicampus.it (G. Cirelli); christian.tamantini@cnr.it (C. Tamantini); cordel@unina.it (L. P. Cordella); f.cordella@unicampus.it (F. Cordella)

0009-0001-4409-9167 (G. Cirelli); 0000-0001-6238-2241 (C. Tamantini); 0000-0002-6946-0377 (F. Cordella)

system autonomously determines the hand gesture, this method allows the user to actively select the grasp while the vision system adjusts wrist orientation and offers correction suggestions.

2. Materials and Methods

Figure 1 provides an overview of the proposed approach. The SCS processes a RGB image acquired by the CVS embedded in the prosthesis, i.e. a Single-Board Computer (SBC) and a miniaturized RGB camera, and the user intended hand gesture obtained from a Human Machine Interface (HMI), i.e. electromyographic sensors. The grasping classes commonly used in commercial prostheses (—*Power*, *Lateral*, *Precision*, and *Pointing*—along with a *Rest*) are obtained by applying an EMG classifier to the EMG signals measured by the HMI. These classes are simulated in this work. By combining these inputs, the system accurately estimates wrist orientation and optimizes hand preshaping based on the object property.

The proposed SCS consists of three main steps: Object Detection, Grasp Selection, and Orientation Estimation. This method operates continuously, providing also visual feedback to the user through a pair of LEDs. The green one reveals successful object detection, while the red one is briefly activated when there is a mismatch between the user intended motion and the vision-based SCS. In this last case, the implemented SCS is capable of selecting the correct hand gesture thanks to the information of the object recognized in the scene. This real-time feedback helps the user to identify errors, allowing them to adjust the EMG output and thus improving user confidence. Each component of the SCS is detailed in the following.

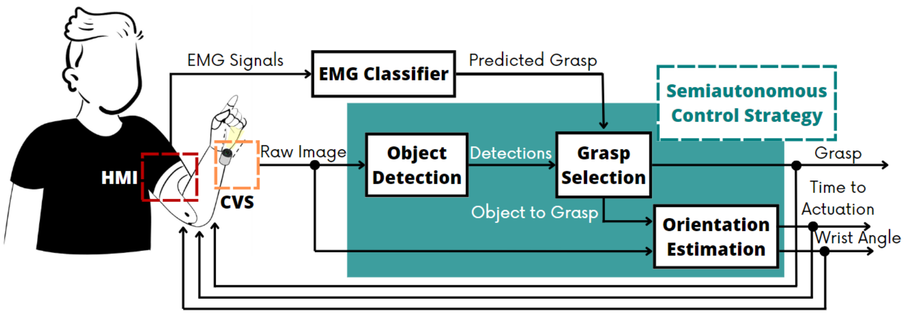

Figure 1: Block diagram of the system. The HMI and the CVS are in the red and orange box, respectively.

2.1. Semiautonomous Control Strategy (SCS)

The image of the scene captured by the CVS is used as input for the Object Detection module. This module identifies objects in the scene and associates each with a bounding box (x and y coordinates of the center, width, and height of the bounding box), a class label, and a confidence score, representing the probability that the object has been correctly classified. This information is structured in a$(N, 6)$ tensor, where N is the number of detected objects and 6 represents the calculated information. Objects are sorted by descending confidence scores.

Among the several CNN models in the literature [10], YOLOv5 (You Only Look Once) [11, 12] in its small version was selected for object detection due to its balance between classification accuracy and processing speed, making it ideal for real-time applications. The model was trained on the Microsoft COCO dataset [13], which contains a wide range of objects in various contexts. The model was trained on the platform Ultralytics [14], with the following parameters: 100 epochs of training with an automatically optimized batch size of 32 and an image size of 640×640, using the Stochastic gradient descent (SGD) optimizer (learning rate 0.01, momentum 0.9) and automatic mixed precision enabled. Data augmentation techniques such as Blur, MedianBlur, Grayscale conversion, and CLAHE (Contrast Limited Adaptive Histogram Equalization) were applied to enhance image contrast and improve model performance. It should be noted that the model was not fine-tuned on a bespoke data set of images of the objects included in the experimental validation of the proposed approach. Instead, the version available online was used for these analyses.

The vision-based control strategy allows for the assignment of object-specific grasp configurations, starting from the macro-categories of grasp detected by the EMG classifier. In particular the hand gestures that the proposed SCS can handle are as follows: *Prismatic*, *Thumb Adducted*, *Thumb Abducted*, *Index Finger Extension*, *Fixed Hook*, *Spherical*, *2-Digits*, *3-Digits*, *Lateral*, *Pointing*, and *Rest* for the hand, *Pronation* and *Supination* for the wrist. This approach ensures high classification performance by limiting the number of EMG classes while enhancing the prosthesis ability to perform specific grasps based on the detected objects.

The information from the Object Detection module is then fed into the Grasp Selection module, where it is compared with the hand gestures predicted by the EMG classifier. The module then handles the following cases: **No objects detected**, in which the control algorithm returns a "Rest" gesture. **Non-coherence**, where the visual data and EMG output do not match. In this case, priority is given to the visual data, selecting the gesture for the object with the highest confidence score. **Coherence**, in which only objects associated with the grasp recognized by the EMG classifier are considered, and priority is given to those in the central part of the image. Moreover, if multiple objects are detected in the central region of the image, only the one with the highest confidence level is selected, because it is assumed that the user targets the camera towards the object that wants to grab [15].

The Orientation Estimation module continuously calculates the wrist P/S angle by segmenting the Region of Interest (ROI) of the selected object and applying Principal Component Analysis (PCA). Focusing on the ROI improves segmentation accuracy by isolating the target object and reducing computational load. The ROI is converted to grayscale and segmented using Otsu's thresholding method[16], followed by morphological operations of closure to remove residual noise.

Contours are detected, and only those within 5–95% of the ROI area are considered for PCA so that it is only applied to the segmented object. PCA is used to determine the wrist P/S angle, which is calculated as the angle between the first Principal Component (PC) of the object and the $\vec{x}$-axis of the image [17]. For objects without a principal axis, like spheres, a default P/S angle of 90° is used. Special objects, such as a mouse, require a different approach where the first PC is aligned with the finger longitudinal axis.

The proposed SCS recognizes the reaching phase when a 50% increase in the object ROI is detected and then triggers the prosthesis to execute the selected hand gesture.

2.2. Experimental Evaluation

The vision-based control algorithm ran on a Raspberry Pi 4 Model B, a SBC which is suitable for embedded applications. Its desktop was remotely managed via Virtual Network Computing (VNC) and powered by a portable USB-C power bank (5V DC, 3A). The Arducam 16MP High-Resolution camera was chosen as an RGB camera for its 16 MP resolution and autofocus. To achieve a good compromise between the computational burden of the algorithm and performance, a 320×240 pixel image resolution was chosen [18, 19].

For evaluating orientation estimation, the Vicon VERO system, a leading motion capture solution, was employed. It consists of eight compact infrared cameras with 2.2 MP resolution, set to acquire data at 100 Hz. The cameras are strategically positioned to track reflective markers on objects, allowing for accurate 3D motion reconstruction against a predefined reference frame.

The Experimental Protocol is composed of two sequential sessions, which are detailed in the following.

The first session focused on evaluating the performance of the proposed control algorithm, in accurately classifying various objects presented individually, and its computational cost in terms of time needed to execute the control pipeline. The camera was positioned above the objects at a predefined angle to simulate real-world conditions, and 16 objects were tested, corresponding to the chosen hand gestures: Lateral (fork and spoon), Pointing (keyboard and mouse), Spherical (sports ball), 2-digits Precision (book and cup), 3-digits Precision (scissors and wine glass), Prismatic (cell phone and remote), Thumb Adducted (bottle), Thumb Abducted (umbrella), Index Finger Extension (knife), and Fixed Hook (backpack and suitcase). Each object was recorded in 5 trials across different configurations, capturing 25 frames per trial for a total of 125 samples. Additionally, two markers were positioned on each object to provide orientation data (except for ball), along with three markers on the camera module to define the image plane. In each test, the object was rotated by a predetermined angular displacement to ensure comparability across different objects. The capacity of the proposed control algorithm to estimate wrist orientation was assessed by comparing its calculated angular displacement with data from the Vicon system. Marker positions were processed in MATLAB to compute the camera plane and the object Principal Component.

Key performance indicators (KPIs) were extracted from the first experimental sessions to evaluate the proposed approach quantitatively. More in detail, the computed KPI were:

- Accuracy in Object and Grasp Classification (*AOC*,*AGC*): it is the correspondence between the real object (or hand gesture) and the predicted one.
- Time to Execute the Control pipeline (TEC): it quantifies, in seconds, the computational load of the algorithm.
- Mean Angular Error (MAE): it assesses the performance of the proposed SCS in correctly estimating angular displacements, where the single Angular Error (AE) is defined as:

$$AE = |\Delta\theta_V - \Delta\theta_P| \,, \tag{1}$$

  where $\Delta\theta_V$ and $\Delta\theta_P$ are the angular displacement computed by Vicon VERO and the SCS, respectively.
- Estimation Stability (*ES*): it is evaluated as the standard deviations of angles computed by the proposed algorithm and the Vicon VERO data. Higher total standard deviation values per setup indicate lower stability

In the second session of the Experimental Protocol, the system was tested in a more complex and real scenario, to evaluate the ability of the algorithm to align user intentions with the vision system, both when there is a discrepancy between the vision system and EMG output, and when multiple object are framed. Hence, the Success Rate (*SR*) was considered to evaluate the system in this phase the number of trials in which a correct grasp is associated with the scene according to the simulated user intention.

The experimental setup in this part of the Experimental Protocol is shown in Figure 2.

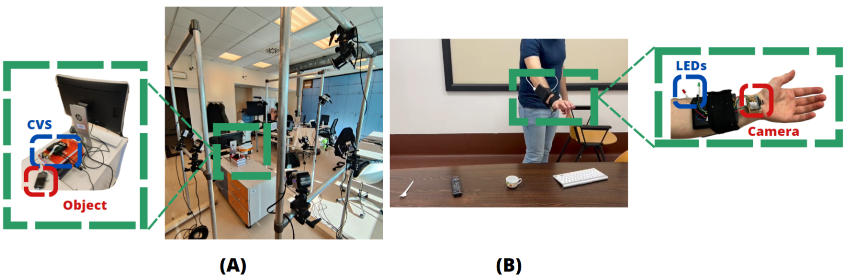

Figure 2: Experimental setup in the first (A) and in the second (B) session of the Experimental Protocol.

3. Results and Discussion

Figure 3 shows the output from each of the modules of the implemented SCS. As illustrated, the system can correctly recognize the object, segment it, and estimate its orientation in the image plane.

Figure 4 reports results obtained in the first session of the experimental protocol for *AOC*, *AGC* and TEC. Specifically, the system reached 97.98% of accuracy in correctly classifying objects, and 99.81% for the grasp. It was found that $AGC \geq AOC$, meaning that objects linked to a specific hand gesture

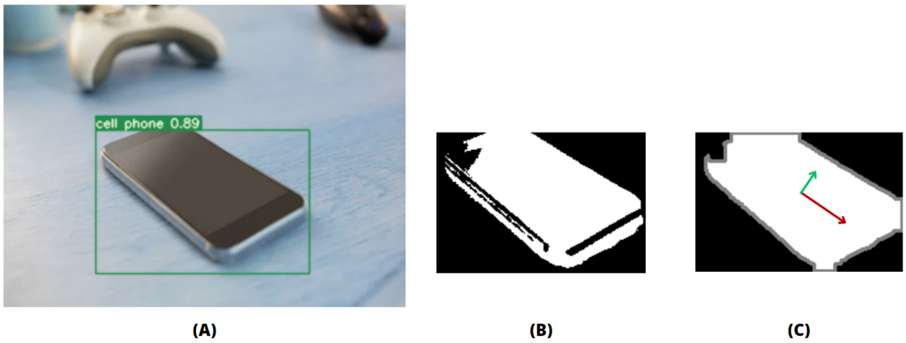

Figure 3: Outputs of the single parts of the image processing pipeline: Object Detection(A) and Segmentation(B), and Morphological Operation of Closure and Orientation Estimation(C).

category are sometimes misclassified with those that can be manipulated using the same gesture. This suggests that the proposed approach is not significantly impacted by object misclassification.

Regarding the TEC, it was found that the average time required to execute the control pipeline was $0.483 \pm 0.169s$. Moreover, Figure 4(C) underlines the differences among single parts of the proposed SCS: the Object Detection module is the one that most heavily impacts the algorithm speed, as evidenced by the three orders of magnitude difference between the times recorded for the object detection module and the others. Considering the percentage of each single step over the total time to execute the control pipeline, the Detection, the Selection, the Segmentation, and the Orientation take 97.55%, 0.05%, 1.1%, 1.30%, respectively.

Concerning the MAE and ES, the proposed SCS grounded on CVS provides promising results, with an average angular of $16.26 \pm 8.62°$ and 0.2, respectively. The obtained results seems to be acceptable for the proposed application, since the small angular error can be easily compensated by the other arm joints.

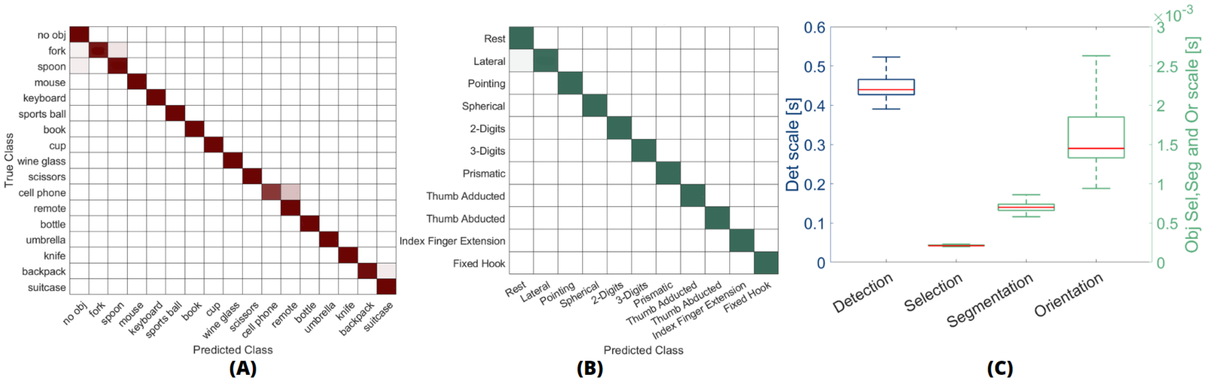

Figure 4: Confusion Matrices of the Accuracy in Object Classification (A) and Grasp Classification (B). Box plot of the time needed by the SBC to execute different parts of the proposed algorithm: Detection, Selection, Segmentation of the object, and Orientation Estimation (C).

In the second session, the ability of the system to respond to two complex scenarios was tested. Specifically, Figure 5 shows frames acquired by the CVS during these two tests. In the first one, multiple objects were framed, and as the user intent changed, the proposed SCS was able to select always the one associated with that macro-category of grasps. Instead, in the second test, the input coming from the EMG classifier was constant, while the framed object changed. In this case, the system was capable of correcting the output by considering the visual information.

Figure 5: Frames acquired by the CVS during the two tests of the Second Phase of the Experimental Protocol.

4. Conclusions

In this study, an AI-based SCS for hand-wrist prostheses grounded on CVS was developed and tested. It combines a convolutional neural network (CNN)-based object detection module, a grasp selection module, and an automatic thresholding algorithm for grasp selection and wrist orientation estimation. The YOLO v5s was chosen as Deep Learning model for detection, and it was trained on COCO dataset. By integrating external visual data from the CVS with user intent through simulated EMG signals, the system aims to improve prosthesis control. The results demonstrate high accuracy in object and grasp classification (over 97%), with an average Time to Execute the Control pipeline of $0.483s$. The system enables the identification of additional grasps beyond those detected by EMG, ensuring the appropriate grasp for different objects. The Mean Angular Error and Estimation Stability in wrist orientation were recorded as $16.26 \pm 8.62^{\circ}$ and 0.2, respectively. While this error may be deemed acceptable, it will be validated on a real prosthetic system.

The strategy successfully performed the tests in the final phase, proving to effectively manages complex situations. Moreover, the system is designed for portability and interoperability, making it easily applicable to various prosthetic hands and robotic grippers capable of replicating a variety of gestures.

Future efforts should be devoted to integrating the proposed CVS into a prosthetic device, and testing the user-friendliness, accuracy, and effectiveness of the overall system in reducing the cognitive workload on a population of users.

Acknowledgments

This work is funded partly by the European Union - Next Generation EU - NRRP M4.C2 - Investment 1.5 Establishing and strengthening of Innovation Ecosystems for sustainability (Project n. ECS00000024 Rome Technopole) and partly by the Italian Ministry of Research, under the complementary actions to the NRRP "Fit4MedRob - Fit for Medical Robotics" Grant PNC0000007, (CUP: B53C22006990001).

References

[1] E. Stefanelli, F. Cordella, C. Gentile, L. Zollo, Hand prosthesis sensorimotor control inspired by the human somatosensory system, Robotics 12 (2023) 136.

[2] E. Stefanelli, M. Lapresa, F. Cordella, D. D'Accolti, C. Cipriani, L. Zollo, A hand-wrist control strategy based on human upper limb kinematics, in: 2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob), IEEE, 2024, pp. 1029–1034.

[3] D. Yadav, K. Veer, Recent trends and challenges of surface electromyography in prosthetic applications, Biomed. Eng. Letters (2023) 1–21.

[4] S. Pancholi, J. P. Wachs, B. S. Duerstock, Use of artificial intelligence techniques to assist individuals with physical disabilities, Annual Review of Biomedical Engineering 26 (2024).

[5] L. Gionfrida, D. Kim, D. Scaramuzza, D. Farina, R. D. Howe, Wearable robots for the real world need vision, Science Robotics 9 (2024) eadj8812.

[6] S. Došen, C. Cipriani, M. Kostić, M. Controzzi, M. C. Carrozza, D. B. Popović, Cognitive vision system for control of dexterous prosthetic hands: experimental evaluation, Journal of neuroengineering and rehabilitation 7 (2010) 1–14.

[7] M. N. Castro, S. Dosen, Continuous semi-autonomous prosthesis control using a depth sensor on the hand, Front. in Neurorobotics 16 (2022) (2022).

[8] M. C. F. Castro, W. C. Pinheiro, G. Rigolin, A hybrid 3d printed hand prosthesis prototype based on semg and a fully embedded computer vision system, Front. in Neurorobotics 15 (2022) 751282.

[9] G. Cirelli, C. Tamantini, L. P. Cordella, F. Cordella, A semiautonomous control strategy based on computer vision for a hand–wrist prosthesis, Robotics 12 (2023) 152.

[10] A. A. Boshlyakov, A. S. Ermakov, Development of a vision system for an intelligent robotic hand prosthesis using neural network technology, in: ITM Web of Conf., volume 35, EDP Sciences, 2020, p. 04006.

[11] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Conf. on CVPR, 2016.

[12] M. Phadtare, V. Choudhari, R. Pedram, S. Vartak, Comparison between yolo and ssd mobile net for object detection in a surveillance drone, Int. J. Sci. Res. Eng. Manag 5 (2021) 1–5.

[13] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco: Common objects in context, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision – ECCV 2014, Springer International Publishing, 2014, pp. 740–755.

[14] G. Jocher, ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements, https://github.com/ultralytics/yolov5, 2020. URL: https://doi.org/10.5281/zenodo.4154370. doi:10.5281/zenodo.4154370.

[15] J. R. Flanagan, Y. Terao, R. S. Johansson, Gaze behavior when reaching to remembered targets, J. of neurophysiol. 100 (2008) 1533–1543.

[16] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Syst., Man, and Cybern. 9 (1979) 62–66.

[17] C. Tamantini, M. Lapresa, F. Cordella, F. Scotto di Luzio, C. Lauretti, L. Zollo, A robot-aided rehabilitation platform for occupational therapy with real objects, in: Converging Clin. and Eng. Res. on Neurorehabilit. IV: 5th ICNR2020, 2020, Springer, 2022, pp. 851–855.

[18] S. Dosen, C. Cipriani, M. Kostić, M. Controzzi, M. C. Carrozza, D. Popović, Cognitive vision system for control of dexterous prosthetic hands: Experimental evaluation, J. of neuroeng. and rehabilit. 7 (2010) 42.

[19] S. Došen, D. B. Popović, Transradial prosthesis: artificial vision for control of prehension, Artif. organs 35 (2011) 37–48.