robotics

MDPI

Article

A Semiautonomous Control Strategy Based on Computer Vision for a Hand–Wrist Prosthesis

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

[1] Research Unit of Advanced Robotics and Human-Centred Technologies, Universitá Campus Bio-Medico di Roma, 00128 Rome, Italy; gianmarco.cirelli@alcampus.it (G.C.); c.tamantini@unicampus.it (C.T.)
[2] Universitá di Napoli Federico II, 80125 Naples, Italy; cordel@unina.it
* Correspondence: f.cordella@unicampus.it

Abstract: Alleviating the burden on amputees in terms of high-level control of their prosthetic devices is an open research challenge. EMG-based intention detection presents some limitations due to movement artifacts, fatigue, and stability. The integration of exteroceptive sensing can provide a valuable solution to overcome such limitations. In this paper, a novel semiautonomous control system (SCS) for wrist–hand prostheses using a computer vision system (CVS) is proposed and validated. The SCS integrates object detection, grasp selection, and wrist orientation estimation algorithms. By combining CVS with a simulated EMG-based intention detection module, the SCS guarantees reliable prosthesis control. Results show high accuracy in grasping and object classification (≥97%) at a fast frame analysis frequency (2.07 FPS). The SCS achieves an average angular estimation error ≤18° and stability ≤0.8° for the proposed application. Operative tests demonstrate the capabilities of the proposed approach to handle complex real-world scenarios and pave the way for future implementation on a real prosthetic device.

Keywords: hand–wrist prosthesis; artificial vision; semiautonomous control

check for updates

Citation: Cirelli, G.; Tamantini, C.; Cordella, L.P.; Cordella, F. A Semiautonomous Control Strategy Based on Computer Vision for a Hand–Wrist Prosthesis. *Robotics* **2023**, *12*, 152. https://doi.org/10.3390/robotics12060152

Academic Editor: Claudio Loconsole

Received: 9 October 2023
Revised: 3 November 2023
Accepted: 11 November 2023
Published: 13 November 2023

1. Introduction

The loss of an upper limb has profound physical, psychological, and social consequences [1]. It significantly impacts an individual's ability to perform activities of daily living and work-related tasks [2], limiting their independence and overall quality of life. Despite advancements in prosthetic technology, statistics show that less than half of amputees (44.7%) use their prosthetic device for more than eight hours per day. Additionally, the majority of amputees (76.9%) use their prostheses for cosmetic purposes rather than for functional ones. This highlights the challenges faced by amputees, where functional limitations and discomfort still influence their decision to utilize the prosthesis [3,4].

To address these issues, alleviating the burden on amputees regarding the control of prostheses is crucial. Hence, researchers are actively developing new strategies to intuitively control multiple degrees of freedom (DoFs) [5].

Traditionally, myoelectric prostheses rely on recognizing user motion intention through electromyographic (EMG) signals [6]. Various strategies based on pattern recognition [7] have been proposed to classify different muscle activation patterns, improving prosthesis control performance with respect to simple threshold-based methods. However, sequential control of different DoFs is still mainly adopted, limiting the naturalness, intuitiveness, reliability, and performance of the prosthesis control [8]. More recent and sophisticated approaches enable parallel control of multiple DoFs to naturally control prosthetic devices [9]. Nevertheless, their performance, as that of all the EMG-based systems, is heavily dependent on a trade-off between the number of possible outputs, i.e., the number of possible classes, and the system robustness: the higher the number of classes, the lower the performance and robustness of the system. Specifically, the classification accuracy drops to 90–95% when the number of motion classes is increased by more than 10, compared to the initial accuracy of 99% achieved when only four classes are considered [10]. This effect can be observed as a limitation arising from both the utilized algorithms, as well as from the inherent difficulty faced by upper-limb amputees in generating precise and consistent contractions. They also exhibit noticeable performance degradation over time, which can be attributed to various factors such as repositioning of the prosthetic socket, sweating, and muscle fatigue resulting from prolonged use or physical exertion [11]. These inaccuracies in detecting user intentions may significantly affect the performance and reliability of the prosthesis control.

To ensure significant improvements in the accuracy, naturalness, and intuitiveness of prosthesis control, and therefore to reduce users' cognitive burden, Semiautonomous Control Strategies (SCS) [12] were introduced to combine user inputs with automated or intelligent systems to control the prosthetic device [13]. Specifically, Computer Vision systems (CVS) can be exploited to predict the grasping configuration from the geometric properties of the framed object, allowing the accomplishment of the task.

The first CVS-based SCS did not consider any user input but only relied on the information retrieved from exteroceptive sensors. For instance, in [14], an RGB camera with an ultrasonic sensor was mounted on the back of a prosthetic hand to control its motion. Geometrical features of the object were extracted from the gray-scale image (320 × 240 pixels) acquired by the camera, using a threshold segmentation method. Wrist rotation, grasp type, and hand opening were determined based on visual and distance information [15]. More specifically, four types of grasp were considered, while the size of the aperture was determined based on the estimated size of the object. The control was run on a standard personal computer, using a DAQ board for communication with sensors. Another approach relies on an RGB-D camera for preshaping the robotic hand and estimating grasp type, size, and wrist orientation [16]. However, in both approaches, the cameras were not integrated into the hand–wrist prosthesis due to their weight and size. Moreover, this approach is based on the estimation of the desired prosthesis configuration by considering the properties of the segmented object, retrieved from the fitting model which well-approximates its shape. Specifically, only three geometrical models were taken into account, i.e., three different grasp types: sphere, cylinder, and cuboid.

Further approaches introduced EMG signals to activate the movement of the prosthesis, and left the identification of the gestures to be performed to the exteroceptive sensors.

In [17], EMG signals were used only to trigger the grasp phase, while Convolutional Neural Networks (CNNs) [18] were trained to categorize the objects framed by an RGB camera into four different hand gestures. All offline and real-time tests were implemented on a personal computer, leaving the integration of the system onto prostheses among future developments. In [19], a similar approach was proposed, where EMG signals recorded from two electrodes were used only to initialize the control strategy and to confirm hand gestures identified by the CNN-based control. Furthermore, the prosthesis user can receive feedback and decide whether to reset, adjust, or continue the movement.

The CNN-based approaches proposed in the literature present an important limitation: a shift towards object detection strategies results in relinquishing the autonomous and fine control of prosthetic hand orientation, i.e., pronation/supination (P/S) rotation, resorting to coarse control solutions or leaving the manual management of the prosthetic hand orientation to the user, significantly impacting user-side control.

Recent studies explored the use of a multimodal system with EMG and CVS in prosthetic control [20,21], proposing a strategy for combining these two types of information to improve the accuracy with respect to an EMG classifier. The hand gesture determined based on information gathered from an environmental CVS was used as an additional feature alongside those extracted from EMG signals to infer the final grasp to be executed [22], neglecting wrist orientation. From the state-of-the-art analysis, the need for a novel SCS that manages both hand and wrist configuration by simultaneously taking into account the user's motion intention is therefore evident. It will foster the inclusion of the user in the prosthesis control loop enhancing the prosthesis embodiment and the naturalness of the control. The objective of this paper is to fill this gap by proposing a novel SCS based on a CVS embedded in a wrist–hand prosthesis and integrating user motion intention, derived from a simulated EMG classifier, and exteroceptive information about the type and orientation of the object to grasp.

The proposed approach compares the output of the EMG classifier with that of the vision system to obtain the optimal grasp to be executed and the corresponding wrist orientation. More specifically, it starts with a limited set of grasp classes, identified by the EMG classifier, to determine object-specific sub-classes, ensuring the appropriate grasp execution. In the vision-based control strategies currently proposed in the literature, the user may feel estranged from controlling the prosthesis, because he/she only triggers the control via EMG sensors, while the grasp to be performed is autonomously determined by the control strategy. The proposed approach addresses this limitation, since the user can decide the grasp to be performed through EMG signals, and the SCS is responsible for determining the most appropriate wrist orientation and for confirming the recognized motion intention or suggesting corrections. More specifically, the EMG classifier is responsible for determining four grasp macro-categories, and the vision-based control strategy enables the assignment of grasp configurations belonging to that macro-category, but object-specific, considering object geometrical characteristics [23]. Furthermore, the implemented control strategy has the capability to handle real-life scenarios, enhancing the overall usability of the system in practical situations.

The proposed approach was validated in two different experimental settings. The first one is a structured environment to quantify the computational burden of the algorithm and its performance in accurately detecting the framed objects and their orientation. The outcome of this phase also guided the choice of the vision system's characteristics, i.e., the camera model to be used and the image resolution. The second experimental setting reproduces a realistic operative scenario. Once the optimal CVS positioning on the hand–wrist prosthesis was determined, the camera was mounted on the corresponding location of a human subject hand to assess the algorithm's capability for handling coherence between the user intention and the vision system, and for properly identifying approaching conditions with the object to be grasped.

The paper is structured as follows: Section 2 details the proposed approach and describes the methodological steps used for validation. Section 3 presents and discusses the results obtained in the different experiments. Lastly, Section 4 draws the conclusion of this work and provides possible future works.

2. Materials and Methods

An overview of the proposed approach is shown in Figure 1. The vision-based hand–wrist control strategy (blue box in Figure 1) takes as input an RGB image of the environment the hand has to interact with, captured from the CVS integrated into the prosthesis, and the user's intention, obtained from an EMG classifier, in terms of desired hand gesture. The combination of these two pieces of information makes it possible to simultaneously cope with coherent and non-coherent situations and accurately estimate wrist orientation to optimize hand preshaping, i.e., to output the hand gesture to be performed by the prosthesis and the wrist orientation (in terms of P/S angles) most appropriate to the object's geometric properties. Specifically, the proposed SCS is composed of 3 sequential processing steps: Object Detection, Grasp Selection, and Orientation Estimation. Visual feedback, based on the output of the control algorithm, is provided to the prosthesis user through two colored LEDs to report whether an object has been detected in the scene and any non-coherence with the output from the EMG classifier. Each block shown in Figure 1 is detailed in the following.

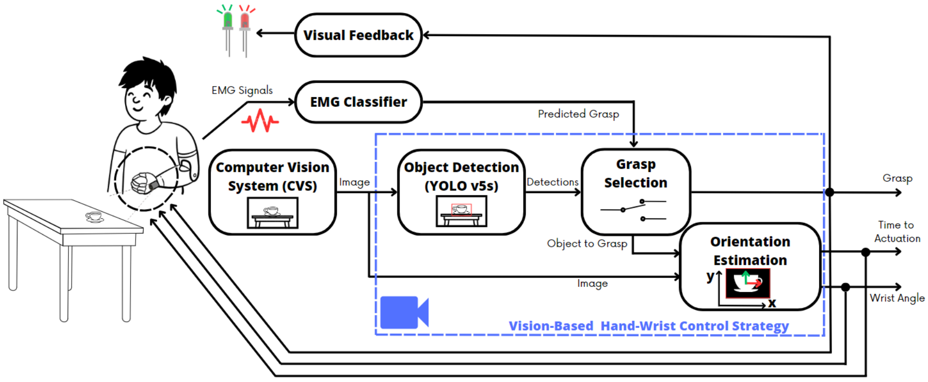


Figure 1. Block scheme of the proposed approach: the control strategy is outlined by the blue box.

2.1. Proposed Approach

2.1.1. EMG Signal Classifier

User motion intention about the grasp to be performed is retrieved by the EMG signal classifier. It is a simulated classifier that, considering the hand gestures commonly performed by commercial prostheses, is able to classify 4 grasp macro-categories in addition to the *Rest*, i.e., *Power*, *Lateral*, *Precision*, and *Pointing*.

2.1.2. Computer Vision System (CVS)

A CVS was introduced to recognize the objects to be grasped and their properties, as explained in Section 2.1.3, with the aim of estimating the most appropriate hand and wrist configuration. The system must be such that it can be integrated into one prosthetic hand. Therefore, the camera sensor and the electronic components should be miniaturized in order not to add weight and bulk to the prosthetic hand. To achieve this goal, a single-board computer (SBC) has been exploited to host the proposed algorithm and perform preliminary tests.

2.1.3. Vision-Based Hand–Wrist Control Strategy

The image of the scene retrieved by the CVS is the input for the Object Detection module. It is responsible for detecting any objects in the scene and associating them with a bounding box (x-coordinate and y-coordinate, width and height of the bounding box), a class label, and a classification confidence level, i.e., the probability the object was correctly classified, organized into an $(N, 6)$ tensor, where N is the number of objects detected and 6 is the computed information. The objects are sorted with descending confidence levels.

To detect the object in the scene, among the several CNN models in the literature, such as VGG16, ResNet50, YOLO, and MobileNet [24], YOLOv5 (You Only Look Once) [25,26] in its small version was chosen. Compared to other models, YOLOv5 provides a favorable trade-off between classification accuracy and processing speed, making it highly suitable for real-time object detection applications. The model was trained on the Microsoft COCO (Common Objects in Context, [27]) dataset, a large-scale image recognition dataset for object detection and segmentation.

Starting from the macro-categories of grasp that can be obtained from the EMG classifier, the vision-based control strategy enables the assignment of grasp configurations belonging to that macro-class, but object-specific, identified on the basis of the Feix taxonomy [28]. Thus, the EMG classifier maintains a high classification performance, as the number of considered classes is limited, while, at the same time, the grasping capabilities of the prosthesis are enhanced, as the implemented control strategy takes into account object-specific grasping sub-classes. Figure 2 shows the list of gesture sub-categories considered in the proposed control strategy and the objects associated with them.

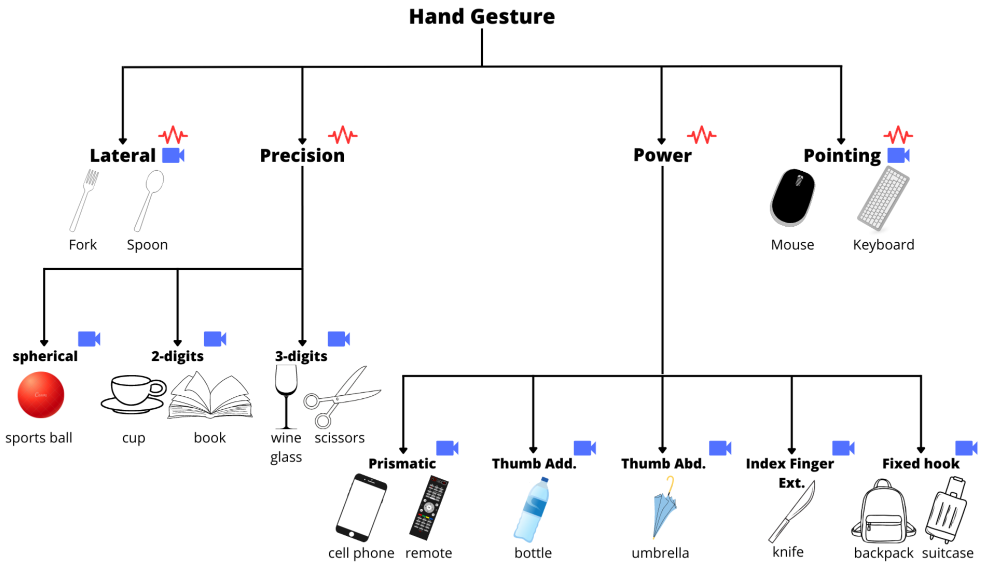


Figure 2. Diagram of all the hand gestures considered and objects associated with them: the symbols indicate the grasps the EMG classifier (red) and the developed SCS (blue) are able to recognize, respectively.

This information is then provided as input to the Grasp Selection module, which compares the hand gesture, obtained from the EMG classifier, with the one estimated by the vision system. The module identifies coherence or non-coherence between the two proposed grasps and the object that the user is most likely intending to grasp. Specifically, a decisional tree is proposed to autonomously respond to all cases that may occur in the scene:

- *If* there are no objects detected in the scene, *then* the control algorithm returns the *Rest* gesture. This information is given back to the user via visual feedback, as explained in Section 2.1.4.
- *If* there is non-coherence between the output from the EMG classifier and the hand gestures associated with the detected objects, *then* priority is given to the visual information. It means that the output of the Grasp Selection module is the hand gesture corresponding to the object with the highest confidence level. The correspondence between the object and the grasp is shown in Figure 2.
- *If* there is coherence between the output of the EMG classifier and the grasp associated with an object in the scene, *then* the object is taken into account, and the other objects not belonging to the class recognized by the EMG classifier are removed from the list of detections. In this case, the number of objects is variable and depends on how many of them can be grasped with the hand gesture obtained from the EMG classifier. In particular, *If* there are multiple objects that can be grasped with the class obtained from the user EMG classifier, *then* only objects framed in the central portion of the scene are considered. It is assumed that the object the user wants to grab is framed in the central region of the image [29], thus the one towards which the prosthetic hand will move. Even in this case, the number of objects is variable and depends on how many of them are in the central portion of the acquired image. *If* there are multiple objects in the central portion of the image, *then* the one that was recognized with a higher confidence level is selected.

The Orientation Estimation module is responsible for continuously estimating the wrist P/S angle by segmenting the Region of Interest (ROI) of the selected object and by applying Principal Component Analysis (PCA).

Selecting only the ROI is crucial, since it improves segmentation accuracy by focusing solely on the target object, eliminating noise areas in the image, and it reduces computational burden by analyzing only the bounding box, thereby increasing analysis speed.

The ROI is transformed from RGB to gray-scale and subjected to threshold segmentation using Otzu's method [30]. It allows for determining the optimal threshold by analyzing the intensity histogram to separate the foreground, i.e., the selected object, from the background. The resulting binary image undergoes the closure morphological operation involving dilation and erosion procedures to remove residual noise in the binary mask.

Contours are detected after segmenting the image, and only contours within the 5–95% area range of the ROI are considered to ensure a single region for the PCA. A method based on PCA has been chosen to retrieve the optimal wrist P/S angle, since PCA reduces the dataset's dimensionality while preserving maximum information and variance. PCA is specifically designed to capture the directions along which the data exhibit the highest variance, making it a valuable method for identifying the principal direction of variability within a segmented region, i.e., the detected object. The proposed method relies on some constraints the camera placement should satisfy: the image plane must be perpendicular to the wrist P/S rotation axis and the y-axis of the image reference system must be aligned with the longitudinal axis of the long fingers, while the x-axis must be perpendicular to it. If the aforementioned constraints are met, the P/S angle α the hand–wrist prosthesis has to reach is determined as the angle between the first Principal Component (PC) of the segmented region in the ROI and the x-axis of the image plane [31]. In this configuration, the first PC is perpendicular to the long finger's longitudinal axis, enabling grasping along the object's smallest dimension. However, some objects require a different procedure. For example, the ones without a principal axis, like spheres, cannot have their wrist orientation estimated. In this case, the optimal P/S angle is the one that allows the palm of the prosthetic hand to be parallel to the plane on which the object is located, which is an angle of 90°. For objects like the mouse, where the pointing hand configuration is required, the first PC should be parallel to the long finger's longitudinal axis, resulting in a rotation angle that complements α.

The approach works continuously and is capable of recognizing the reaching phase. Once the reaching occurs, the hand–wrist prosthesis shapes itself in the configuration estimated by the proposed approach. Specifically, the ROI area is tracked among subsequent frames and, if a 50% area increase is computed with respect to the initial value, the hand gesture obtained with the proposed strategy is actuated.

2.1.4. Visual Feedback

The developed system includes a feedback module that provides visual feedback to the prosthetic user about the detection of an object in the scene. Specifically, it consists of a pair of LEDs, one green and one red, connected to the SBC. The green one indicates the successful detection of an object in the scene. The red one is turned on for a fixed time whenever a mismatch between the user motion intention, obtained by the EMG classifier, and the gesture class, computed by the vision-based hand–wrist control strategy, occurs. This real-time visual feedback aims at improving user confidence, facilitating error detection and enabling effective control of the prosthetic device. More specifically, with the green LED, the user is informed that at least one object in the scene has been recognized by the control algorithm, whereas the red LED indicates a non-coherent condition that might lead to issues in the interaction with the object. This approach allows the user to opt for repeating the muscle contraction to correct the output from the EMG classifier, rather than proceeding with the grasping task solely relying on the information from the CVS.

2.2. Experimental Validation

Extensive tests of the proposed approach performance have been carried out. The experimental setup and protocol, and the key performance indicators used for the system validation are detailed in the following.

2.2.1. Experimental Setup

The vision-based control algorithm was executed on the Raspberry Pi 4 Model B, which was chosen as a single-board computer (SBC) due to its compact size (88 × 58 × 19.5 mm) and computing power, making it an ideal choice for small-scale system development. Moreover, the 64-bit OS Raspbian provides support for Python, the high-level programming language selected for writing the control algorithm. The SBC desktop was remotely controlled via virtual network computing (VNC) and a power bank was used as an energy supply (5V DC and 3A via a USB-C port) to make the system portable.

To develop the CVS, the following requirements were considered for the camera selection: *(i)* integration capability into a hand–wrist prosthesis, which imposes the use of a miniaturized RGB camera; *(ii)* high native resolution to ensure optimal algorithm performance; *(iii)* compatibility with the selected SBC for communication purposes; *(iv)* cost-effectiveness to not raise the prosthesis cost. After evaluating commercial options, two cameras were selected: the Arducam 16MP High-Resolution camera, with 16 MP resolution and autofocus capabilities, and the Joy-it Wide Angle camera, with 160° wide field of view. The technical specifications of the two cameras are reported in Table 1.

Table 1. Cameras technical specifications.

Camera Model	Arducam 16 MP	Joy-It Wide Angle
Sensor	Sony IMX519	OV5647
Resolution [pixels] (Static Images)	16 MP (4656 × 3496)	5 MP (2952 × 1944)
Resolution [pixels] (Video)	1080p30, 720p60, 640 × 480p90	1080p30, 960p 45, 720p60, 640 × 480p90
Field of View(FoV)	80°	160°
Autofocus	✓	N.A.
Dimension [mm]	25 × 23.86 × 9	25 × 24 × 18
Weights [g]	3	5
Cost	33.95 €	20.17 €

Figure 3 shows the scheme of the connections that characterize the CVS, with both the identified cameras.

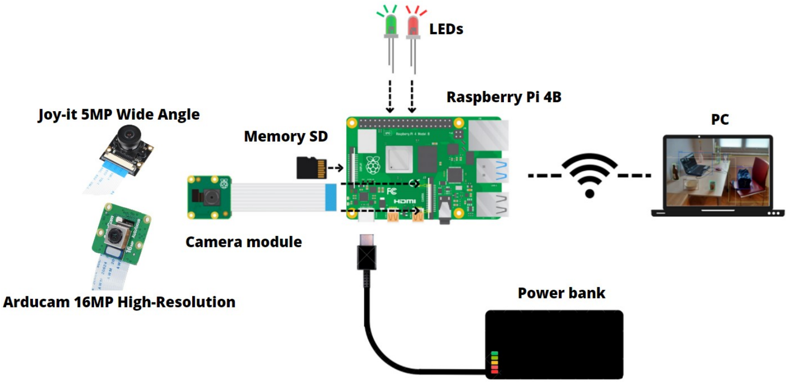


Figure 3. Connections scheme of the CVS.

The PiCamera2 and OpenCV libraries in Python were used to obtain images from the vision sensors. PiCamera2 facilitates communication with the camera and parameter configuration, while OpenCV is used to perform the image processing pipeline described in Section 2.1.3. Specifically, both cameras are configured at their best native resolution, which means an initial frequency of 10 FPS and 15.63 FPS, respectively.

In order to ensure a good compromise between speed and performance, the algorithm was tested by taking as input different image resolutions, i.e., 640 × 480 pixels [32,33], which are commonly used and the largest, and 320 × 240 pixels [14,15], which are widely employed and provide a 50% reduction in size compared to the former option. In this way, it was possible to quantify whether the input image size could affect computational burden and performance.

The Vicon VERO system was introduced for evaluating the orientation estimation of the proposed control strategy. It is an optoelectronic system gold standard in motion capture applications, capable of tracking the movement of objects and/or subjects in a 3D environment; specifically, Vicon is composed of 8 cameras called VERO; they are compact and super-wide infrared cameras with a resolution of 2.2 MP. The acquisition frequency was set to 100 Hz. Cameras are placed in the environment to optimally frame the scene and record the movement of reflective markers attached to the monitored objects. After the acquisition, raw data are used for kinematic reconstruction, giving back the 3D markers' position with respect to a reference frame defined at the beginning.

2.2.2. Experimental Protocol

To evaluate the performance of the proposed control strategy, the experimental protocol was divided into three sessions called Classification Performance, Orientation Estimation Validation, and Operative Tests.

The first session of the experimental protocol aimed at characterizing the performance of the proposed control algorithm, specifically of the Object Detection module, in correctly classifying the various objects shown individually, considering different setup configurations, i.e., various combinations of camera and input image resolution. The environment was structured with the camera framing the objects from above, not perfectly parallel to the image plane: in fact, the camera was held at a pre-defined angle to the plane on which the objects were standing, simulating a real context. In this part of the protocol, 16 objects were taken into account, divided among the 10 considered hand gestures: fork and spoon for Lateral; keyboard and mouse for Pointing; sports ball for Spherical (Precision); book and cup for 2-digits (Precision); scissors and wine glass for 3-digits (Precision); cell phone and remote for Prismatic (Power); bottle for Thumb Adducted (Power); umbrella for Thumb Abducted (Power); knife for Index Finger Extension (Power); backpack and suitcase for Fixed Hook (Power). Five trials were performed for each object per setup configuration, and in each of them, the object was framed for 25 frames for a total of 125 samples. To also verify if the position of the objects with respect to the camera influences the algorithm performance in associating the grasp class to the framed object, the objects were moved within the workspace to predefined positions.

The second session aimed at characterizing the performance of the algorithm in wrist orientation estimation and evaluating its computational burden, i.e., of the Orientation Estimation module, for each of the aforementioned setup configurations. Even in this case, the environment was structured as shown in Figure 4a and the camera was located in a fixed position and framed the scene in which the objects were placed one by one. Moreover, 2 markers were placed on each of the 8 considered objects, 2 per grasp macro-category (shown in Figure 4b) along their largest dimension, to have information on both object orientation, and 3 on the camera module to obtain the image plane.

For each object, 5 trials were performed for each configuration. In each test, the object was rotated by an angular displacement previously fixed to make the results comparable across different configurations. To evaluate the capability of the proposed control algorithm in estimating wrist orientation, the angular displacement obtained from the control algorithm was compared with the one calculated from the Vicon system's data processing. Markers' positions were elaborated on MATLAB and used to compute the camera plane and the Principal Component of the object.

The results obtained from the first two experimental sessions led us to identify the best setup (in terms of camera model and image resolution) among the four considered for obtaining the best algorithm performance. Once the best setup was found, the third part of the experimental protocol focused on finding the best position for the CVS in a hand–wrist prosthesis and evaluating the performance of the proposed control strategy, more specifically of the Grasp Selection module, in 3 more complex operative conditions (i.e., in a less structured environment) named Grasp selection test: Coherent Condition, Grasp selection test: Non-Coherent Condition, and Approaching Condition.

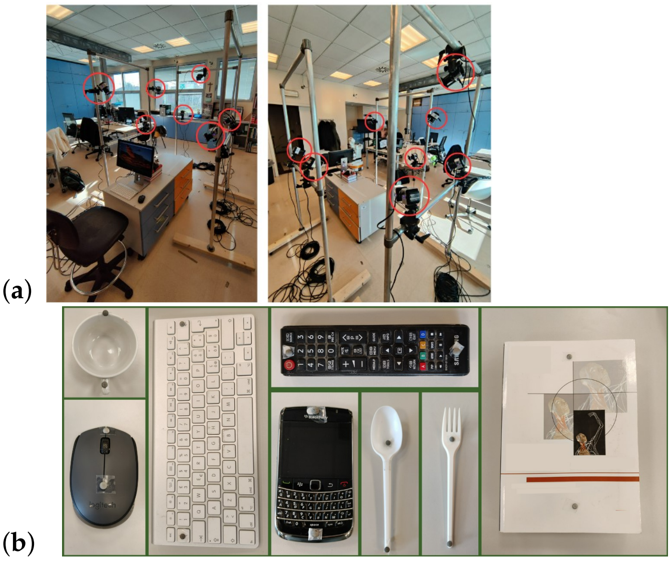


Figure 4. Experimental setup for Control Validation phase; VERO v2.2 cameras are shown in red circles (**a**). Objects used in the second session of the experimental protocol (**b**).

Firstly, the optimal camera placement was evaluated by positioning it in 4 different anatomical locations on the upper limb of 10 healthy subjects and assessing the different levels of hand occlusion. The test was conducted in an operational scenario with 5 men and 5 women (average age = 33.4 ± 13.6 years) of different anthropometric features, in order to characterize the level of occlusion with different wrist and hand sizes. The camera was placed at the wrist level, respecting all the positioning constraints defined in Section 2.1.3. Secondly, considering the so-obtained optimal camera placement, the CVS was placed on the upper limb of a healthy subject, as shown in Figure 5a, and the 3 operative tests were conducted considering the setup in Figure 5b.

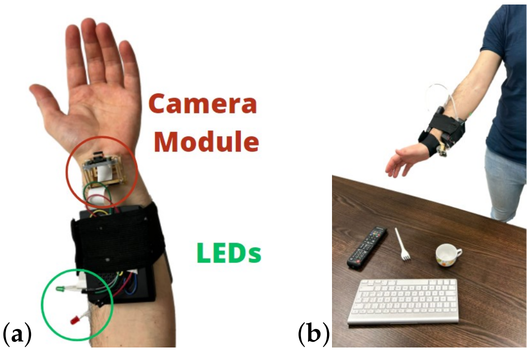


Figure 5. Positioning of CVS on the arm of the subject (**a**) and experimental setup for operative test (**b**).

In the first test, the performance of the control algorithm was analyzed whenever more than one object was detected and the user's intention changed. Four distinct items were displayed, each associated with a particular hand gesture, and the system was evaluated to see if it could select the proper object based on the EMG classifier input. This test aims at validating the capability of the proposed approach to identify the proper object according to the user's intention in a real-life scenario, where multiple tools/objects are often found close together.

The algorithm's ability to determine the proper hand gesture even when there was a discrepancy between the EMG-based user's intention and the hand gesture linked with the framed object was evaluated in the second test. Four items were provided, each representing a distinct grasp type, and the ability to accurately identify the hand gesture when the output from the EMG classifier was constant (Pointing) was tested.

The last test assessed the algorithm's capacity to detect when the prosthetic hand was approaching an object and required to shape the hand to grab it. The goal of this test was to use only information prior to the reaching phase for hand-shaping. Again, 4 objects were considered and the test was repeated twice for each.

In accordance with ethical guidelines and regulations governing research involving human subjects, this study has been determined to be exempt from obtaining approval from a relevant review board since the research poses minimal risk to the participants. The study primarily involves observations and does not involve any invasive or potentially harmful procedures. Participants have not been exposed to any physical, psychological, or social harm.

2.2.3. Key Performance Indicators (KPIs)

As explained in Section 2.2.2, each phase of the experimental protocol aims at assessing different aspects of the proposed control strategy. In order to assess the capability of the proposed pipeline to accurately detect the framed objects, the classification accuracies were quantified during the Classification Performance validation phase. In particular, the following accuracies were computed: *(i)* The Accuracy in Object Classification (AOC) assesses the correspondence between the real object and the predicted one, returned by the Object Detection module. *(ii)* The Accuracy in Grasp Classification (AGC) measures the correspondence between the true hand gesture and the predicted one.

During the Orientation Estimation Validation, the performance of the proposed control strategy in estimating object orientation and the computational burden of the implemented algorithm were characterized. The computed performance indicators were: *(i)* The Angular error (AE), which assesses the accuracy of the proposed system in estimating the object angular displacement and can be computed as

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

where $\Delta\theta_V$ and $\Delta\theta_P$ represent the angular displacement obtained from Vicon VERO and that computed by the proposed SCS, respectively. *(ii)* The Angular Estimation Stability (AES), which was analyzed considering the standard deviations of angles computed by the proposed algorithm and those obtained by the Vicon VERO data elaboration. The higher the value of the total standard deviations, computed per setup configuration, the lower the stability. These two KPIs were used in the first session of the protocol. *(iv)* The Analysis Frequency (AF), which quantifies, in frames per second (FPS), the approach to computational burden, i.e., the frequency at which the control algorithm analyzes the images.

Lastly, during the Operative Tests, the following performance indicators were computed: *(i)* The Occlusion Area (OA) , which is the area of the hand that occludes the scene and it is expressed as

$$OA = \frac{N_H}{N_{tot}} \cdot 100, \tag{2}$$

where N_H and N_{tot} are the number of pixels in which the hand is present and the total number of pixels, respectively. *(ii)* The Success Rate (SR), which is the success rate obtained in the Approaching Condition test, defined as the ratio between the number of trials in which the approaching phase (n) was recognized and the total trials N. It is expressed as

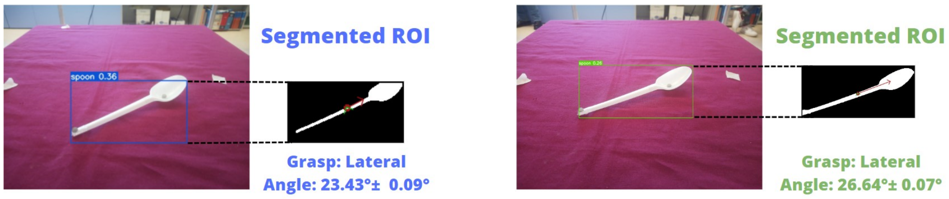
$$SR = \frac{n}{N} \cdot 100. \tag{3}$$

2.2.4. Statistical Analysis

The mean values and standard deviations of the above-mentioned KPIs were computed, specifically for AE and AF. Moreover, the Wilcoxon paired-sample test was performed for AE to analyze the differences between the two input sizes and the two camera sensors. In this way, identifying the best configuration among the four possible ones by comparing the angular errors for each object and each setup configuration was possible. Since all the performed tests are carried out on a couple of datasets, the significance level was set at p-value = 0.05.

3. Results and Discussions

Figure 6 shows the outputs of the proposed SCS for the same object, from Object Detection to Orientation Estimation, for both resolutions. As evident, the proposed approach was capable of *(i)* correctly detecting the framed spoon, *(ii)* segmenting it from the background, and *(iii)* estimating its orientation in the image plane.



(**a**) **240p Resolution** (**b**) **480p Resolution**

Figure 6. Example of Output in the first session of the experimental protocol for the 240p (**a**) and 480p (**b**) resolution and the same object.

Table 2 summarized the KPIs computed during the experimental validation of the proposed approach.

Table 2. Results obtained for each setup configuration in the first two sessions of experimental protocol.

	KPI 1		**KPI 2**	**KPI 3**	**KPI 4**
	AOC	*AGC*	*AE* [deg]	*AES* [deg]	*AF* [FPS]
5 MP 240p	97.80%	98.15%	14.17 ± 10.53	0.70	2.02 ± 0.13
5 MP 480p	97.35%	97.55%	11.40 ± 8.12	0.38	0.87 ± 0.13
16 MP 240p	97.85%	99.80%	16.26 ± 8.62	0.20	2.07 ± 0.15
16 MP 480p	99.35%	99.55%	17.35 ± 8.80	0.34	0.88 ± 0.17

As emerged from Table 2, each setup configuration obtained both AOC and $AGC > 97\%$. Misclassification errors for the 5 MP camera were due to fish-eye lens distortion and low confidence in detected objects. The majority of the errors were encountered for objects with spherical surfaces since fish-eye lens distortion modifies the shape of the objects and they were not detected with high confidence in many frames. On the contrary, the 16 MP camera sensor correctly classified spherical objects since the fish-eye lens distortion is absent. Since the autofocus of this camera requires 3 frames to properly focus the object, the algorithm was not capable of providing stable classification in the initial framing phase. Moreover, it emerged that $AGC \geq AOC$ objects associated with a certain hand gesture category are misclassified with another that can be manipulated with the same hand gesture. This highlights how the proposed approach is not significantly affected by object misclassification.

During the Orientation Estimation session, the average AE of the proposed system in estimating the object orientation did not exceed 18° and the AES was $\leq 0.8°$. The box plots of the AE for each object are divided with respect to the hand gestures and per setup configuration (Figure 7). It is worth observing that the median AE does not exceed 20°, except for the mouse and phone, considering the 16 MP camera with 480p resolution.

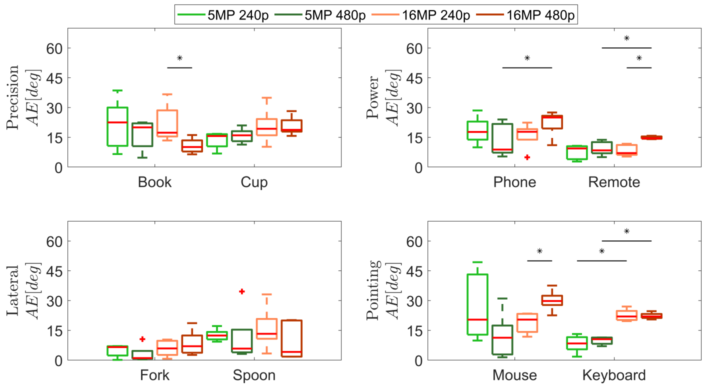


Figure 7. Box plot of angular errors for each object and setup configuration. The * denotes comparison in which p-value < 0.05.

Moreover, the statistical analysis highlights significant differences between the 16 MP camera module with two different resolutions only for the book (p-value = 0.032), remote (p-value = 0.008), and mouse (p-value = 0.032), with a higher AE for 240p resolution only in the first case. Considering 480p resolution for both cameras, the AE is lower for 5 MP one than the 16 MP in the case of the remote (p-value = 0.008), and keyboard (p-value = 0.008), whereas the phone is at the limit of statistical significance (p-value = 0.057). Furthermore, there is a statistical difference between the AE of the two different cameras at 240p resolution only for the keyboard.

Lastly, considering the computational burden of the proposed SCS, as expected, the execution time of the entire control cycle strongly depends on the chosen input image resolution rather than on the camera. AF is 2.02 ± 0.13 FPS and 0.87 ± 0.13 FPS for 240p and 480p resolutions, respectively, for the 5 MP camera sensor. Similarly, AF is 2.07 ± 0.15 FPS and 0.88 ± 0.17 FPS for 240p and 480p resolutions, respectively, for the 16 MP camera. The lower the resolution, the higher the AF of the proposed control algorithm. Regardless of the selected resolution, the Object Detection module takes three orders of magnitude longer than other parts, resulting in being the bottleneck of the control algorithm. In fact, the time required for Object Detection in the scene is >97% of the cycle time.

The comparative analysis highlighted that the 240p resolution outperformed the 480p one. It resulted in having a lower computational burden without affecting the AOC, AGC, AE, and AES. If the camera dimensions are taken into account, the 16 MP one can be considered the best choice since it can be more easily integrated into a prosthesis due to the absence of the fish-eye lens. Furthermore, the 16 MP camera module has the capability of automatically regulating focus: when integrating a camera into a prosthesis that is continuously in movement, automatic focus adjustment reduces the burden on the algorithm, which will always have to work under the condition where the objects of interest are in focus. Thus, the best configuration among the tested ones is the 16 MP Arducam High-Resolution Camera with a 320 × 240 pixels (240p) resolution. This configuration was employed for the last validation phase.

Moreover, the accuracy values in grasp classification are higher than or generally consistent with what has been observed in the literature [17,19], despite a substantial increase in the number of grasps [20,34] and the management of the prosthetic wrist orientation [21,22].

Figure 8a displays the average OA scores obtained for male and female subjects in all the four examined camera positions shown in Figure 8b. It emerged that the ventral position obtained the lowest OA (<15%). Furthermore, in this position, an object is framed until the moment it should be grasped, ensuring a more robust control of hand and wrist preshaping. The Operative tests were carried out by exploiting this camera positioning.

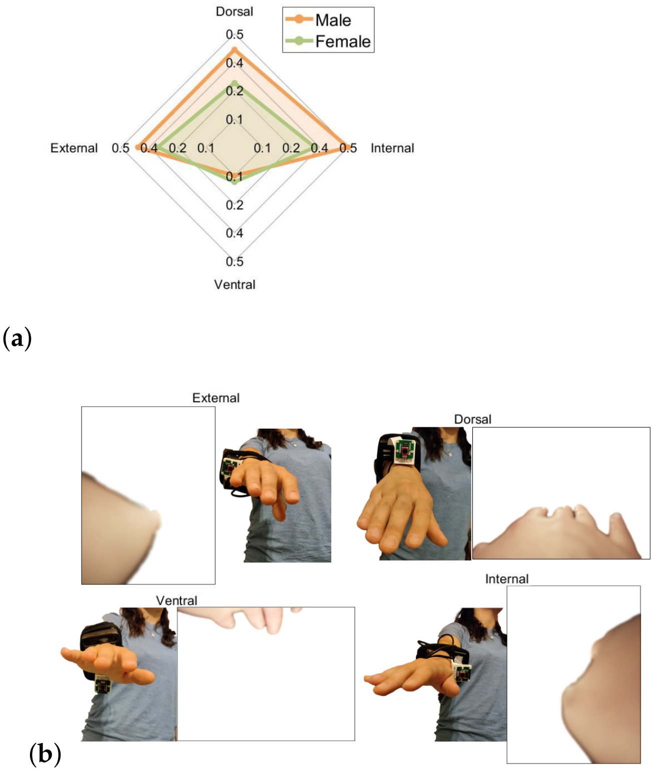


Figure 8. Spider plot of the average OA obtained (**a**). Positioning of the CVS on one subject for all the positions considered, with an example of a frame acquired at each of the four positions (**b**).

In the coherent condition, the proposed algorithm was able to select only the object whose hand gesture matches the EMG classifier output. Figure 9 shows the capability of the proposed SCS to change its output instantly. This occurs because the algorithm always identifies an object in the scene that is graspable with the gesture the user wants to perform. Furthermore, since an object of the user's motion intention class was always present, the red LED never turned on, i.e., the green LED never turned off. Consequently, the orientation estimation was smoothly aligned with the desired object.

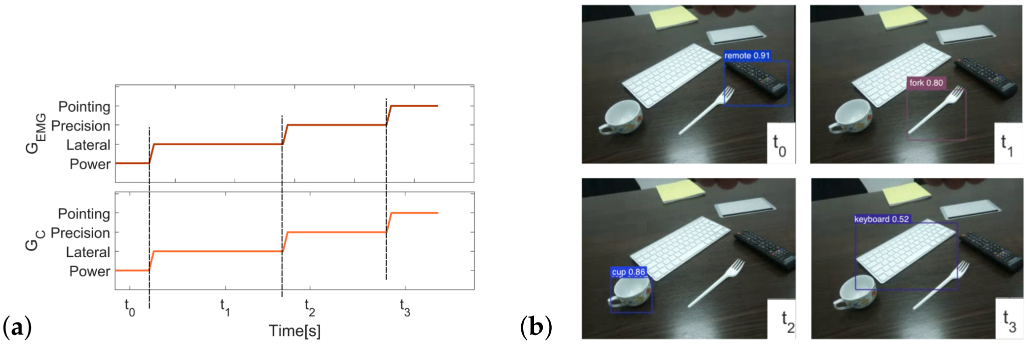


Figure 9. (**a**) Output from the Grasp selection test: coherent condition. (**b**) Frames acquired in four different time instants are reported on the right.

The non-coherent condition results are shown in Figure 10: starting from the initial output of the EMG classifier (dashed line), the SCS associates each object framed in the scene with the correct hand gesture (in orange) to return the correct selected grasp (in red). These discrepancies were fed back to the user by turning on the red LED.

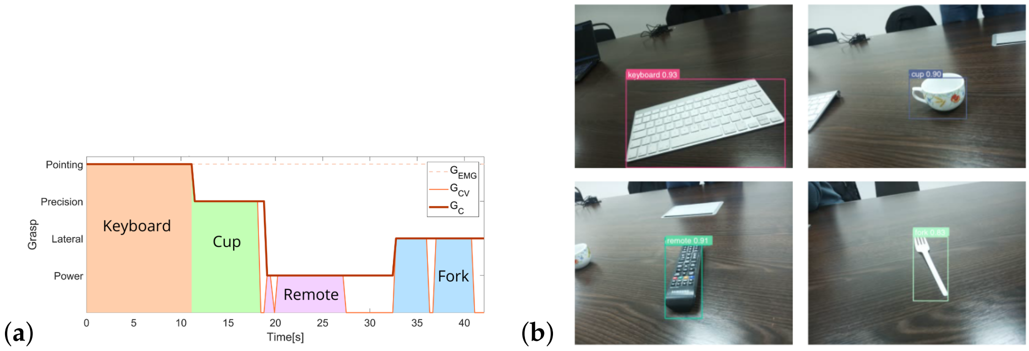


Figure 10. (**a**) Output from the Grasp selection test: non-coherent condition. (**b**) Four frames acquired by the camera are shown on the right.

The *SR* for the approaching condition test was 100%. The proposed algorithm was always capable of detecting the approach of the tested objects. Figure 11 shows a representative test conducted with the fork: the value of Area in pixels along with its variation at each frame is plotted.

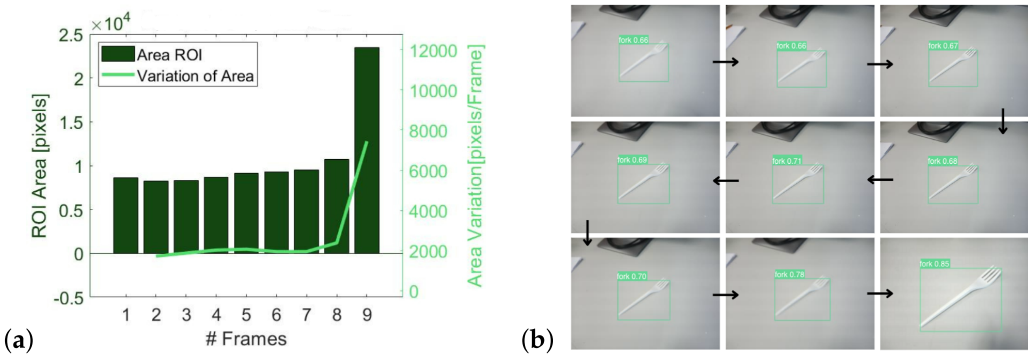


Figure 11. Output from Approaching condition test for the trial with a fork. (**a**) ROI area and its variation over time. (**b**) frames of the acquired video.

A preliminary analysis of the ease of use and the need to perform non-natural movements was carried out on one healthy subject wearing the system. As shown in Figure 5b, the user was asked to position himself in front of the objects in a natural way. In this configuration, the objects are recognized by the proposed algorithm without requiring the user to make non-natural or compensatory movements.

4. Conclusions

In this paper, a semiautonomous control system based on a computer vision system for wrist–hand prostheses was developed and tested. The SCS integrates a CNN-based object detection module, a grasp selection module, and an automatic thresholding algorithm for wrist orientation estimation. By combining exteroceptive information from the CVS with the user's intention via simulated EMG signals, the SCS aims at enhancing prosthesis control performance. The proposed SCS incorporates object detection, selective segmentation of the image, and dynamic thresholding. The results show promising outcomes, including high levels of accuracy in grasping and object classification (above 97%) and an average frame analysis frequency of 2.07 FPS. The developed SCS allows for the recognition of additional grasps beyond those detected by the EMG, ensuring the appropriate grasp for the specific object. The average angular error and angular estimation stability are below 18° and 0.8°, respectively, for all setup configurations. The proposed control strategy is capable of handling more complex situations as demonstrated by the conducted operative tests, reporting a success rate of 100% in recognizing the object approaching phase. Overall, the developed SCS grounded on CVS shows promising results in terms of accuracy, speed, angular estimation, and handling complex situations. Moreover, the proposed approach can be easily applied on different prosthetic hands [35] and robotic grippers [36], as long as they allow a range of hand gestures to be replicated. Indeed, in the design of the proposed solution, particular attention has been paid to build a CVS of reduced dimension to guarantee its portability and interoperability.

In the design of the proposed solution, particular attention has been paid to building a CVS of reduced dimension to guarantee its portability and interoperability.

Future efforts will be devoted to integrating the proposed CVS into a prosthetic hand and testing the developed SCS user-comfort, accuracy, and effectiveness in reducing the cognitive burden on a population of users. Moreover, the implemented SCS will be thoroughly characterized in the use case scenario, to quantify usability aspects in terms of acceptability, ease of use, and performance, i.e., grasp execution times, success rate, and compensatory movements.

Author Contributions: G.C. conceptualized the approach, performed the experiments, analyzed the data, and wrote the paper. C.T. performed the experiments, took care of the visualization of the results, and wrote the paper. L.P.C. conceptualized the approach, and reviewed the paper. F.C. conceptualized the approach, supervised the study, and wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DoF	Degrees of Freedom
EMG	Electromyography
SCS	Semiautonomous Control Strategy
CVS	Computer Vision System
CNN	Convolutional Neural Network
P/S	Pronation/Supination
SBC	Single-Board Computer
COCO	Common Objects in Context
ROI	Region Of Interest
PCA	Principal Component Analysis
PC	Principal Component
VNC	Virtual Network Computing
KPI	Key Performance Indicator
AOC	Accuracy in Object Classification
AGC	Accuracy in Grasp Classification
AE	Angular Error
AES	Angular Estimation Stability
AF	Analysis Frequency
FPS	Frames Per Second
OA	Occlusion Area
SR	Success Rate

References

1. Yamamoto, M.; Chung, K.C.; Sterbenz, J.; Shauver, M.J.; Tanaka, H.; Nakamura, T.; Oba, J.; Chin, T.; Hirata, H. Cross-sectional international multicenter study on quality of life and reasons for abandonment of upper limb prostheses. *Plast. Reconstr. Surg. Glob. Open* **2019**, *7*, e2205. [CrossRef] [PubMed]
2. Tamantini, C.; Cordella, F.; Lauretti, C.; Zollo, L. The WGD—A dataset of assembly line working gestures for ergonomic analysis and work-related injuries prevention. *Sensors* **2021**, *21*, 7600. [CrossRef] [PubMed]
3. Jang, C.H.; Yang, H.S.; Yang, H.E.; Lee, S.Y.; Kwon, J.W.; Yun, B.D.; Choi, J.Y.; Kim, S.N.; Jeong, H.W. A survey on activities of daily living and occupations of upper extremity amputees. *Ann. Rehabilit. Med.* **2011**, *35*, 907–921. [CrossRef]
4. Smail, L.C.; Neal, C.; Wilkins, C.; Packham, T.L. Comfort and function remain key factors in upper limb prosthetic abandonment: Findings of a scoping review. *Disabil. Rehabilit. Assist. Technol.* **2021**, *16*, 821–830. [CrossRef] [PubMed]
5. Igual, C.; Pardo, L.A., Jr.; Hahne, J.M.; Igual, J. Myoelectric control for upper limb prostheses. *Electronics* **2019**, *8*, 1244. [CrossRef]
6. Roche, A.D.; Rehbaum, H.; Farina, D.; Aszmann, O.C. Prosthetic myoelectric control strategies: A clinical perspective. *Curr. Surg. Rep.* **2014**, *2*, 44. [CrossRef]
7. Atzori, M.; Cognolato, M.; Müller, H. Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands. *Front. Neurorobot.* **2016**, *10*, 9. [CrossRef]
8. Hahne, J.M.; Schweisfurth, M.A.; Koppe, M.; Farina, D. Simultaneous control of multiple functions of bionic hand prostheses: Performance and robustness in end users. *Sci. Robot.* **2018**, *3*, eaat3630. [CrossRef]
9. Leone, F.; Gentile, C.; Cordella, F.; Gruppioni, E.; Guglielmelli, E.; Zollo, L. A parallel classification strategy to simultaneous control elbow, wrist, and hand movements. *J. NeuroEng. Rehabilit.* **2022**, *19*, 10. [CrossRef]
10. Yadav, D.; Veer, K. Recent trends and challenges of surface electromyography in prosthetic applications. *Biomed. Eng. Lett.* **2023**, *13*, 353–373. [CrossRef]
11. Zhang, Q.; Zhu, J. The Application of EMG and Machine Learning in Human Machine Interface. In Proceedings of the 2nd International Conference on Bioinformatics and Intelligent Computing, Harbin, China, 21–23 January 2022; pp. 465–469.
12. Tomovic, R.; Boni, G. An adaptive artificial hand. *IRE Trans. Autom. Control* **1962**, *7*, 3–10. [CrossRef]
13. Stefanelli, E.; Cordella, F.; Gentile, C.; Zollo, L. Hand Prosthesis Sensorimotor Control Inspired by the Human Somatosensory System. *Robotics* **2023**, *12*, 136. [CrossRef]
14. Dosen, S.; Cipriani, C.; Kostić, M.; Controzzi, M.; Carrozza, M.C.; Popović, D. Cognitive vision system for control of dexterous prosthetic hands: Experimental evaluation. *J. Neuroeng. Rehabilit.* **2010**, *7*, 42. [CrossRef] [PubMed]
15. Došen, S.; Popović, D.B. Transradial prosthesis: Artificial vision for control of prehension. *Artif. Organs* **2011**, *35*, 37–48. [CrossRef] [PubMed]
16. Castro, M.N.; Dosen, S. Continuous Semi-autonomous Prosthesis Control Using a Depth Sensor on the Hand. *Front. Neurorobot.* **2022**, *16*, 814973. [CrossRef]
17. Ghazaei, G.; Alameer, A.; Degenaar, P.; Morgan, G.; Nazarpour, K. Deep learning-based artificial vision for grasp classification in myoelectric hands. *J. Neural Eng.* **2017**, *14*, 036025. [CrossRef]
18. Dhillon, A.; Verma, G.K. Convolutional neural network: A review of models, methodologies and applications to object detection. *Prog. Artif. Intell.* **2020**, *9*, 85–112. [CrossRef]
19. Weiner, P.; Starke, J.; Rader, S.; Hundhausen, F.; Asfour, T. Designing Prosthetic Hands With Embodied Intelligence: The KIT Prosthetic Hands. *Front. Neurorobot.* **2022**, *16*, 815716. [CrossRef]
20. Perera, D.M.; Madusanka, D. Vision-EMG Fusion Method for Real-time Grasping Pattern Classification System. In Proceedings of the 2021 Moratuwa Engineering Research Conference, Moratuwa, Sri Lanka, 27–29 July 2021; pp. 585–590.
21. Cognolato, M.; Atzori, M.; Gassert, R.; Müller, H. Improving robotic hand prosthesis control with eye tracking and computer vision: A multimodal approach based on the visuomotor behavior of grasping. *Front. Artif. Intell.* **2022**, *4*, 744476. [CrossRef]
22. Deshmukh, S.; Khatik, V.; Saxena, A. Robust Fusion Model for Handling EMG and Computer Vision Data in Prosthetic Hand Control. *IEEE Sens. Lett.* **2023**, *7*, 6004804. [CrossRef]
23. Cordella, F.; Di Corato, F.; Loianno, G.; Siciliano, B.; Zollo, L. Robust pose estimation algorithm for wrist motion tracking. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 3746–3751.
24. Boshlyakov, A.A.; Ermakov, A.S. Development of a Vision System for an Intelligent Robotic Hand Prosthesis Using Neural Network Technology. In Proceedings of the ITM Web of Conference EDP Sciences, Moscow, Russia, 28–29 November 2020; Volume 35, p. 04006.
25. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
26. Phadtare, M.; Choudhari, V.; Pedram, R.; Vartak, S. Comparison between YOLO and SSD mobile net for object detection in a surveillance drone. *Int. J. Sci. Res. Eng. Manag.* **2021**, *5*, b822–b827.
27. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; Fleet, D.; Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
28. Feix, T.; Romero, J.; Schmiedmayer, H.B.; Dollar, A.M.; Kragic, D. The grasp taxonomy of human grasp types. *IEEE Trans. Hum.-Mach. Syst.* **2015**, *46*, 66–77. [CrossRef]

29. Flanagan, J.R.; Terao, Y.; Johansson, R.S. Gaze behavior when reaching to remembered targets. *J. Neurophysiol.* **2008**, *100*, 1533–1543. [CrossRef] [PubMed]
30. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]
31. Tamantini, C.; Lapresa, M.; Cordella, F.; Scotto di Luzio, F.; Lauretti, C.; Zollo, L. A robot-aided rehabilitation platform for occupational therapy with real objects. In Proceedings of the Converging Clinical and Engineering Research on Neurorehabilitation IV: 5th ICNR2020, Vigo, Spain, 13–16 October 2022; pp. 851–855.
32. Gardner, M.; Woodward, R.; Vaidyanathan, R.; Bürdet, E.; Khoo, B.C. An unobtrusive vision system to reduce the cognitive burden of hand prosthesis control. In Proceedings of the 13th ICARCV, Kunming, China, 6–9 December 2004; pp. 1279–1284.
33. DeGol, J.; Akhtar, A.; Manja, B.; Bretl, T. Automatic grasp selection using a camera in a hand prosthesis. In Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 431–434.
34. Castro, M.C.F.; Pinheiro, W.C.; Rigolin, G. A Hybrid 3D Printed Hand Prosthesis Prototype Based on sEMG and a Fully Embedded Computer Vision System. *Front. Neurorobot.* **2022**, *15*, 751282. [CrossRef]
35. Devi, M.A.; Udupa, G.; Sreedharan, P. A novel underactuated multi-fingered soft robotic hand for prosthetic application. *Robot. Auton. Syst.* **2018**, *100*, 267–277.
36. Sun, Y.; Liu, Y.; Pancheri, F.; Lueth, T.C. Larg: A lightweight robotic gripper with 3-d topology optimized adaptive fingers. *IEEE/ASME Trans. Mechatronics* **2022**, *27*, 2026–2034. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.