POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Teleoperation of an Underactuated Bionic Hand: Comparison between Wearable and Vision-based Motion Tracking Methods

Tesi di Laurea Magistrale in
Biomedical Engineering - Ingegneria Biomedica

Author: **Massimiliano Poletti**

Student ID: 945619
Advisor: Prof. Elena De Momi
Co-advisors: Junling Fu
Academic Year: 2021-22

Acknowledgements

Foremost, I would like to express my gratitude to my advisor Prof. Elena De Momi for giving me the opportunity to have a wonderful research experience in a field that I like within a multicultural team, for the precious advice she gave me over the last year and for the helpful revisions she did.

I am sincerely grateful to my co-advisor Junling Fu, who friendly and wisely guided me through the steps and was always there when I needed support, showing patience, altruism and devotion to research.

A special thank goes to my colleague Ilaria Burzo, who was part of the experimental setup as she personified operator 2 in the accuracy test, thus giving me a fundamental help.

Also, thanks to all the people who kindly accepted to be involved in the user experiment, namely Elisa Iovene, Riccardo Bertolotti, Karim Abd el Fatah, Qingsheng Liu and Aldo Marzullo.

Last but not least, heartfelt thanks to my family for supporting and encouraging me throughout all these years.

Abstract

Nowadays, service robotics has to face dynamic and unstructured environments, which make sensorial information less reliable and autonomous decision-making harder. Fast adaptation control strategies and versatile hardware solutions are fundamental to make actions effective and grant safety. The incorporation of human intelligence through teleoperation and the design of bio-inspired components represent a promising solution. In particular, the use of bionic hands as end-effectors of manipulators can make robots successfully interact with highly irregular environments. Underactuated bionic hands are especially suited thanks to their simplified driver structure, compactness and effectiveness. Teleoperation control of bionic hands can be intuitively achieved by capturing the human operator's hand motion and mapping it to the bionic hand. The hand motion capture technologies mainly used in teleoperation are sensorized mechanical gloves and vision-based tracking devices. In general, while the former proved to be more stable and robust, the latter allow more natural movements and grant higher user comfort. Although the characteristics of the two different approaches are clear, literature lacks robust comparative studies between them when applied to teleoperation, as a basis to select the most appropriate method. Moreover, most of the works related to the development of teleoperation systems are oriented toward self-validation, which ultimately prevents a meaningful comparison between different works.

Given such premises, in this thesis, two teleoperation systems were designed to intuitively control the same underactuated bionic hand using a glove-based method and a vision-based technique, respectively. The two systems were tested within the same experimental setup to verify their usability and obtain a robust comparison of the two motion capture technologies, in terms of both control accuracy and resulting performance in a realistic scenario. First of all, the bionic hand fingers motion was calibrated using an optical tracking device. The resulting non-linear inverse kinematic model was employed to develop an accurate actuation controller for the bionic hand. Then, the two tracking systems were configured to extract novel-defined features to describe the motion of the human hand fingers. So, the tracking stages were coupled with the actuation controller to achieve a simple and intuitive mapping between human and robot hand poses. Lastly, the ob-

tained teleoperation systems were compared through both static and dynamic accuracy experiments and a usability grasping test performed by several users.

Both the developed systems proved to be exploitable in both powered and precise grasp tasks. Nevertheless, while the glove-based method proved good accuracy performance, the vision-based technique showed an accuracy level that may be insufficient in real applications concerning fine manipulation of objects. The user experiment underlined that such worse accuracy may also influence the grasping performance. The worse performance is attributable to the limited accuracy of the vision camera device, thus, further research is needed to develop more robust algorithms for image-based hand pose estimation. The user study also underlined that the glove-based method brings a worse comfort and higher effort by the user. Therefore, the glove design must be improved taking user comfort into account.

Keywords: human-robot interaction, hand motion capture, bionic hand, teleoperation

Abstract in lingua italiana

L'allargamento delle applicazioni della robotica di servizio, registrato negli ultimi vent'anni, richiede ai moderni robot di interagire con ambienti sempre più dinamici e meno strutturati, che rendono le loro informazioni sensoriali meno affidabili, quindi i processi decisionali da essi affrontati più complessi. L'adozione di strategie di controllo adattativo e di componenti meccanici versatili è quindi fondamentale per rendere le operazioni dei robot più efficaci e sicure. L'incorporazione dell'intelligenza umana nel sistema di controllo, attraverso la teleoperazione, e la progettazione di componenti bio-ispirati rappresentano una soluzione promettente. In particolare, l'utilizzo di mani bioniche come effettori può rendere più semplice l'interazione fra il robot e ambienti altamente irregolari. Le mani bioniche sottoattuate si rivelano specialmente adatte allo scopo, grazie alla loro semplificata struttura di controllo, compattezza ed efficacia. Il controllo delle mani bioniche in teleoperazione può essere ottenuto intuitivamente catturando il movimento della mano dell'operatore umano e mappandolo sulla mano bionica. Le tecnologie di acquisizione del movimento della mano principalmente utilizzate a tal fine sono guanti meccanici sensorizzati e dispositivi di cattura basati sulle immagini. In generale, mentre i primi hanno dato prova di essere più stabili, i secondi consentono movimenti più naturali e garantiscono un maggiore comfort per l'utente. Sebbene le caratteristiche dei due diversi approcci siano chiare, in letteratura mancano solidi studi comparativi fra i due qualora applicati in teleoperazione, come base per la selezione del metodo più appropriato. Inoltre, la maggior parte degli studi legati allo sviluppo di sistemi di teleoperazione sono orientati all'autovalidazione, che, in definitiva, impedisce un confronto significativo tra studi diversi.

Date tali premesse, in questo lavoro sono stati progettati due sistemi di teleoperazione per controllare intuitivamente la stessa mano bionica sottoattuata, utilizzando rispettivamente un metodo basato sull'utilizzo di guanti sensorizzati e una tecnica basata sulle immagini, in particolare sulla visione stereoscopica. I due sistemi sono stati testati all'interno dello stesso setup sperimentale per verificarne l'utilizzabilità e ottenere un solido confronto tra i due metodi di cattura utilizzati, in termini sia di accuratezza del controllo che della risultante prestazione in uno scenario applicativo realistico. Innanzitutto, il movimento delle dita della mano bionica è stato calibrato utilizzando un sistema di tracciamento ot-

tico. Il modello cinematico inverso non lineare risultante è stato impiegato per sviluppare un accurato controllore per l'attuazione della mano bionica. Dunque, i due sistemi di cattura sono stati configurati per estrarre variabili motorie di nuova definizione capaci di descrivere il movimento delle dita della mano umana, dopodiché sono stati accoppiati con il controllore di attuazione per ottenere una mappatura semplice e intuitiva tra la posa della mano umana e quella della mano bionica. I sistemi di teleoperazione così ottenuti sono stati confrontati tramite tre esperimenti per valutare l'accuratezza statica e dinamica, l'usabilità e l'esperienza dell'utente.

Entrambi i sistemi sviluppati si sono rivelati efficaci nel controllare la mano bionica per afferrare oggetti, tramite sia impugnature di potenza che di precisione. Tuttavia, mentre il metodo basato sull'utilizzo dei guanti ha dimostrato buone prestazioni di accuratezza, la tecnica basata sulla visione ha mostrato un livello di accuratezza che potrebbe essere insufficiente in applicazioni reali riguardanti la manipolazione fine di oggetti. Il grasp test ha sottolineato che tale limite di accuratezza può avere un'influenza sulle prestazioni dell'utente. La limitata accuratezza è attribuibile al software per la cattura del movimento, pertanto, ulteriore ricera si rende necessaria per sviluppare algoritmi più robusti per la stima della posa della mano basata sulle immagini. Lo studio riguardo l'eseperienza degli utenti durante il grasp test ha altresì registrato che il metodo basato sui guanti porta a un comfort ridotto da parte dell'utente. Pertanto, il design del guanto deve essere migliorato tenendo maggiormente conto di tale aspetto.

Parole chiave: interazione uomo-robot, acquisizione del movimento della mano, mano bionica, teleoperazione

Contents

1 | Introduction

1.1. Context of bionic hands teleoperation

Robotics as a research and technological field was born in the early 1960s to relieve humans from hard, risky or unpleasant tasks in the context of industrial production. For thirty years, industrial needs have polarized robotics research towards the optimization of interaction with highly controlled environments. Starting from the 1990s, thanks to technological progress in both hardware and software, robotics spread outside factories evolving into service robotics [1]. According to ISO 8373: 2012 [2] a service robot is a robot that performs useful tasks for humans excluding industrial automation applications. From such a generic definition it is clear that the number of application scenarios increased, as well as the complexity of the operation environment and of the tasks humans ask robots to execute. In the most complex cases, the robot has to interact with dynamic and unstructured environments. Such a situation poses challenges that cannot easily be addressed by approaches developed for highly controlled environments like traditional industrial assembly lines. In unstructured environments, robot sensorial information is less reliable [3]. The robot cannot completely trust its knowledge about the surroundings, thus decision-making becomes a hard challenge if it is based on robotics techniques that rely on perfect knowledge of the world. Then, variability related to dynamicity may lead to unexpected events the robot cannot promptly react or adapt to. In such situation, providing the machine with robust adaptive control and flexible hardware solutions is fundamental to make its actions effective and to prevent it from damaging itself, the environment or surrounding people, if present.

Concerning the control, although artificial intelligence reached good results in the last decades, standard intelligent programming based on automatic optimizing planners is not enough reliable to manage the aforementioned situations yet [4]. Teleoperation control of the robot by a remote human operator is still the best solution because it incorporates human intelligence to manage variability, poor sensing and unexpected events. The assistance provided by humans depends on the uncertainty degree and it can be classified into three levels [5]: supervised control, if the machine performs tasks autonomously and

the human operator merely supervises the execution; direct control, if the robot entirely follows the commands coming from the human operator, lacking any autonomous capability; shared control, if the robot is only partially autonomous and the operator has an active but not complete role in control.

Concerning the hardware, an innovative approach to implement robotic mechanical structures able to effectively interact with complex environments is provided by bionics, which is the application of structural and functional rules found in nature to design engineering systems. In particular, the human hand is an extremely adaptable mechanical structure. Millions of years of evolution have shaped it to be very sophisticated and versatile, allowing our species to interact profitably with a wide range of objects and environments, thus making it one of the most adaptable on the planet. It consists of 27 bones, reciprocally articulated in such a way to achieve 20 local degrees of freedom (DoF), as depicted in figure 1.1. Additionally, six global DOF are provided by the wrist to change the hand orientation. Besides the mechanical structure, the actuation and control systems are intricate too. Thirty-four extrinsic and intrinsic muscles provide strength and precision respectively, receiving control signals from the central nervous system via three different nerves, which also send somatosensory feedback generated by 17000 mechanoreceptors [6].

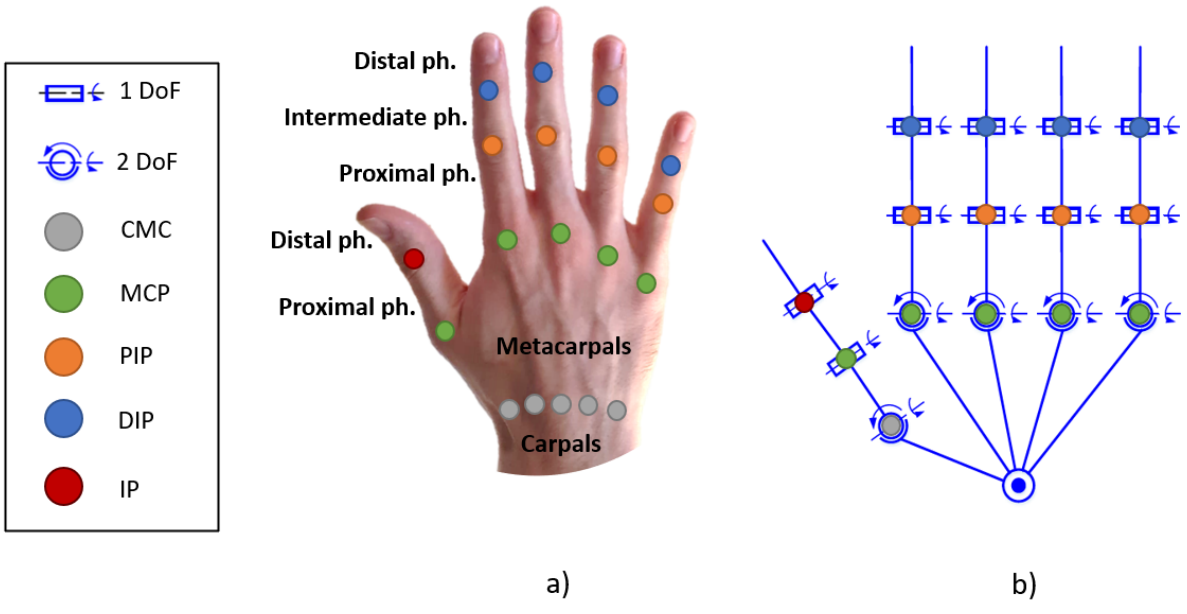


Figure 1.1: Human hand structure. (a) Bones and joint classification. (b) Kinematic model with 20 local DoF (image from [4]).

Bionic hands try to imitate the structure and the function of the human hand. However, replicating such a complex system is nearly impossible. The design must take into account

many factors and usually, a compromise must be found. Important design factors include the number of DoF, actuation type, applicable force, tactile and force sensing, weight, size, power, materials, cost, dexterity, compliance, and controllability [4]. D. Rebollo et al. [7] provided a review of multi-fingered robotic hands between 1983 and 2016, including hardware specifications. The existing bionic hands are classified as fully actuated and underactuated. In the former case, the degree of actuation (DoA) is equal to the DoF. For example, the bionic hand implemented in the DLR system [8] has 19 DoF, each one driven by a couple of antagonistic motors placed in the forearm of the robot, as depicted in figure 1.2.a. The main advantage of this solution lies in dexterity. However, this brings to higher design complexity, cost, larger size and weight, given the higher number of actuators needed. On the contrary, underactuated bionic hands are characterized by a DoA smaller than the DoF. For example, Shunk SVH [9] depicted in figure 1.2.b is endowed with 20 DoF controlled by a set of 9 motors contained inside the hand structure. The lack of a complete actuation is partially compensated by higher compliance that allows self-adaptive grasping, granting good grasping performance on a huge variety of objects. For these reasons, they represent a good trade-off between simplicity and versatility, thus, they are gaining widespread attention in research. The main issue with underactuation is that the resulting nonlinear motion characteristics impede easy mathematical modeling of the device mechanics [10].

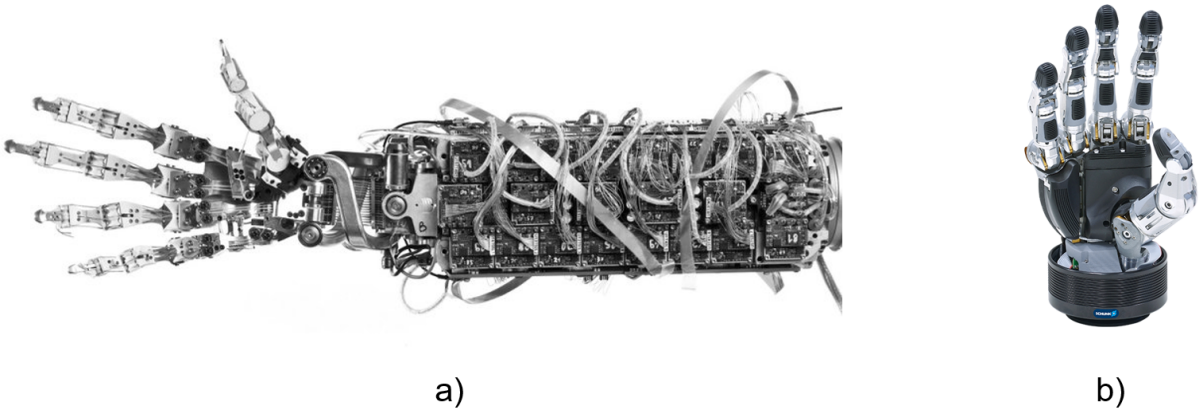


Figure 1.2: Bionic hand examples. (a) Fully actuated bionic hand employed in the DLR system. (b) Underactuated Shunk SVH bionic hand.

The use of bionic hands, in particular underactuated ones, as end-effectors of robotic manipulators could improve the skills of teleoperated robots, making them able to interact with unstructured environments like humans, which means manipulating practically any kind of object or tool and eliminating the need to design and mount specialized end-effectors for different scenarios. Such application of bionic hands has been investigated in

several scenarios, employing, for example, the robots depicted in figure 1.3. In particular, Robo Sally [11] and DRH-HUBO [12] robots were tested for emergency response activities, while NASA Valkyrie [13] and Robonaut [14] humanoid robots were used for maintenance of nuclear plants and space applications, respectively. These studies demonstrated that, although bionic hands have the same basic function as traditional two-fingered grippers, as the ones used in [15] and [16], however, their complex kinematics resembling the human one provides a higher grasping versatility.

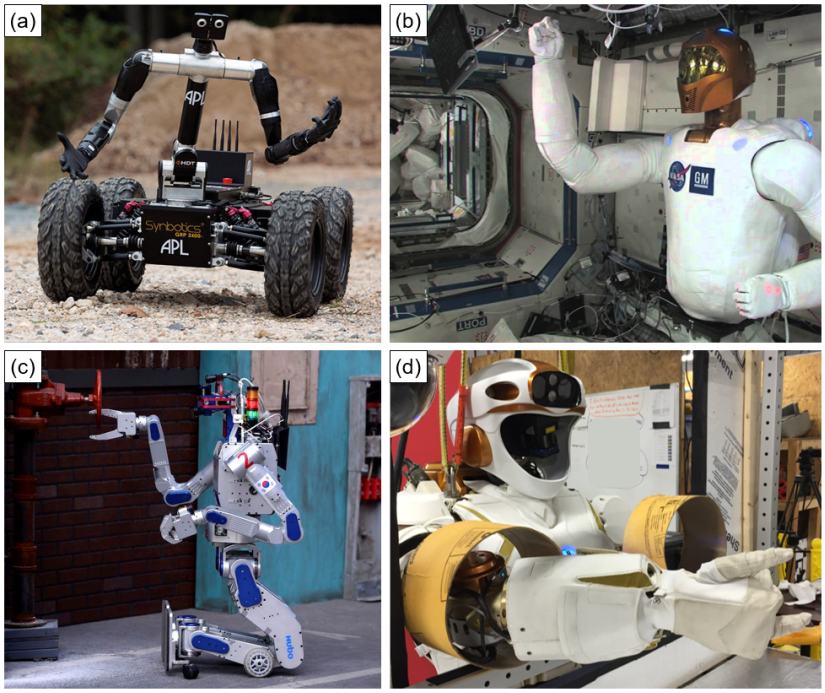


Figure 1.3: Application examples of bionic hand teleoperation. (a) Robo Sally. (b) NASA Robonaut. (c) DRH-HUBO. (d) NASA Valkyrie

1.2. State of the art

Motion retargeting Joysticks are the typical controller devices used by human operators in teleoperation robotics. However, the local DoF a bionic hand is not so intuitive to control with such traditional approach given the high number of parallel joints. Innovative controllers have been implemented to capture the motion of the human operator's fingers and then map it to the bionic hand fingers motion in real-time to obtain a similar gesture, taking advantage of the kinematic correspondence between the two structures. Such method is referred to as "human-to-robot motion retargeting" and can make the control much more intuitive, immersive and de-alienating. A fundamental issue of motion

retargeting is that the kinematic differences between human and robot structures forbid an identical reproduction of the human gesture on the robot, in particular, if the bionic hand is underactuated. Incongruities may affect DoF, actuation strategy and workspace. Therefore, the real goal of mapping is that the resulting correspondence between the human hand gesture and the bionic hand motion is intuitively understandable by the user without any mental effort or long-time learning needed. In other words, the movement of the bionic hand resulting from teleoperation must be highly semantically correlated to the human one. The concept of semantic correlation is depicted in figure 1.4.

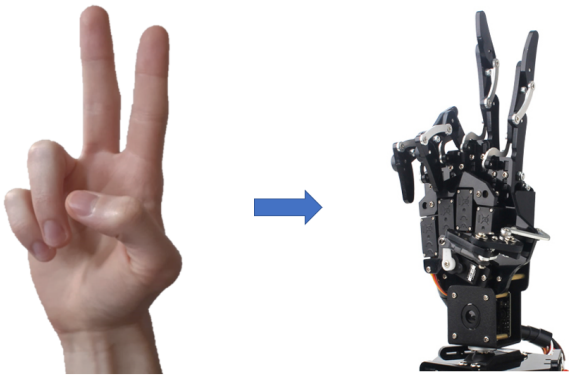


Figure 1.4: Example of semantic correlation between human hand and bionic hand.

Current mapping strategies propose solutions that can be classified into gesture recognition mapping and hand pose estimation mapping [4]. In the former case, a gesture recognition algorithm is trained to classify discrete human hand poses that are associated with predefined robotic hand poses. Although a limited set of commands makes the system largely controllable and reduces the risk of damaging the robotic hand and the grasped objects, it is also a great limitation in terms of control smoothness and grasping versatility. On the contrary, in hand pose estimation mapping, both human and robot poses are represented in continuous parameter spaces and a continuous function is used to derive the desired robot parameters according to the measured human ones, thus achieving high-resolution teleoperation. The main limitation is that inaccurate pose estimation may lead to abnormal commands, making the teleoperation results unpredictable. However, thanks to its higher versatility, continuous control is preferable to a discrete one. The main classes of hand pose estimation mapping methods are joint-to-joint mapping and point-to-point mapping, although not all methods fit strictly in this classification. In the former case, the angles related to human hand joints are mapped to robotic joints angles. As a result, the overall gesture similarity is high, but fingertip positions are not well replicated. Consequently, this method is suitable for power grasp. In point-to-point

mapping, human fingertips positions are mapped to robotic fingertips positions. This method is more complex because it needs intense computations in terms of forward and inverse kinematics, and calibration between human and robot frames, but it provides the system with precision grasp skills.

Hand motion capture: sensorized gloves Many technological solutions have been investigated to capture hand motion in recent years. Mainly they can be classified as wearable tracking devices and contactless vision-based tracking devices. According to Li et al. survey [4], the most commonly used hand tracking devices in the teleoperation context are wearable systems equipped with mechanical sensors, that can be referred to as mechanical gloves. The main commercially available mechanical gloves are listed by [17] and split into three main classes: exoskeletal, fabric and striped. One example for each category is reported in figure 1.5. Exoskeletal gloves consist of a base rigid structure located in the back of the hand, coupled with some fabric strings to make it adhere, and some rigid links connected to the fingers, which typically transmit the motion to rotational sensors placed in the base. This kind of glove is suitable to deliver kinesthetic (force) feedback through proper embedded actuators. Fabric gloves are pieces of fabric that fully cover the hand and its fingers, endowed with small mechanical sensors like IMU, bend sensors and strain sensors. Striped gloves are composed of strips of fabric, plastic or other materials that do not entirely cover the hand but only the places that have to be sensed. The same kind of sensors as fabric gloves are typically used in this case.

Figure 1.5: Examples of sensorized mechanical gloves. (a) Exoskeleton: SensGlove DK. (b) Fabric: CaptoGlove. (c) Striped: Rapael.

Liu and Zhang [18] use a Cyberglove to measure human fingertips cartesian coordinates with respect to a local reference frame attached to the human hand. Tips positions are then mapped to the reference frame of a virtual bionic hand considering a proper scaling

factor. Inverse kinematics is applied to obtain the robot motor input. A teleoperation experiment is conducted to assess the distance between the dynamic trajectories of human and robot fingertips respectively in terms of mean distance. Colasanto et al. [19] use a mechanical glove provided with 22 angle sensors to teleoperate a Shunk Anthropomorphic Hand endowed with 13 independent DoF. The authors merged a traditional joint-to-joint mapping and a fuzzy logic mapping that identify some poses of the human hand and associates them with predefined poses of the mechanical hand. The system is tested through grasping trials performed by different users on different types of objects. Fani et al. [20] use a fabric glove equipped with five goniometers to control an underactuated bionic hand endowed with 19 DoF and just one DoA. Glove data are used to extract one motion feature that represents the level of synergy of the human hand, which is then mapped to the input of the hand motor. The hand motion retargeting is effectively integrated with a teleoperation system used to move a Kuka arm which employs the bionic hand as an end-effector. The overall system is evaluated through a user study dealing with a drilling task. Success rate and user subjective evaluation are recorded.

Mechanical gloves outperform the vision-based tracking methods in terms of stability and robustness and do not suffer from occlusion problems. However, some practical limitations still hinder their application. The one-size-fits-all design, the cumbersome wearing process, and the adoption of unergonomic structures lead to poor comfort and affect the intuitiveness and the transparency of the operation. It has been well documented that the overall measurement performance for these devices is influenced by how well they fit the user's hand [21]. Fitting quality also influences user comfort. Indeed, bad fitting impedes natural movements and makes the user feel a higher effort during the activity. Moreover, given that each hand is different from the others, a calibration procedure is always required and, in some cases, it might take a long time and be not precise. Another typical problem with mechanical gloves is wear-and-waste.

Other types of hand tracking wearable systems are available but less used because of their specific limitations. Marker gloves are composed of a set of optical markers positioned in such a way to retrieve joints 3D position using infrared stereo cameras. Such tracking systems can be very accurate but the hardware setup can be very cumbersome. In [22] for example, a set of eleven cameras is needed to assure continuous monitoring. Moreover, ambient light and bright objects may interfere and the field of view around the operator needs to be always clear. Electromyographic sensors have been studied too, but the captured data are poor and cannot be used alone to achieve continuous and precise control of the fingers motion. In [23] for example, just a limited set of predefined human gestures could be recognized by EMG data and mapped to predefined poses of a bionic hand.

Hand motion capture: vision-based methods Vision-based tracking devices propose an interesting alternative to wearable devices [24]. In this case, the hardware is reduced to the minimum, typically consisting of just one depth camera which obtains images of the bare human hand. No wearable devices are usually needed, thus overcoming the problems related to size fitting, intrusiveness, discomfort and wear-and-waste typical of mechanical gloves and allowing also more natural movements. Moreover, the recovery of the global DoF of the human hand has been proven to be feasible, while in the wearable tracking approach, gloves can only capture the local DoF, and recovering the global DoF requires other expensive devices.

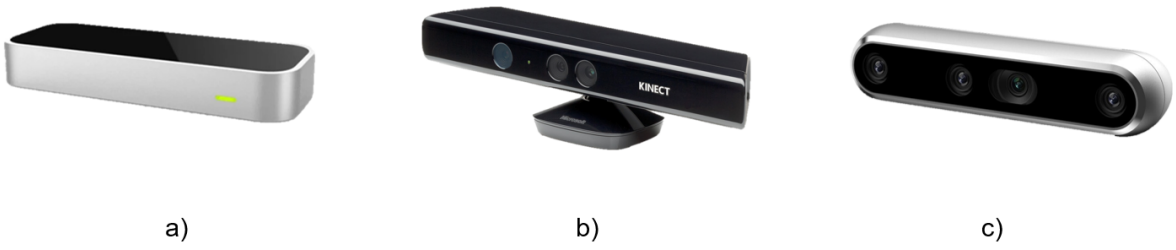


Figure 1.6: Examples of vision tracking devices: (a) Leap Motion Controller, (b) Kinect, (c) Realsense.

In most cases, the depth images are processed by an advanced algorithm that must segment the hand and recognize the 3D hand joints locations with respect to a fixed reference frame. The sensing principle is usually based on time of flight (ToF), structured light, or stereo vision. ToF cameras sense depth by emitting a modulated light signal and measuring the time delay of the echo wavefront. In structured light technology, a pattern of infrared dots is projected onto the object to be tracked and is captured by an infrared camera, then stereo triangulation is used to obtain the point position of the projections. Stereo vision systems generate two images from slightly different perspectives. The disparity map between the images is used to infer depth information. The depth of the corresponding pixel location is inversely proportional to the disparity value. According to Li survey[4], the commercially available vision tracking devices that are mostly used in teleoperation of bionic hands are the ones depicted in figure 1.6, that are Leap Motion Controller (Ultraleap, United States), Kinect (Microsoft, United States) and RealSense (Intel, United States). Leap Motion (LMC) is gaining traction in research due to its high accuracy (0.7mm on average) in comparison to other devices that are not specifically designed for hand tracking but rather for more generic human motion capture.

Scarcia et al. [25] utilized an LMC to measure human hand fingertips positions with respect to a local reference frame attached to the hand and map them with proper scaling to the fingertips positions of a UB-Hand IV bionic hand endowed with 20 local DoF and 15 DoA. Grasping experiments were conducted to assess the performance and the intuitiveness of the control, quantifying the success rate. Bimbo et al. [26] employed an LMC in order to teleoperate a Pisa/IIT SoftHand. In this case, LMC was also used to control the global DoF of the hand by teleoperating a UR5 robotic arm. A wearable haptic armband was implemented to deliver vibrotactile feedback regarding the torque experienced by the arm. A pick and place test was designed to assess the performance in terms of success rate and completion time. Handa et al. [27] implemented a teleoperation system to control a Wonik Allegro bionic hand connected to a Kuka manipulator. The user's hand motion was tracked by a set of four synchronized RealSense cameras. The mapping method aimed at replicating the relative distances between the fingertips. Pick and place tests were conducted to assess the performance, measuring success rate and completion time.

Although vision-based methods overcome typical issues of wearable devices, there are still some bottlenecks that prevent them to be systematically used in real applications [24]. Environmental factors may drastically affect vision-based tracking performance, such as the presence of bright objects, cluttered background, presence of objects with shapes or colors similar to the fingers ones, and camera occlusion. Even in an ideal environment, self-similarities between the fingers and self-occlusion can bring great inaccuracies. Moreover, the user often has a limited workspace. Wrong estimation of human hand pose by vision tracking algorithms may lead to a challenge in ensuring controllability during teleoperation.

1.3. Motivation and objectives

Teleoperation of underactuated bionic hands has proven to be feasible using both wearable tracking devices and vision-based tracking devices. However, usability and accuracy limitations of the two tracking methods respectively, still restrict the application of such teleoperation systems. Moreover, the underactuation nature of bionic hands poses challenges in terms of mathematical modeling, due to the resulting non-linear motion, and mapping, due to an increased kinematic discrepancy, which must be addressed. Therefore, further research is needed to improve the result. On the human hand tracking side, quantitative comparison between different hand tracking methods can be useful as a basis for selection in research. Nevertheless, comparative studies found in the literature

mainly address the evaluation of the accuracy of the motion capture system alone, with respect to a ground truth tracking method [28] [29] [30], without considering the resulting performance of the teleoperation systems based on such methods. Moreover, most of the works related to the development of bionic hand teleoperation systems are oriented to self-validation, which ultimately prevents a meaningful comparison between different works[4]. The evaluation tests to be performed for self-validation and the metrics to be measured are widely agreed upon. Generally, grasping tasks are performed by different users in several conditions to demonstrate the usability of the system. Different kinds of objects, in terms of size and shape, are grasped in order to test if the user interface can exploit the versatility of the bionic hand. Success rate or completion time are usually recorded as performance metrics. However, there are no standards to define the grasping task, thus the comparison between the metrics obtained in different works has little significance. Moreover, such tests address only usability and not the overall accuracy of the teleoperation.

Given such premises, the first main aim of this work is to implement two teleoperation systems, designed to intuitively control the same underactuated bionic hand in a continuous pose space, but based on two different hand motion tracking devices, belonging to the two most competitive categories employed in the teleoperation field, namely the wearable mechanical devices and the contactless vision-based devices. The usability of the implemented systems is demonstrated and quantified through a teleoperation experiment where different users were asked to accomplish some grasping tasks employing the developed systems. The second main objective is to compare the two implemented systems in terms of accuracy, usability and user experience, within the same experimental setup, in order to obtain a meaningful comparison of the two tracking devices when applied to teleoperation. The experimental results can be a useful basis for selection. Furthermore, the implemented evaluation method can be used as a reference to meaningfully compare bionic hand teleoperation systems based on different hand-tracking technologies. Lastly, the construction of the motion model of the utilized bionic hand, employed by the aforementioned teleoperation control systems, can be seen as a third subsidiary contribution of the work. Indeed, such model overcomes the accuracy limitations of previous linear approaches to model the same bionic hand [31], by considering the nonlinear features of the fingers mechanics.

To achieve such goals, four main activities are performed. First of all, the inverse kinematic model of the bionic hand is derived from a calibration procedure performed with an external ground truth tracking system. Then, two different hand tracking systems are implemented using an exoskeleton glove and a Leap Motion Controller, respectively.

Consequently, the two teleoperation systems are developed by establishing a suitable and simple mapping strategy between the capture of human data and the desired motion of the bionic hand and combining the tracking stages with an actuation control stage that is based on the bionic hand inverse kinematic model previously derived. Two teleoperation experiments are performed using a ground truth tracking system to to evaluate and compare the accuracy of the developed systems in static and dynamic scenarios, respectively. Lastly, a user study on practical grasp tasks is conducted to demonstrate the usability of the frameworks and to further compare their performance in terms of success rate and subjective evaluation of the workload.

2 | Materials and methods

2.1. Bionic hand features and modeling

This section introduces the bionic hand utilized in this work. The first part describes the main features characterizing the bionic hand. The second part presents the method used to build the inverse kinematic models of its fingers.

2.1.1. Bionic hand features overview

U-Hand-Pi (Hiwonder Technology Co., China) is an underactuated anthropomorphic robotic hand composed of a rigid case and five articulated fingers, as shown in figure 2.1. The rigid case can be associated to the metacarpal part of the human hand and serves as a container for the five small servo motors that move the fingers. In turn, it is connected to another servomotor that, when attached to a rigid base, can be used to perform a movement corresponding to the human wrist rotation. However, this functionality is never used within this work because the focus is on the local DoF mapping only (fingers motion). Each finger is composed of three rigid links and three hinge joints including one connected to the rigid case. The joints can be associated to the ones of the human hand described in figure 1.1.a. To simplify notation, in all the fingers, including the thumb, the joints are labeled as the ones belonging to longer fingers: metacarpophalangeal (MCP), proximal interphalangeal (PIP) and distal interphalangeal (DIP). Each joint has one DoF, leading to 15 local DoF in the entire hand. However, the underactuation paradigm imposes only one degree of actuation for each finger, with the servomotor directly driving only the MCP joint. The flexion movement of the first link is then passively transmitted to the other two links thanks to two bi-articular metallic bars. The resulting structure, as a first approximation, can be schematized as a composition of two rigid link quadrangles having only one degree of freedom, allowing the finger to do only two kinds of movement: bending on itself or opening. These two antithetical movements will be referred to as “flexion” and “extension” respectively. It should be noted that the underactuation paradigm determines some kinematic differences with respect to the human finger: the

robotic finger cannot decompose the flexion/extension movement into three independent sub-movements (one for each joint), nor it can perform abduction/adduction around the MCP joint.

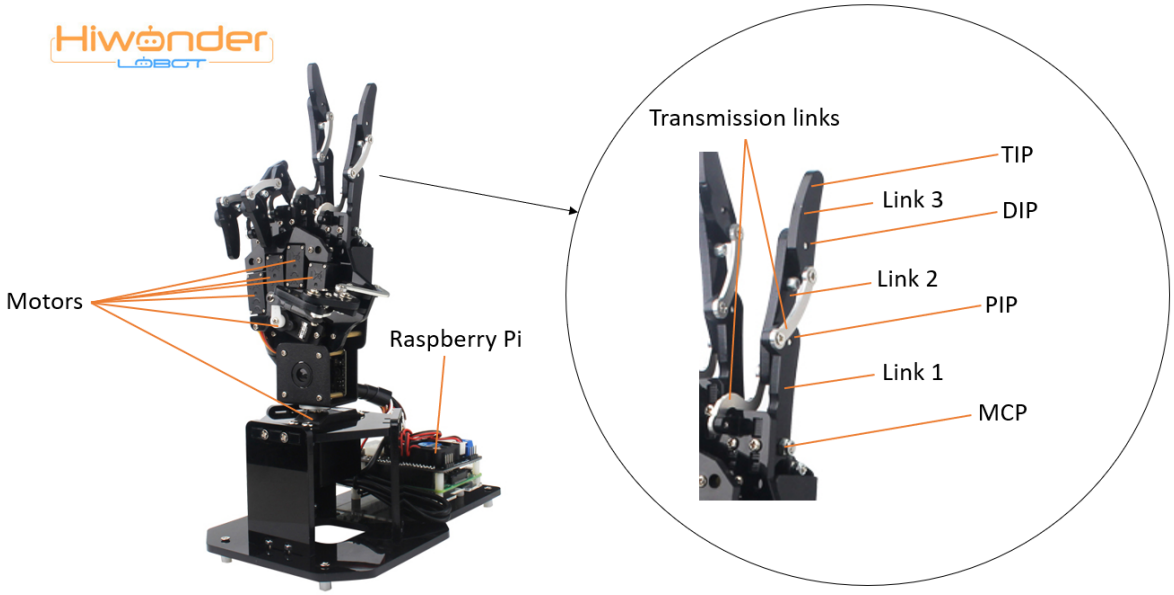


Figure 2.1: Hiwonder U-Hand Pi with index finger details

The motors are Hiwonder LFD-01 servos (Hiwonder Technology Co., China) with a built-in anti-blocking mechanism. Each motor is controlled by a Pulse Width Modulation signal, which is a periodic square wave with a fixed period and a variable width of the switch-on sub-period (pulse width). The higher the pulse width, the higher the current average value sent to the motor, which in turn encodes the rotation of the servo motor axis. The pulse width of the signal delivered to the LFD-01 ranges from 500 μs (PW_{min}) to 2500 μs (PW_{max}), corresponding to 180° and 0° rotations of the motor, respectively. However, given the structure of the hand, the motors cannot fulfill their entire rotation range and the true range is unknown. Motors input is driven by a Raspberry Pi unit (Raspberry Pi Ltd., England) endowed with both a WI-FI module and an Ethernet module, which can be used to access the Raspberry desktop from a remote computer through a remote access software. Raspberry can run a desktop user interface application that can be used to manually set the input value of each motor in real-time. Motor input values can be also set from a remote computer by sending a specific message to Raspberry using TCP/IP protocol, without passing from the desktop application.

2.1.2. Inverse kinematics modeling

Problem statement The inverse kinematics of a bionic hand finger is the mathematical relationship between the desired finger pose and the input value of the corresponding motor, in terms of PWM signal width (PW), that should bring to a resulting finger pose equal to the desired one. In order to transfer the human fingers motion to the robotic fingers, in the context of teleoperation, the inverse kinematics of each finger had to be modeled. Prior to the models synthesis, one independent kinematic variable was identified for each model to exhaustively represent the finger pose. Given that each finger has only one DoF, one variable for each finger was enough. In particular, as shown by the example in figure 2.2 for the middle finger, the angle described by a virtual link that goes from the MCP joint to the fingertip (TIP) was selected as the independent variable, because it is representative of the whole finger pose. Such angle was referred to as θ and called "bending angle", given that as it increases, the finger bends on itself. The zero-reference of the bending angle corresponds to the pose of maximum extension in order to have a monotonic relationship with the motor input.

Figure 2.2: Definition of bionic hand finger bending angle, θ, for the middle finger with respect to the related maximum extension pose

In a previous work [31], a direct linear function was used to define the inverse kinematics of the finger, in terms of relationship between the PWM signal width and the aforementioned bending angle:

$$PW = PW_{max} - \frac{\theta}{\theta_{max}} \cdot (PW_{max} - PW_{min}) \tag{2.1}$$

where θ_{max} is the maximum bending angle reachable by the finger. However, such method is insufficiently accurate because it overlooks some critical issues: (1) motor saturation prevents covering the whole PW range, introducing nonlinearity; (2) non-rigid phenomena caused by the non-idealities of the joints, such as compliance, friction and backlash, change the sensitivity of the finger to the motor input, causing further nonlinearity; (3) the nonlinear features depend on whether the motion is flexion or extension, thus introducing hysteresis. In order to synthesize the models, a data-driven approach was preferred to capture such nonlinearity. Following this approach, an accurate calibration procedure was carried out for each finger separately, consisting of acquiring kinematic data related to finger motion and fitting a mathematical model on the obtained data, capable of capturing the nonlinear behavior of the finger.

Calibration setup The complete experimental setup for the acquisition of calibration data is depicted in figure 2.3. The bionic hand was fixed on a table and an optical measurement system was set up to acquire the bending angles of its fingers, namely an Optotrak Certus System (Norhtern Digital Inc, United States). Such measuring system is composed of a position sensor, a set of active markers, a markers strober, a control unit, and a desktop application (NDI First Principle) that can be executed on Windows.

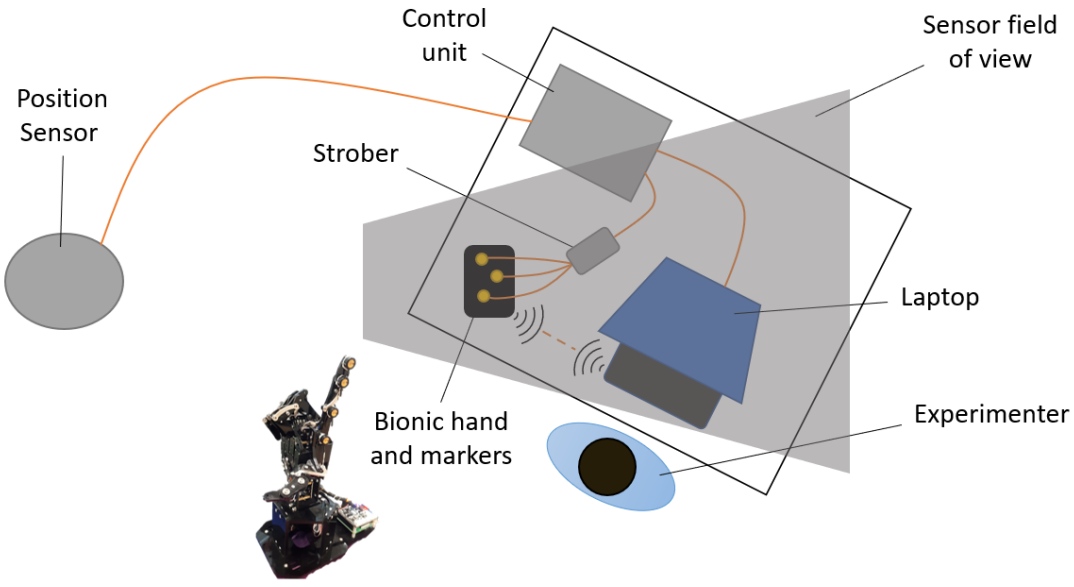


Figure 2.3: View from above of the experimental setup used for the inverse kinematics calibration

The position sensor consists of a set of three infrared stereoscopic cameras piled in a columnar rigid case 1.5 m long (figure 2.5.a). It can measure the 3D coordinates of

active markers with 0.1 mm accuracy and 0.01 mm resolution. All the markers must be within the field of view, depicted in figure 2.4, and clearly visible from each camera. The measured coordinates are referred to a default global coordinate system, shown in figure 2.4, which is determined at the factory and has its origin at the center of the middle camera. During the calibration experiment, the position sensor was fixed at about 2m from the bionic hand. Relative motion between the hand and the sensor was not present.

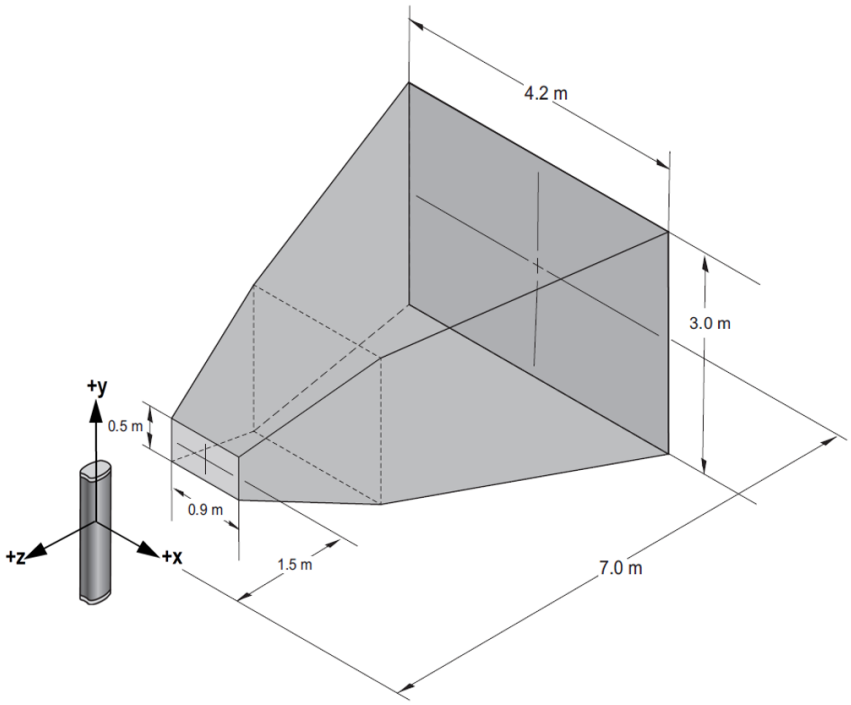


Figure 2.4: Field of view and reference system of the NDI Optotrak Certus position sensor

Markers consist of a near-infrared light-emitting diode, 11 mm large, and are connected to the strober through flexible cables. The strober activates markers sequentially according to the timing and power settings specified inside the desktop application. When the marker is activated, its position is measured by the sensor. Activation timing is determined by the frame frequency, which sets the number of recordings for each marker in 1 s, and by the marker frequency, which sets the recording period for each marker. Each finger was calibrated separately, thus markers were moved each time from one finger to the other to minimize the number of markers placed on the hand at the same time. Even though only TIP and MCP positions are required to compute the bending angle, also DIP and PIP positions were acquired to obtain a complete set of motion data that may be useful for future works. At first, one optical marker was placed on the side of each joint, as shown in figure 2.5.c for the index finger. However, due to mechanical occlusion of the hand

structure, the position of the MCP joints was difficult to be directly measured through markers. For such reason, MCP joint marker was removed (figure 2.5.b shows the markers positioning on the thumb after the MCP marker removal) and MCP joint position was measured using a pointing probe endowed with two aligned markers, labeled as A and B. For example, figure 2.5.d. shows the pointing probe while measuring the MCP joint position of the middle finger.

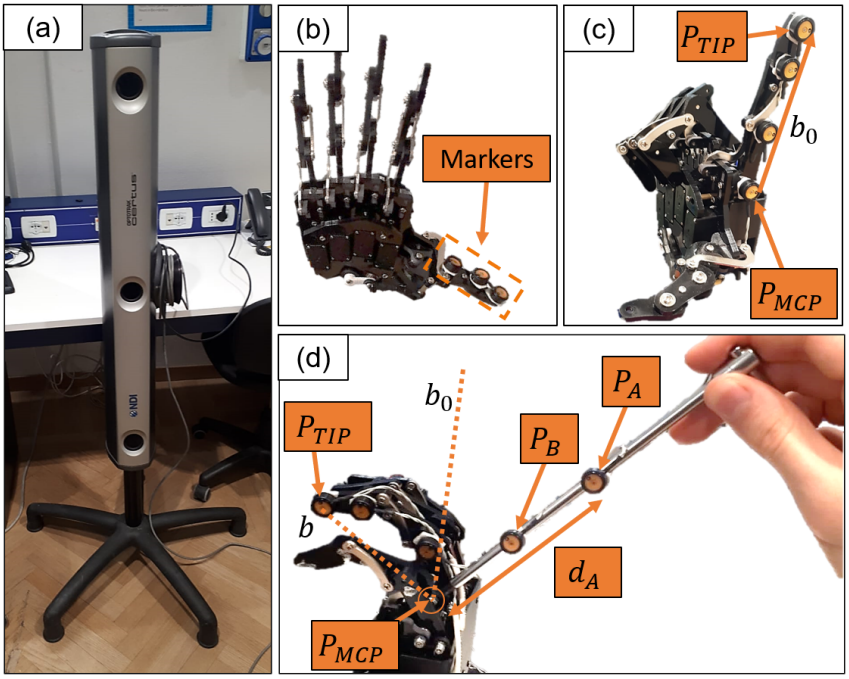


Figure 2.5: Images from the experimental setup for inverse kinematics calibration. (a) Optical position sensor. (b) Position of markers on the thumb (MCP joint excluded). (c) Position of markers on the index (MCP joint included) and definition of the bending vector b_0 corresponding to the maximum extension pose. (d) Probe pointing to the MCP joint of the middle finger.

The position of the two probe markers, namely $\boldsymbol{P_A}$ and $\boldsymbol{P_B}$ respectively, are measured while the probe points to the MCP joint of the finger. Then, by knowing the distance d_A from A to the probe tip, the joint position $\boldsymbol{P_{MCP}}$ can be calculated as follows:

$$\boldsymbol{P_{MCP}} = \boldsymbol{P_A} + \frac{\boldsymbol{P_B} - \boldsymbol{P_A}}{|\boldsymbol{P_B} - \boldsymbol{P_A}|} \cdot d_A \tag{2.2}$$

By knowing the MCP joint position, $\boldsymbol{P_{MCP}}$, and the fingertip position measured by the corresponding marker, $\boldsymbol{P_{TIP}}$, the bending vector $\boldsymbol{b}$ (figure 2.5.d) from the MCP joint to the fingertip can be computed.

$$\boldsymbol{b} = \boldsymbol{P_{TIP}} - \boldsymbol{P_{MCP}} \tag{2.3}$$

Then, given the bending vector corresponding to the maximum extension pose of the finger, $\boldsymbol{b}_0$, that is measured using the markers as well, as shown in figure 2.5.c for the index, the bending angle θ can be expressed as:

$$\theta = \text{acos}\left(\frac{\boldsymbol{b} \cdot \boldsymbol{b}_0}{|\boldsymbol{b}|\,|\boldsymbol{b}_0|}\right) \tag{2.4}$$

As depicted in figure 2.3, the control unit of the measuring system was an essential part of the setup, as it drives the strober, receives the raw position data from the sensor, processes them, and sends them to the host computer where they are handled by the software application. The host computer was used to manually control the bionic hand finger motor values during the calibration through a WI-FI connection using the Raspberry desktop application.

Data acquisition For each finger, four complete calibration cycles were recorded. In each cycle, 41 values equally distributed along the motor input range were delivered sequentially, first in ascending order (extension), then in descending order (flexion). For each value, once the finger reached a stationary pose, the coordinates of the markers were recorded. Given that the bionic hand did not move during the data acquisition, the positions of the MCP joints were measured only once at the beginning of the calibration. The resulting bending angle samples for each finger, associated with the corresponding motor input value, are shown in figure 2.6. Thumb is the only finger in which the PW signal has a positive correlation with bending angle (increasing PW causes angle increasing), in all the other cases the correlation is negative (increasing PW causes angle decreasing). Each finger has two saturation regions that start around 1000 and 2000 respectively, in line with the fact that the range of the servo motors cannot be fulfilled due to the hand structure. The extension characteristic appears to be linear, at least in the central region. For all the fingers a significant non-linearity associated with hysteresis arises in the flexion direction. For the thumb, this phenomenon regards all the operative region, while for the other fingers it affects only the poses characterized by little flexion.

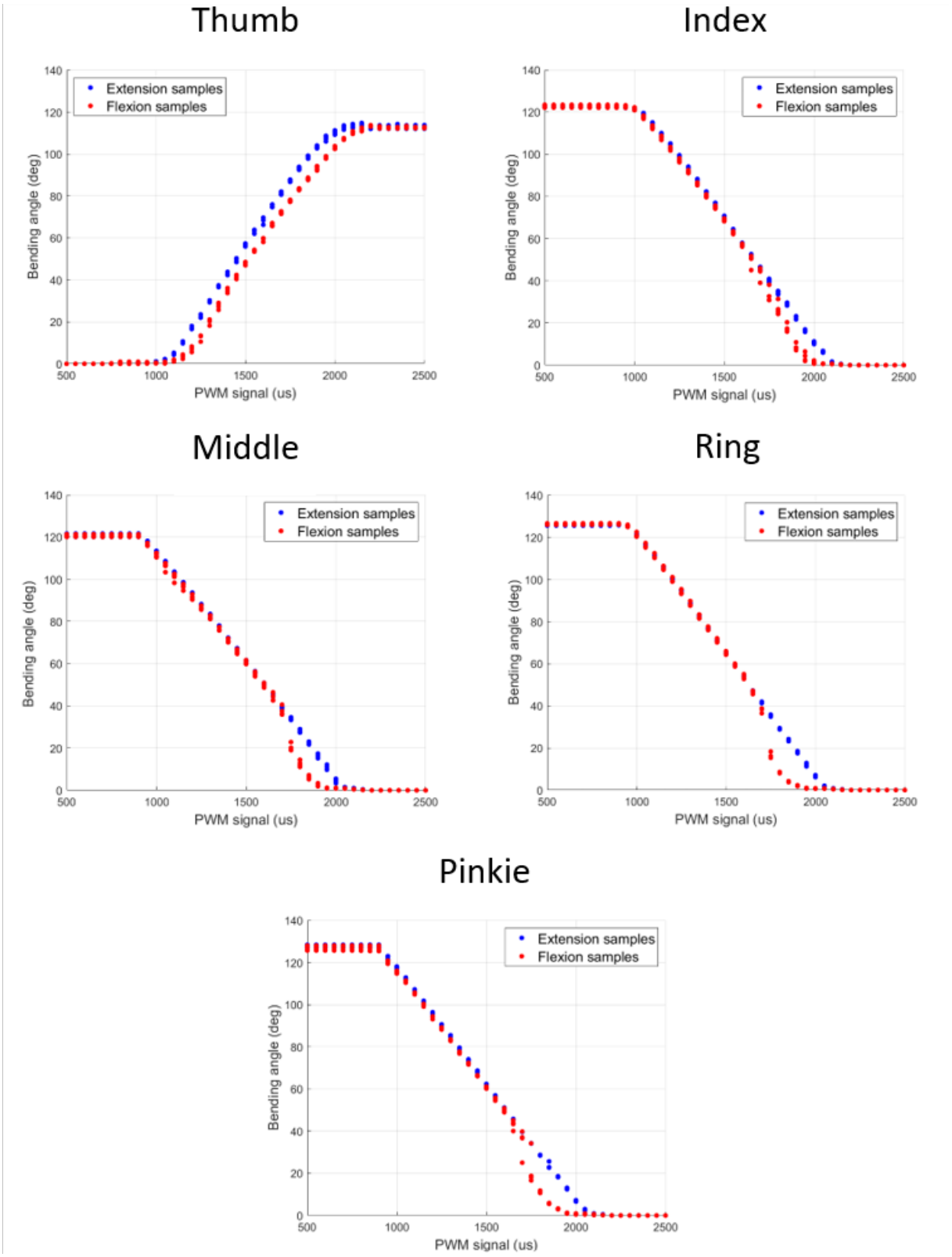


Figure 2.6: Calibration cycles acquired for inverse kinematics modeling

Model synthesis Neural networks were utilized to extract an inverse kinematic model for each finger from the acquired calibration cycles, in the form of a couple of functions f_{ext} and f_{flex} which model the extension and flexion behaviors respectively. In particular, they give the motor input value (PW) as a function of the desired bending angle θ:

$$\begin{cases} PW = f_{ext}(\theta) \text{ , during extension} \\ PW = f_{flex}(\theta) \text{ , during flexion} \end{cases} \tag{2.5}$$

The generic structure of a neural network able to learn an arbitrary function expression from training data is depicted in figure 2.7 [32].

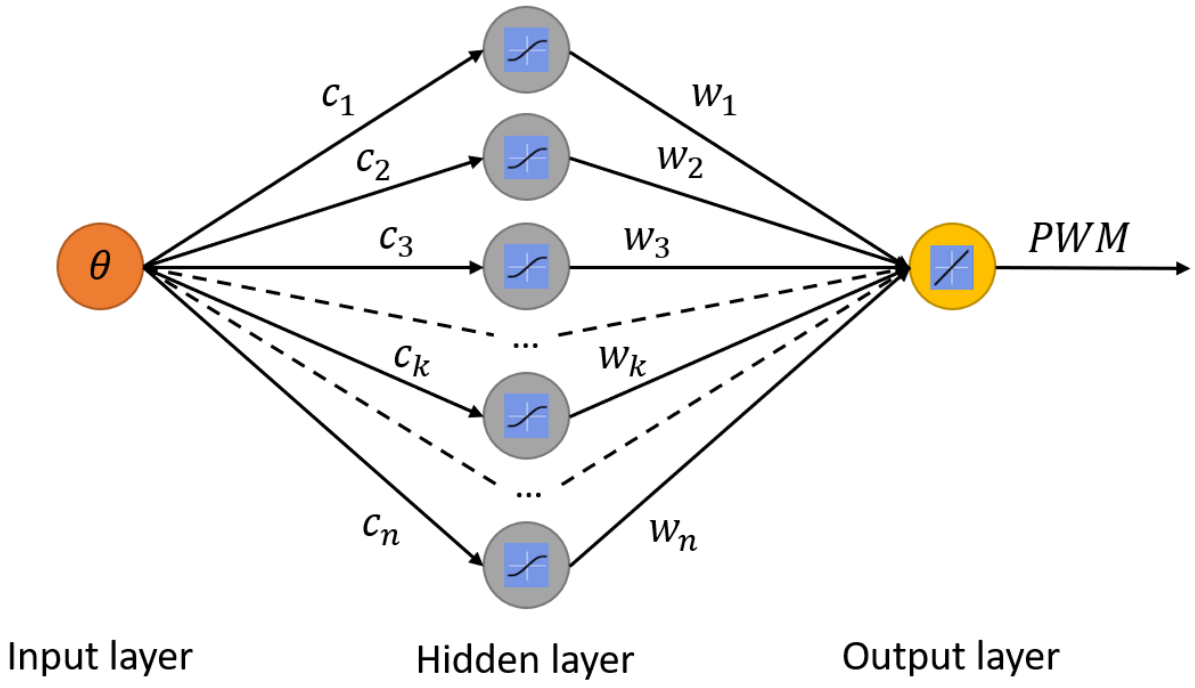


Figure 2.7: General neural network structure for functions approximation

It is composed by one hidden layer comprising n neurons, one input neuron and one output neuron. The independent variable of the function (θ in our case) is given as input to all the neurons of the hidden layer. Each k^{th} (k=1,...,n) hidden neuron weights the input for a weight c_k and subtracts a threshold T_k to it, obtaining the neuron potential P_k. Then it applies the hyperbolic tangent function to obtain the output u_k.

$$P_k = c_k\theta - T_k \tag{2.6}$$

$$u_k = \tanh(P_k) = \frac{e^{P_k} - e^{-P_k}}{e^{P_k} + e^{-P_k}} \tag{2.7}$$

The results from all the hidden neurons are weighted for a value w_k and summed by the output neuron. The result is the output of the network (PWM value in our case).

$$PWM = \sum_{k=1}^{n} w_k u_k \tag{2.8}$$

The number of hidden neurons is an important hyper-parameter to be set. Indeed, by increasing it, the fitting on the training data will be better, but if it is too large, over-fitting may arise from the high complexity of the resulting function which loses generality. For each curve to be found, 7 networks were trained, each one having a different number of hidden neurons, ranging from 2 to 8. This range was chosen after some preliminary tests. Saturation samples were excluded from the data set. The network which achieved the best performance in terms of mean squared error on the validation set was selected. The domain of the resulting functions was defined from zero to the average angle among the recordings corresponding to the minimum motor input (maximum bending angle).

2.2. Wearable glove - based motion tracking

In this section, the implemented wearable glove-based tracking system is presented. Firstly, the main features of the sensorized tracking glove and its readout signals are described. Then, the real-time processing applied to the raw motion readout of the glove is explained and the resulting human motion features are defined. Lastly, the detailed implementation of a suitable Kalman filter to smooth the tracking stage output is shown.

2.2.1. Glove structure and functioning

A Hiwonder Glove (figure 2.8, Hiwonder Technology Co., China) was employed to design the teleoperation system based on wearable tracking technology. In particular, such glove can be classified as an exoskeleton glove. It is composed of a rigid structure that must be placed on the user's hand back and fastened around by two Velcro laces. A printed circuit board is attached to the rigid structure. The most significant board components are an ATmega328P microcontroller, a Bluetooth module, five potentiometers, a gyroscope acceleration sensor, an ON/OFF switch, a battery unit and a USB port which can be also used to power the board. Each potentiometer is connected to one finger through a

kinematic chain that must be fastened around the first phalanx of the finger going from the MCP joint to the PIP joint by a Velcro lace. This mechanism is able to convert the flexion angle φ of the first phalanx of the finger, defined in figure 2.9, into a proportional rotation and voltage readout V of the sensor. The range of φ allowed by the chain structure is around 40°.

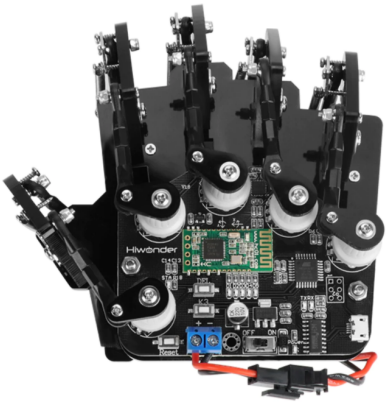


Figure 2.8: Top view of Hiwonder Glove

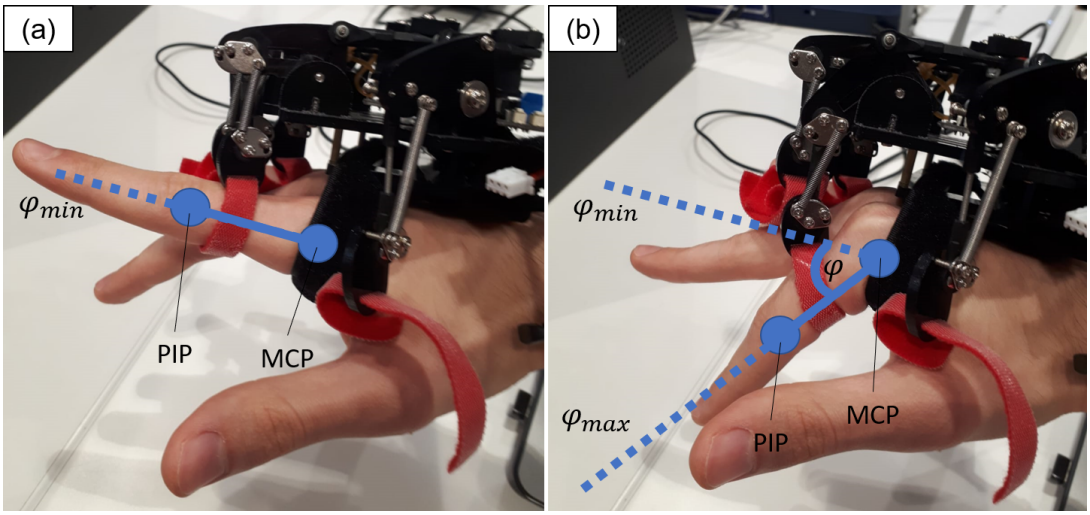


Figure 2.9: Definition of flexion angle: minimum (a) and maximum (b) flexion poses

The data acquisition firmware is developed in Arduino environment and it is made accessible and customizable by the producer. Basically, it implements two main routines. The first is a calibration phase that runs whenever the glove is switched on. During this phase, the user has to flex and extend all the fingers one time covering the whole range of motion allowed by the kinematic chain. For each finger, the CPU records the maximum and minimum voltages V_{max} and V_{min} reached by the potentiometer, which corresponds

to the minimum and maximum finger flexion angles φ_{min} and φ_{max} reached by the finger respectively and depicted in figure 2.9. The second routine is a measurement loop that starts automatically after the calibration and continuously samples the output of the sensors. The loop repetition frequency is around 250Hz. For each potentiometer, the voltage V is read and scaled into the range 500-2500 considering the previously calibrated voltage range:

$$S = \frac{(V - V_{\text{min}})}{V_{\text{max}} - V_{\text{min}}} \cdot 2000 + 500 \tag{2.9}$$

By considering φ_{min} equal to zero, the resulting signal readout S has the following linear relationship with respect to the flexion angle φ:

$$S = (1 - \frac{\varphi}{\varphi_{max}}) \cdot 2000 + 500 \tag{2.10}$$

2.2.2. Real-time motion data processing

The complete framework of the wearable glove-based motion tracking system is shown in figure 2.10. The glove sends its readout to a computer that runs Ubuntu 16.04 and Robot Operating system (ROS [33]) Kinetic via USB connection. The communication between the glove and the ROS network was made possible by adding a ROS library and a ROS publisher inside the data acquisition firmware of the glove. Rosserial interface package is employed to make the ROS network handle the serial communication with the external glove device.

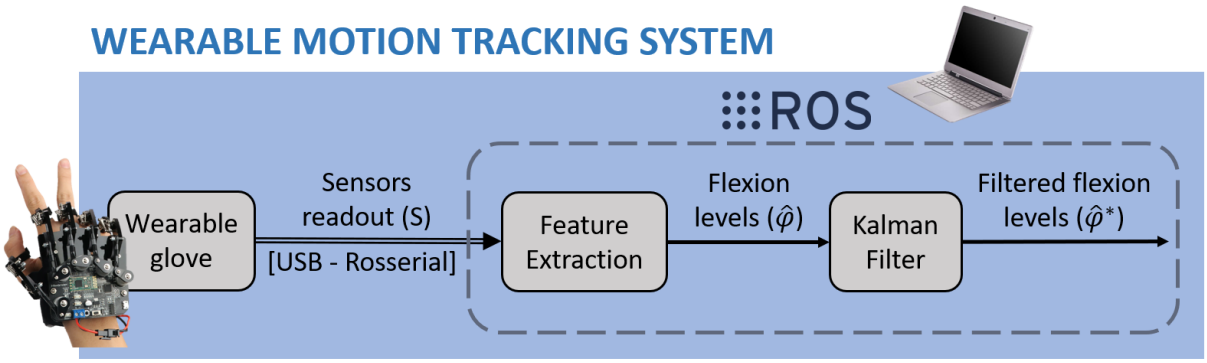


Figure 2.10: Components of the wearable glove - based motion tracking system

The role of the ROS network is to process the raw glove readout in order to extract the human motion features to be mapped into the five robot fingers bending angles when

the tracking system is applied to remotely control the bionic hand. In particular, for each finger, the feature extraction node of the ROS network samples the signal readout S at 40Hz and computes the so-called "flexion level", $\hat{\varphi}$. The sampling frequency was selected considering a quality margin from the maximum signal frequency that U-Hand Pi bionic hand can handle which is 50 Hz. The flexion level is a dimensionless quantity, ranging from zero to one, that concisely represents the human finger motion and is easily and intuitively mappable to the bionic hand finger motion, as described in section 2.4. Considering the relationship represented by equation 2.10 between the finger flexion angle φ and the potentiometer readout S, the flexion level is calculated as follows:

$$\hat{\varphi} = 1 - \frac{S - 500}{2000} = \frac{\varphi}{\varphi_{\max}} \tag{2.11}$$

As a final step of the processing stage, a real-time linear Kalman filter [34] is applied to smooth each flexion level, thus obtaining the filtered flexion level, $\hat{\varphi}^*$. The implementation details are discussed in the next section.

2.2.3. Kalman filter implementation

In order to apply Kalman filtering to each flexion level, a state equation-based representation of the finger motion was conceived. In line with a previous work [31], the linearized motion model of the variable to be filtered, namely the flexion level, was adopted:

$$\begin{cases} \hat{\varphi}_k = \hat{\varphi}_{k-1} + T_S\dot{\hat{\varphi}}_{k-1} + \dfrac{T_S^2}{2}\ddot{\hat{\varphi}}_{k-1} \\ \dot{\hat{\varphi}}_k = \dot{\hat{\varphi}}_{k-1} + T_s\ddot{\hat{\varphi}}_{k-1} \end{cases} \tag{2.12}$$

Where $\hat{\varphi}_k$ and $\hat{\varphi}_{k-1}$ are the flexion level at time k and time $k-1$, respectively, $\dot{\hat{\varphi}}_k$ and $\dot{\hat{\varphi}}_{k-1}$ are the first degree time derivative of the flexion level at time k and time $k-1$, respectively, $\ddot{\hat{\varphi}}_{k-1}$ is the second degree time derivative of the flexion level at time $k-1$, T_s is the sampling period. By approximating $\ddot{\hat{\varphi}}$ as a white gaussian noise (WGN) w with zero mean and σ_w variance, and considering $\hat{\varphi}$ and $\dot{\hat{\varphi}}$ as state variables, the following state stochastic equation can be derived:

$$\boldsymbol{x}_k = \boldsymbol{\Phi}\boldsymbol{x}_{k-1} + \boldsymbol{\Gamma}w_{k-1} \tag{2.13}$$

where:

$$\boldsymbol{x} = \begin{bmatrix} \hat{\varphi} \\ \dot{\hat{\varphi}} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad , \quad \boldsymbol{\Phi} = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix} \quad , \quad \boldsymbol{\Gamma} = \begin{bmatrix} \frac{T_s^2}{2} \\ T_s \end{bmatrix} \tag{2.14}$$

Ideally, the flexion level computed by the feature extraction node, should be equal to the flexion level state variable, x_1. Nevertheless, the measurement is affected by random fluctuations that can be modeled by an additive WGN v with zero mean and σ_v variance that roughens the signal. Hence, the measurement stochastic equation can be defined as follows:

$$y_k = \boldsymbol{H}\boldsymbol{x}_k + v_k \tag{2.15}$$

where y_k is the flexion level sample measured by the system at time k and:

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{2.16}$$

Equations 2.13 and 2.15 are used by the Kalman filter algorithm to smooth the flexion level signal. The algorithm acts recursively on each measurement sample and can be divided into a prediction phase and an update phase. During the prediction step, the state vector value at time k, $\widehat{\boldsymbol{x}}_{k|k-1}$, along with its covariance matrix, $\boldsymbol{P}_{k|k-1}$, are predicted according their current estimation values, $\widehat{\boldsymbol{x}}_{k-1|k-1}$ and $\boldsymbol{P}_{k-1|k-1}$ respectively, and to the state equation parameters:

$$\widehat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{\Phi}\widehat{\boldsymbol{x}}_{k-1|k-1} \tag{2.17}$$

$$\boldsymbol{P}_{k|k-1} = \boldsymbol{\Phi}\boldsymbol{P}_{k-1|k-1}\boldsymbol{\Phi}^T + \sigma_w \tag{2.18}$$

Consequently, the flexion level measurement at time k is predicted according to the measurement equation:

$$\widehat{y}_k = \boldsymbol{H}\widehat{\boldsymbol{x}}_{k|k-1} \tag{2.19}$$

In the update step, the actual measured flexion level y_k, sampled at time k, is compared to the predicted one, and the difference is added to the state vector prediction, weighted for the Kalman gain, K_k, defined in equation 2.20. Hence, state vector and state covariance

matrix estimates are updated to $\widehat{\boldsymbol{x}}_{k|k}$ and $\boldsymbol{P}_{k|k}$, respectively:

$$K_k = \boldsymbol{P}_{k|k-1}\boldsymbol{H}^T \left(\boldsymbol{H}\boldsymbol{P}_{k|k-1}\boldsymbol{H}^T + \sigma_v\right)^{-1} \tag{2.20}$$

$$\widehat{\boldsymbol{x}}_{k|k} = \widehat{\boldsymbol{x}}_{k|k-1} + K_k \left(y_k - \widehat{y}_k\right) \tag{2.21}$$

$$\boldsymbol{P}_{k|k} = \left(\boldsymbol{I} - K_k\boldsymbol{H}\right)\boldsymbol{P}_{k|k-1} \tag{2.22}$$

The updated estimate of the flexion level state variable, $\widehat{\boldsymbol{x}}_{k|k}$, stands for the filtered flexion level, $\hat{\varphi}^*$, and is published as the final output of the tracking system.

2.3. Vision-based motion tracking

In this section, the implemented contactless vision-based tracking system is presented. Firstly, the main features of the utilized camera and of the related company software for real-time image processing are described. Then, the further processing stage that was specifically developed for this application to extract the driving human motion features is analyzed.

2.3.1. Leap Motion structure and functioning

A Leap Motion Controller (LMC, Ultraleap, California, United States) was chosen as representative of vision-based hand tracking devices. Figure 2.11 depicts the size and the internal components of such device. It is composed of a rigid case 80 x 30 x 11.3 mm large and 32g weighing endowed with an anti-slip pad on the bottom surface. On the upper surface, there are three near-infrared (IR) light-emitting diodes (LEDs) that illuminate the human hand and two stereo IR cameras that measure the reflected light and capture two grayscale images. The field of view of the cameras extends for 140° x 120° and from 10cm to 60cm far from the device, which must be fixed on a plane surface facing the human hand palm. The couples of images are acquired at 120Hz and sent via USB connection to a computer that must run the Leap Motion Service (LMS), which is a non-customizable company software for image processing. After removing background objects and environmental light, the LMS reconstructs a 3D representation of the hand surface by applying the principles of stereo vision. Depth information is obtained by measuring the disparity between the two images, caused by the distance between the two

slightly different points of view. Disparity values are inversely proportional to the depth of the corresponding pixel location.

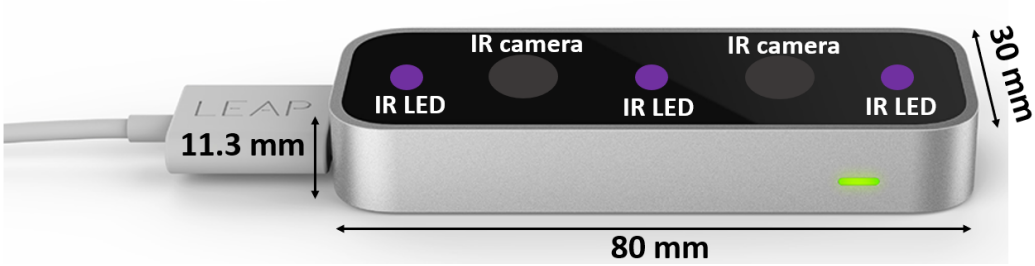


Figure 2.11: Size and internal components of the Leap Motion Controller device

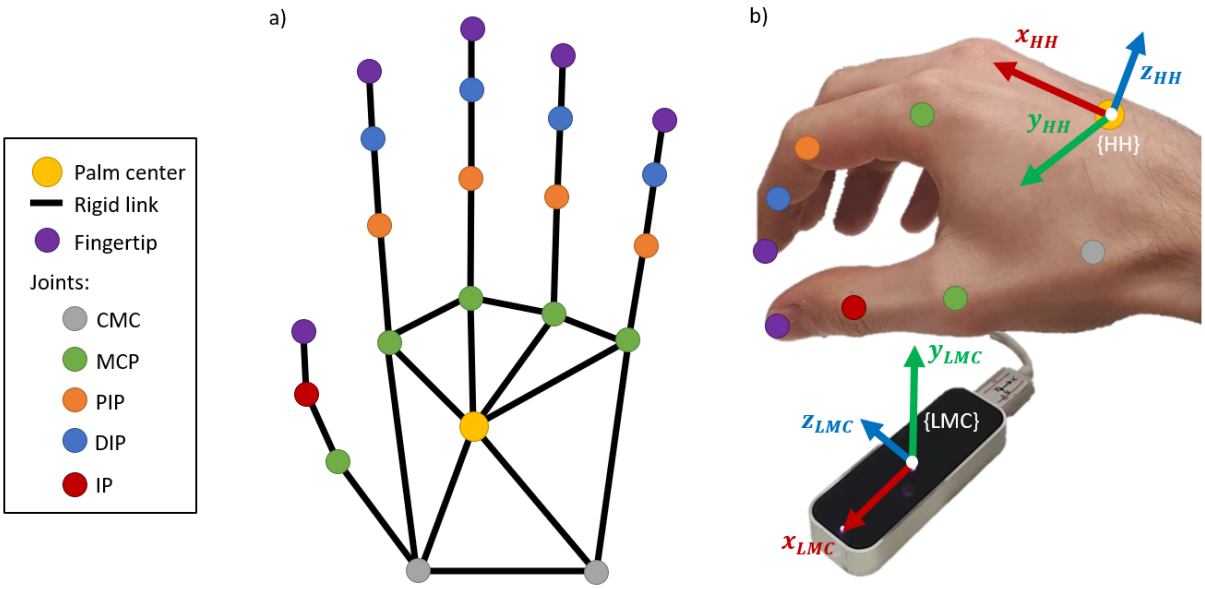


Figure 2.12: Leap Motion Controller readout. (a) Employed human hand model. (b) Coordinates reference frames attached to the camera device, {LMC}, and to the hand, {HH}, respectively.

Next, a tracking layer extracts the 3D model of the hand depicted in figure 2.12.a, which specifies the 3D coordinates of fingers joints and tips with respect to a reference frame attached to the LMC device, labeled as $\{LMC\}$ and shown in figure 2.12.b. The algorithm can infer with limited accuracy the positions of occluded parts of the hand. In such a hand model, the joints and the tip belonging to the same finger are connected by rigid links with a fixed length equal to the estimated length of the corresponding bone, as shown in figure 2.12.a. The MCP joints of the index, middle, ring and pinkie, together with the CMC joints of thumb and pinkie do not move with respect to each other and form a rigid

hexagonal structure that reproduces the hand profile. Furthermore, the model defines a point called palm center, which is rigidly connected to the hand hexagonal profile and is used as the origin of a local reference frame attached to the hand, labelled as {HH}. As shown in figure 2.12.b, such frame has the x-axis directed towards the MCP joint of the middle finger, the y axis lying on the palm plane and directed towards the thumb, the z-axis obtained with the right-hand rule. The LMS directly provides the unit vectors of {HH} axes with respect to {LMC} frame, which can be used, together with the palm center, as transformation parameters to pass from one frame to the other.

2.3.2. Real-time motion data processing

The complete framework of the implemented vision-based motion tracking system is shown in figure 2.13. The LMC sends via USB connection the two images to a computer that runs Ubuntu 16.04, the LMS and ROS Kinetic.

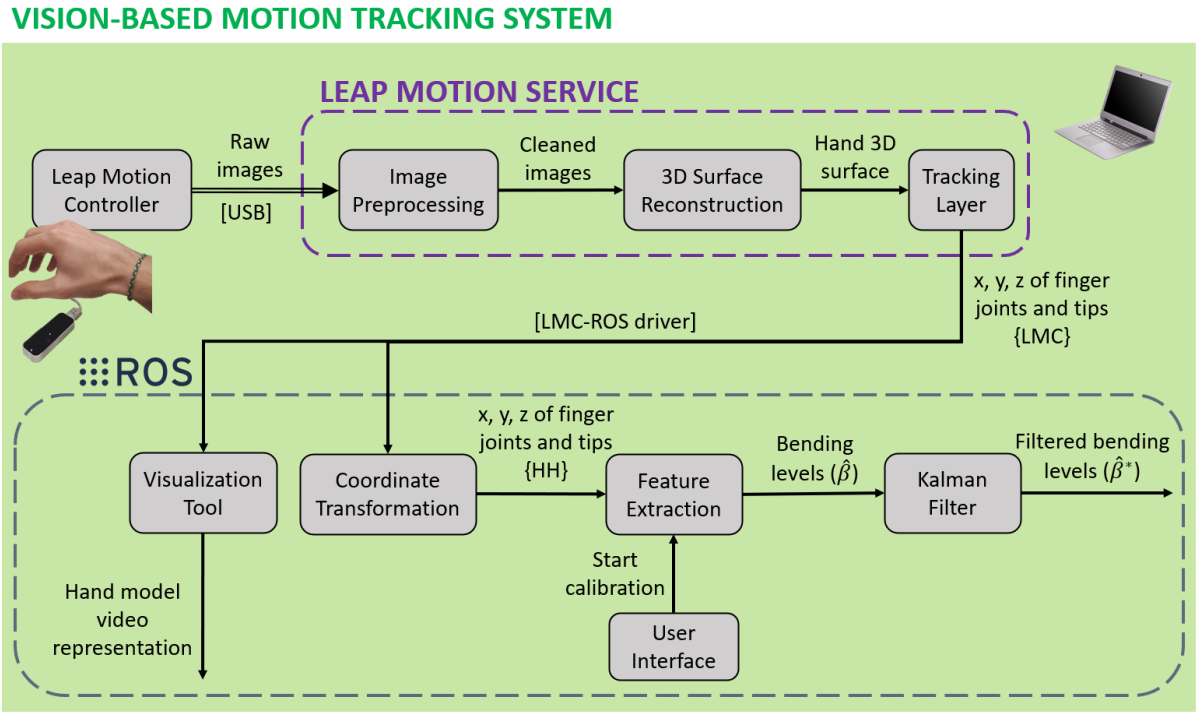


Figure 2.13: Hardware components and software nodes of the vision motion tracking system

The LMS handles the serial communication and computes the raw motion data from the received pair of images as described in section 2.3.1. The readout, consisting of fingers joints and tips coordinates with respect to the {LMC} frame, is published into the ROS network by a Leap Motion – ROS interface driver which is implemented by an open-source

package available on GitHub. Besides the publishing, such package implements a basic visualization tool to observe the human hand model evolution in real-time using RViz. The main role of the ROS network is to process the LMS readout in order to extract the human motion features to be mapped into the five robot fingers bending angles when the vision-based tracking system is applied to remotely control the bionic hand.

Coordinate transformation Given that the goal of this work focuses on the remote control of the bionic hand local DoF only, a prior step to extract the motion features is to refer the fingers joints and tips positions to the local reference frame {HH} in order to single out the motion due to the local DoF of the human hand. Hence, the coordinate transformation node of the ROS network samples the LMS readout at 40Hz. The sampling frequency was selected considering a quality margin from the maximum signal frequency that U-Hand Pi bionic hand can handle, that is 50 Hz. Then, such node transforms joints and tips coordinates from {LMC} to {HH}, considering the geometric elements depicted in figure 2.14.

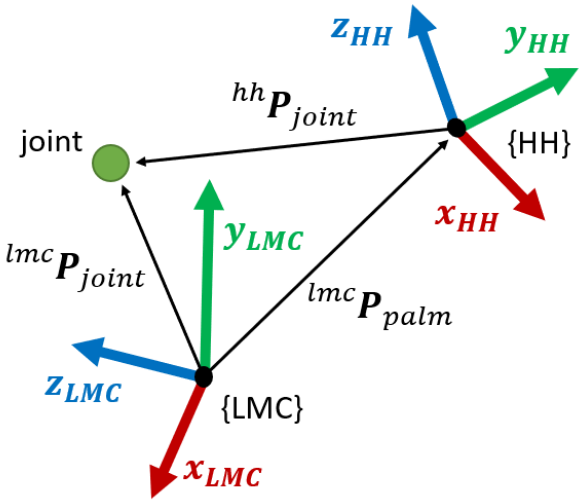


Figure 2.14: Vision-based tracking method: geometric elements used to transform fingers joints and tips coordinates.

The position of the the local frame origin with respect to the LMC frame, ${}^{lmc}\boldsymbol{P}_{palm}$, and the column unit vectors of the local frame axes with respect to the LMC frame, $\widehat{\boldsymbol{x}}_{HH}$, $\widehat{\boldsymbol{y}}_{HH}$ and $\widehat{\boldsymbol{z}}_{HH}$ are used to define the rotation matrix ${}^{hh}_{lmc}\boldsymbol{R}$ and the translation vector ${}^{hh}_{lmc}\boldsymbol{Q}$ that describe {LMC} with respect to {HH}:

$$ {}^{hh}_{lmc}\boldsymbol{R} = \begin{bmatrix} \widehat{\boldsymbol{x}}_{HH} & \widehat{\boldsymbol{y}}_{HH} & \widehat{\boldsymbol{z}}_{HH} \end{bmatrix}^{T} \tag{2.23} $$

$$ {}^{hh}_{lmc}\boldsymbol{Q} = -{}^{lmc}_{hh}\boldsymbol{R}^{T} \cdot {}^{lmc}\boldsymbol{P}_{palm} \tag{2.24} $$

The resulting transformation terms are applied to each joint position referred to {LMC}, ${}^{lmc}\boldsymbol{P}_{joint}$, in order to carry out the coordinate transformation, obtaining the joint position referred to {HH}, ${}^{hh}\boldsymbol{P}_{joint}$:

$$ {}^{hh}\boldsymbol{P}_{joint} = {}^{hh}_{lmc}\boldsymbol{R} \cdot {}^{lmc}\boldsymbol{P}_{joint} + {}^{hh}_{lmc}\boldsymbol{Q} \tag{2.25} $$

Definition of human finger bending angle The transformed coordinates of joints and fingertips are used by the feature extraction node to compute the human hand motion features to be mapped to the bionic hand ones in the context of remote control. The completeness of the data provided by the LMC device can be exploited to extract motion features capable of increasing the similarity between the human hand pose and the bionic hand pose, in comparison to the glove-based system, which instead refers to a poorer human hand model that only considers the first phalanxes of the fingers. Of course, the underactuation of the bionic hand limits the pose similarity, regardless of the completeness of motion tracking data, because of the higher DoF of the human finger compared to the bionic one. In order to exploit the richness of tracking data, taking into account the underactuation of the bionic hand and making a simple mapping possible, one feature for each finger, similar to the one used in section 2.1.2 to describe the robotic finger movement, is extracted, namely the finger bending angle. The definition of such angle in the case of the human hand is slightly different compared to the one that suits the bionic hand, considering the kinematic discrepancies. In particular, given that the robotic finger MCP joint has only one DOF, which is dedicated to flexion, the robotic bending angle, defined as the angle described by the MCP-to-TIP vector with respect to a maximum extension pose, depends only on the MCP joint flexion motion. In humans instead, the MCP joint has two DoF, one dedicated to flexion and one to abduction. Therefore, defining the bending angle as described above would make it prone to be influenced also by MCP joint abduction movements. In other words, robot MCP joint flexion would be sensitive to human MCP abduction, thus causing a counter-intuitive control. To reduce the sensitivity to abduction, in the human case the bending angle β is described by the projection of the MCP-to-TIP vector on the plane where fingertip, PIP joint and MCP joint of the corresponding finger lie when MCP joint abduction is null. Such plane is labeled as a plane of flexion and is fixed with respect to the hand frame. β is defined as the angle between the aforementioned projection vector and another vector belonging to the plane of flexion, which corresponds to the minimum bending pose and the minimum

bending angle β_{min}. The geometric representation of such definition is shown in figure 2.15.

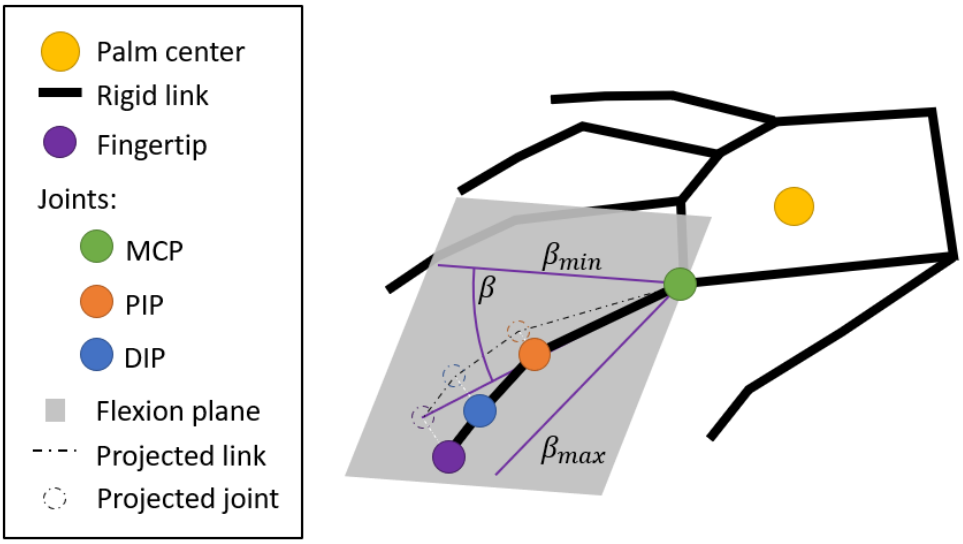


Figure 2.15: Vision-based tracking method: geometric definition of the human bending angle. The example depicted in the image regards the index finger.

Calibration of the tracking system In order to measure such bending angle, a calibration phase must be performed when the tracking system starts, to obtain for each finger, separately, the plane of flexion, the minimum bending pose and the maximum bending angle. The implemented calibration routine is divided into two phases, both accomplished by the feature extraction node. In the first phase, the user has to put each finger in the desired minimum bending pose, represented in figure 2.16.a, then enter "m" in the computer terminal through the keyboard. A user interface node handles such external interaction and makes the feature extraction node read one data frame from the coordinate transformation node and carry out the first calibration step on the measured data. For each finger, a specific reference frame $\{ff\}$ is defined. As shown in figure 2.16.a, such reference system has the origin in the measured MCP joint position, the x-axis on the direction of the measured MCP-to-TIP vector, the z-axis perpendicular to the plane defined by MCP, PIP and TIP and directed towards the thumb direction. The y axis follows the right-hand rule. The xy plane of such frame serves as the finger flexion plane, while the direction of the x-axis unit vector is considered as the direction of the minimum bending pose. Calling ${}^{hh}\boldsymbol{P}_{tip}$, ${}^{hh}\boldsymbol{P}_{pip}$ and ${}^{hh}\boldsymbol{P}_{mcp}$ the position vectors of fingertip, PIP joint and MCP joint according to the hand frame $\{HH\}$, respectively, the unit vector of $\{ff\}$ axes, $\widehat{\boldsymbol{x}}_{ff}$, $\widehat{\boldsymbol{y}}_{ff}$ and $\widehat{\boldsymbol{z}}_{ff}$, can be computed as follows:

$$\begin{cases} \widehat{\boldsymbol{x}}_{ff} = \dfrac{{}^{hh}\boldsymbol{P}_{tip} - {}^{hh}\boldsymbol{P}_{mcp}}{\|{}^{hh}\boldsymbol{P}_{tip} - {}^{hh}\boldsymbol{P}_{mcp}\|} \\ \widehat{\boldsymbol{z}}_{ff} = \dfrac{\left({}^{hh}\boldsymbol{P}_{pip} - {}^{hh}\boldsymbol{P}_{mcp}\right) \times \widehat{\boldsymbol{x}}_{ff}}{\|\left({}^{hh}\boldsymbol{P}_{pip} - {}^{hh}\boldsymbol{P}_{mcp}\right) \times \widehat{\boldsymbol{x}}_{ff}\|} \\ \widehat{\boldsymbol{y}}_{ff} = \widehat{\boldsymbol{z}}_{ff} \times \widehat{\boldsymbol{x}}_{ff} \end{cases} \tag{2.26}$$

The feature extraction node saves the finger frame parameters in form of rotation matrix ${}^{ff}_{hh}\boldsymbol{R}$ and translation vector ${}^{ff}_{hh}\boldsymbol{Q}$ that describe the hand frame according to the finger frame:

$${}^{ff}_{hh}\boldsymbol{R} = \left[\widehat{\boldsymbol{x}}_{ff} \quad \widehat{\boldsymbol{y}}_{ff} \quad \widehat{\boldsymbol{z}}_{ff}\right]^T \tag{2.27}$$

$${}^{ff}_{hh}\boldsymbol{Q} = -{}^{ff}_{hh}\boldsymbol{R} \cdot {}^{hh}\boldsymbol{P}_{mcp} \tag{2.28}$$

In the second phase of the calibration, the user has to put each finger in the desired maximum bending pose, depicted in figure 2.16.b, then enter "m" in the computer terminal through the keyboard. The user interface node handles the interaction and make the feature extraction node read one data frame from the coordinate transformation node and carry out the second calibration step on the measured data. In this subroutine the maximum bending angle of each finger is defined. The measured fingertip coordinates are transformed from {HH} to $\{ff\}$ by applying the rotation matrix and the translation vector previously computed in calibration phase one. Figure 2.16.b reports the main geometric elements involved in the computation.

$${}^{ff}\boldsymbol{P}_{tip} = {}^{ff}_{hh}\boldsymbol{R} \cdot {}^{hh}\boldsymbol{P}_{tip} + {}^{ff}_{hh}\boldsymbol{Q} \tag{2.29}$$

Then, the maximum bending angle $\beta_{\max}$ is computed as the angle described by the vector going from the origin of $\{ff\}$ (the MCP joint) to the projection of the fingertip on the plane of flexion. Considering the x and y components of ${}^{ff}\boldsymbol{P}_{tip}$, namely ${}^{ff}P_{tip,x}$ and ${}^{ff}P_{tip,y}$ respectively, the four-quadrant inverse tangent is applied to compute $\beta_{\max}$:

$$\beta_{\max} = \operatorname{atan}\left(\frac{{}^{ff}P_{tip,y}}{{}^{ff}P_{tip,x}}\right) + \frac{\pi}{2}\operatorname{sign}\left({}^{ff}P_{tip,y}\right)\left(1 - {}^{ff}P_{tip,x}\right) \tag{2.30}$$

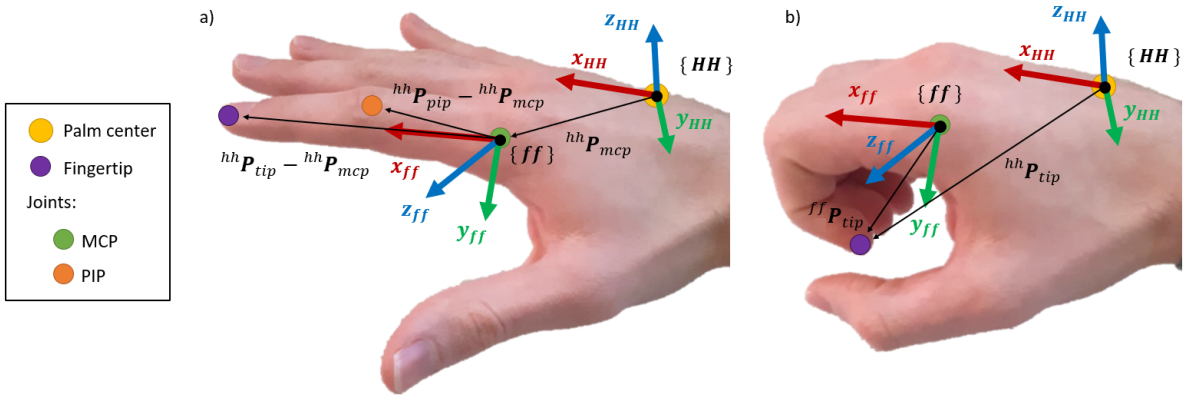


Figure 2.16: Vision-based tracking method: geometric elements involved in the system calibration. (a) First phase: computation of the finger frame $\{ff\}$ in the minimum bending pose. (b) Second phase: computation of the maximum bending angle in the maximum bending pose.

Feature extraction After the calibration, the tracking system starts running automatically. The feature extraction node reads the data coming from the coordinate transformation node and processes the fingertip coordinates of each finger as done in the second calibration phase and described by equations 2.29 and 2.30. The resulting bending angle β is cut if it is over the maximum angle or under zero. Given that the human bending angle range is generally different from the robotic one, β is then normalized for its own range to obtain a dimensionless variable called "bending level", $\hat{\beta}$, that ranges from zero to one, concisely represents the human finger motion and is easily and intuitively mappable to the bionic hand finger motion, as described in section 2.4. Considering β_{min} equal to zero, the bending angle range can be defined as the maximum bending angle β_{max}, shown in figure 2.15. Thus, the bending level can be obtained as follows:

$$\hat{\beta} = \frac{\beta}{\beta_{max}} \tag{2.31}$$

Additionally, a real-time linear Kalman filter is utilized to smooth the bending level value of each finger, thus improving robustness and steadiness during the teleoperation control of the bionic hand. The same method utilized for the glove-based tracking system was implemented, as described in section 2.2.3. The result from the filtering, $\hat{\beta}^*$ is published as the final output of the tracking system.

2.4. Teleoperation control frameworks

In this section, two distinct bionic hand teleoperation systems with the same basic framework, shown in figure 2.17, are constructed to remotely control the bionic hand analyzed in section 2.1. A motion tracking stage is used to measure the human hand pose in real-time, which, in turn, is mapped to the desired bionic hand pose. Then, an actuation controller stage drives the bionic hand motors to obtain a resulting bionic hand pose as similar as possible to the desired one. It should be noted that just motion mapping is implemented, while force control is not investigated because it is outside the scope of this study. Both the mapping and the actuation controller stages are the same for both the teleoperation systems, while the distinctive stage is the motion tracking one. In particular, the two motion tracking methods previously described in sections 2.2 and 2.3, namely the wearable glove-based and the vision-based tracking methods, are respectively employed. In this section, first, the implementation of the actuation control stage that drives the bionic hand motors, independently from the utilized hand tracking device, is explained; then, the method applied to map the tracked human hand motion to the desired bionic hand motion is presented.

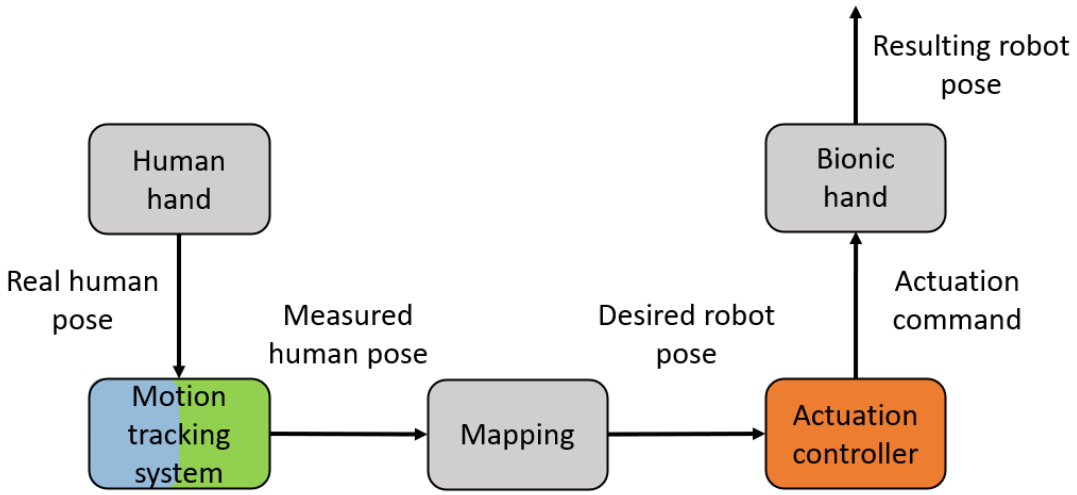


Figure 2.17: Schematic framework of the two implemented teleoperation systems.

2.4.1. Actuation control stage

The structure of the actuation control stage, employed by both the developed teleoperation system to drive the motion of the U-HandPi bionic hand, is depicted in figure 2.18. The control software is implemented on a computer that runs Ubuntu 16.04 and ROS Kinetic. The communication link between the computer and the bionic hand is estab-

lished using an Ethernet link. A simple open-loop control solution was adopted given that the bionic hand is not provided with internal feedback position sensors. The core of the open-loop is the inverse kinematics node, which computes the motors control commands corresponding to the desired robot pose. In particular, for each finger, such node applies the corresponding inverse kinematics model synthesized in section 2.1.2 to the desired bending angle, θ, to obtain the pulse width of the PWM signal of the motor, PW. The bending angle is read at 40Hz, in accordance with the sampling frequency set by the human hand motion tracking systems. The simplest way to apply the inverse kinematic model is presented by algorithm 2.1. At each k^{th} frame, the difference between the current value θ_k of the bending angle and the previous one θ_{k-1} is checked to recognize whether the direction of movement is extension or flexion and the PW_k value is consequently computed employing the corresponding modeled function $f_{ext}(\theta)$ or $f_{flex}(\theta)$ respectively.

Algorithm 2.1 Basic algorithm for inverse kinematic application

```
Initialize θ_{k-1}
while Inverse Kinematics node is running do
    θ_k = new desired angle
    if θ_k < θ_{k-1} then
        PW_k = f_ext(θ_k)
    else
        PW_k = f_flex(θ_k)
    end if
    θ_{k-1} = θ_k
end while
```

However, sudden changes from one function to the other would cause great instability, in particular for low bending angle values, where the two functions are significantly distant from each other and in static situations when noise on the input signal would also be recognized as a direction change. Such instability is due to the fact that the real system passes from one behavior to the other by a transient. To avoid a sudden change from one function to the other, and thus instability, the average bending angle in the last 40 frames (moving average on 1s), $\bar{\theta}_{40}$, is computed.

$$\bar{\vartheta}_{40} = \frac{1}{40} \sum_{i=k-40}^{k-1} \vartheta_i \tag{2.32}$$

Only if θ deviates from $\bar{\theta}_{40}$ for more than 1% of the bending angle range, θ_{max}, the direc-

tion change, if present, is recognized. Then, according to the current value of direction (extension or flexion), the corresponding inverse kinematic function, f_{ext} or f_{flex} respectively, is applied to θ in order to obtain the motor control signal value PW. The details are summarized by algorithm 2.2.

Algorithm 2.2 Advanced algorithm for inverse kinematics application

```
Initialize θ̄_40
Initialize direction
while Inverse Kinematics node is running do
    θ = new desired angle
    if θ < θ̄_40 − 0.01 · θ_max then
        direction = extension
    else
        if θ > θ̄_40 + 0.01 · θ_max then
            direction = flexion
        end if
    end if
    if direction == extension then
        PW = f_ext(θ)
    else
        PW = f_flex(θ)
    end if
    Update θ̄_40
end while
```

Once the pulse widths are computed, the communication handler node packs them in a message characterized by a proper TCP/IP format and sends them to the bionic hand through the Ethernet link. Each message specifies the six motor values (including the one related to the wrist motor, which is fixed at 1500) and the time in ms in which the motors must achieve those values. The minimum time that has proven to be manageable by the hand is 20ms, so, in order to keep a safety margin, a time equal to 25ms (40Hz) was fixed. Control messages are unpacked by the bionic hand processor and motors are consequently driven. A user interface manages the opening and the closing of the communication link according to the user's will. Once the user gets ready to start the communication, he/she has to enter "c" in the terminal through the keyboard. The communication handler creates a TCP/IP client socket using the Ethernet IP address that is assigned to the bionic hand. Whenever the user wants to stop the communication, he/she has to enter

"d" and the communication handler shutdowns the socket.

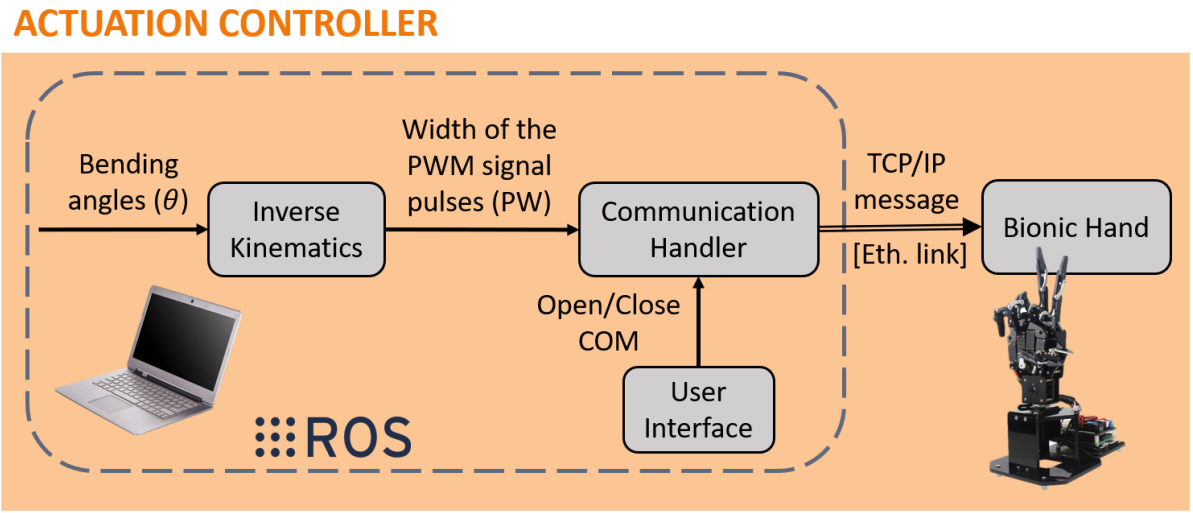


Figure 2.18: Components of the actuation control system, shared by both the implemented teleoperation systems.

2.4.2. Mapping method

A mapping stage computes the desired bending angles of the bionic hand fingers based on the pose of the corresponding human hand finger. As described in sections 2.2 and 2.3 respectively, the two developed tracking systems define the human finger pose in two different ways, according to the sensorial capabilities of the corresponding tracking device. For each finger, each tracking system extracts one feature ranging from 0 to 1, namely the flexion level $\hat{\varphi}^*$ and the bending level $\hat{\beta}^*$ respectively. Although such variables are different from each other, they can be mapped in the same intuitive and easy way to the robot finger bending angle. Indeed, the mapping node just multiplies the output of the utilized tracking system, whether $\hat{\varphi}^*$ in the case of the wearable system or $\hat{\beta}^*$ in the case of the vision system, for the bending angle range, θ_{max}, measured during the inverse kinematics calibration.

$$\begin{cases} \theta = \theta_{max} \cdot \hat{\varphi}^* \text{ , for the wearable system} \\ \theta = \theta_{max} \cdot \hat{\beta}^* \text{ , for the vision-based system} \end{cases} \tag{2.33}$$

As a result, a one-to-one mapping is established between the human finger characteristic angle (flexion angle in the case of the wearable tracking system and bending angle in the case of the vision tracking system) and the robot's bending angle, normalized for their own ROM.

2.5. Experimental evaluation

In this section, the three experiments which were carried out to evaluate and compare the two implemented teleoperation systems are discussed. Such experiments addressed the evaluation of the static accuracy, the dynamic accuracy and the usability of the implemented systems, respectively. In particular, the results of the first and second experiments concern the index finger. This section is organized as follows: first, the experimental setup used for the first and second experiments is described; then, protocols and extracted metrics of the first and second experiment are shown sequentially; lastly, the third experiment regarding the user study is analyzed in terms of setup, protocol and extracted metrics.

2.5.1. Ground truth measurements setup

During the first and second experiments, as described in detail in sections 2.5.2 and 2.5.3, a human operator had to control the bionic hand using the two developed systems. In the meanwhile, two sets of active, near-IR, optical markers were used to measure the ground truth of the resulting bending angle θ of the bionic hand (see section 2.1 for the definition of robot bending angle) and the ground truth of the characteristic human finger angle used to control the bionic hand, namely the real human flexion angle, φ_{real} (see section 2.2 for the definition of human flexion angle), for the wearable system and the real human bending angle, β_{real} (see section 2.3 for the definition of human bending angle), for the vision system.

The complete experimental setup is schematized in figure 2.19. To consider the optical measurements as "ground truth", the optical tracking system utilized to capture the markers 3D position had to be more accurate than both developed tracking systems. The NDI Optotrak Certus tracking system was employed, given its sub-millimeter accuracy and high robustness. As such device was already used during the calibration of the bionic hand inverse kinematics, see section 2.1.2 for further details. The position sensor was located at about 2m from the human hand of the operator and from the robotic hand to make them appear inside the field of view. One laptop was used to manage the optical tracking system and to record its measurements. A second laptop was connected to the bionic hand and the hand tracking device under study (glove or LMC) and was used to run the teleoperation software and to record ROS bag files containing the tracking stage output. The experiment was executed by two operators: operator 1 used the teleoperation system, while operator 2 handled markers and ROS bag recordings.

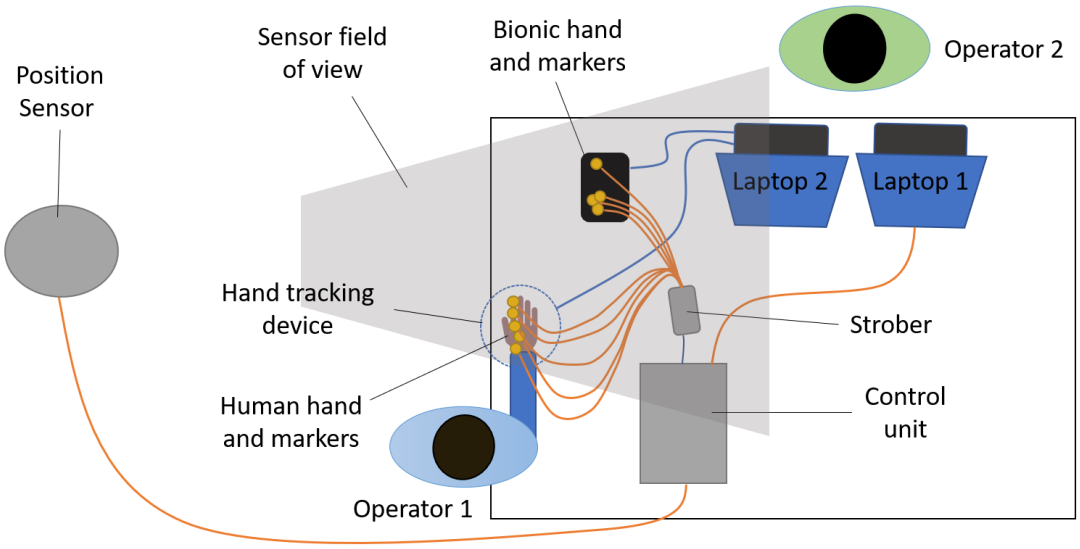


Figure 2.19: Setup scheme of human ROM estimation and accuracy evaluation experiments

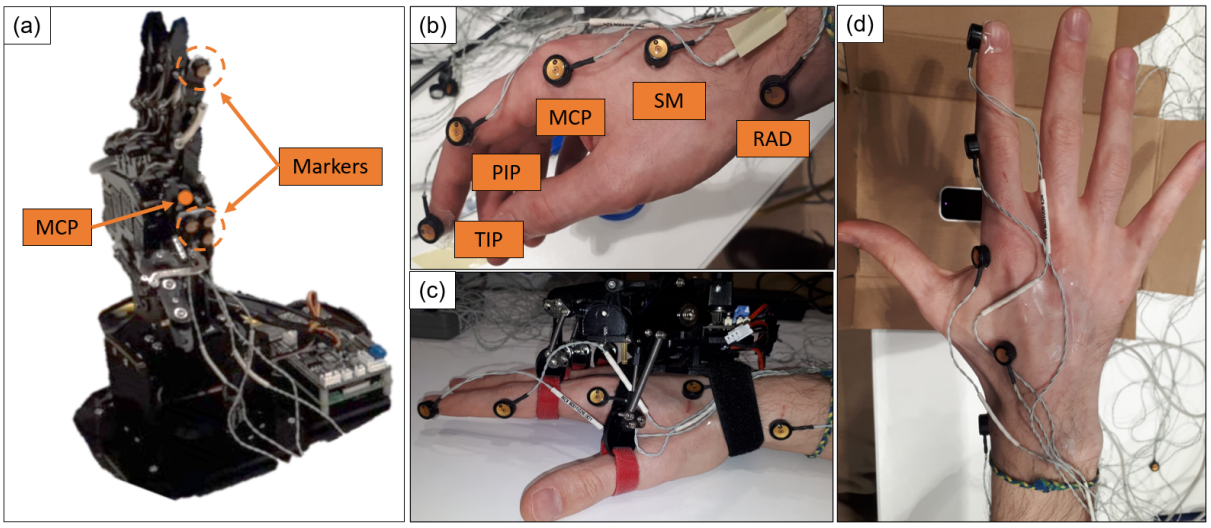


Figure 2.20: Markers positioning during for the ground truth tracking. (a) Marker position on bionic hand. (b) Markers position on the human hand. (c) Human hand with markers in the glove-based method. (d) Human hand with markers in vision tracking-based method.

As illustrated in Figure 2.20.(a), one marker was located on the side of the bionic hand index fingertip and three markers were placed on the side of the hand rigid case to define a 3D local reference frame, labeled as $\{RH\}$ and shown in figure 2.21. The MCP joint position was measured using the same method described in section 2.1.2 for the bionic

hand calibration, employing a pointing probe like the one shown in figure 2.5.d. As a prior step to compute the bending angle, MCP joint and fingertip positions were transformed from the Optotrak reference frame, $\{OPTO\}$, to $\{RH\}$. Calling ${}^{opto}\boldsymbol{P}_{m1}$, ${}^{opto}\boldsymbol{P}_{m2}$ and ${}^{opto}\boldsymbol{P}_{m3}$ the positions of the three markers attached to the rigid case with respect to $\{OPTO\}$, as shown in figure 2.21, the unit vectors $\widehat{\boldsymbol{x}}_{rh}$, $\widehat{\boldsymbol{y}}_{rh}$ and $\widehat{\boldsymbol{z}}_{rh}$ of the local reference frame axes can be defined as follows:

$$\begin{cases} \widehat{\boldsymbol{x}}_{rh} = \dfrac{{}^{opto}\boldsymbol{P_{m2}} - {}^{opto}\boldsymbol{P_{m1}}}{\|{}^{opto}\boldsymbol{P_{m2}} - {}^{opto}\boldsymbol{P_{m1}}\|} \\ \widehat{\boldsymbol{z}}_{rh} = \dfrac{\widehat{\boldsymbol{x}}_{rh} \times ({}^{opto}\boldsymbol{P_{m3}} - {}^{opto}\boldsymbol{P_{m1}})}{\|\widehat{\boldsymbol{x}}_{rh} \times ({}^{opto}\boldsymbol{P_{m3}} - {}^{opto}\boldsymbol{P_{m1}})\|} \\ \widehat{\boldsymbol{y}}_{rh} = \widehat{\boldsymbol{z}}_{rh} \times \widehat{\boldsymbol{x}}_{rh} \end{cases} \tag{2.34}$$

Given such definitions, the rotation matrix and the translation vector, which represent the Optotrak reference frame with respect to $\{RH\}$, can be computed and applied to fingertip and MCP joint positions to transform them. Such method is explained in detail in section 2.3.2 when the vision-based tracking system passes from the LMC reference frame to the hand frame. In particular, the computations are described by equations 2.23, 2.24 and 2.25. Once the local coordinates of the markers are obtained, the bending angle is computed as described in section 2.1.2 by equations 2.3 and 2.4.

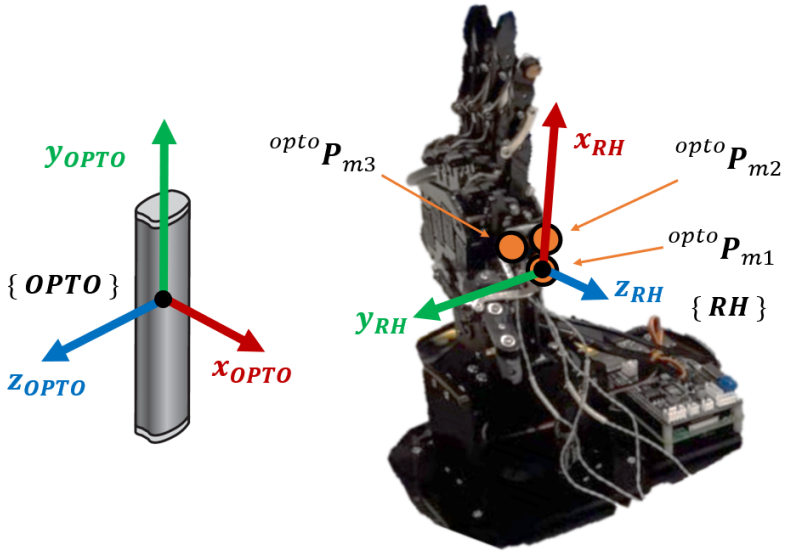


Figure 2.21: Ground truth tracking of bionic hand index finger: Optotrak reference frame, bionic hand reference frame and marker posistions labelling.

Finding a stable and meaningful positioning of the markers on the human hand was

fundamental to obtain a good estimation of human flexion and bending level. In [35] the authors describe a simplified but still accurate method to reconstruct finger movements in real-time, based on tracking of twenty-four optical markers by a set of six cameras. One marker is placed on each finger joint and a rigid body containing four markers is fixed to the wrist. Through the rigid tool, the authors define a 3D local reference frame that moves together with the hand and then refers the coordinates of the other markers to it. The transformed coordinates can be used to compute the angles of each joint. Getting inspired by this work, the starting idea was to place one marker on fingertip, PIP joint and MCP joint and to place a set of at least three rigidly connected markers on the wrist or the hand back to define the local reference frame. However, three issues had to be faced. The first was that during the recordings corresponding to the glove-based system, positioning possibilities were limited by the wide space covered by the glove rigid case and strings. This was not the case when using Leap Motion, however, the positioning of the markers had to be the same in both experiments to make the comparison more robust. The second issue was that only one optical sensor was used to simplify the setup, thus all the markers had to be placed on the same side of the hand to be seen together by the sensor. The third issue, which is typical of marker-based tracking, was that the markers had to be placed where skin motion was not so significant to minimize as much as possible relative motion between markers and bones.

Therefore, a novel solution was employed to overcome the problems related to this specific experimental scenario. As shown in figure 2.20.b., five markers were placed on the fingertip side, the PIP joint side, the MCP joint side, the second metacarpal bone (SM) in correspondence with the styloid process and the radial styloid process (RAD), respectively. Figures 2.20.(c) and 2.20.(d) show how the markers were integrated when using the glove-based and the vision-based methods, respectively. MCP, SM and RAD were used to define a local reference frame, $\{HH\}$. While testing the glove-based system, MCP and PIP joints were used to obtain the flexion angle; during the test of the vision-based method, MCP joint and fingertip were used to obtain the bending angle.

As a prior step to compute the human finger angle, MCP joint, PIP joint and fingertip positions were transformed from the Optotrak reference frame, $\{OPTO\}$, to $\{HH\}$. Calling ${}^{opto}\boldsymbol{P}_{mcp}$, ${}^{opto}\boldsymbol{P}_{sm}$ and ${}^{opto}\boldsymbol{P}_{rad}$ the positions of the three reference markers shown in figure 2.20.b, with respect to $\{OPTO\}$, the unit vectors $\widehat{\boldsymbol{x}}_{rh}$, $\widehat{\boldsymbol{y}}_{rh}$ and $\widehat{\boldsymbol{z}}_{rh}$ of the local reference frame axes can be defined as follows:

$$\begin{cases} \widehat{\boldsymbol{x}}_{hh} = \dfrac{{}^{opto}\boldsymbol{P_{mcp}} - {}^{opto}\boldsymbol{P_{rad}}}{\|{}^{opto}\boldsymbol{P_{mcp}} - {}^{opto}\boldsymbol{P_{rad}}\|} \\ \widehat{\boldsymbol{z}}_{hh} = \dfrac{({}^{opto}\boldsymbol{P_{sm}} - {}^{opto}\boldsymbol{P_{rad}}) \times \widehat{\boldsymbol{x}}_{hh}}{\|({}^{opto}\boldsymbol{P_{sm}} - {}^{opto}\boldsymbol{P_{rad}}) \times \widehat{\boldsymbol{x}}_{hh}\|} \\ \widehat{\boldsymbol{y}}_{hh} = \widehat{\boldsymbol{z}}_{hh} \times \widehat{\boldsymbol{x}}_{hh} \end{cases} \tag{2.35}$$

Given such definitions, the rotation matrix and the translation vector, which represent the Optotrak reference frame with respect to $\{HH\}$, can be computed and applied to fingertip, MCP joint and PIP joint positions to transform them. Again, such method is explained in detail in section 2.3.2 by equations 2.23, 2.24 and 2.25.

Once the local coordinates of the markers are obtained, they are transformed to a second local reference system, the finger frame $\{ff\}$, that is fixed with respect to the hand frame. The xy plane of such frame is considered as the plane where fingertip, PIP joint and MCP joint lie when MCP joint abduction is null. The finger frame is derived from a calibration procedure similar to the one implemented by the vision-based tracking system, as described in section 2.3.2. In such calibration phase, operator 1 has to put the finger in a neutral pose, with minimum bending, while operator 2 records the markers position. Thus, $\{ff\}$ is defined as having the origin in the measured MCP joint position, the x-axis on the direction of the measured MCP-to-PIP or MCP-to-TIP vector, according to whether the tracking system used in the test is the wearable one or the vision-based one respectively, the z-axis perpendicular to the plane defined by MCP, PIP and TIP and directed towards the thumb direction. The y axis follows the right-hand rule. Rotation matrix and translation vector of $\{ff\}$ are computed through equations 2.26, 2.27 and 2.28, while the coordinate transformation of the tip or PIP joint positions, according to the system under study, is executed as done by equation 2.29. Lastly, the flexion angle and the bending angle are computed by applying equation 2.30 to the tip or PIP position respectively.

2.5.2. Static accuracy assessment

The static accuracy experiment consisted in a static calibration procedure and was executed first on the glove-based teleoperation system and then on the vision-based one. During such procedure, several samples of ground-truth human angles, φ_{real} or β_{real}, homogeneously distributed in the huaman finger ROM, were measured through the markers and, at the same time, the output of the tracking system, flexion level $\hat{\varphi}^*$ and human bending level $\hat{\beta}^*$ respectively, was recorded. Moreover, the robotic bending angle, θ, resulting from the teleoperation control, was measured too. Four complete calibration cycles were

obtained for each system. A preprocessing step was needed to normalize the acquired φ_{real}, β_{real} and θ samples for the respective ranges, thus obtaining $\hat{\varphi}_{real}$, $\hat{\beta}_{real}$ and $\hat{\theta}$:

$$\hat{\varphi}_{real} = \frac{\varphi_{real}}{\varphi_{real,max} - \varphi_{real,min}} \tag{2.36}$$

$$\hat{\beta}_{real} = \frac{\beta_{real}}{\beta_{real,max} - \beta_{real,min}} \tag{2.37}$$

$$\hat{\theta} = \frac{\theta}{\theta_{max}} \tag{2.38}$$

While θ_{max} was known from the bionic hand calibration, the angle limits of φ_{real} and β_{real} were unknown, and given the difficulty of precisely recording the poses which are associated with the human finger limits by the tracking systems, an estimation of such values was needed, based on the acquired samples. In particular, a set of calibration cycles of the tracking stage were obtained by coupling the measured samples of φ_{real} and $\hat{\varphi}$, for the glove-based method, and the samples of β_{real} and $\hat{\beta}$, for the vision-based method. Then, the linear function that best fits the cycles (excluding saturation samples) was computed through least-squares linear regression. Such function is assumed to be the ideal characteristic of the tracking system. This approximation is the most optimistic one, given that it minimizes the distance of the samples from the ideal behavior. φ_{real} and β_{real} values that according to the ideal characteristic should correspond to the minimum and the maximum $\hat{\varphi}^*$ and $\hat{\beta}^*$ were considered as the human range limits, that are $\varphi_{real,max}$ and $\varphi_{real,min}$, $\beta_{real,max}$ and $\beta_{real,min}$, respectively. After normalization, three kinds of error were investigated to evaluate and compare the accuracy performance of the two systems, namely the ones listed and computed as follows and shown in figure 2.22:

- Overall teleoperation control accuracy error

$$\begin{cases} \varepsilon_{tele,glove} = \hat{\varphi}_{real} - \hat{\theta} \\ \varepsilon_{tele,vision} = \hat{\beta}_{real} - \hat{\theta} \end{cases} \tag{2.39}$$

- Human hand motion tracking accuracy error

$$\begin{cases} \varepsilon_{track,glove} = \hat{\varphi}_{real} - \hat{\varphi}^* \\ \varepsilon_{track,vision} = \hat{\beta}_{real} - \hat{\beta}^* \end{cases} \tag{2.40}$$

- Bionic hand actuation control accuracy error

$$\begin{cases} \varepsilon_{act,glove} = \hat{\varphi}^* - \hat{\theta} \\ \varepsilon_{act,vision} = \hat{\beta}^* - \hat{\theta} \end{cases} \tag{2.41}$$

Note that, the desired robot pose can be expressed directly by the output of the motion tracking stage, given that the resulting robot pose is described not by the bending angle but by the dimensionless bending level and that the mapping is one-to-one in this case.

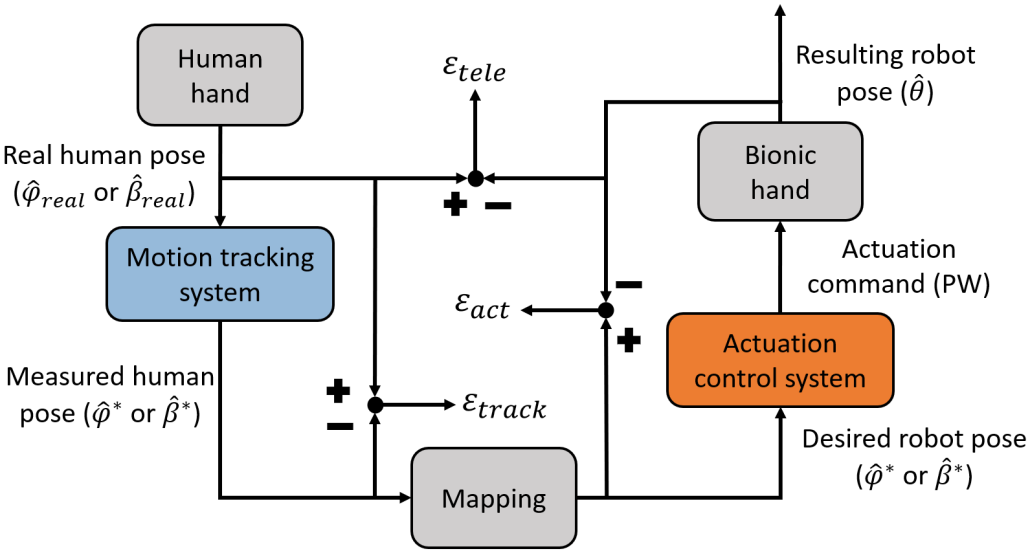


Figure 2.22: Illustration of the errors in the comparison experiments, including the overall teleoperation error, human hand motion tracking error, and the bionic hand actuation control error.

For each kind of error, a standard approach to evaluate the accuracy of an input-output system static response was applied [36] [37]. The couple of variables determining the error were compared by assessing their calibration cycles. A prior step was to compute the real static characteristic of the system under study, defined by the couple of curves (one for flexion and one for extension) that best fit the samples. In other words, it represents the average static behavior of the system in terms of how the input is translated into the output. The two curves were found by fitting polynomials or rational functions (ratios of polynomials) on the samples through least squares regression. Adjusted R-squared (R^2_{adj}) metric was used to set polynomials degree, which is a measure of explained variance that penalizes complex functions to reduce over-fitting. Given N samples y_i with mean $\bar{y}$, k regressors and N prediction $\hat{y}_i$ from the fitted model, R^2_{adj} is defined as:

$$R^2_{adj} = 1 - \frac{N-1}{N-k-1} \cdot \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{2.42}$$

The followed rule was to start from a linear function, try to increase step by step the order of the polynomial and each time check if the R^2_{adj} improved. If R^2_{adj} did not improve or the function showed clear over-fitting behavior (for example, huge oscillations or a curvature such that the function does not reach the range limits), then the degree was reduced to the previous step and the resulting function was selected. In this way, attention was paid so as not to increase too much the function complexity and risking to get over-fitting while grasping the significant non-linearity of the system. Once the curves were obtained, two static accuracy metrics were extracted:

- Non-linearity: maximum distance between the real characteristic and the ideal one (output equal to input)
- Hysteresis: maximum distance between the two curves that compose the hysteresis cycle

Note that, as regards the analysis of the bionic hand actuation control performance, since the actuation controller is the same for both the teleoperation systems, data from glove and leap motion experiments were joint together to obtain a single static characteristic.

2.5.3. Dynamic accuracy assessment

The dynamic experiment was done just after the static one without restarting the sensors or changing their calibration. In this phase, operator 1 performed six different movements such that the flexion level, for the wearable system case, or the bending level, for the vision system case, showed a sinusoidal behavior in time (sinusoidal movements). Each movement differed from the others in terms of frequency or amplitude. Frequency was qualitatively set by the operator to low ($<$0.4Hz) or high ($>$ 0.4Hz), while amplitude was set to small (40% of $\hat{\theta}$ range explored), medium (60% of $\hat{\theta}$ range explored) or large ($\hat{\theta}$ reaches saturation). For each movement, the variables which determine the errors defined by equations 2.39, 2.40 and 2.41 were recorded at 40Hz, which is equal to the sampling frequency of the teleoperation systems. The same angle normalization as the one employed in the static experiment was used to retrieve such variables. For each couple of variables, x and y, the cross-correlation function was applied to the signals corresponding to the same kind of movement. Calling m the time lag between the samples of the two signals and N the number of samples of the two signals, the cross-correlation function can be defined as:

$$C_{xy}(m) = \sum_{n=0}^{N-m-1} x_{n+m} y_n \tag{2.43}$$

The two signals were then synchronized by back-shifting in time the delayed one for the time lag corresponding to the maximum peak of cross-correlation. Two dynamic accuracy metrics were computed for each couple of signals:

- Mean Time Delay: time lag corresponding to the maximum peak of the cross-correlation function. Note that such metric was analyzed only on the overall teleoperation system because ROS bag recordings did not start at the same time instant as the markers recordings, so the lag detected by the cross-correlation includes also the delay between the starting times and cannot be used to characterize nor the tracking system delay or the motion controller delay. In such case, it can be just used to synchronize the signals.
- Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2} \tag{2.44}$$

  where N is the number of samples, x_i and y_i are the i^{th} samples of the two signals respectively. This metric was extracted after time synchronization. It is widely used in literature to assess the accuracy of teleoperation systems like the ones implemented in this work [30] [28] [29].

2.5.4. User study

The third experiment aimed at demonstrating the usability of the two teleoperation systems during both precision and power grasp tasks, and, at the same time, evaluating the user experience. The experiment design was inspired by similar tests implemented in related works. Indeed, the execution of a grasp experiment is always present in previous related works, as it is fundamental to assess if the developed system is actually able to do what it is designed for and how good its performance is.

In this experiment, six right-handed users, in good health condition and with no motor deficiency, were invited to perform the grasp tasks. Figure 2.23.(a) shows the overall experimental setup. The bionic hand was fixed to a table, and the user sat comfortably in front of it. In order to test different kinds of grasp, four different objects, depicted in

figure 2.23.b, were selected for the test. Depending on their shape and size, each one of them required a different kind of grasp posture in order to be caught, according to Feix taxonomy [38]. Furthermore, they were divided into two groups depending on whether the grasp required precision or not:

- Parallelepiped: large diameter power grasp (figure 2.23.b.1)
- Cork stopper: tripod precision grasp (figure 2.23.b.2)
- Cylinder: medium wrap power grasp (figure 2.23.b.3)
- Cube: quadpod precision grasp (figure 2.23.b.4)

The four objects were placed on the table in front of the user as well. A laptop was used to execute the implemented teleoperation software, acquiring the human hand motion data, whether from the glove or the LMC, and sending the proper command signals to the bionic hand.

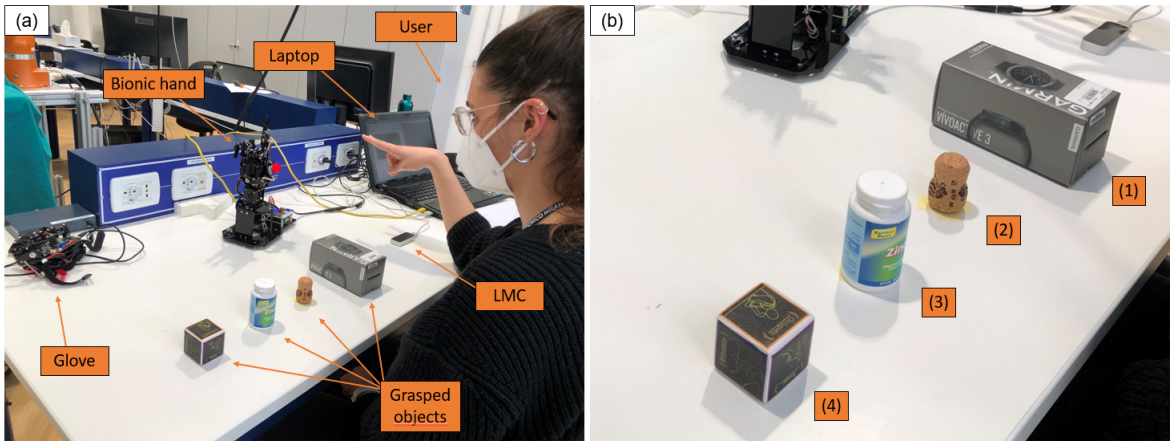


Figure 2.23: User study: (a) experimental setup, (b) grasped objects.

According to the designed experimental protocol, which was carefully explained to the user before the experiment, the user had to make the bionic hand grasp the objects from the right to the left, starting from the parallelepiped and ending with the cube, as shown in figure 2.23.a, in order to alternate power and precision grasps. For each object, the user had to take the object with the left hand, make it approach the robotic hand, and use the right hand to control the robot grasp through the hand motion tracking device. Once the user was confident enough, he/she left the object to the bionic hand, which in turn had to hold it for 5 seconds. The objects belonging to the precision grasp group (cork stopper and cube) had to be grasped using only the fingertips of the bionic hand, without touching the palm, in order to assure a precision grasp. Once five seconds passed, the user

could take the object back to its original position using the left hand. If the object fell or touched the palm (only for the precision grasp) during any phase of the trial, the trial was deemed a failure, otherwise, it was considered a success. The user had to repeat the grasp sequence six times. Half of the users experimented first with the glove-based system and then with the vision-based system, while the other half did vice versa. Hence, 144 trials were done for each method. Before performing the experiment, the users were allowed to familiarize themself with the teleoperation control. During such learning phase, grasp practice was not allowed. Following the completion of each experiment, the user was asked to complete a NASA TLX questionnaire [39] to assess their perceived workload. Such questionnaire is divided into two parts. In the first step, six different indexes of workload had to be subjectively rated by the user on a scale from 0 to 100 endowed with twenty-one grades, namely:

- Mental demand: required mental and perceptual activity.
- Physical demand: required physical activity.
- Temporal demand: felt time pressure due to the task pace.
- Effort: overall hardness of the work
- Performance: felt level of success (in this case, the scale value is reversed, id est a low index value indicates a high performance)
- Frustration: discouragement, irritation, stress and annoyance.

The user was provided with a description of each workload index. In the second phase of the questionnaire, the user was shown the fifteen combinations of the six indexes and, for each pair, he/she was asked to choose the index he/she thought was more important to determine the total workload of a generic activity. In this way, each index gets a weighting that is proportional to the tally t_i (i = 1,2,...,6) of the times in which it is chosen. Then, the adjusted rating can be computed by multiplying the raw rating R_i for the corresponding weight. Lastly, a total workload index, WL, can be computed as the sum of all the adjusted ratings:

$$WL = \sum_{i=1}^{6} R_i \cdot \frac{t_i}{15} \tag{2.45}$$

The application of weights accounts for the subjectivity of the user assessment, reducing its influence on the final result.

3 | Results

This section presents the results of the designed experiments, including the human hand range of motion estimation, the accuracy performance comparison, and the user demonstration results.

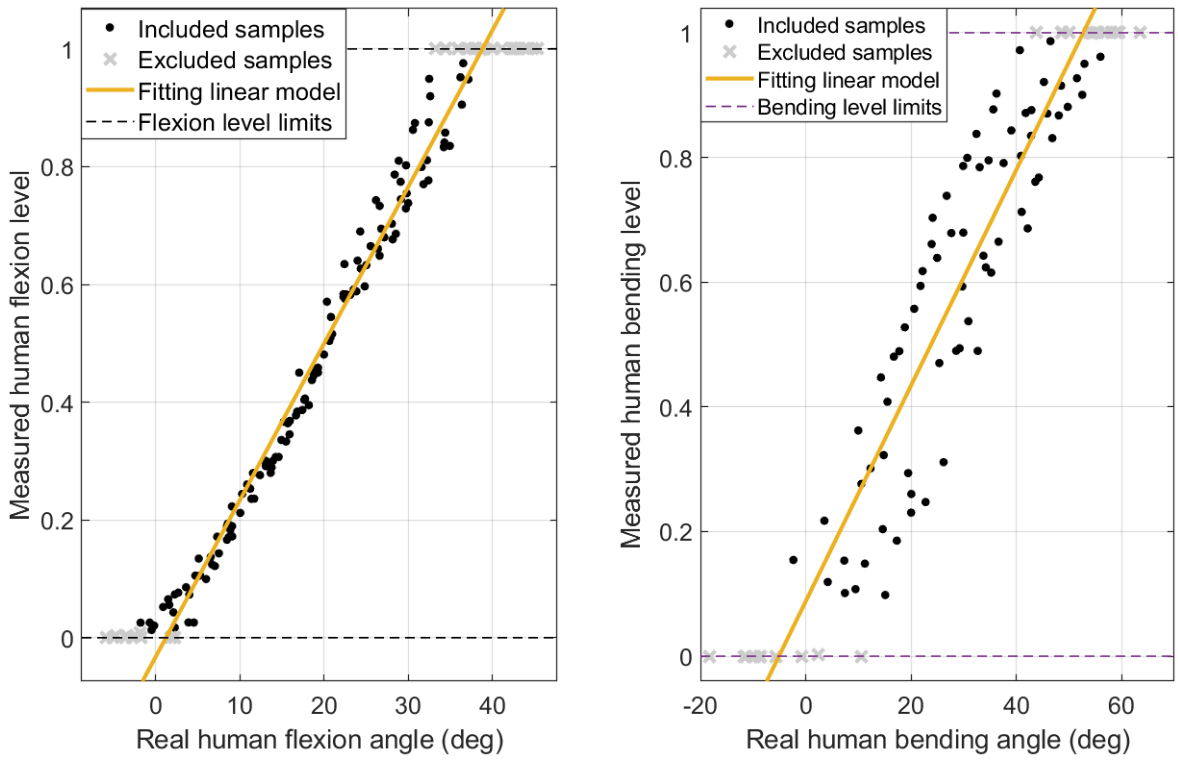
3.1. Human range of motion estimation

Four complete static calibration cycles were measured for each system. In particular: 152 static poses were obtained while using the wearable system (38 per cycle on average), 56 of which related to extension, 52 to flexion and 44 to saturation; as regards the vision-based system, one calibration cycle was excluded because it was affected by markers occlusion, thus obtaining 96 static poses (32 per cycle on average), 33 of which related to extension, 34 to flexion and 27 to saturation. Figure 3.1 shows, for each tracking system, the fitting result of linear regression and the fitted samples. Table 3.1 summarizes the metrics of fitting goodness, that are RMSE and R^2_{adj}, together with resulting functions parameters which are referred to equation 3.1. The resulting ROM limits were $[1.22°; 38.8°]$ for φ_{real} (figure 3.1.a) when the operator used the glove and $[-5.09°; 52.7°]$ for β_{real} (figure 3.1.b) when the operator used the vision-based tracking method. Therefore, the resulting ROM (difference between the limits) are 37.6° and 57.8°, respectively. Such values are coherent with the qualitative observations.

System	RMSE	R^2_{adj}	p1	p2
Wearable	0.0397	0.980	0.0266	-0.0325
Vision	0.121	0.797	0.0173	0.0881

Table 3.1: Fitting results and parameters of tracking system linear model

$$y = p1x + p2 \tag{3.1}$$



(a) (b)

Figure 3.1: Motion tracking: linear fitting results for wearable system (a) and vision system (b)

3.2. Accuracy performance

3.2.1. Overall teleoperation control performance

Polynomial functions were used to find the static characteristics of the two overall teleoperation systems, that are the expressions that give $\hat{\theta}$ as a function of $\hat{\varphi}_{real}$ or $\hat{\beta}_{real}$, for the glove-based system and the vision-based one respectively. The fitting curves resulting from the adopted criteria are represented in figures 3.2a and 3.2b, together with the fitted samples. Non-linearity and hysteresis of the static characteristics are underlined by figures 3.2c and 3.2d, respectively. Table 3.2 summarizes the fitting results and the resulting metrics of non-linearity and hysteresis. Table 3.3 shows the parameters values of the fitted functions.

The wearable system behavior was modeled by a second-degree polynomial (equation 3.2) in extension and a fifth-degree polynomial (equation 3.3) in flexion. The system is characterized by a linear region from 25% to 60% of the flexion level range, having a high

agreement with the ideal characteristic, both in flexion and extension. The shift from the ideal characteristic is mainly present in the extension curve at the range borders, while the flexion curve seems quite aligned. Non-linearity and hysteresis peak both at 89% of the range.

The vision-based system behavior was modeled by a second degree polynomial (equation 3.2) and a linear function (equation 3.1). Fitting results in this case are poorer because the dispersion of the samples is higher. The ideal characteristic runs from end to end of the hysteretic cycle and is completely detached from the two curves, implying that non-linearity exists throughout the range. Such behavior gives worse error metrics than the glove-based method. Non-linearity and hysteresis peak at 20% and 26% of the range, respectively.

System	Direction	Degree	RMSE	R^2_{adj}	Non-linearity	Hysteresis
Wearable	Flexion	5^{th}	0.0225	0.993	11.1%	14.6%
	Extension	2^{nd}	0.0273	0.989		
Vision	Flexion	1^{st}	0.0555	0.964	20.0%	27.0%
	Extension	2^{nd}	0.0609	0.934		

Table 3.2: Fitting results and error metrics of the two teleoperation systems.

System	Direction	p1	p2	p3	p4	p5	p6
Wearable	Flexion	3.43	-6.41	3.53	-0.362	0.883	-0.0251
	Extension	0.531	0.563	0.0799	-	-	-
Vision	Flexion	1.20	-0.240	-	-	-	-
	Extension	-0.269	1.34	0.0124	-	-	-

Table 3.3: Parameters of the teleoperation systems characteristics.

$$y = p1x^2 + p2x + p3 \tag{3.2}$$

$$y = p1x^5 + p2x^4 + p3x^3 + p4x^2 + p5x + p6 \tag{3.3}$$

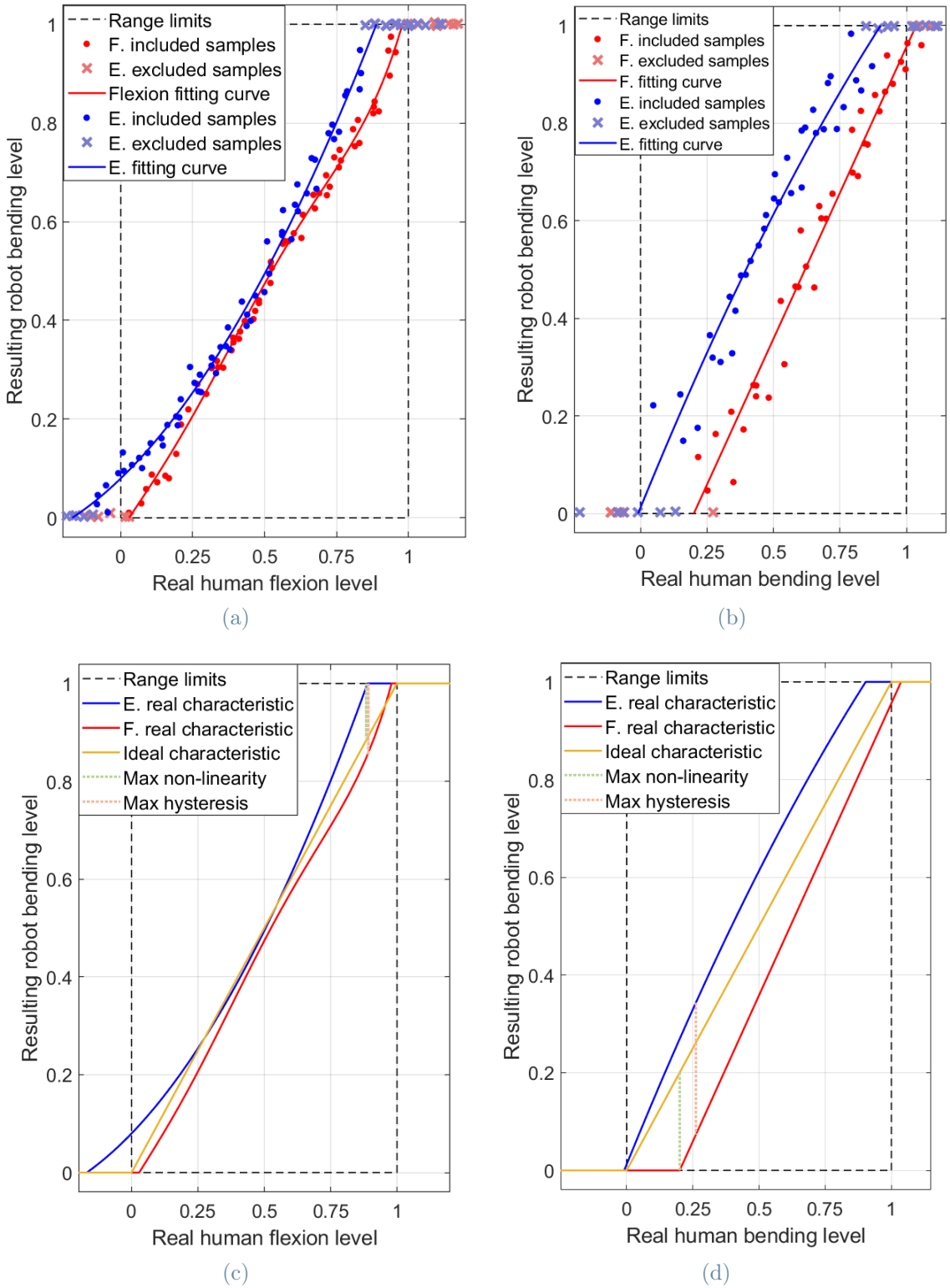


Figure 3.2: Static testing of overall teleoperation control systems: fitting results of glove-based (a) and vision-based (b) systems; non-linearity and hysteresis of glove-based (c) and vision-based (d) systems.

The six couples of synchronized signals from the dynamic trials of the experiment are shown in figure 3.3.a and 3.3.b for the glove-based system and the vision-based one, respectively. Delay and RMSE metrics are summarized in table 3.4.

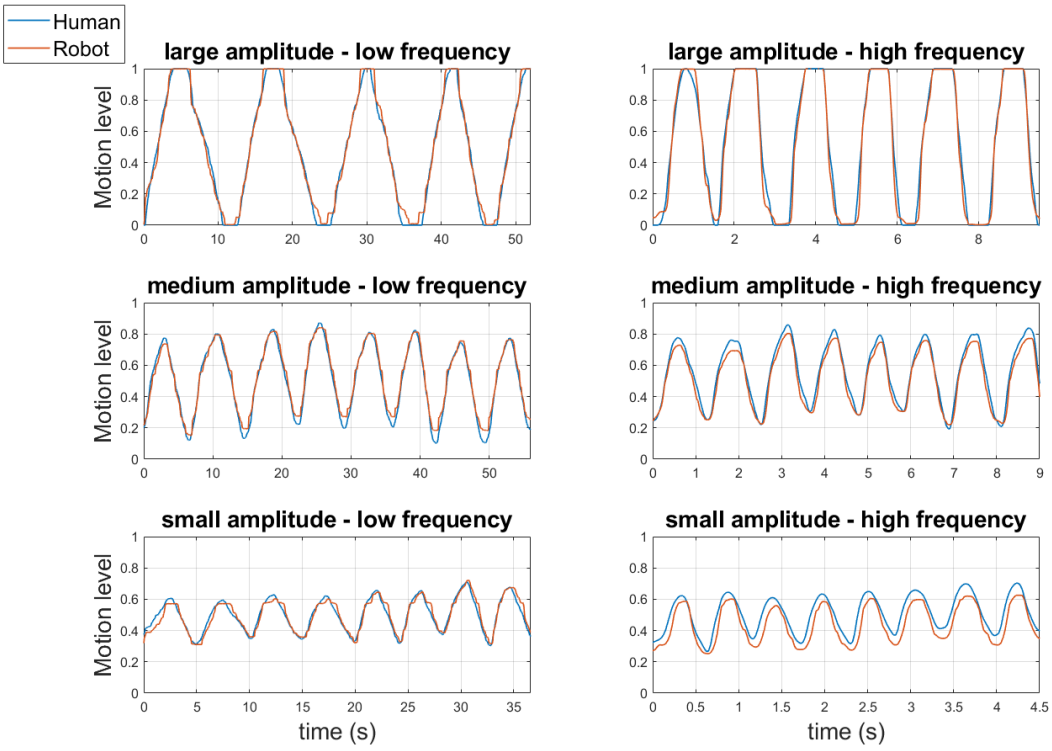
In the wearable system case, human and robot signals are well overlapped except few missed peaks and a systematic vertical shift for the small-amplitude - high-frequency trial. From RMSE values it seems evident that the best results are obtained for small and slow movements, while increasing speed or amplitude makes tracking harder.

As regards the dynamic trials of the vision-based system, although the signals are highly correlated, they are less superimposed than in the wearable system case and missing peaks are more visible. The small-amplitude - high-frequency trial is the worst case, as happened for the glove system. In this case, the general trend appears to be that movements with both high frequency and high amplitude are more easily transferred to the robot. RMSE values are higher in the vision-based system in all the six conditions, indicating greater teleoperation error than in the glove case.

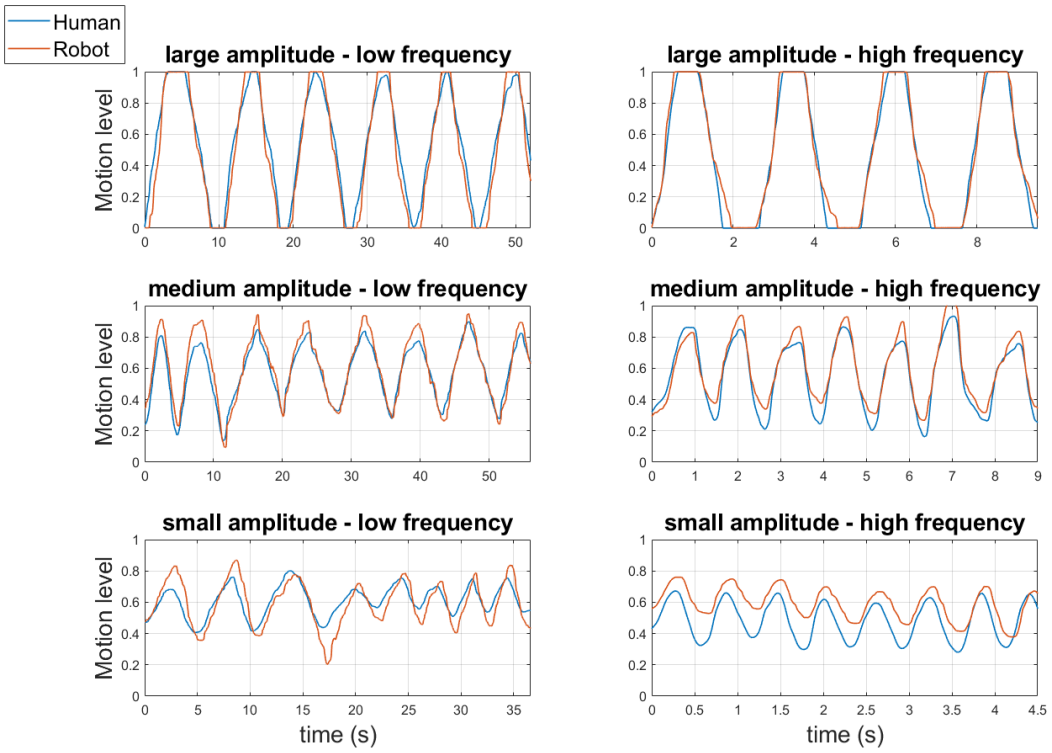
There is insufficient evidence to discern a systematic difference in delay values between the two systems. However, the delay is always less than 0.4 seconds in all trials for both systems.

	System	L - l	M - l	S - l	L - h	M - h	S - h
Delay (ms)	Wearable	400	225	125	175	175	150
	Vision	350	325	0	250	225	150
RMSE	Wearable	4.01%	3.40%	2.80%	5.67%	4.44%	7.33%
	Vision	8.72%	6.56%	8.51%	6.42%	7.19%	12.6%

Table 3.4: Mean delay and RMSE of the two teleoperation systems for each motion condition. The symbols L, M and S stand for large, medium and small amplitude, respectively; l and h stand for low and high frequency, respectively.



(a)



(b)

Figure 3.3: Dynamic testing of overall teleoperation control under 6 different conditions: synchronized human and robot signals for wearable system (a) and vision system (b).

3.2.2. Human hand motion tracking performance

Polynomial functions were explored to find the static characteristics of the motion tracking stages of the two teleoperation systems, which are the expressions that give $\hat{\varphi}^*$ or $\hat{\beta}^*$ as a function of $\hat{\varphi}_{real}$ or $\hat{\beta}_{real}$, for the wearable system and the vision-based one respectively. The fitting curves resulting from the adopted criteria are represented in figures 3.4a and 3.4b, together with the fitted samples. Non-linearity and hysteresis of the static characteristics are underlined by figures 3.4c and 3.4d, respectively. Table 3.5 summarizes the fitting results and the resulting metrics of non-linearity and hysteresis. Table 3.6 shows the parameters values of the fitted functions.

The best-fitting curves have the same polynomial degrees as those used in section 3.2.1 to model the characteristics of the two overall teleoperation systems. In both the systems, the main differences from the overall teleoperation system case are a lower dispersion of the samples around the curves and a lower hysteresis below 20% of the human angle range. Such distinctions result in better fitting and slightly different metrics of non-linearity and hysteresis, which are higher in the wearable system case and lower in the vision system case. Once again, samples dispersion, nonlinearity and hysteresis in the vision system case are significantly worse than in the wearable case.

System	Direction	Degree	RMSE	R^2_{adj}	Non-linearity	Hysteresis
Wearable	Flexion	5^{th}	0.0184	0.995	12.5%	16.2%
	Extension	2^{nd}	0.0252	0.992		
Vision	Flexion	1^{st}	0.0548	0.963	18.0%	25.1%
	Extension	2^{nd}	0.0590	0.946		

Table 3.5: Fitting results and error metrics of the two implemented hand tracking stages.

System	Direction	p1	p2	p3	p4	p5	p6
Wearable	Flexion	0.0120	0.0145	-0.0389	-0.0370	0.292	-0.516
	Extension	0.515	0.638	0.0465	-	-	-
Vision	Flexion	1.17	-0.215	-	-	-	-
	Extension	-0.494	1.61	-0.0590	-	-	-

Table 3.6: Parameters of the tracking systems characteristics.

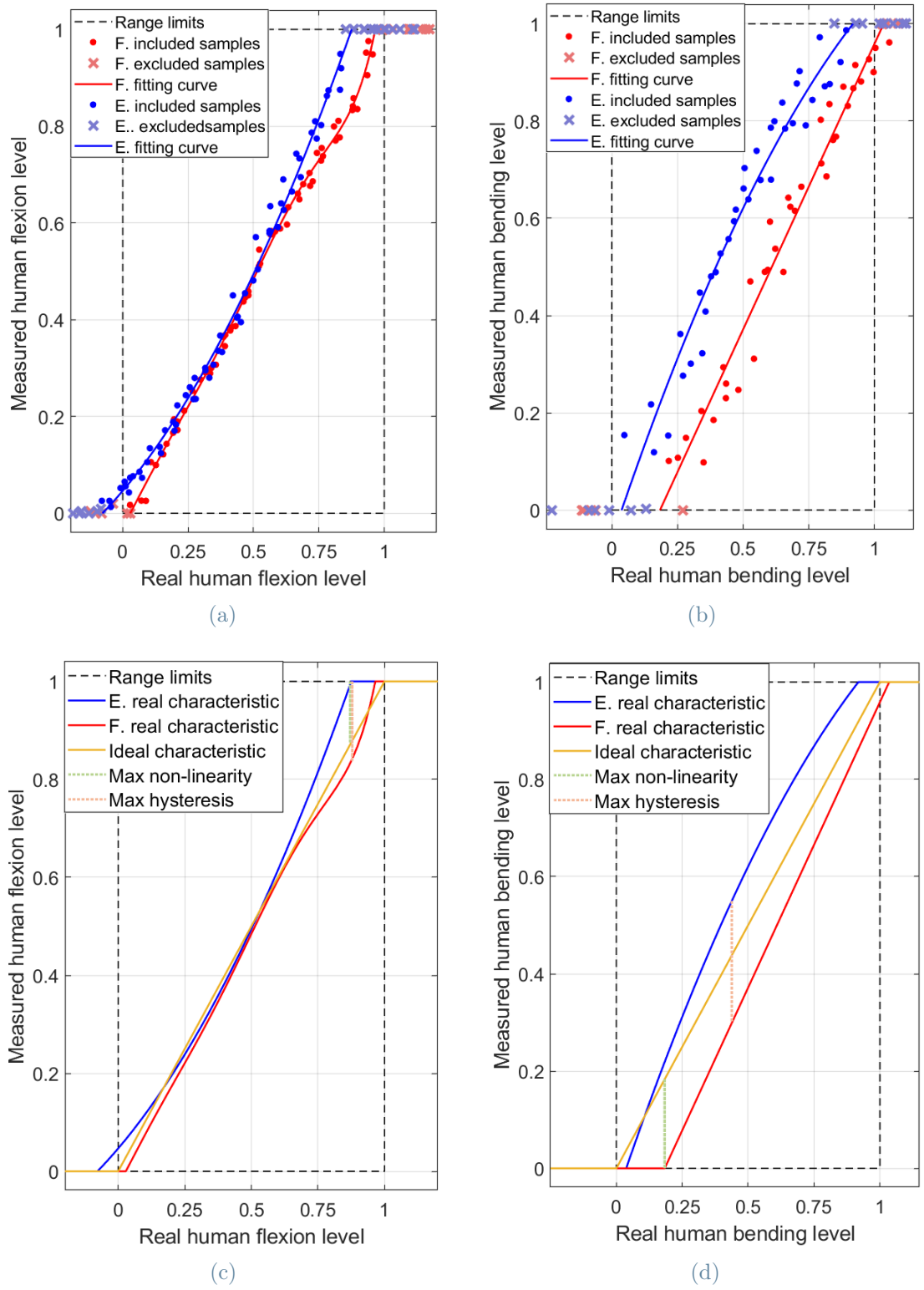


Figure 3.4: Static testing of the tracking systems: fitting results of glove-based (a) and vision-based (b) systems; non-linearity and hysteresis of glove-based (c) and vision-based (d) systems.

The six couples of synchronized signals from the dynamic trials of the experiment are shown in figure A.1.a and A.1.b for the glove-based system and the vision based one, respectively. RMSE values are summarized in table 3.7. The graphical results are very similar to the ones regarding the overall teleoperation systems, although, in most of the conditions, the RMSE is slightly lower. Once again, the vision-based system is affected by more evident missing peaks and greater RMSE values in all the six conditions.

System	L - l	M - l	S - l	L - h	M - h	S - h
Wearable	3.20%	3.05%	2.23%	3.89%	4.71%	7.50%
Vision	8.14%	6.62%	7.44%	5.70%	7.01%	12.5%

Table 3.7: RMSE of the two tracking systems for each motion condition. The symbols L, M and S stand for large, medium and small amplitude, respectively; l and h stand for low and high frequency, respectively.

3.2.3. Bionic hand actuation control performance

Rational functions were explored to find the static characteristic of the actuation control stage, common to both teleoperation systems, that is the expressions which give $\hat{\theta}$ as a function of the output of the tracking stage, $\hat{\varphi}^*$ or $\hat{\beta}^*$ depending on the system. The fitting curves resulting from the adopted criteria are represented in figure 3.5a together with the fitted samples. Non-linearity and hysteresis of the static characteristic are underlined by figure 3.5b. Table 3.8 summarizes the fitting results and the resulting metrics of non-linearity and hysteresis. Table 3.9 shows the parameters values of the fitted functions. Equations 3.4 and 3.5 represent the function classes used to model extension and flexion behaviors, respectively.

Above 20% of the range, both the extension and flexion curves are almost linear and superimposed to the ideal behavior. Non-linearity and hysteresis are present only under 20% of the range. Samples dispersion is also higher in that region. RMSE, non-linearity and hysteresis are significantly lower than the ones measured for the overall teleoperation systems.

Direction	Degree	RMSE	R^2_{adj}	Non-linearity	Hysteresis
Flexion	$3^{rd}/4^{th}$	0.0139	0.998	4.11%	7.37%
Extension	$3^{rd}/1^{st}$	0.0132	0.999		

Table 3.8: Fitting results and error metrics of the actuation control system.

Direction	p1	p2	p3	p4	q1	q2	q3	q4
Flexion	45.3	191	266	122	-5.06	134	359	229
Extension	0.152	0.783	0.0835	0	0.0178	-	-	-

Table 3.9: Parameters of the actuation control system

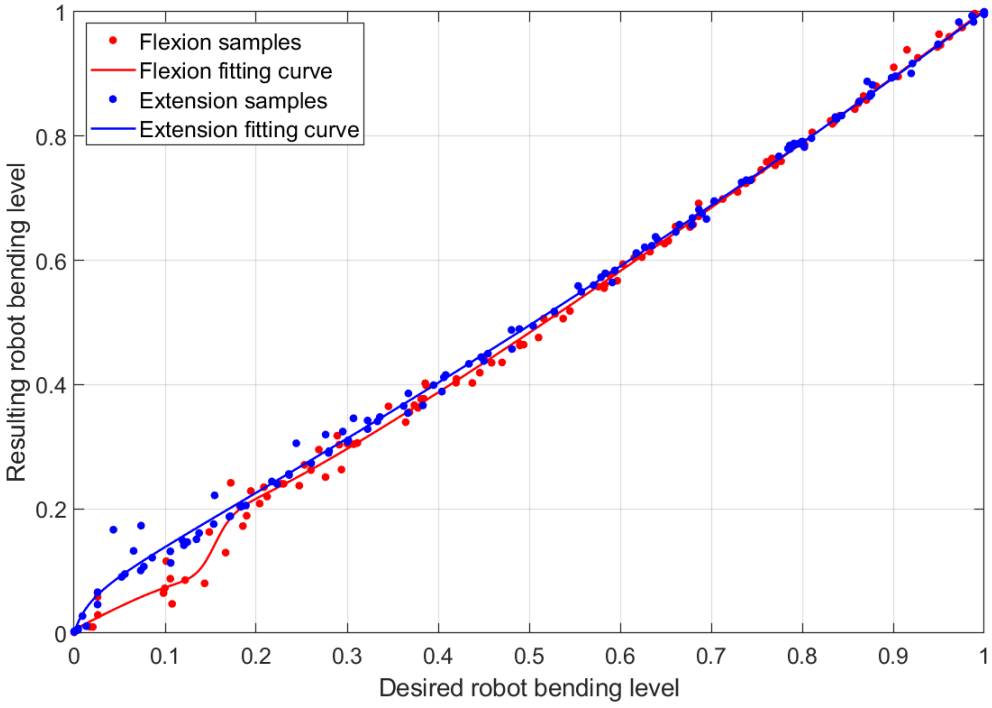
$$y = \frac{p_1 x^3 + p_2 x^2 + p_3 x + p_4}{x + q_1} \tag{3.4}$$

$$y = \frac{p_1 x^3 + p_2 x^2 + p_3 x + p_4}{x^4 + q_1 x^3 + q_2 x^2 + q_3 x + q_4} \tag{3.5}$$

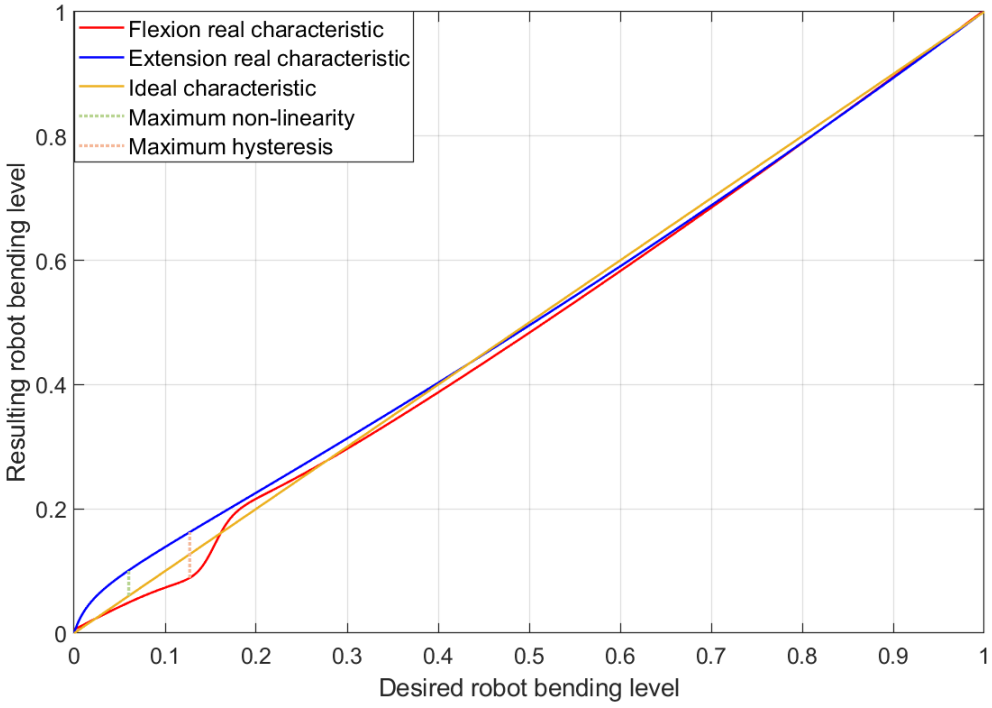
The six couples of synchronized signals from the dynamic trials of the experiment are shown in figure A.2.a and A.2.b for the glove-based system and the vision-based one, respectively. RMSE values are summarized in table 3.10. In all conditions, the resulting robot signal follows precisely the desired one defined by the measured human pose. RMSE is always under 3% of the range except for one trial.

System	L - l	M - l	S - l	L - h	M - h	S - h
Wearable	1.78%	1.90%	1.00%	4.36%	1.63%	2.58%
Vision	1.52%	1.27%	0.78%	2.03%	1.68%	3.00

Table 3.10: RMSE of the actuation control system for each motion condition. The symbols L, M and S stand for large, medium and small amplitude, respectively; l and h stand for low and high frequency, respectively.



(a)



(b)

Figure 3.5: Static testing of the actuation control systems: fitting results (a); non-linearity and hysteresis (b).

3.3. User study

All the people involved in the user study succeeded in performing all four grasp tasks using both the teleoperation systems. Figures 3.8 (a)-(d) report one demonstration example for each grasp scenario when using the glove-based tracking method, while figures 3.8 (e)-(h) illustrate one demonstration example for each grasp scenario when utilizing the vision-based tracking method. The success rate is high in both cases, in particular, it is 98.6% for the glove-based method and 96.5% for the vision-based method, which are not significantly different values in statistics terms (Fisher's exact test). Figure 3.6 illustrates the average raw ratings for each workload index resulting from the NASA-TLX questionnaires compiled by the six users. The profiles of the two systems are similar. Nevertheless, the vision-based system is affected by lower mean perceived performance, while the glove-based system brings a higher mean effort.

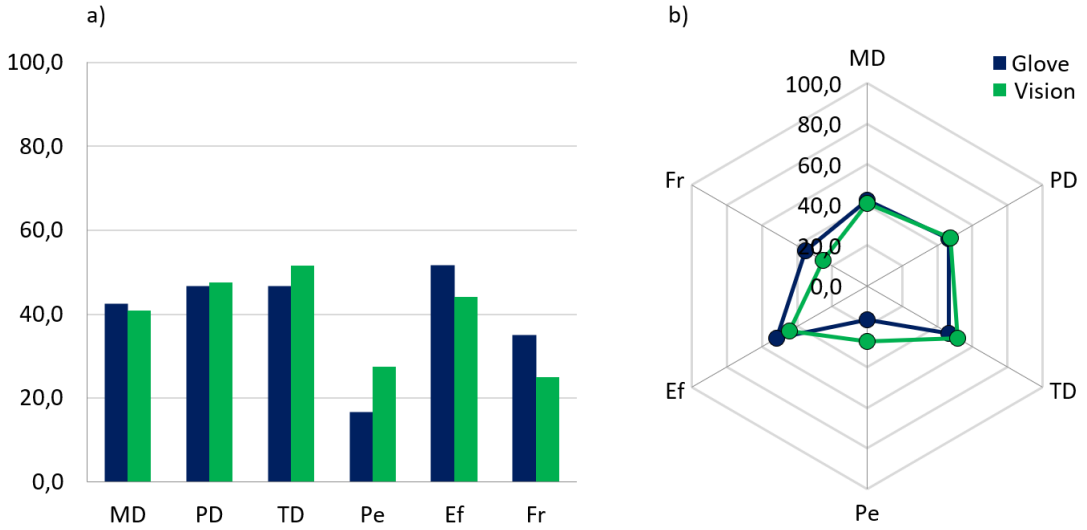


Figure 3.6: User study: graphical representation of the average raw scores through bar diagram (a) and radar diagram (b). MD = mental demand, PD = physical demand, TD = temporal demand, Pe = performance, Ef = effort, Fr = frustration.

The scores, adjusted for the subjective weights, are represented in figure 3.7. The main differences between wearable and vision-based methods noticed for the raw ratings remain valid. The total perceived workload, that is the sum of the adjusted ratings, is practically the same in the two systems. For further details, table 3.11 reports the values of average raw rating, adjusted rating and total workload. Free user feedback is in line with questionnaire results: users agreed in feeling the glove as better performing but at the same time as less comfortable, causing a greater general effort.

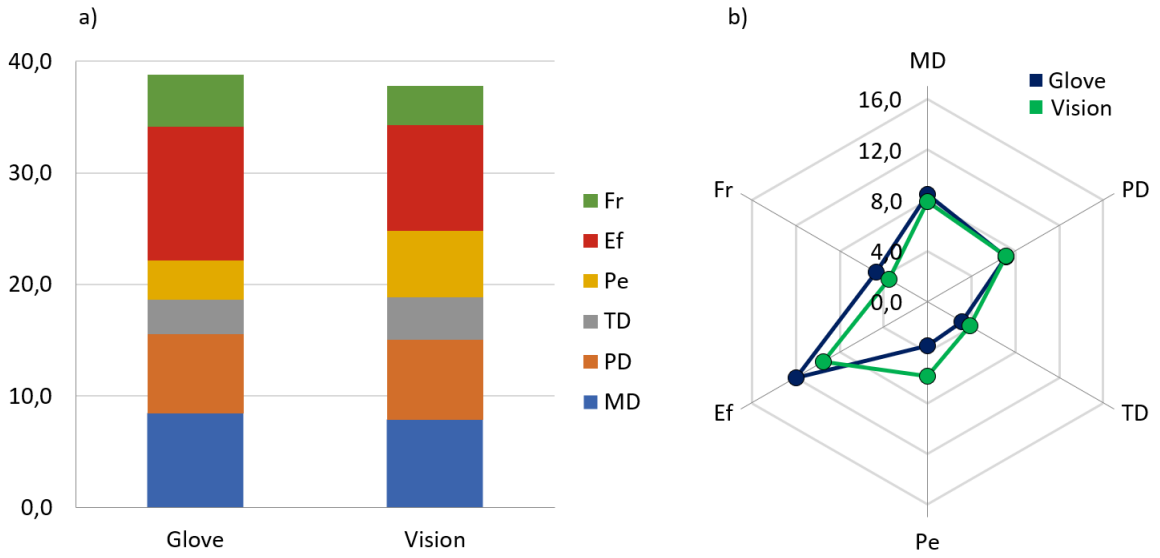


Figure 3.7: User study: graphical representation of the average adjusted scores through pile diagram (a) and radar diagram (b). MD = mental demand, PD = physical demand, TD = temporal demand, Pe = performance, Ef = effort, Fr = frustration.

	Index	Glove		Vision	
		μ	σ	μ	σ
Raw ratings	Mental demand	42.5	17.0	40.8	22.3
	Physical demand	46.7	22.9	47.5	20.7
	Temporal demand	46.7	17.8	51.7	8.2
	Performance	16.7	10.3	27.5	19.7
	Effort	51.7	21.1	44.2	12.4
	Frustration	35.0	21.7	25.0	22.6
Adj. ratings	Mental demand	8.4	5.6	7.9	6.1
	Physical demand	7.1	4.8	7.2	5.7
	Temporal demand	3.1	3.7	3.8	4.8
	Performance	3.5	2.1	5.9	4.1
	Effort	12	5.9	9.5	2.4
	Frustration	4.7	4	3.5	4.2
	Total workload	38.8/100		37.8/100	

Table 3.11: Comparison result of NASA TLX questionnaire. The average (μ) and standard deviation (σ) of raw and adjusted ratings and the total workload are reported.

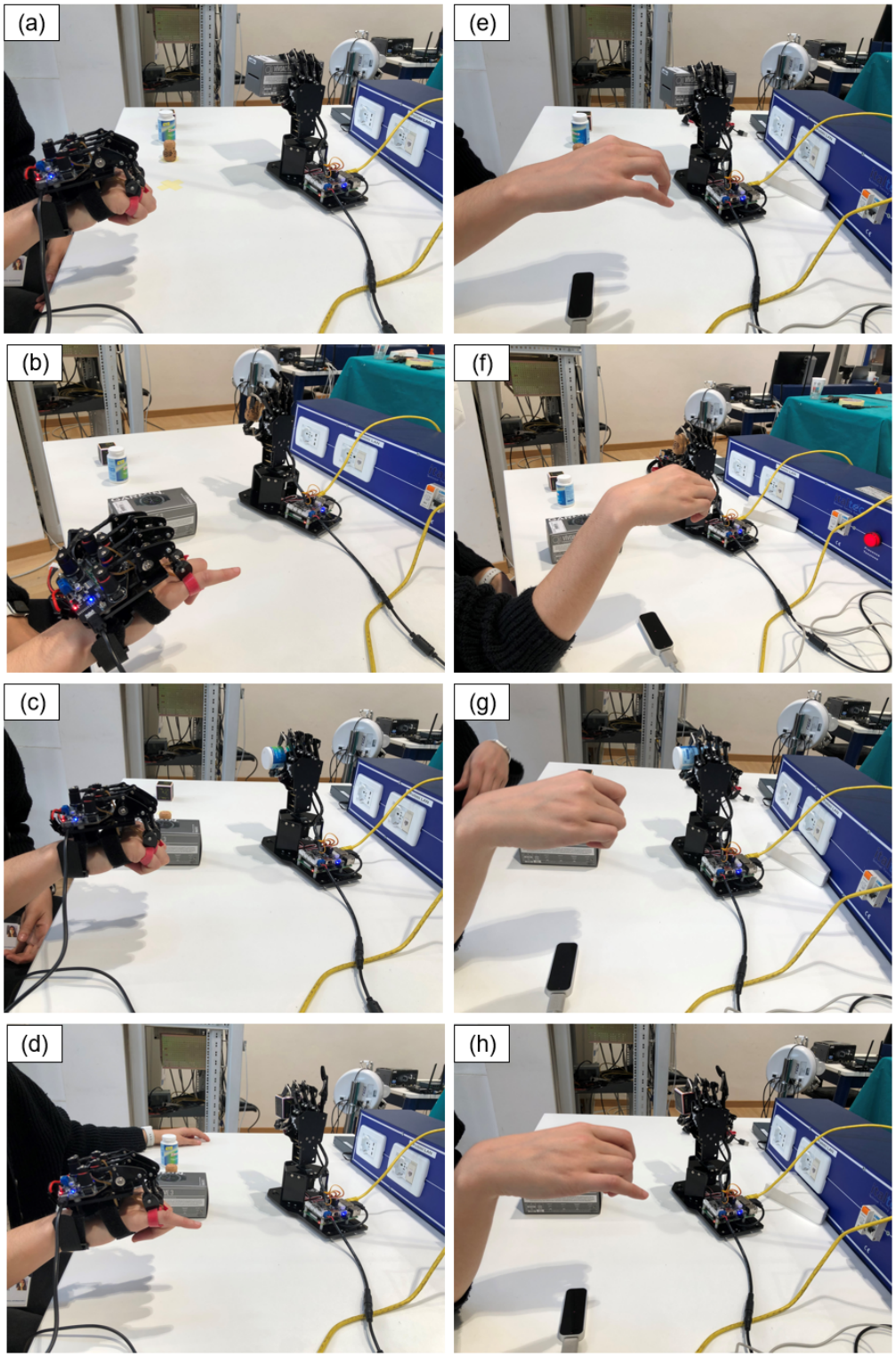


Figure 3.8: User study and demonstration of the implemented teleoperation control frameworks. Figures from (a) to (d) represent the four grasping performed through glove-based teleoperation. Figures from (e) to (h) report the four grasping performed through vision-based teleoperation.

4 | Discussion

In this section, the aim of the work and the performed activities are briefly recalled, then the results from the experiments are interpreted and discussed according to the objectives, lastly, the conclusions of the work are drawn.

Objectives recall and performed activities This work aimed to develop and compare two teleoperation frameworks designed to intuitively control the same underactuated bionic hand but based on two different human hand tracking devices that belong to the two most competitive categories in the field, namely the wearable mechanical gloves and the contactless vision-based tracking devices. The workflow included four main steps. First, the inverse kinematic model of the bionic hand was derived from a calibration procedure performed with an external ground truth tracking system. Then, an exoskeleton sensorized glove and a Leap Motion Controller were employed to implement the wearable glove-based and the vision-based hand tracking systems, respectively. Consequently, two teleoperation systems were developed by combining one of the two tracking systems, respectively, with an actuation control stage that is based on the bionic hand inverse kinematic model previously derived. The connection between the tracking and actuation stages was achieved by establishing a simple mapping strategy between the captured human data and the desired motion of the bionic hand. Lastly, three experiments were executed to evaluate the performance of the implemented teleoperation systems in terms of static and dynamic accuracy, usability and user experience.

Demonstration of usability The high success rates achieved by the glove-based teleoperation system and the vision-based one during the usability experiment are 98.6% and 96.5% respectively, as reported in section 3.3. Such values prove that both systems can be used to control the underactuated bionic hand to successfully perform both power and precision grasp tasks on objects of various shapes and sizes. In both cases, the average level of performance perceived by users and measured by the NASA-TLX questionnaire was high, with scores of 16.7/100 and 27.5/100, respectively, confirming the good result (note that the less the performance index is the better the perceived performance is).

Therefore, both the systems may be used as a basis to develop a teleoperation system endowed with a complete manipulator.

Accuracy and performance analysis The results of the static part of the accuracy experiment regarding the overall teleoperation control, reported in section 3.2.1, show that both the overall teleoperation systems, are affected by non-linearity caused by hysteresis and non-uniform sensitivity of the hysteretic curves. However, both non-linearity and hysteresis are greater in the vision-based system. In particular, the vision-based teleoperation system is characterized by a non-linearity that peaks the 20% of the finger ROM, which corresponds to 24° that is quite remarkable. The hysteretic cycle is notable too, since it reaches 27% of the ROM, corresponding to 32°. The wearable method showed more reasonable non-linearity and hysteresis values, which are 11% and 15% of the ROM respectively, corresponding to 13° and 17°. Moreover, in this case, non-linearity is confined at the ROM borders, while, in the vision-based system, it affects the whole ROM. Given such results, we can state that the accuracy of the vision-based teleoperation method is worse with respect to the glove-based one. The dynamic trials confirmed the better performance of the glove-based system since the RMSE metric is between 2.8% and 7.3% for the glove-based method and between 6.4% and 12.6% for the vision-based one. In particular, the vision-based system performs worse for slow and small movements, which characterize fine and delicate operations. Furthermore, the dispersion of the data around the static characteristic is significantly higher in the vision-based system, hinting at lower behavior predictability, lower mapping repeatability and lower precision. The usability experiment demonstrated that the inaccuracy issues characterizing the vision-based method do not compromise the teleoperation activity. However, it also underlined that, with respect to the glove-based method, the average performance perceived by the users is worse. Such result suggests that the lower accuracy of the vision-based method may influence teleoperation performance.

Since the actuation control stage is the same for both systems, such a difference in the accuracy performance is attributable to the different motion tracking stages. Indeed, the results reported in section 3.2.2 for the analysis of the human hand motion tracking performance confirm that the vision-based tracking method has worse accuracy than the glove-based tracking method at determining the desired robot pose according to the human one, both in static and dynamic scenarios. Furthermore, given that in both the developed tracking systems the processing to extract the human motion feature from the sensor data and map it to the robot feature is reduced to the minimum, the worse performance of the vision-based tracking method with respect to the glove-based one suggests that the LMC device applied as a controller for the teleoperation of a bionic hand brings

to a less accurate motion retargeting than the glove utilized in this work, at least if advanced and heavy processing techniques on the acquired data are not performed. The source of such accuracy error may be due to the limits of the image processing procedure performed by the LMS while reconstructing the joints position. Since LMC is one of the most accurate vision-based tracking devices, such a conclusion may be extended also to the vision-based tracking device category as a whole. Such conclusion regarding the vision devices category is important, considering that the design of the glove utilized in this work requires a far simpler design than vision tracking devices since it has a simple structure and it does not need complex software, like the LMS, to derive the tracking data. Such limits may be not significant when vision technology is applied for the interaction with virtual reality; however, in the case of robot teleoperation in real scenarios, higher accuracy is advisable to grant safety and effectiveness of the control, thus further research is needed to develop more robust algorithms for joints position estimation.

User comfort During the usability experiment, thanks to its better accuracy, the glove-based system performs better. However, the user study results, reported in section 3.3, also underline a greater effort felt on average by the users when using the glove. As suggested by the literature, the reasons may be the glove intrusiveness, which hinders natural movements, and the mechanical frictions of the glove structure, which enhance the effort needed to move the fingers. Such issue may seem less impacting on the overall system performance than the accuracy issue characterizing the vision-based system. However, it is interesting to note that the higher effort and frustration bring the average perceived total workload practically equal to the one felt when using the vision-based system. Such a result confirms that, although wearable mechanical tracking devices are more accurate than vision-based ones, they can be strongly limited in their application by a mechanical structure that is designed without giving high priority to the user comfort, thus making vision-based systems so attractive despite their accuracy limitations. Therefore, in order to make wearable systems more competitive, mechanical hardware must be improved taking user comfort into account.

Time delay As reported in section 3.2.1, there is no discernible systematic difference between the time delays which characterize the two systems. In both cases, the delay is always under 0.5s, with a mean value across the six trials of about 0.2s, which is acceptable and does not compromise the teleoperation activity.

Performance of the actuation control system Lastly, comparing figures 3.4 and 3.2, it is evident that, in both the cases, the overall teleoperation system behavior is very

similar to the tracking system stage behavior. The only significant difference, for both the systems, is a lower hysteresis in the lower part of the range. This brings an increment in the linear behavior of the glove-based system and decreased non-linearity and hysteresis for the vision-based system. The overall dispersion around the curves, in such region, is slightly lower too. Also for the dynamic trials, the difference is minimal.

Such high similarity brings to the hypothesis that the motion tracking stage tends to impact more than the actuation control stage on the accuracy of the overall system. The results from the analysis of the actuation control system alone, reported in section 3.2.3, confirm such hypothesis. Indeed, the actuation control system is characterized by a nice linear behavior over the 20% of the ROM with low dispersion. In such region, the actuation control does not significantly affect the accuracy of the overall teleoperation systems. The actuation controller has an active part in reducing the accuracy only below the 20% of the range, where it is affected by non-linearity. This can be the cause of a larger hysteresis in that region of the overall teleoperation systems. However, non-linearity does not exceed 4.1% of the ROM, which corresponds to a bending angle of 5°, which is irrelevant for finger teleoperation. Dynamic trials confirm that the desired bionic hand pose is accurately transferred to the bionic hand. Therefore, the inverse kinematic model obtained through finger motion calibration, applied through a proper algorithm, proved to grant a generally accurate and reliable actuation control system, demonstrating improvements with respect to the linear approach used in previous works.

Conclusion Mechanical wearable gloves and vision cameras are the most used hand tracking devices in the context of bionic hand teleoperation. The lack of robust comparative studies between the two classes, in terms of the impact on the overall teleoperation control, prevents researchers to have a quantitative basis for selection, which could be useful to address the specific limitations of the two methods, which ultimately restrict their application.

For these reasons, this work firstly aimed at developing two teleoperation systems, based on the two methods respectively, and designed to intuitively control the same under-actuated bionic hand in a continuous pose space. A user experiment was executed to demonstrate the usability of the developed systems. Results show that both systems can be used to effectively perform grasp tasks on objects of various shapes and sizes, applying both power and precision grasp skills.

The second main objective was to compare the two implemented systems in terms of both accuracy and usability within the same experimental setup, to obtain a meaningful comparison of the two tracking devices when applied to teleoperation. The accuracy

experiment revealed a significantly lower accuracy of the vision-based teleoperation with respect to the glove-based one. The user experiment underlined that such worse accuracy may influence the grasping performance perceived by the user. Given the structure and the functioning of the two teleoperation systems, the worse accuracy and performance of the vision-based system is attributable to the limited accuracy of the vision camera device, suggesting the lower suitability of such technology for performing fine and precise tasks. Thus, further research is needed to develop more robust algorithms for image-based joints positions estimation. On the other side, the user study suggested that the glove-based method brings a worse comfort and higher effort by the user. Therefore, mechanical hardware must be improved taking user comfort into account.

The third subsidiary objective, consisting in constructing an accurate non-linear motion model of the bionic hand, was also achieved. Indeed, the actuation control stage of the developed teleoperation systems, which is based on the obtained inverse kinematic model of the fingers, proved to be highly accurate.

Future research may focus on three main aspects: (1) the design of a sensorized glove which improves the user comfort while keeping a high accuracy performance; (2) the application of sensor fusion techniques to improve the accuracy performance of the vision-based method, while not increasing too much the system encumbrance; (3) the implementation of a complete hand-arm teleoperation system.

Bibliography

[1] Elena Garcia, Maria Antonia Jimenez, Pablo Gonzalez De Santos, and Manuel Armada. The evolution of robotics research. *IEEE Robotics & Automation Magazine*, 14(1):90–103, 2007.

[2] International Organization for Standardization (ISO). Robots and robotic devices–vocabulary (iso 8373: 2012), 2012.

[3] Dov Katz, Jacqueline Kenney, and Oliver Brock. How can robots succeed in unstructured environments. In *In Workshop on Robot Manipulation: Intelligence in Human Environments at Robotics: Science and Systems*. Citeseer, 2008.

[4] Rui Li, Hongyu Wang, and Zhenyu Liu. Survey on mapping human hand motion to robotic hands for teleoperation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[5] Jing Luo, Wei He, and Chenguang Yang. Combined perception, control, and learning for teleoperation: key technologies, applications, and challenges. *Cognitive Computation and Systems*, 2(2):33–43, 2020.

[6] Hirakjyoti Basumatary and Shyamanta M Hazarika. State of the art in bionic hands. *IEEE Transactions on Human-Machine Systems*, 50(2):116–130, 2020.

[7] Daniel R Ramírez Rebollo, Pedro Ponce, and Arturo Molina. From 3 fingers to 5 fingers dexterous hands. *Advanced Robotics*, 31(19-20):1051–1070, 2017.

[8] Markus Grebenstein, Maxime Chalon, Gerd Hirzinger, and Roland Siegwart. Antagonistically driven finger design for the anthropomorphic dlr hand arm system. In *2010 10th IEEE-RAS International Conference on Humanoid Robots*, pages 609–616. IEEE, 2010.

[9] Steffen W. Ruehl, Christoper Parlitz, Georg Heppner, Andreas Hermann, Arne Roennau, and Ruediger Dillmann. Experimental evaluation of the schunk 5-finger gripping hand for grasping tasks. In *2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014)*, pages 2465–2470, 2014.

[10] Bin He, Shuai Wang, and Yongjia Liu. Underactuated robotics: A review. *International Journal of Advanced Robotic Systems*, 16(4), 2019.

[11] Jared Wormley, Matthew Johannes, David Handelman, Christopher Dohopolski, Reed Young, Michael McLoughlin, and Rodrigo Rimando. High dexterity robotics for safety and emergency response-17104. In *Proc. 43rd Annu. Waste Management Conf.(WM2017)*, volume 4, pages 2064–2076, 2017.

[12] Jeongsoo Lim, Inho Lee, Inwook Shim, Hyobin Jung, Hyun Min Joe, Hyoin Bae, Okkee Sim, Jaesung Oh, Taejin Jung, Seunghak Shin, et al. Robot system of drc-hubo+ and control strategy of team kaist in darpa robotics challenge finals. *Journal of Field Robotics*, 34(4):802–829, 2017.

[13] Jordan Allspaw, Jonathan Roche, Michael Yannuzzi, Holly A Yanco, and Nicholas Lemiesz. Remotely teleoperating a humanoid robot to perform fine motor tasks with virtual reality-18446. WM Symposia, Inc., PO Box 27646, 85285-7646 Tempe, AZ (United States), 2018.

[14] M.A. Diftler, J.S. Mehling, M.E. Abdallah, N.A. Radford, L.B. Bridgwater, A.M. Sanders, R.S. Askew, D.M. Linn, J.D. Yamokoski, F.A. Permenter, B.K. Hargrave, R. Platt, R.T. Savely, and R.O. Ambrose. Robonaut 2 - the first humanoid robot in space. In *2011 IEEE International Conference on Robotics and Automation*, pages 2178–2183, 2011.

[15] Geng Yang, Honghao Lv, Zhiyu Zhang, Liu Yang, Jia Deng, Siqi You, Juan Du, and Huayong Yang. Keep healthcare workers safe: application of teleoperated robot in isolation ward for covid-19 prevention and control. *Chinese Journal of Mechanical Engineering*, 33(1):1–4, 2020.

[16] Dongseok Ryu, Chang-Soon Hwang, Sungchul Kang, Munsang Kim, and Jae-Bok Song. Wearable haptic-based multi-modal teleoperation of field mobile manipulator for explosive ordnance disposal. In *IEEE International Safety, Security and Rescue Rototics, Workshop, 2005.*, pages 75–80. IEEE, 2005.

[17] Manuel Caeiro-Rodríguez, Iván Otero-González, Fernando A Mikic-Fonte, and Martín Llamas-Nistal. A systematic review of commercial smart gloves: Current status and applications. *Sensors*, 21(8):2667, 2021.

[18] Jie Liu and Yuru Zhang. Mapping human hand motion to dexterous robotic hand. In *2007 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 829–834. IEEE, 2007.

[19] Luca Colasanto, Raúl Suárez, and Jan Rosell. Hybrid mapping for the assistance of teleoperated grasping tasks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 43(2):390–401, 2012.

[20] Simone Fani, Simone Ciotti, Manuel G Catalano, Giorgio Grioli, Alessandro Tognetti, Gaetano Valenza, Arash Ajoudani, and Matteo Bianchi. Simplifying telerobotics: Wearability and teleimpedance improves human-robot interactions in teleoperation. *IEEE Robotics & Automation Magazine*, 25(1):77–88, 2018.

[21] Laura Dipietro, Angelo M Sabatini, and Paolo Dario. A survey of glove-based systems and their applications. *Ieee transactions on systems, man, and cybernetics, part c (applications and reviews)*, 38(4):461–482, 2008.

[22] Alexander Crammond, Crystal Boggio, Jared C. Delinger, Sean I. Jones, Adam Katafiasz, and Douglas R. Isenberg. Commanding an anthropomorphic robotic hand with motion capture data. In *2019 SoutheastCon*, pages 1–8, 2019.

[23] Michael T Wolf, Christopher Assad, Matthew T Vernacchia, Joshua Fromm, and Henna L Jethani. Gesture-based robot control with variable autonomy from the jpl biosleeve. In *2013 IEEE International Conference on Robotics and Automation*, pages 1160–1165. IEEE, 2013.

[24] Rui Li, Zhenyu Liu, and Jianrong Tan. A survey on 3d hand pose estimation: Cameras, methods, and datasets. *Pattern Recognition*, 93:251–272, 2019.

[25] Umberto Scarcia, Roberto Meattini, and Claudio Melchiorri. Mapping human hand fingertips motion to an anthropomorphic robotic hand. In *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 774–779. IEEE, 2017.

[26] Joao Bimbo, Claudio Pacchierotti, Marco Aggravi, Nikos Tsagarakis, and Domenico Prattichizzo. Teleoperation in cluttered environments using wearable haptic feedback. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3401–3408. IEEE, 2017.

[27] Ankur Handa, Karl Van Wyk, Wei Yang, Jacky Liang, Yu-Wei Chao, Qian Wan, Stan Birchfield, Nathan Ratliff, and Dieter Fox. Dexpilot: Vision-based teleoperation of dexterous robotic hand-arm system. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9164–9170. IEEE, 2020.

[28] C Mizera, T Delrieu, V Weistroffer, C Andriot, A Decatoire, and J-P Gazeau. Evaluation of hand-tracking systems in teleoperation and virtual dexterous manipulation. *IEEE Sensors Journal*, 20(3):1642–1655, 2019.

[29] Godwin Ponraj and Hongliang Ren. Sensor fusion of leap motion controller and flex sensors using kalman filter for human finger tracking. *IEEE Sensors Journal*, 18(5):2042–2049, 2018.

[30] Josien C van den Noort, Henk G Kortier, Nathalie van Beek, DirkJan HEJ Veeger, and Peter H Veltink. Measuring 3d hand and finger kinematics—a comparison between inertial sensing and an opto-electronic marker system. *PLoS One*, 11(11):e0164889, 2016.

[31] Hang Su, Junhao Zhang, Junling Fu, Salih Ertug Ovur, Wen Qi, Guoxin Li, Yingbai Hu, and Zhijun Li. Sensor fusion-based anthropomorphic control of under-actuated bionic hand in dynamic environment. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2722–2727. IEEE, 2021.

[32] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[33] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Y Ng, et al. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5. Kobe, Japan, 2009.

[34] Greg Welch, Gary Bishop, et al. An introduction to the kalman filter. 1995.

[35] Pietro Cerveri, Elena De Momi, N Lopomo, Gabriel Baud-Bovy, RML Barros, and Giancarlo Ferrigno. Finger kinematic modeling and real-time hand motion estimation. *Annals of biomedical engineering*, 35(11):1989–2002, 2007.

[36] John P Bentley. *Principles of measurement systems*. Pearson education, 2005.

[37] Paul Regtien and Edwin Dertien. 3 - uncertainty aspects. In Paul Regtien and Edwin Dertien, editors, *Sensors for Mechatronics (Second Edition)*, pages 39–60. Elsevier, second edition edition, 2018.

[38] Thomas Feix, Javier Romero, Heinz-Bodo Schmiedmayer, Aaron M Dollar, and Danica Kragic. The grasp taxonomy of human grasp types. *IEEE Transactions on human-machine systems*, 46(1):66–77, 2015.

[39] Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage publications Sage CA: Los Angeles, CA, 2006.

A | Appendix

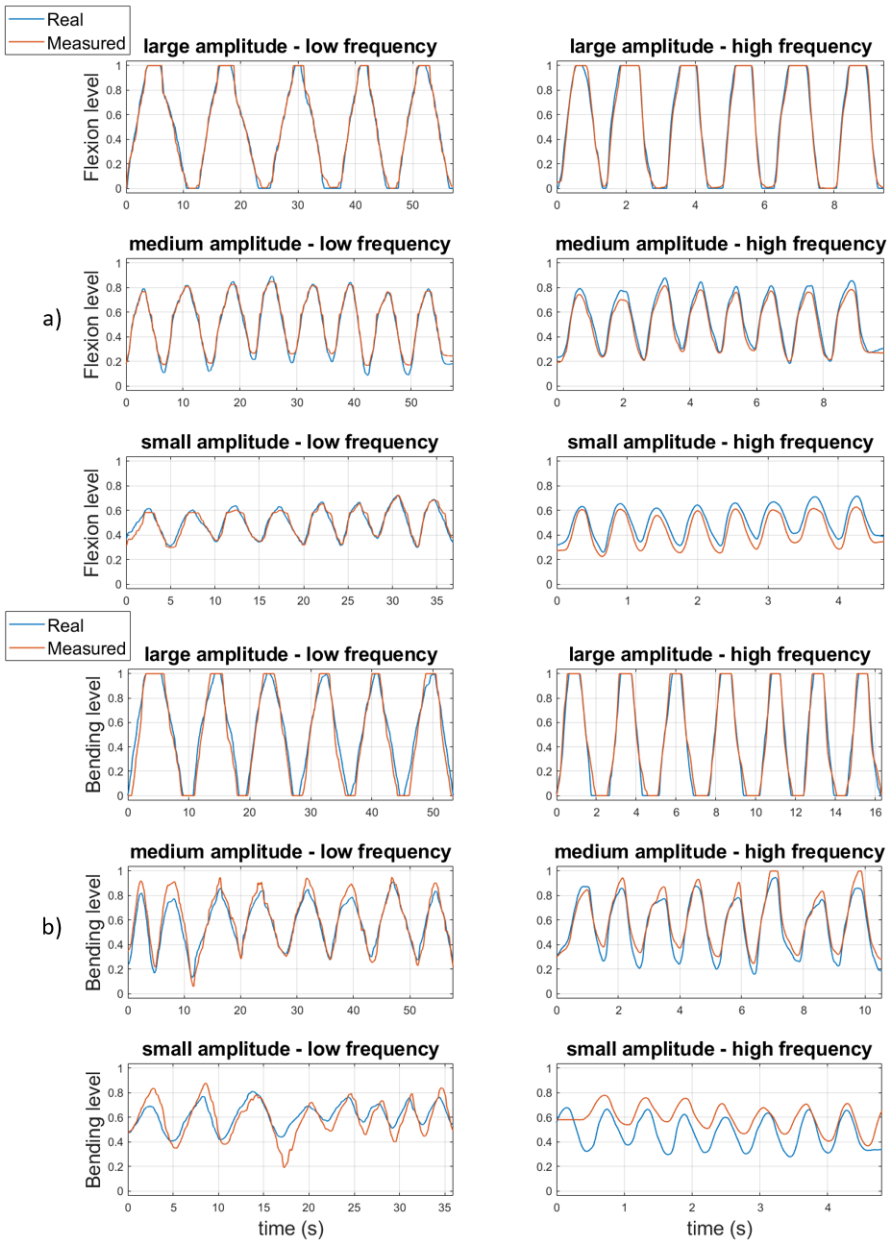


Figure A.1: Dynamic trials of the tracking systems: synchronized human and robot signals for wearable (a) and vision (b) systems.

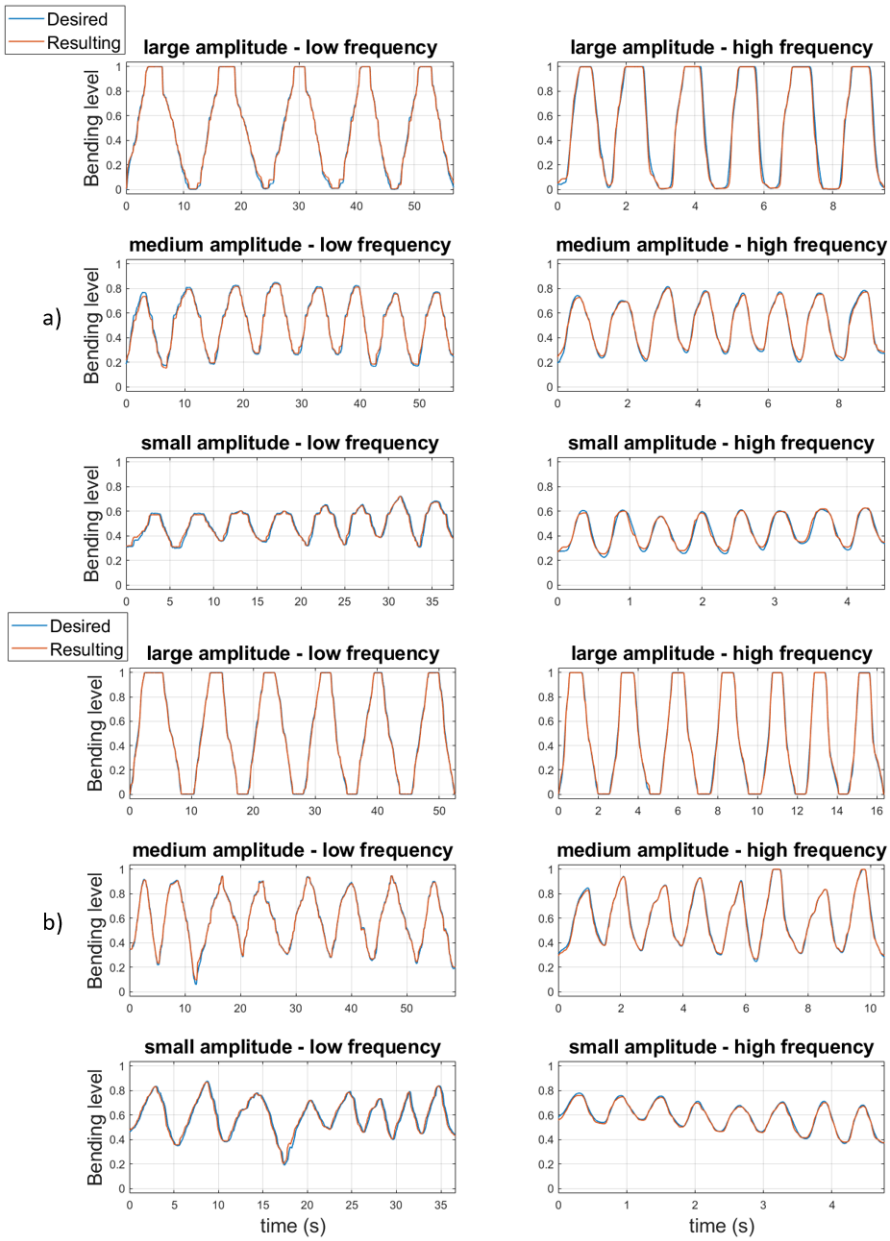


Figure A.2: Dynamic trials of the actuation control system: synchronized human and robot signals for wearable (a) and vision (b) teleoperation systems.

List of Figures

List of Tables