POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

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

Laurea Magistrale in Biomedical Engineering - Ingegneria Biomedica

Author: Massimiliano Poletti

Advisor: Prof. Elena De Momi

Co-advisor: Junling Fu

Academic year: 2021-2022

1. Introduction

Service robotics often has to face dynamic and unstructured environments, which make sensorial information less reliable and autonomous decision-making harder. The incorporation of human intelligence through teleoperation and the design of bio-inspired components, like bionic hands, can grant fast adaptation and effective interactions with irregular environments [1]. Underactuated bionic hands are especially promising thanks to their simplified driver structure and compactness.

Such devices can be intuitively controlled by capturing the human operator's hand motion and mapping it to the bionic hand in such a way to grant semantic correlation between human and robot movements. To this end, the mainly used hand motion capture technologies are mechanical gloves [2] and vision-based tracking devices [3]. The former outperform the latter in terms of stability and robustness. Indeed, optical environmental factors, self-occlusion and self-similarities between fingers may affect vision-based tracking performance. However, in the case of gloves, the single fit size, the cumbersome wearing process and unergonomic structures lead to poor comfort and affect the intuitiveness and transparency of the operation, thus making vision-based methods competitive thanks to their contactless nature.

Quantitative comparisons of wearable and vision-based tracking methods can be a basis for selection. Nevertheless, studies found in literature mainly address the evaluation of the accuracy of the motion capture system alone, without considering the performance of the resulting teleoperation system [4]. Moreover, most of the studies related to the development of bionic hand teleoperation systems are oriented to self-validation, preventing a meaningful comparison between different works [5]. Indeed, although the tasks and metrics are widely agreed upon (grasping task, success rate and completion time, respectively) there are no standards to define the test.

The objective of this work was to design two teleoperation systems to intuitively control the same underactuated bionic hand, using a glove-based method and vision-based technique, respectively, and compare them in terms of teleoperation accuracy and usability. First, the bionic hand inverse kinematics was modeled. Then, a glove-based and a vision-based hand tracking systems were designed to retrieve human motion

parameters suitable to control the bionic hand. Thus, an actuation stage to control the bionic hand motors, based on inverse kinematics, was developed. Lastly, 3 experiments were executed to evaluate the two resulting teleoperation systems: (1) static accuracy evaluation of the two overall teleoperation systems and their internal stages (the hand tracking systems and the bionic hand actuation controller); (2) dynamic accuracy evaluation of the two overall teleoperation systems and their internal stages; (3) evaluation of usability and user experience.

2. Methods

2.1. Bionic hand calibration

Figure 1 depicts the structure of the underactuated bionic hand used in this work. Each finger is composed by 3 primary links resembling the human finger phalanxes and connected by 3 hinge joints. Additionally, 2 transmission links are used to apply the underactuation paradigm by reducing the chain freedom. Each finger is independently driven by one servo motor controlled by a Pulse Width Modulation signal.

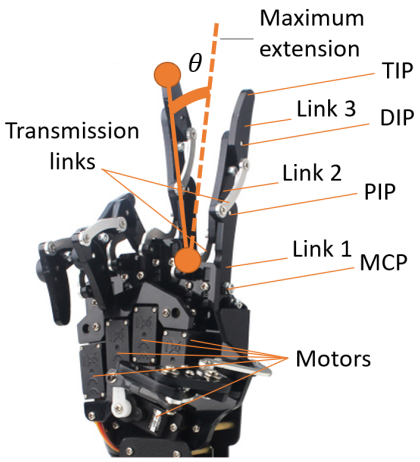

Figure 1: Bionic hand structure. Distal Interphalangeal joint (DIP), Proximal Interphalangeal joint (PIP), Metacarpophalangeal joint (MCP), bending angle (θ).

For each finger, a bending angle θ is defined as the angle between the current vector going from MCP to TIP and the finger direction corresponding to the maximum extension pose, as shown in Figure 1. The inverse kinematic model of a finger is a mathematical relationship that gives the PWM signal width, PW, of the corresponding motor as a function of the finger bending angle. In this work, for each finger, a nonlinear inverse kinematic model was found in order to catch the nonlinearity introduced by motor saturation, mechanical backlash, and hysteresis. To this end, a calibration procedure was performed using a set of markers and an optical sensor (NDI Optotrak Certus) to measure θ. For each finger, two nonlinear functions, $PW = f_{ext}(\theta)$ and $PW = f_{flex}(\theta)$, were fitted on the acquired samples to model the extension (decreasing θ) and flexion (increasing θ) motions, respectively. A neural network-based fitting method was employed. For each finger, the maximum bending angle, θ_{max}, was defined as the one associated to the maximum PW value.

2.2. Glove-based motion tracking

In the glove-based tracking method, an exoskeleton glove endowed with five potentiometers was employed. Each sensor is connected to the first phalanx of the corresponding finger and can be used to retrieve the flexion angle φ of the phalanx normalized for its range $\varphi_{\max}$ (Figure 2).

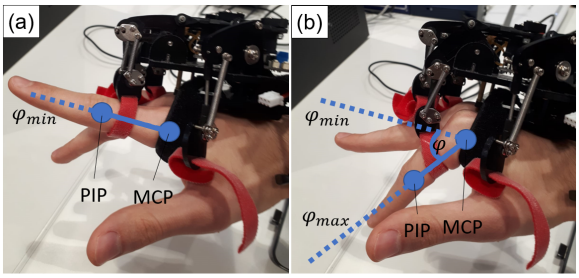

Figure 2: Definition of flexion angle: minimum (a) and maximum (b) flexion poses allowed by the glove.

The glove readout is sent to a computer running Ubuntu 16.04 and Robot Operating system (ROS) Kinetic via USB cable. For each finger, the ROS network samples the signal S of the corresponding potentiometer at 40Hz and extracts the normalized flexion angle $\hat{\varphi}$ as follows:

$$\hat{\varphi} = \frac{\varphi}{\varphi_{\max}} = 1 - \frac{S - 500}{2000} \qquad (1)$$

A real time linear Kalman Filter, which adopts a discrete-time state space representation of the finger motion, is employed to smooth $\hat{\varphi}$ and improve the robustness of the control. The implementation details can be found in a previous work about sensor fusion control [6].

2.3. Vision-based motion tracking

In the vision-based method, a Leap Motion Controller (LMC) is utilized to retrieve two infrared stereo images of the bare hand, that are then sent via USB cable to a computer that runs the Leap Motion Service (LMS), Ubuntu 16.04, and ROS Kinetic. The LMS reconstructs the 3D coordinates of the human hand joints and fingertips with respect to an absolute reference system $\{LMC\}$, as illustrated in Figure 3. Moreover, it provides the parameters describing a local reference frame $\{HH\}$ attached to the hand. The ROS network samples the LMS output data at 40Hz and transforms joints and fingertip coordinates to $\{HH\}$. Then, it computes the finger bending angle β defined by the projection of the MCP-to-fingertip vector on the flexion plane P, fixed to the hand frame, with respect to the vector belonging to P that represents the minimum bending pose, as shown in Figure 3. Such angle is mainly influenced by MCP flexion, like the bionic hand finger bending angle θ, and it is minimally influenced by MCP abduction, which is not present in the bionic hand MCP joints. P, the minimum bending pose and the maximum bending pose, which ultimately defines the maximum bending angle β_{max}, are measured during a calibration phase when the system starts.

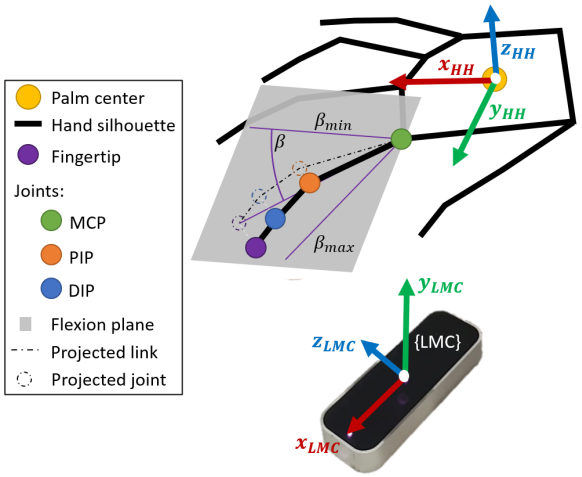

Figure 3: Leap Motion readout and bending angle computation.

Then, β is normalized for β_{max} to obtain a dimensionless variable $\hat{\beta}$:

$$\hat{\beta} = \frac{\beta}{\beta_{max}} \tag{2}$$

Lastly, a real time linear Kalman filter is used to smooth the value of $\hat{\beta}$ and improve the robustness of the control. The approach is the same as the one applied in the glove-based system case.

2.4. Teleoperation control

The two developed hand tracking stages are used to implement two distinct teleoperation systems to control the bionic hand. In both cases, for each finger, the normalized angle measured by the tracking system ($\hat{\varphi}$ for the glove-based method and $\hat{\beta}$ for the vision-based method) is multiplied for the bending angle range, θ_{max}, of the corresponding bionic hand finger, thus obtaining the desired bending angle θ:

$$\theta = \begin{cases} \hat{\varphi} \cdot \theta_{max} \text{ (glove-based system)} \\ \hat{\beta} \cdot \theta_{max} \text{ (vision-based system)} \end{cases} \tag{3}$$

Then, an actuation control stage applies the inverse kinematic model of the bionic hand finger, described in Section 2.1. The two functions constituting the model, $f_{flex}(\theta)$ and $f_{ext}(\theta)$, are used according to the current direction of motion. Calling θ_k, PW_k and d_k the desired bending angle, the pulse width of the motor control signal and the motion direction, respectively, at time k, the following algorithm is applied:

$$PW_k = \begin{cases} f_{flex}(\theta_k) \text{ , if } d_k = flex \\ f_{ext}(\theta_k) \text{ , if } d_k = ext \end{cases} \tag{4}$$

2.5. Experimental evaluation

Static accuracy assessment A calibration procedure was performed on each developed teleoperation system to evaluate their static accuracy, In particular, while a human operator was controlling the bionic hand, a set of markers and the NDI optical sensor were used to measure the bending angle θ of the bionic hand index finger and, at the same time, the ground truth of the flexion angle, φ_{real}, and the bending angle, β_{real}, of the human index finger. φ_{real}, $\hat{\varphi}$ and θ were recorded for the glove-based system, β_{real}, $\hat{\beta}$ and θ were measured for the vision-based 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 $\hat{\theta}$:

$$\hat{\varphi}_{real} = \frac{\varphi_{real}}{\varphi_{real,max} - \varphi_{real,min}} \tag{5}$$

$$\hat{\beta}_{real} = \frac{\beta_{real}}{\beta_{real,max} - \beta_{real,min}} \quad (6)$$

$$\hat{\theta} = \frac{\theta}{\theta_{max}} \quad (7)$$

While θ_{max} was known from Section 2.1, the angle limits of φ_{real} and β_{real} were unknown, therefore, they were estimated. 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. The linear function that best fits such cycles was computed through least-squares linear regression. The angle values that according to the linear model correspond to the minimum and the maximum tracking system readout were considered as the limits of the angle. After normalization, the resulting data-set contained samples of the input-output static characteristics listed below and represented in Figure 4, which should ideally give output equal to input:

- Overall teleoperation system:
  $\hat{\theta} = g_{tele,1}(\hat{\varphi}_{real})$, $\hat{\theta} = g_{tele,2}(\hat{\beta}_{real})$
- Hand motion tracking system:
  $\hat{\varphi} = g_{track,1}(\hat{\varphi}_{real})$, $\hat{\beta} = g_{track,2}(\hat{\beta}_{real})$
- Actuation control system (independent from the tracking system used):
  $\hat{\theta} = g_{act,1}(\hat{\varphi}) = g_{act,2}(\hat{\beta})$

For each characteristic, $g(x)$, the best approximating mathematical function was found by fitting polynomials on the corresponding acquired samples through nonlinear least-squares regression. In particular, the samples related to flexion motions were fitted separately from the ones related to extension motions, thus obtaining two functions, $f(x)$ and $e(x)$, which model flexion and extension, respectively:

$$g(x) = \begin{cases} f(x) \text{ , for flexion} \\ e(x) \text{ , for extension} \end{cases} \quad (8)$$

Figure 5 reports a result example. Then, for each characteristic $g(x)$, two metrics were computed, which assess the accuracy of the corresponding system:

- Nonlinearity, that is the maximum distance between the real characteristic and the ideal one (output equal to input):

$$NL = \max ||g(x) - x|| \quad (9)$$

- Hysteresis, that is the maximum distance between the two hysteretic curves.

$$Hys = \max ||f(x) - e(x)|| \quad (10)$$

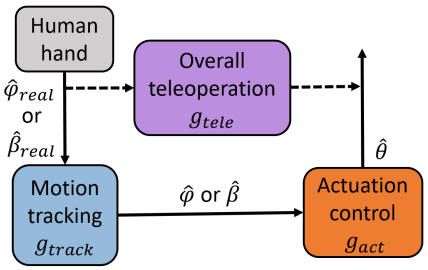

Figure 4: Input-output characteristics scheme.

Dynamic accuracy assessment In the dynamic experiment, the human operator had to continuously move his index finger in such a way that the corresponding flexion or bending angle varied sinusoidally. Such motion was mapped to the bionic hand using the implemented systems. The optical system used for the static assessment was employed to measured the same variables at 40Hz, and the same angle normalization was performed on the acquired signals. The input and output signals of the overall system, the tracking stage and the actuation controller stage were paired and synchronized, as shown by the result example reported by Figure 6. The distance between the synchronized signals was extracted as a metric of dynamic accuracy, computed as root mean square error:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2} \quad (11)$$

where N is the number of samples, x_i and y_i are the i^{th} samples of the two signals respectively. The experiment was repeated in six different conditions, each one characterized by a different amplitude or frequency of the flexion or bending angle signal in order to explore the accuracy dependency on such parameters.

Usability assessment To evaluate and compare the feasibility of the two implemented systems and assess the user experience, a third experiment was executed. 6 users performed a grasp task on 4 objects requiring different grasping postures that can be divided into two groups:

power grasp and precision grasp. While the bionic hand was fixed to a table, the user had to bring one object to it using the left hand and then make it grasp such object using one of the developed systems. Once the user was confident enough, he/she had to leave the object to the bionic hand and hold it for 5 seconds. If the object fell or touched the palm (only when a precision grasp was required), the trial was deemed a failure, otherwise it was a success. Each user performed 6 trials for each object with both the glove and vision tracking methods. The success rate of all the trials was measured. Furthermore, after using each system, each user was asked to complete a NASA TLX questionnaire to assess the workload perceived during the task.

3. Results and Discussion

3.1. Accuracy evaluation

Overall teleoperation accuracy The resulting static characteristic of the overall teleoperaion systems are represented in Figure 5. Table 1 reports the RMSE of the samples from the flexion and extension curves, the non-linearity and the hysteresis metrics.

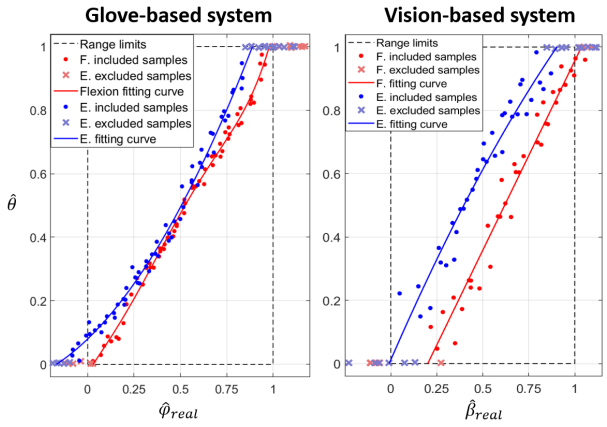

Figure 5: Static characteristics of the overall glove-based and the vision-based teleoperation systems and fitted samples.

	RMSE		NL	Hys
	f	e		
Glove	0.023	0.027	11.1%	14.6%
Vision	0.056	0.061	20.0%	27.0%

Table 1: Goodness of fitting and error metrics of the two teleoperation systems. f: flexion; e: extension; NL: nonlinearity; Hys: hysteresis.

The glove-based system is characterized by a linear region from 25% to 60% of the flexion level range, overlapped to the ideal characteristic. Non-linearity and hysteresis are significantly present only outside this region, however, the respective metrics are acceptable. In the vision-based system, instead, 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. Non-linearity and hysteresis are significantly worse than in the glove system case. Fitting results, represented by the RMSE, are poorer too, hinting at lower behavior predictability, repeatability and precision.

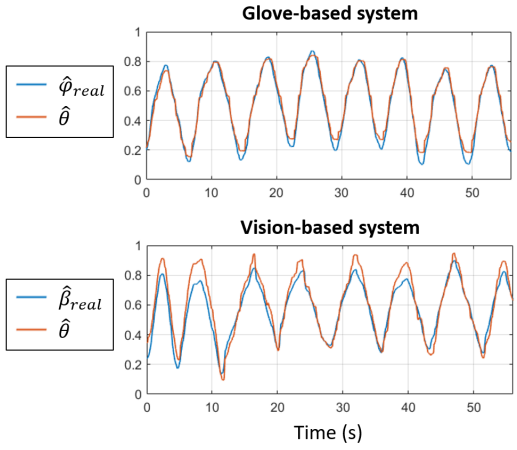

Figure 6: Dynamic accuracy comparison.

In the dynamic trials, the RMSE values measured for the glove system are lower in all six conditions, confirming the higher accuracy of such method. Indeed, as shown by the example in Figure 6, the glove-based system shows better-overlapped signals and less pronounced missed peaks. Moreover, the vision-based system performs worse for slow and small movements, indicating that it is less suitable for fine and delicate operations.

Tracking accuracy The resulting static characteristics of the tracking systems are very similar to the ones of the overall teleoperation systems. The main differences consist of a slightly lower dispersion around the curves and a lower hysteresis in the lower part of the human range. The vision-based tracking method gives significantly worse accuracy metrics for both static and dynamic scenarios with respect to the glove-based one. Since the actuation control stage is

the same for both systems, the lower accuracy of the vision-based teleoperation method is solely due to the worse accuracy of its tracking stage. Furthermore, given that in both the developed teleoperation systems the processing on the sensor data is reduced to the minimum, such worse performance can be attributable to the limited accuracy of the Leap Motion device in estimating joints position, implying a reduced applicability of such technology for being used, with simple data processing, to control bionic hands for fine and precise tasks.

Actuation control accuracy Lastly, the static characteristic of the actuation control system displays an excellent superimposition to ideal behavior over 20% of the range, with low sample dispersion. Below such region, nonlinearity and hysteresis reach 4.1% and 7.4% of the range, which are not relevant values (about 5° and 9° respectively). Furthermore, in all six dynamic situations, the resulting robot path closely follows the desired one, and RMSE is low. The good performance of the actuation control system is confirmed also by the high similarity between the results of the overall teleoperation and the ones related to motion tracking alone. As a result, the inverse kinematic model resulting from finger motion calibration, if applied through a suitable algorithm, provides generally accurate and reliable actuation control system.

3.2. Usability evaluation

All the participants succeeded in performing all four grasp tasks using both the developed systems. 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. According to a Fisher's exact test, there is no evidence to claim a significant statistical difference between the two values. Table 2 reports the results from the questionnaire in terms of the average (μ) and the standard deviation (σ) of the perceived workload components. The total workload is the sum of the average workload components and it is practically the same in the two systems. However, in the case of the glove, on average, the global effort is perceived as higher, and performance is perceived as higher too. Free user feedback is in line with such result: users felt the glove was better performing but, at the same time, less comfortable, causing greater effort. Therefore, the inaccuracy issue of the vision-based method does not compromise the teleoperation control, nevertheless, it influences the perceived performance. As regards the glove-based method, user's comfort is a critical issue.

Index	Glove		Vision	
	μ	σ	μ	σ
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 2: User study: average (μ) and standard deviation (σ) of the workload indexes, total workload (sum of the average workload indexes).

4. Conclusions

In this work, two teleoperation systems were developed to intuitively control the same underactuated bionic hand, based on the motion capture of the human hand through a wearable mechanical glove and vision-based tracking device, respectively. Both proved to be able to effectively perform both power and precision grasp tasks successfully. However, the vision-based system revealed a significantly lower accuracy than the glove-based one, which is due to the limited accuracy of the vision tracking method in estimating joints position, implying a reduced applicability for completing fine and precise tasks. Such worse accuracy does not compromise grasp task success but has a negative influence on the user's perception of performance. The glove-based solution, instead, reduced user comfort by impeding natural movements, thus, although wearable devices are more accurate than vision-based ones, their use may be limited if the user's comfort is important for the application. Future research may focus on three aspects: (1) the design of a sensorized glove that improves the user comfort; (2) the employment of sensor fusion techniques to improve the accuracy performance of the vision-based method; (3) the realization of a comprehensive hand-arm teleoperation system.

References

[1] 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.

[2] 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.

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

[4] 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.

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

[6] 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 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2722–2727. IEEE, 2021.