A natural interface based on intention prediction for semi-autonomous micromanipulation

Laura Cohen, Mohamed Chetouani, Stéphane Régnier, Sinan Haliyo

Intelligent Service Robotics manuscript No.
(will be inserted by the editor)

A natural interface based on predictive intention extraction for micromanipulation

Received: date / Accepted: date

Abstract Micromanipulation tools are not yet commonly used in the industry or in the research due to the lack of natural and intuitive human-computer interfaces. This work proposes a "metaphor-free" interface for a pick-and-place operation for semi-operated teleoperation. A predictive intention extraction technique is proposed through a computational model inspired from cognitive sciences and implemented through a *Kinect* depth sensor. This allows a more natural interaction without any prior instructions to the operator. The model is compared to a gesture recognition technique in terms of naturalness and intuitiveness. It shows an improvement in user performance in duration and success of the task, and a qualitative preference for the proposed approach evaluated by a user survey.

Address(es) of author(s) should be given

Keywords human-robot interaction · microrobotics · intention extraction · gesture recognition · pick-and-place operation · kinect

1 Introduction

Manipulation of micro objects is a key issue for further developments in nanotechnology and biology. At this scale, typical manipulation tasks involve pick-and-place of micro objects, in unstructured environments. Robotic manipulation open access to these scales, however full automated operation is a still a standing issue due to complex and counter-intuitive force fields, and the high influence of environmental conditions [6] [15]. Teleoperation is hence the generally adopted approach to give the control of the robotic system to the user to carry out the task through the scale barrier. Nevertheless, the vision feedback and sensor data of micro-

manipulation systems are generally limited. Dedicated interfaces between the user and the micro world are called for to overcome the difficulties of remote human interaction to small scales.

Recent works focus on the development of novel interfaces to provide information from the microworld toward the natural human sensory modalities. Haptic interfaces enable the operator to touch these intangible scales [4], and virtual reality to see and interact with the microworld in 3D [18]. These techniques assist the user to apprehend the task in spite of uncertainties in micro and nanoscale physical environments.

Despite the recent progress, the lack of intuitive interfaces limits the adoption of micromanipulation systems. The task is still conducted manually, and requires the operator to directly control the position or velocity of the robot with a joystick or a haptic device. From the operator's point of view, these interfaces remain complex to use, unhandy and hard to manage. For example, naive users mention the difficulty to identify the interface's handle to the real end-effector. Furthermore, these interfaces require a learning phase because of the use of a specific symbolic language. Several works in the literature agree on the necessity to create more natural and intuitive interfaces for widespread adoption of microrobotic technologies, but this specific issue has not been addressed yet. A promising approach is semi-automated manipulation: the whole operation is divided in simple automated tasks. The operator controls the transition between those tasks and their setpoints.

New solutions for the detection of user actions are currently emerging in the field of macroscale Human Computer Interface (HCI), with low-cost RGB-Depth sensors (e.g. Microsoft Kinect) based on natural interaction modalities such as hand gesture recognition. This kind of interfaces, dedicated to macroscale interaction, appears also as a promising solution for micromanipulation as they are intuitive and does not require markers or gloves.

Current macroscale approaches widely use gesture recognition methods to detect predefined commands. Ren and O'Neill [16] propose a 3D selection method that uses the gesture direction to determine the target object and confirm selection by raising the other hand. This kind of interfaces therefore requires to learn and remember a complex symbolic language [12]. Furthermore, their low flexibility makes them poorly robust to gestures that differ from the predefined dictionary. Even if the gesture modality seems "natural", the need to produce a discrete predefined gesture remains as symbolic as clicking on a button to trigger an action. It implies a learning phase and thus is ill-fitted for non-specialist users.

To address this issue, "symbol-free" interfaces have been proposed [14]. The aim of this approach is to enable the user to interact with virtual objects in the same way as they interact with the real world. Hence, the operator relies on already mastered skills and doesn't need any instruction.

This approach raises many challenges. An interface that would need no instructions to specify the constraints of the system to the user has to handle the complexity of free interaction. A symbol-free interface thus requires interpreting naturally human actions. Humans are experts in understanding others' behavior. Thus, a promising approach is to exploit models issued from cognitive sciences on how humans infer others' intentions by observing their actions.

This work provides a natural interface to assist the teleoperation of a micromanipulation system. Proposed system is experimented on a dedicated AFM (*Atomic Force Microscopy*) simulator from the literature [10]. User movements are tracked using a Kinect sensor to allow free interaction. A symbol-free intention extraction model is proposed to grab and drop a microsphere, based on cognitive sciences. This approach is compared to a classical hand gesture recognition system in terms of naturalness and intuitiveness.

2 AFM based micromanipulation

At microscale, typical interactive manipulation tasks consist in picking, transfering and placing micro objects on a substrate. An example strategy of micromanipulation that has been experimentally demonstrated for this purpose is the pick-and-place by adhesion with an AFM cantilever [7] [8]. AFM has the advantage to provide force sensing in the micronewton range and is the most used manipulation tool in micro and nanoscales [2].

2.1 Manipulation by adhesion

For objects with dimensions less than 100μm, adhesion forces (namely van der Waals, capillary and electrostatic) are stronger than gravitation forces. These forces can be advantageously used for manipulation. An AFM cantilever is used for the manipulation as shown on fig. 1. Objects are picked by simple contact due to the superiority of the adhesion vs. weight. Then, objects are placed on a substrate with a higher adhesion than the cantilever's [7].

For an AFM based micromanipulation, the semi-automation can be implemented with these three elementary tasks:

- picking an object by adhesion: the cantilever touches down the sample and removes it from the substrate,

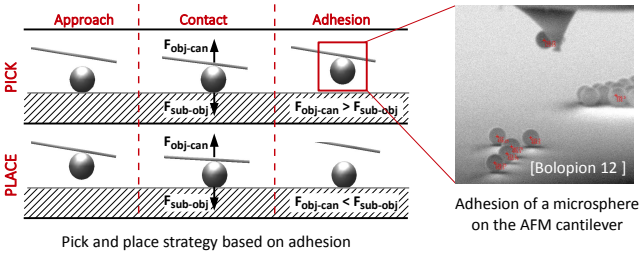

Fig. 1 Pick (up) and place (down) of a microsphere with an AFM cantilever by adhesion. F_{i-j} is the adhesion force between i and j.

- transfering the object while avoiding contact with the substrate or other samples,
- placing the object on a target site: touching down the object then release from cantilever.

Due to the force sensing of the AFM cantilever, the contact and adhesion force measurements are used to detect the success or failure of the pick and place tasks.

2.2 Teleoperation of the AFM cantilever

Some recent works have focused on providing both haptic and 3D virtual reality feedbacks to enable the operator to manipulate micro objects intuitively in a wide range of applications [3] [10]. Virtual reality (*VR*) environments are interesting solutions to provide users with a reconstructed 3D view of a manipulation scene, which can be displayed at the human scale and from different vantage points. This overcomes the limitations of vision sensors such as optical or electron microscopy which lacks the sense of depth. This VR layer approach is depicted in fig. 2.

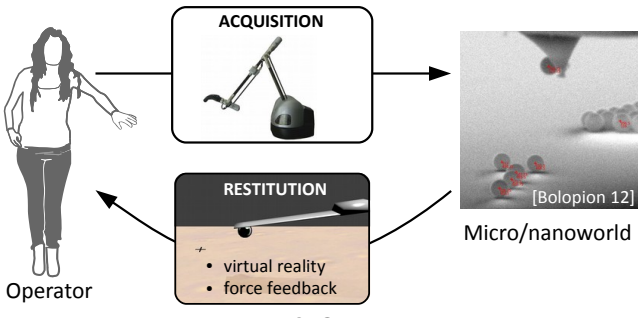

Fig. 2 Teleoperation of a micromanipulation system through a VR environment. The operator is 600 km away from the manipulation chamber [3].

In a previous work [9], authors used a haptic arm to teleoperate an AFM cantilever in a VR environment. This includes a simulator which realistically reproduces the adhesion and dynamics of a real micro manipulation setup and was used to evaluate haptic teleoperation. It's composed of a flexible virtual cantilever that can be displaced depending on the system kinematics. The adhesion forces between the object, the substrate and the cantilever shown on fig. 1 are realistically modeled. The interaction is measured from the cantilever deflection as in a real AFM system. A computational physics engine enables real-time interaction. The haptic feedback allows the user to feel interactions like contact, adhesion and pull-off phenomena.

This simulator is used here to impersonate a real micromanipulation system, in order to circumvent the difficulties to set up repetitive real-world operations and to focus on the development of the interface.

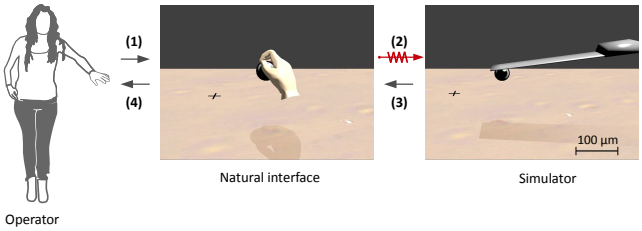

Fig. 3 Teleoperation of a VR micromanipulation simulator. (1) The user's hand motions are used to move the virtual hand in the natural interface. (2) a spring-and-damper coupling is used to teleoperate the AFM cantilever. (3) The failure or success of the pick/place task is fed back towards the natural interface to give visual feedback to the operator (4).

2.3 Virtual reality interface and coupling

To establish a direct link between the user's hand and the end-effector, a natural interface layer is added to the system. In this interface, the user interacts through a hand displayed directly in the VR environment. The Kinect skeleton data of the operator is used to transcribe the hand's spatial position and rotation onto the interface. The hand position is then used to teleoperate the end-effector as described in fig 3 through a spring-and-damper coupling in transfer phases, and to select the target on pick and place phases. During the transfer phase, the cantilever holding the object is kept at a certain distance above to substrate to avoid collisions; only its planar motion is coupled to the hand position.

The transition between pick and transfer, and between transfer and place phases are controlled through the detection of actions of the operator. The working principle is to detect his intention to "grab" (respectively to "drop") and move to end-effector accordingly to contact with the object/substrate and back. For each a corresponding animation of the virtual hand is created to provide visual feedback. When an action is recognized, the corresponding animation is played and the grab/drop of the microsphere is triggered. This detection is the key component of the semi-automated approach presented here and is the focus of this manuscript.

Two approaches are compared to detect actions performed by the operator. A classical gesture recognition method, based solely on visual feedback is presented in the next section and is then compared to a novel predictive intention extraction approach, detailed in the following section. The AFM simulator is used to validate and compare these two approaches.

3 Hand gesture recognition

To detect the grab and drop actions, an initial method is to use gesture recognition techniques available in the

literature. In order to avoid the constraint of learning symbolic gestures, a first approach is to use "natural" gestures. As the proposed task involves the pick-and-place of a microsphere, two natural gestures are used: a closing hand near a sphere triggers the pick and an opening hand near a target triggers the drop. The Microsoft Kinect SDK hand gesture recognition library is used as it is a robust and position-invariant method.

A preliminary evaluation of this system consists in the observation of users interacting without any instruction on the method to pick-and-place the sphere. Some operators are observed while performing pick-and-place tasks using the hand gesture based interface.

A first issue is that the animation can be triggered only when the gesture is recognized, that is when the gesture is already completed. Therefore, users report an impression of delay between the performed gesture and the virtual hand animation, even if the gesture recognition operates in real-time.

The second point is that users tend to not fully close their hand, as if to adapt it to the size of the virtual object to be grabbed. Furthermore, the rest position of the hand is neither totally open nor closed. Despite the robustness of the hand gesture recognition method, many false negatives are thus observed.

The interface thus seems sufficiently close to a real interaction for the user to behave as if to interact with a real object. However, this likeness harms the robustness of the detection.

There is therefore a duality between the natural behavior of the user towards the interface and the symbolic language used by it. This observation shows that it is not sufficient to mimic a natural behavior, like grabbing/dropping an object by associating to it a hand open/closed recognition. To address this issue, a symbol-free intention prediction model is proposed in the next section. The gesture recognition approach is used in this work as a baseline to evaluate this intention prediction model.

4 Computational model of intention prediction

The observations from the previous section show that designing a natural interface requires fitting to the natural behavior of the user.

In order to perform the task, two pieces of information are required: what object the user wants to interact with, and what action to perform on this object. An action is composed of two parts: an intention and a movement. In cases of deliberative action, the action is caused by what Searle [19] terms a prior intention, that is, an intention to act formed in advance of the action itself. A same movement can be caused by different prior intentions, for example a person can grasp an apple to eat it or to hand it to another person. Becchio

et al. [1] show that human observers can infer the prior intention of others by observing their actions. In this paper, this capacity is termed as intention extraction.

Building a computational model of a predictive intention extraction is an interesting perspective. In fact, it is centered on the goal of the user and it allows an early detection, thus avoiding the delay between the gesture recognition and the visual feedback. Intention is a high level notion and thus cannot be directly extracted by a sensor. It is however correlated to low level data that are quantifiable. In this work, a high level intention prediction model is proposed based on low level features extracted from the Kinect sensor.

In order to establish a computational model of intention prediction, the first step consists in determining the relevant low level quantifiable data upon which to model the intention of the user.

4.1 Low level features to model the intention

The determination of the relevant low level features must meet two requirements. First, the chosen feature must contain enough information about the intention of the user. The second one is that the feature has to be sufficiently invariant regarding the intention of the user.

4.1.1 Features of human intention prediction

The first requirement to select the pertinent features is to determine if they contain enough information about the intention of the user. Humans are experts in intention extraction. Even before grabbing an object, during the reach phase, a human observer can infer what action the actor wants to perform, and what is their target object. This capacity is essential to interact and collaborate with others.

Several works in the field of cognitive science have studied this capacity. Becchio et al. [1] synthesize various experimental data showing that for a same gesture, for example a prehensile gesture, the kinematics features vary depending on the prior intention of the actor. Sartori et al. [17] demonstrate that kinematics features are sufficient for an observer to infer the prior intention. In their experiment, white spots placed on an actor's articulations are the only visible elements, and observers are asked to infer the intention. The performances are close to classical video results. Other studies [20] show the importance of the context in intention extraction. The prediction is more precise if the action is constrained by a target and a precise context. However, human observers principally rely on kinematic information rather than direct visual information on the target object and the context.

Gesture kinematics contains both invariant features and sufficient information to enable a human observer to infer other intention. Thus, the kinematic features are selected as low level information upon which to build the intention prediction model. For this purpose, hand gesture velocity is extracted by derivation of the position of the hand joints of the Kinect SDK skeleton library.

4.1.2 Gesture invariant features

Despite the great variability of human goal-directed gestures, there exist some invariant features. The velocity has an invariant log normal shape, that can be approximated by a Gaussian for the movements with an intermediate speed [11]. This classical profile is shown in fig. 4 from a gesture towards a target in the virtual reality simulator described in the previous section. This classical profile is expressed by :

$$v(t) = v_{max} \cdot \mathrm{e}^{-\frac{(t-\mu)^2}{2 \cdot \sigma^2}} \quad (1)$$

where μ is the mean, σ the standard deviation, v the hand velovity and v_{max} the maximum velocity (fig. 4).

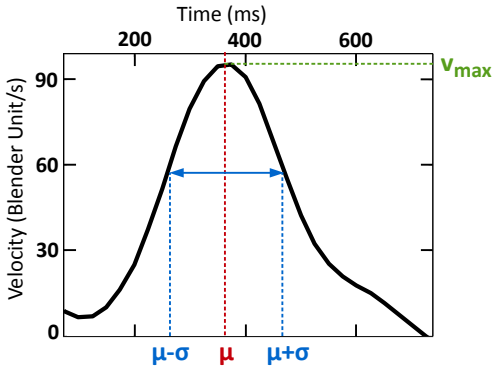

Fig. 4 Invariant velocity profile for goal-directed gesture, with mean μ, standard deviation σ and maximum v_{max}. A Blender unit corresponds to the Blender simulator default unit of measurement.

Another property of goal-directed gestures is the isochrony principle [21]. The duration of a gesture to reach a target is demonstrated to be a constant. Thus, the speed depends linearly on the distance to the target. This result is validated in the proposed architecture from 60 trials of reaching a target in the virtual reality simulator. The result is shown in fig. 5.

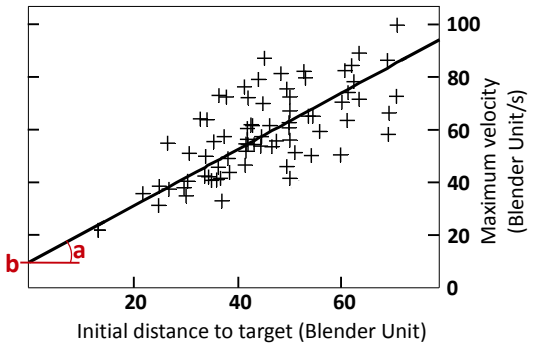

Fig. 5 Isochrony principle for goal-directed gesture. A Blender unit corresponds to the Blender simulator default unit of measurement.

This principle is expressed by the following linear equation:

$$v_{max}(d_0) = a \cdot d_0 + b \quad (2)$$

where a is the slope, b the intercept et d_0 the initial distance between the virtual hand and the target (fig. 5).

Thus, the gaussian profile of the movement and the isochrony principle are selected as invariant features in the virtual reality micromanipulation simulator context.

4.2 High level model of intention prediction

Despite the fact that humans are experts at intention extraction, cognitive models of this capacity are not exploited yet to design natural interfaces. The proposed solution is based on a model of the intention extraction by a human observing an actor performing an action.

Oztop et al. [13] propose a computational model that could give an explanation of how the same cerebral areas are used in motor planning and in understanding others' action. The observer forms an intention hypothesis, based on sensory consequences of actor's motion, according to his own learned motor behavior. At the next step, a difference is computed between the hypothesis and the observation. If this error is too high, a new intention hypothesis is then made. This model is predictive and symbol-free: observer guesses the other intention before the action is completed, and relies only on his past experiences. This model is interesting solution to answer the raised issues.

This work proposes to adapt this model to predict the intention of a user in the context of the micromanipulation interface. Figure 6 depicts the plant derived from the cognitive model of an observer extracting the intention of an actor. The plant is constructed such as the information exploited by the observer is the joint velocities of the actor.

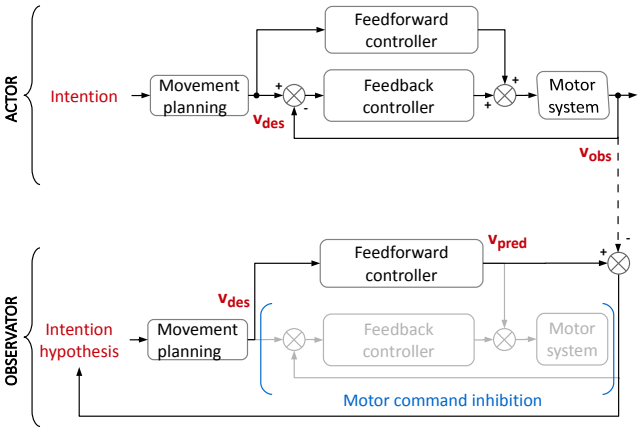

Fig. 6 Intention prediction cognitive model

In the teleoperation context, two intentions are considered: the picking of a microsphere and its placing on a specific target site on the substrate. Hence, the states space is reduced to two possible intentions. Furthermore, the intention hypothesis also depends on the context, i.e. what objects are present and what are the possible actions to perform on these objects. In the micromanipulation interface, when the hand is free, its only possibility of action is to grab an object and when a microsphere is grabbed, the only possible intention is to drop it on a site. Hence, the system takes into account of this context to establish a relevant intention hypothesis. This observation is consistent with the fact

that human intention prediction is more precise when the context is known [20].

4.3 Construction of predictors

The first step is to learn the basic motor behavior of grab and drop gestures upon which to compute the intention predictors. These predictors corresponds to the forward model (fig. 6). Gesture kinematics is a good feature for intention prediction as shown previously, thus this work exploits hand velocity in the proposed model as the sensory input signal. The invariant laws of the goal-directed movements are used to compute the predictor. The predictor profile is approximated by a Gaussian according to the invariant speed profile law. The isochrony principle states that gesture velocity increases with the initial distance to the target. The amplitude of the Gaussian should hence be proportional to the distance to the target.

Humans are apparently able to infer intention from kinematics features. Thus, this work hypothesis is that the invariant laws are modulated by the intention. An experiment is proposed to validate this hypothesis in the micromanipulation context.

The learning phase protocol consists in 20 grab gestures and 20 drops at various distances from the target performed by 3 subjects. A microsphere and a target dropping site are randomly placed on the substrate for each round. The grabbing and dropping are automatically triggered when the hand is sufficiently close to the target, in order to avoid any a priori hypothesis that could affect the naturalness of the interaction. No information is given to the user about how to grab and drop the object. The hand velocity and distance toward the target are recorded for each task.

A Gaussian fit is then performed on the hand velocity for each task to compute the amplitude, expected value and standard deviation parameters.

4.3.1 Task influence

An example result for a user learning phase is shown on fig. 7. The three parameters (amplitude (A), expected value (EV) and standard deviation (SD)) of the Gaussian are plotted as functions of the initial distance to the target. Each point corresponds to one round (in black for the grab task and in blue for the drop task). A polynomial interpolation is then performed on the points for each parameter and each task.

As expected, the amplitude depends linearly on the distance to the target. Only the first order of the polynomial interpolation is significant. This result is consistent with the isochrony law. Furthermore, the expected value is also a linear function of the distance to the target. On the other hand, the standard deviation doesn't seem to depend on the distance to the target.

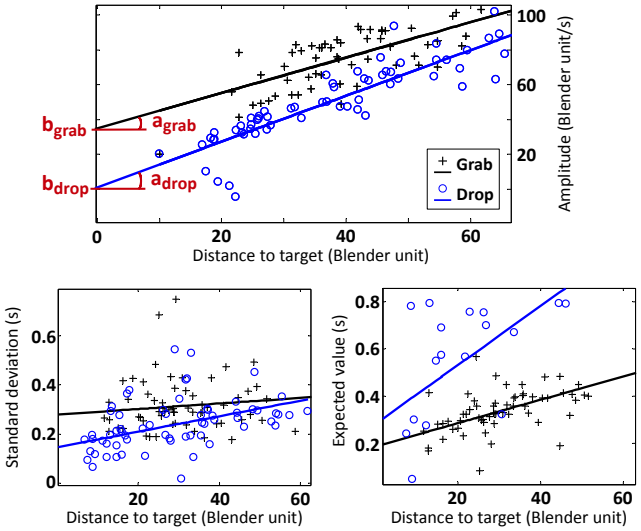

Fig. 7 Task influence on the Gaussian predictor parameters

Another interesting result is that the curves associated with the grab and the drop tasks are different. Despite the fact that both are goal-directed gestures, the amplitude and the expected value of the speed vary depending on the intention. This observation supports the fact that prior intention modifies the movement kinematics even for the same reach gesture [1,17]. Thus, different predictors need to be computed for these actions.

These observations lead to a reformulation of the isochrony principle (equation 2) and the gaussian profile (equation 1) laws taking the intention parameter into account:

$$v_{max}(d_0, task) = a_{task} \cdot d_0 + b_{task} \tag{3}$$

where a_{task} and b_{task} are the parameters interpolated linearly of the maximum velocity as a function of the initial distance to target, depending on the grad or drop task (fig. 7).

$$v(t) = v_{max}(d_0, task) \cdot \mathrm{e}^{-\frac{(t-\mu_{task}(d_0))^2}{2\cdot\sigma^2}} \tag{4}$$

where $\mu_{task}(d_0)$ is the mean value estimater for each task and for the initial distance d_0, from the linear interpolation proposed on fig.7. The standard deviation σ doesn't depend on the task.

4.3.2 User influence

Fig. 8 shows the influence of the user on the Gaussian parameters as a function of the initial distance to the target.

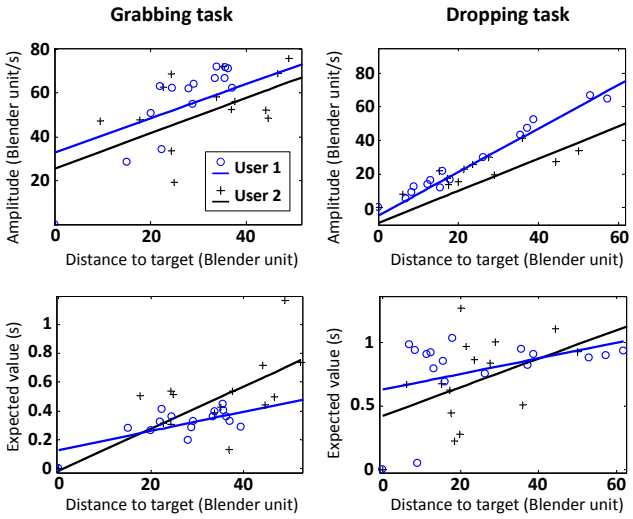

Fig. 8 User influence on the Gaussian predictor parameters

The curves are increasing and linear for all the users. Moreover, the amplitudes for the grab task are higher than for the drop task for all the users. Conversely, the

expected value is lower for the grab task. Therefore, despite the numeric differences, the Gaussian curves have similar characteristics, independent of the user. This observation shows the generality of the model.

4.3.3 Computation of predictors

The predictors $pred(t)$ are computed based on these observations according to the current possible task $\{grab, drop\}$ and to the current distance to target ($dist$), based on equations 4 and 3,

$$v_{pred}(t) = v_{max}(d_0, task) \cdot \mathrm{e}^{-\frac{(t-\mu)^2}{2 \cdot \sigma^2}} \quad (5)$$

where v_{pred} is the velocity predictor, σ and μ are the means of respectively the standard deviation and mean learned for each possible task, and $v_{max}(d_0, tâche)$ is the maximum of the gaussian predictor computed from the linear equation 3.

As the standard deviation doesn't seem to depend on the distance to target, its value is computed as the mean value of learning samples. Furthermore, the gesture segmentation is emergent from the proposed model. A gesture starts when it is predicted well enough. Thus, the mean of the gaussian is simply compared to the mean of learning samples.

4.4 Prediction method

In the first phase of the task, the user grabs the microsphere. According to the proposed model, the intention hypothesis is set on the grab intention, and the corresponding predictor is used. Algorithm 1 presents the proposed prediction method for the grab task. At each step, if the prediction mean error is under a fixed threshold ($errThresh$), the predictor is considered as good enough and it is kept. The score (N) is increased. An intention hypothesis is validated if the predictor has been good for $Npred$ points. This threshold is fixed depending on the frame rate of the sensor, shortly after the maximum of the Gaussian is reached. Once the grab intention is predicted, the intention hypothesis is set to the drop task and the following steps are repeated.

Fig. 9 shows an example of a grabbing task intention prediction. The first 2s correspond to non-goal-directed random gestures. The figure shows that the predictor is not good enough and is discarded. Following predictors are then recalculated. At t = 2s, the mean error between the predictor and the hand mean velocity ($\overline{err}$) is bellow the threshold $errThresh$ and after $NPred$ points, the grab intention is predicted, corresponding to the red dotted line. This gesture is effectively the user reaching to grab the object.

Algorithm 1 Grab intention prediction

Extract current distance to target $dist(t)$ from sensor

$v(t) = \frac{d(dist(t))}{dt}$

if $v(t) > 0$ **then**

$\overline{err} = \frac{1}{N} \cdot \sum_{i=1}^{N} (v_{pred}(N) - v(t))^2$

if $\overline{err} > errThresh$ **then**

$v_{pred}(t) = v_{max}(d_0, grab) \cdot \mathrm{e}\left(-\frac{(t-\mu)^2}{2 \cdot \sigma^2}\right)$

$N = 1$

else

$N = N + 1$

end if

end if

if $N > Npred$ **then**

Grab intention predicted

end if

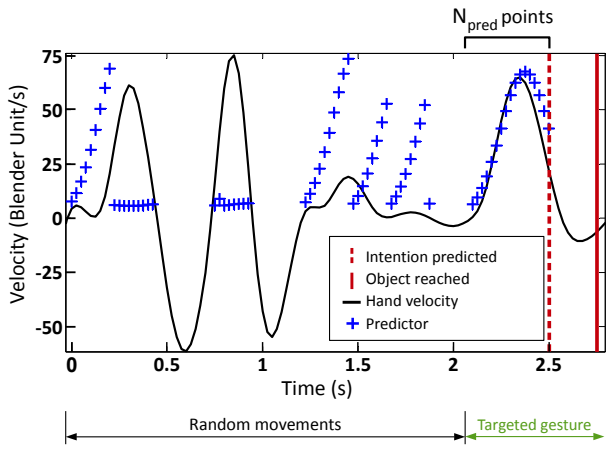

Fig. 9 Intention prediction example

Once the microsphere grab intention is predicted, the interface starts the corresponding animation of the virtual hand, in order to avoid the delay observed with the classical gesture recognition method.

5 Experimental results

5.1 User test protocol

A user test protocol is set up to evaluate the proposed interactive system and compare it to the baseline hand gesture recognition approach. The task consists in grabbing a microsphere and displacing it to a target on the substrate marked as a cross. A new configuration of the microsphere and the target is drawn randomly at the end of each round. To validate the natural aspect of the interaction, no instruction is given to the operator on the method to use to perform the task.

The experimental protocol is conducted in 3 phases. The first one is a case-control in which the grabbing and dropping is triggered only by a proximity criterion as in the learning phase. The second uses the gesture recognition method presented in section III. The third uses the intention prediction approach to test the model proposed in this work.

The order of the phases proposed to the user are changed randomly in order to avoid a learning effect on the results. Each phase is constituted of 15 rounds of grab and drop. At the end of each phase, a user questionnaire is submitted to the operator. This questionnaire is based on the System Usability Scale (SUS) [5]. Nine adult subjects took part in the experiment.

5.2 Intention prediction and gesture recognition comparative results

5.2.1 Quantitative results

The success of the task is evaluated on a proximity and duration criterion : if the hand stays near the target for more than $0.5s$ and the grab/drop is not detected, the task is considered as a failure. The success percentage for each task is shown in fig. 10 on the left. The proposed intention prediction method shows a significant improvement of 33.8% over the classical gesture recognition method for the dropping task.

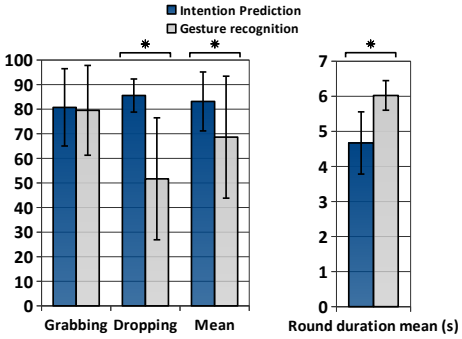

Fig. 10 Percentage of success for the two tasks with the intention prediction and classical gesture recognition approaches (left) and round duration mean (right). One asterisk (*) indicates a p value smaller than 0.05 ($p < 0.05$). The statistical analysis used is the ANOVA test.

A second quantitative result is the mean duration of each round. The intention prediction methods mean duration is $4.7s$, while the gesture recognition is $6s$ as shown in fig. 10 on the right. Thus, the proposed method allows for significant improvement in the duration of the task.

5.2.2 Qualitative results

Finally, the SUS questionnaire shows a strong preference for the model proposed in this work of 30.8% over the gesture recognition method as shown in fig. 11. In particular, the users' evaluation of the ease to manipulate the microspheres shows better results than both the hand gesture recognition method and the case-control method. The evaluation of comfort is consistent with these results, showing a significant preference for the proposed approach.

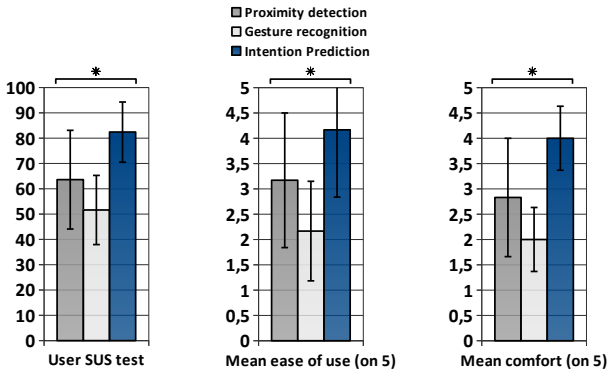

Fig. 11 Qualitative SUS user survey results. One asterisk (*) indicates a p value smaller than 0.05 ($p < 0.05$). The statistical analysis used is the ANOVA test.

5.3 Discussion

The results show the improvements of the proposed method in a context of natural interaction without any

prior instruction, both quantitatively and qualitatively. Hence, the proposed approach is validated as an appropriate solution to provide an assistance for AFM micromanipulation. However, some users mention the fact that the method seems sometimes too predictive, weakening the impression of being in control of the task.

For further improvements, the operator's focus of attention can be integrated to generalize the model to multitarget contexts. For this purpose, a predictor would be computed for each target and the attentional module would select the more pertinent predictors depending on the region of focus of attention of the user.

6 Conclusions

In this paper, a symbol-free interface is proposed for the teleoperation of an AFM cantilever in a virtual drag and drop task of microspheres. To achieve a non symbolic interaction, an intention prediction model is proposed, based on the natural invariant features of human goal-directed movement. Without any instruction to users, the predictivity of this model significantly improves the task duration over the classical hand gesture recognition method. In addition, the System Usability Scale questionnaire shows a 30.8% preference of the users for the predictive intention extraction model over the hand gesture recognition approach. Hence, these results confirm the symbol-free approach as an interesting solution for natural interaction. Furthermore, the proposed approach enables manipulation tasks independently of the hand pose and the fingers. Thus, this method can be generalized to various systems, such as a mouse or a haptic arm.

References

1. Becchio, C., Manera, V., Sartori, L., Cavallo, A., Castiello, U.: Grasping intentions: from thought experiments to empirical evidence. Frontiers in human neuroscience **6** (2012)
2. Binnig, G., Quate, C.F., Gerber, C.: Atomic force microscope. Physical review letters **56**(9), 930 (1986)
3. Bolopion, A., Stolle, C., Tunnell, R., Haliyo, S., Régnier, S., Fatikow, S.: Remote microscale teleoperation through virtual reality and haptic feedback. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 894–900 (2011)
4. Bolopion, A., Xie, H., Haliyo, S., Régnier, S.: Haptic teleoperation for 3-d microassembly of spherical objects. IEEE/ASME Transactions on Mechatronics **17**(1), 116–127 (2012)
5. Brooke, J.: SUS-A quick and dirty usability scale. Usability evaluation in industry **189**, 194 (1996)
6. Gauthier, M., Régnier, S.: Robotic Micro-assembly. IEEE press (2010)
7. Haliyo, D.S., Regnier, S., Guinot, J.C.: [¡ i¿ mü¡/i¿] mad, the adhesion based dynamic micro-manipulator. European Journal of Mechanics-A/Solids **22**(6), 903–916 (2003)

8. Haliyo, S., Dionnet, F., Régnier, S.: Controlled rolling of microobjects for autonomous manipulation. Journal of Micromechatronics **3**(2), 75–102 (2004)
9. Millet, G., Lécuyer, A., Burkhardt, J.M., Haliyo, S., Régnier, S.: Improving perception and understanding of nanoscale phenomena using haptics and visual analogy. In: Haptics: Perception, Devices and Scenarios, pp. 847–856. Springer (2008)
10. Millet, G., Lécuyer, A., Burkhardt, J.M., Haliyo, S., Régnier, S.: Haptics and graphic analogies for the understanding of atomic force microscopy. International Journal of Human-Computer Studies **71**(5), 608–626 (2013)
11. Nagasaki, H.: Asymmetric velocity and acceleration profiles of human arm movements. Experimental Brain Research **74**(2), 319–326 (1989)
12. Norman, D.A.: Natural user interfaces are not natural. Interactions **17**(3), 6–10 (2010). DOI 10.1145/1744161.1744163
13. Oztop, E., Wolpert, D., Kawato, M.: Mental state inference using visual control parameters. Cognitive Brain Research **22**(2), 129–151 (2005)
14. Ravasio, P., Tscherter, V.: Users theories of the desktop metaphor, or why we should seek metaphor-free interfaces. Beyond the desktop metaphor: designing integrated digital work environments pp. 265–294 (2007)
15. Régnier, S., Chaillet, N.: Microrobotics for Micromanipulation. Wiley-ISTE (2010)
16. Ren, G., O'Neill, E.: 3d selection with freehand gesture. Computers & Graphics **37**(3), 101–120 (2013)
17. Sartori, L., Becchio, C., Castiello, U.: Cues to intention: the role of movement information. Cognition **119**(2), 242–252 (2011)
18. Sauvet, B., Ouarti, N., Haliyo, S., Régnier, S.: Virtual reality backend for operator controlled nanomanipulation. In: IEEE International Conference on Manipulation, Manufacturing and Measurement on the Nanoscale (3M-NANO), pp. 121–127 (2012)
19. Searle, J.R.: Intentionality: An essay in the philosophy of mind. Cambridge University Press (1983)
20. Stapel, J.C., Hunnius, S., Bekkering, H.: Online prediction of others actions: the contribution of the target object, action context and movement kinematics. Psychological research **76**(4), 434–445 (2012)
21. Viviani, P., Flash, T.: Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. Journal of Experimental Psychology: Human Perception and Performance **21**(1), 32 (1995)