Within-Object User Intention Prediction Model in Virtual Reality

Elodie Bouzbib* Gilles Bailly†

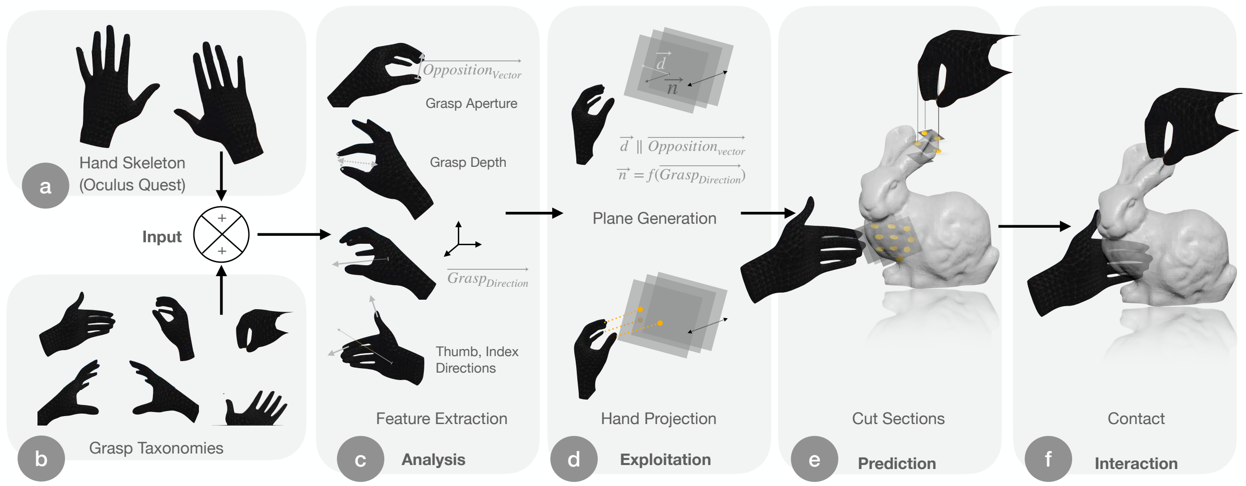

Fig. 1. We present a novel Within-Object Intention Prediction model for Bare-hands interactions in VR (eg manipulation tasks). Based on the (a) users' skeleton, extracted from the Oculus Quest, and (b) Grasp taxonomies, (c) we analyze and extract 4 key geometrical features to capture the users' grasp behaviour. (d) We exploit these features to generate planes, on which the users' skeleton is projected. (e) These planes act as Cut Sections over a virtual object of interest, and predict the users' *future contact locations*, prior to performing an interaction. (f) The user interacts with the virtual object at the predicted positions.

Abstract— We present a novel computational model to favour bare-hands interactions with haptic technologies in Virtual Reality environments. Using grasp taxonomies defined in the literature, we broke down users' gestures into four key geometrical features and developed a model which dynamically predicts the users future within-object grasp intentions locations. The model supports a wide range of grasps including precision and power grasps, pulling or pushing as well as two-handed interactions. Moreover, its implementation does not require calibration (no parameter, user-independent) nor specific devices such as eye-trackers. We evaluate the model in a user study involving various shapes, sizes and gestures. The results show our model provides a great accuracy for predicting the future interaction locations (below 30mm) more than one second prior to interaction. Finally, we propose use-cases for our model - using redirection techniques or encountered-type of haptic devices.

Index Terms—Virtual Reality, Haptics, Grasp, Bare-Hands Interactions, Intentions, Grasp Location Position, Computational Model

1 Introduction

In the real world, we perform various and complex grasps and manipulations with one or two hands to interact with objects in our environments [8]. These interactions are generally performed bare-hands. For instance, we can manipulate a teapot through its base or its handle (eg *displacing it*) depending on the task. It can also be held so its top can be taken off (eg *opening it*). *Bare-hands manipulations* allow rich interactions as humans precisely feel tactile and kinesthetic properties of the real objects.

While researchers in Virtual Reality (VR) often refer to VR interactions as "being natural", we observe that most of the systems do not support bare-hand interactions and the real-world variety of manipulations. Typically, a controller enables the acquisition of a target by pushing on a button, but does not enable the users to explore or manipulate an object with various hand configurations.

The goal of this technology paper is to favour the elaboration of haptic solutions in VR supporting rich bare-hand interactions. We believe that the future of Haptics in VR relies on the users freedom to manipulate - with their bare-hands - any object of interest with no regards to the scenario's progress, referred to as *non-deterministic scenarios* [12, 13]. In this direction, recent technologies such as redirection techniques [48] or robotic graphics [53] let the users unencumbered [70] - free of any contraption (wearable, controller, exoskeleton) - while they interact with physical props. While these technologies are promising, providing the users with the adequate physical prop actually requires their intentions to be predicted. Anticipating these intentions prior to interaction is crucial for the interfaces to properly overlay physical props over virtual objects in real-time [57], and for the adequate haptic (tactile and kinesthetic) feedback to be provided over each hand individual parts (thumb, fingers, palm).

We present a novel user-intention model to improve unencumbered bare-hands interactions with haptic feedback in VR. The input of the model is the users' real hands *(inputs, Figure 1-a,b)*, captured by an Oculus Quest; the model provides the users future contact locations *(outputs, Figure 1-e)*. In contrast with previous user-intention models [11, 12, 20], we do not predict *which* object the user will interact with (*between-objects*), but *where* (*within-object*). This is essential as there are multiple ways to grasp and manipulate a single object depending

*Elodie Bouzbib is with Centre Inria de l'Université de Rennes, Rennes, France.
e-mail: elodie.bouzbib@inria.fr
†Gilles Bailly is with CNRS, ISIR, Sorbonne Université, Paris, France.
e-mail: gilles.bailly@sorbonne-universite.fr

on the task [8]. Our model thus provides predictions with a finer level of granularity.

An originality of our model is to take root from grasp taxonomies [23, 29, 31, 35] and bare-hand interactions from the real world. These taxonomies depict and describe many hand/fingers configurations for manipulating objects and provide a great understanding of the human behaviour. From our analysis, we extracted four geometrical key features that contain the necessary information to predict the future user grasp contact points.

Our user intention model has several properties. It supports a wide range of grasps, including precision and power grasps, pushing or pulling as well as two-handed (bi-manual) grasps. It is easy to implement and fast (works in real-time) as it relies on simple geometrical features and does not require calibration (no parameter, user-independent) nor specific devices such as eye-trackers.

We quantify our model capabilities in a user study involving 135 different grasping configurations (various objects, sizes, performed manipulations). Our results show our model provides a great accuracy (below 3cm) more than one second prior the interaction. The user study also confirmed the benefits of bare-hands interactions in VR as well as the variety of grasps spontaneously performed when interacting naturally. These results should encourage the design of novel haptic technologies (e.g. using robotic graphics, redirection techniques) to support unencumbered bare-hands interactions in VR. The main contributions of this work are:

- A model predicting the user *within-object* grasp intention locations for bare-hands interactions and manipulations in VR and its evaluation.
- Foundations for the deployment of VR bare-hands experiences with haptic feedback in non-deterministic scenarios.

2 Motivations & Use-Cases

Our model is motivated by the deployment of unencumbered haptic technologies, supporting bare-hands experiences in VR. We foresee use-cases with two main class of systems in these regards:

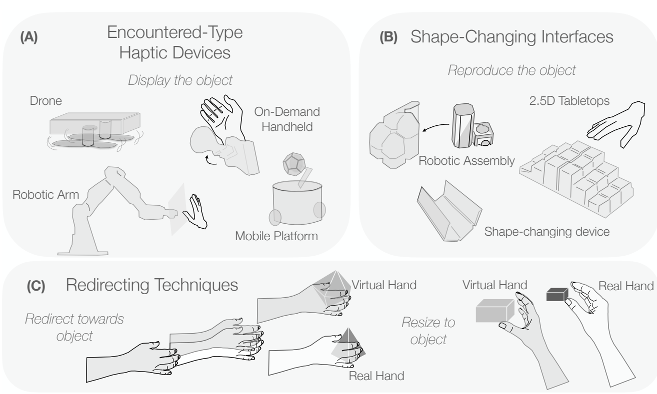

Fig. 2. Different classes of technologies that benefit from our model to predict future contact locations within objects: Robotic Graphics interface can (A) *Display* the chosen object (with Encountered-Type of Haptic Devices, such as drones [3]; mobile platform [36]; on-demand handheld [26]; robotic arm [47]) or (B) *Reproduce* the chosen object (Shape-Changing interfaces; robotic assembly [78]; 2.5D tabletop [32]). (C) Redirection techniques can also be applied to redirect the user hand towards the correct object [48] or can be exploited to resize the grasp [10].

Redirection techniques. Redirection techniques such as *"Haptic Retargeting"* [6] (Figure 2 - C) modify the users' virtual hand trajectory and/or grasp size (e.g. [10]) so that the physical hand (or fingers) meets the physical prop when the virtual hand meets the corresponding virtual object. In deterministic scenarios, i.e. scenarios where the users actions are dictated by the scenario, the degree of redirection can be set offline. However, in more realistic scenarios (non-deterministic scenarios), the users are free to grasp and manipulate any objects with no regards to the scenario's progress. In these scenarios, it is thus necessary to predict how the user will interact with, to adapt the degree of redirection accordingly. For instance, late predictions will introduce a larger amplitude of redirection, which is more likely to break the illusion and therefore, the immersion.

Robotic Graphics. Robotic Graphics [53] are interfaces encountering the users locally, with their "desired object size, shape". Instead of modifying the location or size/shape of the virtual hand such as in the hand redirection, the approach, here, consists of either using a robot to physically change the location of the physical prop (Figure 2 - A), or to literally reconfigure itself to reproduce the corresponding object shape (Figure 2 - B). Similarly to hand redirection, the robot should know in advance which part of the object and how the user with interact with to adapt (move, reconfigure) accordingly.

The aim of this paper is to provide foundations for the deployment of VR bare-hands interactions with haptic feedback in VR, through a replicable and usable user-intention model. We provide the future contact *location* information prior to interaction, and can extract the *haptic properties* (e.g. shapes) even when interacting with a complex shaped object such as a teapot (Figure 3 - A). For instance, Figure 3 - B illustrates the benefits of our model where a robotic arm anticipates the future user grasp to move the physical prop accordingly: it therefore overlays the adequate part of its associated virtual object.

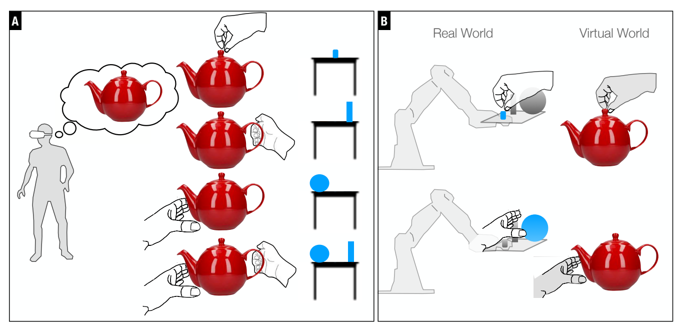

Fig. 3. A. A user wants to interact with a teapot: a complex shaped object - primitives are different *within* the object. Depending on his manipulation, the user will not interact with the same primitive; it can be single or two-handed, from the top or from the sides. B. As an example of our model use-cases, a robotic arm displays the correct object primitive to represent a teapot in the virtual environment.

3 Related Work

3.1 Bare-Hands Interactions in Virtual Reality

Bare-hands interactions in Virtual reality have been subject to various implementations, mostly for target acquisition tasks and selecting the 3D virtual objects (eg Go-Go, World-in-Miniature) [5, 9, 55, 58, 71]. They however do not depict the complexity of the human grasp nor offer to exploit the objects' affordances. Consequently, *direct manipulation* [16], "the ability for a user to control objects in a virtual environment in a direct and natural way, much as objects are manipulated in the real world", is usually unavailable.

Efforts start to be made in this direction. For example, Oculus [2] is currently proposing avatar hands replicating the users' ones, but only two gestures are available for interacting, with basic *raycasting* and *pinching* techniques. Despite the use of their real hands, users still need to be trained to be able to interact with their environment and cannot grasp with multiple fingers [39] such as the grasp taxonomies would suggest it. One main reason is probably the lack of haptic feedback which is needed for natural grasping movements [63], in particular in VR as distance and depth perceptions are altered [15, 52].

Fig. 4. Illustrations of some definitions from the literature [31]: Virtual Fingers - VFs - are an abstract representation of fingers applying forces in the same direction and working as a unit. (a) Precision Grasp with Pad Opposition: the hand surfaces are parallel to the palm direction (dotted arrow). (b) Power Grasp: there is a rigid relationship between the object and the hand. The grasp is performed with a Palm Opposition: the hand surfaces are perpendicular to the palm (dotted arrow). In both these configurations, the thumb is abducted, e.g. it is opposing the fingertips. (c) Non-Prehensile Grasp: the whole hand works as a unit, with a single Virtual Finger. (d) The hand is shaped as a ”Hook”. The thumb is adducted: its direction follows the palm one.

3.2 Bare-Hands Interactions with Haptics in Virtual Reality

Haptic technologies for Virtual reality applications have increased drastically in the last few years, as a substitution for the common controller, which forces the user to perform a single predefined grasp type (holding the controller).

Modular hands-free controllers have been proposed to widen the grasp types [73]. For instance, [66] proposes a controller with a ’pop-up’ prop simulating three different shapes primitives (sphere, cube, cylinder) within the users’ palm, while [77] renders them using a pin-based array. In the same regards, wearable technologies or exoskeletons also provide haptic feedback for manipulation purposes, by constraining the users’ fingers according to objects’ boundaries [1, 4, 18, 19, 30]. These interfaces (exoskeletons, wearables, controllers) aim to let the user feels what he sees [74], yet discrepancies between the expected and provided haptic feedback remain. Their use is limited compared to the real world interaction opportunities and they are opposed to the postulate that users should be unencumbered in artificial realities [70].

On the opposite, direct manipulation [16] and passive props [42] are shown to have the potential to haptically enhance the virtual environments. Physical 3D objects theoretically enable the users to perform any grasp type from the literature, and even to perform bi-manual interactions. In this direction, Hinckley et al. demonstrated that using 3D objects and directly manipulating with both hands is an advantage for learning skills in virtual environments [38].

Enabling the use of real objects in Virtual Reality is therefore promising, however it requires strategies to map the physical environment. This mapping can be done ”offline” [57], for instance by having the physical and virtual worlds perfectly matched [60], or in ”real-time”. The real-time strategies refer to our aforementioned use-cases (Section 2): Robotic Graphics interfaces or Redirection techniques. In order to provide the users with their correct object/primitive, we hence are required to predict their interactions intentions [12, 17, 34].

This paper aims at facilitating the extraction of the users’ grasp intentions to leverage the design of **bare-hands interactions** in Virtual Reality, with no contraption, wearable nor controller of any type (*unencumbered* [70]). We believe anticipating the users behaviour is the key to providing them with the adequate physical objects/primitives.

3.3 Exploiting User-Intentions

User intentions are generally used in fields such as teleoperation [21,62], assisted grasping [46,49] or shared interactions with robotic devices [25, 27, 28, 61, 68, 72, 75, 76]. For instance, predicting the next interactions enables the tele-operated device to react faster and to reduce the latency during interactions. It thus enables a smoother experience, with less jerk, saccades or latency from the devices.

In Virtual Reality, user intentions have previously been used to move users within the virtual scene [51, 65] or to provide them with haptic feedback, anticipating the contact timestamp [14]. User behaviours are usually anticipated for path-planning [74–76]: knowing the future contact point location, users and/or interfaces trajectories can be planned.

Once an algorithm predicts which object of interest is about to be interacted with, an interface can then provide an exact physical mapping of the virtual object [12] or propose to explore an adequate shape primitive [34]. In these regards, *Sparse Haptics* [17] coupled a tangible interface with multiple shape primitives with an eye-hand coordination model for an exploration/target acquisition task [11]. It predicts the users intentions in order to redirect them to the correct object primitive.

Direct manipulation with bare-hands is however a trickier task. First, the proposed *exploration* tasks in [17, 34] involve a *single finger* in a physical 2-Dimensional world. A *manipulation* task would require *multiple fingers* contact points, in a 3-Dimensional physical world, with potentially multiple various shape primitives or depths.

While the few available models for haptics in VR determine which object of interest is about to be interacted with (*between-objects*), a model for a manipulation task is required to predict a much finer area of interest, *within-object* of interest. Compared to Machine Learning models [67], we propose in this work an computational and analytical model based on the understanding of the human grasp.

4 Understanding Grasp

Feix et al. [31] defined a grasp as ”every static hand posture with which an object can be held securely, irrespective of the hand orientation”. A manipulation task is the modification of an object position or orientation, and hence requires a grasp to be performed beforehand.

Grasp is task and object dependent [8, 56] - a bottle will be grasped differently whether a user want to drink it or to transmit it to someone else - yet, it still can be discriminated based on hand configurations.

Many grasp taxonomies have been drawn [23, 35, 44, 56], using object types or sizes. A systematic review of grasp taxonomies was depicted by Feix et al. in 2016, as the ”GRASP Taxonomy”, resulting in 33 coherent human hand configurations, according to 4 properties: (1) Virtual fingers, (2) Grasp types, (3) Opposition space, (4) Thumb position. We will define these properties as a basis to develop a human-centered model for grasp.

4.1 Virtual Fingers

Iberall defined the *Virtual Finger* as an abstract representation of a combination of fingers applying forces in the same direction, and working as a unit [41]. For instance, the Index and Middle finger of the Figure

4-a or the four long fingers of the Figure 4-c constitute a single virtual finger as they apply forces in the same direction.

Grasps are by essence "prehensile", they provide the ability to hold things, especially by curling around them. All of the Prehensile grasps count at least two virtual fingers, as fingers do need to be constraining an object from two directions to perform a "clamping" mechanism and enable its manipulation. However, Cutkosky defined a non-prehensile grasp, formed from a single Virtual finger. This grasp hence involves the whole hand, as a unit, and can be used to perform the translation of an object for instance, while pushing or pulling on it [24] (Figure 4-c-d).

The virtual fingers are a key in the understanding of human grasp [7, 33, 40], as they combine both the hand biomechanics and the interactions hand/objects.

4.2 Grasp Types

Two main types of grasp are currently depicted in the literature.

- *Precision Grasp (Figure 4 - a)* In a precision grasp, "the hand is able to perform intrinsic movements" [23, 31, 35, 44]. This means that the manipulation of the object relies on a few phalanges and the fingers are able to displace an object without involving the arm or wrist displacement. This type of grasp is usually performed through the fingertips, and with decreased object sizes [23].

- *Power Grasp - (Figure 4 - b)* On the opposite, a power grasp is qualified by "a rigid relationship between the object and the hand" [23, 31, 35, 44]. This means that in order to manipulate the object and modify its position/orientation, the entire hand is involved. Gestures then result from the wrist or arm displacements. This type of grasp usually involves the palm, and/or multiple phalanges from a finger, and increased object sizes [23].

4.3 Opposition Space

The opposition space corresponds to the direction applied between the hand and the object. There are three types of opposition: pad, palm, and side. Pad opposition corresponds to a grasp "where hand surfaces are parallel to the palm" (Figure 4 -a [31], the arrow between the index and thumb is parallel to the dotted arrow representing the palm); Palm opposition's direction is generally perpendicular to the palm (Figure 4 - b); Side opposition is in a direction transverse to the palm.

4.4 Thumb Position

The last property is the thumb position. It can either be adducted, following the palm direction (see Figure 4-c-d), or abducted, able to oppose the fingertips (see Figure 4-a-b). An adducted position will for instance allow for a "hook" posture [44] (Figure 4 - d). All of the pad opposition grasps require the thumb to be abducted, to be able to perform the "pinching" of the objects when manipulating them.

Taxonomies then exploit the Virtual Fingers numbers, opposition types and thumb positions to determine the type of grasp a user is performing [31].

4.5 Current uses of Taxonomies

Understanding human prehension and grasp behaviours is currently being used in multiple fields: avatar designs, dexterous robots, prosthetic hands. Many datasets have extracted human grasps to feed avatar simulations [22, 64], so they would replicate real human behaviours; or robot control [45, 54], for instance for teleoperation purposes. They help to choose how to efficiently manipulate objects, based on their topology. Indeed, an object can communicate multiple ways to be manipulated through its affordances (its geometry) and can be manipulated differently depending on the user's intended task. This can be illustrated with the teapot, whether a user wants to fill it with water, serve a cup of tea, or put it back in the cupboard. Objects with simpler geometries can also be used in this example: an apple will be held differently whether it is to be eaten or given to someone else [8, 64].

5 Feature Extraction

From our previous "Understanding Grasp" section (Section 4), we define 4 geometrical key features from the hand.

We define P_T, P_I, P_P, as the respective 3D positions of the thumb tip, index tip and palm center.

5.1 Feature 1: Opposition Vector - (Figure 5 - a)

The index and thumb often are to be considered in a grasp: they usually form two of the "Virtual Fingers" composing a grasp; their orientation inform us of the "Opposition space". It was demonstrated that the formation of the finger grip occurs during the hand transportation in natural prehension movements [43].

Consequently, we define a vector, called the **"Opposition Vector"**, between these two finger pads (L is its norm).

$$\overrightarrow{Opp} = \overrightarrow{P_T P_I}$$

$$L = ||\overrightarrow{Opp}||$$

5.2 Feature 2: Thumb, Index and Palm directions (Figure 5 - b)

The angle from the thumb and palm directions can suggest some types of grasps and manipulations (for instance, pushing or pulling with a whole hand). For instance, if the palm and thumb direction are similar, but opposed to the index direction, we can suggest that a "hook" grasp is likely to be performed (Figure 4 - d).

Similarly, if the global hand only possesses a single Virtual Finger (same shared directions for these three hand parts), this suggests that either the user will push/pull the object, or that both hands are going to be used for manipulation, so the object is securely held.

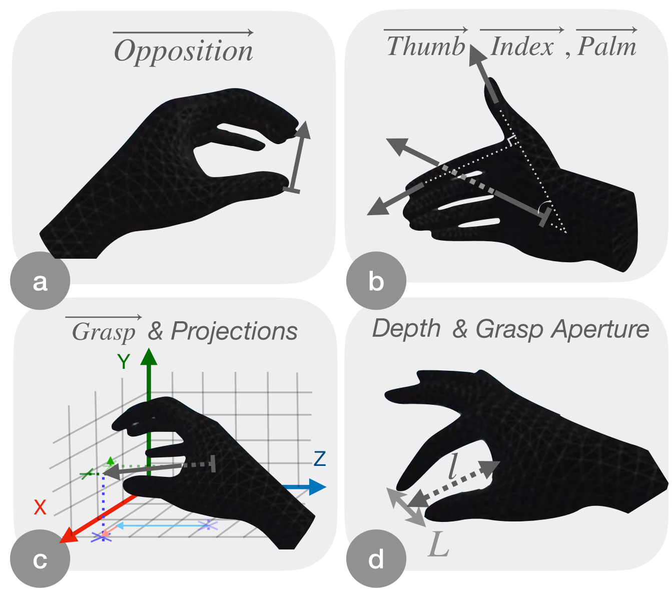

Fig. 5. Features extracted from users' hands: (a) The "opposition vector": It links the thumb pad to the index one. (b) The thumb, index, and palm directions: we here can note the palm and thumb are following the same directions, with a small angle separating them, while the thumb and index directions are perpendicular. (c) We extract the palm orientation and project it over the coordinate system to define the Grasp direction. Here, the upward component is significantly smaller than the other ones: the grasp will be performed from the sides. (d) The "depth" and "grasp aperture" distances. We depict the grasp aperture L, being the "opposition vector" norm, and the grasp depth l, being the distance between the grasp aperture midpoint and the palm center.

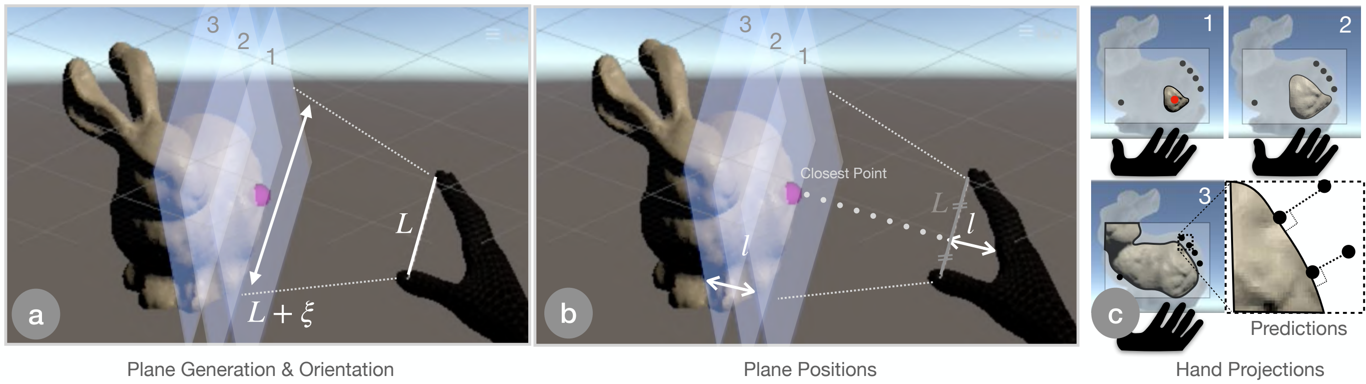

Fig. 6. (a) Plane Generation: The Cut sections (1,2,3) lengths are an extension of the Grasp Aperture L. They are parallel to it. (b) Plane Positions: They are located from the Grasp aperture mid-point's closest point over the object of interest (pink sphere) and spread over a Grasp Depth length l. (c) Hand Projections: Each plane cuts the object of interest. The hand phalanges are projected onto each of them (1, 2, 3). The spheres represent the hand projections: they are the hand phalanges projections on the cut sections. *Predictions*: The hand projections are projected onto the intersection between the Object of interest and the Cut Sections.

5.3 Feature 3: Grasp Direction (Figure 5 - c)

We also need a feature to define whether a grasp is performed from the top or the side of an object. This decision is often made very soon in the gesture leading to the grasp [43]. We define the grasp direction as the palm center's trajectory[1]:

$$\overrightarrow{Grasp} = \overrightarrow{P_{P,t-1}P_{P,t}}$$

5.4 Feature 4: Grasp Depth (Figure 5 - d)

Geometrically, a grasp always occur within the volume defined between the Opposition vector and the palm. Even if a hand were to be non-folded (e.g. non-prehensile grasp), this theory remains valid as the contact would be colinear to the Opposition vector.

Also, as the grasp aperture and finger grip are formed during the movement [43], we deducted that the object "in-depth" contact point should be extracted at an early stage of the grasp. We thus introduce the grasp depth $\overrightarrow{l}$:

$$\overrightarrow{l} = \overrightarrow{P_p P_m}$$

where P_m is the midpoint of the grasp aperture:

$$P_m = \frac{P_T + P_I}{2}$$

6 Model Implementation

We now present our computational model, which predicts the future contact points within the virtual object of interest. It relies on the four mathematical features elaborated from grasp taxonomies. The general approach is summarized in the illustrated pipeline **Figure 1**.

6.1 General approach

Our previous features are hand geometry-based. We hence decided to explore this perspective more thoroughly, and to create a geometric tool to leverage their use. We decided to create planes, acting as **cut sections** over the objects of interest.

6.2 Inputs

The inputs are 3D representations (positions and orientations) of the hand palm, index, thumb; and the object of interest *OOI*. We use the Oculus Quest hands to extract the users whole hands.

6.3 Plane Generation

We first define C_{OOI} as the Closest point from the Object of interest (OOI) to the grasp aperture midpoint P_m (see Section 5.4).

[1] t represents the time

6.3.1 Orienting the Planes

To define the cut sections global orientation, we extract the palm direction. We project it over the XYZ coordinate system, and find its greatest component (over X, Y, or Z) (see Figure 5 - c, Figure 7) (see Algorithm 1).

This defines whether the plane generation should be *horizontal* or *vertical*. This gives us our first plane directing vector.

The cut sections are then rotated to be **colinear** to the Opposition vector (see Figure 7), which is the plane second directing vector (see Figure 5 - a). The plane normal $\overrightarrow{n}$ is therefore defined by the cross product of these two directing vectors.

When the palm information is not available to extract, for instance if only two trackers (on the thumb and index) are being used in a VR environment, we can trade the palm orientation for the following:

$$\overrightarrow{Palm} \approx \overrightarrow{P_T P_I} \wedge \overrightarrow{\theta_{z,I}}$$

where $\overrightarrow{\theta_{z,I}}$ is the index' local $\overrightarrow{Z}$ vector (see Figure 7 - c).

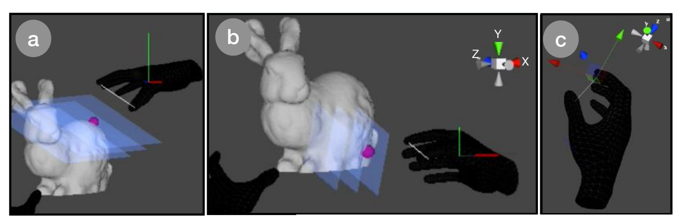

Fig. 7. The same Opposition Vector is involved in these two gestures. The Palm direction is here projected over the XYZ coordinate system, and their respective component are displayed according to Unity coloring system (X = red, Y = green, Z = blue). The planes are parallel to the Opposition Vector (displayed in white). (a) The Y component is greater than the other ones, the grasp will be performed from the top. (b) The X component is greater than the other two: the grasp will be performed from the X side. The planes are cutting the object along this direction. (c) The index local referential and the opposition vector (white).

6.3.2 Positioning the Planes

The first plane's origin is located at C_{OOI} (pink sphere in Figure 6). We spread two other planes over the Grasp Depth length l (total number of planes $k=3$). C_{OOI} is defined to be dependent on the Object of interest's closest point from P_m. To keep the "depth" property even

Algorithm 1 Orienting the Cut Sections; The inputs are the palm direction $\overrightarrow{Palm}$ and the opposition vector $\overrightarrow{Opp}$. We define the cut sections normal vector $\overrightarrow{n}$.

procedure ORIENTATION PLANE ($\overrightarrow{Palm}$,$\overrightarrow{Opp}$) ▷ Defining two non colinear directing vectors and the plane normal $\overrightarrow{n}$.
  if $proj_{\mathbf{y}}\overrightarrow{Palm} > proj_{\mathbf{x,z}}\overrightarrow{Palm}$ **then**
    $\overrightarrow{d_1} \leftarrow (0, 0, 1)$ ▷ Plane is Horizontal
  else
    $\overrightarrow{d_1} \leftarrow (0, 1, 0)$ ▷ Plane is Vertical
  end if
  $\overrightarrow{n} \leftarrow \overrightarrow{d_1} \wedge \overrightarrow{Opp}$ ▷ Plane normal
end procedure

when the user's hands are getting closer to the object, we inverse the direction of the spread in this configuration (see Algorithm 2).
We define the plane i position $P_{plane,i}$ in Algorithm 2. The cut sections lengths (in their virtual representation) are an extension of the grasp aperture.

Algorithm 2 Positioning the Cut Section i among the k ones. The inputs are: P_m, midpoint of the grasp aperture; $\overrightarrow{l}$, grasp depth (norm = l); C_{OOI} the closest point from the Object of interest to P_m.

procedure POSITION PLANE i, ($\overrightarrow{l}$, C_{OOI}, k, P_m)
  if $P_m - C_{OOI} > l$ **then**
    ▷ Predict within the object's depth
    $P_{plane,i} \leftarrow C_{OOI} - \overrightarrow{l} * \frac{i}{(k-1)}$
  else ▷ Keep the Depth Predictions when Grasping
    $P_{plane,i} \leftarrow C_{OOI} + \overrightarrow{l} * \frac{i}{(k-1)}$
  end if
  return $P_{plane,i}$ ▷ Position of the Plane #i
end procedure

6.4 Projection of the Hand onto the Cut Sections

Once the planes are oriented and positioned correctly, we project the hand phalanges onto each of them (see Figure 6 - c). We define this projection as $Proj_{k,phalanx}$.

We also project them at the intersection between the object of interest and the cut sections $Proj_{OOI,k,phalanx}$.

6.5 Output: Predictions of Future Contact Points

We compare the distances from each phalanx projection among the three planes (see Figure 6 - c). For each cut section k, we compare the distances between $Proj_{k,phalanx}$ and $Proj_{OOI,k,phalanx}$. The final phalanx prediction is $Proj_{OOI,k,phalanx}$ with the smallest distance. This projection gives the position of the future contact points. We can also extract its local shape, and its different substance-related properties [50].

7 MODEL PROPERTIES

This model shows multiple benefits. First, it is user-independent. Indeed, the hand geometry and its dynamism while grasping does not change among users. This is a direct consequence of using taxonomies: it is adapted to any grasp and any user. Second, it can be used as a real-time model for predicting grasp intentions locations in VR. Third, it is adapted for both hands, and for bi-manual interactions. Fourth, as the contact point is predicted, the zone of interest shape can also be deducted quite easily by extracting the future contact point primitive.

This provides the haptic solution with a physical future contact location *and* the associated local haptic properties of the virtual object.

We will evaluate our model accuracy prior to contact in the next section. Note that the model does not predict the number of future contact points: **if a contact is to occur, the model predicts its future absolute position**. The model does not discriminate the grasp type to analyse the number of future contact points.

8 EVALUATION

We conducted a user study to test the capacity of our model to accurately predict the user's grasp positions when performing different manipulations (e.g. hold, push, pull) on various objects (e.g. cube, cylinder). More precisely, we estimate the distance between the actual and predicted touch contacts at different times prior to the interactions.

8.1 Participants and Apparatus

8.1.1 Participants

We recruited 7 participants from our acquaintances and laboratory (3 female), aged from 25 to 37 (mean = 30, std = 4). Five users were familiar with VR technologies, 2 of them were experiencing VR for the first time, and none of the users had ever used a head-mounted display (HMD) without controllers. No rewards were attributed to the participants. All the users were right-handed.

8.1.2 Apparatus

The participants wore an Oculus Quest HMD without any supplementary sensors, controllers nor wearable devices. They adjusted their HMD to their convenience, and remained in a standing position during the whole experiment.

The 3D scene was designed on Unity3D, and compiled as an Android application. The scene contained the avatar hands (e.g. Figure 8) available on the Oculus Quest as well as the virtual object to manipulate (white) and its target location (red) as shown on Figure 9. Walls were surrounding the scene, and users were standing in front of a virtual table. The virtual objects were not subject to Unity3D's physics engine (gravity) and were attached to the hands[2] when a collision occurred, to move accordingly with them.

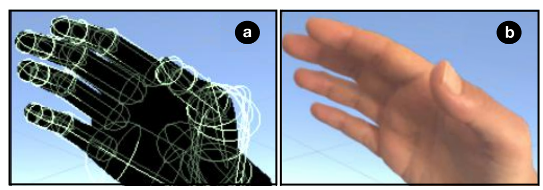

Fig. 8. Hand Representations: (a) Virtual hand, with associated Phalanges (in green); (b) Real hand during the game: the user does not wear any tracker nor holds a controller.

8.2 Method

8.2.1 Conditions

We controlled two factors related to the objects (SIZE and SHAPE) as well as one factor related to the task MANIPULATION, as the nature of the grasp is object/task-dependent.

The three object SIZES were 5cm, 10cm and 25cm. These sizes were chosen to be relatively small, medium or large compared to an average hand size ($\approx$ 18cm).

We considered nine object SHAPES. Two shapes were simple: Cube and Cylinder (Figure 9 - a,b). Seven shapes, corresponding to the seven first digits, were more complex $\{0, 1, 2, 3, 4, 5, 6\}$. The complex shape objects are combinations of simple primitives with different radii, or sharp/round angles. They offer more grasp opportunities. For instance, the digit 1 can be grasped by its rectangle base or its cylinder trunk or from its spherical top (see Figure 9).

Finally, we considered five MANIPULATIONS – "Hold, Pull, Push, Raise and Push down" – (Figure 9) involving different grasp types, opposition space and thumb position (see Section 4), and representing the common real world day-to-day manipulations.

[2] More specifically, the objects were attached to P_m to enable a smooth manipulation.

8.2.2 Task

The instruction was displayed on the virtual wall and indicated the manipulation to perform (e.g. hold, push) to place the white object into the red target location (see Figure 9). Once the object was positioned in the target location, this latter became green.

As we are interested in the grasp intentions *prior* to contact, the virtual objects disappeared 3 seconds once the grasp was performed (ie when a contact between the object and any hand phalanx was maintained) - even if the user was not finished moving it to accomplish the given MANIPULATION.

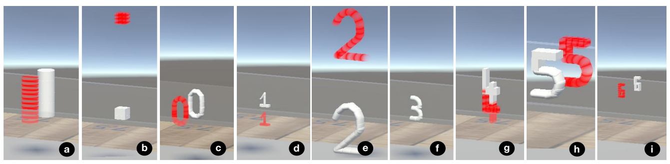

Fig. 9. The task consists of placing the white virtual object with various shape and size into the red phantom target location with a given manipulation (e.g. hold, push, pull): (a) Medium Cylinder to be pulled; (b) Small Cube to be raised; (c) Medium 0 to be pulled; (d) Small 1 to be pushed down; (e) Large 2 to be raised; (f) Small 3 to be simply touched; (g) Medium 4 to be pushed down; (h) Large 5 to be pushed; (i) Small 6 to be pulled.

8.2.3 Procedure

Participants were first asked to fill in a consent form, validated by the university Ethics Committee (CER 2020-60). They were informed about the experience, its goal and duration.

During the experiment, they were also asked to interact as naturally as possible. In particular, they were free to use one or two hands and the desired number of fingers. The participants HMD view was cast over the experimenter's phone, for her to overview the experiment's progress and ensure the task was well understood. The duration of the experiment was approximately 20 minutes per participant (mean = 21, std = 1.2mn). No participant felt any discomfort during the experiment.

8.2.4 Design

We used a within-participant design. Each participant tested the 135 conditions, corresponding to 5 MANIPULATION $\times$ 9 SHAPE $\times$ 3 SIZE in a random order. The total number of trials is 7 participants $\times$ 135 conditions = 945 trials.

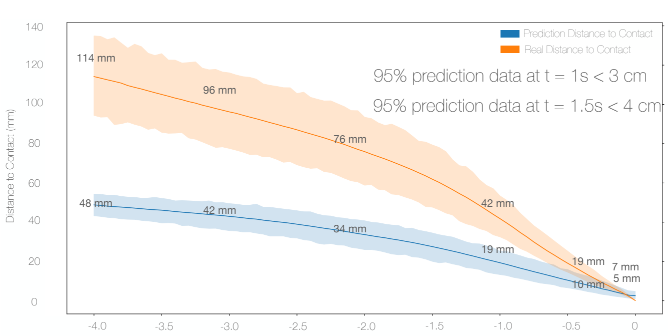

Fig. 10. Right Index and Thumb Prediction and Real Phalanges distances to the contact point (in mm). The predictions remain in the contact point vicinity. Intervals show 95-CI. Real phalanges distances and predictions distances intersect only 200ms prior to contact.

8.2.5 Measures

We collected each of Oculus avatar hands' phalanges name, position, orientation and created their associated colliders[3] (see Figure 8). The position and orientation of the Oculus Quest HMD was also recorded. We recorded each configuration number, and each objects' of interest position and orientation. We collected all the data at 60 fps.

8.3 Results

Reach-to-Grasp Duration. When a reach-to-grasp duration was above 4s, either the tracking was lost and the user had to wait to get their virtual hands back, or the users were exploring the environment. In the following results, we hence summarize the data over a 4s scale and truncate above it. More than 80% of the grasps were performed under 4s (mean = 3.2s, std = 1.6s).

Analysis Procedure. When a contact occurred between the virtual hands and the virtual objects, we recorded it as the contact point, at t_0. We then analysed the distance between this contact point and (a) the mean prediction position and (b) the real phalanx position, from $t = -4s$ up to t_0, over all users and configurations. We first cleaned our data from the lost tracking, and verified that the grasps were maintained for at least 5 frames. This resulted in a total of 934 grasps to analyse.

mean (std)	Predictions				Real Positions			
in mm	All Hand Phalanges	R-Index	R-Thumb	R-Index + R-Thumb	All Hands Phalanges	R-Index	R-Thumb	R-Index + R-Thumb
t = 0	3.7 (2.0)	2.6 (3.2)	2.6 (2.6)	2.6 (2.8)	0 (0)	0 (0)	0 (0)	0 (0)
t = 0.25s	8.7 (3.8)	6.4 (5.3)	5.9 (5.1)	6.1 (5.1)	10.5 (2.9)	10.1 (4.9)	8.6 (3.8)	9.4 (4.3)
t = 0.5s	17.0 (10.3)	11.2 (7.4)	9.7 (6.5)	10.5 (6.8)	20.8 (12.0)	20.6 (8.8)	17.7 (7.8)	19.1 (8.2)
t = 1.0s	30.2 (15.9)	21.0 (9.4)	17.7 (9.0)	19.3 (9.0)	41.6 (21.3)	44.7 (17.6)	38.9 (14.9)	41.8 (15.8)
t = 2.0s	47.6 (18.2)	35.6 (10.5)	31.2 (10.9)	33.7 (10.5)	71.2 (27.5)	81.3 (28.1)	70.6 (24.5)	76.0 (25.7)
t = 3.0s	57.7 (17.8)	44.3 (11.6)	41.7 (10.4)	43.0 (10.2)	89.4 (28.6)	102.4 (31.7)	90.0 (25.9)	96.2 (28.3)
t = 4.0s	64.3 (16.0)	49.8 (10.8)	47.8 (7.8)	48.8 (8.2)	101.2 (26.0)	121.8 (36.8)	106.6 (27.0)	114.2 (31.3)

Table 1. Distances from the predictions to the contact point VS Real phalanx distances to the contact point. The index/thumb data belong to the users right hands (all users were right-handed). As a side note, the sample sizes can be different: the thumb and index pads are not necessarily involved in all the configurations.

8.3.1 Global Results

We ran simulations using the users data and analysed the distance between the contact points and our predictions prior to interaction. We define the accuracy of the model as the distance between the future contact point and the prediction. It is presented in the Table 1. Because all of our participants were right-handed, we also display the results for the users' right thumb/index pads.

The prediction slope is drastically lower than the real phalanges ones: the predictions are indeed always within the vicinity of the future contact points. To validate this, we can notice that the 95-confidence interval of the real positions only meets the Predictions one around 200 ms prior to contact.

We discuss ways to improve these results (e.g. increasing the model accuracy with a longer delay) in Section 9.

8.3.2 Analysis per Manipulation

The manipulation tasks also had an impact on our model accuracy. Indeed, the "Raise" prediction to contact distance is drastically higher than the other ones. As the users were placing their hands below the objects to raise them, we believe the hand projections could not reach the right contact points from the beginning. In a scenario where the users would be free to perform any gesture, we do believe that predicting and exploiting the "task" type could help improve the prediction time (see Section 9).

[3] We kept the initial properties of the colliders proposed by Oculus, however they often lost tracking and were not following the users' skeleton accurately. We associated our colliders to skeleton positions, which worked well and allowed various interactions.

8.3.3 Qualitative Feedback

Our experiment confirms that interacting with a simple avatar hand and bare-hands interactions is **fun** and **“natural”** [69]. P1 and P6 reported that it was fun to be allowed to perform any types of interactions with the objects. They took advantage of the freedom that was given to them to grasp objects in many various and unexpected ways (Figure 11). For instance, P1 interacted with the large cube (Figure 11- a) by pushing it with his fists. Similarly, P6 experienced holding objects through two of her fingertips (Figure 11 - c,d); and was satisfied to be able to literally move the objects with these unexpected types of grasps. All participants interacted with both of their hands to manipulate large objects, notably the big cube (Figure 11 - b). Five participants instinctively spoke up during the experiments, globally reporting that ”they knew it was silly to hold big objects with two hands, but they felt they were heavier and had to manipulate them with both of their hands and through their geometries”. Users had a tendency of considering lots of different grasps with medium sized objects, and changed their grasp configurations during the same grasp trial. We believe that unencumbered bare-hands interactions enable a **large variety of movements**.

The users behaviours suggest that our motivations and data collection are valid. Grasping in VR through objects’ affordances and in a ”natural” way is important for immersive experiences. We noticed a strong **correlation** between the **objects** types, the **manipulation** tasks, and the **users hand configurations** and grasps. We discuss how to take advantage of it to improve our model detection time in Section 9.

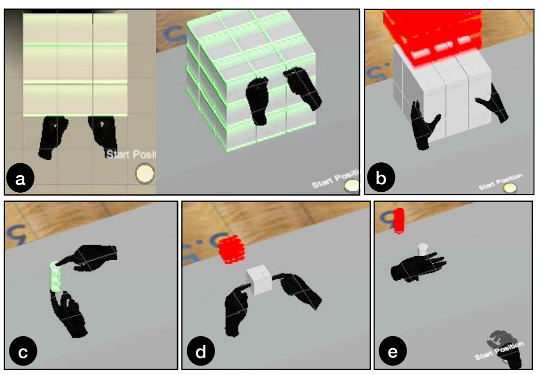

Fig. 11. Unexpected Grasps from our Data collection: (a) a user decided to push the cube using its fists; (b) All users used at least once both their hands to manipulate the big cube, here while raising it; (c) P6 locked the small cylinder position by holding it between two fingertips; (d) The same phenomenon happened with the small cube, being moved over with two single fingers; (e) A user pushed the small cylinder with an upside-down hand. All of these users reported they had fun moving the objects freely.

9 Discussion

In this section, we summarise our main findings, provide directions to extend the model and discuss another usage of the model (hardware specifications).

9.1 Main findings

The model we presented is user-independent and does not require any parameter. Qualitatively, we show that users appreciate the opportunity to perform any types of grasps, and take advantage of bare-hands interactions to manipulate virtual objects in a natural manner, much as with real-life interactions. Quantitatively, the results are promising as our model identifies the future contact points vicinity early during the reach-to-grasp phase, and provides a refined location (accuracy below $<$ 3cm) more than a second prior to interaction. It also shows to be working with no regards to the users’ grasp configurations, the manipulated objects or the manipulation tasks, even involving both of the users hands (two-handed interactions).

9.2 Extending the model

As future work, we see three main directions to extend our user-intention model.

9.2.1 Using Semantic Information

Using some semantic information about the object or the task can be used as *priors* to refine the predictions of the model both in terms of time and accuracy. Indeed, the environments constrain the available users interactions, therefore allowing the designers to add ”prior probabilities” over the different object areas. For instance, if we consider the teapot of Figure 3, three main areas of interest are more likely to be interacted with: its base, its handle and its top. Moreover, the base is more likely to be interacted with the non-dominant hand while the handle is more likely to be interacted with the dominant one. Similarly, if an object is placed over a desktop, the probability for its bottom to be touched would be null, as this area will be unavailable for direct contact. Prior probabilities can therefore be added accordingly to refine the predictions.

9.2.2 Using Grasp types

Another promising direction is to extend our model to return the grasp type and refine the predictions. The type of grasps can be determined from our (or additional) mathematical features or by using a Machine Learning algorithm. This information can potentially provide the future number of contact points, for a multi-fingered tangible interface to either display this exact number of physical contact points (such as in [59]) or for a redirection to be applied *per finger*.

9.2.3 Coupling Between- and Within-Objects Models

Finally, our model can be coupled with other models. In particular, it can easily be used with *between-objects* algorithms, in target acquisition tasks for instance [12, 17]. Interfaces will then able to determine *which* object of interest is about to be interacted with, and *where* within.

We can also build upon the identification of passive props using a computer vision algorithm (such as [37]) to provide the users with their local objects of interest.

9.3 Hardware Specifications

We currently used our user-intention model to predict the future contact points as soon as possible so that the given system can adapt accordingly in *real time*. However, we can also use our model the other way around. Indeed, our empirical results suggest that for a given precision (e.g. 3cm), a correct prediction can be made t seconds (respectively 1s) before the contacts occur. This information can be useful, *offline*, in the early stage of the design of the system to help the designers define the hardware specifications, ie. motor speed, maximum distance between/to the physical props, etc.

10 Conclusion

We presented a new model to favor the use of bare-hands interaction in VR. The model takes as input 3D representations of the hand and a virtual object and predict the future contact locations with an accuracy below 3cm about 1s before the contact. The model relies on the use of four key geometrical features that we defined and extracted from an analysis of grasp taxonomies and reach-to-grasp behaviours. The model is user-independent and supports two-handed interactions. Moreover, it can be coupled with *between-objects* user intention prediction models, such as the ones used for target acquisition [11, 17, 20, 34] or for VR non-deterministic scenarios [12].

Our user study also confirmed the importance and the variety of bare-hand interactions in VR. Indeed, participants performed ”natural” interactions in VR even when not requested to (e.g. holding the large-scale objects with both hands). These results should encourage the design of unencumbered haptic interaction techniques in VR, for instance using Robotic Graphics interfaces (Figure 2-A,B) or redirection techniques (Figure 2-C). Indeed, Haptics and VR designers can use our model to anticipate the users grasp intentions locations in real time to improve the control of the devices, e.g. refining the next location or configuration of the robot, the amplitude of redirection, etc.

11 ACKNOWLEDGEMENTS

Authors would like to thank all the user experience participants.

12 FUNDING INFORMATION

This work was supported by Sorbonne Université (ISIR, ISCD) and the CNRS.

13 DECLARATION OF INTEREST STATEMENT

The authors report there are no competing interests to declare.

REFERENCES

[1] CyberGrasp, 2019.

[2] Oculus Rift S, 2019.

[3] P. Abtahi, B. Landry, J. J. Yang, M. Pavone, S. Follmer, and J. A. Landay. Beyond The Force: Using Quadcopters to Appropriate Objects and the Environment for Haptics in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pp. 1–13. ACM Press, Glasgow, Scotland Uk, 2019. doi: 10.1145/3290605.3300589

[4] M. Achibet, A. Girard, A. Talvas, M. Marchal, and A. Lecuyer. Elastic-Arm: Human-scale passive haptic feedback for augmenting interaction and perception in virtual environments. In *2015 IEEE Virtual Reality (VR)*, pp. 63–68. IEEE, Arles, Camargue, Provence, France, Mar. 2015. doi: 10.1109/VR.2015.7223325

[5] M. Al-Kalbani, I. Williams, and M. Frutos-Pascual. Analysis of Medium Wrap Freehand Virtual Object Grasping in Exocentric Mixed Reality. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 84–93. IEEE, Merida, Yucatan, Mexico, Sept. 2016. doi: 10.1109/ISMAR.2016.14

[6] M. Azmandian, M. Hancock, H. Benko, E. Ofek, and A. D. Wilson. Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pp. 1968–1979. ACM Press, Santa Clara, California, USA, 2016. doi: 10.1145/2858036.2858226

[7] G. Baud-Bovy and J. F. Soechting. Two Virtual Fingers in the Control of the Tripod Grasp. *Journal of Neurophysiology*, 86(2):604–615, Aug. 2001. doi: 10.1152/jn.2001.86.2.604

[8] C. Becchio, V. Manera, L. Sartori, A. Cavallo, and U. Castiello. Grasping intentions: from thought experiments to empirical evidence. *Frontiers in Human Neuroscience*, 6, May 2012. doi: 10.3389/fnhum.2012.00117

[9] J. Bergström, T.-S. Dalsgaard, J. Alexander, and K. Hornb. How to Evaluate Object Selection and Manipulation in VR? Guidelines from 20 Years of Studies. p. 20, 2021.

[10] J. Bergström, A. Mottelson, and J. Knibbe. Resized Grasping in VR: Estimating Thresholds for Object Discrimination. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 1175–1183. ACM, New Orleans LA USA, Oct. 2019. doi: 10.1145/3332165.3347939

[11] G. Binsted, R. Chua, W. Helsen, and D. Elliott. Eye–hand coordination in goal-directed aiming. *Human Movement Science*, 20(4-5):563–585, Nov. 2001. doi: 10.1016/S0167-9457(01)00068-9

[12] E. Bouzbib, G. Bailly, S. Haliyo, and P. Frey. CoVR: A Large-Scale Force-Feedback Robotic Interface for Non-Deterministic Scenarios in VR. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 209–222. ACM, Virtual Event USA, Oct. 2020. doi: 10.1145/3379337.3415891

[13] E. Bouzbib, G. Bailly, S. Haliyo, and P. Frey. "Can I Touch This?": Survey of Virtual Reality Interactions via Haptic Solutions. In *32e Conférence Francophone sur l'Interaction Homme-Machine (IHM '20.21), April 13–16, 2021, France*. Virtual Event, France, Apr. 2021.

[14] E. Bouzbib, M. Teyssier, T. Howard, C. Pacchierotti, and A. Lécuyer. Palmex: Adding palmar force-feedback for 3d manipulation with haptic exoskeleton gloves. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–8, 2023. doi: 10.1109/TVCG.2023.3244076

[15] C. Bozzacchi and F. Domini. Lack of depth constancy for grasping movements in both virtual and real environments. *Journal of Neurophysiology*, 114(4):2242–2248, Oct. 2015. doi: 10.1152/jn.00350.2015

[16] S. Bryson. Direct Manipulation in Virtual Reality. In *Visualization Handbook*, pp. 413–430. Elsevier, 2005. doi: 10.1016/B978-012387582-2/50023-X

[17] L.-P. Cheng, E. Ofek, C. Holz, H. Benko, and A. D. Wilson. Sparse Haptic Proxy: Touch Feedback in Virtual Environments Using a General Passive Prop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pp. 3718–3728. ACM Press, Denver, Colorado, USA, 2017. doi: 10.1145/3025453.3025753

[18] I. Choi, H. Culbertson, M. R. Miller, A. Olwal, and S. Follmer. Grabity: A Wearable Haptic Interface for Simulating Weight and Grasping in Virtual Reality. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology - UIST '17*, pp. 119–130. ACM Press, Québec City, QC, Canada, 2017. doi: 10.1145/3126594.3126599

[19] I. Choi, E. W. Hawkes, D. L. Christensen, C. J. Ploch, and S. Follmer. Wolverine: A wearable haptic interface for grasping in virtual reality. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 986–993. IEEE, Daejeon, South Korea, Oct. 2016. doi: 10.1109/IROS.2016.7759169

[20] A. Clarence, J. Knibbe, M. Cordeil, and M. Wybrow. Unscripted Retargeting: Reach Prediction for Haptic Retargeting in Virtual Reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 150–159. IEEE, Lisboa, Portugal, Mar. 2021. doi: 10.1109/VR50410.2021.00036

[21] L. Cohen, M. Chetouani, S. Régnier, and S. Haliyo. A natural interface based on intention prediction for semi-autonomous micromanipulation. *Journal on Multimodal User Interfaces*, 12(1):17–30, Mar. 2018. doi: 10.1007/s12193-018-0259-1

[22] E. Corona, A. Pumarola, G. Alenya, F. Moreno-Noguer, and G. Rogez. GanHand: Predicting Human Grasp Affordances in Multi-Object Scenes. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5030–5040. IEEE, Seattle, WA, USA, June 2020. doi: 10.1109/CVPR42600.2020.00508

[23] M. R. Cutkosky. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on Robotics and Automation*, 5(3):269–279, June 1989. doi: 10.1109/70.34763

[24] M. R. Cutkosky and R. D. Howe. Human Grasp Choice and Robotic Grasp Analysis. In S. T. Venkataraman and T. Iberall, eds., *Dextrous Robot Hands*, pp. 5–31. Springer New York, New York, NY, 1990.

[25] R. de Souza, S. El-Khoury, J. Santos-Victor, and A. Billard. Recognizing the grasp intention from human demonstration. *Robotics and Autonomous Systems*, 74:108–121, Dec. 2015. doi: 10.1016/j.robot.2015.07.006

[26] X. de Tinguy, T. Howard, C. Pacchierotti, M. Marchal, and A. Lécuyer. WeATaViX: WEarable Actuated TAngibles for VIrtual reality eXperiences. p. 9, 2020.

[27] O. Dermy, F. Charpillet, and S. Ivaldi. Multi-modal Intention Prediction with Probabilistic Movement Primitives. In F. Ficuciello, F. Ruggiero, and A. Finzi, eds., *Human Friendly Robotics*, vol. 7, pp. 181–196. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-319-89327-3_14

[28] A. Edsinger. *Human-Robot Interaction for Cooperative Manipulation: Handing Objects to One Another*. 2007.

[29] J. M. Elliott and K. J. Connolly. A Classification of Manipulative Hand Movements. *Developmental Medicine & Child Neurology*, 26(3):283–296, 1984. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8749.1984.tb04445.x. doi: 10.1111/j.1469-8749.1984.tb04445.x

[30] C. Fang, Y. Zhang, M. Dworman, and C. Harrison. Wireality: Enabling Complex Tangible Geometries in Virtual Reality with Worn Multi-String Haptics. p. 10, 2020.

[31] T. Feix, J. Romero, H.-B. Schmiedmayer, A. M. Dollar, and D. Kragic. The GRASP Taxonomy of Human Grasp Types. *IEEE Transactions on Human-Machine Systems*, 46(1):66–77, Feb. 2016. doi: 10.1109/THMS.2015.2470657

[32] S. Follmer, D. Leithinger, A. Olwal, A. Hogge, and H. Ishii. inFORM: dynamic physical affordances and constraints through shape and object actuation. In *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*, pp. 417–426. ACM Press, St. Andrews, Scotland, United Kingdom, 2013. doi: 10.1145/2501988.2502032

[33] R. Gilster, C. Hesse, and H. Deubel. Contact points during multidigit grasping of geometric objects. *Experimental Brain Research*, 217(1):137–151, Mar. 2012. doi: 10.1007/s00221-011-2980-9

[34] E. J. Gonzalez, P. Abtahi, and S. Follmer. REACH+: Extending the Reachability of Encountered-type Haptics Devices through Dynamic Redirection in VR. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 236–248. ACM, Virtual Event USA, Oct. 2020. doi: 10.1145/3379337.3415870

[35] F. Gonzalez, F. Gosselin, and W. Bachta. A framework for the classification

of dexterous haptic interfaces based on the identification of the most frequently used hand contact areas. In *2013 World Haptics Conference (WHC)*, pp. 461–466. IEEE, Daejeon, Apr. 2013. doi: 10.1109/WHC.2013.6548452
[36] Z. He, F. Zhu, and K. Perlin. PhyShare: Sharing Physical Interaction in Virtual Reality. *arXiv:1708.04139 [cs]*, Aug. 2017. arXiv: 1708.04139.
[37] A. Hettiarachchi and D. Wigdor. Annexing Reality: Enabling Opportunistic Use of Everyday Objects as Tangible Proxies in Augmented Reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pp. 1957–1967. ACM Press, Santa Clara, California, USA, 2016. doi: 10.1145/2858036.2858134
[38] K. Hinckley, R. Pausch, D. Proffitt, and N. F. Kassell. Two-handed virtual manipulation. *ACM Transactions on Computer-Human Interaction*, 5(3):260–302, Sept. 1998. doi: 10.1145/292834.292849
[39] D. Holz, S. Ullrich, M. Wolter, and T. Kuhlen. Multi-Contact Grasp Interaction for Virtual Environments. *JVRB - Journal of Virtual Reality and Broadcasting*, (7), 2008. doi: 10.20385/1860-2037/5.2008.7
[40] T. Iberall. *Grasp Planning for Human Prehension*. 1987.
[41] T. Iberall. Human Prehension and Dexterous Robot Hands. *The International Journal of Robotics Research*, 16(3):285–299, June 1997. Publisher: SAGE Publications Ltd STM. doi: 10.1177/027836499701600302
[42] B. E. Insko. Passive Haptics Significantly Enhances Virtual Environments. p. 111, 2001.
[43] M. Jeannerod. The Timing of Natural Prehension Movements. *Journal of Motor Behavior*, 16(3):235–254, Sept. 1984. Publisher: Routledge _eprint: https://doi.org/10.1080/00222895.1984.10735319. doi: 10.1080/00222895.1984.10735319
[44] N. Kamakura, M. Matsuo, H. Ishii, F. Mitsuboshi, and Y. Miura. Patterns of Static Prehension in Normal Hands. *American Journal of Occupational Therapy*, 34(7):437–445, July 1980. Publisher: American Occupational Therapy Association. doi: 10.5014/ajot.34.7.437
[45] K. Karunratanakul, J. Yang, Y. Zhang, M. Black, K. Muandet, and S. Tang. Grasping Field: Learning Implicit Representations for Human Grasps. *arXiv:2008.04451 [cs]*, Aug. 2020. arXiv: 2008.04451.
[46] K. Khokar, R. Alqasemi, S. Sarkar, K. Reed, and R. Dubey. A novel telerobotic method for human-in-the-loop assisted grasping based on intention recognition. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4762–4769. IEEE, Hong Kong, China, May 2014. doi: 10.1109/ICRA.2014.6907556
[47] Y. Kim, H. J. Kim, and Y. J. Kim. Encountered-type haptic display for large VR environment using per-plane reachability maps: Encountered-type Haptic Display for Large VR Environment. *Computer Animation and Virtual Worlds*, 29(3-4):e1814, May 2018. doi: 10.1002/cav.1814
[48] L. Kohli. Redirected touching: Warping space to remap passive haptics. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 129–130. IEEE, Waltham, MA, USA, Mar. 2010. doi: 10.1109/3DUI.2010.5444703
[49] P. Kratzer, N. B. Midlagajni, M. Toussaint, and J. Mainprice. Anticipating Human Intention for Full-Body Motion Prediction in Object Grasping and Placing Tasks. *arXiv:2007.10038 [cs]*, July 2020. arXiv: 2007.10038.
[50] S. J. Lederman and R. L. Klatzky. Hand movements: A window into haptic object recognition. *Cognitive Psychology*, 19(3):342–368, July 1987. doi: 10.1016/0010-0285(87)90008-9
[51] Z. Li, D. Akkil, and R. Roope. Gaze-Based Kinesthetic Interaction for Virtual Reality, 2020.
[52] L. Makin, G. Barnaby, and A. Roudaut. Tactile and kinesthetic feedbacks improve distance perception in virtual reality. In *Proceedings of the 31st Conference on l'Interaction Homme-Machine - IHM '19*, pp. 1–9. ACM Press, Grenoble, France, 2019. doi: 10.1145/3366550.3372248
[53] W. A. McNeely. Robotic graphics: a new approach to force feedback for virtual reality. In *Proceedings of IEEE Virtual Reality Annual International Symposium*, pp. 336–341, Sept. 1993. doi: 10.1109/VRAIS.1993.380761
[54] A. Miller and P. Allen. GraspIt! A Versatile Simulator for Robotic Grasping. *IEEE Robotics & Automation Magazine*, 2004.
[55] M. R. Mine. ISAAC : A Virtual Environment Tool for the Interactive Construction of Virtual Worlds. p. 11, 1995.
[56] J. Napier. The Prehensile Movement of The Human Hand. *The Journal of Bone and Point Surgery*, 1956.
[57] N. C. Nilsson, A. Zenner, and A. L. Simeone. Propping up Virtual Reality with Haptic Proxies. *IEEE Computer Graphics and Applications*, p. 10, 2021.
[58] I. Poupyrev and T. Ichikawa. Manipulating Objects in Virtual Worlds: Categorization and Empirical Evaluation of Interaction Techniques. *Journal of Visual Languages & Computing*, 10(1):19–35, Feb. 1999. doi: 10.1006/jvlc.1998.0112
[59] K. Shigeta, Y. Sato, and Y. Yokokohji. Motion Planning of Encountered-type Haptic Device for Multiple Fingertips Based on Minimum Distance Point Information. In *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)*, pp. 188–193. IEEE, Tsukaba, Mar. 2007. doi: 10.1109/WHC.2007.85
[60] A. L. Simeone, E. Velloso, and H. Gellersen. Substitutional Reality: Using the Physical Environment to Design Virtual Reality Experiences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pp. 3307–3316. ACM Press, Seoul, Republic of Korea, 2015. doi: 10.1145/2702123.2702389
[61] D. Song, N. Kyriazis, I. Oikonomidis, C. Papazov, A. Argyros, D. Burschka, and D. Kragic. Predicting human intention in visual observations of hand/object interactions. In *2013 IEEE International Conference on Robotics and Automation*, pp. 1608–1615. IEEE, Karlsruhe, Germany, May 2013. doi: 10.1109/ICRA.2013.6630785
[62] N. Stefanov, A. Peer, and M. Buss. Online intention recognition for computer-assisted teleoperation. In *2010 IEEE International Conference on Robotics and Automation*, pp. 5334–5339. IEEE, Anchorage, AK, May 2010. doi: 10.1109/ROBOT.2010.5509432
[63] K. Stockmeier, H. Horton, and V. H. Franz. How do we grasp (virtual) objects in three-dimensional space? *Journal of Vision*, 3(9):383–383, 2003. doi: 10.1167/3.9.383
[64] O. Taheri, N. Ghorbani, M. J. Black, and D. Tzionas. GRAB: A Dataset of Whole-Body Human Grasping of Objects. p. 19, 2020.
[65] Y. Tanaka, K. Kanari, and M. Sato. Interaction with virtual objects through eye-tracking. In *International Workshop on Advanced Imaging Technology (IWAIT) 2021*, vol. 11766, p. 1176624. International Society for Optics and Photonics, Mar. 2021. doi: 10.1117/12.2590989
[66] S.-Y. Teng, T.-S. Kuo, C. Wang, C.-h. Chiang, D.-Y. Huang, L. Chan, and B.-Y. Chen. PuPoP: Pop-up Prop on Palm for Virtual Reality. In *The 31st Annual ACM Symposium on User Interface Software and Technology - UIST '18*, pp. 5–17. ACM Press, Berlin, Germany, 2018. doi: 10.1145/3242587.3242628
[67] D. Valkov, P. Kockwelp, F. Daiber, and A. Krüger. Reach prediction using finger motion dynamics. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544549.3585773
[68] W. Wang, R. Li, Y. Chen, and Y. Jia. Human Intention Prediction in Human-Robot Collaborative Tasks. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, pp. 279–280. ACM Press, Chicago, IL, USA, 2018. doi: 10.1145/3173386.3177025
[69] M. Weise, R. Zender, and U. Lucke. How Can I Grab That?: Solving Issues of Interaction in VR by Choosing Suitable Selection and Manipulation Techniques. *i-com*, 19(2):67–85, Aug. 2020. Publisher: De Gruyter Section: i-com. doi: 10.1515/icom-2020-0011
[70] A. Wexelblat. Virtual reality: applications and explorations, 1993. Myron Krueger, Artificial reality 2 An easy entry to Virtual reality Chap 7.
[71] C. Wilkes and D. A. Bowman. Advantages of velocity-based scaling for distant 3D manipulation. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology - VRST '08*, p. 23. ACM Press, Bordeaux, France, 2008. doi: 10.1145/1450579.1450585
[72] K. Xuel, P. Di, H. Wang, and F. Zou. Predicting human trajectory with virtual HRI environment. In *2018 11th International Workshop on Human Friendly Robotics (HFR)*, pp. 19–24. IEEE, Shenzhen, China, Nov. 2018. doi: 10.1109/HFR.2018.8633515
[73] J. J. Yang, H. Horii, A. Thayer, and R. Ballagas. VR Grabbers: Ungrounded Haptic Retargeting for Precision Grabbing Tools. In *The 31st Annual ACM Symposium on User Interface Software and Technology - UIST '18*, pp. 889–899. ACM Press, Berlin, Germany, 2018. doi: 10.1145/3242587.3242643
[74] Y. Yokokohji, R. L. Hollis, and T. Kanade. WYSIWYF Display: A Visual/Haptic Interface to Virtual Environment. *Presence: Teleoperators and Virtual Environments*, 8(4):412–434, Aug. 1999. doi: 10.1162/105474699566314
[75] Y. Yokokohji, J. Kinoshita, and T. Yoshikawa. Path planning for encountered-type haptic devices that render multiple objects in 3D space. In *Proceedings IEEE Virtual Reality 2001*, pp. 271–278, Mar. 2001. doi: 10.1109/VR.2001.913796
[76] Y. Yokokohji, N. Muramori, Y. Sato, and T. Yoshikawa. *Haptic Display for*

Multiple Fingertip Contacts Based on the Observation of Human Grasping Behaviors. 2005.

[77] S. Yoshida, Y. Sun, and H. Kuzuoka. PoCoPo: Handheld Pin-based Shape Display for Haptic Rendering in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. ACM, Honolulu HI USA, Apr. 2020. doi: 10.1145/3313831.3376358

[78] Y. Zhao, L. H. Kim, Y. Wang, M. Le Goc, and S. Follmer. Robotic Assembly of Haptic Proxy Objects for TangibleInteraction and Virtual Reality. In *Proceedings of the Interactive Surfaces and Spaces on ZZZ - ISS '17*, pp. 82–91. ACM Press, Brighton, United Kingdom, 2017. doi: 10.1145/3132272.3134143